Ann. Inst. Statist. Math.
35 (1983), Part A, 425-438

THE LEAST SQUARES ESTIMATION OF THE TRANSITION PROBABILITIES OF BINARY PROCESSES ON THE BASIS OF SAMPLE PATHS

HIROHISA KISHINO

(Received April 3, 1982; revised Sept. 6, 1983)

Summary

We consider the weighted least squares (WLS) estimation of the transition probabilities of binary processes on the basis of given sample paths in connection with log linear and logistic model analyses. We investigate, in particular, its effectiveness in the analyses supported by a Bayesian method with a smoothness prior over the time domain.

1. Introduction

Practically, we frequently face the data based on a sample from a binary process such as the pattern of giving birth to a child after marriage and the data of the spasms of asthmatic patients. The first can be formulated as a renewal process if the lengths of the intervals between births are of interest, and the second as a Markov chain, since patients are likely to have spasms if they have those on the previous day. In the latter case, the transition probability would vary in connection with seasonal fluctuations.

These data are formally expressed as the $N\times T$ matrix, $Z=\{z_t^{(n)};\ n=1,\cdots,N,\ t=1,\cdots,T\}$. Here, each element of the matrix takes the values 1 or 0, according as the event occurs or does not. Each row corresponds to a sample path. For given random matrix $\tilde{Z}=\{\tilde{z}_t^{(n)};\ n=1,\cdots,N,\ t=1,\cdots,T\}$ corresponding to Z, we estimate the transition probability

$$r_n(t, \boldsymbol{z}_{t-1}; \boldsymbol{\theta})=P_{\boldsymbol{\theta}}(\tilde{z}_t^{(n)}=1\,|\,\tilde{\boldsymbol{z}}_{t-1}^{(n)}=\boldsymbol{z}_{t-1}) .$$

Here $\boldsymbol{z}_{t-1}=(z_1,\cdots,z_{t-1})'$ and $\tilde{\boldsymbol{z}}_{t-1}^{(n)}=(z_1^{(n)},\cdots,z_{t-1}^{(n)})'$. $\boldsymbol{\theta}$ denotes an unknown parameter vector. Throughout this paper, $\tilde{\boldsymbol{z}}_T^{(n)}$ $(n=1,\cdots,N)$ are assumed to be mutually independent.

If the processes are Markovian and identically distributed to each others (i.e. $r_n(t, \boldsymbol{z}_{t-1})=r(t, z_{t-1})$—independent of n and $\boldsymbol{z}_{t-2}$—for $n=1,\cdots,N$, $t=2,\cdots,T$), Anderson and Goodman [2] discusses the maximum likelihood (ML) estimation of the transition probability. In the case where only the aggregate data $n_i(t)$ is at hand, Miller [10] and Madansky [9] get by the equality

$$(1) \qquad p_i(t)=\sum_{j=1}^{M} q_{ji}(t;\boldsymbol{\theta})p_j(t-1)$$

the regression model

$$(2) \qquad n_i(t)=\sum_{j=1}^{M} q_{ji}(t;\boldsymbol{\theta})n_j(t-1)+e_{ij}(t)\ ,$$

and estimate q_{ji}'s by the weighted least squares (WLS) method, since each sample path in this case cannot directly be observed. Here, $n_i(t)$ denotes the size of the sample on the state i at time t, and $q_{ji}(t)$ is the transition probability from j to i. However, it should be noticed that the equality (1) holds only when the process is Markovian. For many other binary processes, $p_i(t)$ cannot be expressed by the transition probabilities $q_{ji}(t;\boldsymbol{\theta})$ in the way stated in (1). Therefore the applicability of the method expressed in the model (2) is limited to the analysis of data for a Markov chain.

In this paper, we consider the WLS method for estimating the parameter $\boldsymbol{\theta}$ in the log linear and logistic models of the transition probabilities of various binary processes, when each sample path is observed. Under the situation where a sequence of sample paths can be used as a data, we estimate the parameter $\boldsymbol{\theta}$ using the log linear regression model for a general process,

$$(3) \qquad \log \hat{r}(t, \boldsymbol{z}_{t-1})=\log r(t, \boldsymbol{z}_{t-1};\boldsymbol{\theta})+e(t, \boldsymbol{z}_{t-1}) \quad \text{for } \boldsymbol{z}_{t-1}\text{'s},\ t=1,\cdots,T\,.$$

Here, $\hat{r}(t, \boldsymbol{z}_{t-1})$ is defined as $m(\boldsymbol{z}_{t-1}, 1)/m(\boldsymbol{z}_{t-1})$ and $m(\boldsymbol{z}_t)=m(z_1,\cdots, z_t)$ denotes the number of the sample paths in the vector expressions of which the first t elements are identical with the vector $\underline{z}_t$. In Section 2, we consider the asymptotic distribution of $\hat{r}(t, \boldsymbol{z}_{t-1})$ in the aforementioned model and also in the logistic model in order to make the applicability of these models clear. Further we obtain the asymptotic distribution of the estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. In Section 3, we treat the log linear and logistic models for multiple Markov chains. (We deal with renewal processes in [7]). In 3.1 and 3.2, we consider the asymptotic distributions of the estimators for the parameters in multiple Markov chains. In 3.3, we consider the non-parametric estimation of the transition probability $r(t, \boldsymbol{z}_k)$ of a non-homogeneous multiple Markov chain under the constraint that this transition probability changes slowly. We show that this approach is equivalent to the Byes method with a smoothness prior over the time domain. In this case the WLS method is also available to calculate the marginal likelihood of hyper-parameters. (Cf. Shiller [11], Akaike [1] and Ishiguro and Akaike [6]). We have that the covariance matrices of the error terms of the regression models are of the almost common form in all these binary processes. Finally, in 3.4, a numerical example is given.

2. The asymptotic distribution of $\hat{r}(t, z_{t-1})$ and the log linear models

We first show that the asymptotic distribution of $(\log \hat{r}(t, z_{t-1}); t, z_{t-1})'$ as stated in Section 1 is a normal distribution and its covariance matrix is diagonal. This result makes it possible to use the WLS method for the estimation of the parameter $\boldsymbol{\theta}$. Put $\hat{p}(z_t)=m(z_t)/N$, and let $\hat{\boldsymbol{p}}$ denote, for short, the $(2^{T+1}-2)$-dimensional vector $(\hat{p}(z_t)$: for all z_t; $t=1,\cdots,T)'=(\hat{p}(0), \hat{p}(1), \hat{p}(00), \hat{p}(01), \hat{p}(10), \hat{p}(11),\cdots,\hat{p}(11\cdots1))'$. The $2^{T+1}-1$ dimensional vector $\boldsymbol{p}$ is defined by $(p(z_t)$: for all z_t; $t=1,\cdots,T)'$ in the same way where $p(z_t)$ is $P(\tilde{z}_t=z_t)$. Further we write the vector $(\log r(t, z_{t-1})$: for all z_{t-1}, $t=1,\cdots,T)'$ $(=(\log r(1), \log r(2, (0)), \log r(2, (1)), \cdots, \log r(T, (11\cdots1)')))')$ as $\mathbf{log}\,(\boldsymbol{r})$, and $\mathbf{log}\,(\hat{\boldsymbol{r}}, \boldsymbol{r})$ is defined as $\mathbf{log}\,(\hat{\boldsymbol{r}})-\mathbf{log}\,(\boldsymbol{r})$.

First, we consider a natural extension of the asymptotic property of the multinomial variate for $N\hat{\boldsymbol{p}}$. For vectors $z_t^{[1]}$ and $z_s^{[2]}$ $(t\leqq s)$ we write $z_t^{[1]}\leqq z_s^{[2]}$, if the first t elements of $z_s^{[2]}$ are the same as the elements of $z_t^{[1]}$. Also let

$$\delta(z_t^{[1]}\leqq z_s^{[2]})=\begin{cases} 1 & \text{if } z_t^{[1]}\leqq z_s^{[2]} \\ 0 & \text{otherwise}\,. \end{cases}$$

Theh we have:

LEMMA.

$$\sqrt{N}\,(\hat{\boldsymbol{p}}-\boldsymbol{p})\underset{d}{\rightarrow} N(\boldsymbol{0}, \Sigma) \qquad (N\rightarrow\infty)\,, \tag{4}$$

where each $(z_t^{[1]}, z_s^{[2]})$*-component* $(t\leqq s)$ *of the matrix* Σ *is expressed as*

$$N\,\mathrm{Cov}\,(\hat{p}(z_t^{[1]}), \hat{p}(z_s^{[2]}))=\delta(z_t^{[1]}\leqq z_s^{[2]})p(z_s^{[2]})-p(z_t^{[1]})p(z_s^{[2]})\,. \tag{5}$$

PROOF. Since

$$\hat{p}(z_t)=\frac{1}{N}\sum_{n=1}^{N}\delta_{z_t^{(n)}, z_t} \tag{6}$$

where

$$\delta_{z_t^{[1]}, z_t^{[2]}}=\begin{cases} 1 & \text{if } z_t^{[1]}=z_t^{[2]} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathrm{E}\,[\hat{p}(z_t)]=p(z_t)\ ,$$

the convergence follows directly from the central limit theorem. So it is sufficient for the proof to find the covariance Σ. For given $z_t^{[1]}$ and $z_s^{[2]}$ $(t\leqq s)$, we have that

$$\mathrm{E}\,[\delta_{\tilde{z}_t z_t^{[1]}}\delta_{\tilde{z}_s z_s^{[2]}}]=\delta(z_t^{[1]}\leqq z_s^{[2]})p(z_s^{[2]})\ ,$$

and so

$$\mathrm{Cov}\,(\delta_{\tilde{z}_t z_t^{[1]}},\ \delta_{\tilde{z}_s z_s^{[2]}})=\delta(z_t^{[1]}\leqq z_s^{[2]})p(z_t^{[1]})-p(z_t^{[1]})p(z_s^{[2]})\ .$$

This implies the equation (2) stated in the lemma.

Then we get;

THEOREM 2.1.

$$\sqrt{N}\,\mathbf{log}\,(\hat{\boldsymbol{r}},\ \boldsymbol{r})\underset{d}{\rightarrow}N(0,\ \Sigma_0)\qquad(N\rightarrow\infty)\ , \tag{7}$$

where the matrix Σ_0 is diagonal and its $(t,\ z_{t-1})$th diagonal element is

$$\sigma^2(t,\ z_{t-1})=(p(z_{t-1},\ 1))^{-1}-(p(z_{t-1}))^{-1}=(p(z_{t-1}))^{-1}((r(t,\ z_{t-1}))^{-1}-1)\ . \tag{8}$$

Remark. The result stated in (4) and (5) can be easily extended to a general finite space, but is not mentioned here.

PROOF OF THEOREM 2.1. We have by Taylor's formula that

$$\log\hat{p}(z_t)=\log p(z_t)+f(z_t)+\varepsilon(z_t)\ ,$$

where

$$f(z_t)=(p(z_t))^{-1}(\hat{p}(z_t)-p(z_t))\quad\text{and}\quad\sqrt{N}\,\varepsilon(z_t)\rightarrow O\ ,\qquad(N\rightarrow\infty)\ .$$

Hence

$$\begin{aligned}\log\hat{r}(t,\ z_{t-1})=\log r(t,\ z_{t-1})&+f(z_{t-1},\ 1)-f(z_{t-1})\\&+\varepsilon(z_{t-1},\ 1)-\varepsilon(z_{t-1})\ .\end{aligned} \tag{9}$$

Let $g(z_t)$ be defined as $\sqrt{N}\{f(z_t,1)-f(z_t)\}$. By the lemma, we have only to check the covariance structure of $g(z_t)$'s.

We can assume $t\leqq s$ without the loss of generality. Since

$$\begin{aligned}\mathrm{Cov}\,\{g(z_t^{[1]}),\ g(z_s^{[2]})\}=N\{&\mathrm{E}\,((f(z_t^{[1]},\ 1)f(z_s^{[2]},\ 1))-\mathrm{E}\,(f(z_t^{[1]},\ 1)f(z_s^{[2]}))\\&-\mathrm{E}\,(f(z_t^{[1]})f(z_s^{[2]},\ 1))+\mathrm{E}\,(f(z_t^{[1]})f(z_s^{[2]}))\}\ ,\end{aligned}$$

we write the right-hand side of the abovemensioned equality as $I_1-I_2-I_3+I_4$.

Noticing that we have by the lemma

$$\mathrm{Cov}\,\{f(z_t^{[1]}),\, f(z_s^{[2]})\} = \frac{1}{N}\{\delta(z_t^{[1]} \leqq z_s^{[2]})(p(z_t^{[1]}))^{-1} - 1\}$$

$$\mathrm{E}\,\{f(z_t)\} = 0\ ,$$

we calculate the terms I_1, I_2, I_3 and I_4 in the following way;

Case 1 (when $t<s$).

(1a) We first show the case where $z_t^{[1]} \leqq z_s^{[2]}$:

(1a$_1$) When $(z_t^{[1]}, 1) \leqq z_s^{[2]}$,

$$I_1 = I_2 = (p(z_t^{[1]}, 1))^{-1} - 1\ , \qquad I_3 = I_4 = (p(z_t^{[1]}))^{-1} - 1\ .$$

(1a$_2$) When $(z_t^{[1]}, 1) \lneqq z_s^{[2]}$ (i.e. $z_{t+1}^{[2]} = 0$),

$$I_1 = I_2 = -1\ , \qquad I_3 = I_4 = (p(z_t^{[1]}))^{-1} - 1\ .$$

(1b) In the case where $z_t^{[1]} \lneqq z_s^{[2]}$:

$$I_1 = I_2 = I_3 = I_4 = -1\ .$$

Case 2 (when $t=s$).

(2a) When $z_t^{[1]} = z_s^{[2]} = z_t$,

$$I_1 = (p(z_t, 1))^{-1} - 1\ , \qquad I_2 = I_3 = I_4 = (p(z_t))^{-1} - 1\ ,$$

(2b) When $z_t^{[1]} \neq z_t^{[2]}$,

$$I_1 = I_2 = I_3 = I_4 = -1\ .$$

These equalities give the covariance of $g(z_t)$.

Thus the proof of this theorem is complete.

In order to estimate the parameter $\boldsymbol{\theta}$, we can use the WLS method for the model;

$$\log \hat{r}(t, z_{t-1}) = \log r(t, z_{t-1}; \boldsymbol{\theta}) + e(t, z_{t-1})\ ,$$

as the asymptotic normality of $\boldsymbol{e}$ and the diagonality of its covariance matrix are provided in Theorem 2.1. We shall give an example of the above method for the birth process after marriage. Let $h(t; a, b, c)$ be defined as

$$h(t; a, b, c) = at^b \exp(-ct)\ . \tag{10}$$

The peaks of these curves are atained at $t=b/c$. Let $e_{k,t}^{i_1,\cdots,i_k}$ be the t-dimensional vector whose $i_1, \cdots, i_k$-components are 1 and the others are 0 as $e_{k,t}^{i_1,\cdots,i_k} = (0\cdots0\overset{i_1}{\check{1}}0\cdots0\overset{i_2}{\check{1}}0\cdots0\overset{i_k}{\check{1}}0\cdots0)'$. When $k=0$, $\boldsymbol{e}_{0,t} = \boldsymbol{0}$ and $i_0 = 0$. The length of intervals between births is important for the birth process, so we consider the models of the types;

(i) $r(t, e_{k,t-1}^{i_1,\cdots,i_k})=h(t-i_k; a_k, b_k, c_k)$
$(1\leqq i_1<\cdots<i_k<t\leqq T,\ k=0, 1,\cdots, M)$,

(ii) $r(t, e_{k,t-1}^{i_1,\cdots,i_k})=h(t-i_k; a_{k,i_k}, b_{k,i_k}, c_{k,i_k})$.

For example, we set $M=2$ if we want to know about 1st, 2nd and 3rd birth.

Taking Theorem 2.1 into account, we have the following WLS estimation method; For the first model (i),

$$\hat{\boldsymbol{\theta}}_k=\hat{A}^{-1}\hat{\boldsymbol{x}}\ , \tag{11}$$

where

$$\boldsymbol{\theta}_k=(\tilde{a}_k, b_k, c_k)'\ ,\qquad \tilde{a}_k=\log a_k\ ,$$

$$\hat{A}=\sum_{i_1,\cdots,i_k,t}(\hat{\sigma}^2(t, e_{k,t-1}^{i_1,\cdots,i_k}))^{-1}\boldsymbol{d}(t-i_k)\boldsymbol{d}(t-i_k)'$$

$$\hat{\boldsymbol{x}}=\sum_{i_1,\cdots,i_k,t}(\hat{\sigma}^2(t, e_{k,t-1}^{i_1,\cdots,i_k}))^{-1}(\log\hat{r}(t, e_{k,t-1}^{i_1,\cdots,i_k}))\boldsymbol{d}(t-i_k)\ ,$$

$$\boldsymbol{d}(t)=(1, \log t, -t)'\ ,$$

$$\hat{\sigma}^2(t, \boldsymbol{z}_{t-1})=(\hat{p}(\boldsymbol{z}_{t-1}, 1))^{-1}-(\hat{p}(\boldsymbol{z}_{t-1}))^{-1}$$

and the summation is taken for all the quantities $i_1,\cdots, i_k$ and t for $i_1<\cdots<i_k<t\leqq T$.

3. In the case of multiple Markov chains

When T becomes large, $P(\tilde{\boldsymbol{z}}_{t-1}=\boldsymbol{z}_{t-1})$ decreases and it is seen from (8) that the estimators expressed in the preceding section become unstable. Therefore, when the character of a process—for example, Markovian property or renewal property—is specified from another information, the WLS method stated in the same section should be accomodated to the formulation of this specification. This corresponds to the grouping in the contingency table analysis. As an example, we get into details for a Markov chain in the following subsections.

Here we write

$$r_n(t, \boldsymbol{z}_{t-1})=P(\tilde{z}_t^{(n)}=1\,|\,\tilde{\boldsymbol{z}}_{t-1}^{(n)}=\boldsymbol{z}_{t-1})\ . \tag{12}$$

If the process is k-multiple Markovian, then the $r_n(t, \boldsymbol{z}_{t-1})$ can be expressed as

$$r_n(t, \boldsymbol{z}_{t-1})=r_n(t, \boldsymbol{z}_{t-1}^k)\ , \tag{13}$$

where

$$\boldsymbol{z}_{t-1}^k=(z_{t-k},\cdots, z_{t-1})'\ .$$

3.1. *The time-homogeneous case*

If the process is time-homogeneous, the model is formulated as

$$r_n(t, \boldsymbol{z}_k; \boldsymbol{\theta}) = r_n(\boldsymbol{z}_k; \boldsymbol{\theta}) \qquad \text{for all } n,\ t, \text{ and } \boldsymbol{z}_k\text{'s}, \tag{14}$$

in the k-multiple Markovian case, where $\boldsymbol{\theta}$ is an unknown parameter. Here and hereafter, in the description of $r_n(t, \boldsymbol{z}_k)$ we abbreviate the letter $\boldsymbol{\theta}$ for simplicity. Writing the k dimensional vector $(z_{t-k}^{(n)}, \cdots, z_{t-1}^{(n)})'$ as $\boldsymbol{z}_{t-1}^{k,(n)}$, let

$$m_1(n; \boldsymbol{z}_k) = \frac{1}{T-k} \sum_{t=k+1}^{T} \delta(\boldsymbol{z}_{t-1}^{k,(n)}, \boldsymbol{z}_k),$$

$$m_1(n; \boldsymbol{z}_k, 1) = \frac{1}{T-k} \sum_{t=k+1}^{T} \delta(\boldsymbol{z}_{t-1}^{k,(n)}, \boldsymbol{z}_k) z_t,$$

and

$$\hat{r}_n(\boldsymbol{z}_k) = m_1(n; \boldsymbol{z}_k, 1)/m_1(n; \boldsymbol{z}_k),$$

where

$$\delta(\boldsymbol{z}_{t-1}^k, \boldsymbol{z}_k) = \begin{cases} 1 & \text{if } \boldsymbol{z}_{t-1}^k = \boldsymbol{z}_k, \\ 0 & \text{otherwise}. \end{cases}$$

Then we get the $2^k \times N$ dimensional vector $(\log \hat{r}_n(\boldsymbol{z}_k)/r_n(\boldsymbol{z}_k);\ n=1, \cdots, N,\ \boldsymbol{z}_k\text{'s})' = (\boldsymbol{s}_1', \cdots, \boldsymbol{s}_N')'$ where $\boldsymbol{s}_n = (\log \hat{r}_n(00\cdots0)/r_n(00\cdots0), \log \hat{r}_n(0\cdots01)/r_n(0\cdots01), \cdots, \log \hat{r}_n(11\cdots1)/r_n(11\cdots1))'$. From now on, some ordering is assumed in the same way in the similar expressions. Then, we have that,

THEOREM 3.1. *The statistic $\hat{r}_n(\boldsymbol{z}_k)$ is a consistent estimator of $r_n(\boldsymbol{z}_k)$ and*

$$\sqrt{T}(\log \hat{r}_n(\boldsymbol{z}_k)/r_n(\boldsymbol{z}_k);\ n=1, \cdots, N,\ \boldsymbol{z}_k\text{'s})' \xrightarrow[d]{} N(\boldsymbol{0}, \Sigma_2) \qquad (T \to \infty) \tag{15}$$

where Σ_2 is the diagonal matrix whose diagonal elements $\sigma_n^2(\boldsymbol{z}_k)$ are expressed as

$$\sigma_n^2(\boldsymbol{z}_k) = \frac{1}{Em_1(n; \boldsymbol{z}_k, 1)} - \frac{1}{Em_1(n; \boldsymbol{z}_k)} + o(1). \tag{16}$$

PROOF. Since $(\hat{r}_n(\boldsymbol{z}_k); \boldsymbol{z}_k)'$ $(n=1, \cdots, N)$ are mutually independent, we consider $\hat{r}_n(\boldsymbol{z}_k)$ for a fixed n. By the Taylor's theorem,

$$\log \hat{r}_n(\boldsymbol{z}_k) = \log r_n(\boldsymbol{z}_k) + \frac{m_1(n; \boldsymbol{z}_k, 1) - Em_1(n; \boldsymbol{z}_k, 1)}{Em_1(n; \boldsymbol{z}_k, 1)} - \frac{m_1(n; \boldsymbol{z}_k) - Em_1(n; \boldsymbol{z}_k)}{Em_1(n; \boldsymbol{z}_k)} + \varepsilon_n(\boldsymbol{z}_k).$$

Writing the second and third terms of the right-hand side as $f_{1n}(z_k)$ and $f_{2n}(z_k)$, we see that $\varepsilon_n(z_k)=(1/2)(f_{1n}^2(z_k)-f_{2n}^2(z_k))h$ $(0<h<1)$, so $\sqrt{N}\cdot\varepsilon_n(z_k)\to 0$ $(N\to\infty)$. The consistency and the asymptotic normality of $\hat{r}_n(z_k)$ follow directly from the law of large numbers and the central limit theorem for Markov processes (Doob [4], Chapter V, Theorem 6.2, 7.5—Considering the process $\tilde{v}_t^{(n)}=\tilde{z}_t^{(n)}=(\tilde{z}_t^{(n)},\cdots,\tilde{z}_{t-k+1}^{(n)})'$, which is Markovian, we can apply these theorems to our case). Therefore let us express the covariances. In the above equation, $f_{1n}(z_k)-f_{2n}(z_k)$ are combiend to

$$\frac{m_1(n;z_k,1)-r_n(z_k)m_1(n;z_k)}{\mathrm{E}\,m_1(n;z_k,1)}=\frac{w(n;z_k)}{\mathrm{E}\,m_1(n;z_k,1)}\ .$$

From now on, we delete the symbol n for simplicity.

(i) In the case where $z_k^{[1]}\neq z_k^{[2]}$;

$$\begin{aligned}\mathrm{E}\,w(z_k^{[1]})w(z_k^{[2]})&=\left(\frac{1}{T-k}\right)^2\mathrm{E}\left[\sum_{t=k+1}^{T}\sum_{s=k+1}^{T}\delta(\tilde{z}_{t-1}^k,z_k^{[1]})\right.\\&\qquad\left.\times\delta(\tilde{z}_{s-1}^k,z_k^{[2]})(\tilde{z}_t-r(z_k^{[1]}))(\tilde{z}_s-r(z_k^{[2]}))\right]\\&=\left(\frac{1}{T-k}\right)^2\left\{\sum_{t=k+1}^{T}\mathrm{E}\,[\delta(\tilde{z}_{t-1}^k,z_k^{[1]})\delta(\tilde{z}_{t-1}^k,z_k^{[2]})\right.\\&\qquad\times(\tilde{z}_t-r(z_k^{[1]}))(\tilde{z}_t-r(z_k^{[2]}))]\\&\qquad+\sum_{t\neq s}\sum\mathrm{E}\,[\delta(\tilde{z}_{t-1}^k,z_k^{[1]})\delta(\tilde{z}_{s-1}^k,z_k^{[2]})\\&\qquad\left.\times(\tilde{z}_t-r(z_k^{[1]}))(\tilde{z}_s-r(z_k^{[2]}))]\right\}\ .\end{aligned}$$

The first term of the last side is 0 because $\delta(\tilde{z}_{t-1}^k,z_k^{[1]})\delta(\tilde{z}_{t-1}^k,z_k^{[2]})=0$. Concerning to the second term, for $t<s$,

$$\begin{aligned}&\mathrm{E}\,[\delta(\tilde{z}_{t-1}^k,z_k^{[1]})\delta(\tilde{z}_{s-1}^k,z_k^{[2]})(\tilde{z}_t-r(z_k^{[1]}))(\tilde{z}_s-r(z_k^{[2]}))\,|\,\tilde{z}_{s-1}]\\&=\delta(\tilde{z}_{t-1}^k,z_k^{[1]})\delta(\tilde{z}_{s-1}^k,z_k^{[2]})(\tilde{z}_t-r(z_k^{[1]}))\,\mathrm{E}\,[\tilde{z}_s-r(z_k^{[2]})\,|\,\tilde{z}_{s-1}]\\&=\delta(\tilde{z}_{t-1}^k,z_k^{[1]})(\tilde{z}_t-r(z_k^{[1]}))\delta(\tilde{z}_{s-1}^k,z_k^{[2]})(r(\tilde{z}_{s-1}^k)-r(z_k^{[2]}))\ ,\end{aligned}$$

which is equal to 0 from the last two factors. The situation is similar for $t>s$. Therefore, we have that $\mathrm{E}\,w(z_k^{[1]})w(z_k^{[2]})=0$.

(ii) In the case where $z_k^{[1]}=z_k^{[2]}=z_k$; Similarly we have that

$$\begin{aligned}\mathrm{E}\,w^2(z_k)&=\left(\frac{1}{T-k}\right)^2\sum_{t=k+1}^{T}\mathrm{E}\,[\delta(\tilde{z}_{t-1}^k,z_k)(\tilde{z}_t-r(z_k))^2]\\&=\frac{1}{T-k}\{\mathrm{E}\,m_1(z_k,1)-2r(z_k)\,\mathrm{E}\,m_1(z_k,1)+r^2(z_k)\,\mathrm{E}\,m_1(z_k)\}\\&=\frac{1}{T-k}\{\mathrm{E}\,m_1(z_k,1)-r(z_k)\,\mathrm{E}\,m_1(z_k,1)\}\ .\end{aligned}$$

These results shown in (i) and (ii) conclude the proof.

As in the preceding section, we consider the following expression;

$$\sqrt{T}\log\hat{r}_n(z_k)=\sqrt{T}\log r_n(z_k;\theta)+e_n(z_k) \qquad n=1,\cdots,N, \text{ for all } z_k, \tag{17}$$

where the variance-covariance matrix of e is Σ_2 in Theorem 3.1. From the normality of e and the diagonality of Σ_2, we can consider the WLS method for the model (17). In this case, Σ_2 is replaced by its estimator $\hat{\Sigma}_2$ where $m_1(n;z_k)$ and $m_1(n;z_k,1)$ are substituted for $\mathrm{E}\,m_1(n;z_k)$ and $\mathrm{E}\,m_1(n;z_k,1)$, respectively.

For the logistic type model, we provide the following proposition. Let $r'_n(z_k)=\log r_n(z_k)/(1-r_n(z_k))$ and $\hat{r}'_n(z_k)=\log\hat{r}_n(z_k)/(1-\hat{r}_n(z_k))$.

PROPOSITION 3.1. The statistic $\hat{r}'_n(z_k)$ is a cosistent estimator of $r'_n(z_k)$ and

$$\sqrt{T}(\hat{r}'_n(z_k)-r'_n(z_k);n,z_k)'\xrightarrow[d]{}N(0,\Sigma_3) \qquad (T\to\infty). \tag{18}$$

Here Σ_3 is the diagonal matrix whose diagonal elements are

$$\sigma_n^{2\prime}(z_k)=(\mathrm{E}\,m_1(n;z_k,1))^{-1}+(\mathrm{E}\,m_1(n;z_k,0))^{-1}+o(1). \tag{19}$$

3.2. *The identically distributed case*

In the identically distributed case, the model is formulated as

$$r_n(t,z_k;\theta)=r(t,z_k;\theta) \qquad \text{for all } n,\ t \text{ and } z_k\text{'s} \tag{20}$$

for the k-multiple Markovian chain.

Here we consider the following statistics:

$$\hat{r}(t,z_k)=m_2(t;z_k,1)/m_2(t;z_k),$$

where

$$m_2(t;z_k,1)=\frac{1}{N}\sum_{n=1}^{N}\delta(z_{t-1}^{k,(n)},z_k)z_t^{(n)}$$

$$m_2(t;z_k)=\frac{1}{N}\sum_{n=1}^{N}\delta(z_{t-1}^{k,(n)},z_k).$$

Then we have:

THEOREM 3.2. *The statistic $\hat{r}(t,z_k)$ is a consistent estimator of $r(t,z_k)$ and*

$$\sqrt{N}(\log\hat{r}(t,z_k)/r(t,z_k);t,z_k)'\xrightarrow[d]{}N(0,\Sigma_4) \qquad (N\to\infty). \tag{21}$$

Here Σ_4 is the diagonal matrix whose diagonal elements are

$$\sigma^2(t, z_k)=(P(\tilde{z}_{t-1}^k=z_k, \tilde{z}_t=1))^{-1}-(P(\tilde{z}_{t-1}^k=z_k))^{-1} . \tag{22}$$

PROOF. We have

$$\log \hat{r}(t, z_k)=\log r(t, z_k)+w(t, z_k)/P(\tilde{z}_{t-1}^k=z_k, \tilde{z}_t=1)+\varepsilon(t, z_k)$$

where

$$w(t, z_k)=m_2(t; z_k, 1)-r(t, z_k)m_2(t; z_k) ,$$

and

$$\sqrt{N}\,\varepsilon(t, z_k)\to 0 \qquad (N\to\infty) .$$

From the analogous discussion to Theorem 3.1, the result follows.

The result stated in Theorem 3.2 leads to the regression analysis as follows;

$$\sqrt{N}\log \hat{r}(t, z_k)=\sqrt{N}\log r(t, z_k; \boldsymbol{\theta})+e(t, z_k) ,$$

where the covariance matrix of $\boldsymbol{e}$ is Σ_4.

For the logistic type model, let $\hat{r}'(t, z_k)=\log \hat{r}(t, z_k)/(1-\hat{r}(t, z_k))$. Then,

PROPOSITION 3.2. *The statistic $\hat{r}'(t, z_k)$ is a consistent estimator of $r'(t, z_k)$, and*

$$\sqrt{N}(\hat{r}'(t, z_k)-r'(t, z_k); t, z_k)' \xrightarrow[d]{} N(0, \Sigma_5) \qquad (N\to\infty) . \tag{23}$$

Here Σ_5 is diagonal, and the diagonal elements are

$$\sigma^2(t, z_k)=(P(\tilde{z}_{t-1}^k=z_k, \tilde{z}_t=1))^{-1}+(P(\tilde{z}_{t-1}^k=z_k, \tilde{z}_t=0))^{-1} . \tag{24}$$

3.3. *An application of the WLS method to a Bayes estimation with a smoothness prior over the time domain*

In the case where we can suppose that the $r(t, z_k)$ stated in Subsection 3.2 changes slowly as we, for example, have to take a seasonal influence to the process into account, the following cost function would be reasonable to be considered (Shiller [11], Ishiguro and Akaike [6]);

$$L(z_k)=\sum_{t=k+1}^{T} N_t/\sigma^2(t, z_k)(\log \hat{r}(t; z_k)-\log r(t; z_k))^2 + a^2\sum_{t=k+2}^{T}(\log r(t; z_k)-\log r(t-1; z_k))^2 . \tag{25}$$

Here, $\log r(t; z_k)$ $t=k+1,\cdots, T$ are unknown parameters to be estimated, and the second term stands for the constraint that the transition probability changes slowly. The logistic type model is similarly constructed. The problem of minimizing $L(z_k)$ is equivalent to that of minimizing the following formulation;

$$(26)\quad \begin{aligned} &\log \hat{r}(t; z_k)=\log r(t; z_k)+e_1(t; z_k) ,\\ &\log r(t; z_k)=\log r(t-1; z_k)+e_2(t; z_k) \qquad t=(k+1,\cdots,T) , \end{aligned}$$

or the one is the vector form,

$$\log \hat{\boldsymbol{r}}(z_k)=\log \boldsymbol{r}(z_k)+\boldsymbol{e}_1(z_k) \qquad \boldsymbol{O}=D \log \boldsymbol{r}(z_k)+\boldsymbol{e}_2(z_k) ,$$

where $\log \boldsymbol{r}(z_k)=(\log r(t, z_k),\ t=k+1,\cdots,T)$ and D is the $(T-k-1)\times(T-k)$ matrix

$$\begin{bmatrix} 1 & -1 & 0 & \\ & \ddots & \ddots & \\ 0 & & 1 & -1 \end{bmatrix} .$$

Also, $\boldsymbol{e}_1(z_k)$ and $\boldsymbol{e}_2(z_k)$ are the mutually independent and normally distributed variables whose covariance matrices are $\Sigma_4^N(z_k)$ and a^2I, respectively. Here, $\Sigma_4^N(z_k)$ corresponds to Σ_4 and the diagonal elements are

$$\sigma^2(t, z_k)=1/N_t(P(\tilde{z}_{t-1}^k=z_k, \tilde{z}_t=1)^{-1}-P(\tilde{z}_{t-1}^k=z_k)^{-1}) .$$

For simplicity, we write $\Sigma_4^N(z_k)$ by the same form as Σ_4. We here notice that the probability distribution of $\boldsymbol{e}_2(z_k)$ stands for a smoothness prior over the time domain and a stands for a hyper-parameter (see Akaike [1] and Ishiguro and Akaike [6]).

Once a is determined, the Bayes estimators are obtained by the WLS method for

$$\widehat{(\log \boldsymbol{r}(z_k))}=(\hat{\Sigma}_4^{-1}+a^2D'D)^{-1}\hat{\Sigma}_4^{-1}\log \hat{\boldsymbol{r}}(z_k) .$$

Here, we note that the diagonal elements of $\hat{\Sigma}_4$ are

$$\hat{\sigma}^2(t, z_k)=1/N_t(m_4(t, z_k, 1)^{-1}-m_4(t-1, z_k)^{-1}) ,$$

and the value of a is selected as the one maximizing the marginal likelihood

$$(27)\qquad L'(\log \hat{\boldsymbol{r}}\,|\,a)=\int f(\log \hat{\boldsymbol{r}}\,|\log \boldsymbol{r})g(\log \boldsymbol{r}\,|\,a)d(\log \boldsymbol{r})$$

(see also Akaike [1] and Ishiguro and Akaike [6]), where f and g are the density functions of $\boldsymbol{e}_1(z_k)$, $\boldsymbol{e}_2(z_k)$, respectively. As f and g are normally distributed, the calculation of the right-hand side of the expression (27) is feasible. In fact, we have

$$(28)\quad \log L'(\log \hat{\boldsymbol{r}}(z_k)\,|\,a)$$
$$=\text{const.}-\frac{1}{2}\{\widehat{(\log \boldsymbol{r}(z_k))}-\log \hat{\boldsymbol{r}}(z_k)\}'\hat{\Sigma}_4^{-1}\{\widehat{(\log \boldsymbol{r}(z_k)}-\log \hat{\boldsymbol{r}}(z_k)\}$$

$$+(T-k-1)\log a-\frac{1}{2}\log\det\hat{\Sigma}_4-\frac{1}{2}\log\det(\hat{\Sigma}_4^{-1}+a^2D'D)\ .$$

3.4. *A numerical example*

Here, we give a numerical example for Section 3.2.

Let $N=50$, $T=200$, $k=1$,

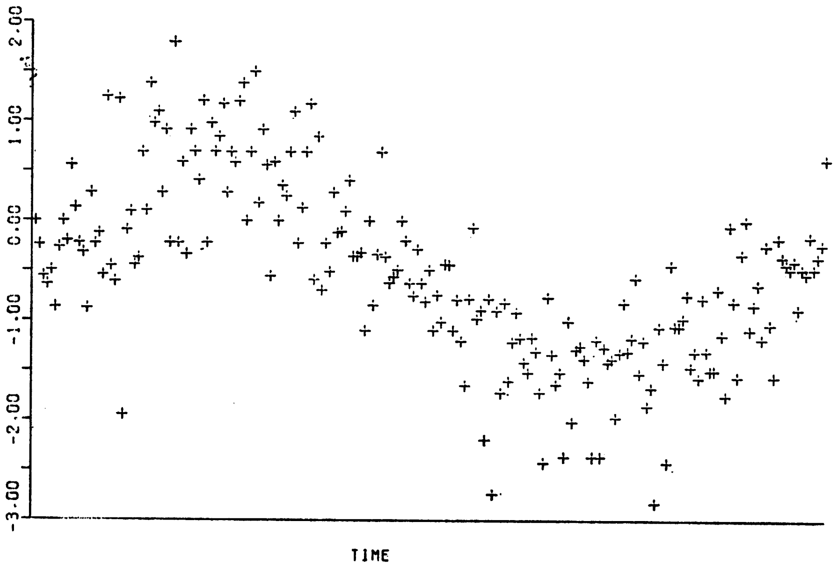

Fig. 1 $\hat{r}'(t,0)=\log\hat{r}(t,0)/(1-\hat{r}(t,0))$.

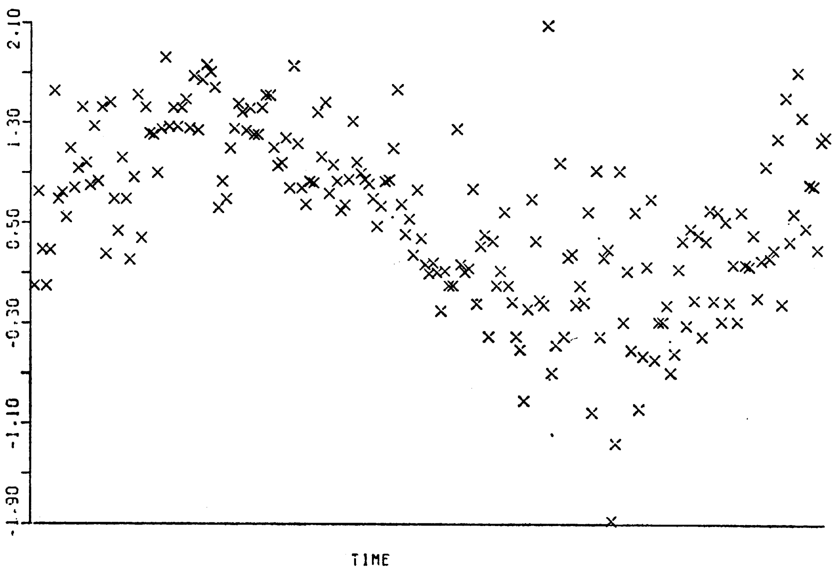

Fig. 2 $\hat{r}'(t,1)=\log\hat{r}(t,1)/(1-\hat{r}(t,1))$.

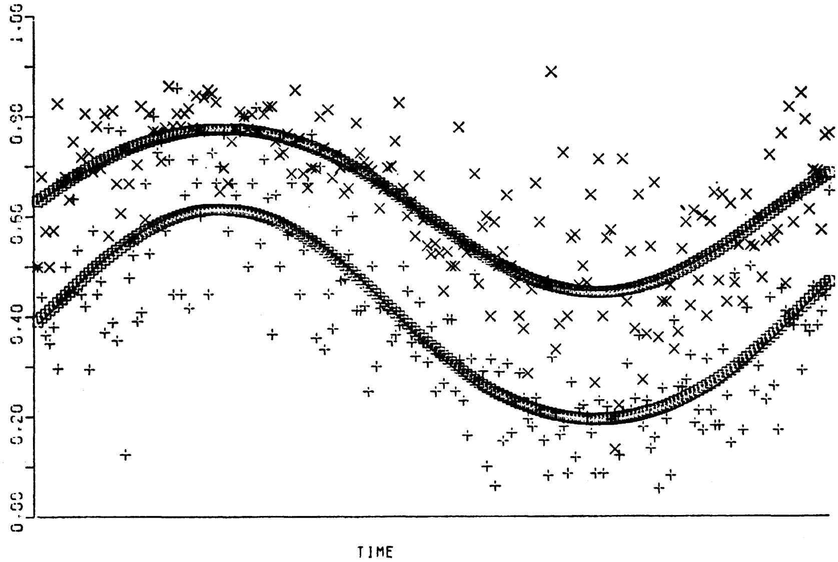

Fig. 3 $r(t,i;\boldsymbol{\theta})$, $r(t,i;\hat{\boldsymbol{\theta}})$ and $\hat{r}(t,i)$ $(i=0,1)$.

——: $r(t,i;\hat{\boldsymbol{\theta}})$

⊓⊓⊓: $r(t,i;\hat{\boldsymbol{\theta}})$ $i=0,1$

+: $\hat{r}(t,0)$

×: $\hat{r}(t,1)$

$$r(t,0;\boldsymbol{\theta})=1/(1+\exp\{-(\theta_{00}+\theta_{01}\sin t/30)\}) ,$$

$$r(t,1;\boldsymbol{\theta})=1/(1+\exp\{-(\theta_{10}+\theta_{11}\sin t/30)\}) ,$$

where

$$\theta_{00}=-0.5 , \quad \theta_{01}=1.0 , \quad \theta_{10}=0.5 , \quad \theta_{11}=0.7 .$$

Figs. 1 and 2 shows $\hat{r}'(t,0)=\log \hat{r}(t,0)/(1-\hat{r}(t,0))$ and $\hat{r}'(t,1)$.

Using Proposition 3.2, we estimate $\boldsymbol{\theta}$ by

$$\hat{\theta}_{00}=-0.475 , \quad \hat{\theta}_{01}=0.943 , \quad \hat{\theta}_{10}=0.513 , \quad \hat{\theta}_{11}=0.723 .$$

Fig. 3 shows $r(t,z_{t-1})$, $\hat{r}(t,z_{t-1})$ and their estimates $r(t,z_{t-1};\hat{\boldsymbol{\theta}})$. We see there that the structure of these transition probabilities is well estimated.

Acknowledgement

The author wishes to acknowledge his indebtedness to Dr. K. Noda and to the referees for their valuable suggestions for improving the presentation.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1980). Likelihood and Bayes procedure in *Bayesian Statistics*, (eds. J. M. Bernado, M. H. De Groot, D. U. Lindley and A. F. M. Smith), University Press, Valencia, Spain.

[2] Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chain, *Ann. Math. Statist.*, **28**, 89-110.

[3] Basawa, I. V. and Rao, B. L. S. P. (1980). *Statistical Inference for Stochastic Processes*, Academic Press, London.

[4] Doob, J. L. (1953). *Stochastic Processes*, John Wiley and Sons, New York.

[5] Imrey, P. B., Koch, G. G. and Stokes, M. E. (1981). Categorical data analysis; some reflections on the log linear model and logistic regression, Part I; Historical and methodological overview, *Int. Statist. Rev.*, **49**, 265-283.

[6] Ishiguro, M. and Akaike, H. (1980). Trading day adjustment for the Beyesian seasonal adjustment program BAYSEA, *Research Memorandom*, No. 189, The Institute of Statistical Mathematics, Tokyo.

[7] Kishino, H. (1982). Statistical analysis of sample paths, *Research Memorandom*, No. 229, The Institute of Statistical Mathematics, Tokyo.

[8] Lee, T. C., Judge, G. G. and Zellner, A. (1970). Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North Holland, Amsterdam.

[9] Madansky, A. (1959). Least squares estimation in finite Markov processes, *Psychometrika*, **24**, 137-144.

[10] Miller, G. A. (1952). Finite Markov processes in psychology, *Psychometrika*, **17**, 149-167.

[11] Shiller, R. J. (1973). A distributed lag estimation derived from smoothness priors, *Econometrica*, **41**, 775-788.