응용통계연구 제 7권 2호
1994년 9월 pp. 53-77

On the Analysis of Multistate Survival Data using Cox's Regression Model[1)]

Sung-chil Yeo[2)]

Abstract

In a certain stochastic process, Cox's regression model is used to analyze multistate survival data. From this model, the regression parameter vectors, survival functions, and the probability of being in response function are estimated based on multistate Cox's partial likelihood and nonparametric likelihood methods. The asymptotic properties of these estimators are described informally through the counting process approach. An example is given to illustrate the results in this paper.

1. Introduction

In clinical trials, we 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, in addition to death, we may also be interested in the fact that a patient has reached a specific illness state and the amount of time spent in that state. The statistical analysis of such resulting data is called multistate survival analysis.

Over the last few decades, various kinds of stochastic models have been proposed for analyzing multistate survival data. For example, Lagakos(1976, 1977) applied a homogeneous Markov model to analyze survival data in the presence of auxiliarly 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. Lagakos et al.(1978) suggested a nonparametric likelihood method for the analysis of partially censored data based on a semi-Markov process model. 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. 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. Hsieh et al.(1983) studied some extended nonparametric test statistics related to the log rank test based on semi-Markov model.

1) This research was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1992.

2) Department of Applied Statistics, Kon-Kuk University, Seoul, 133-701, KOREA.

Voelkel and Crowley(1984) applied a counting process approach to certain hierarchical semi-Markov processes and obtained some useful asymptotic results for the PBRF. Yeo(1989) proposed an extended model which combines the non-homogeneous Markov and semi-Markov models and obtained the generalized maximum likelihood estimators of the transition rates under this model. Recently, Pepe(1991) developed some nonparametric inference for events with dependent risks in multiple endpoint studies.

On the other hand, for a simple two state survival data model, say alive and dead, Cox(1972) suggested a regression model(so-called the proportional hazards model) to study the effects of explanatory variables(or covariates) upon which the individual's survival time may depend. Cox suggested a method of estimating the regression parameters in the absence of knowledge of the unspecified underlying hazard function and also then estimating the underlying hazard function. Since Cox's original paper, many others have contributed to this model.(e.g., Kalbfleisch and Prentice(1973), Breslow(1974), Cox(1975), Efron(1977), Tsiatis(1981), etc.) On the other hand, Aalen(1975,1978) introduced the multiplicative intensity model for counting processes and showed how this model provides a general framework for analyzing data on events observed over time. His approach has been proved remarkably successful in yielding important results about statistical methods for many problems arising in censored data.(e.g., Aalen and Johansen(1978), Gill(1980), Andersen et al.(1982), Andersen and Gill(1982), Ramlau-Hansen(1983), etc.) This counting process approach, which relies heavily on modern theory of martingales and stochastic integrals, also has the considerable advantage of providing straightforward, but rigorous, proofs for the distributional properties of the various estimators and test statistics under very general censoring patterns.

In this paper, we employ Cox's regression model to the non-homogeneous Markov model for analyzing multistate survival data. However, in order to simplify ideas, we define our model in a four-state space given by Temkin(1978).

In a four-state space, each patient who starts in the inital state, say 0, may be assigned at random to receive a drug treatment. Upon receiving this treatment, each patient may entered the response state, say 1, a transient state showing a certain improvement, or may entered the progressive state, say 2, an absorbing state deteriorating or dying without showing any improvement or may entered the relapse state, say 3, an absorbing state which can be reached only from the response. We assume that for each patient one of these transitions would occur with probability one, but that patient may be censored before the transition occus. We also assume that no patients can be in the response state after either progression or relapse. Thus, the only possible direct transitions are $0 \rightarrow 1$, $0 \rightarrow 2$, $1 \rightarrow 3$. Figure 1 exhibits this four-state space.

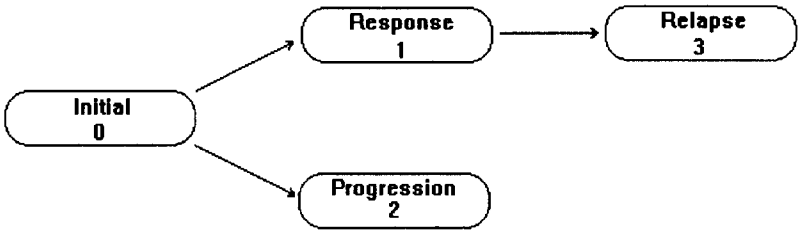


Figure 1. Four-state space

Let T_1, T_2, and T_3 be the transition times from state 0 to 1, 0 to 2, 1 to 3, respectively. And let h_1, h_2, and h_3 be the transition rates (or hazard rates) from state 0 to 1, 0 to 2, 1 to 3, respectively. Then, for the case of continuous random variables h_i $(i=1, 2, 3)$ are defined as follows:

$$h_i(t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(t \le T_i < t+\Delta t \mid T_i \ge t), \quad 0 \le t < \infty, \quad i = 1, 2 \qquad (1.1)$$

and

$$h_3(t \mid s) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(t \le T_3 < t+\Delta t \mid T_3 \ge t, T_1 = s), \quad 0 \le s < t < \infty. \qquad (1.2)$$

That is, $h_i(t)$ $(i=1, 2)$ are the instantaneous rates of transitions to state i at time t and $h_3(t \mid s)$ is the instantaneous rate of transition to state 3 at time t, conditional on having entered state 1 at time s. In this four-state stochastic process, we may be interested in the distribution of time spent in each state i before next transition if censoring were eliminated. In general, this distribution may depend on the transition time to each state, however, certain specific assumptions may be given. In particuler, homogeneous Markov, non-homogeneous Markov, and semi-Markov models can be described in terms of hazard rates, completely analogous to the transition probabilities in a more general stochastic process. If h_i $(i=1, 2, 3)$ are constants, then the model is called the homogeneous Markov model. If $h_3(t \mid s) = h_3(t)$, a function of t only, then the model is called the non-homogeneous Markov model. If $h_3(t \mid s) = h_3(t-s)$, then the model is called the semi-Markov model.

Now, in this paper, we apply Cox's regression model to the non-homogeneous Markov model for the analysis of multistate survival data. In our model which we would call the Markov regression model, the hazard rates h_i $(i=1, 2, 3)$ are defined as follows:

$$h_i(t; \underline{z}(t)) = h_{0i}(t) \exp(\underline{\beta}_i' \underline{z}(t)), \quad i = 1, 2 \qquad (1.3)$$

and

$$h_3(t|s\,;\underline{z}(t)) = h_{03}(t|s)\exp(\underline{\beta}_3'\underline{z}(t)) = h_{03}(t)\exp(\underline{\beta}_3'\underline{z}(t)), \tag{1.4}$$

where $\underline{z}(t) = (z_1(t), \cdots, z_p(t))$ is the vector of, possibly time-dependent, covariates associated with each patient, and $\underline{\beta}_i = (\beta_{i1}, \cdots, \beta_{ip})'$ $(i=1, 2, 3)$ are the vectors of cause-specific regression parameters, and $h_{0i}(\cdot)$ $(i=1, 2, 3)$ are the unspecified underlying hazard functions for each patient with covariate $\underline{z}(\cdot) = \underline{0}$. We note that it is convenient to let p, the number of covariates, to be the same for all i. This can always be possible, if necessary, by introducing extra type-specific covariates.

We next describe the plan of this paper. In Section 2, in the Markov regression model we estimate the regression parameter vectors and the underlying cumulative hazard functions or survival functions based on Multistate Cox's partial likelihood and multistate nonparametric likelihood methods, respectively. We then estimate the PBRF. In Section 3, we describe informally the large sample properties of the estimators given in Section 2 through the counting process approach. In Section 4, we illustrate our results to the real data obtained from a certain clinical trial. Finally, in Section 5, we give some concluding remarks.

2. Multistate likelihoods

2.1 Multistate Cox's partial likelihood

Suppose that there are n individuals in the initial state at time 0, the time to start for study. Let t_{ij} $(i=1, 2, 3\,;\, j=1, 2, \cdots, n)$ be the ordered transition time to state i for the jth individual. As usual we do not observe all t_{ij}'s, but possibly censored transition times $\tilde{t}_{ij}$ $(i=1, 2, 3;\, j=1, 2, \cdots, n)$ and censoring indicators $\delta_{ij} = I(\tilde{t}_{ij} = t_{ij})$, where $I(A)$ denotes the indicator function of a set A. Then, by the analogous arguments used for Cox's partial likelihood in the continuous case, under the non-homogeneous Markov model the multistate Cox's partial likelihood is of the form

$$L(\underline{\beta}_1, \underline{\beta}_2, \underline{\beta}_3) = \prod_{i=1}^{3}\prod_{j=1}^{n}\left(\frac{\exp(\underline{z}_j(\tilde{t}_{ij}))}{\sum_{l\in R_i(\tilde{t}_{ij})}\exp(\underline{\beta}_3'\underline{z}_l(\tilde{t}_{ij}))}\right)^{\delta_{ij}}, \tag{2.1}$$

where $R_i(t) = \{\, l : \tilde{t}_{il} \geq t \,\}$ $(i=1, 2)$, i.e. the set of individuals who are observed to be

at risk just before time t and have a potential to enter state i, and $R_3(t) = \{ l : \tilde{t}_{3l} \geq t > \tilde{t}_{1l}, \delta_{1l} = 1 \}$, i.e. the set of individuals who are observed to be at risk just before time t and have a potential to enter state 3, but already responded before time t.

We note that the likelihood (2.1) can be factored for each i. Thus, each $\underline{\beta}_i$ can be estimated separately. One of the advantages of the multistate survival data models is that formally we can analyze each transition separately and treat other types of transitions as censoring. As a result, it can be handled as in the usual survival data. On the other hand, on the continuous variables case, survival data frequently include ties because of rounding off or grouping. If ties are prestnt in the observed transition times t_{ij}, then the likelihood (2.1) may be modified as follows;

$$L(\underline{\beta}_1, \underline{\beta}_2, \underline{\beta}_3) = \prod_{i=1}^{3} \prod_{j=1}^{k_i} \frac{\exp\left(\sum_{l \in D_{ij}} \underline{\beta}_i' \underline{z}_l(t_{ij}) \right)}{\left[\sum_{l \in R_{ij}} \exp(\underline{\beta}_i' \underline{z}_l(t_{ij})) \right]^{d_{ij}}}, \tag{2.2}$$

where k_i is the total number of individuals who are observed to enter state i, and $R_{ij} = R_i(t_{ij})$, and D_{ij} is the set of individuals who are observed to transit to state i at time t_{ij}, and d_{ij} is the number of individuals in the set D_{ij}. If there are no ties, then all $d_{ij} = |D_{ij}| = 1$, and the modified likelihood (2.2) reduces to (2.1).

From (2.2), the log likelihood is given by

$$\log L(\underline{\beta}_1, \underline{\beta}_2, \underline{\beta}_3) = \sum_{i=1}^{3} \log L_i(\underline{\beta}_i) , \tag{2.3}$$

where

$$\log L_i(\underline{\beta}_i) = \sum_{j=1}^{k_i} \left[\sum_{l \in D_{ij}} \underline{\beta}_i' \underline{z}_l(t_{ij}) - d_{ij} \log \left\{ \sum_{l \in R_{ij}} \exp(\underline{\beta}_i' \underline{z}_l(t_{ij})) \right\} \right], \; i = 1, 2, 3. \tag{2.4}$$

And the first derivatives of (2.3) with respect to (w.r.t.) $\underline{\beta}_i$ ($i = 1, 2, 3$), respectively give the score vectors

$$\begin{aligned} \underline{U}_i(\underline{\beta}_i) &= \frac{\partial}{\partial \underline{\beta}_i} \log L_i(\underline{\beta}_i) \\ &= \sum_{j=1}^{k_i} \{ \underline{S}_{ij} - d_{ij} \underline{A}_{ij}(\underline{\beta}_i) \}, \; i = 1, 2, 3 , \end{aligned} \tag{2.5}$$

where

$$\underline{S}_{ij} = \sum_{l \in D_{ij}} \underline{z}_l(t_{ij}) \tag{2.6}$$

and

$$\underline{A}_{ij}(\underline{\beta}_i) = \frac{\sum_{l \in R_{ij}} \underline{z}_l(t_{ij}) \exp(\underline{\beta}_i' \underline{z}_l(t_{ij}))}{\sum_{l \in R_{ij}} \exp(\underline{\beta}_i' \underline{z}_l(t_{ij}))} \quad (2.7)$$

Similarly, minus the second derivatives of (2.3) w.r.t. $\underline{\beta}_i$ $(i = 1, 2, 3)$, respectively give the information matrices

$$I_i(\underline{\beta}_i) = -\frac{\partial}{\partial \underline{\beta}_i} \underline{U}_i(\underline{\beta}_i)$$

$$= \sum_{j=1}^{k_i} d_{ij} C_{ij}(\underline{\beta}_i), \quad i=1,2,3, \quad (2.8)$$

where

$$C_{ij}(\underline{\beta}_i) = \frac{\sum_{l \in R_{ij}} \underline{z}_l(t_{ij})^{\otimes 2} \exp(\underline{\beta}_i' \underline{z}_l(t_{ij}))}{\sum_{l \in R_{ij}} \exp(\underline{\beta}_i' \underline{z}_l(t_{ij}))} - A_{ij}(\underline{\beta}_i)^{\otimes 2}, \quad (2.9)$$

and where $X^{\otimes 2} = \underline{X}\,\underline{X}'$ for any vector $\underline{X}$.

The values $\hat{\underline{\beta}}_i$ $(i = 1, 2, 3)$, called the maximum partial likelihood estimates (MPLE's), that maximize (2.1) or (2.2) can usually be obtained by the Newton-Rapson method as the solutions of $\underline{U}_i(\underline{\beta}_i) = \underline{0}$. In order to make inferences about $\underline{\beta}_i$, we mainly have to rely on large sample procedures. In the next section, we sketch that under mild conditions on the covariates and censoring, $\hat{\underline{\beta}}_i$ $(i=1,2,3)$ have the asymptotic normalities with means $\underline{\beta}_{i0}$ and covariance matrices $I_i^{-1}(\underline{\beta}_{i0})$, where $\underline{\beta}_{i0}$ is the vector of true values of $\underline{\beta}_i$. Inferences can also be based on the score vectors $U_i(\underline{\beta}_{i0})$ which are asymptotically normal with means $\underline{0}$ and covariance matrices $I_i(\underline{\beta}_{i0})$. For estimation of $I_i(\underline{\beta}_{i0})$ or $I_i^{-1}(\underline{\beta}_{i0})$, we replace $\underline{\beta}_{i0}$ by $\hat{\underline{\beta}}_i$. The third possiblity is to use the likelihood ratio method. For example, if we want to test $H_0 : \underline{\beta}_i = \underline{0}$ for all $i = 1, 2, 3$, then the above three methods give the following test statistics

$$W_1 = \sum_{i=1}^{3} \hat{\underline{\beta}}_i' I_i(\hat{\underline{\beta}}_i) \hat{\underline{\beta}}_i, \quad (2.10)$$

$$W_2 = \sum_{i=1}^{3} U_i(\underline{0})' I_i^{-1}(\underline{0}) U_i(\underline{0}), \quad (2.11)$$

and

$$W_3 = -2 \sum_{i=1}^{3} \log \left(\frac{L_i(\underline{0})}{L_i(\hat{\underline{\beta}}_i)} \right), \quad (2.12)$$

respectively. These three test statistics have the same asymptotic chi-square distribution with 3p degree of freedom.

2.2 Multistate nonparametric likelihood

In Section 1, the Markov regression model was described for the continuous variables case. However, as Johansen (1978) pointed out, if we define the model only for the continuous case, then the nonparametric maximum likelihood estimates (NMLE's) of the underlying survival functions do not exist. Thus, in order to obtain the NMLE's of the underlying survival functions, We rewrite our model given in (1.3) and (1.4) in terms of survival functions so that the model includes both continuous and discrete cases. That is, we define the Markov regression model as

$$S_i(t\,;\underline{z}(t)) = S_{0i}(t)^{\exp(\underline{\beta}_i'\underline{z}(t))}, \quad i = 1,2 \tag{2.13}$$

$$S_3(t|s\,;\underline{z}(t)) = S_{03}(t|s)^{\exp(\underline{\beta}_3'\underline{z}(t))} = \left(\frac{S_{03}(t)}{S_{03}(s)}\right)^{\exp(\underline{\beta}_3'\underline{z}(t))}, \tag{2.14}$$

where $S_{0i}(t) = P(T_i > t\,;\underline{z}(t) = \underline{0})$ $(i = 1,2,3)$ are the underlying survival functions corresponding to the transition times to state i. For the discrete case, (2.13) and (2,14) can be expressed, in terms of hazard functions, as

$$h_i(t\,;\underline{z}(t)) = 1-(1-h_{0i}(t))^{\exp(\beta_i'\underline{z}(t))}, \qquad i=1,2 \tag{2.15}$$

and

$$h_3(t\,|s\,;\underline{z}(t)) = 1-(1-h_{03}(t))^{\exp(\beta_3'\underline{z}(t))}. \tag{2.16}$$

We now proceed to find the NMLE's of $S_{0i}(t)$ $(i = 1,2,3)$ as follows : As with the same notations given in Section 2.2, let $t_{i1} < \cdots < t_{ik_i}$ be the ordered observed transiton times to state i, and define $t_{i0} = 0$, $t_{i,k_i+1} = \infty$ for convenience. And let C_{ij} be the set of patients with censoring times in $[t_{ij}, t_{i,j+1})$ $(i = 1,2,3;\ j = 0,1,\cdots,k_i)$. The censoring times in C_{ij} are denoted by c_{ijl}. Then, under the assumption of indepent censoring mechanisms (see Kalbfleisch and Prentice's (1980) book p.120) the nonparametric likelihood function for four-state survival data is of the form

$$L = \prod_{i=1}^{2}\prod_{j=1}^{k_i}\left[\prod_{l\in D_{ij}}\{S_i(t_{ij}-0\,;\underline{z}_l) - S_i(t_{ij}\,;\underline{z}_l)\}\prod_{l\in C_{ij}} S_i(c_{ijl}\,;\underline{z}_l)\right]$$

$$\times \prod_{j=1}^{k_3}\left[\prod_{l\in D_{3j}}\{S_3(t_{3j}-0|t_{1j}'\,;\underline{z}_l) - S_3(t_{3j}|t_{1j}'\,;\underline{z}_l)\}\prod_{l\in C_{3j}} S_3(c_{3jl}|t_{1j}'\,;\underline{z}_l)\right] \tag{2.17}$$

where $S_i(t-0) = \lim_{h\to 0^+} S_i(t-h)$, and $\underline{z}_l = \underline{z}_l(t_{ij})$ for each time t_{ij}. t_{1j}' $(j = 1,2,\cdots,k_3)$ is the response time of the jth individual corresponding to t_{3j} or c_{3jl}.

If we substitute (2.13) and (2.14) into (2.17), then we have

$$L = \prod_{i=1}^{2}\prod_{j=1}^{k_i}\left[\prod_{l\in D_{ij}}\{S_{0i}(t_{ij}-0)^{\exp(\underline{\beta}_i'\underline{z}_l)} - S_{0i}(t_{ij})^{\exp(\underline{\beta}_i'\underline{z}_l)}\}\prod_{l\in C_{ij}} S_{0i}(c_{ijl})^{\exp(\underline{\beta}_i'\underline{z}_l)}\right]$$

$$\times \prod_{j=1}^{k_3}\left[\prod_{l\in D_{3j}}\left\{\frac{S_{03}(t_{3j}-0)^{\exp(\underline{\beta}_3'\underline{z}_l)} - S_{03}(t_{3j})^{\exp(\underline{\beta}_3'\underline{z}_l)}}{S_{03}(t_{1j}')^{\exp(\underline{\beta}_3'\underline{z}_l)}}\right\}\right.$$

$$\left.\times \prod_{l\in C_{3j}}\left\{\frac{S_{03}(c_{3jl})}{S_{03}(t_{1j}')}\right\}^{\exp(\underline{\beta}_3'\underline{z}_l)}\right]. \qquad (2.18)$$

In the likelihood function L, we note that each product term involves only each one of the functions $S_{0i}(\cdot)$ and $\underline{\beta}_i$ $(i = 1, 2, 3)$, respectively. Since we assume no interrelationships among $S_{0i}(\cdot)$ and $\underline{\beta}_i$'s, L can be maximized by maximizing each product term. Thus, we can write L in (2.18) as

$$L = \prod_{i=1}^{3} L_i \qquad (2.19)$$

where L_i $(i = 1, 2, 3)$ are the product terms corresponding to state i in (2.18).

Now in order to find NMLE's of $S_{0i}(\cdot)$, we employ the same approach given by Kalbfleisch and Prentice (1973). Since $S_{0i}(t)$ is one minus the distribution function, it is nonincreasing and right continuous, and thus as with the Kaplan-Meier estimate, each L_i is maximized by taking $S_{0i}(t) = S_{0i}(t_{i,j+1}-0)$ for $t_{ij} \le t < t_{i,j+1}$ and allowing probability mass to fall only at the observed transition times $t_{i1}, \cdots, t_{i,k_i}$. These observations lead to the consideration of a discrete model with hazard contribution h_{ij} $(=h_i(t_{ij}))$ at time t_{ij} $(j = 1, \cdots, k_i)$. Thus, we take

$$S_{oi}(t) = \prod_{j:t_{ij}\le t}(1-h_{ij}), \quad i = 1, 2, 3. \qquad (2.20)$$

If we denote $\alpha_{ij}(=\alpha_i(t_{ij})) = 1-h_{ij}$, then we have

$$S_{oi}(t_{ij}-0) = S_{0i}(t_{i,j-1})$$

$$= \prod_{k=1}^{j-1}\alpha_{ik}, \quad j = 1, 2, \cdots, k_i, \qquad (2.21)$$

where $\alpha_{i0} = 1$. If we substitute (2.21) into (2.18) and rearrange terms, then we obtain

$$L_i' = \prod_{j=1}^{k_i}\left[\prod_{l\in D_{ij}}\left(1-\alpha_{ij}^{\exp(\underline{\beta}_i'\underline{z}_l)}\right)\prod_{l\in R_{ij}-D_{ij}}\alpha_{ij}^{\exp(\underline{\beta}_i'\underline{z}_l)}\right]. \qquad (2.22)$$

The estimation of $S_{0i}(\cdot)$ as well as $\underline{\beta}_i$ can be obtained by jointly maximizing (2.22) w.r.t. both α_{ij}'s and $\underline{\beta}_i$. However, due to Kalbfleisch and Prentice (1973), a simpler approach is to take $\underline{\beta}_i = \hat{\underline{\beta}}_i$, the MPLE of $\underline{\beta}_i$ and then to maximize (2.22) w.r.t. α_{ij}'s.

If we differentiate the logarithm of (2.22) w.r.t. each α_{ij} $(j = 1, \cdots, k_i)$ and rearrange terms, then we obtain

$$\sum_{l \in D_{ij}} \frac{\exp(\underline{\beta}_i' \underline{z}_l)}{1-\alpha_{ij}^{\exp(\underline{\beta}_i' \underline{z}_l)}} = \sum_{l \in R_{ij}} \exp(\underline{\beta}_i' \underline{z}_l), \quad i = 1, 2, 3 \;\; ; \;\; j = 1, \cdots, k_i \; . \tag{2.23}$$

If $d_{ij} = |D_{ij}| = 1$, then (2.23) gives a direct solution of the form

$$\hat{\alpha}_{ij} = \left(1 - \frac{\exp(\hat{\underline{\beta}}_i' \underline{z}_j)}{\sum_{l \in R_{ij}} \exp(\hat{\underline{\beta}}_i' \underline{z}_l)}\right)^{\exp(-\hat{\underline{\beta}}_i' \underline{z}_j)} \tag{2.24}$$

$$\approx 1 - \frac{1}{\sum_{l \in R_{ij}} \exp(\hat{\underline{\beta}}_i' \underline{z}_l)} \; . \tag{2.25}$$

And if $d_{ij} > 1$, then (2.23) should be solved iteratively for α_{ij}. By noting that $\hat{\alpha}_{ij}^{\exp(\hat{\underline{\beta}}_i' \underline{z}_l)} \approx 1 + \exp(\hat{\underline{\beta}}_i' \underline{z}_l) \log \alpha_{ij}$ and substituting this into (2.23), we get a good initial value to $\widehat{\alpha_{ij}}$ such that

$$\tilde{\alpha}_{ij} = \exp\left(\frac{-d_{ij}}{\sum_{l \in R_{ij}} \exp(\hat{\underline{\beta}}_i' \underline{z}_l)}\right) \tag{2.26}$$

$$\approx 1 - \frac{d_{ij}}{\sum_{l \in R_{ij}} \exp(\hat{\underline{\beta}}_i' \underline{z}_l)} \; . \tag{2.27}$$

From the above results, the NMLE's of $S_{0i}(t)$ $(i = 1, 2, 3)$ are then

$$\hat{S}_{oi}(t) = \prod_{j : t_{ij} \le t} \hat{\alpha}_{ij} \;\; , \;\; i = 1, 2, 3 \; . \tag{2.28}$$

And from (2.28), the cumulative underlying hazard functions of $H_{0i}(t) = \int_0^t h_{0i}(u)du$ $(i = 1, 2, 3)$ may be estimated by

$$\hat{H}_{0i}(t) = -\log \hat{S}_{0i}(t)$$
$$= \sum_{j:\, t_{ij} \le t} -\log \hat{\alpha}_{ij} \quad . \tag{2.29}$$

In (2.28), we note that when the covariates vector $\underline{z}_l = \underline{0}$ for all individuals, $\hat{S}_{0i}(t)$ $(i = 1, 2, 3)$ reduce to the Kaplan-Meier estimates. On the other hand, from (2.13) and (2.14), the estimated survival functions with the specified covariates vector $\underline{z}_0(t)$ are given by

$$\hat{S}_i(t\,;\underline{z}_0(t)) = \prod_{j:\, t_{ij} \le t} \hat{\alpha}_{ij}^{\exp(\hat{\underline{\beta}}_i \underline{z}_0(t))} \quad , \quad i = 1, 2 \tag{2.30}$$

$$\widehat{S_3}(t|s\,;\underline{z}_0(t)) = \prod_{j:\, s < t_{ij} \le t} \hat{\alpha}_{3j}^{\exp(\hat{\underline{\beta}}_3 \underline{z}_0(t))} \quad . \tag{2.31}$$

In the next section, we discuss briefly the convergence of $\hat{H}_{0i}(t)$ to a Gaussian process which was essentially proved by Anderson and Gill (1982). In specific form, $\sqrt{n}\,(\hat{H}_{0i}(t) - H_{0i}(t))$ $(i = 1, 2, 3)$ at each fixed t are asymptotically normal with means zero and the estimated variances

$$n \sum_{j:\, t_{ij} \le t} \frac{1}{\left(\sum_{l \in R_{ij}} \exp(\hat{\underline{\beta}}_i' \underline{z}_l) \right)^2} + n\, Q_i'(\hat{\underline{\beta}}_i, t)\, I_i^{-1}(\hat{\underline{\beta}}_i)\, Q_i(\hat{\underline{\beta}}_i, t), \quad i = 1, 2, 3, \tag{2.32}$$

where

$$Q_i(\hat{\underline{\beta}}_i, t) = \sum_{j:\, t_{ij} \le t} \frac{\sum_{l \in R_{ij}} \underline{z}_l \exp(\hat{\underline{\beta}}_i' \underline{z}_l)}{\left(\sum_{l \in R_{ij}} \exp(\hat{\underline{\beta}}_i' \underline{z}_l) \right)^2} \tag{2.33}$$

and $\underline{z}_l = \underline{z}_l(t_{ij})$. It follows then that for a specifed vector $\underline{z}_0(t)$, $\hat{H}_i(t\,;\underline{z}_0(t)) = \int_0^t \exp(\hat{\underline{\beta}}_i' \underline{z}_0(u))\, d\hat{H}_{0i}(u)$, $\hat{S}_{0i}(t)$, and $\hat{S}_i(t\,;\underline{z}_0(t)) = \hat{S}_{0i}(t)^{\exp(\hat{\underline{\beta}}_i \underline{z}_0(t))}$ would also converge weakly to Gaussian processes.

Based on the preceding results, we now estimate the PBRF. As mentioned in Section 1, Temkin(1978) introduced the PBRF to provide a unified view of response-related endpoints. For no covariates case, the PBRF is defined as

$$P(t) = P(T_1 \le t < T_4)$$
$$= \int_0^t S_3(t|s)\, S_1(s)\, S_2(s)\, dH_1(s) \tag{2.35}$$

where $S_3(t \mid s) = \exp(-\int_s^t h_3(u|s)du)$, and T_4 is the transition time to state 2 or 3. Under the Markov regression model, the PBRF is given by

$$P(t) = \int_0^t \left(\frac{S_{03}(t)}{S_{03}(s)} \right)^{\exp(\underline{\beta}_3 \underline{z}(t))} S_{01}(s)^{\exp(\underline{\beta}_1 \underline{z}(s))} \times S_{02}(s)^{\exp(\underline{\beta}_2 \underline{z}(s))} \exp(\underline{\beta}_1 \underline{z}(s)) dH_{01}(s). \quad (2.36)$$

In order to estimate the PBRF in (2.36), we can use (2.28) and (2.29) for the estimates of $S_{0i}(t)$ and $H_{0i}(t)$ (i=1,2,3), respectively. Thus, the estimated PBRF is given by

$$\widehat{P}(t) = \int_0^t \left(\frac{\hat{S}_{03}(t)}{\hat{S}_{03}(s)} \right)^{\exp(\hat{\underline{\beta}}_3 \underline{z}(t))} \hat{S}_{01}(s)^{\exp(\hat{\underline{\beta}}_1 \underline{z}(s))} \times \hat{S}_{02}(s)^{\exp(\hat{\underline{\beta}}_2 \underline{z}(s))} \exp(\hat{\underline{\beta}}_1 \underline{z}(s)) d\hat{H}_{01}(s). \quad (2.37)$$

We note that $\widehat{P}(t)$ is a step function with jumps at each observed transition time t_{ij}. Thus, in practice $\widehat{P}(t)$ can be evaluated as follows : For all observed transition times t_{ij} 's (i=1,2,3; j=1,2,3, $\cdots$ k_i), we rearrange these times in ascending order, say $0 \equiv t_0 < t_1 < \cdots < t_k$.

Then for $t \in [t_i, t_{i+1})$, $\widehat{P}(t) = \widehat{P}(t_i)$ and

$$\widehat{P}(t_i) = \sum_{j=1}^{i} \left(\frac{\hat{S}_{03}(t_i)}{\hat{S}_{03}(t_j)} \right)^{\sum_{l \in G_{3j}} \exp(\hat{\underline{\beta}}_3 \underline{z}_l(t_i))} \hat{S}_{01}(t_j)^{\sum_{l \in G_{3j}} \exp(\hat{\underline{\beta}}_1 \underline{z}_l(t_j))} \hat{S}_{02}(t_j)^{\sum_{l \in G_{3j}} \exp(\hat{\underline{\beta}}_2 \underline{z}_l(t_j))} \times \sum_{l \in G_{3j}} \exp(\hat{\underline{\beta}}_1 \underline{z}_l(t_j)) (\hat{H}_{01}(t_j) - \hat{H}_{01}(t_{j-1}) \quad (2.38)$$

or, after a small amount of algebraic manipulation, $\widehat{P}(t_i)$ can be calculated recursively of the form

$$\widehat{P}(t_i) = \frac{\hat{S}_{03}(t_i)^{\sum_{l \in G_{3i}} \exp(\hat{\underline{\beta}}_3 \underline{z}_l(t_i))}}{\hat{S}_{03}(t_{i-1})^{\sum_{l \in G_{3,i-1}} \exp(\hat{\underline{\beta}}_3 \underline{z}_l(t_{i-1}))}} \widehat{P}(t_{i-1}) + \hat{S}_{01}(t_i)^{\sum_{l \in G_{3i}} \exp(\hat{\underline{\beta}}_1 \underline{z}_l(t_i))} \hat{S}_{02}(t_i)^{\sum_{l \in G_{3i}} \exp(\hat{\underline{\beta}}_2 \underline{z}_l(t_i))} \times \sum_{l \in G_{3i}} \exp(\hat{\underline{\beta}}_1 \underline{z}_l(t_i)) (\hat{H}_{01}(t_i) - \hat{H}_{01}(t_{i-1})), \quad (2.39)$$

where G_{ij} is the set of individuals whose transition times to state i are observed to be

t_j, and

$$\hat{S}_{0i}(t_j) = \prod_{k=1}^{j} \hat{\alpha}_i(t_k),\ i=1,2,3,$$

$$\hat{H}_{01}(t_j) = \sum_{k=1}^{j} -\log \hat{\alpha}_1(t_k),$$

and where $\hat{\alpha}_i(t_k)=1$ if $t_k \notin \{t_{ij} : j=1,2,\cdots,k_i\}$.

For the asymptotic property of $\hat{P}(t)$, as mentioned before, we note that all estimated integrand functions of $\hat{P}(t)$ given in (2.37) converge weakly to Gaussian processes. Thus, $\hat{P}(t)$ would also be asymptotically Gaussian.

3. Asymptotic properties

In this section we discuss the asymptotic properties of the estimators obtained from the previous section. Following the lines of Andersen and Gill(1982), the rigorous proofs for the asymptotic properties of these estimators can be given in detail. However due to Gill(1984), we informally sketch how the proofs can be followed by the counting process approach.

For excellent reviews of the counting process methods, see Andersen et al.(1982), Andersen and Borgan(1985), and others. Recently, Fleming and Harrington(1991) and Andersen et al.(1993) wrote the books for the survival analysis based on counting processes, respectively. For the background definitions and results used in this section, we refer to these materials.

3.1 Counting process formulation of the model

For the four-state survival data model described in Section 1, Let $N_{ij}(t)$ $(i=1,2,3\ ;\ j=1,2,\cdots,n)$ count when the jth individual is observed to enter state i in $[0,t]$ in the presence of censoring, i.e. $N_{ij}(t)=I\{\tilde{T}_{ij}\le t,\ \delta_{ij}=1\}$. It is assumed that the individuals behave independently of each other, and that no two transitions occur simultaneously. Then $\underline{N}=\{(N_{ij}(t)\ ;\ i=1,2,3,\quad j=1,2,\cdots,n),\ 0\le t<\infty\}$ is regarded as a 3n-variate counting process, i.e. each N_{ij} is a counting process, and that no two component processes jump at the same time.

Under certain regularity conditions, which need not concern us, this multivariate counting process $\underline{N}$ is governed by its (random) intensity process

$\underline{\lambda} = \{(\lambda_{ij}(t) ; i = 1, 2, 3, j = 1, 2, \cdots, n), 0 \le t < \infty\}$ which is argued as follows : Let I_{dt} be a small time interval of length dt around time t, and let $dN_{ij}(t)$ be the increments of N_{ij} over I_{dt}. If we let F_{t-} represent everything that has happened just before time t, then we have

$$\lambda_{ij}(t)dt = P\{dN_{ij}(t)=1 \mid F_{t-}\} \tag{3.1}$$

here F_{t-} , called the history, includes a complete specification of the paths of $\underline{N}(s)$, $0 \le s < t$, as well as all other events implicitly or explicitly included in the model which have happened up to (but not at) time t.

Under the assumptions of the independent censoring mechanisms and our Markov regression model, the specific form of the above intensity process $\lambda_{ij}(t)$ is given as follows : At any time t, given what has happened just before the time interval I_{dt}, we know that the jth individual has been observed to enter state i, or he has been censored, or he is to be at risk for making a transition to state i $(i = 1, 2, 3)$. For the first two cases, the conditional probability of observing N_{ij} to jump in the interval I_{dt} is zero. For the latter case, this conditional probability is $h_i(t ; \underline{z}_{ij}(t))\,dt = h_{0i}(t)\exp(\underline{\beta}_i' \underline{z}_{ij}(t))$ $(i = 1, 2, 3)$, where $h_3(t|s ; \underline{z}(t)) = h_3(t ; \underline{z}(t))$ is denoted for convenience (s is the transition time to state 1), and we assume that the underlying transition intensities (or hazards) for different individuals are identical.

We define $Y_{ij}(t)=1$ $(i = 1, 2 ; j = 1, 2, \cdots, n)$ if the jth individual is observed just before time t and have a potential to enter state i ; $Y_{ij}(t) = 0$ otherwise, i.e.

$$\begin{aligned} Y_{ij}(t) &= I\{\widetilde{T}_{ij} \ge t\} \\ &= I\{N_{ij}(t-) = 0\} \end{aligned} \tag{3.2}$$

and define $Y_{3j}(t) = 1$ if the jth individual is observed and have a potential to enter state 3, but already entered state 1 just before time t; $Y_{3j}(t) = 1$ otherwise, i.e.

$$\begin{aligned} Y_{3j}(t) &= I\{\widetilde{T}_{3j} \ge t > \widetilde{T}_{1j},\ \delta_{1j} = 1\} \\ &= I\{N_{1j}(t-)=1,\ N_{3j}(t-)=0\}\ . \end{aligned} \tag{3.3}$$

From the above arguments and from (3.1), the counting process $N_{ij}(t)$ has the intensity process $\lambda_{ij}(t)$ w.r.t. F_t, where

$$\lambda_{ij}(t)dt = Y_{ij}(t)\,h_{0i}(t)\exp(\underline{\beta}_i' \underline{z}_{ij}(t))\,dt, \quad i = 1, 2, 3 \ ; \ j = 1, 2, \cdots, n. \tag{3.4}$$

In (3.4), we note that $h_{0i}(t)\exp(\underline{\beta}_i' \underline{z}_{ij}(t))$ is a non-negative deterministic function, while

$Y_{ij}(t)$ is a non-negative observable stochastic process whose value at any time t is known just before time t. we say that a process with these properties is predictable. Formally, if a process is adapted and has left-continuous sample paths, then it is predictable and locally bounded. Moreover, any deterministic process is predictable.

Now, we give an extension of (3.4) of the form

$$\lambda_{ij}(t)dt = Y_{ij}(t)\, h_{0i}(t) \exp(\underline{\beta}_i' \underline{Z}_{ij}(t))\, dt \ , \tag{3.5}$$

here we have replaced the fixed covariates vector $\underline{z}_{ij}(t)$ by the random covariates vector $\underline{Z}_{ij}(t)$. We no longer require that each N_{ij} jumps at most once, nor do we require that each Y_{ij} is of the special form given in (3.2) and (3.3). All we require is that N_{ij}, Y_{ij} and $\underline{Z}_{ij}$ are observable stochastic processes and that Y_{ij} and $\underline{Z}_{ij}$ are predictable. So, Y_{ij} and $\underline{Z}_{ij}$ are indicator and covariate processes, respectively which are fixed given what has happened just before time t, i.e. given F_{t-}, we know the values of $Y_{ij}(t)$ and $\underline{Z}_{ij}(t)$ (but not yet $N_{ij}(t)$ for instance), where F_{t-} is formally given by

$$F_{t-} = \sigma\{N_{ij}(s),\ Y_{ij}(s+),\ \underline{Z}_{ij}(s+)\ ;\ 0 \le s < t,\ i = 1, 2, 3,\ j = 1, 2, \cdots, n\}. \tag{3.6}$$

The above condition is forced on us by the meaning of $\lambda_{ij}(t)$ as the transition intensity with which N_{ij} jumps given F_{t-}. This also restricts Y_{ij} to being non-negative.

In order to find the asymptotic properties of the estimators given in Section 2, we now make a link between the counting processes and martingales. Since the increment $dN_{ij}(t)$ of N_{ij} over I_{dt} is a 0-1 variable, we have, from (3.1)

$$E\{dN_{ij}(t)|F_{t-}\} = \lambda_{ij}(t)dt \ . \tag{3.7}$$

Thus, if we define stochastic processes M_{ij} $(i = 1, 2, 3\ ; j = 1, 2, \cdots, n)$ by having increments

$$dM_{ij}(t) = dN_{ij}(t) - \lambda_{ij}(t)\, dt \tag{3.8}$$

over I_{dt} (and satisfying $M_{ij}(0) = 0$), then

$$E\{dM_{ij}(t)|F_{t-}\} = 0 \ . \tag{3.9}$$

This implies that (3.5) is equivalent to the assertion that M_{ij}'s are martingales, where

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s)\, h_{0i}(s) \exp(\underline{\beta}_i' \underline{Z}_{ij}(s))\, ds \ . \tag{3.10}$$

In particular $E\{M_{ij}(t)\} = 0$ for all $0 \le t < \infty$. The relation (3.10) is the key to the counting process approach to our Markov regression model. In fact, This is known to the Doob-Meyer decomposition theorem of the local submartingale N_{ij}, and by this theorem

$M_{ij}(t)$ of the form (3.10) are local square integrable martingales and hence they are orthogonal, i.e. $<M_{ij}, M_{ik}> = 0$ for $j \neq k$. In general, for two martingales, say M_1, M_2, the predictable covariation process is defined by $<M_1, M_2> = \int_0^t d<M_1, M_2>(s)$, where $d<M_1, M_2>(s) = Cov\{dM_1(s), dM_2(s) \mid F_{t-}\}$ over I_{dt}. For a martingale M, the predictable variation process is defined by having increments $d<M>(t) = Var\{dM(t) \mid F_{t-}\}$ over I_{dt}. For the counting process martingales M_{ij} given in (3.10), the predictable variation processes are given by

$$<M_{ij}>(t) = \int_0^t Y_{ij}(s) h_{0i}(s) \exp(\underline{\beta}_i' Z_{ij}(s)) ds \ . \tag{3.11}$$

3.2 Asymptotic properties of the estimators

In the following we discuss the asymptotic properties of the estimators given in Section 2. The proofs of Andersen and Gill(1982) go through almost unchanged for our model. Therefore, we briefly sketch the ideas.

From the previous arguments, the multistate Cox's partial likelihood (2.1) or (2.2) can be re-presented as

$$L(\underline{\beta}_1, \underline{\beta}_2, \underline{\beta}_3) = \prod_{i=1}^{3} L_i(\underline{\beta}_i) \tag{3.12}$$

where

$$L_i(\underline{\beta}_i) = \prod_{j=1}^{n} \prod_{s \geq 0} \left\{ \frac{Y_{ij}(s) \exp\{\underline{\beta}_i' Z_{ij}(s)\}}{\sum_{k=1}^{n} Y_{ik}(s) \exp\{\underline{\beta}_i' Z_{ik}(s)\}} \right\}^{dN_{ij}(s)} , \tag{3.13}$$

and where the product over s is a product over disjoint intervals. So, (3.13) reduces to a finite product over all j and s for which N_{ij} jumps at time s ($dN_{ij}(s) = 1$) ; elsewhere $dN_{ij}(s) = 0$. We also define $L_i(\underline{\beta}_i, t)$ as the likelihood function in which the product over $s \geq 0$ in (3.13) is replaced by a product over $o \leq s \leq t$. Since the likelihood $L_i(\underline{\beta}_i)$ is based on observations of $\{(N_{ij}, Y_{ij}, Z_{ij}) ; i=1, 2, 3, j=1, 2, \cdots, n\}$ on the interval $[0, \infty)$, we may regard it as the value of the process $L_i(\underline{\beta}_i, t)$ as $t \to \infty$. So, we denote $L_i(\underline{\beta}_i) = L_i(\underline{\beta}_i, \infty)$.

The logarithm of the likelihood $L_i(\underline{\beta}_i, t)$ is of the form

$$\log L_i(\underline{\beta}_i, t)$$

$$= \sum_{j=1}^{n} \int_0^t \underline{\beta}_i' \underline{Z}_{ij}(s)\, dN_{ij}(s) - \int_0^t \log \{ \sum_{k=1}^{n} Y_{ik}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ik}(s)\} \}\, dN_i(s) , \quad (3.14)$$

where $N_i = \sum_{j=1}^{n} N_{ij}$. Since the score vetor $U_i(\underline{\beta}_i)$ is the value of the process $\underline{U}_i(\underline{\beta}_i, t)$ as $t \to \infty$, where

$$\underline{U}_i(\underline{\beta}_i, t) = \frac{\partial}{\partial \underline{\beta}_i} \log L_i(\underline{\beta}_i, t)$$

$$= \sum_{j=1}^{n} \int_0^t \{ \underline{Z}_{ij}(s) - \underline{E}_i(\underline{\beta}_i, s) \}\, dN_{ij}(s) , \quad (3.15)$$

and where

$$\underline{E}_i(\underline{\beta}_i, s) = \frac{\sum_{j=1}^{n} Y_{ij}(s)\, \underline{Z}_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}}{\sum_{j=1}^{n} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}} , \quad (3.16)$$

the estimators $\hat{\underline{\beta}}_i$ $(i=1, 2, 3)$ can be defined as the solutions to equations

$$\underline{U}_i(\underline{\beta}_i, \infty) = \underline{0}, \quad i = 1, 2, 3 . \quad (3.17)$$

In (3.15) each integrand term $\underline{Z}_{ij}(s) - \underline{E}_i(\underline{\beta}_i, s)$ may be thought of as a covariate centered by its empirical average calculated with the probability mass function which assigns a weight proportional to $Y_{ij}(s) \exp\{\beta_i' Z_{ij}(s)\}$ to the jth individual. The average of these centered terms w.r.t. this same discrete probability function is then zero. That is,

$$\frac{\sum_{j=1}^{n} \{ \underline{Z}_{ij}(s) - \underline{E}_i(\underline{\beta}_i, s) \} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}}{\sum_{j=1}^{n} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}} = \underline{0}$$

or equivalently,

$$\sum_{j=1}^{n} \{ \underline{Z}_{ij}(s) - \underline{E}_i(\underline{\beta}_i, s) \} Y_{ij}(s)\, h_{0i}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\} = 0 .$$

This last equation implies that we have $U_i(\underline{\beta}_i, t)$ of the form

$$\underline{U}_i(\underline{\beta}_i, t) = \sum_{j=1}^{n} \int_0^t \underline{H}_{ij}(s)\, dM_{ij}(s) , \quad i = 1, 2, 3, \quad (3.18)$$

where $M_{ij}(s)$ is given by (3.10) and $\underline{H}_{ij}(s) = \underline{Z}_{ij}(s) - \underline{E}_i(\underline{\beta}_i, s)$ is the vector of predicta ble processes (it only depends on the fixed parameter $\underline{\beta}_i$ and the predictable processes

Y_{ij}, $\underline{Z}_{ij}$ $(j = 1, 2, \cdots, n)$.

We note that by the martingale transform theorem, a stochastic integral of a predictable process w.r.t. a martingale is itself a martingale, and that a sum of martingales is also a martingale. Thus, if we write

$$\underline{U}_i(\underline{\beta}_i, t) = \sum_{j=1}^{n} \underline{U}_{ij}(\underline{\beta}_i, t) \ , \tag{3.19}$$

where

$$\begin{aligned} \underline{U}_{ij}(\underline{\beta}_i, t) &= \int_0^t \underline{H}_{ij}(s)\, dM_{ij}(s) \\ &= \int_0^t \{\underline{Z}_{ij}(s) - \underline{E}_i(\underline{\beta}_i, s)\}\, dM_{ij}(s) \ , \end{aligned} \tag{3.20}$$

then $\underline{U}_i(\underline{\beta}_i, t)$ considered as a stochastic process in t, is the sum of n (vector) martingales which is also a martingale.

In (3.19), we note that for each i $(i = 1, 2, 3)$, $\underline{U}_{ij}(\underline{\beta}_i, t)$ $(j = 1, 2, \cdots, n)$ are not independent, due to the term $\underline{E}_i(\underline{\beta}_i, s)$, however $\underline{U}_i(\underline{\beta}_i, t)$ can be seen as a sum of uncorrelated terms from the martingale representation for $\underline{U}_i(\underline{\beta}_i, t)$. Therefore, the martingale central limit theorem can be applied to prove that as $n \to \infty$, the processes $n^{-1/2}\underline{U}_i(\underline{\beta}_i, t)$ $(i=1,2,3)$ are asymptotically distributed as the Gaussian martingales with mean vectors $\underline{o}$ and covariance function matrices $n^{-1}I_i(\underline{\beta}_i, t)$, where

$$\begin{aligned} I_i(\underline{\beta}_i, t) &= -\frac{\partial}{\partial \underline{\beta}_i} U_i(\underline{\beta}_i, t) \\ &= \int_0^t V_i(\underline{\beta}_i, s)\, dN_i(s) \ , \quad i = 1, 2, 3 \end{aligned} \tag{3.21}$$

and where

$$\begin{aligned} V_i(\underline{\beta}_i, s) = &\frac{\sum_{j=1}^{n} Y_{ij}(s)\,(\underline{Z}_{ij}(s))^{\otimes 2} \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}}{\sum_{j=1}^{n} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}} \\ &- \left(\frac{\sum_{j=1}^{n} Y_{ij}(s)\, \underline{Z}_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}}{\sum_{j=1}^{n} \underline{Z}_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}} \right)^{\otimes 2} , \quad i = 1, 2, 3. \end{aligned} \tag{3.22}$$

To prove the asymptotic normality of $\sqrt{n}(\hat{\underline{\beta}}_i - \underline{\beta}_i)$, we use a Taylor expansion of $U_i(\hat{\underline{\beta}}_i, t)$ around $\underline{\beta}_{i0}$, much in the same way as for a standard maximum likelihood theory. As a result, $\sqrt{n}(\hat{\underline{\beta}}_i - \underline{\beta}_{i0})$ $(i=1,2,3)$, as $n \to \infty$, have the asymptotically multivariate normal distributions with mean vectors $\underline{o}$ and covariance matrices $n I_i^{-1}(\underline{\beta}_i)$, where

$I_i(\underline{\beta}_i) = I_i(\underline{\beta}_i, \infty)$. To prove the consistencies of $\hat{\underline{\beta}}_i$ $(i=1,2,3)$, which need to be shown before the above result, we can use Lenglart's inequality and the fact that $\log L_i(\underline{\beta}_i, t)$ is a concave function with a unique maximum at $\underline{\beta}_i = \hat{\underline{\beta}}_i$. Sufficient conditions for the consistency and asymptotic normality of $\hat{\underline{\beta}}_i$ were given by Andersen and Gill(1982).

On the other hand, the estimates $\widehat{H\alpha}(t)$ $(i=1,2,3)$ given in (2.27) can be re-presented as

$$\widehat{\Lambda}_{\alpha}(t) = \int_0^t \frac{I\{Y_i(s)>0\}}{\sum_{j=1}^{n} Y_{ij}(s) \exp\{\hat{\underline{\beta}}_i' Z_{ij}(s)\}} dN_i(s), \quad i=1,2,3, \qquad (3.23)$$

where $Y_i = \sum_{j=1}^{n} Y_{ij}$. Under the same conditions for the asymptotic properties of $\hat{\underline{\beta}}_i$, $\sqrt{n}(\widehat{\Lambda}_{0i}(t) - \Lambda_{0i}(t))$ $(i=1,2,3)$ on $0 \le t \le T$, as $n \to \infty$, are asymptotically distributed as a Gaussian processes with means zero, independent increments, and the estimated variance functions

$$n \int_0^t \frac{dN_i(s)}{\left[\sum_{j=1}^{n} Y_{ij}(s) \exp(\underline{\beta}_i' \underline{Z}_{ij}(s))\right]^2} + H_i'(\underline{\beta}_i, t)\, I_i^{-1}(\underline{\beta}_i, T)\, H_i(\underline{\beta}_i, t), \quad i=1,2,3, \quad (3.24)$$

where the time T is such that $\int_0^T h_{0i}(s)\,ds < \infty$, and

$$H_i(\underline{\beta}_i, t) = -\int_0^t \frac{\sum_{j=1}^{n} Y_{ij}(s)\, \underline{Z}_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}}{\left[\sum_{j=1}^{n} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}\right]^2} dN_i(s), \quad i=1,2,3. \quad (3.25)$$

The proof of this result is as given in Andersen and Gill(1982), and the main step is being to notice that

$$\int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}} - \Lambda_{0i}^*(t), \quad i=1,2,3$$

are local square integrable martingales, say $W_i(t)$ which are orthogonal to $U_i(\underline{\beta}_i, t)$, where

$$\Lambda_{0i}^*(t) = \int_0^t I\{Y_i(s)>0\}\, h_{0i}(s)\, ds, \quad i=1,2,3,$$

$$W_i(t) = \int_0^t \frac{dM_i(s)}{\sum_{j=1}^{n} Y_{ij}(s) \exp\{\underline{\beta}_i' \underline{Z}_{ij}(s)\}}, \quad i=1,2,3,$$

and where

$$M_i = N_i - \Lambda_i, \qquad \Lambda_i = \sum_{j=1}^{n} \Lambda_{ij},$$

$$\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)\, ds .$$

Based on the above results, the PBRF given in (2.36) can be estimated by

$$\hat{P}(t) = \int_0^t \left(\frac{\hat{S}_{03}(t)}{\hat{S}_{03}(s)} \right)^{\exp(\beta_3' Z(t))} \hat{S}_{01}(s)^{\exp(\beta_1' Z(s))}\, \hat{S}_{02}(s)^{\exp(\beta_2' Z(s))} \times \exp(\beta_1' Z(s))\, d\hat{\Lambda}_{01}(s) , \tag{3.26}$$

where

$$\hat{S}_{0i}(t)^{\exp(\beta_i' Z(t))} = \prod_{j=1}^{n} \prod_{s \le t} \left(1 - \frac{1}{\sum_{k=1}^{n} Y_{ik}(s) \exp(\beta_i' Z_{ij}(s))} \right)^{Y_{ij}(s) \exp(\beta_i' Z_{ij}(s))\, dN_{ij}(s)}, \quad i = 1, 2, 3 ,$$

$$d\hat{\Lambda}_{01}(s) = \frac{I\{Y_1(s) > 0\}}{\sum_{j=1}^{n} Y_{1j}(s) \exp(\hat{\beta}_1' Z_{1j}(s))}\, dN_1(s) ,$$

and

$$\exp(\beta_1' Z(s)) = \sum_{j=1}^{n} Y_{1j}(s) \exp(\beta_1' Z_{ij}(s)) .$$

We see that this estimate coincides with the estimate given in (2.37). Following similar methods of Voelkel and Crowley(1984), we can obtain the asymptotic properties of the estimate $\hat{P}(t)$. As a result $\sqrt{n}(\hat{P}(t) - P(t))$ would converge weakly to a mean zero Gaussian process.

4. Example

In this section, we illustrate the results in the previous sections to the real data given in Hsieh(1980). In a clinical trial performed at the Wisconsin Clinical Cancer Center, 135 patients with advanced breast cancer had been entered for the study. A primary goal of this study was to evaluate the relative effectiveness of Adriamycin and Dibromodulcitol, with and without Tamoxifen(denoted by DAT and DA, respectively) for these patients.

In this study, the ECOG(Eastern Cooperative Oncology Group) criteria were used to evaluate response categories of the patients. Patients who have complete response or partial response or improvement were classified as having a beneficial response. Of the 135 patients, 55 and 80 were randomly treated with DA and DAT, respectively. Among these 55 patients, 18 had relapsed after having response, two had responded without relapse, and 35 progressed. Among those 80 patients, 26 had relapsed after having response, 18 had responded without relapse, 35 had progressed, and one had neither progressed nor

responded.

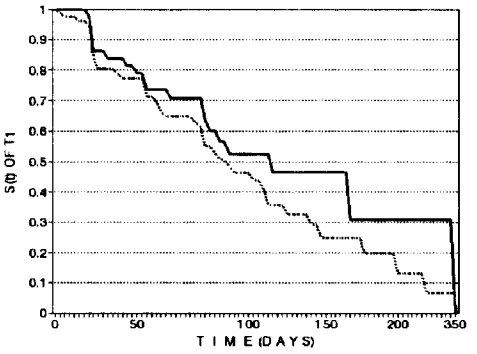
In order to apply the results in this paper, we consider 135 patients with DA and DAT treatments as a single population. So, in the relations (1.3) and (1.4), let the covariate z be a single dummy variable to classify these two groups such as $z = 0$ if patients are treated with DA, and $z = 1$ if patients are treated with DAT. Then (1.3) and (1.4) are simply reduced to $h_i(t;z) = h_{0i}(t)e^{\beta_i z}$ $(i = 1, 2)$ and $h_3(t|s;z) = h_{03}(t)e^{\beta_3 z}$, where $z = 0$ or 1.

For the advanced breast cancer data in Hsieh(1980), solving $U_i(\beta_i) = 0$ given in (2.5), then the estimates of β_i $(i = 1, 2, 3)$ are $\hat{\beta}_1 = -0.0511$, $\hat{\beta}_2 = -0.4197$, and $\hat{\beta}_3 = -0.5755$, respectively. And from (2.8), their estimated standard deviations, i.e. $I_i^{-1/2}(\hat{\beta}_i)$ $(i = 1, 2, 3)$ are 0.3071, 0.2397, and 0.3098, respectively.

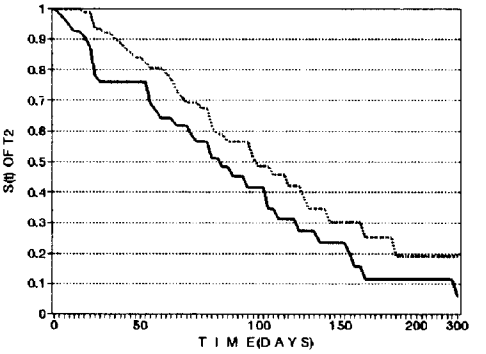
Thus, if we use (2.10) for testing $H_0 : \beta_i = 0$ for all $i = 1, 2, 3$, then we have $W_1 = 6.793$ which gives the one-tailed p-value between 0.05 and 0.1. This implies that the relative effectiveness of the treatment DAT to the treatment DA seems to be not great but slightly better over all sense. In particular, the values of standard normal statistics, i.e. $Z_i = \hat{\beta}_i / I_i^{-1/2}(\hat{\beta}_i)$ $(i = 1, 2, 3)$ are -0.1664, -1.7509, and -1.8577, respectively. And the corresponding p-values for testing $H_0 : \beta_i = 0$ *vs* $H_1 = \beta_i < 0$ for each $i = 1, 2, 3$ are 0.434, 0.04, and 0.032, respectively. From this result, we see that the addition of Tamoxifen to Adriamycin and Dibromodulcitol does not seem to improve the probability of getting a beneficial response, but appears to diminish the probability of reaching progression or relapse.

On the other hand, figures 2.(a)-(c) show the estimated survival functions of the transition times to each state. These figures also support the above conclusions. Specifically, Figure 2.(a) shows that the DA patients have the longer duration times in the response state than the DAT patients. Figure 2.(b) and (c) show that the DAT patients have the larger survival percentages than the DA patients in the progressive and the relapse states, respectively. Figure 3 represents the curve of the estimated PBRF for the patients with the treatments DA and DAT, respectively. From this figure, we see that the steeper rise of the DAT curve reflects the shorter time for response to occur. And the higher DAT curve indicates that the DAT patients comprise the larger percentage of responders than the DA patients.

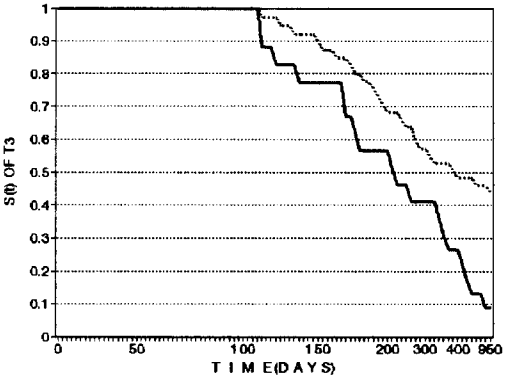
Therefore, in the overall sense we conclude that treatment DAT is slightly more effective than treatment DA.



(a) Response Time



(b) Progressive Time



(c) Relapse Time

Figure 2. Estimated Survival Functions

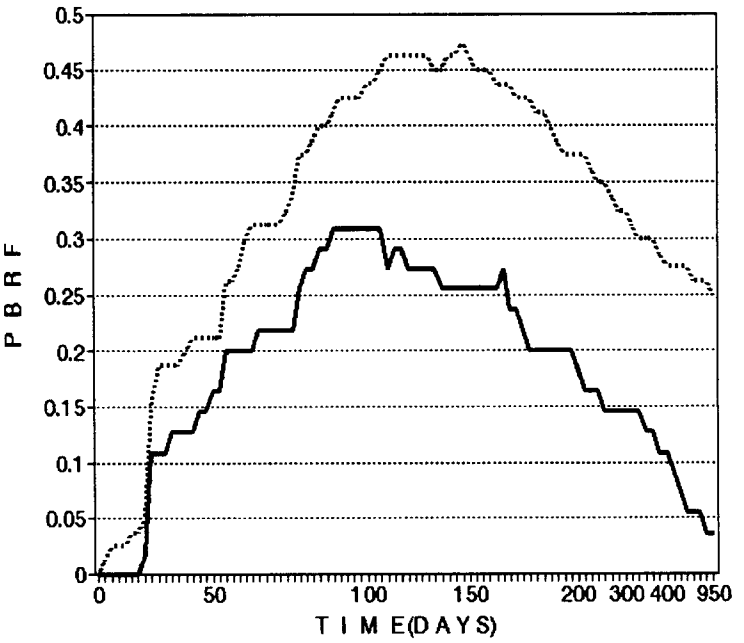


Figure 3. Estimated PBRF

5. Concluding Remarks

In this paper, we have used Cox's regression model to analyze multistate survival data. we have seen that the multistate Cox's regression model can be useful to a unified treatment for a complex experiment such as a clinical trial described in this paper.

On the other hand, we have informally shown that how the multistate Cox's regression model can be formulated within the framework of the multiplicative intensity model due to Aalen(1975, 1978). Based on this counting process formulation, we have briefly discussed the large sample properties of the estimators given in this paper.

However, in this paper we have been only concerned with the non-homogeneous Markov model. In fact, Cox's regression model may also be used for the semi-Markov model. And then we may want to develop methods of distinguishing between the Markov and the semi-Markov models.

In conclusion, we hope that the results in this paper would be useful to the analysis of data arising from clinical trials. We also hope that an attempt at the counting process formulation in this multistate survival data model would provide a good opportunity for

other statisticians to use these ideas in their own model buildings and analyses.

References

[1] Aalen, O. O.(1975). Statistical Inference for a Family of Counting Processes, Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.

[2] Aalen, O. O.(1978). Non-parametric Inference for a Family of Counting Processes, *Annals of Statistics*, Vol. 6, 701-726.

[3] Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian J. Statist,* Vol. 5, 141-150.

[4] Andersen, P. K. and Borgan, Ø.(1985). Counting Process Models for Life History Data : a Review (with discussian). *Scandinavian Journal of Statistics*, Vol. 12, 97-158.

[5] Andersen, P. K. and Gill, R. D.(1982). Cox's Regression Model for Counting Processes : a Large Sample Study, *Annals of Statistics*, Vol. 10, 1100-1120.

[6] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N.(1982). Linear Non-parametric Tests for Comparison of Counting Processes, with Applications to Censored Survival Data(with discussion). *International Statistical Review*, Vol. 50, 219-258.

[7] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N.(1993). Statistical Models Based on Counting Processes, Springer-Verlag, New York.

[8] 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*, Vol. 38, 59-66.

[9] Breslow, N. E.(1974). Covariance Analysis of Censored Survival Data, *Biometrics*, Vol. 30, 89-99.

[10] Cox, D. R.(1972). Regression Models and Life Tables (with discussion). *Journal of the Royal Statistical Society*, B, Vol. 34, 187-220.

[11] Cox, D. R.(1975). Partial Likelihood, *Biometrika*, Vol. 62, 269-276.

[12] Efron, B.(1977). Efficiency of Cox's Likelihood Function for Censored Data, *Journal of the American Statistical Association*, Vol. 72, 557-565.

[13] Fleming, T. R. and Harrington, D. P.(1991). *Counting Processes and Survival Analysis*, Wiley, New York.

[14] Gill, R. D.(1980). Censoring and Stochastic Integrals, *Mathematical Centre Tracts*, 124, Mathematical Centre, Amsterdam.

[15] Gill, R. D.(1984). Understanding Cox's Regression Model: a Martingale Approach, *Journal of the American Statistical Association*, Vol. 79, 441-447.

[16] Hsieh, F. Y.(1980). Multistate Survival Analysis, Ph.D. Dissertation, Department of Statistics, University of Wisconsin-Madison.

[17] Hsieh, F. Y., Crowley, J. and Tormey, D. C.(1983).Some Test Statistics for Use in Multistate Survival Analysis, *Biometrika*, Vol. 70, 111-119.

[18] Johansen, S.(1978). The Product Limit Estimate as a Maximum Likelihood Estimate,

Scandinavian Journal of Statistics, Vol. 5, 195-199.

[19] Kalbfleisch, j. D. and Prentice, R. L.(1973). Marginal Likelihoods Based on Cox's Regression and life model. *Biometrika*, Vol. 60, 267-278.

[20] Kalbfleisch, J. D. and Prentice, R. L.(1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.

[21] Lagakos, S. W.(1976). A Stochastic Model for Censored Survival Data in the Presence of an Auxiliary Variable, *Biometrics*, Vol. 32, 551-559.

[22] Lagakos, S. W.(1977). Using Auxiliary Variables for Improved Estimates of Survival Time, *Biometrics*, Vol. 33, 399-404.

[23] Lagakos, S. W., Sommer, C. J. and Zelen, M.(1978). Semi-Markov Models for Partially Censored Data, *Biometrika*, Vol. 65, 311-317.

[24] Pepe, M. S.(1991). Inference for Events with Dependent Risks in Multiple Endpoint Studies, *Journal of the American Statistical Association*, Vol. 86, 770-778.

[25] Ramlau-Hansen, H.(1983). Smoothing Counting Process Intensities by Means of Kernel Functions, *Annals of Statistics*, Vol. 11, 453-466.

[26] Temkin, N. R.(1978). An Analysis for Transient States with Application to Tumor Shrinkage, *Biometrics*, Vol. 34, 571-580.

[27] Tsiatis, A. A.(1981). A Large Sample Study of Cox's Regression Model, *Annals of Statistics*, Vol. 9, 93-108.

[28] Voelkel, J. G. and Crowley, J.(1984). Non-parametric Inference for a Class of Semi-Markov Processes with Censored Observations, *Annals of Statistics*, Vol. 12, 142-160.

[29] Yeo, S.(1989). Generalized Maximum likelihood Estimation in a Multistate Stochastic Model, *Journal of the Korean Operations Research and Management Science Society*, Vol. 14, 1-15.

Cox 회귀모형을 이용한 다중상태의 생존자료분석에 관한 연구[1)]

여성칠[2)]

요약

병원의 임상연구실험에서 종종 환자들의 치료에 따른 병세의 호전상태를 여러단계로 분류하여 상이한 치료방법에 따른 치료효과간의 차이를 알고자 하는 경우가 있다. 이와 같이 다중상태의 생존자료분석을 위한 한가지 방법으로 본 논문에서는 비동형의 Markov 모형에 Cox 회귀모형을 적용하여 회귀계수와 기저생존함수, 그리고 이를 바탕으로 반응확률함수를 추정하고 아울러 이들 추정량들의 대표본 성질들을 셈과정(Counting process)기법을 이용하여 알아 보았다. 그리고 본 논문의 결과에 대해 실제 예를 들어 보였다.

1) 이 논문은 1992년도 교육부지원 한국학술진흥재단의 자유공모과제 학술연구 조성비에 의하여 연구되었음.

2) (133-701) 서울특별시 성동구 모진동 93-1, 건국대학교 응용통계학과.