International Journal of Management Science
Vol 12, No 2, November 2006

Bayesian Analysis of Binary Non-homogeneous Markov Chain with Two Different Time Dependent Structures

Minje Sung*
School of Business Administration, Ajou University
San5, Woncheon-Dong, Yeongtong-Gu, Suwon, 443-749, Korea

(Received Jun. 2006; Revised Sep. 2006; Accepted Sep. 2006)

ABSTRACT

We use the hierarchical Bayesian approach to describe the transition probabilities of a binary non-homogeneous Markov chain. The Markov chain is used for describing the transition behavior of emotionally disturbed children in a treatment program. The effects of covariates on transition probabilities are assessed using a logit link function. To describe the time evolution of transition probabilities, we consider two modeling strategies. The first strategy is based on the concept of exchangeability, whereas the second one is based on a first order Markov property. The deviance information criterion (DIC) measure is used to compare models with two different time dependent structures. The inferences are made using the Markov chain Monte Carlo technique. The developed methodology is applied to some real data.

Keywords: Markov Model, Hierarchical Bayesian Model, Binary Markov Chain

1. INTRODUCTION

Binary scale data can arise in many areas, including economics, engineering, social, political, and biomedical sciences. When such data are collected, over a span of years at discrete time points, a Markov chain model is used to describe the transitional behavior of the data. Most of the literature on statistical inference, concerning the transition probabilities of a Markov chain, is limited to the traditional methods, based on the likelihood ratio tests presented in Anderson and Goodman [1], Billingsley [2] and Chatfield [4]. However, the traditional approaches are based on the large sample approximations to sampling distributions.

Bayesian approaches for analyzing Markov chains were considered by Lee *et*

* Email: sungmj@ajou.ac.kr

al. [9], wherein the authors presented conjugate Bayesian methodology to estimate transition probabilities of a homogeneous chain. Meshkani [10] proposed empirical Bayes estimates of transition probabilities for both homogeneous and nonhomogeneous Markov chains. Recently, Erkanli *et al.* [6] considered a binary Markov chain using a Markov logistic regression setup. However, none of these previous approaches considered a formal treatment of nonhomogeneous Markov chains. In other words, the time non-homogeneity was described by time dependent covariates, but the model parameters were time-invariant static. In the present study, formal treatments of time non-homogeneous Markov chains are considered with time varying parameters.

Let $\{s_0, s_1, \ldots\}$ be a sequence of random variables, indexed by time, taking binary values in $\varepsilon = \{0, 1\}$. The sequence of states $\{s_0, s_1, \ldots\}$ form a binary Markov chain as

$$p(s_t = j \mid s_{t-1} = i, \ldots, s_0 = k) = p(s_t \mid s_{t-1}).$$

The probabilistic evolution of the Markov chain is governed by the transition probabilities

$$p(s_t = j \mid s_{t-1} = i) = \pi_{ijt},$$

where $i, j \in \varepsilon$, $0 \le \pi_{ijt} \le 1$, $t = 1, \ldots, T$, and $\sum_{j=0}^{1} \pi_{ijt} = 1$. The matrix of transition probabilities is

$$\Pi(t) = \begin{bmatrix} \pi_{00t} & \pi_{01t} \\ \pi_{10t} & \pi_{11t} \end{bmatrix}, \tag{1}$$

where $\pi_{i1t} = 1 - \pi_{i0t}$ for all $i = 0, 1$ and $t = 1, \ldots, T$.

The Markov chain model that allows for incorporation of covariate effects is the so called Markov regression model setup, which was originally suggested by Cox [5] as the observation driven logistic autoregression model. For the above binary Markov chain in (1), Muenz and Rubinstein [11] related the covariates to the transition probabilities as

$$\pi_{00t}(\tilde{\alpha}'\tilde{z}_m) = \frac{\exp(\tilde{\alpha}'\tilde{z}_m)}{1 + \exp(\tilde{\alpha}'\tilde{z}_m)},$$

$$\pi_{10t}(\tilde{\beta}'\tilde{z}_m) = \frac{\exp(\tilde{\beta}'\tilde{z}_m)}{1 + \exp(\tilde{\beta}'\tilde{z}_m)},$$

where the vector $\tilde{z}_m$ contains p covariates for the mth person in the study of the psychological impact of breast cancer, and $\tilde{\alpha} = (\alpha_0, ..., \alpha_p)$ and $\tilde{\beta} = (\beta_0, ..., \beta_p)$ are the two parameter vectors associated with two rows of the $\Pi(t)$. The parameters were estimated using maximum likelihood estimates.

Erkanli *et al.* [6] presented Bayesian methods, where the initial values of the Markov chain, treated as unknown quantities, were described by a probability distribution and obtained unconditional inferences for the Markov logistic regression models as

$$\operatorname{logit}(\pi_{mt}) = \tilde{\beta}' z_{mt} + \gamma s_{m,t-1}. \tag{2}$$

However, the treatment of nonhomogeneous Markov chains in (2) is via the use of time variant covariates, but not by time varying parameters. In the Bayesian literature, the term Markov regression model may be used to refer to two classes of Markov models, which can be classified as parameter driven and observation driven Markov models, using the terminology of Cox [5]. Markov chain models are observation driven models, whereas the models, such as Cargnoni *et al.* [3], where the parameters evolve over time according to a first-order Markov property, are parameter driven models. In the present study, we consider models that combine both parameter driven and observation driven models for time nonhomogeneous binary Markov chains. We use a full Bayesian approach and propose two dependence structures for describing the time dependence of transition probabilities of the nonhomogeneous chains, which are exchangeability and first-order Markov dependence. The goodness of model fit between two dependence structures will be compared, using Deviance Information Criterion (DIC) measure [13], in the context of application in the data of mental health. The proposed models are implemented using a real data collected in a psychiatric treatment program in Virginia, USA [12].

2. MODELS

2.1 Exchangeable Model

Let y_{mijt} as a binary variable representing the transition of the mth individual from state i at time $(t-1)$ to state j at time t, that is,

$$y_{mijt} = 1(s_{mt} = j \mid s_{mt-1} = i),$$

where $1(A)$ takes value 1 if event A occurs and 0 otherwise. Then, y_{mijt} is a Bernoulli random variable with probability π_{mijt} denoted as

$$(y_{mijt} \mid \pi_{mijt}) \sim Bernoulli(\pi_{mijt}), \tag{3}$$

Now we define a logit transform on π_{mi0t} as

$$\operatorname{logit}(\pi_{mi0t}) = \gamma_t + z_{mt}\beta_{it} \tag{4}$$

for $m = 1, \ldots, M$, $i = 0,1$, and $t = 1, \ldots, T$. Exchangeability of time dependent regression parameters $\{\gamma_t\}$ and $\{\beta_{it}\}$, over time, can be achieved by assuming that $\{\gamma_t\}$ and $\{\beta_{it}\}$ are conditionally independent, over time, given the hyperparameters $\{\mu_\gamma\}$ and $\{\mu_{\beta_i}\}$ respectively and by specifying priors for $\{\mu_\gamma\}$ and $\{\mu_{\beta_i}\}$. For $\{\gamma_t\}$, the exchangeability is achieved by assuming the conditional independence given μ_γ and τ_γ where

$$\gamma_t \mid \mu_\gamma, \tau_\gamma \sim N(\mu_\gamma, \tau_\gamma), \tag{5}$$

and assuming $\tau_\gamma \sim Gamma(a, b)$ with a and b specified. In (5) τ_γ, there is so called 'precision,' which is the inverse of the variance. Further, μ_γ is assumed to have a normal prior. Similarly, for $\{\beta_{it}\}$, we assume that β_{it}'s are conditionally independent, given μ_{β_i} and τ_{β_i} as

$$\beta_{it} \mid \mu_{\beta_i}, \tau_{\beta_i} \sim N(\mu_{\beta_i}, \tau_\beta), \tag{6}$$

where μ_{β_i} is assumed to have a normal prior and $\tau_\beta \sim Gamma(a, b)$ with a and b specified.

Figure 1 is the graphical representation of the hierarchical Bayesian model with the exchangeable time-dependent parameters. This illustrates the model structure, and how parameters at a given time point are related to those at other time points. In this figure, a plate represents repeated components for the range, for example (i in $1 : M$). The arrow shows the specific relationships between two nodes. The descriptions of nodes and arrows are as follows [14]: A constant node $\boxed{c}$ describes a quantity fixed by the study design; stochastic node (y) describes a variable that is given a distribution; a deterministic node describes all logical functions of other nodes; an arrow "$\longrightarrow$" represents the relationship between parent node and descendants; an arrow "$\Rightarrow$" connects two nodes by logical functions.

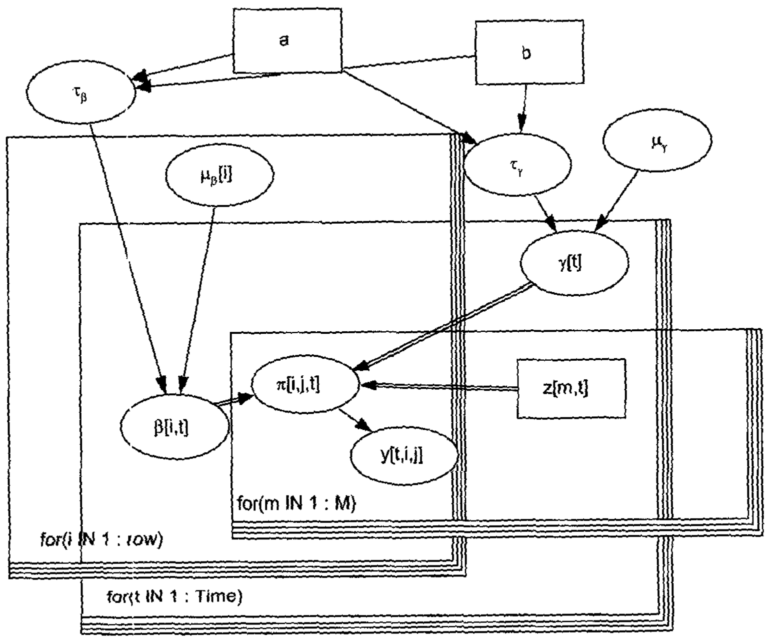

Figure 1. Graphical representation of the exchangeable parameter model

The conditional independence assumptions, as illustrated in Figure 1, are

$$y_{mijt} \perp (\beta_{it}, \gamma_t, \mu_{\beta_i}, \tau_\beta, \mu_\gamma, \tau_\gamma) \mid \pi_{mijt},$$
$$\pi_{mijt} \perp (\mu_{\beta_i}, \tau_\beta, \mu_\gamma, \tau_\gamma) \mid \beta_{it}, \gamma_t.$$

The plates in the figure represent the repetitive structure with respect to the corresponding index. Each node is assumed to be conditionally independent with respect to indices $m = 1, \ldots, M$, $i = 0, 1, j = 0, 1$, and $t = 1, \ldots, T$.

To summarize, the hierarchical setup for the exchangeable model can be represented as

First Level:

$$(y_{mijt} \mid \pi_{mijt}) \sim Bernoulli(\pi_{mijt}),$$

Second Level;

$$\log\text{it}(\pi_{mi0t}) = \gamma_t + z_{mt}\beta_{it},$$
$$(\gamma_t \mid \mu_\gamma, \tau_\gamma) \perp (\beta_{it} \mid \mu_{\beta_i}, \tau_\beta),$$

Third Level:

$$\mu_\gamma \perp \tau_\gamma \perp \mu_{\beta_i} \perp \tau_\beta.$$

2.2 Markov Model

We also consider a stronger form of dependence than exchangeability: a first order Markov property assessed on the model parameters in (4), following the Markov structure used by Grunwald *et al.* [8] and Cargnoni *et al.* [3]. For γ_t and β_{it}, we assume a random walk type of Markov structure as

$$\gamma_t = \gamma_{t-1} + w_{\gamma_t}, \tag{7a}$$

$$\beta_{it} = \beta_{it-1} + w_{\beta_t}, \tag{7b}$$

where w_{γ_t} and w_{β_t} follow independent normal distributions $N(0, \tau_\gamma)$ and $N(0, \tau_\beta)$, respectively for all t. From the above, it follows that

$$\gamma_t \mid \gamma_{t-1}, \tau_\gamma \sim \begin{cases} N(\gamma_{t-1}, \tau_\gamma) \text{ if } t > 0 \\ N(0, \tau_\gamma) \text{ if } t = 0 \end{cases}, \tag{8a}$$

$$\beta_{it} \mid \beta_{it-1}, \tau_\beta \sim \begin{cases} N(\beta_{it-1}, \tau_\beta) \text{ if } t > 0 \\ N(0, \tau_\beta) \text{ if } t = 0 \end{cases}, \tag{8b}$$

where $\tau_\gamma \sim Gamma(a, b)$ and $\tau_\beta \sim Gamma(a, b)$, independently with a and b specified.

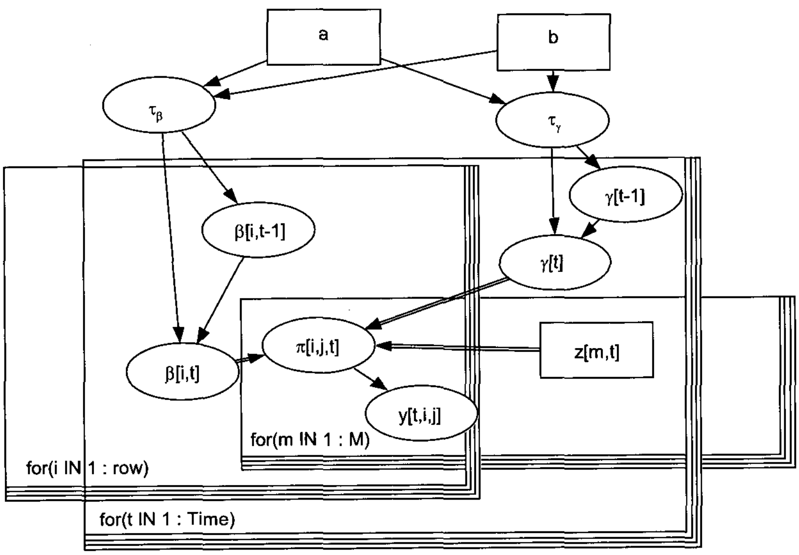

Figure 2. Graphical representation of the Markov parameter model

Figure 2 is the graphical representation of the hierarchical Bayesian model with the first order Markov structure on time-varying parameters.

The conditional independence assumptions, as illustrated in Figure 2, are

$$y_{mijt} \perp (\beta_{it}, \beta_{it-1}, \gamma_t, \gamma_{t-1}, \tau_\beta, \tau_\gamma) \mid \pi_{mijt},$$

$$\pi_{mijt} \perp (\beta_{it-1}, \gamma_{t-1}, \tau_\beta, \tau_\gamma) \mid \beta_{it}, \gamma_t.$$

The plates in the figure represent the repetitive structure with respect to the corresponding index. Each node is assumed to be conditionally independent with respect to indices $m = 1, \ldots, M$, $i = 0, 1$, $j = 0, 1$, and $t = 1, \ldots, T$.

To summarize, the hierarchical setup for the exchangeable model can be represented as

First Level:

$$(y_{mijt} \mid \pi_{mijt}) \sim Bernoulli(\pi_{mijt}),$$

Second Level;

$$\log\text{it}(\pi_{mi0t}) = \gamma_t + z_{mt}\beta_{it},$$

$$(\gamma_t \mid \gamma_{t-1}, \tau_\gamma) \perp (\beta_{it} \mid \beta_{it-1}, \tau_\beta),$$

Third Level:

$$(\tau_\gamma \mid a, b) \perp (\tau_\beta \mid a, b).$$

3. POSTERIOR ANALYSIS

In section 1, two classes of models were presented, employing two dependence structures for model parameters. Bayesian analyses of these models require the evaluation of posterior distribution of unknown parameters. One of the most popular inference procedures in Bayesian analysis is to take the Monte Carlo based simulation technique called Markov chain Monte Carlo (MCMC) methods, which replace the analytic integrations with Monte Carlo integration. Especially, the Gibbs sampler (see Gelfand and Smith [7]), a special MCMC method, is widely used to generate samples from the posterior distributions of interest. Once the Gibbs sampler is implemented and a posterior sample of size G, $\{\pi_{mijt}^1, \ldots, \pi_{mijt}^G\}$ is obtained from the posterior distribution $p(\pi_{mijt} \mid \tilde{S}_1, \ldots, \tilde{S}_M)$, where $\tilde{S}_m = (s_{m0}, \ldots, s_{mT})$, a variety of quantities of interest can be obtained using the

Monte Carlo estimates. For example, the marginal posteriors can be obtained using the full conditionals as

$$p(\pi_{mijt} \mid \tilde{S}_1, \ldots, \tilde{S}_M) \approx \frac{1}{G}\sum_{g=1}^{G} p(\pi_{mijt} \mid \pi_{mijt}^{g(-)}, \tilde{S}_1, \ldots, \tilde{S}_M).$$

Also, the posterior mean of π_{mijt} can be computed as

$$E[\pi_{mijt} \mid \tilde{S}_1, \ldots, \tilde{S}_M] \approx \frac{1}{G}\sum_{g=1}^{G} \pi_{mijt}^{g}.$$

In the following sections, the implementation of the Gibbs sampler is discussed under two different dependence structures of the model parameters.

3.1 Posterior Analysis for Exchangeable Model

The Bayesian analysis requires the full joint posterior distribution, which can be written as

$$\begin{aligned} &p(\Pi_1(1), \ldots, \Pi_M(T), \gamma_1, \ldots, \gamma_T, \beta_{01}, \ldots, \beta_{1T}, \mu_{\beta_0}, \mu_{\beta_1}, \mu_\gamma, \tau_\beta, \tau_\gamma \mid \tilde{S}_1, \ldots, \tilde{S}_M) \\ &\propto \prod_{m=1}^{M}\prod_{i=0}^{1}\prod_{t=1}^{T} p(\tilde{y}_{mit} \mid \gamma_t, \beta_{it}) p(\gamma_t \mid \mu_\gamma, \tau_\gamma) p(\mu_\gamma) p(\tau_\gamma) \\ &\times p(\beta_{it} \mid \mu_{\beta_i}, \tau_\beta) p(\mu_{\beta_i}) p(\tau_\beta), \end{aligned} \tag{9}$$

where $\Pi_m(t) = (\tilde{\pi}_{m0t}, \tilde{\pi}_{m1t})$, $\tilde{S}_m = (s_{m0}, \ldots, s_{mT})$, $\tilde{y}_{mit} = (y_{mi0t}, y_{mi1t})$, $\tilde{\pi}_{mit} = (\pi_{mi0t}, \pi_{mi1t})$ for $m = 1, \ldots, M$, $I = 0, 1$, $t = 1, \ldots, T$.

In what follows, the full conditional posterior distribution of any random quantity ϕ will be denoted by $p(\phi \mid \tilde{S}, \phi^{(-)})$, where $\tilde{S} = (\tilde{S}_1, \ldots, \tilde{S}_M)$. For the full conditional distribution of γ_t, it follows from (9) that

$$\begin{aligned} p(\gamma_t \mid \tilde{S}, \gamma_t^{(-)}) &\propto \prod_{m=1}^{M}\prod_{i=0}^{1} p(\tilde{x}_{mit} \mid \gamma_t, \beta_{it}) p(\gamma_t \mid \mu_\gamma, \tau_\gamma) \\ &\propto \prod_{m=1}^{M}\prod_{i=0}^{1}\left[\left(\frac{\exp(\gamma_t + z_{mt}\beta_{it})}{1+\exp(\gamma_t + z_{mt}\beta_{it})}\right)^{x_{mi0t}}\left(\frac{1}{1+\exp(\gamma_t + z_{mt}\beta_{it})}\right)^{x_{mi1t}}\right] \\ &\times \exp\left\{-\frac{\tau_\gamma}{2}(\gamma_t - \mu_\gamma)^2\right\}. \end{aligned} \tag{10}$$

The above is not a standard form of known distributions and the Gibbs sampler is used to simulate posterior samples.

The full conditional distribution of β_{it} can be written as

$$p(\beta_{it} \mid \tilde{S}, \beta_{it}^{(-)}) \propto \prod_{m=1}^{M} p(\tilde{x}_{mit} \mid \gamma_t, \beta_{it}) p(\beta_{it} \mid \mu_{\beta_i}, \tau_\beta)$$

$$\propto \prod_{m=1}^{M} \left[\left(\frac{\exp(\gamma_t + z_{mt}\beta_{it})}{1 + \exp(\gamma_t + z_{mt}\beta_{it})} \right)^{x_{mi0t}} \left(\frac{1}{1 + \exp(\gamma_t + z_{mt}\beta_{it})} \right)^{x_{mi1t}} \right]$$

$$\times \exp\left\{ -\frac{\tau_\beta}{2} (\beta_{it} - \mu_{\beta_i})^2 \right\}. \tag{11}$$

In (11), the product with respect to the row index i is suppressed, as β_{it} is independent for $i \neq k$.

The full conditional distribution of μ_{β_i} can be written as

$$p(\mu_{\beta_i} \mid \tilde{S}, \mu_{\beta_i}^{(-)}) \propto p(\beta_{it} \mid \mu_{\beta_i}, \tau_\beta) p(\mu_{\beta_i})$$

$$\propto \exp\left[-\frac{1}{2} \{ \tau_\beta (\beta_{it} - \mu_{\beta_i})^2 + \tau_{\mu_\beta} (\mu_{\beta_i} - \mu_{\mu_\beta})^2 \} \right]$$

$$\propto \exp\left[-\frac{C}{2} (\mu_{\beta_i} - B)^2 \right], \tag{12}$$

where $C = \tau_\beta + \tau_{\mu_\beta}$ and $B = C^{-1} \tau_{\mu_\beta} \mu_{\mu_\beta}$. Thus, the full conditional distribution of μ_{β_i} follows the normal distribution with mean B and precision C.

The full conditional distribution of μ_γ follows the normal distribution, similar to (12). The full conditional distribution of τ_β can be written as

$$p(\tau_\beta \mid \tilde{S}, \tau_\beta^{(-)}) \propto p(\beta_{it} \mid \mu_{\beta_i}, \tau_\beta) p(\tau_\beta)$$

$$\propto \tau_\beta^{1/2} \exp\{ -\frac{\tau_\beta}{2} (\beta_{it} - \mu_{\beta_i})^2 \} \tau_\beta^{a-1} \exp(-\tau_\beta / b)$$

$$\propto \tau_\beta^{a^*-1} \exp(-\tau_\beta / b^*), \tag{13}$$

which is a Gamma with $a^* = a + 1/2$, $b^* = (\frac{(\beta_{it} - \mu_{\beta_i})^2}{2} + \frac{1}{b})^{-1}$. The full conditional distribution of τ_γ has similar Gamma distribution to (13).

3.2 Posterior Analysis for Markov Model

The Bayesian analysis requires the full joint posterior distribution, which can be written as

$$\begin{aligned} &p(\Pi_1(1), \ldots, \Pi_M(T), \gamma_1, \ldots, \gamma_T, \beta_{01}, \ldots, \beta_{1T}, \tau_\beta, \tau_\gamma \mid \tilde{S}_1, \ldots, \tilde{S}_M) \\ &\quad \propto \prod_{m=1}^{M} \prod_{i=0}^{1} \prod_{t=1}^{T} p(\tilde{y}_{mit} \mid \gamma_t, \beta_{it}) p(\gamma_t \mid \gamma_{t-1}, \tau_\gamma) p(\tau_\gamma) \\ &\quad \times p(\beta_{it} \mid \beta_{it-1}, \tau_\beta) p(\tau_\beta), \end{aligned} \tag{14}$$

where $\Pi_m(t) = (\tilde{\pi}_{m0t}, \tilde{\pi}_{m1t})$, $\tilde{S}_m = (s_{m0}, \ldots, s_{mT})$, $\tilde{y}_{mit} = (y_{mi0t}, y_{mi1t})$, $\tilde{\pi}_{mit} = (\pi_{mi0t}, \pi_{mi1t})$ for $m = 1, \ldots, M$, $i = 0, 1$, $t = 1, \ldots, T$.

For the full conditional distribution of γ_t, it follows from (14) that

$$\begin{aligned} p(\gamma_t \mid \tilde{S}, \gamma_t^{(-)}) &\propto \prod_{m=1}^{M} \prod_{i=0}^{1} p(\tilde{x}_{mit} \mid \gamma_t, \beta_{it}) p(\gamma_t \mid \gamma_{t-1}, \tau_\gamma) p(\gamma_{t+1} \mid \gamma_t, \tau_\gamma) \\ &\propto \prod_{m=1}^{M} \prod_{i=0}^{1} \left[\left(\frac{\exp(\gamma_t + z_{mt}\beta_{it})}{1 + \exp(\gamma_t + z_{mt}\beta_{it})} \right)^{x_{mi0t}} \left(\frac{1}{1 + \exp(\gamma_t + z_{mt}\beta_{it})} \right)^{x_{mi1t}} \right] \\ &\quad \times \exp\left[-\frac{\tau_\gamma}{2} \left\{ (\gamma_t - \gamma_{t-1})^2 + (\gamma_{t+1} - \gamma_t)^2 \right\} \right]. \end{aligned} \tag{15}$$

The above is not a standard form of known distributions, and the Gibbs sampler is used to simulate posterior samples.

The full conditional distribution of β_{it} can be written as

$$\begin{aligned} p(\beta_{it} \mid \tilde{S}, \beta_{it}^{(-)}) &\propto \prod_{m=1}^{M} p(\tilde{x}_{mit} \mid \gamma_t, \beta_{it}) p(\beta_{it} \mid \beta_{it-1}, \tau_\beta) p(\beta_{it+1} \mid \beta_{it}, \tau_\beta) \\ &\propto \prod_{m=1}^{M} \left[\left(\frac{\exp(\gamma_t + z_{mt}\beta_{it})}{1 + \exp(\gamma_t + z_{mt}\beta_{it})} \right)^{x_{mi0t}} \left(\frac{1}{1 + \exp(\gamma_t + z_{mt}\beta_{it})} \right)^{x_{mi1t}} \right] \\ &\quad \times \exp\left[-\frac{\tau_\gamma}{2} \left\{ (\beta_{it} - \beta_{it-1})^2 + (\beta_{it+1} - \beta_{it})^2 \right\} \right]. \end{aligned} \tag{16}$$

In (16), the product with respect to the row index i is suppressed as β_{it} is independent for $i \neq k$.

The full conditional distributions of τ_β and τ_γ are similar to the Gamma form given in (13), but with a modified scale parameter involving quadratic form $(\beta_{it} - \beta_{it-1})^2$ and $(\gamma_t - \gamma_{t-1})^2$, respectively.

4. APPLICATION TO THE REAL DATA

In this section, the proposed two classes of models are implemented with the real life data reported in Nhan [12]. The longitudinal data presented in Nhan [12] is from a psychiatric treatment study of children and young adolescents in Virginia, USA. The present study uses M=240 subjects for T=4 time periods going through a treatment program to assess the change of patients' functional status over time. The subjects who participated in the study cover ages from 8 to 17 at the time they entered the program. The status of patient functioning is represented by 0 (lower state) and 1 (higher state) and was evaluated every three months after the patients entered the program.

4.1 Prior Specifications

Non-informative, but proper, priors were used to describe the uncertainty about the parameters of proposed models. In the exchangeable model, $(\gamma_t \mid \mu_\gamma, \tau_\gamma) \sim N(\mu_\gamma, \tau_\gamma)$ and $(\beta_{it} \mid \mu_{\beta_i}, \tau_\beta) \sim N(\mu_{\beta_i}, \tau_\beta)$, μ_γ and μ_{β_i} are assumed to follow $N(0, 0.01)$ independently for i's; τ_γ and τ_β are assumed to follow Gamma distribution with parameters (0.01, 0.01).

In the Markov model, $(\gamma_t \mid \gamma_{t-1}, \tau_\gamma) \sim N(\gamma_{t-1}, \tau_\gamma)$ and $(\beta_{it} \mid \beta_{it-1}, \tau_\beta) \sim N(\beta_{it-1}, \tau_\beta)$ for $t > 0$. In this case, τ_γ and τ_β are assumed to follow the Gamma distribution with parameters (0.01, 0.01).

4.2 Results

The inferences about the transition probabilities, as well as the parameters of the proposed models in section 2, are based on the posterior distributions of parameters of interest, and the posteriors were obtained from the Gibbs sampling technique. The models were implemented in the WinBUGS programming environment [14]. After an initial burn-in sample of 100,000 iterations, 2,000 simulated samples of all parameters were saved for parameter inferences. It took about 950 seconds for 100,000 iterations in a Pentium IV 3 GHz Windows XP PC, and the results did not exhibit any serious convergence problems.

The effects of age on the transitional behaviors are captured in β_{it}'s and the posterior densities of β_{it}'s are compared between the Exchangeable and Markov models in Figure 3. β_{it}'s for the Exchangeable model are represented by a line and the Markov model by a dotted line. β_{it}'s for the Exchangeable model are slightly bigger than those of the Markov models, but this difference is compensated by the fixed effect parameter γ_t's, which are smaller for the Exchangeable model than for the Markov model, as shown in Figure 4. Both the fixed effect parameters γ_t's and the age effect parameters β_{it} did not vary much by time, and they increased slightly with time, representing increased age effects over time. Thus, the patients' likelihood of moving to state 1 increased with time. Also, posterior distribution of β_{it}'s showed mostly positive values; this represents that the likelihood of moving to state 1 also increased with age. Thus, the treatment program is more effective for older age groups.

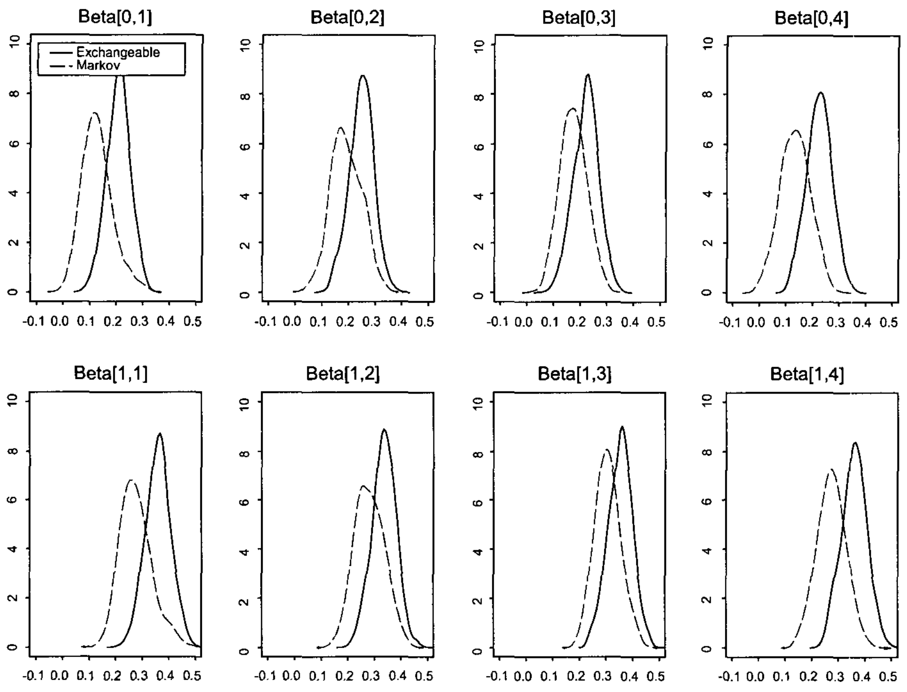

Figure 3. Posterior distributions of β_{it} for i = 0, 1, t = 1,...,4 for the Exchangeable and Markov models

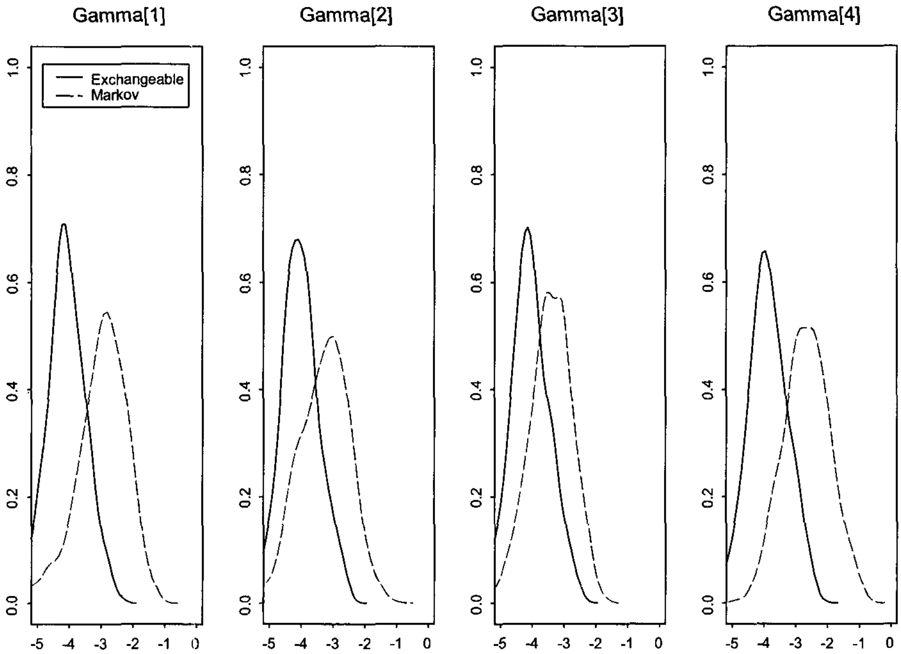

Figure 4. Posterior distributions of γ_t for $t = 1,\ldots,4$ for the Exchangeable and Markov models

While the transition to move from state 0 to state 1 increased both with time and age, the likelihood of remaining in the same state when they were in state 1 was higher than the likelihood of moving from state 0 to state 1.

The goodness of fit was evaluated using the deviance information criterion (DIC), developed by Spiegelhalter *et al.* [13]. Table 1 shows that the DIC is in favor of the Exchangeable model, as implied by the lower DIC value. In the table, D-bar is the posterior mean of parameters, D-hat is a point estimate of the deviance, evaluated using the posterior means of parameters, and P_D is the effective number of parameters. Then, DIC=D-bar + P_D. The actual number of parameters for γ_t and β_{it} is 12, but the effective number of parameters, P_D, is slightly fewer (9.39 and 10.02) than 12. This means that not all of the 12 parameters contributed to explaining the transitional behavior.

Table 1. Comparison of model fit using DIC

Model	D-bar	D-hat	P_D	DIC
Exchangeable	1077.71	1068.32	9.39	1087.10
Markov	1079.84	1069.82	10.02	1089.87

Even though the DIC favored the Exchangeable model, the difference in DIC was not considerable, and the posterior distributions of transition probabilities π_{ijt}'s are almost identical between two models for given i, j and t. Thus, for the illustration purpose, the results from the exchangeable models are presented to make inferences about the transitions. Figure 5 presents the posterior distributions of π_{01t}'s for $t = 1,\ldots,4$ by age. While the effect of age on the transition probabilities was minimal for the younger group, the effect of age became more obvious for the older group and increased with time. Thus, older patients have a higher probability of moving to state 1 from state 0 when they stay longer at the hospital.

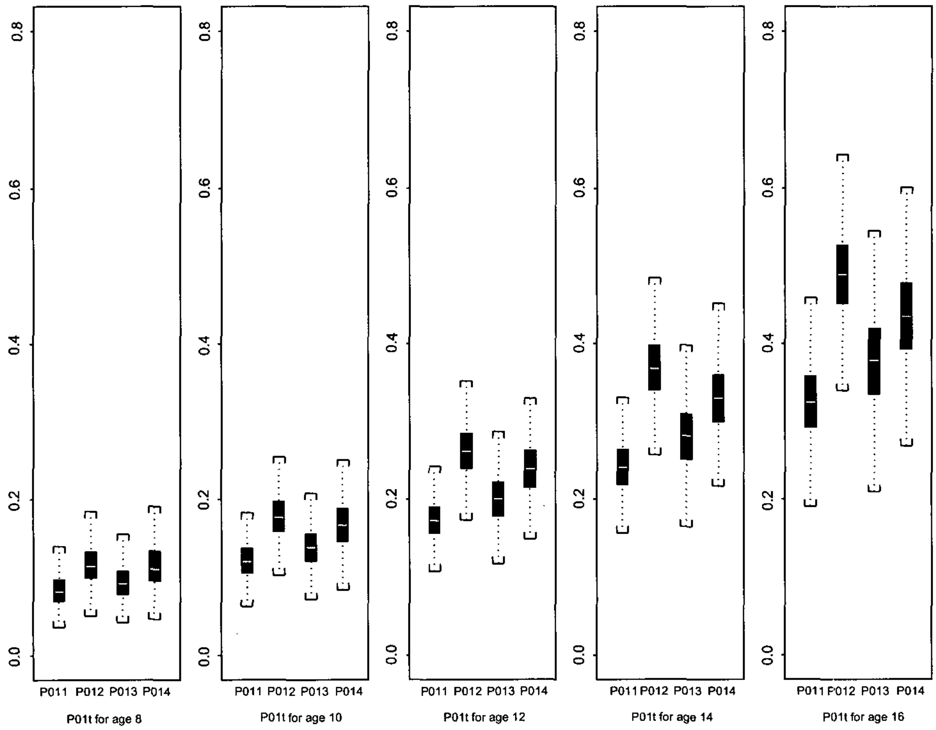

Figure 5. Posterior distributions of π_{01t}'s for $t = 1, \ldots, 4$ by age

Transitional behavior from state 1 to state 0 by age and time is presented in Figure 6. This is the transition moving from a good state to a worse state, and Figure 6 shows that the likelihood of moving from state 1 to state 0 decreased with time and age. Thus, older patients are less likely to be deteriorated than younger ones, and the likelihood of deterioration decreased with time.

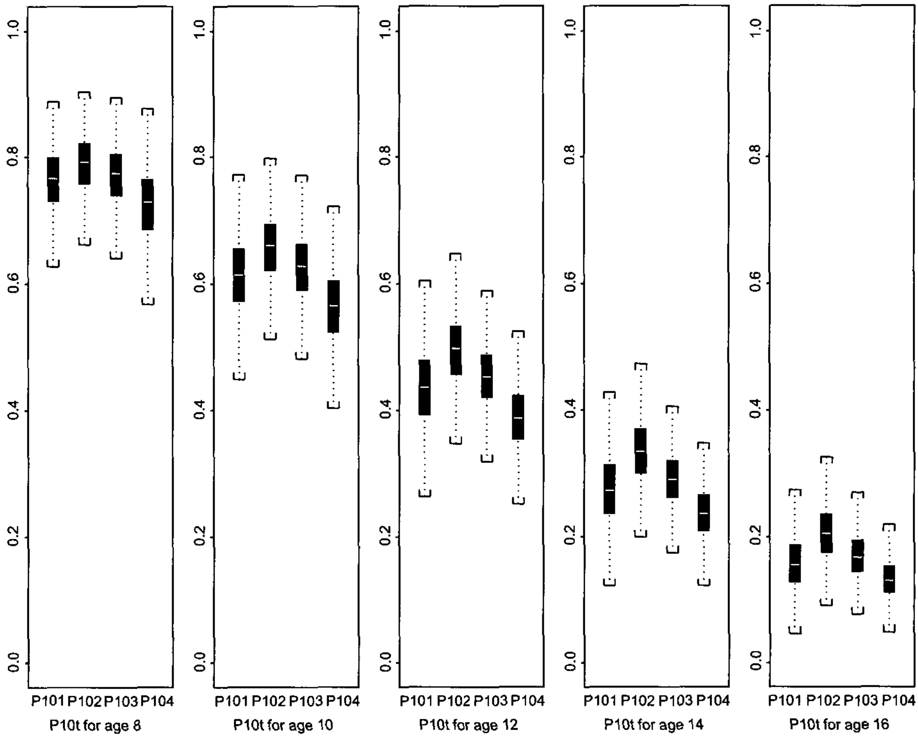

Figure 6. Posterior distributions of π_{10t}'s for $t = 1, \ldots, 4$ by age

5. CONCLUSIONS

In the present study, the parametric modeling strategies to make inferences about transition probabilities of a binary non-homogeneous Markov chain were developed. In so doing, the formal treatment of time non-homogeneity was assessed using two different dependence structures on the time-variant parameters of the logistic model. The exchangeable model described a mild dependence of parameters, while the Markov model described a stronger first order dependence of parameters across time. The proposed modeling strategies were implemented in the WinBUGS programming environment using a real data set collected from a psychiatric treatment program in Virginia, USA.

Posterior distributions of parameters of interest were obtained by the Gibbs sampling technique. DIC measure revealed a slightly better model fit for the exchangeable model. The exchangeability represented in (6) involves a judgment of complete symmetry among $\beta_{i1}, \ldots, \beta_{iT}$, and this means that $\{\beta_{i1}, \ldots, \beta_{iT}\}$ are a random sample from a normal distribution with the unknown mean μ_{β_i} and the unknown precision, the inverse of the variance, τ_β. Thus, we can infer form the DIC measure that the age effects, at different time points, on the transition behavior are similar. In other words, the DIC indicates that β_{it}'s are independent of the time order in which they are collected, rather than having first order dependence in terms of the time. Age effects on transition probabilities were assessed by posterior distributions of transition probabilities. Transitions to a good state from a bad state increased with time and age, whereas transitions to a bad state decreased with time and age.

In the present study, we used a hierarchical Bayesian approach, as the non-Bayesian framework does not allow for modeling of the dynamically evolving structure of the parameters. Although the model was implemented using a data set from a medical treatment program, the application is not limited to the medical area, but can be extended to a variety of fields. For example, in marketing research, changes in preferences for a particular product, represented in a binary scale, can be formulated using a Markov chain.

REFERENCES

[1] Anderson, T. W. and L. A. Goodman, "Statistical inference about Markov chains," *Annals of Mathematical Statistics* 28 (1957), 89-110.

[2] Billingsley, P., "Statistical Methods in Markov chains," *Annals of Mathematical Statistics* 32 (1961), 12-40.

[3] Cargnoni, C., P. Mueller, and M. West, "Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models," *Journal of American Statistical Association* 92 (1997), 640-647.

[4] Chatfield, C., "Statistical inference regarding Markov chain models," *Applied Statistics* 22, 1 (1973), 7-20.

[5] Cox, D. R., "Statistical analysis of time series: Some recent developments," *Scandinavian Journal of Statistics* 8 (1981), 93-115.

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

[7] Gelfand, A. E. and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of American Statistical Association* 85

(1990), 972-985.

[8] Grunwald, G., A. Raftery, and P. Guttorp, "Time series for continuous proportions," *Journal of the Royal Statistical Society, Ser. B* 55 (1993), 103-116.

[9] Lee, T.C., G. G. Judge, and A. Zellner, "Maximum likelihood and Bayesian estimation of transition probabilities," *Journal of American statistical Association* 63 (1968), 1162-1179.

[10] Meshkani, M., "Empirical Bayes estimation of transition probabilities for Markov chains," Ph.D. Dissertation, Florida State University, 1978.

[11] Muenz, L. and L. Rubinstein, "Markov models for covariate dependence of binary sequences," *Biometrics* 41 (1985), 91-101.

[12] Nhan, N., "Effects and outcome of residential treatment," Technical report, Graydon Manor Research Department, USA, 1999.

[13] Spiegelhalter D., N. Best, B. Carlin, and Van der Linde, "Bayesian Measures of Model Complexity and Fit (with Discussion)," *Journal of the Royal Statistical Society, Series B* 64, 4 (2003), 583-616.

[14] Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks, "Bayesian Inference Using Gibbs Sampling Manual (version ii)," MRC Biostatistics Unit, Cambridge University, UK, 1996.