Existence and Consistence of MLEs for Piecewise Constant Intensity Model on Lexis Diagram

HANAYAMA Nobutane

Summary

Hanayama (2001) proposed a simple two-stage (STS) model for cancer death rates on the Lexis diagram. STS model belongs to neither the class of the generalized linear models nor the class of generalized additive models which are used in ordinary studies of survival data, hence, the properties of MLEs for those models are not relevant for STS model. So, in this paper, it is seen that the problem of getting MLEs for STS model results in that for the models on right-censored survival data and the existence and consistency of MLEs for the proposed model are proved.

Key words: Nonlinear model, Poisson regression model, Piecewise constant intensity model, Right-censored survival data.

1. Introduction

Hanayama (2001) proposed a simple two-stage (STS) model for cancer death rates on the Lexis diagram to project cancer risk in the environment referring to the two-sage cancer growing process which is elaborately studied in the field of biostatistics. (See Moolgavkar and Venzon, 1979, or Dewanji et al, 1999, for recent studies of the two-stage model.) Though it is seen that STS model belongs to the class of Poisson regression models, the model is nonlinear even if we adopt some link-functions used in the generalized linear model (McCullagh and Nelder, 1989). Because of that, the tools for GLIM are not available, hence, we have to get MLEs by solving the maximum likelihood equation directly. So, our aim in this paper is to make sure the existence and consistence of MELs of the parameters in the model.

In the next Section 2, we introduce STS model and show its properties for MLEs, where it is seen that the problem of getting MLEs for STS model results in that for models on the right-censored survival data. In Section 3, the existence and consistency of MLEs are shown in the manner of Lehmann (1983, pp. 427-432).

2. STS Model

2.1. Basic Idea

Consider cancer incidence rates of age u and time (year) v denoted by $\mu(u, v)$, where $(u, v) \in [0, a) \times [0, b)$. Now notice that the intensity function $\mu(u, v)$ is the one for individuals who were born at $v - u \in (-a, b)$ and are still alive at $v \in [0, b)$. Then, hanayama (2001) proposed a model which assumed that

$$\mu(u,v)=\int_{v-u}^{v}\xi(t)\psi(v-t)dt, \tag{2.1}$$

where $\xi(t)$ $(t \in (-a, b))$ is the intensity with which cells in a body of individual alive at t have been primed by carcinogen existing in environment during $[t, t+dt)$, and $\psi(s-t)$ $(s \in [0, b), s-t \in [0, a)$ and $s \geq t)$ is the intensity with which a cell primed at $t \in (-a, b)$ has grown to be terminal cancer during $[s, s+ds)$. Although (2.1) is presenting the basic idea for the model, we cannot apply it to data given by 5-year age groups. For the analysis of such data, the piecewise constant intensity (PCI) models on the Lexis diagram is known as a useful tool. (See Keiding, 1990.) So we reconstruct the model (2.1) as one belonging to the class of PCI models in the followings.

2.2. STS Model for Piecewise Constant Intentisy

Suppose that the age and time intervals $[0, a)$ and $[0, b)$ are divided into I and J 5-year intervals respectively so that a and b are written as $a=5I$ and $b=5J$, where I and J are fixed integers. Assume that

$$\mu(u,v)=\mu_{ij} \; if \; (u,v) \in W_{ij} \equiv \{(u,v) \mid (u,v) \in [5(i-1),5i) \times [5j-1),5j)\}$$

$$\mu(u,v)=\mu_{ij} \; if \; (u,v) \in W_{ij} \equiv \{(u,v) \mid v-u \in [5(j-i-1),5(j-i)) \text{ and } u \in [5(j-1),5j)\}$$

where $i=1,\cdots,I$ and $J=1,\cdots,J$. (Figure 1 illustrates W_{ij}'s on the Lexis diagram. In the figure the abscissa and ordinate represent time and age respectively, solid oblique lines represent individuals being alive, and the squares figured by dotted lines represent W_{ij}'s.) Then, the model is reconstructed as

$$\mu_{ij} = \mu_{ij}(\theta) \equiv \sum_{k=j-i+1}^{j} \xi_k \psi_{j-k+1}, \tag{2.2}$$

where $\theta \equiv (\psi_1, \cdots, \psi_I, \; \xi_{2-I}, \cdots, \xi_J)' \in \Theta \equiv \{\theta \mid \psi_1 > 0, \psi_i \in (-\infty, \infty); i = 2, \cdots, I, \xi_k > 0; k=2-I, \cdots, J, \mu_{ij}(\theta) > 0 \text{ for } i=1,\cdots,I; j=1,\cdots,J\}$. Now notice that $\psi_1 > 0$ because it is supposed that $\mu_{1j}(\theta) > 0$ and $\xi_j > 0$. (Figure 2 illustrate parameters in the model on the Lexis diagram.)

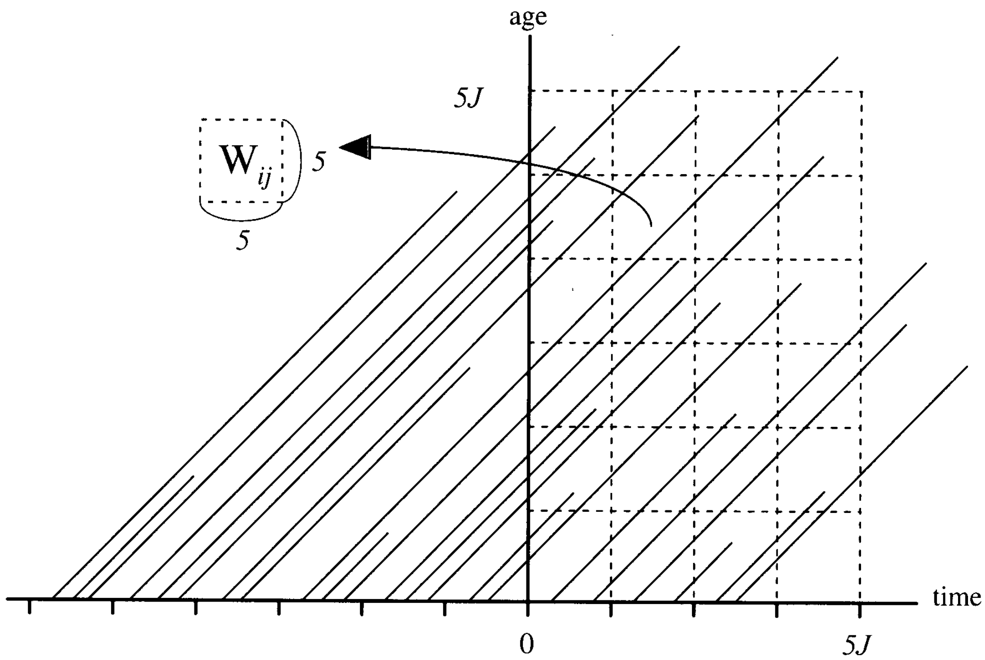

Figure 1. W_{ij}'s on the Lexis diagram when I=6 and J=4.

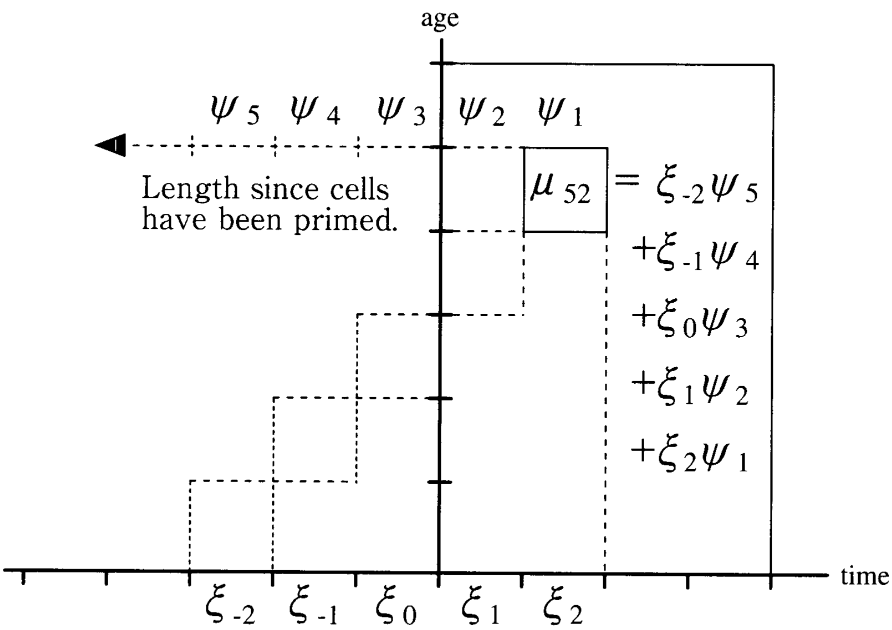

Figure2. The parameters in STS model on the Lexis diagram.

2.3. Likelihood Function and Identifiability

Let $X_{ij}^{(n)}$ ($n=1,\cdots,N_{ij}$) be a random variable representing the survival time for the n-th member of those who are alive over W_{ij}, and $D_{ij}^{(n)}$ be an indicator describing whether he/she has died of cancer over W_{ij} ($D_{ij}^{(n)}=1$) or not ($D_{ij}^{(n)}=0$). And assume that

$$\Pr\left\{ X_{ij}^{(n)} \in [x, x+dx) \mid X_{ij}^{(n)} \geq x \right\} = (\mu_{ij} + \lambda_{ij})dx. \tag{2.3}$$

where λ_{ij}'s represent the death rates of other causes. Then, the contribution to the likelihood based on an observation ($x_{ij}^{(n)}$, $d_{ij}^{(n)}$) on ($X_{ij}^{(n)}$, $D_{ij}^{(n)}$) is

$$L_{i,j}^{(n)}(\theta) \equiv \{\mu_{ij}(\theta)\}^{d_{ij}^{(n)}} \exp\left[-\left\{\mu_{ij}(\theta)+\lambda_{ij}\right\} x_{ij}^{(n)}\right]$$

Thus the likelihood based on observations ($x_{ij}^{(n)}$, $d_{ij}^{(n)}$)'s is proportional to

$$L(\theta) \equiv \prod_{i=1}^{I}\prod_{j=1}^{J} L_{i,k+i}^{(n)}(\theta) = \prod_{i=1}^{I}\prod_{j=1}^{J} \mu_{ij}^{d_{ij}^{\bullet}}(\theta) \exp\left[-\left\{\mu_{ij}(\theta)+\lambda_{ij}\right\} x_{ij}^{\bullet}\right], \tag{2.4}$$

where $d_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} d_{ij}^{(n)}$ and $x_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} x_{ij}^{(n)}$. (See Keiding, 1975.) The function is of well known form in previous studies of PCI models on the Lexis diagram. (See Keiding, 1990, Berzuini and Clayton, 1994, Robertson and Boyle, 1998, for example.)

Though MLEs of are obtained by maximizing $L(\theta)$, it is a function of θ only through $\mu(\theta) \equiv (\mu_{1,1}(\theta),\cdots,\mu_{1,J}(\theta),\mu_{2,1}(\theta),\cdots,\mu_{2,J}(\theta),\cdots\cdots,\mu_{I,1}(\theta),\cdots,\mu_{I,1}(\theta))'$. Hence, we have to see if θ is uniquely determined for every $\mu \equiv (\mu_{1,1},\cdots,\mu_{1,J},\mu_{2,1},\ldots,\mu_{2,J},\cdots\cdots,\mu_{I,1},\cdots,\mu_{I,J})' \in \mathrm{M} \equiv \{\mu \mid \mu = \mu(\theta), \theta \in \Theta\}$ from the equation $\mu = \mu(\theta)$. As for this problem, Hanayama (2001) gave a proposition where it is shown that the value of $\boldsymbol{q}(\psi_1, \theta)$ is uniquely determined from the equation

$$\mu(\theta) = \boldsymbol{q}(w_1, \theta) \equiv \left(\frac{\psi_1}{w_1}, \cdots, \frac{\psi_I}{w_1}, w_1\xi_{2-I}, \cdots, w_1\xi_J\right)'$$

for every $\mu \in \mathrm{M}$ unless μ satisfies

$$\mu_{i,k+i-1} - \mu_{i-1,k+i-1} (= \xi_k \psi_1) = 0\ ;\ i = a_k, \cdots, b_k - 1, \tag{2.5}$$

for more than one k, where w_1 is a positive constant and $a_k = \max\{1, 2-k\}$, $b_k = \min\{I, J-k+1\}$.

Because of the above proposition, we may assume that $\theta = (1, \theta_*')'$ where $\theta^* \equiv (\psi_2, \cdots, \psi_I, \xi_{2-I}, \cdots \xi_J)'$ and have the likelihood equation with respect to θ like

$$\frac{\partial \log L}{\partial \theta_*}\Big|_{\theta_* = \hat{\theta}_*} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\partial \mu_{ij}((1,\theta_*')')}{\partial \theta_*}\left(\frac{d_{ij}^{\bullet}}{\mu_{i,j}((1,\theta_*')')} - x_{ij}^{\bullet}\right)\Big|_{\theta_* = \hat{\theta}_*} = 0_{2I+J-2}, \quad (2.5)$$

where

$$\frac{\partial \mu_{ij}((1,\theta_*')')}{\partial \theta_*} = \begin{cases} \left(0_{I-1}', 0_{j-1+I-i}', \psi_i, \cdots, \psi_2, 1, 0_{J-j}'\right)' & \text{if } i=1;\ j=1,\cdots,J \\ \left(\xi_{j-1}, \cdots, \xi_{j-i+1}, 0_{I-i-1}', 0_{j-1+I-i}', \psi_i, \cdots, \psi_2, 1, 0_{J-j}'\right)' & \text{if } \begin{array}{l} i=2,\cdots,I; \\ j=1,\cdots,J, \end{array} \end{cases}$$

and 0_l indicates a l-dimensional row vector whose componenents are all zero. Thus, the solution $\hat{\theta}_*$ is obtained as a MLE consistent for $(\psi_2^0/\psi_1^0,\cdots,\psi_I^0/\psi_1^0,\psi_1^0\xi_{2-I}^0,\cdots,\psi_1^0\xi_J^0)'$, where $\xi_k^{(0)}$'s and $\psi_i^{(0)}$'s are the true parameters.

3. Existence and Consistence of MLE

The likelihood equation (2.5) has the same form as would have been obtained when $D_{ij}^{\bullet}$'s ($\equiv \sum_{n=1}^{N_{ij}} D_{ij}^n$) are assumed to be Poisson distributed with mean $\mu_{ij}(\theta)x_{ij}^{\bullet}$ so that STS model belongs to the class of Poisson regression models. However, the model is nonlinear even if we transform μ_{ij} by some link-functions used in GLIM (McCullagh and Nelder, 1989). So, the tools for GLIM are not available, hence, we have to get MLEs by solving (3.4) directly. Because of that, it is needed to make sure the existence and consistence of $\hat{\theta}_*$. Thus, in the following, the existence and consistency of MLEs are shown as the totals of survival time are regarded as random variables. The way of discussion is in the manner of Lehmann (1983, pp. 427-432).

The complexity of the proof of the existence and consistency is due to the fact that N_{ij}'s are random variables, and $(X_{ij}^{\bullet}, D_{ij}^{\bullet})$ and $(X_{i-1,j-1}^{\bullet}, D_{i-1,j-1}^{\bullet})$, where $X_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} X_{ij}^{(n)}$, are not independent each other. So, for the convenient to prove the existence and consistency, we introduce the following notations. First, let m_{j-i} ($=1,\cdots,M_{j-i}$) be a given constant which is the numbers of those who were born in the time interval $[5(j-i-1), 5(j-i+1))$. Further, let $Y_{ij}^{(m_{j-i})}$ be a random variable representing the survival time for the m_{j-i}th member of those who were born in $[5(j-i-1), 5(j-i+1))$ over W_{ij}, and $E_{ij}^{(m_{j-i})}$ be an indicator describing whether he/she has died of cancer over W_{ij} ($E_{ij}^{(m_{j-i})}=1$) or not ($E_{ij}^{(m_{j-i})}=0$).

Then, in the followings, we will show that there exists a solution which is consistent for $\theta_*^0 \equiv \psi_2^0/\psi_1^0,\cdots,\psi_I^0/\psi_1^0,\psi_1^0,\xi_{2-I}^0,\cdots,\psi_1^0\xi_J^0)'$, with probability tending to 1, as the minimum value of M_{j-i}'s, $M_{\min}$ say, is infinitely large, as long as it is assumed that

$$\frac{M_l}{M_{\min}} \to h_l^0 > 0 \text{ for } l = 2-I,\ldots,J \text{ as } \mathrm{M}_{\min} \to \infty. \tag{3.1}$$

Now, for the purpose of showing the existence and consistency, we state the following properties of the likelihood function:

(P1)
$$E\Big[\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{m_{j-i}=1}^{M_{j-i}} \frac{\partial \log L_{ij}^{(m_{j-i})}((1,\theta_*')')}{\partial\theta_*}\Big] = 0,$$

where $L_{ij}^{(m_{j-i})}(\theta) \equiv \{\mu_{ij}(\theta)\}^{e_{ij}^{(m_{j-i})}} \exp\Big[-\{\mu_{ij}(\theta)+\lambda_{ij}\}e_{ij}^{(m_{j-i})}\Big]$.

(See Appendix A for the proof.)

(P2)
$$-E\Big[\frac{\partial^2 \log L_{ij}^{(m_{j-i})}((1,\theta_*')')}{\partial\theta_*\partial\theta_*}\Big]$$

$$= E\Big[\Big\{\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\frac{E_{ij}^{(m_{j-i})}}{\{\mu_{ij}((1,\theta_*')')\}^2}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}{}'$$

$$-\Big(\frac{E_{ij}^{(m_{j-i})}}{\mu_{ij}((1,\theta_*')')} - Y_{ij}^{(m_{j-i})}\Big)\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_*\partial\theta_*}\Big\}\Big]$$

$$= \frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\frac{\Big[1-\exp\Big\{-(\mu_{ij}((1,\theta_*')')+\lambda_{ij})\tau_{ij}^{(m_{j-i})}\Big\}\Big]}{\mu_{ij}((1,\theta_*')')(\mu_{ij}((1,\theta_*')')+\lambda_{ij})}$$

$$\times\exp\Big\{-\sum_{m_{j-i}=1}^{M_{j-i}}(\mu_{ij}(1,\theta_*)+\lambda_{ij})\tau_{ij}^{(m_{j-i})}\Big\}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*'},$$

where $\tau_{ij}^{(m_{j-i})}$ is the maximum value of $Y_{ij}^{(m_{j-i})}$ for fixed m_{j-i}.

(The proof of this is essentially given in the proof of (P1).)

(P3) There exists a functions $C_{ij}^{(m_{j-i})}$ of $(E_{ij}^{(m_{j-i})}, Y_{ij}^{(m_{j-i})})$ $(l,m,n=1,\cdots,I+2J\text{-}2)$ such that

$$\left|\frac{\partial^3 \log L_{ij}^{(m_{j-i})}((1,\theta_*')')}{\partial\theta_l^*\partial\theta_m^*\partial\theta_n^*}\right| \le C_{l,m,n}^{(ij)}(E_{ij}^{(m_{j-i})}, Y_{ij}^{(m_{j-i})})$$

for all $\theta_*\in\Theta_*\equiv\{\theta_*|\psi_i\in(-\infty,\infty)$; $i=2,\cdots,I, \xi_k>0$; $k=2\text{-}I,\cdots,J$, $\mu_{ij}((1,\theta_*')')>0$ for $i=1,\cdots,I$; $j=1,\cdots,J\}$, where

$$\theta_l^* = \begin{cases}\psi_{l+1} & \text{if } 1\le l\le I-1\\ \xi_{l+1-I} & \text{if } I\le l\le 2I+J-2\end{cases}$$

and $c_{l,m,n}^{(ij)} = E\left[C_{l,m,n}^{(ij)}(D_{ij}^{(n)}X_{ij}^{(n)})\right] < \infty$ for all at neighborhood of θ_*^0.
(See Appendix B for the proof.)

Now we give the following proposition.

Proposition 3.1. Consider a sphere

$$Q_a(\theta_*^0) = \left\{ \theta_* \;\middle|\; \theta_* \in \Theta_* \;\; and \;\; \left\|\theta_* - \theta_*^0\right\| = a \right\}.$$

Then, for any sufficiently small $a>0$, the probability that

$$\log L((1,\theta_*')') < \log L((1,\theta_*^{0'})') \;\; \text{at all} \;\; \theta_* \in Q_a(\theta_*^0) \tag{3.2}$$

tends to 1 as $N_{\min} \to \infty$.
(See Appendix C for the proof.)

Hence the function $\log L((1, \theta_*)')$ must have a local maximum point interior of $Q_a(\theta_*^0)$ with probability tending to 1 as $M_{\min} \to \infty$. Further, because the above proposition holds for any small $a>0$, such a local maximum point consists with θ_*^0 as $M_{\min} \to \infty$.

4. Concluding remarks

We have shown that the existence and consistency of MLEs of the parameters in the proposed model in Section 3. Miao and Hahn (1997) showed the existence and consistency of likelihood estimates for multi-dimensional exponential families. Although the Poisson model is a special one belonging to multi-dimensional exponential families, the way of discussion in the manner of them does not suit to our model. This is due to the fact that the joint distribution function for $(X_{ij}^{\bullet}, D_{ij}^{\bullet})$'s can not be expressed by the canonical form of the exponential families immediately because of the fact described there.

Appendix A (Proof of (P1))

Let $Y_{ij}^{(n)}$ be a random variable representing the survival time in W_{ij} for m_{j-i}th member ($m_{j-i}=1,\cdots,M_{j-i}, j-i=1-I,\cdots,J-1$) those who were born in the time interval $[5(j-i-1), 5(j-i+1))$ and $E_{ij}^{(m_{j-i})}$ be an indicator describing whether he/she has died of cancer over W_{ij} ($E_{ij}^{(m_{j-i})}=1$) or not ($E_{ij}^{(m_{j-i})}=0$). Then we have ....

$$E\left[\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{n=1}^{N_{ij}} \frac{\partial \mu_{ij}((1,\theta_*')')}{\partial \theta_*} \left(\frac{D_{ij}^{(n)}}{\mu_{i,j}((1,\theta_*')')} - X_{ij}^{(n)} \right)\right]$$

$$= E\left[\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{m_{j-i}=1}^{M_{j-i}} \frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\left(\frac{E_{ij}^{(m_{j-i})}}{\mu_{i,j}((1,\theta_*')')} - Y_{ij}^{(m_{j-i})}\right)\right]$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{m_{j-i}=1}^{M_{j-i}} \frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\left[E\left[\left(\frac{E_{ij}^{(m_{j-i})}}{\mu_{i,j}((1,\theta_*')')} - Y_{ij}^{(m_{j-i})}\right)\middle| Y_{ij}^{(m_{j-i})} = 0\right]\Pr\left\{Y_{ij}^{(m_{j-i})} = 0\right\}\right.$$

$$\left. E\left[\left(\frac{E_{ij}^{(m_{j-i})}}{\mu_{i,j}((1,\theta_*')')} - Y_{ij}^{(m_{j-i})}\right)\middle| Y_{ij}^{(m_{j-i})} > 0\right]\Pr\left\{Y_{ij}^{(m_{j-i})} > 0\right\}\right] \tag{A.1}$$

In the above formula, it is apparently found that

$$E\left[\left(\frac{E_{ij}^{(m_{j-i})}}{\mu_{ij}((1,\theta_*')')} - Y_{ij}^{(m_{j-i})}\right)\middle| Y_{ij}^{(m_{j-i})} = 0\right] = 0. \tag{A.2}$$

Besides, $E_{ij}^{(m_{j-i})}$ and $Y_{ij}^{(m_{j-i})}$ are independent each other because both of μ_{ij} and λ_{ij} are constant in $W_{ij}^{(m_{j-i})}$. Thus, other terms in (A.1) are calculated separately. First the expectations of $E_{ij}^{(m_{j-i})}$ are written as

$$E\left[E_{ij}^{(m_{j-i})}\middle| Y_{ij}^{(m_{j-i})} > 0\right]$$

$$= \int_{0<y<\tau_{ij}^{(m_{j-i})}} \Pr\left\{Y_{ij}^{(m_{j-i})} \in [y, y+dy), E_{ij}^{(m_{j-i})} = 1 \middle| Y_{ij}^{(m_{j-i})} > 0\right\}$$

$$= \int_{0<y<\tau_{ij}^{(m_{j-i})}} \mu_{ij}((1,\theta_*')')\exp\{-(\mu_{ij}((1,\theta_*')') + \lambda_{ij})y\}dx \tag{A.3}$$

$$= \frac{\mu_{ij}((1,\theta_*')')}{\mu_{ij}((1,\theta_*')') + \lambda_{ij}}\left[1 - \exp\left\{-(\mu_{ij}((1,\theta_*')') + \lambda_{ij})\tau_{ij}^{(m_{j-i})}\right\}\right],$$

where $\tau_{ij}^{(m_{j-i})} = 5(i-j+1) + b_{ij}^{(m_{j-i})}$. Next, the expectations of $Y_{ij}^{(m_{j-i})}$ is written as

$$E\left[Y_{ij}^{(m_{j-i})}\middle| Y_{ij}^{(m_{j-i})} > 0\right]$$

$$= \int_{0<y<\tau_{ij}^{(m_{j-i})}} y(\mu_{ij}((1,\theta_*')') + \lambda_{ij})\exp\left\{-(\mu_{ij}((1,\theta_*')') + \lambda_{k+i}^{i})y\right\}dy$$

$$+ \tau_{ij}^{(m_{j-i})}\exp\left\{-\left(\mu_{ij}((1,\theta_*')') + \lambda_{ij}\right)\tau_{ij}^{(m_{j-i})}\right\} \tag{A.4}$$

$$= \frac{1}{\mu_{ij}((1,\theta_*')') + \lambda_{ij}}\left[1 - \exp\left\{-\left(\mu_{ij}((1,\theta_*')') + \lambda_{ij}\right)\tau_{ij}^{(m_{j-i})}\right\}\right]$$

From (A.1), (A.2), (A.3) and (A.4) and we have the equation (P1).

Appendix B (Proof of (P3))

Supposed that the nth member of those who are alive over W_{ij} is the same as $m_{j,i}$th member of those who were born in the time interval $[5(j-i-1), 5(j-i+1))$. Then, we find that

$$\frac{\partial^3 \log L_{ij}^{(n)}(n)}{\partial\theta_l\,\partial\theta_m\partial\theta_n} = \frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_l}\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_m\partial\theta_n}\frac{E_{ij}^{(m_{j-i})}}{\{\mu_{ij}((1,\theta_*')')\}^2}$$

$$-\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_l\,\partial\theta_m}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_n}\frac{E_{ij}^{(m_{j-i})}}{\{\mu_{ij}((1,\theta_*')'\}^2} \quad \text{(B.1)}$$

$$-\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_l\,\partial\theta_m}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_n}\frac{E_{ij}^{(m_{j-i})}}{\{\mu_{ij}((1,\theta_*')')\}^2}$$

$$+2\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_l}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_m}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_n}\frac{E_{ij}^{(m_{j-i})}}{\{\mu_{ij}((1,\theta_*')')\}^3}$$

where

$$\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_l^*} = \begin{cases} \xi_{j-l} & if\ 1 \le l \le i-1 \\ \psi_{k+i-(l-I)} & if\ j-i+1 \le l-2I+2 \le i \\ 0 & otherwise\ , \end{cases}$$

$$\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_l^*\partial\theta_m^*} = \begin{cases} 1 & if\ j-l = m-2I+2 \\ 0 & otherwise\ , \end{cases}$$

and

$$\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_l^*\partial\theta_m^*\partial\theta_n^*} = 0 \text{ for all } l,\ m,\ n.$$

Thus, by considering a functiion, $C_{l,m,n}^{(i,j)}$ say, whose order is greater than $\mu_{ij}((1,\theta_*')')$ when $\mu_{ij}((1,\theta_*')') \ge 1$ or less than $\{\mu_{ij}((1,\theta_*')')\}^{-3}$, we have

$$\left|\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\partial^2 \log L_{ij}((1,\theta_*')')}{\partial\theta_l^*\partial\theta_m^*\partial\theta_n^*}\right| \le C_{l,m,n}^{(i,j)}(E_{ij}^{(m_{j-i})}{}'s, Y_{ij}^{(m_{j-i})}{}'s) \quad \text{(B.2)}$$

for all for all $\theta_* \in \Theta_*$, and

$$c_{l,m,n}^{(i,j)} = E\left[C_{l,m,n}^{(i,j)}(E_{ij}^{(m_{j-i})}{}'s, Y_{ij}^{(m_{j-i})}{}'s)\right] < \infty \text{ for all } l,\ m,\ n.$$

Appendix C

To obtain needed facts concerning the behavior of the likelihood on $Q_a(\theta_*^0)$ for small a>0, we expand the log likelihood about θ_* and divide by $M_{\min}$ to find

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m_{j-i}=1}^{M_{ij}}\left\{\frac{1}{M_{\min}}\log L_{ij}^{(m_{j-i})}(1,\theta_*)-\frac{1}{M_{\min}}\log L_{ij}^{(m_{j-i})}(1,\theta_*^0)\right\}$$

$$=(\theta_*-\theta_*^0)'\frac{1}{M_{\min}}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m_{j-i}=1}^{M_{j-i}}\left.\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\right|_{\theta_*=\theta_*^0}A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})$$

$$+\frac{1}{2}(\theta_*-\theta_*^0)'\frac{1}{N_{\min}}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m_{j-i}=1}^{M_{j-i}}\left\{A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})\left.\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta\partial\theta'}\right|_{\theta_*=\theta_*^0}\right. \tag{C.1}$$

$$\left.-\left.\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\right|_{\theta_*=\theta_*^0}B_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})\left.\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\right|_{\theta_*=\theta_*^0}\right\}(\theta_*-\theta_*^0)$$

$$+\frac{1}{6}\sum_{l=1}^{I+2J-2}\sum_{m=1}^{I+2J-2}\sum_{n=1}^{I+2J-2}\left\{(\theta_l^*-\theta_l^{*0})(\theta_m^*-\theta_m^{*0})(\theta_n^*-\theta_n^{*0})\right.$$

$$\left.\times\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m_{j-i}=1}^{M_{j-i}}\frac{1}{M_{\min}}\gamma_{l,m,n}^{(i,j)}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})C_{l,m,n}^{(i,j)}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})\right\}$$

$$=S_1(\theta_*,\theta_*^0)+S_2(\theta_*,\theta_*^0)+S_3(\theta_*,\theta_*^0),$$

where

$$A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})=\left(\frac{e_{ij}^{(m_{j-i})}}{\mu_{ij}((1,\theta_*^{0'})')}-y_{ij}^{(m_{j-i})}\right),\quad B_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})})=\frac{e_{ij}^{(m_{j-i})}}{\{\mu_{ij}((1,\theta_*^{0'})')\}^2}$$

and, by the property (P3),

$$0\le\left|\gamma_{l,m,n}^{(i,j)}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})C_{l,m,n}^{(i,j)}(e_{ij}^{(m_{j-i})},y_{ij}^{(m_{j-i})})\right|\le 1.$$

To prove that the maximum of the difference

$$\max_{\theta_*\in Q_a(\theta_*^0)}\left\{\frac{1}{M_{\min}}\log L_{ij}^{(m_{j-i})}(1,\theta_*)-\frac{1}{M_{\min}}\log L_{ij}^{(m_{j-i})}(1,\theta_*^0)\right\}<0$$

with probability tending to 1 for any sufficiently small $a>0$, we will show that with high probability

$$\max_{\theta_*\in Q_a(\theta_*^0)}\left\{S_2(\theta_*,\theta_*^0)\right\}<0 \text{ while } \max_{\theta_*\in Q_a(\theta_*^0)}\left\{S_1(\theta_*,\theta_*^0)\right\} \text{ and } \max_{\theta_*\in Q_a(\theta_*^0)}\left\{S_3(\theta_*,\theta_*^0)\right\}<0$$

are small compared to S_2. The basic tools for showing this are the facts that by (P1), (P2) and the law of large numbers

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{M_{\min}}\sum_{m_{j-i}=1}^{M_{ij}}\left.\frac{\partial\log L_{ij}^{(m_{j-i})}(1,\theta_*)}{\partial\theta_*}\right|_{\theta_*=\theta_*^0} \tag{C.2}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{M_{j-i}}{M_{\min}}\sum_{m_{j-i}=1}^{M_{j-i}}\frac{1}{M_{j-i}}\frac{\partial\mu_{ij}(1,\theta_*)}{\partial\theta_*}\Bigg|_{\theta_*=\theta_*^0} A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})}, y_{ij}^{(m_{j-i})})$$

$$\rightarrow \sum_{i=1}^{I}\sum_{j=1}^{J} h_{j-i}^0 \frac{\partial\mu_{ij}(1,\theta_*)}{\partial\theta}\Bigg|_{\theta_*=\theta_*^0} E\left[A_{ij}^{(m_{j-i})}(E_{ij}^{(m_{j-i})}, Y_{ij}^{(m_{j-i})})\right]=0$$

in probability, and

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{M_{\min}}\sum_{m_{j-i}=1}^{M_{j-i}}\frac{\partial^2 \log L_{ij}^{(m_{j-i})}(1,\theta_*)}{\partial\theta_*\partial\theta_*'}\Bigg|_{\theta_*=\theta_*^0}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{M_{j-i}}{M_{\min}}\sum_{m_{j-i}=1}^{M_{j-i}}\frac{1}{M_{j-i}}\left\{A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})}, y_{ij}^{(m_{j-i})})\frac{\partial^2\mu_{ij}(1,\theta_*)}{\partial\theta_*\partial\theta_*'}\right.$$

$$\left. - \frac{\partial\mu_{ij}(1,\theta_*)}{\partial\theta_*} B_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})})\frac{\partial\mu_{ij}(1,\theta_*)}{\partial\theta_*'}\Bigg|_{\theta_*=\theta_*^0}\right\} \tag{C.3}$$

$$\rightarrow \sum_{i=1}^{I}\sum_{j=1}^{J} h_{j-i}^0 \sum_{m_{j-i}=1}^{M_{j-i}}\left\{E\left[A_{ij}^{(m_{j-i})}(E_{ij}^{(m_{j-i})}, Y_{ij}^{(m_{j-i})})\right]\frac{\partial^2\mu_{ij}(1,\theta_*)}{\partial\theta_*\partial\theta_*'}\right.$$

$$\left. - \frac{\partial\mu_{ij}(1,\theta_*)}{\partial\theta_*} E\left[B_{ij}^{(m_{j-i})}(E_{ij}^{(m_{j-i})})\right]\frac{\partial\mu_{k+i}^{i}(\theta)}{\partial\theta_*'}\Bigg|_{\theta_*=\theta_*^0}\right\}$$

$$= -\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\partial\mu_{ij}(1,\theta_*)}{\partial\theta_*} h_{j-i}^0 I_{ij}(\theta_*^0)\frac{\partial\mu_{k+i}^{i}(\theta)}{\partial\theta_*'}\Bigg|_{\theta_*=\theta_*^0}$$

in probability, where

$$I_{ij}(\theta_*^0) = \frac{\left[1-\exp\left\{-(\mu_{ij}(1,\theta_*)+\lambda_{ij})\tau_{ij}^{(m_{j-i})}\right\}\right]}{\mu_{ij}(1,\theta_*)(\mu_{ij}(1,\theta_*)+\lambda_{ij})} \times \exp\left\{-\sum_{m_{j-i}=1}^{M_{j-i}}(\mu_{ij}(1,\theta_*)+\lambda_{ij})\tau_{ij}^{(m_{j-i})}\right\}$$

Let us begin with S_1. On $Q_a(\theta_*^0)$ we have

$$\left|S_1(\theta_*,\theta_*^0)\right| < \left|\frac{1}{M_{\min}}(\underbrace{a;a,\dots,a}_{I+2J-2})\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m_{j-i}=1}^{M_{j-i}}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\Bigg|_{\theta_*=\theta_*^0} A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})}, y_{ij}^{(m_{j-i})})\right|.$$

With probability tending to 1, for any given a>0, it follows from (E.2) that

$$h_{j-i}^0\left|\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\Bigg|_{\theta_*=\theta_*^0} A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})}, y_{ij}^{(m_{j-i})})\right| < a^2 \text{ for } l=1,\dots,I+2J-2$$

and hence that

$$\left|S_1(\theta_*,\theta_*^0)\right| < \sum_{k=2-I}^{J} h_k^0(I+2J-2)a^3 \tag{C.4}$$

as $M_{\min}\rightarrow\infty$. Next, consider

$$2S_2(\theta_*,\theta_*^0)$$

$$= -(\theta_* - \theta_*^0)'\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{m_{j-i}=1}^{M_{j-i}} \frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\frac{M_k}{M_{\min}} I_{ij}(\theta_*^0)\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\bigg|_{\theta_*=\theta_*^0}(\theta_* - \theta_*^0)$$

$$+ (\theta_* - \theta_*^0)'\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{m_{j-i}=1}^{M_{j-i}} \left\{\frac{1}{M_{\min}} A_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})}, y_{ij}^{(m_{j-i})})\frac{\partial^2\mu_{ij}((1,\theta_*')')}{\partial\theta_*\partial\theta_*}\bigg|_{\theta_*=\theta_*^0}\right.$$

$$\left. -\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\frac{1}{M_{\min}}\left(B_{ij}^{(m_{j-i})}(e_{ij}^{(m_{j-i})}) - I_{ij}(\theta_*^0)\right)\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\bigg|_{\theta_*=\theta_*^0}\right\}(\theta_* - \theta_*^0)$$

For the second term, $S_2^{(2)}(\theta_*,\theta_*^0)$ say, it follows from (C.3) and an argument analogous to that for S_I that

$$\left|S_2^{(2)}(\theta_*\ \theta_*^0)\right| < \sum_{i=1}^{I}\sum_{j=1}^{I} h_{j-i}^0(I+2J-2)^2 a^3 \tag{C.5}$$

with probability tending to 1 as $M_{\min}\rightarrow\infty$. The first term, $S_2^{(1)}(\theta_*,\theta_*^0)$ say, is a quadratic form in the variable $(\theta_* - \theta_*^0)$. Because the matrix

$$\sum_{i=1}^{I}\sum_{j=1}^{I}\frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*}\ \frac{\partial\mu_{ij}((1,\theta_*')')}{\partial\theta_*'}\bigg|_{\theta_*=\theta_*^0}$$

is positive definite as shown in hanayama (2001), by an orthogonal transformation, $S_2^{(1)}(\theta_*,\theta_*^0)$ can be reduced to diagonal form

$$2S_2^{(1)}(\theta_*\ \theta_*^0) = \sum_{l=1}^{I+2J-2}\omega_l\sigma_l^2 \text{ becoming } \sum_{l=1}^{I+2J-2}\sigma_l^2 = a^2 \text{ on } \theta_*\in Q_a(\theta_*^0) \tag{C.6}$$

where ω_l's are negative and numbered so that $\omega_1\leqq\omega_2\leqq\cdots\leqq\omega_{I+2J-2}<0$. Then it is found that

$$\sum_{l=1}^{I+2J-2}\omega_l\sigma_l^2 \geq \omega_1\sum_{l=1}^{I+2J-2}\sigma_l^2 = \omega_l\sigma^2$$

Combining the first term and second term, that is, (C.5) and (C.6), we see that there exist $c>0, a_0>0$ such that for $a<a_0$

$$S_2(\theta_*,\theta_*^0) < -ca^2 \text{ on } \theta_*\in Q_a(\theta_*^0) \tag{C.7}$$

with probability tending to 1 as $M_{\min}\rightarrow\infty$. Finally, with probability tending to 1,

$$\left|\frac{1}{M_{\min}}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{m_{j-i}=1}^{M_{j-i}} C_{l,m,n}^{(i,j)}(e_{ij}^{(m_{j-i})}, y_{ij}^{(m_{j-i})})\right| < 2\sum_{i=1}^{I}\sum_{j=1}^{I} h_{j-i}^{0} c_{l,m,n}^{(i,j)}$$

and hence

$$\left|S_3(\theta_*, \theta_*^0)\right| < ba^3, \tag{C.8}$$

where

$$b = \frac{1}{3}\sum_{l=1}^{I+2J-2}\sum_{m=1}^{I+2J-2}\sum_{n=1}^{I+2J-2}\sum_{i=1}^{I}\sum_{j=1}^{J} h_{j-i}^{0} c_{l,m,n}^{(i,j)}$$

on $\theta_* \in Q_a(\theta_*^0)$. Combining the three inequalities (C.6), (C.7) and (C.8), we see that with probability tending to 1 as $M_{\min}\to\infty$,

$$\begin{aligned} &\max\left\{S_1(\theta_*, \theta_*^0) + S_2(\theta_*, \theta_*^0) + S_3(\theta_*, \theta_*^0)\right\} \\ &< -ca^2 + \left\{b + (I + 2J - 2)\sum_{i=1}^{I}\sum_{j=1}^{J} h_{j-i}^{0}\right\} a^3 \end{aligned} \tag{C.9}$$

which is less than zero if

$$a < \frac{c}{b + (I + 2J - 2)\sum_{i=1}^{I}\sum_{j=1}^{J} h_{j-i}^{0}}, \tag{C.10}$$

and this completes the proof of Proposition 3.1.

□

References

Andersen, P. K., Borgan, ϕ., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag: New York.

Berzuini, C. and Clayton, D. (1994). 'Baysian analysis of survival on multiple time scale'. *Statistics in Medicine* 13, 823-838.

Breslow, N. E. (1984). 'Extra-Poisson variation in log-linear models'. *Applied Statistics* 33, 38-44.

Dewandji, A., Goddard, M.J. and Moolgavkar, S.H. (1999). 'Two stage model for carcinogenesis: number and size distributions of premalignant clones in longitudinal studies'. *Mathematical Biosciences* 155, 1-12.

Hanayama, N. (2001). 'A Simple Model for Cancer Risks in the Environment.. *Environmetrics*, being prepared for publishing.

Holford, T. R. (1983). 'The estimation of age, period and cohort effects for vital rate'. *Biometrics* 39, 311-324.

Holford, T. R., Zhang, Z. and McKay, L. A. (1994). 'Estimating age, period and cohort effects using the multistage model for cancer'. *Statistics in Medicine* 13, 23-41.

Keiding, N. (1975). 'Maximum likelihood estimation in the birth-and-death process'. *Annals of Statistics* 3, 363-372.

Keiding, N. (1990). 'Statistical inference in the Lexis diagram'. *Philosophical Transactions of the Royal Society of London Series* A 332, 487-509.

Kripke, M.L. (1986). *The University of Texas M.D. Anderson Symposium on Fundamental Cancer Research vol. 38, Immunology and Cancer,* edited by M.L. Kripke and M.D.P Frost. University of Texas Press: Austin. 113-120.

Kuroki, T. et al. (1984). *An Outline of Medical Science vol. 5, Cancer,* Nakayama-shoten: Tokyo. (in Japanese).

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Model 2nd ed.,* Chapman & Hall: London.

Moolgavkar, S.H. and Venzon, D.J. (1979). 'Two event models for carcinogenesis: Incidence curves for childhood and adult tumors'. *Mathematical Biosciences* 47, 55.

Osmond, C. and J. Gardener M. (1982). 'Age, period and cohort models applied to cancer mortality'. *Statistics in Medicine* 1, 245-259.

Robertson, C. and Boyle, P. (1986). 'Age, period and cohort model : The use of individual record'. *Statistics in Medicine* 5, 527-538.

Robertson, C. and Boyle, P. (1998). 'Age-period-cohort Analysis of Chronic Disease Rates. I: Modeling Approach'. *Statistics in Medicine* 17, 1305-1323.

Tango, T. and Kurashina, S., (1987), Age period and cohort analysis of trends in mortality from major disease in Japan, 1955 to 1979: Peculiarity of the cohort born in the early Showa era,. *Statistics in Medicine* 6, 709-726.

Voet, D. and Voet, J. G. (1995). *Biochemistry,* John Wiley & Sons: New York.