The Korean Journal of Applied Statistics (2020) DOI: https://doi.org/10.5351/KJAS.2020.33.1.047
33(1), 47–59

Bayesian inference of longitudinal Markov binary regression models with *t*-link function

Bohyun Sim[a] · Younshik Chung[a,1]

[a]Department of Statistics, Pusan National University

(Received November 21, 2019; Revised December 30, 2019; Accepted January 8, 2020)

Abstract

In this paper, we present the longitudinal Markov binary regression model with t-link function when its transition order is known or unknown. It is assumed that logit or probit models are considered in binary regression models. Here, t-link function can be used for more flexibility instead of the probit model since the t distribution approaches to normal distribution as the degree of freedom goes to infinity. A Markov regression model is considered because of the longitudinal data of each individual data set. We propose Bayesian method to determine the transition order of Markov regression model. In particular, we use the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) of possible models in order to determine the transition order of the Markov binary regression model if the transition order is known; however, we compute and compare their posterior probabilities if unknown. In order to overcome the complicated Bayesian computation, our proposed model is reconstructed by the ideas of Albert and Chib (1993), Kuo and Mallick (1998), and Erkanli *et al.* (2001). Our proposed method is applied to the simulated data and real data examined by Sommer *et al.* (1984). Markov chain Monte Carlo methods to determine the optimal model are used assuming that the transition order of the Markov regression model are known or unknown. Gelman and Rubin's method (1992) is also employed to check the convergence of the Metropolis Hastings algorithm.

Keywords: deviance information criterion (DIC), Markov chain Monte Carlo method, Markov binary regression with t-link, Gelman and Rubin diagnostic

1. 서론

일반적으로 선형회귀분석은 관측치들이 독립이고, 오차항이 정규분포를 따른다는 가정하에 분석이 수행된다. 종단 자료(longitudinal data)는 시간이 지남에 따라 동일한 개인 혹은 집단들을 반복적으로 측정한 자료로, 때로는 이러한 데이터를 패널 데이터라고도 한다. 이러한 반복적으로 측정된 종단자료에 일반적인 선형회귀분석의 가정들을 적용하기에는 어려움이 따른다. 따라서 종단 자료는 상이한 시점에서 수집된 관측치가 상호 연관되는 경향이 있으므로, 좀 더 유연한 오차항을 가진 마코프 회귀모형을 고려하는 것이 합리적이다.

Cox (1970)은 가우시안 이진 시계열에 대한 로지스틱 회귀의 확장인 Markov 체인을 제안했으며 Korn과 Whittemore (1979)는 이 모델을 패널 데이터에 적용했다. Kalbfleisch와 Lawless (1985)는

[1]Corresponding author: Department of Statistics, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Korea. Email: yschung@pusan.ac.kr

범주형 패널 데이터에 대한 Markov 회귀 모델에 대해 논의했다. Lee 등 (1968, 1970)는 동질적 체인의 전이 확률을 추정하기 위해 공액(conjugate) 베이지안 방법론을 제시했고, Meshkani (1978)는 동질적 및 비동질적 Markov 체인에 대한 전이 확률의 경험적 Bayes 추정값(empirical Bayes estimates)을 제안했다. Erkanli 등 (2001)는 Markov 로지스틱 회귀 설정을 사용하여 이진 Markov 체인을 고려하고 모델의 베이지안 분석을 제시했다. 그 외 Markov 모델의 분석은 Azzalini (1982), Bartholomew (1983), Cox (1981), Singer와 Spilerman (1976a, 1976b), Wasserman (1980), Zeger와 Qaqish (1988) 등에 의해 논의되었다.

또한 작은 표본 문제에 관심이 있었던 Gosset (1908)은 'Student'라는 가명으로 t 분포를 소개했고, Fisher (1925)는 이 분포를 'Student 분포'라 명명하였다. 이후 Gosset의 필명에 따라 Student t 분포라고 알려지게 되었다. 이 분포는 종 모양이며 대칭이고 자유도가 증가함에 따라 정규 분포로 근사하는 특징을 가진다. 따라서 본 논문에서는 t 분포의 자유도가 증가함에 따라 정규 분포로 근사하는 특징을 이용해 오차항이 정규분포를 따른다는 가정의 유연한 대안으로 오차항이 Student t 분포를 따른다고 설정한다. Albert와 Chib (1993)는 이진회귀 모형에서 t-링크 함수를 이용한 베이지안 모델을 제시했다. 그러므로 우리는 종단자료에 대해 t 오차항을 가진 마코프 이항 회귀 모형을 제안하고, 이 때 마코프 회귀 모형의 전이 시차(transition order)에 대한 베이지안 분석을 수행하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 Erkanli 등 (2001)과 Albert와 Chib (1993)의 접근 방식을 사용하여 t-링크 함수를 갖는 Markov 이항 회귀 모형을 제안한다. 3장에서는 2장에서 제안한 t-링크 함수를 갖는 Markov 이항 회귀 모형의 전이 시차가 알고 있는 경우와 그렇지 않은 경우에 따른 베이지안 분석에 대해 다룬다. 4.1절에서 시뮬레이션 데이터를 이용해 제안된 모형의 성능을 확인하고, 4.2절에서는 제안된 모형을 Somer 등 (1984)에 의해 연구된 코호트 데이터에 적용한다. 마지막으로, 5장에서 본 연구의 결론을 제시하고 향후 과제를 논의한다.

2. t-링크를 갖는 마코프 이항 회귀 모형

i번째 대상이 서로 다른 상황 T에서 관측되었다고 가정할 때, $y_i = (y_{i1}, \ldots, y_{iT})$, 이벤트가 발생하면 $y_{it} = 1$, 그렇지 않으면 $y_{it} = 0$이라 한다면, y_{it}는 확률이 $p_{it} = P(y_{it} = 1|h_{it})$인 베르누이 분포(Bernoulli distribution)를 따른다고 할 수 있다.

$$f(y_{it}|h_{it}) = p_{it}^{y_{it}}(1 - p_{it})^{1-y_{it}},$$

여기서 $h_{it} = \{y_{i1}, \ldots, y_{i,t-1}\}$인 히스토리 벡터이다.

Erkanli 등 (2001)에 따르면, p_{it}를 히스토리 벡터 h_{it}와 공변량 벡터 x_{it}의 가법함수(additive function)의 형태로 나타낼 수 있다. 예를 들어, 프로핏 모델에서 p_{it}는 다음과 같이 표현된다.

$$p_{it} = \Phi\left(\mu + x_{it}'\beta + g(h_{it})\right),$$

여기서 $x_{it} = (x_{i1t}, \ldots, x_{ikt})'$, $\beta = (\beta_1, \ldots, \beta_k)'$ 그리고 $\Phi(\,)$는 표준 가우시안 누적 분포 함수(cumulative distribution function; cdf)이다.

또한, Erkanli 등 (2001)에 따르면, 함수 $g(\,)$에 따라 관측치의 전이 행동의 믿음의 정도를 나타낸다. 특별히 우리는 $g(\,)$ 함수를 다음과 같이 정의한다.

$$g(h_{it}) = \sum_{k=1}^{s} \alpha_k y_{i,t-k}, \quad \text{for } s \in \{1, \ldots, s_{\max}\}.$$

이는 최대 순서(maximal order) $1 \le s_{\max} \le T-1$인 s번째 순서를 가진 마코프 회귀모형을 나타낸다. George와 McCulloch (1993, 1997)와 Kuo와 Mallick (1998)의 아이디어에 따라, $g(h_{it})$를 다음과 같은 가법형태(additive structure)로 확장시킬 수 있다.

$$g(h_{it}) = \sum_{s=1}^{s_{\max}} \alpha_s \delta_s y_{i,t-s}.$$

$y_{i,t-s}$가 모델에 포함되면 $\delta_s = 1$이고, 그렇지 않으면 $\delta_s = 0$이다. 따라서 $\delta_1 = 1$ 그리고 $s > 1$에 대해 $\delta_s = 0$은 $g(h_{it}) = \alpha_1 y_{i,t-1}$인 1차 모델을 의미한다. 유사하게 $\delta_1 = \delta_2 = 1$, $s > 2$일 때, $\delta_s = 0$은 $g(h_{it}) = \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2}$인 2차 모델을 의미한다. 즉, $\delta = (\delta_1, \ldots, \delta_{s_{\max}})$, $\delta_i = 0, 1$, $i = 1, \ldots, s_{\max}$를 이용하여 전이 시차를 결정할 수 있다. 일반성을 잃지 않고 우리는 $s_{\max} = T-1$를 설정한다. 이 때, 시점에 따른 영향을 받지 않는 경우와 $s_{\max} = T-1$인 경우를 고려하여 모든 Markov 모델의 공간 M은 아래와 같이 $(T-1)$-튜플로 이루어진 T 차원이 된다.

$$M = \{(0, \ldots, 0), (1, 0, \ldots, 0), (1, 1, 0, \ldots, 0), \ldots, (1, \ldots, 1)\}. \tag{2.1}$$

이때, 첫 번째 벡터 $(0, \ldots, 0)$은 시점에 따른 영향을 받지 않는 제로순서 모델 M_0이다. 두 번째 벡터 $(1, 0, \ldots, 0)$은 첫 번째 순서 모델 M_1이다.

Albert와 Chib (1993)에서 이진 회귀 모델을 $p_{it} = C(x_{it}^T \beta)$, $i = 1, \ldots, N$로 정의 할 수 있다. 여기서 $C(\,)$는 확률 p_{it}을 선형 구조 $x_{it}^T\beta$와 연결하는 알려진 cdf로, 본 연구에서는 자유도가 ν인 t 분포의 cdf, $T_\nu(\,)$를 사용하고자 한다. 이 때, 확률 p_{it}는 아래와 같다.

$$\begin{aligned} p_{it} &= T_\nu(x_{it}'\beta + \alpha_1\delta_1 y_{i,t-1} + \alpha_2\delta_2 y_{i,t-2} + \cdots + \alpha_{t-1}\delta_{t-1} y_{i1}) \\ &= \int_{-\infty}^{x_{it}'\beta + \sum_{s=1}^{t-1} \alpha_s\delta_s y_{i,t-s}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt. \end{aligned}$$

이 때, t 분포가 정규분포와 감마분포의 혼합 형태로 표현되어지므로,

$$Z_{it}|\lambda_{it} \sim N\left(x_{it}'\beta + \sum_{s=1}^{t-1} \alpha_s\delta_s y_{i,t-s}, \lambda_{it}^{-1}\right) \text{ 과 } \lambda_{it} \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

를 이용하여 확률 p_{it}를 아래와 같이 나타낸다.

$$\begin{aligned} p_{it} &= \int_0^\infty \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(z - x_{it}'\beta - \sum_{s=1}^{t-1} \alpha_s\delta_s y_{i,t-s})^2}{\nu}\right)^{-\frac{\nu+1}{2}} dz \\ &= \int_0^\infty \int_0^\infty \sqrt{\frac{\lambda_{it}}{2\pi}} \exp\left(-\frac{\lambda_{it}}{2}\left(Z_{it} - x_{it}'\beta - \sum_{s=1}^{t-1} \alpha_s\delta_s y_{i,t-s}\right)^2\right) \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_{it}^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_{it}\right) d\lambda_{it} dZ_{it}. \end{aligned}$$

Tanner과 Wong (1987)이 제안한 데이터 확장 방법(data augmentation)을 이용해, $Z_{it} > 0$이면 $Y_{it} = 1$, $Z_{it} < 0$이면 $Y_{it} = 0$인 방정식을 알 수 있다. 즉, 우도함수(likelihood function)는

$$\begin{aligned} L(y_{11}, \ldots, y_{N,T}) \propto \prod_{i=1}^{N}\prod_{t=1}^{T} &\{I(Z_{it} > 0)I(y_{it} = 1) + I(Z_{it} < 0)I(y_{it} = 0)\} \\ &\times \sqrt{\frac{\lambda_{it}}{2\pi}} \exp\left(-\frac{\lambda_{it}}{2}\left(Z_{it} - x_{it}'\beta - \sum_{s=1}^{T-1} \alpha_s\delta_s y_{i,t-s}\right)^2\right) \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_{it}^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_{it}\right). \end{aligned}$$

와 같이 나타낼 수 있고, $x'_{it}\beta + \sum_{s=1}^{T-1} \alpha_s \delta_s y_{i,t-s}$가 선형 조합 구조(linear combination structure)이므로, ${X^*}'_{it}\gamma$로 재구성한다.

$$X^* = \left({X^*}'_{11}, {X^*}'_{12}, \ldots, {X^*}'_{N,T}\right)^T = \begin{bmatrix} \underline{x}'_{11} & 0 & \cdots & \cdots & 0 \\ \underline{x}'_{12} & y_{11} & 0 & \cdots & 0 \\ \underline{x}'_{13} & y_{12} & y_{11} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \end{bmatrix} \text{과 } \gamma = \begin{pmatrix} \beta \\ \alpha \end{pmatrix},$$

여기서 $\underline{x}'_{it}$는 i번째 대상의 k차원 공변량 벡터, $\beta = (\beta_1, \ldots, \beta_k)^T$, $\alpha = (\alpha_1\delta_1, \ldots, \alpha_{T-1}\delta_{T-1})^T$이다. 즉, 우도함수는 아래와 같다.

$$\begin{aligned} L\left(y_{11}, \ldots, y_{N,T}\right) \propto & \prod_{i=1}^{N}\prod_{t=1}^{T} \{I(Z_{it} > 0)I(y_{it} = 1) + I(Z_{it} < 0)I(y_{it} = 0)\} \\ & \times \sqrt{\frac{\lambda_{it}}{2\pi}} \exp\left(-\frac{\lambda_{it}}{2}\left(Z_{it} - {X^*}'_{it}\gamma\right)^2\right) \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_{it}^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_{it}\right). \end{aligned} \tag{2.2}$$

3. Bayesian approach

2장에서 다룬 t-링크를 갖는 마코프 이항 회귀 모형의 전이 시차가 알려진 경우와 그렇지 않은 경우일 때, 각각에서 최적의 전이 시차 모델을 알아보고자 베이지안 분석을 실시한다. 마코프 모델의 공간 M은 식 (2.1)과 같이 $(T-1)$개의 튜플로 이루어진 T차원 공간이다. M_i는 j가 i보다 작거나 같은 경우 $\delta_j = 1$, 그렇지 않으면 $\delta_j = 0$, $i = 0, \ldots, T-1$, $j = 1, \ldots, T-1$으로, 마지막 관측치가 가장 최근 i번째 관측치까지 영향을 받는 모델을 의미한다.

3.1. 전이 시차가 알려진 경우

먼저 자유도 ν는 고정하지 않고, 전이 시차에 대해서만 알고 있다고 가정한다. 이 때, γ의 사전분포(prior distribution)는 $\gamma \sim N_p(\mu, \Sigma^{-1})$, where $p = k + s$, $s = 1, \ldots, T$, ν의 사전분포는 평균과 분산이 각각 a/b, a/b^2인 Gamma(a, b)로 가정한다. 따라서 γ, ν의 결합사전분포 $\pi(\gamma, \nu)$는 다음과 같다.

$$\begin{aligned} \pi(\gamma, \nu) &= \pi(\gamma)\pi(\nu) \\ &\propto \frac{1}{\sqrt{(2\pi)^p}} |\Sigma^{-1}|^{-\frac{1}{2}} \frac{b^a}{\Gamma(a)} \nu^{a-1} \exp\left(-\frac{1}{2}\left(\gamma - \mu\right)' \Sigma \left(\gamma - \mu\right) - b\nu\right). \end{aligned}$$

그러므로 Z, γ, λ와 ν의 결합사후밀도함수(joint posterior density function)는

$$\begin{aligned} \pi(Z, \gamma, \lambda, \nu | y, X^*) \propto & \prod_{i=1}^{N}\prod_{t=1}^{T} \{I(Z_{it} > 0)I(Y_{it} = 1) + I(Z_{it} < 0)I(Y_{it} = 0)\} \\ & \times \sqrt{\frac{\lambda_{it}}{2\pi}} \exp\left(-\frac{\lambda_{it}}{2}\left(Z_{it} - {X^*}'_{it}\gamma\right)^2\right) \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_{it}^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_{it}\right) \\ & \times \frac{1}{\sqrt{(2\pi)^p}} |\Sigma^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\gamma - \mu\right)' \Sigma \left(\gamma - \mu\right)\right) \\ & \times \frac{b^a}{\Gamma(a)} \nu^{a-1} \exp(-b\nu) \end{aligned} \tag{3.1}$$

로 나타낼 수 있고, 사후분포(posterior distribution) $\pi(Z,\gamma,\lambda,\nu|y,X^*)$에 대한 추론을 위해 깁스샘플링 방법을 사용한다. γ의 완전 조건부 밀도(full conditional density; FCD)는

$$\gamma|Z,\lambda,\nu,y,X^* \sim N_p\left(\left(X^{*\prime}WX^*+\Sigma\right)^{-1}\left(X^*WZ+\Sigma\mu\right),\left(X^{*\prime}WX^*+\Sigma\right)^{-1}\right) \tag{3.2}$$

이며, $W=\mathrm{diag}(\lambda_1,\ldots,\lambda_{1200})$이다. Z_{it}의 완전 조건부 밀도는

$$Z_{it}|\gamma,\lambda,\nu,y,X^* \sim N\left({X^*}'_{it}\gamma,\lambda_{it}^{-1}\right). \tag{3.3}$$

이 때, $y_{it}=1$이면 0의 왼쪽이 절단된 분포이고, $y_{it}=0$이면 그 반대이다. λ_{it}의 완전 조건부 밀도는 다음과 같다.

$$\lambda_{it}|Z,\gamma,\nu,y,X^* \sim \Gamma\left(\frac{\nu+1}{2},\frac{(Z_{it}-{X^*}'_{it}\gamma)^2+\nu}{2}\right). \tag{3.4}$$

ν의 완전 조건부 밀도는 다음과 같다.

$$\begin{aligned}\pi(\nu|\gamma,Z,\lambda,y,X^*) &\propto \prod_{i=1}^{N}\prod_{t=1}^{T}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\lambda_{it}^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_{it}\right)\nu^{a-1}\exp(-b\nu)\\ &\propto \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}\times NT}}{\Gamma\left(\frac{\nu}{2}\right)^{NT}}\nu^{a-1}\exp\left(-\left\{b+\frac{1}{2}\sum_{i,t}\left(\lambda_{it}-\log\lambda_{it}\right)\right\}\nu-\sum_{i,t}\log\lambda_{it}\right).\end{aligned} \tag{3.5}$$

3.2. 전이 시차가 알려지지 않은 경우

3.2절에서는 마코프 모델의 전이 시차와 자유도 ν를 모른다고 가정한다. 이 때, γ의 사전분포는 $\gamma\sim N_p(\mu,\Sigma^{-1})$, $p=k+T$, ν의 사전분포는 평균과 분산이 각각 a/b , a/b^2인 Gamma(a,b)로 가정한다. 그리고, $\delta\in M$의 사전분포는 각 모델에 $p_k=1/T$, $k=0,1,\ldots,T-1$, $\sum_{k=0}^{T-1}p_k=1$인 균일분포를 가정한다. 따라서 γ,δ,ν의 결합사전분포 $\pi(\gamma,\delta,\nu)$는 다음과 같이 구해진다.

$$\begin{aligned}\pi(\gamma,\delta,\nu) &= \pi(\gamma)\pi(\delta)\pi(\nu)\\ &\propto \frac{1}{\sqrt{(2\pi)^p}}|\Sigma^{-1}|^{-\frac{1}{2}}\frac{b^a}{\Gamma(a)}\nu^{a-1}\times\exp\left(-\frac{1}{2}\left(\gamma-\mu\right)'\Sigma\left(\gamma-\mu\right)-b\nu\right)\mathrm{Unif}(0,T).\end{aligned} \tag{3.6}$$

그러므로 Z, γ, λ, δ와 ν의 결합사후밀도함수는

$$\begin{aligned}\pi\left(Z,\gamma,\delta,\lambda,\nu|y,X^*\right) \propto &\prod_{i=1}^{N}\prod_{t=1}^{T}\left\{I(Z_{it}>0)I(Y_{it}=1)+I(Z_{it}<0)I(Y_{it}=0)\right\}\\ &\times\sqrt{\frac{\lambda_{it}}{2\pi}}\exp\left(-\frac{\lambda_{it}}{2}\left(Z_{it}-{X^*}'_{it}\gamma\right)^2\right)\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\lambda_{it}^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_{it}\right)\\ &\times\frac{1}{\sqrt{(2\pi)^p}}\left|\Sigma^{-1}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\gamma-\mu\right)'\Sigma\left(\gamma-\mu\right)\right)\\ &\times\frac{b^a}{\Gamma(a)}\nu^{a-1}\exp(-b\nu)\mathrm{Unif}(0,T)\end{aligned} \tag{3.7}$$

로 나타낼 수 있고, 사후분포 $\pi(Z,\gamma,\delta,\lambda,\nu|y,X^*)$에 대한 추론을 위해 깁스샘플링 을 수행하기 위한 γ의 완전조건부 밀도는 아래와 같다.

$$\gamma|Z,\lambda,\delta,\nu,y,X^* \sim N_p\left(\left(X^{*\prime}WX^*+\Sigma\right)^{-1}\left(X^*WZ+\Sigma\mu\right),\left(X^{*\prime}WX^*+\Sigma\right)^{-1}\right), \tag{3.8}$$

여기서 $W = \mathrm{diag}(\lambda_1, \ldots, \lambda_{NT})$이다. Z_{it}의 완전조건부 밀도는

$$Z_{it}|\gamma, \lambda, \delta, \nu, y, X^* \sim N\left({X^*}'_{it}\gamma, \lambda_{it}^{-1}\right) \tag{3.9}$$

이며, $y_{it} = 1$이면 0의 왼쪽이 절단된 분포이고, $y_{it} = 0$이면 그 반대이다.

λ_{it}, δ와 ν의 완전조건부 밀도는 각각 다음과 같다.

$$\lambda_{it}|Z, \gamma, \delta, \nu, y, X^* \sim \Gamma\left(\frac{\nu+1}{2}, \frac{(Z_{it} - {X^*}'_{it}\gamma)^2 + \nu}{2}\right) \tag{3.10}$$

$$\delta = \begin{cases} M_0, & \text{w.p. } p_0, \\ M_1, & \text{w.p. } p_1, \\ M_2, & \text{w.p. } p_2, \\ M_3, & \text{w.p. } p_3, \\ M_4, & \text{w.p. } p_4, \\ M_5, & \text{w.p. } p_5, \end{cases} \quad p_i = \frac{\delta_{\mathbf{i}}}{\sum_{i=0}^{5}\delta_{\mathbf{i}}}, \quad \sum_{i=0}^{5} p_i = 1, \tag{3.11}$$

여기서 $\pi(\delta_{\mathbf{k}}) \propto \prod_{i=1}^{N}\prod_{t=1}^{T}\sqrt{\lambda_{it}/2\pi}\exp(-(\lambda_{it}/2)(Z_{it} - x'_{it}\beta - \sum_{s=0}^{k}\alpha_s y_{i,t-s})^2)$, $\alpha_0 = 0$, $k = 0, \ldots, 5$.

$$\begin{aligned}\pi(\nu|\gamma, Z, \lambda, \delta, y, X^*) &\propto \prod_{i=1}^{N}\prod_{t=1}^{T}\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\lambda_{it}^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_{it}\right)\nu^{a-1}\exp(-b\nu) \\ &\propto \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}\times NT}}{\Gamma\left(\frac{\nu}{2}\right)^{NT}}\nu^{a-1}\exp\left(-\left\{b + \frac{1}{2}\sum_{i,t}(\lambda_{it} - \log\lambda_{it})\right\}\nu - \sum_{i,t}\log\lambda_{it}\right). \end{aligned} \tag{3.12}$$

4. 시뮬레이션과 실제 데이터

4.1. 시뮬레이션 데이터

제안한 모형을 실제 데이터에 적용하기 전에 성능을 확인하기 위해 시뮬레이션 데이터를 생성한다. 이 때, t 분포의 자유도 $\nu = 1, 3, 5, 7, 10$으로 고정하고, 전이 시차를 알고 있다고 가정한다. 반복 측정된 시뮬레이션 데이터를 생성하기 위해 공변량 X는 4.2절의 인도네시아 어린이 종단 자료 중 6번 모두 성실히 응답한 122명을 대상으로 30번 복원 추출하였으며, 4.2절의 인도네시아 어린이 종단 데이터의 프로빗 모형에서의 최소제곱추정치(least square estimator; LSE)를 이용해 y를 새로 추정했다. 따라서 $n = 30$, $T = 6$, $s_{\max} = 5$, 생성된 데이터를 이용해 모델 M_i, $i = 0, \ldots, 5$에 대한 데이터를 재구성하여 각각 깁스샘플링을 10,000번 수행한다. 모델 비교를 위해 Spiegelhalter 등 (2002)이 제안한 편차 정보기준(deviance information criterion; DIC)을 사용한다.

$$\mathrm{DIC} = D\left(\bar{\theta}\right) + 2p_D, \tag{4.1}$$

여기서 $D(\theta) = -2\log L(y, X^*|\theta)$, $\bar{\theta}$는 사후 평균(posterior mean), $p_D = E(D(\theta)|y, X^*) - D(\bar{\theta})$이다.

t 분포의 자유도 $\nu = 1, 3, 5, 7, 10$일 때, 모델 M_i, $i = 0, \ldots, 5$에 대한 $\bar{D}(\theta)$, p_D, DIC 값은 Table 4.1에 나타나있다. $\nu = 1$인 경우에는 $M_0 = 1{,}501.25e + 260$, $M_1 = 1{,}505.39e + 260$, $M_2 = 1{,}501.55e + 260$, $M_3 = 1{,}502.12e + 260$, $M_4 = 1{,}505.63e + 260$, $M_5 = 1{,}496.90e + 260$으로, M_5 모델의 DIC값이 $1{,}496.90e + 260$으로 가장 낮게 나타났다. $\nu = 3$인 경우에는 $M_0 = 1{,}844.44e{+}260$, $M_1 = 1{,}842.21e + 260$, $M_2 = 1{,}839.35e{+}260$, $M_3 = 1{,}848.26e{+}260$, $M_4 = 1{,}851.34e{+}260$, $M_5 = 1{,}829.71e{+}260$으로,

Table 4.1. Simulation result using the DIC (Unit: $1e+260$)

ν		$\bar{D}(\theta)$	p_D	DIC
1	M_0	1,030.59	470.66	1,501.25
	M_1	1,032.83	472.56	1,505.39
	M_2	1,030.63	470.92	1,501.55
	M_3	1,030.61	471.51	1,502.12
	M_4	1,032.91	472.72	1,505.63
	M_5	1,027.52	469.38	1,496.90
3	M_0	1,158.26	686.18	1,844.44
	M_1	1,156.92	685.29	1,842.21
	M_2	1,155.73	683.63	1,839.35
	M_3	1,160.35	687.91	1,848.26
	M_4	1,162.26	689.08	1,851.34
	M_5	1,150.82	678.89	1,829.71
5	M_0	766.82	520.83	1,287.65
	M_1	767.74	521.94	1,289.68
	M_2	770.47	523.99	1,294.47
	M_3	768.00	521.83	1,289.83
	M_4	766.90	520.79	1,287.69
	M_5	763.76	517.90	1,281.66
7	M_0	403.73	279.48	683.21
	M_1	407.99	283.23	691.22
	M_2	410.09	284.66	694.76
	M_3	406.93	282.64	689.57
	M_4	404.19	279.96	684.15
	M_5	402.38	277.72	680.10
10	M_0	153.26	91.40	244.66
	M_1	152.13	90.46	242.59
	M_2	153.11	90.76	243.87
	M_3	152.44	90.51	242.95
	M_4	152.89	90.96	243.85
	M_5	151.95	90.03	241.98

DIC = deviance information criterion.

M_5 모델의 DIC값이 $1{,}829.71e+260$으로 가장 낮게 나타났다. $\nu=5$인 경우에는 $M_0=1{,}287.65e+260$, $M_1=1{,}289.68e+260$, $M_2=1{,}294.47e+260$, $M_3=1{,}289.83e+260$, $M_4=1{,}287.69e+260$, $M_5=1{,}281.66e+260$으로, M_5 모델의 DIC값이 가장 낮게 나타났다. 마찬가지로, $\nu=7$과 10인 경우에도 M_5의 DIC 값이 다른 모형에 비해 가장 낮은 것으로 나타났다.

4.2. 실제 데이터

Sommer 등 (1984)에 의해 보고된 인도네시아 아동의 호흡기질환이 있는 이진 종단 연구는 6개의 Javanese 마을의 미취학 아동 4,600명이 18개월 동안 지속된 종단 연구이다. 안과 의사, 소아과 의사, 영양사, 간호사 및 현장 근로자로 구성된 팀이 3개월마다 각 어린이를 재검사하려고 시도했으며, 총 6번의 간격으로 총 7번의 시험이 이루어졌다. 그 지역에서 사망하거나 영구적으로 떠난 아이들은 연구에서 제외되었고, 최소 2개월 동안 떠난 아이들 역시 연구에서 제외되었다. 모든 어린이가 항상 참석한 것은 아니며, 지역 금기로 일반적으로 3개월 미만의 아동에 대한 검사는 배제했다.

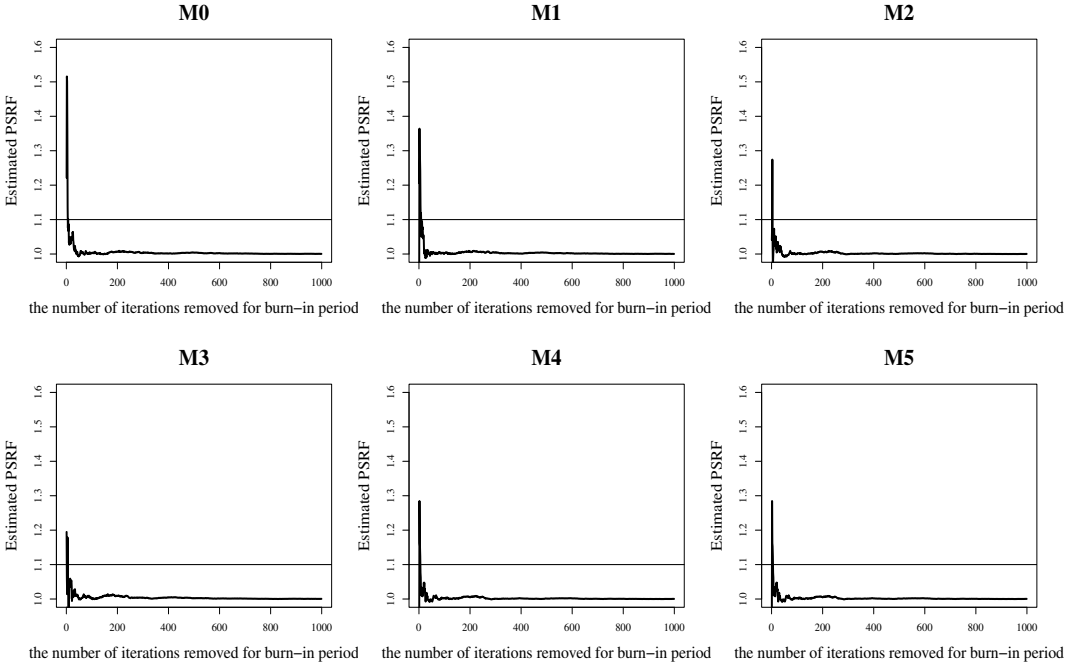

Figure 4.1. Plots of the estimated PSRF for models. PSRF = positive scale reduction factor.

우리는 R의 `DPpackage`에 내장된 `indon` 데이터를 사용했다. 이는 Sommer 등 (1984)의 코호트 자료의 일부로, 275명의 인도네시아 미취학 아동들의 호흡기 감염 여부에 대해 최대 연속 6분기 동안 조사가 이루어졌다. 따라서 마코프 모델의 공간 M은 5개의 튜플로 이루어진 6차원 공간이다. 이 자료는 1,200개의 행(row)으로 이루어져 있으며, 공변량은 다음과 같다: 성별; 90 %를 중심으로 하는 미국 국가 건강 통계 센터(National Center for Health Statistics; NCHS) 표준의 백분율로서 나이의 신장; 연간주기의 계절 코사인 및 사인; 야맹증 또는 Bitot's spot과 같은 안과 질환 유무; 등록 당시 나이와 아동의 나이는 36개월에서 중심화되었다.

4.2.1. 전이 시차가 알려진 경우 전이 시차를 알고 있다고 가정하면, 3.1절의 계산식을 따른다. 이 때, ν의 완전 조건부 밀도는 해석적으로 잘 알려진 분포 형태가 아니므로 Metropolis Hastings 알고리즘을 사용한다. 식 (3.5)에서 샘플링을 위해 제안분포(proposal distribution)는 $(1, 10)$에서 절단된 감마분포를 이용한다. Metropolis Hastings 알고리즘의 수렴성을 확인하기 위해, Gelman과 Rubin (1992)이 제안하고 Brooks와 Gelman (1997)이 개선한 수렴 진단을 사용했다. 모두 동일한 목표 분포에 수렴되는지를 진단하기 위해 다양한 초기 값을 가진 병렬 체인을 사용하고 체인 내 분산과 체인 간 분산을 계산한다. 잠재적 규모 축소 계수(positive scale reduction factor; PSRF)와 multivariate PSRF (MPSRF)를 각각 추정하기 위해 2,000회 반복으로 5개의 체인을 실행했다. Gelman과 Rubin (1992)은 PSRF가 1에 가까우면 후자의 절반 시퀀스가 수렴되었다고 판단하고 PSRF가 1.1보다 작은 경우에는 표본 변동성을 무시할 수 있도록 충분한 관측치가 있다고 했다.

Figure 4.1은 6개의 모델 $M_0, \ldots, M_5$에 대한 추정 PSRF로, 번인(burn-in) 기간 이후 약 500회 반복 후 1에 도달하는 것을 알 수 있다. 따라서 Metropolis Hastings 알고리즘을 500회 반복 한 후 수렴에 도달했다는 결정을 내리고 총 반복 횟수를 1,000으로 설정했다.

Table 4.2. The DIC for the six models (Unit: 1e + 260)

M_0	M_1	M_2	M_3	M_4	M_5
11,384.28	11,380.91	11,440.89	11,410.53	11,505.76	11,380.78

Gibbs 샘플러 알고리즘은 다음과 같다.

Step 1. 초기값 설정 $\gamma^{(0)}, \lambda^{(0)}, W^{(0)}$, and $\nu^{(0)}$.

Step 2. $j = 1$.

Step 3. 식 (3.3)의 $\pi(Z_{it}|\gamma^{(j-1)}, \lambda^{(j-1)}, \nu^{(j-1)}, y, X^*)$으로부터 $Z^{(j)}$ 추출.

Step 4. 식 (3.4)의 $\pi(\lambda_{it}|Z^{(j)}, \gamma^{(j-1)}, \nu^{(j-1)}, y, X^*)$으로부터 $\lambda^{(j)}$.

Step 5. $(1, 10)$에서 절단된 감마분포를 제안분포로 한 Metropolis Hastings 알고리즘을 이용해, 식 (3.5)의 $\pi(\nu|Z^{(j)}, \lambda^{(j)}, \gamma^{(j-1)}, y, X^*)$으로부터 $\nu^{(i)}$ 추출.

Step 6. 식 (3.2)의 $\pi(\gamma|Z^{(j)}, \lambda^{(j)}, \nu^{(j)}, y, X^*)$으로부터 $\gamma^{(j)}$ 추출.

Step 7. $j = j + 1$.

Step 8. Step 3부터 Step 7의 과정을 각 모델 M_i, $i = 0, \ldots, 5$에 대해 10,000번 반복실행.

깁스샘플링을 수행하기 위해 초기 값은 $\lambda_i = 1$, $i = 1, \ldots, 1200$, β는 프로빗 회귀 모델에서 LSE와 같고, 각 모델에 따라 $\alpha = 1$ 벡터로 설정한다. 모델의 전이 시차를 알고 있으므로 식 (4.1)에서 정의된 DIC를 계산해 최적의 모델을 선택하며, 그 결과는 Table 4.2에 나타나있다. M_0 = 11,384.28e + 260, M_1 = 11,380.91e+260, M_2 = 11,440.89e+260, M_3 = 11,410.53e+260, M_4 = 11,505.76e+260, M_5 = 11,380.78e + 260으로, M_5 모델의 DIC값이 가장 낮은 것으로 나타났다. 이 때 M_5 모델에서의 자유도 ν의 추정값은 4.83이며, 회귀계수 γ값은 −167,498.41, −71,793.10, −162,381.07, 26,977.92, 31,883.80, −6,951.31, 357,611.13, −849,144.21, −9,478.14, −4,577.56, −3,040.41, −1,546.63, −1,526.17으로 나타났다. 즉, 바로 직전의 관측치가 주는 영향이 제일 크고, 관측 시점이 멀어질수록 영향력이 떨어지는 것을 알 수 있다.

4.2.2. 전이 시차가 알려지지 않은 경우 Metropolis Hastings 알고리즘의 수렴성을 확인하기 위해 앞서 언급한 Gelman과 Rubin (1992)이 제안하고 Brooks와 Gelman (1997)이 수정한 수렴 진단을 사용했다. 또한 PSRF를 추정하기 위해 2,000회 반복으로 5개의 체인을 실행한다. Figure 4.2는 번인 기간 이후 약 500회 반복 후 1에 도달하는 것을 나타낸다. 따라서 Metropolis Hastings 알고리즘을 500회 반복 한 후 수렴에 도달했다는 결정을 내리고 총 반복 횟수를 1,000으로 설정했다.

Gibbs 샘플러 알고리즘은 다음과 같다.

Step 1. 초기값 설정 $\gamma^{(0)}, \lambda^{(0)}, W^{(0)}$, $\delta^{(0)}$, 그리고 $\nu^{(0)}$.

Step 2. $j = 1$.

Step 3. 식 (3.9)의 $\pi(Z_{it}|\gamma^{(j-1)}, \lambda^{(j-1)}, \delta^{(j-1)}, \nu^{(j-1)}, y, X^*)$으로부터 $Z^{(j)}$ 추출.

Step 4. 식 (3.10)의 $\pi(\lambda_{it}|Z^{(j)}, \gamma^{(j-1)}, \delta^{(j-1)}, \nu^{(j-1)}, y, X^*)$으로부터 $\lambda^{(j)}$ 추출.

Step 5. $(1, 10)$에서 절단된 감마분포를 제안분포로 한 Metropolis Hastings 알고리즘을 이용해, 식 (3.12)의 $\pi(\nu|Z^{(j)}, \lambda^{(j)}, \gamma^{(j-1)}, \delta^{(j-1)}, y, X^*)$으로부터 $\nu^{(i)}$ 추출.

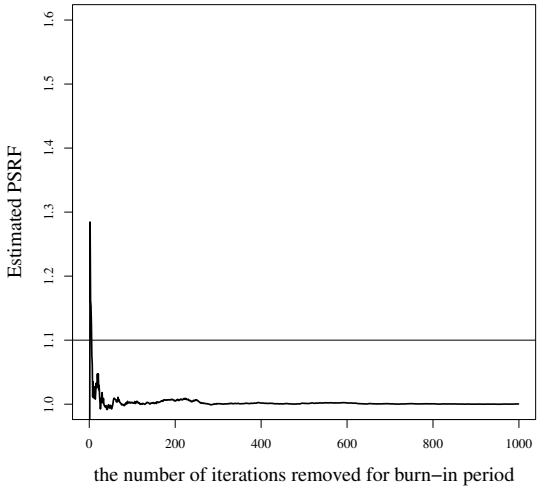

Figure 4.2. Plot of the estimated PSRF of ν.

Table 4.3. The frequency of δ for the six models

M_0	M_1	M_2	M_3	M_4	M_5
0	33	265	616	916	3,170

Step 6. 식 (3.11)의 $\pi(\delta|Z^{(j)}, \lambda^{(j)}, \nu^{(j)}, \gamma^{(j-1)}, y, X^*)$으로부터 $\delta^{(j)}$ 추출.

Step 7. 식 (3.8)의 $\pi(\gamma|Z^{(j)}, \lambda^{(j)}, \nu^{(j)}, \delta^{(j)}, y, X^*)$으로부터 $\gamma^{(j)}$ 추출.

Step 8. $j = j + 1$.

Step 9. Step 3 부터 Step 7의 과정을 10,000번 반복실행

깁스샘플링을 수행하기 위해 초기 값은 $\lambda_i = 1$, $i = 1, \ldots, 1200$, β는 프로 빗 회귀 모델에서 LSE와 같고, $\alpha = 1_5$ 벡터, $\delta = \delta_5$ 그리고 $\nu = 5$로 설정한다. 모형의 사후 확률을 이용하여 5,000번 번인 기간 후의 δ의 빈도값을 구한 결과는 Table 4.3에 나타나있다. $M_0 = 0$, $M_1 = 33$, $M_2 = 265$, $M_3 = 616$, $M_4 = 916$, $M_5 = 3{,}170$으로, M_5 모델의 빈도가 가장 많은 것을 알 수 있다. 이 때 M_5 모델에서의 자유도 ν의 추정값은 4.78이며, 회귀계수 γ값은 $-167{,}937.18$, $-71{,}933.57$, $-163{,}816.44$, 26,832.56, 32,061.20, $-6{,}964.18$, 363,215.58, $-846{,}709.99$, $-9{,}505.87$, $-4{,}598.17$, $-2{,}876.81$, $-1{,}270.06$, -962.15로 나타났다. 이를 통해 인도네시아 어린이 종단 자료의 경우, 이전 관측치를 모두 고려한 t-링크 함수를 갖는 마코프 이항 회귀 모형이 가장 최적의 모형이며, 바로 직전의 관측치가 주는 영향이 제일 크고, 관측 시점이 멀어질수록 영향력이 떨어지는 것을 알 수 있다.

5. 결론

본 논문에서는 2절에 표현 된 바와 같이 Albert와 Chib (1993)의 접근법을 사용하여 Erkanli가 제안한 마코프 이항 회귀 모형을 재구성하였다. t-링크를 이용해 데이터 확장 방법을 사용하고 매개 변수의 FCD를 계산한다. FCD가 분석적으로 잘 알려진 분포로 나타낼 수 없는 경우 Metropolis Hastings 알고리즘을 사용하고 PSRF를 사용하여 Metropolis Hastings 알고리즘의 수렴성을 확인한다. 시뮬레이션 데이터를 통해 모형의 성능을 확인하고, 제안 된 모형을 인도네시아 어린이 종단 데이터에 적용하였

다. 최적의 모형을 찾기 위해 DIC와 모형 $P(M_i|y, X^*)$의 사후 확률을 사용하여 각 모형을 비교하였다. 최적의 모델은 $T-1$ 시점까지 모든 시점을 고려한 마지막 모델로 나타났다.

본 논문에서는 전이 시차가 알려지지 않은 경우에는 모형의 차원을 $(k+T-1)$로 고정하고, 전이 시차를 알고 있다고 가정한 경우에는 각 모형의 차원을 각각 $k, k+1, \ldots, k+T-1$ 로 고정한 상태에서 마코프체인 몬테카를(Markov chain Monte Carlo; MCMC)를 수행했는데, 모형의 차원을 고정하지 않고 튜플 δ에 따라 회귀계수 의 차원을 다르게 하는 방법도 고려해볼 수 있을 것이다. 이 경우 차원이 다른 공간을 탐색하기 위해서는 역점프 MCMC(reversible jump MCMC; RJMCMC)와 같은 기법이 고려되어야 한다. Green (1995)의 RJMCMC 방법은 모형의 선택 및 모수 추정의 문제를 동시에 해결해 주는 장점을 갖고 있어서 회귀분석에서의 변수 선택, 모수의 수가 서로 다른 모형들 간의 베이지안 모형 선택 문제, 다중 변화점 문제 등 많은 분야에서 응용되어지고 있다.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and Polychotomous response data, *Journal of the American Statistical Association*, **88**, 669–679.

Azzalini, A. (1982). Approximate filtering of parameter driven processes, *Journal of Time Series Analysis*, **3**, 219–223.

Bartholomew, D. J. (1983). Some recent developments in social statistics, *International Statistical Review*, **51**, 1–9.

Brooks, S. P. and Gelman, A. (1997). General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, **7**, 434–455.

Cox, D. R. (1970). *The Analysis of Binary Data*, Methuen, London.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments, *Scandinavian Journal of Statistics*, **8**, 93–115.

Erkanli, A., Soyer R., and Angold A. (2001). Bayesian analyses of longitudinal binary data using Markov regression models of unknown order, *Statistics in Medicine*, **20**, 755–770.

Fisher, R. A. (1925). Applications of " Student's" distribution, *Metron*, **5**, 90–104.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–511.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling, *Journal of the American Statistical Association*, **88**, 881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection, *Statistica Sinica*, **7**, 339–373.

Gosset, W. S. (1908). The probable error of a mean, *Biometrika*, **6**, 1–25.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711–732.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption, *Journal of the American Statistical Association*, **80**, 863–871.

Korn, E. L. and Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution, *Biometrics*, **35**, 795–802.

Kuo, L. and Mallick, B. (1998). Variable selection for regression models, *Sankhyā: The Indian Journal of Statistics*, **B 60**, 65–81.

Lee, T. C., Judge, G. G., and Zellner, A. (1968). Maximum likelihood and Bayesian estimation of transition probabilities, *Journal of the American Statistical Association*, **63**, 1162–1179.

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 and Pub. Co., Amsterdam.

Meshkani, M. (1978). *Empirical Bayes estimation of transition probabilities for Markov chains* (Ph.D. Dissertation), Florida State University.

Singer, B. and Spilerman, S. (1976a). The Representation of Social Processes by Markov Models, *American*

Journal of Sociology, **82**, 1–54.

Singer, B. and Spilerman, S. (1976b). Some Methodological Issues in the Analysis of Longitudinal Surveys, *Annals of Economic and Sociological Measurement*, **5**, 447–474.

Sommer, A., Katz, J. and Tarwotjo, I. (1984). Increased risk of respiratory infection and diarrhea in children with pre-existing mild vitamin A deficiency, *American Journal of Clinical Nutrition*, **40**, 1090–1095.

Spiegelhalter, D. A., Best, N. G., Carlin, B. P., and Linde, A. V. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B*, **64**, 583–639.

Tanner, T. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–549.

Wasserman, S. (1980). Analyzing social networks as stochastic processes, *Journal of the American Statistical Association*, **75**, 280–294.

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach, *Biometrics*, **44**, 1019–1031.

t-링크를 갖는 마코프 이항 회귀 모형을 이용한 인도네시아 어린이 종단 자료에 대한 베이지안 분석

심보현[a] · 정윤식[a,1]

[a]부산대학교 통계학과

(2019년 11월 21일 접수, 2019년 12월 30일 수정, 2020년 1월 8일 채택)

요 약

본 논문에서는 마코프 이항 회귀 모형의 시차가 알려져 있거나 그렇지 않은 경우일 때, t-링크 함수를 갖는 종단적 마코프 이항 회귀 모형을 제시한다. 일반적으로, 이항 회귀 모형에서는 로직 모형이나 프로빗 모형이 주로 사용된다. t-링크 함수는 t 분포가 자유도가 커질수록 정규분포로 근사하기 때문에 프로빗 모형을 대신 더 많은 유연성을 위해 사용될 수 있다. 게다가 마코프 회귀모형은 종단 자료에 대해 사용될 수 있다. 우리는 마코프 회귀 모형의 시차를 결정하기 위해 베이지안 방법을 제시하고자 한다. 특히, 각 모델의 차수에 대해 알고 있는 경우에는 DIC를 기준으로 모델 비교를 실시하였다. 모델의 차수에 대해 모르는 경우에는 가능한 모델들의 사후 확률을 이용하였다. 복잡한 베이지안 계산을 해결하기 위하여 Albert와 Chib (1993), Kuo와 Mallick (1998)과 Erkanli 등 (2001)의 방법을 이용하여 모델을 재설정하였다. 제안하는 방법은 시뮬레이션 데이터와 Somer 등 (1984)에 의해 조사된 인도네시아 어린이 종단 데이터에 적용했다. 마코프 이항 회귀모형의 순서에 대해서 아는 경우와 모르는 경우를 각각 가정하여 최적의 모델을 알아보기 위해 MCMC 방법을 사용하였다. 또한, 매트로폴리스 해스팅 알고리즘의 수렴성을 점검하기 위해 Gelman과 Rubin의 진단을 이용했다.

주요용어: DIC, MCMC 방법, t-링크를 갖는 마코프 이항 회귀모형, Gelman과 Rubin 진단

[1]교신저자: (46241) 부산광역시 금정구 부산대학로63번길 2 (장전동), 부산대학교 통계학과.
E-mail: yschung@pusan.ac.kr