BAYESIAN ANALYSIS OF NONHOMOGENEOUS MARKOV CHAINS: APPLICATION TO MENTAL HEALTH DATA

Minje Sung[1*], Refik Soyer[2], Nguyen Nhan[3]

[1]*School of Business Administration, Ajou University, Suwon, Korea.*
[2]*Department of Management Science, The George Washington University, Washington, DC 20052, USA.*
[3]*Research Director (now retired), Graydon Manor Psychiatric Center, Leesburg, VA, USA.*

**Corresponding author: Tel +82 31 219 2912; fax + 82 31 219 1616; e-mail sungmj@ajou.ac.kr.*

SUMMARY

In this paper we present a formal treatment of nonhomogeneous Markov chains by introducing a hierarchical Bayesian framework. Our work is motivated by the analysis of correlated categorical data which arise in assessment of psychiatric treatment programs. In our development, we introduce a Markovian structure to describe the nonhomogeneity of transition patterns. In so doing, we introduce a logistic regression setup for Markov chains and incorporate covariates in our model. We present a Bayesian model using Markov chain Monte Carlo methods and develop inference procedures to address issues encountered in the analyses of data from psychiatric treatment programs. Our model and inference procedures are implemented to some real data from a psychiatric treatment study.

Key words: Markov models, Bayesian inference, longitudinal data, dynamic models.

1. Introduction

Categorical type longitudinal data often arises in studies of psychiatric treatment programs where measurements describe either the mental status of the patients or their functioning status in the program at different points in time. Modeling the states of the subjects over time, understanding the changing behavior of the patients and related analyses are of interest to scientists who are involved in these studies. Nhan [1] presented an example of data from a psychiatric treatment study of children and young adolescents and discussed such issues of interest. In modeling this type of data, the states measured at discrete points in time are considered as a sequence of correlated discrete random variables. Thus, a Markov chain is typically used to describe the correlation structure. An earlier example of this is the homogeneous Markov chain model proposed by Meredith [2] for evaluation of a treatment program. However, the analysis of this type of data from treatment programs often suggests nonhomogeneous transition patterns for patients. For example, in his study Nhan [1] observed strong evidence in favor of nonhomogeneity in transition probabilities for patients.

In this paper we present a formal treatment of nonhomogeneous Markov chains by introducing a hierarchical Bayesian framework. In the Bayesian literature, the term Markov model may be used to refer to two different classes of models which can be classified as *parameter driven* and *observation driven* Markov models using the terminology of Cox [3]. Both classes of models are used for analysis of categorical time series data. The observation driven Markov models are the Markov chains where the Markov structure is on the observables such as the state occupancies of the individuals. As pointed out by Erkanli et al. [4], most of the work in Bayesian literature concentrated on the parameter driven Markov models such as Cargnoni et al. [5] where the parameters evolve over time according to a first-order Markov model. These models are in the same class as the dynamic linear models (DLM's) of Harrison and Stevens [6] and general DLM's of West et al. [7]. Even though the parameter driven Markov models are not

Markov chains, they are of interest to us in modeling transition matrices for our analysis of nonhomogeneous Markov chains.

Earlier efforts to make inferences on the transition probabilities of a Markov chain can be found in Anderson and Goodman [8] where the maximum likelihood methods are used and in Lee et al. [9] where a Bayesian analysis of the homogeneous Markov chains is presented using a Dirichlet prior distribution on transition probabilities. An empirical Bayes approach is introduced by Meshkani [10] for homogenous chains who considered extensions to nonhomogeneous Markov chains by viewing the problem as a parametric empirical Bayes problem in the sense of Morris [11]. These earlier approaches have not considered the effects of covariates on transition probabilities.

Muenz and Rubinstein [12] presented a logistic regression setup for a binary Markov chain, and obtained the maximum likelihood estimates for the transition probabilities. Zeger and Qaqish [13] presented the Markov logistic regression setup for correlated longitudinal data and discussed maximum likelihood estimation (MLE) for the model. This setup fits into the transition models of Diggle et al. [14] where the Markovian structure on the observations is introduced via the logistic link function. Recently, Erkanli et al. [4] pointed out some of the problems in applying MLE methods in Markov logistic regression setup with only a few number of observations and presented Bayesian methods. However, the work of Erkanli et al. [4] is based on binary Markov logistic regression models and their treatment of nonhomogeneity is via inclusion of time dependent deterministic covariates.

In this paper we present Bayesian methods for modeling and analyses of nonhomogeneous Markov chains, and develop inference procedures to be able to address issues encountered in the analyses of data from psychiatric treatment programs. In so doing, we introduce a class of models for describing nonhomogeneity in the transition probabilities. Our modeling strategy is based on the logistic regression setup of Muenz and Rubinstein [12] and uses a Markovian structure for describing time evolution of

Markov chain's transition matrix. Thus, in the sense of Cox [3], our models can be classified as *parameter* and *observation driven* Markov models.

In section 2, we present a hierarchical Bayes representation of the static logistic regression setup for nonhomogeneous Markov chains using time-dependent covariates. We extend our setup by introducing a first order Markov structure for describing the time dependence of transition probabilities of the nonhomogeneous Markov chains. Bayesian inferences for these models are fully developed in section 3. In section 4, the models are applied to real data from a psychiatric treatment program. Extensions of these models such as analysis of ordinal data and probit regression setup are considered in section 5 and conclusions are presented in section 6.

2. Models for Nonhomogeneous Markov Chains

In this section, we present the Markov chain model and introduce a hierarchical Bayesian representation of the logistic regression setup for Markov chains. We first present the static hierarchical Bayesian representation for nonhomogeneous Markov chains based on time-dependent covariates and then introduce a dynamic Markovian modeling strategy for describing uncertainty about transition probabilities of nonhomogeneous Markov chains.

2.1. Notation and preliminaries

Define $\{s_{m0}, s_{m1}, s_{m2}, ...\}$ as a sequence of random variables indexed by time taking finite values in $\mathcal{E} = \{1, ..., J\}$. We assume that the sequence $\{s_{m0}, s_{m1}, s_{m2}, ...\}$ forms a first-order Markov chain as the conditional probability distribution of s_{mt} given $s_{m,t-1}, \ldots, s_{m0}$ depends only on the value of $s_{m,t-1}$. Here, s_{mt} represents the state of a patient m at time t. Let x_{mijt} represent the transition of the m-th individual from state i at time $(t-1)$ to state j at time t, that is,

$$x_{mijt} = 1(s_{mt} = j | s_{m,t-1} = i), \tag{1}$$

where $1(A)$ takes the value 1 if event A occurs and 0 otherwise. Then, the vector $\boldsymbol{x}_{mit} = (x_{mi1t}, \ldots, x_{miJt})$ is a multinomial random variable with probability vector $\boldsymbol{\pi}_{mit} = (\pi_{mi1t}, \ldots, \pi_{miJt})$ where $\pi_{mijt} = p(s_{mt} = j | s_{m,t-1} = i)$ and $\sum_{j=1}^{J} \pi_{mijt} = 1$. The multinomial model for the transitions from the i-th state of the chain is given by

$$(\boldsymbol{x}_{mit} | \boldsymbol{\pi}_{mit}) \sim Multinomial(\boldsymbol{\pi}_{mit}, 1), \tag{2}$$

for $i, j = 1, \ldots, J$, $t = 1, \ldots, T$. The matrix of transition probabilities $\pi_{mijt}, i, j \in \mathcal{E}$, for individual m is

$$\mathbf{\Pi}_{mt} = \begin{bmatrix} \pi_{m11t} & \cdots & \pi_{m1Jt} \\ \vdots & \ddots & \vdots \\ \pi_{mJ1t} & \cdots & \pi_{mJJt} \end{bmatrix}. \tag{3}$$

where the (i, j)-th entry of the matrix, π_{mijt}, represents a subject's probability of making transition from i-th state to j-th state at time t. If the transition probabilities π_{mijt}'s are not dependent on time t, that is, if $\mathbf{\Pi}_{mt} = \mathbf{\Pi}_m$ for all $t = 1, \ldots, T$, then the Markov chain is called a time *homogeneous Markov chain* whereas the case with time dependent transition probabilities is referred to as a *nonhomogeneous Markov chain*.

2.2. Static logistic regression setup for nonhomogeneous Markov chains

The logistic regression setup of Muenz and Rubinstein [12] for the Markov chains incorporates covariate effects on the transition pattern by using a logit transformation on the transition probabilities of the chain. The earlier treatment of these models presented by Muenz and Rubinstein [12] only deals with binary Markov chains. Their setup can be easily extended for a Markov chain with $J > 2$ states using a multinomial logit transform for the elements of the probability transition vector $\boldsymbol{\pi}_{mit} = (\pi_{mi1t} \ldots \pi_{miJt})'$ for the nonhomogeneous Markov chain by using time dependent covariates while model parameters are static. In what follows we will present the static Bayesian logistic regression setup for the J dimensional Markov chain.

We define the multinomial logit transformation for the elements of the transition vector $\boldsymbol{\pi}_{mit}$ as

$$\eta_{mijt} = logit(\pi_{mijt}) = log(\frac{\pi_{mijt}}{\pi_{miJt}}) = \boldsymbol{F}_{mt}\,\boldsymbol{\theta}^{ij}, \tag{4}$$

for $i = 1, \ldots, J,\; j = 1, \ldots, J-1, t = 1, ..., T$ where $\boldsymbol{F}_{mt}$ is a $1 \times K$ covariate vector for the m-th individual, and $\boldsymbol{\theta}^{ij} = (\theta_{ij1}, \ldots, \theta_{ijK})'$ is a $K \times 1$ vector of regression parameters. We use the J-th category as a baseline category in (4). Thus, the transition probability π_{mijt} is given by

$$\pi_{mijt} = \frac{exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}^{ij})}{\sum_{j=1}^{J} exp(\boldsymbol{F}_{mt}\,\boldsymbol{\theta}^{ij})}. \tag{5}$$

We can write (4) in a more general form as a multivariate logit transformation as

$$\boldsymbol{\eta}_{mit} = \boldsymbol{F}_{mt}\,\boldsymbol{\Theta}^{\boldsymbol{i}}, \tag{6}$$

by defining the $1 \times (J-1)$ logit vector $\boldsymbol{\eta}_{mit} = (\eta_{mi1t} \ldots \eta_{miJt})$ and the $K \times (J-1)$ regression parameter matrix $\boldsymbol{\Theta}^{\boldsymbol{i}}$ as

$$\boldsymbol{\Theta}^{i} = \begin{bmatrix} \theta_{i11} & \cdots & \theta_{i,J-1,1} \\ \vdots & \ddots & \vdots \\ \theta_{i1K} & \cdots & \theta_{i,J-1,K} \end{bmatrix}. \tag{7}$$

We note that $\boldsymbol{\theta}^{ij}$, the regression parameter vector for transition probabilities from state i to j represents the j-th column of (7). Each row of matrix $\boldsymbol{\Theta}^{i}$ represents the effect of the k-th covariate on transitions from state i. We will define the k-th row of (7) as $\boldsymbol{\theta}_k^i = (\theta_{i1k} \ldots \theta_{i,J-1,k})$ and assume that each row of (7) is a multivariate normal vector defined as

$$\boldsymbol{\theta}_k^i | \boldsymbol{\mu}_k^i, \boldsymbol{W}_k \sim MVN(\boldsymbol{\mu}_k^i, \boldsymbol{W}_k), \tag{8}$$

with specified $(J-1) \times 1$ mean vector $\boldsymbol{\mu}_k^i$ and $(J-1) \times (J-1)$ unknown covariance matrix $\boldsymbol{W}_k$. In the model implementation, the prior uncertainty for each $\boldsymbol{\theta}_k^i$ will be described by zero mean vector and unknown covariance matrix. We specify an inverse Wishart prior for $\boldsymbol{W}_k$ as

$$\boldsymbol{W}_k^{-1}|\boldsymbol{R}, v \sim Wish(\boldsymbol{R}, v), \tag{9}$$

where $\boldsymbol{R}$ and v are known quantities and assume that $\boldsymbol{\theta}_k^i$'s, the rows of (7), as well as $\boldsymbol{W}_k$'s are independent of each other for $k = 1, \ldots K$. Furthermore, $\boldsymbol{\Theta}^i$'s are conditionally independent of each other given specified mean vector $\boldsymbol{\mu}_k^i$'s and unknown covariance matrix $\boldsymbol{W}_k$'s for $i = 1, \ldots, J$.

In summary, the static logistic regression setup for nonhomogeneous Markov chains can be represented as a hierarchical Bayesian model as

$$\begin{gathered}
\boldsymbol{x}_{mit}|\,\boldsymbol{\pi}_{mit} \sim Multinomial(\boldsymbol{\pi}_{mit}, 1), \\
\eta_{mijt} = logit(\pi_{mijt}) = \boldsymbol{F}_{mt}\,\boldsymbol{\theta}^{ij}, \\
\boldsymbol{\theta}_k^i|\boldsymbol{\mu}_k^i, \boldsymbol{W}_k \sim MVN(\boldsymbol{\mu}_k^i, \boldsymbol{W}_k), \\
\boldsymbol{W}_k^{-1}|\boldsymbol{R}, v \sim Wish(\boldsymbol{R}, v).
\end{gathered} \tag{10}$$

The hierarchical setup (10) associated with the ith row of the transition matrix $\boldsymbol{\Pi}_{mt}$ is generalized to include $i = 1, \ldots, J$, that is, at the first level of the hierarchy, $\boldsymbol{x}_{mit}$'s are conditionally independent given $\boldsymbol{\pi}_{mit}$'s for $i \neq j$. At the second level, $\boldsymbol{\pi}_{mi}$'s are conditionally independent given $\boldsymbol{\theta}^{ij}$'s for $i \neq j$. The unknown quantities $\boldsymbol{W}_k$**,** that are common for all i's, will induce some form of dependence across the rows of the transition probability matrix. The Bayesian analysis of the hierarchical model (10) will be presented in Section 3.

2.3. Dynamic models for nonhomogeneous Markov chains

The logistic regression setup of the Markov chain described in the previous section is an observation driven Markov model. We next extend the static hierarchical Bayesian representation given by (10) to the dynamic model for nonhomogeneous Markov chains. The time nonhomogeneity of transition probabilities were incorporated into the model by using time dependent covariates $\boldsymbol{F}_{mt}$ in (4). However, in what follows, we consider a formal treatment of nonhomogeneity by introducing a Markovian structure to describe the evolution of transition probabilities over time. The resulting models can be classified as parameter and observation driven Markov models.

In our development we consider the regression parameter matrix of (7) and index it by time as

$$\boldsymbol{\Theta}_t^i = \begin{bmatrix} \theta_{i11t} & \cdots & \theta_{i,J-1,1t} \\ \vdots & \ddots & \vdots \\ \theta_{i1Kt} & \cdots & \theta_{i,J-1,Kt} \end{bmatrix}. \tag{11}$$

We assume a Markov structure on the k-th row of $\boldsymbol{\Theta}_t^i$, that is, on $\boldsymbol{\theta}_{kt}^i = (\theta_{i1kt} \ldots \theta_{i,J-1,kt})$. More specifically following Grunwald et al. [15] and Cargnoni et al. [5], to describe a first order dependence of the time evolving parameters, we assume that the parameter vector $\boldsymbol{\theta}_{kt}^i$ follows a random walk model as

$$\boldsymbol{\theta}_{kt}^i = \boldsymbol{\theta}_{k,t-1}^i + \boldsymbol{\omega}_{kt}^i, \tag{12}$$

where $\boldsymbol{\omega}_{kt}^i$ is a $1 \times (J-1)$ vector of uncorrelated error terms for the parameter vector $\boldsymbol{\theta}_{kt}^i$. We also assume that $\boldsymbol{\omega}_{kt}^i$'s are normally distributed with mean vector **0** and unknown covariance matrix $\boldsymbol{W}_k$ where $\boldsymbol{W}_k^{-1}|\boldsymbol{R}, v \sim Wish(\boldsymbol{R}, v)$ as in (9). The choice of the nonstationary random walk model in (12) reflects a locally constant mean of parameters over time and the model is referred to as the steady model; see for example West et. al [7]. This is a reasonable assumption in the type of applications considered here where the parameters are expected to change in a slow manner from one time period to another. Otherwise, a more general time evolution of parameters can be described by using a first or higher order vector auto-regressive (AR) process on $\boldsymbol{\theta}_{kt}^i$.

Thus, the multivariate logit transformation for the nonhomogeneous chain is given by

$$\boldsymbol{\eta}_{mit} = \boldsymbol{F}_{mt}\,\boldsymbol{\Theta}_t^i,$$

where $\boldsymbol{\eta}_{mit} = (\eta_{mi1t} \ldots \eta_{mi,J-1,t})$. Thus, the logit transform of time dependent transition probability π_{mijt} is defined as

$$\eta_{mijt} = logit(\pi_{mijt}) = log(\frac{\pi_{mijt}}{\pi_{miJt}}) = \boldsymbol{F}_{mt}\,\boldsymbol{\theta}_t^{ij}, \tag{13}$$

where $\boldsymbol{\theta}_t^{ij}$ is the time dependent version of the $K \times 1$ vector of regression parameters in (4), for $i = 1, \ldots, J$, $j = 1, \ldots, J-1$, and $t = 1, \ldots, T$. Again we use the Jth category as a baseline category in (13). We note that time dependence is assumed on a given row of the parameter matrix (11) whereas at a given point in time $\boldsymbol{\theta}_{kt}^i$'s, the rows of (11) are independent for $k = 1, \ldots, K$. As in section 2.2, $\boldsymbol{W}_k$'s are independent of each other for $k = 1, \ldots K$ and at time t, $\boldsymbol{\Theta}_t^i$'s are conditionally independent given $\boldsymbol{\theta}_{k,t-1}^i$'s and $\boldsymbol{W}_k$'s for $i = 1, \ldots, J$. It follows from (12) that

$$(\boldsymbol{\theta}_{kt}^i | \boldsymbol{\theta}_{k,t-1}^i, \boldsymbol{W}_k) \sim N(\boldsymbol{\theta}_{k,t-1}^i, \boldsymbol{W}_k) \quad \text{if } t > 0 \tag{14}$$

and for $t = 0$ we assume that $(\boldsymbol{\theta}_{k0}^i | \boldsymbol{W}_k) \sim N(\mathbf{0}, \boldsymbol{W}_k)$.

Thus, the dynamic logistic regression setup for nonhomogeneous Markov chains can be represented as a hierarchical Bayesian model as

$$\boldsymbol{x}_{mit} | \, \boldsymbol{\pi}_{mit} \sim Multinomial(\boldsymbol{\pi}_{mit}, 1),$$

$$\eta_{mijt} = logit(\pi_{mijt}) = \boldsymbol{F}_{mt}\,\boldsymbol{\theta}_{jt}^i,$$

$$\boldsymbol{\theta}_{kt}^i | \boldsymbol{\theta}_{k,t-1}^i, \boldsymbol{W}_k \sim N(\boldsymbol{\theta}_{k,t-1}^i, \boldsymbol{W}_k),$$

$$\boldsymbol{W}_k^{-1} | \boldsymbol{R}, v \sim Wish(\boldsymbol{R}, v) \text{ and } \boldsymbol{\theta}_{k0}^i | \boldsymbol{W}_k \sim N(\mathbf{0}, \boldsymbol{W}_k). \tag{15}$$

The hierarchical Bayes setup (15) is associated with the ith row of the transition matrix $\boldsymbol{\Pi}_{mt}$ in (3). It can be generalized to include $i = 1, \ldots, J$, that is, at the first level of the hierarchy, $\boldsymbol{x}_{mit}$'s are independent given $\boldsymbol{\pi}_{mit}$'s for $i \neq j$. As before at the second level, $\boldsymbol{\pi}_{mit}$'s are conditionally independent given $\boldsymbol{\theta}_{kt}^i$'s for $i \neq j$. As in the static case, (15) represents the hierarchical setup for individual m.

3. Posterior Analysis of Markov Chain Models

We note that the hierarchical Bayesian setups (10) and (15) are shown for the transitions from the i-th state of the Markov chain for a specific individual m. The generalization of the setup to all states, $i = 1, ..., J$, for M individuals, $m = 1, ..., M$, is straightforward due to the conditional independence of $\boldsymbol{x}_{mit}$'s given the transition probability vectors $\boldsymbol{\pi}_{mit}$'s. In what follows, we will present the Bayesian analyses of both the static and dynamic nonhomogeneous Markov chain models.

3.1. Posterior analysis for static nonhomogeneous Markov chains

Given the transition data on M individuals for T time periods, the joint posterior distribution needed for the Bayesian analysis of homogeneous Markov chains is

$$p(\Pi_{11}, \ldots, \Pi_{MT}, \boldsymbol{\Theta}^1, \ldots, \boldsymbol{\Theta}^J, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_K | \boldsymbol{S}_1, ..., \boldsymbol{S}_M)$$

$$\propto \prod_{m=1}^{M} \prod_{i=1}^{J} \left[\prod_{t=1}^{T} p(\boldsymbol{x}_{mit} | \boldsymbol{\Theta}^i) \right] \prod_{k=1}^{K} p(\boldsymbol{\theta}_k^i | \boldsymbol{\mu}_k^i, \boldsymbol{W}_k) p(\boldsymbol{W}_k), \tag{16}$$

where the components of Π_{mt} represents the transition matrix of the m-th subject, $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_M$ are the observed transitions of M individuals over $t = 1, \ldots, T$ time periods, with $\boldsymbol{S}_m = \{s_{m0}, \ldots, s_{mT}\}$. Since the joint posterior distribution in (16) can not be obtained in any analytically tractable form, we will use a Gibbs sampler to draw samples from the full conditional distributions:

$$(\boldsymbol{\Theta^i} | \boldsymbol{S}, \boldsymbol{\Theta^i}^{(-)}), (\boldsymbol{W}_k | \boldsymbol{S}, \boldsymbol{W}_k{}^{(-)}), \tag{17}$$

where $\boldsymbol{S} = (\boldsymbol{S}_1, ..., \boldsymbol{S}_M)$ and for notational convenience, we denote the full conditional posterior distribution of a random quantity ϕ_i by $p(\phi_i | \boldsymbol{S}, \phi_i^{(-)})$ where $\phi_i^{(-)}$ includes all random quantities except ϕ_i.

For simulating $\boldsymbol{\Theta}^i$, the $K \times J$ matrix of the regression parameters in (7), we have

$$p(\mathbf{\Theta}^i|\boldsymbol{S},\mathbf{\Theta}^{i(-)}) \propto \prod_{m=1}^{M}\left[\prod_{t=1}^{T}p(\boldsymbol{x}_{mit}|\mathbf{\Theta}^i)\right]\prod_{k=1}^{Q}p(\boldsymbol{\theta}_k^i|\boldsymbol{\mu}_k^i,\boldsymbol{W}_k), \tag{18}$$

which can be rewritten as proportional to

$$\prod_{m=1}^{M}\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\frac{exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^i)}{\sum_{j=1}^{J}exp(\boldsymbol{F}_{mt}\,\boldsymbol{\theta}_j^i)}\right)^{x_{mijt}}exp\{-\frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\theta}_k^i-\boldsymbol{\mu}_k^i)'\boldsymbol{W}_k^{-1}(\boldsymbol{\theta}_k^i-\boldsymbol{\mu}_k^i)\}. \tag{19}$$

To draw from $p(\boldsymbol{W}_k|\boldsymbol{S},\boldsymbol{W}_k{}^{(-)})$, we note that the full conditional of $\boldsymbol{W}_k^{-1}$ can be written as

$$\propto |\mathbf{W}_k^{-1}|^{v/2}exp\{-\frac{1}{2}tr\big[\big(\boldsymbol{R}+\sum_{i=1}^{J}(\boldsymbol{\theta}_k^i-\boldsymbol{\mu}_k^i)\,(\boldsymbol{\theta}_k^i-\boldsymbol{\mu}_k^i)'\big)\mathbf{W}_k^{-1}\big], \tag{20}$$

which is a Wishart density with degree of freedom, $v+J+1$, and scale matrix $\frac{1}{2}\big(\boldsymbol{R}+\sum_{i=1}^{J}(\boldsymbol{\theta}_k^i-\boldsymbol{\mu}_k^i)\,(\boldsymbol{\theta}_k^i-\boldsymbol{\mu}_k^i)'\big)$.

3.2. Posterior analysis for dynamic nonhomogeneous chains

Given the transition data on M individuals for T time periods, for the nonhomogeneous Markov chains setup, we need to obtain the joint posterior distribution

$$p(\Pi_{11},\ldots,\Pi_{MT},\mathbf{\Theta}_1^1,\ldots,\mathbf{\Theta}_T^J,\boldsymbol{W}_1,\ldots,\boldsymbol{W}_K|\,\boldsymbol{S}_1,...,\boldsymbol{S}_M)$$

$$\propto \prod_{m=1}^{M}\prod_{i=1}^{J}\prod_{t=1}^{T}p(\boldsymbol{x}_{mit}|\mathbf{\Theta}_t^i)\prod_{k=1}^{K}p(\boldsymbol{\theta}_{kt}^i|\boldsymbol{\theta}_{k,t-1}^i,\boldsymbol{W}_k)p(\boldsymbol{W}_k). \tag{21}$$

For simulating $\mathbf{\Theta}_t^i$, we can use the Markov property as implied by (12) and write

$$p(\mathbf{\Theta}_t^i|\boldsymbol{S},\mathbf{\Theta}_t^{i(-)}) \propto \prod_{m=1}^{M}p(\boldsymbol{x}_{mit}|\mathbf{\Theta}_t^i)\prod_{k=1}^{K}p(\boldsymbol{\theta}_{kt}^i|\boldsymbol{\theta}_{k,t-1}^i,\boldsymbol{W}_k)p(\boldsymbol{\theta}_{k,t+1}^i|\boldsymbol{\theta}_{kt}^i,\boldsymbol{W}_k), \tag{22}$$

implying that $p(\mathbf{\Theta}_t^i|\boldsymbol{S},\mathbf{\Theta}_t^{i(-)})$ is

$$\propto \prod_{m=1}^{M}\Big[\prod_{j=1}^{J}\pi_{mijt}^{x_{mijt}}\Big]exp\Big[-\frac{1}{2}\sum_{k=1}^{K}\Big((\boldsymbol{\theta}_{kt}^{i}-\boldsymbol{\theta}_{k,t-1}^{i})'\boldsymbol{W}_{k}^{-1}(\boldsymbol{\theta}_{kt}^{i}-\boldsymbol{\theta}_{k,t-1}^{i})$$

$$+(\boldsymbol{\theta}_{k,t+1}^{i}-\boldsymbol{\theta}_{kt}^{i})'\boldsymbol{W}_{k}^{-1}(\boldsymbol{\theta}_{k,t+1}^{i}-\boldsymbol{\theta}_{kt}^{i})\Big)\Big]. \qquad (23)$$

Note that the conditional posterior distribution of $\boldsymbol{\Theta}_{t}^{i}$ has a similar form as in (19) except that the product with respect to the time index *t* is suppressed.

To draw from $p(\boldsymbol{W}_k|\boldsymbol{S},\boldsymbol{W}_k{}^{(-)})$, we note that the full conditional of $\boldsymbol{W}_k^{-1}$ can be written as proportional to

$$|\mathbf{W}_k^{-1}|^{(v+T)/2}exp\Big[-\frac{1}{2}tr\Big\{\Big(\boldsymbol{R}+\sum_{i=1}^{J}\sum_{t=1}^{T}(\boldsymbol{\theta}_{kt}^{i}-\boldsymbol{\theta}_{k,t-1}^{i})(\boldsymbol{\theta}_{kt}^{i}-\boldsymbol{\theta}_{k,t-1}^{i})'\Big)\boldsymbol{W}_k^{-1}\Big\}\Big], \qquad (24)$$

which is again a Wishart density with degree of freedom, $v+T+J+1$, and scale matrix $\Big(\boldsymbol{R}+\sum_{i=1}^{J}\sum_{t=1}^{T}(\boldsymbol{\theta}_{kt}^{i}-\boldsymbol{\theta}_{k,t-1}^{i})(\boldsymbol{\theta}_{kt}^{i}-\boldsymbol{\theta}_{k,t-1}^{i})'\Big)/2$.

4. Application to the Data from a Psychiatric Treatment Study

In this section, we will illustrate the implementation of the models introduced in the previous section using the real life longitudinal data reported in Nhan [1]. The data is from a psychiatric treatment study of children and young adolescents in Virginia. The goal of the data analysis is to assess the change of patients' functional status over time. The subjects who participated in the study cover a wide age range of 8-17 years old at the time they entered the program. The treatment program is based on psychodynamic principles and is interdisciplinary in approach. The treatment process involves psychiatry, psychology, social work, special education, child care, nursing, and comprehensive medical services.

The data on various aspects of patient functioning was collected from the treatment team members at regular time intervals during the period of treatment. There are four states that a patient can occupy at each time point where state one indicates the lowest level and state four indicates the highest level of functioning. The data collection started from 30 days after the admission, which was considered time 0, and continued every three months thereafter until the patient was discharged. In our analysis, we use the data on 348 patients for 7 time periods. During this period some patients are discharged from the treatment program and inferring the reasons for discharge is of great interest to psychiatrists. For example, it is important to be able to infer whether patients are discharged because they have responded positively to the treatment.

To reflect the discharges, in our setup we define the $(J+1)$-th state as an absorbing state in the Markov chain implying $p(s_{mt} = J+1 | s_{m,t-1} = J+1) = 1$. Here we assume that the reentry is not allowed. Then, the transition probability matrix of (3) can be modified for the absorbing chain as

$$\Pi_{mt} = \begin{bmatrix} \pi_{m11t} & \cdots & \pi_{m1Jt} & \pi_{m1,J+1,t} \\ \vdots & \ddots & & \vdots \\ \pi_{mJ1t} & & \ddots & \pi_{m1,J+1,t} \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \tag{25}$$

where $\pi_{m,J+1,j,t} = 0$ for $j \neq J+1$.

In the multinomial logit transform (6), we specify $\boldsymbol{F}_{mt} = (1,\, 1, z_{mt})$, and $\boldsymbol{\theta}^{ij} = (\gamma_j,\, \gamma_{ij},\, \beta_{ij})'$ for the static model and $\boldsymbol{\theta}_{\boldsymbol{t}}^{ij} = (\gamma_{jt},\, \gamma_{ijt},\, \beta_{ij})'$ for the dynamic nonhomogeneous chains, where $z_{mt} = Age_{mt}$ is the age of the m-th patient at time t. Thus, we can write

$$\begin{pmatrix} \eta_{mi1t} \\ \vdots \\ \eta_{miJt} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_J \end{pmatrix} + \begin{pmatrix} \gamma_{i1} \\ \vdots \\ \gamma_{iJ} \end{pmatrix} + Age_{mt} \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iJ} \end{pmatrix}, \tag{26}$$

for the static case and

$$\begin{pmatrix} \eta_{mi1t} \\ \vdots \\ \eta_{miJt} \end{pmatrix} = \begin{pmatrix} \gamma_{1t} \\ \vdots \\ \gamma_{Jt} \end{pmatrix} + \begin{pmatrix} \gamma_{i1t} \\ \vdots \\ \gamma_{iJt} \end{pmatrix} + Age_{mt} \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iJ} \end{pmatrix} \tag{27}$$

for the dynamic case.

In (26) vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)'$ represents factors common across the rows whereas the vector $\boldsymbol{\gamma}_i = (\gamma_{i1}, ..., \gamma_{iJ})'$ is row specific and thus describes the row effects on transition probabilities. Time-variant versions of these are defined for (27). In both cases the vector $\boldsymbol{\beta}_i = (\beta_{i1}, ..., \beta_{iJ})'$ represents the covariate effect for row i in the model. Since the discharge state is used as the baseline category, we set $\gamma_{J+1} = \gamma_{i,J+1} = \beta_{i,J+1} = 0$ and $\gamma_{J+1,t} = \gamma_{i,J+1,t} = 0$ for all i's and t's. We note that (26) and (27) can be easily generalized to include $q > 1$ covariates.

4.1. Prior distributions for logistic parameters

In describing prior uncertainty about the unknown model parameters, in all cases we used non-informative but proper priors. More specifically, in the static model we assume independent multivariate normal distributions for parameter vectors $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}_i$, and $\boldsymbol{\beta}_i$. In each of the multivariate normal distributions, we specified zero-mean vectors and the unknown precision matrices, In all cases the scale matrix $\boldsymbol{R}$ of the Wishart was assumed to be $diag(.01,\ .01,\ .01,\ .01)$ implying a high degree of uncertainty.

In the dynamic nonhomogeneous Markov chain model, for time homogeneous parameters we used the same priors as given above for the static case. For the Markovian dependence on parameters, we specified $(\boldsymbol{\gamma}_0 | \boldsymbol{W}_1) \sim MVN(\mathbf{0}, \boldsymbol{W}_1)$ for $t = 0$ and $(\boldsymbol{\gamma}_t | \boldsymbol{\gamma}_{t-1}, \boldsymbol{W}_1) \sim MVN(\boldsymbol{\gamma}_{t-1},\ \boldsymbol{W}_1)$ for $t > 0$. In this case, $\boldsymbol{W}_1^{-1}$ has the same Wishart prior as specified in the above for the static case. For the row specific vector $\boldsymbol{\gamma}_{it}$, we assume that $(\boldsymbol{\gamma}_{i0} | \boldsymbol{W}_2) \sim MVN(\mathbf{0}, \boldsymbol{W}_2)$ for $t = 0$ and $(\boldsymbol{\gamma}_t | \boldsymbol{\gamma}_{t-1}, \boldsymbol{W}_2) \sim MVN(\boldsymbol{\gamma}_{t-1},$ $\boldsymbol{W}_2)$ for $t > 0$, where $\boldsymbol{W}_2^{-1}$ has the same Wishart prior as in the static case.

4.2. Analysis and results

In our analysis, we used a single run of the Gibbs sampler with an initial burn-in sample of 50,000 iterations. After the burn-in sample we simulated an additional 20,000 iterations and obtained a sample of 2,000 realizations from the posterior distributions after thinning at 10th iteration of this sample. This approach was taken to ensure the convergence of the Gibbs sampler. We ran models using 'Age' as a time dependent covariate. The models were implemented using WinBugs 1.4 [17]. The posterior simulated samples of transition probabilities and model parameters did not show any convergence problems. Figure 1 shows trace examples of the three independent sampled chains for $\gamma_{jt}, j = 1, 2,\ t = 1$ and $\beta_{ij},\ i = 1,\ j = 1, ..., 4$ of non-homogeneous model in (27). The Gelman-Rubin convergence statistics [18], approached to 1 after 1,500 monitored iterations in all cases, which indicates convergence of both the pooled and within interval widths to stability. Other parameters also exhibit similar patterns.

FIGURE 1 ABOUT HERE *

We use the deviance information criterion (DIC), a generalization of AIC, developed by Spiegelhalter et al. [19] as a measure of goodness of fit when we compare the static and dynamic nonhomogeneous models. Table 1 shows that the DIC is in favor of the dynamic nonhomogeneous model as implied by the lower DIC value. In the table, D-bar is the posterior mean of the deviance, 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'. The criterion is computed as DIC = D-bar + p$_D$. Note that the effective number of parameters is close to the number of parameters in the static case, but it is considerably smaller in the dynamic nonhomogeneous model indicating that not all time dependent parameters effectively contribute to explaining the transition behavior of the subjects.

*** TABLE 1 ABOUT HERE***

Analysis of the data shows strong evidence in favor of dynamic nonhomogeneity as indicated by the DIC criterion in Table 1. Thus, in the remainder of this section, the results from the dynamic nonhomogeneous Markov chain model will be presented.

In modeling transitions from state i, the effects common to all the rows of the transition matrix are described by γ_{jt}'s whereas γ_{ijt}'s represent the row specific effects on transition to the j-th state at time t. Using the logit transform defined in (26), we can write the odds ratio of making transition to the j-th state from a given row i at time t as

$$\frac{\pi_{mijt}}{\pi_{mi5t}} = exp(\gamma_{jt} + \gamma_{ijt} + \beta_{ij} Age_{mt})$$

which is the odds relative to the transition to the discharge state, that is, state 5 in our case. The above can also be represented as a change in log of the probabilities as

$$log(\pi_{mijt}) - log(\pi_{mi5t}) = \gamma_{jt} + \gamma_{ijt} + \beta_{ij} z_{mt}$$

and the component $(\gamma_{jt} + \gamma_{ijt})$ can be interpreted as the expected change in log probabilities what is not described by the age covariate.

TABLE 2 ABOUT HERE

In Table 2, we present the posterior means and 95% credible intervals of $(\gamma_{jt} + \gamma_{1jt})$ for transitions from state 1. Each posterior summary represents the values of log odds with respect to the discharge state, that is, state 5. We note that as we move from left to right in a given row of the table, the mean of the posterior distribution decreases. This implies that when we control the age effect, as we move to better states, the log transition probability difference between that state and the discarge state, that is, state 5, becomes smaller. Furthermore, this also implies that for transitions from state 1, when we control the age effect, log odds in favor of staying in state 1 is higher than that of moving to a higher state. For example, at time period 4, the subjects are most likely to remain at the same state (that is, at state 1), but they are more likely to exit than move to state 4 as reflected by the negative log odds term. Similar insights can be obtained from posterior summaries associated with transitions from other rows.

Figure 2 shows how the transition probabilities from state 2, that is, π_{2jt}'s for $j = 1, \ldots, 5$ differ by time ($t = 1, \ldots, 7$), and age (ranging from 8 to 17). From state 2, the transition probabilities to state 3 or 4 increase with age but not with time, and the age effect is stronger at earlier time points than later. Transition probabilities to state 1 or 2 decrease with age, implying that the older subjects are more likely to make progress in the treatment program. The likelihood of discharge rapidly increases with time and age, and this implies that as time passes the subjects will exit the program either because they get better or because they do not show much improvement.

*** FIGURE 2 ABOUT HERE***

From the analysis, it appears that older children are more likely to make an improvement than younger children. To assess the effect of age on improvement when the time effect is controlled, we can compare posterior summaries for β_{ij}'s that are given in Table 3. The table shows posterior means and 95% credible intervals of age effect. We note that, in a given row i, as we move from left to right columns, the values become less negative implying that the likelihood of moving to a better state increases with age.

TABLE 3 ABOUT HERE

In evaluating a treatment program, it is of interest to infer how likely to be discharged from a given state as well as to infer the reason of these disharges. In other words, given that a patient is at state i at time $t - 1$, we are interested in assessing how likely it is for this patient to be discharged at time t. Note that this is helpful to be able to infer whether patients are discharged because they have responded positively to the treatment program. The posterior distributions of discharge probabilities from each state are illustrated in Figure 3 for time periods $t = 1, \ldots, 7$ and ages from 8 to 17. We note that in each case in Figure 3 the discharge probability increases with time and age regardless of the prior state. The discharge probabilities do not seem to differ much from one state to the other upto period $t = 3$. After period $t = 4$, increase in discharge probability seems to be accelerated from states 1 and 4. As before, this implies that as time passes patients will exit the program either because they get better or because they do not show much improvement.

FIGURE 3 ABOUT HERE

In our experienct we have found that the posterior results presented in this section were robust. We note that the use of different forms of priors on the parameters have resulted in very similar results.

5. Model Extensions

The models presented in previous sections used a multinomial transformation on the transition probabilities of nonhomogeneous Markov chains. In this section, we incorporate the ordinal nature of the data and consider a cumulative logit model for the nonhomogeneous Markov chains. The cumulative logit model can be specified in terms of cumulative transition probabilities

$$Q_{mijt} = p(s_{mt} \leq j | s_{m,t-1} = i) = \sum_{j=1}^{J} \pi_{mijt},\ j = 1, ..., J.$$

as

$$logit(Q_{mijt}) = log(\frac{Q_{mijt}}{1-Q_{mijt}})$$

for $m = 1, ..., M, i = 1, ..., J, j = 1, ..., J, t = 1, ..., T$ where Q_{mijt} is the probability that the mth individual will move to state j or a worse state at time t from state i. The cumulative logit can be written as

$$logit(Q_{mijt}) = \delta_{ij} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i, \tag{28}$$

where δ_{ij} is cut point parameter, and $\boldsymbol{\theta}_t^i = (\theta_{i1t}, ..., \theta_{iKt})'$ is a $K \times 1$ vector of regression parameters. To explain how the ordinal response is generated, we introduce a latent variable Y_{mit} having cdf $G(\delta_{ij} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i)$ and $s_{mt} = j$ if $\delta_{i,j-1} < Y_{mit} < \delta_{ij}$ [20]. As Johnson and Albert [21] pointed out, the multivariate normal densities are not well suited for generating candidate vectors for ordinal parameters due to the ordering constraints imposed on the cutoff points. Thus, we used independent truncated normal distributions for each component of the cut parameter vector $\boldsymbol{\delta}_i = (\delta_{i1}, ..., \delta_{iJ})$ so that $\delta_{i0} \leq \delta_{i1} \leq ... \leq \delta_{iJ}$ for $i = 1, ..., J$. For identifiability we set $\delta_{i0} = 0$. The precision parameter τ_δ is assumed to follow Gamma(a, b) with a and b specified. We use the similar Markov structure on $\boldsymbol{\theta}_t^i$ to the form in (14). We assume a multivariate normal distribution for $\boldsymbol{\theta}_t^i$ as

$$(\boldsymbol{\theta}_t^i | \boldsymbol{\theta}_{t-1}^i, \boldsymbol{W}_\theta) \sim N(\boldsymbol{\theta}_{t-1}^i, \boldsymbol{W}_\theta) \text{ if } t > 0 \tag{29}$$

and for $t = 0$, $(\boldsymbol{\theta}_0^i | \boldsymbol{W}_\theta) \sim N(\mathbf{0}, \boldsymbol{W}_\theta)$, where $\boldsymbol{W}_\theta$ is a K-dimensional diagonal matrix defined as $\boldsymbol{\tau_\theta I}_K$ and $\boldsymbol{\tau}_\theta = (\tau_{\theta_1}, ..., \tau_{\theta_K})$, where $\tau_{\theta_k} = 1/\sigma^2_{\theta_k}$.

As $\boldsymbol{\theta}_t^i < \mathbf{0}$ in (28), s_{mt} tends to be larger at higher values of covariates, which is more meaningful than the case with $\boldsymbol{\theta}_t^i > \mathbf{0}$. The transition probability π_{mijt} is obtained as $\pi_{mijt} = Q_{mijt} - Q_{mi,j-1,t}$ for $j = 2, ..., J+1$, and $\pi_{mi1t} = Q_{mi1t}$.

The full conditional posterior distribution of the cutpoint δ_{ij} can be obtained as

$$p(\delta_{ij}|\boldsymbol{S}, \delta_{ij}^{(-)}) \propto \prod_{m=1}^{M}\prod_{j=1}^{J}\Big[\prod_{t=1}^{T} p(\boldsymbol{x}_{mit}|\delta_{ij}, \boldsymbol{\theta}_t^i)\Big] p(\delta_{ij}|\tau_\delta)$$

$$\propto \prod_{m=1}^{M}\prod_{j=1}^{J}\Big[\prod_{t=1}^{T}\Big(Q(\delta_{ij} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i) - Q(\delta_{i,j-1} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i)\Big)^{x_{mijt}} exp(-\frac{\tau_\delta}{2}(\delta_{ij} - \mu_\delta)^2).$$

The full conditional posterior distribution of $\boldsymbol{\theta}_t^i$ is given by

$$p(\boldsymbol{\theta}_t^i|\boldsymbol{S}, \boldsymbol{\theta}_t^{i(-)}) \propto \prod_{m=1}^{M} p(\boldsymbol{x}_{mit}|\delta_{ij}, \boldsymbol{\theta}_t^i)\prod_{k=1}^{K} p(\theta_{ikt}|\theta_{ik,t-1}, \tau_{\theta_k})p(\theta_{ik,t+1}|\theta_{ikt}, \tau_{\theta_k})$$

$$\propto \prod_{m=1}^{M}\Big[\prod_{j=1}^{J}\Big(Q(\delta_{ij} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i) - Q(\delta_{i,j-1} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i)\Big)^{x_{mijt}}\Big]$$

$$\times\, exp\Big[-\frac{1}{2}\sum_{k=1}^{K}\Big((\theta_{ikt} - \theta_{ik,t-1})\tau_{\theta_k} + (\theta_{ik,t+1} - \theta_{ikt})\tau_{\theta_k}\Big)\Big].$$

The full conditional of τ_δ can be written as proportional to

$$\tau_\delta^{J^2/2+a-1} exp\Big[-\frac{\tau_\delta}{2}(\sum_{i,j}(\delta_{ij} - \mu_\delta)^2 + 2b)\Big], \tag{30}$$

which is a gamma distribution with parameters $(J^2/2 + a,\ \frac{1}{2}(\sum_{i,j}(\delta_{ij} - \mu_\delta)^2 + 2b))$. The unknown precision τ_{θ_k} for θ_{ikt} has the same form as in (29) with parameters appropriately defined.

For implementation, we used non-informative but proper priors. The ordering $\delta_{i0} \le \delta_{i1} \le ... \le \delta_{i,J-1}$ for $i = 1, ..., J$ of the is described by independently and normally distributed increments N(0, τ_δ). We used Gamma(0.01, 0.01) priors for precision parameters, τ_δ and τ_{θ_k}. The model was implemented in WinBugs 1.4 and the mixing was very fast compared to multinomial logit models. Figure 4 was obtained based on the 2,000 simulated posterior samples after 50,000 burn-in iterations. Basically, the transition patterns in Figure 4 are very similar to those of Figure 2. Thus, cumulative logit model provided almost identical results as multinomial logit model.

*** Figure 4 about here****

An alternative modeling strategy for nonhomogeneous Markov chains can be developed by using the ordinal nature of the data and using a dynamic probit model. This can be achieved by introducing latent variables in our setup similar to Albert and Chib's [22] treatment of probit models. More specifically, we define a continuous latent variable Z_{mit} such that

$$x_{mijt} = \begin{cases} 1(s_{mt} = j|s_{m,t-1} = i) \times 1(\delta_{i,j-1} < Z_{mit} \leq \delta_{ij}) \\ 0 \ \text{ otherwise,} \end{cases} \tag{31}$$

where δ_{ij}'s are the unknown cut points as before. If we assume that $Z_{mit} \sim (\boldsymbol{F}_{mt}\boldsymbol{\theta}_t^{\boldsymbol{i}}, 1)$, then we can write the cumulative probabilities as

$$Q_{mijt} = \Phi(\delta_{ij} - \boldsymbol{F}_{mt}\boldsymbol{\theta}_t^i),$$

where $\boldsymbol{F}_{mt}$ is a $1 \times K$ covariate vector and $\boldsymbol{\theta}_t^i$ is a $K \times 1$ vector of regression parameters. We can assume that each vector $\boldsymbol{\theta}_t^i$ follows (29), where $\boldsymbol{W}_\theta$ has an inverse Wishart prior with scale matrix $\boldsymbol{R}$ and degrees of freedom $r > K$.

The above setup can be represented as a dynamic linear model (DLM) in the sense of West and Harrison [23] as

$$\boldsymbol{Z}_{it} = \boldsymbol{F}_t\boldsymbol{\theta}_t^i + \boldsymbol{u}_{it}$$

$$\boldsymbol{\theta}_t^i = \boldsymbol{\theta}_{t-1}^i + \boldsymbol{w}_{it}$$

where $\boldsymbol{Z}_{it}$ is a $M \times 1$ vector of observations, $\boldsymbol{F}_t$ is a $M \times K$ matrix of covariates, $\boldsymbol{u}_{it}$'s are uncorrelated and independent multivariate normal error vectors with mean $\mathbf{0}$ and variance-covariance matrix is given by M dimensional identity matrix $\boldsymbol{I}_M$, and $\boldsymbol{w}_{it}$'s are normal with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{W}_\theta$. Given the DLM structure it is possible to develop a Gibbs sampler using ideas of Albert and Chib [22] and posterior simulation methods presented in Chapter 15 of West and Harrison [23] for DLMs. More specifically given the transition data on M individuals for T time periods, latent variables Z_{mit}'s, and the cut points we can design a Gibbs sampler we can directly draw from the joint posterior distribution of $\boldsymbol{\theta}_t^i$'s using the *forward filtering backward sampling* algorithm which is given in West and Harrison [23] to draw from multivariate normal densities. In this case the full conditional of $\boldsymbol{W}_\theta$ can be shown to be a Wishart density and the full conditionals of Z_{mit}'s are conditionally independent truncated normal densities. The full conditionals of cutpoints δ_{ij}'s can not be obtained in any known form when we use independently and normally distributed increments as in the logistic setup case. However, one can use the type of priors considered by Albert and Chib [22] and implement the Gibbs sampler.

The attractive feature of the probit setup is in that it enables us to draw from known full conditional distributions with the exception of the cutpoints δ_{ij}'s. Thus, in this case the Gibbs sampler can be implemented in programming environments other than WinBUGS. However, our experience has shown that draws from the full conditionals of cutpoints may turn out to be unstable and problematic. Thus, implementation of the method requires further study and this will be considered elsewhere.

6. Conclusions

In this paper, we presented Bayesian methods for modeling and analyses of nonhomogeneous Markov chains. In so doing, we developed inference procedures to be able to address issues encountered in the analyses of data from psychiatric treatment programs. As posterior distributions of parameters of interest could not be obtained in analytically tractable forms, we used simulation (MCMC) based approaches in developing inferences for the models. The proposed models were implemented using real data from a psychiatric treatment program and various type of insights that can be obtained from the Bayesian analysis were illustrated.

The application of the methodology developed in the present study is not limited to psychiatry and can be extended to other application areas in engineering and sciences.

REFERENCES

1. Nhan N. *Effects and outcome of residential treatment*. Technical Report. Graydon Manor Research Department, VA, 1999.
2. Meredith J. Program evaluation in a hospital for mentally retarded persons. *American Journal of Mental Deficiency* 1974; 78:471-481.
3. Cox DR. Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics* 1981; 8:93-115.
4. Erkanli A, Soyer R, Angold A. Bayesian analyses of longitudinal binary data using Markov regression models of unknown order. *Statistics in Medicine* 2001; 20:755-770.
5. Cargnoni C, Mueller P, West M. Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association* 1997; 92:640-647.
6. Harrison P, Stevens C. Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Ser. B* 1976; 38:205-247.
7. West M, Harrison J, Migon H. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* 1985; 80:73-97.
8. Anderson TW, Goodman LA. Statistical inference about Markov chains. *Annals of Mathematical Statistics* 1957; 28:89-110.
9. Lee TC, Judge GG, Zellner A. *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*. North-Holland and Pub. Co.: Amsterdam, 1970.
10. Meshkani M. *Empirical Bayes estimation of transition probabilities for Markov chains*. Ph.D. Dissertation. Florida State University, 1978.
11. Morris CN. Parametric empirical Bayes inference: Theory and applications. *Journal of American Statistical Association* 1983; 78:47-65.

12. Muenz L, Rubinstein L. Markov models for covariate dependence of binary sequences. *Biometrics* 1985; 41:91-101.
13. Zeger S, Qaqish B. Markov regression models for time series: A quasi-likelihood approach. *Biometrics* 1988; 44:1019-1031.
14. Diggle P, Liang K, Zeger S. *Analysis of Longitudinal Data*. Oxford Science Publications: Oxford, 1994.
15. Grunwald G, Raftery A, Guttorp P. Time series for continuous proportions. *Journal of the Royal Statistical Society, Ser. B* 1993; 55:103-116.
16. Gilks W, Wild P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Ser. B* 1992; 41:337-348.
17. Spiegelhalter D, Thomas A, Best N, Gilks W. Bayesian Inference Using Gibbs Sampling Manual (version ii). MRC Biostatistics Unit, Cambridge University, 1996.
18. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science 1992; 7:457-511.
19. Spiegelhalter D, Best N, Carlin BR,van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Ser. B* 2002; 64:583-616.
20. Agresti A. *Categorical Data Analysis*. John Wiley & Sons: New York, 1990.
21. Johnson VE, Albert JH. *Ordinal data modeling*. Springer: New York, 1999.
22. Albert JH, Chib, S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 1993; 88:669-679.
23. West M, Harrison J. *Bayesian forecasting and dynamic models*. Springer: New York, 1997.

Table 1 DIC comparison between two classes of models

	D-bar	D-hat	p_D	DIC
static model	5070.47	5047.40	23.07	5093.54
dynamic model	4875.13	4830.87	44.26	4919.40

Table 2. Posterior means and 95% credible interval of fixed effects for transitions from State 1.

	$\gamma_{1t}+\gamma_{11t}$		$\gamma_{2t}+\gamma_{12t}$		$\gamma_{3t}+\gamma_{13t}$		$\gamma_{4t}+\gamma_{14t}$	
	Mean	95%CI	Mean	95%CI	Mean	95%CI	Mean	95%CI
$t=1$	7.78	6.14,9.53	6.24	4.63,7.60	3.88	2.54,5.39	2.13	0.15,4.11
$t=2$	6.25	4.25,7.62	4.90	3.51,6.14	2.67	1.04,4.04	0.96	-0.85,2.44
$t=3$	5.30	3.63,6.90	3.87	2.39,5.32	1.28	-0.56,2.66	-0.82	-2.85,0.62
$t=4$	5.09	3.37,6.64	3.70	2.09,5.14	1.19	-0.88,2.61	-0.87	-3.10,0.47
$t=5$	4.73	3.11,6.21	3.38	1.80,4.84	0.95	-0.59,2.49	-0.12	-3.14,0.63
$t=6$	3.85	2.02,5.30	2.45	0.70,3.82	-0.05	-1.81,1.32	-2.30	-4.63,-0.84
$t=7$	3.42	1.63,4.92	2.01	0.26,3.38	-0.44	-2.32,0.91	-2.72	-4.83,-1.06

Table 3. Comparison of posterior means and 95% credible interval for age effect β_{ij}

	β_{i1}		β_{i2}		β_{i3}		β_{i4}	
	Mean	95%CI	Mean	95%CI	Mean	95%CI	Mean	95%CI
$i=1$	-0.25	$-0.35,-0.13$	-0.18	$-0.28,-0.07$	-0.08	$-0.17,0.04$	-0.04	$-0.12,0.09$
$i=2$	-0.41	$-0.51,-0.33$	-0.27	$-0.34,-0.20$	-0.16	$-0.23,-0.08$	-0.13	$-0.20,-0.04$
$i=3$	-0.49	$-0.57,-0.41$	-0.31	$-0.39,-0.23$	-0.15	$-0.23,-0.09$	-0.10	$-0.19,-0.01$
$i=4$	-0.63	$-0.71,-0.51$	-0.43	$-0.52,-0.34$	-0.26	$-0.35,-0.17$	-0.12	$-0.20,-0.04$

Figure 1. Convergence monitoring plots of 3 independent chains for $\gamma_{jt}, j = 1, 2, t = 1$, $\beta_{ij}, i = 1, j = 1, ..., 4$ in non-homogeneous model: GR=Gelman Rubin's [18] scale reduction factor, 97.5%=97.5 percentile of GR

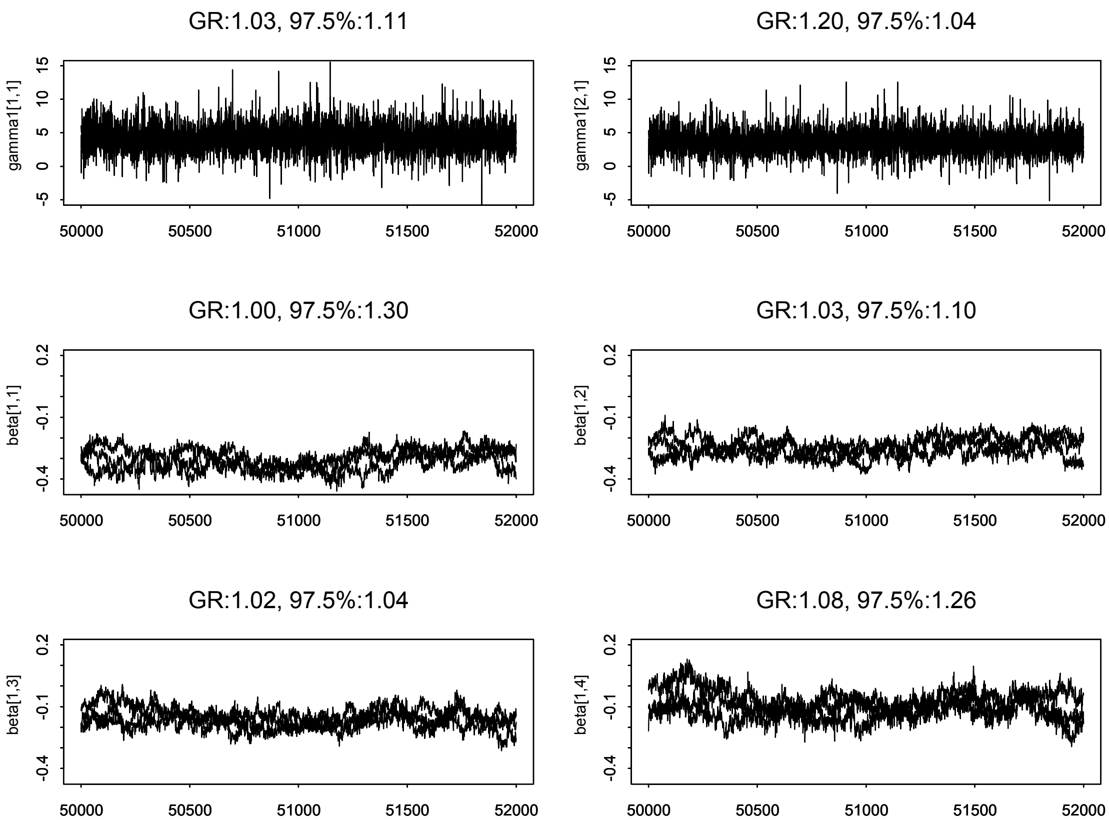

Figure 2. Posterior means of transition probabilities from state 2 at different time points and ages

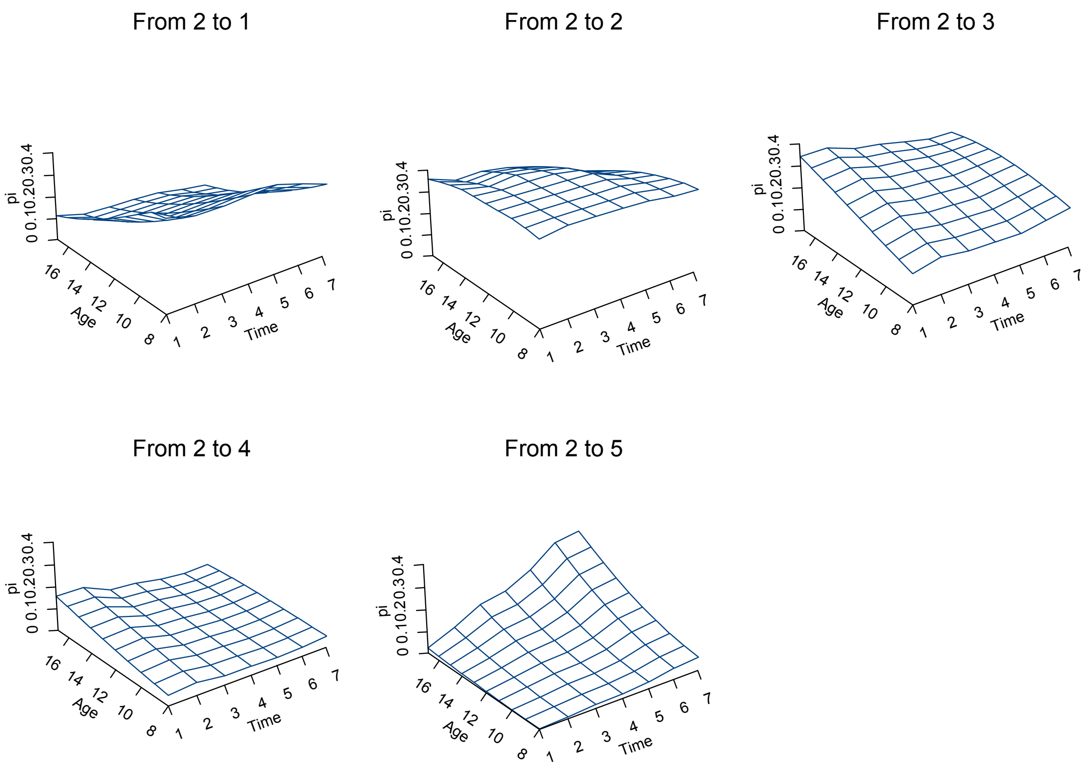

Figure 3. Posterior distributions of exit probability (π_{i5t}'s) from each state by age and time

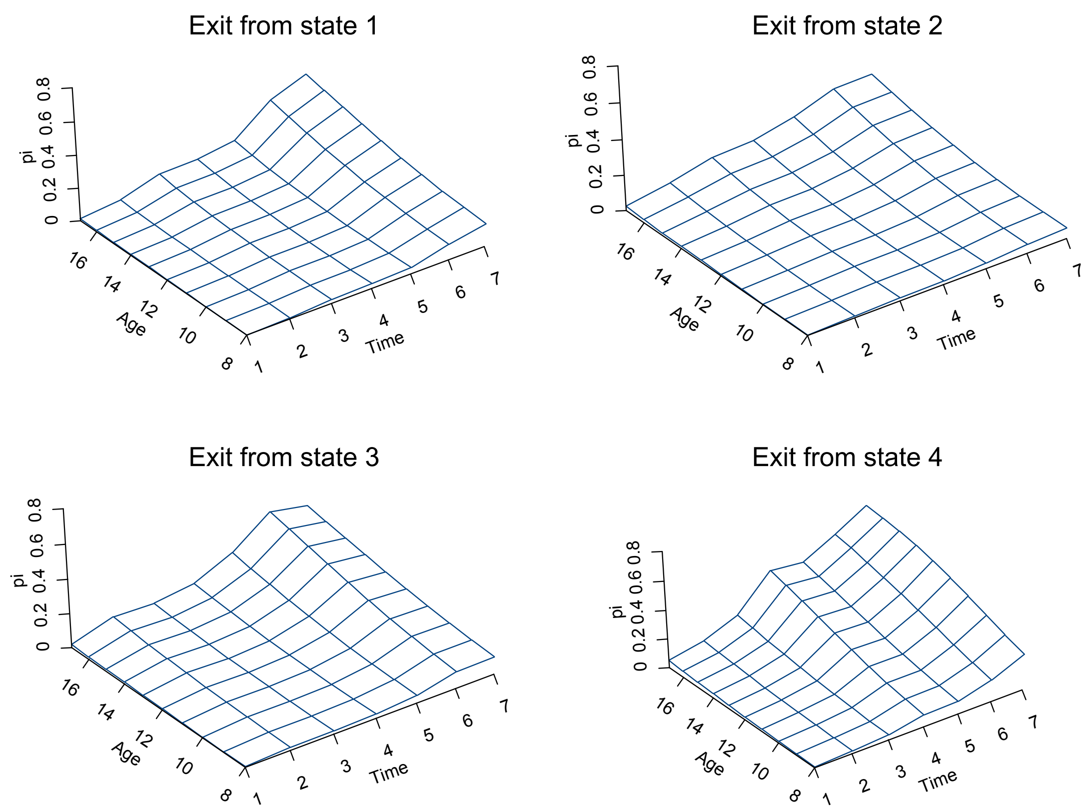

Figure 4. Posterior mean transition probabilities obtained from the cumulative logit model: Transitions from state 2 at different time points and ages