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

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

[1]*Department of Biostatistics and Bioinformatics, Box 3454, Duke University Medical Center, Durham, NC 27710, USA.*
[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 +1 919 687 4686, x-294; fax + 1 919 687 4737; e-mail msung@psych.duhs.duke.edu.*

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 of these models are used for 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 homogeneous Markov chains. 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 and conclusions are presented in section 5.

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 hierarchical Bayesian representation for homogeneous Markov chains 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. Logistic regression setup for homogeneous 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 $\boldsymbol{\pi}_{mi} = (\pi_{mi1} \ldots \pi_{miJ})'$ for the homogeneous Markov chain, where $\mathbf{\Pi}_{mt} = \mathbf{\Pi}_m$ for all $t = 1, \ldots, T$. In what follows we will present the Bayesian logistic regression setup for the J dimensional Markov chain.

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

$$\eta_{mij} = logit(\pi_{mij}) = log(\frac{\pi_{mij}}{\pi_{miJ}}) = \boldsymbol{F}_m \boldsymbol{\theta}^{ij}, \tag{4}$$

for $i = 1, \ldots, J,\ j = 1, \ldots, J-1,$ where $\boldsymbol{F}_m$ is a $1 \times Q$ covariate vector for the m-th individual, and $\boldsymbol{\theta}^{ij} = (\theta_{ij1} \ldots \theta_{ijQ})'$ is a $Q \times 1$ vector of regression parameters. We use the J-th category as a baseline category in (4). Thus, the transition probability π_{mij} is given by

$$\pi_{mij} = \frac{exp(\boldsymbol{F}_m \boldsymbol{\theta}^{ij})}{\sum_{j=1}^{J} exp(\boldsymbol{F}_m \boldsymbol{\theta}^{ij})}. \tag{5}$$

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

$$\boldsymbol{\eta}_{mi} = \boldsymbol{F}_m \boldsymbol{\Theta}^{\boldsymbol{i}}, \tag{6}$$

by defining the $1 \times J$ logit vector $\boldsymbol{\eta}_{mi} = (\eta_{mi1} \ldots \eta_{miJ})$ and the $Q \times J$ regression parameter matrix $\boldsymbol{\Theta}^{\boldsymbol{i}}$ as

$$\boldsymbol{\Theta}^i = \begin{bmatrix} \theta_{i11} & \cdots & \theta_{iJ1} \\ \vdots & \ddots & \vdots \\ \theta_{i1Q} & \cdots & \theta_{iJQ} \end{bmatrix}. \tag{7}$$

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 q-th covariate on transitions from state i. We will define the q-th row of (7) as $\boldsymbol{\theta}_q^i = (\theta_{i1q} \ldots \theta_{iJq})$ and assume that each row of (7) is a multivariate normal vector defined as

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

with specified $J \times 1$ mean vector $\boldsymbol{\mu}_q^i$ and $J \times J$ unknown covariance matrix $\boldsymbol{W}_q$. We specify an inverse Wishart prior for $\boldsymbol{W}_q$ as

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

where $\boldsymbol{R}$ and k are known quantities and assume that $\boldsymbol{\theta}_q^i$'s, the rows of (7), as well as $\boldsymbol{W}_q$'s are independent of each other for $q = 1, \ldots Q$. Furthermore, $\boldsymbol{\Theta}^i$'s are conditionally independent of each other for $i = 1, \ldots, J$.

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

$$\boldsymbol{x}_{mit}|\,\boldsymbol{\pi}_{mi} \sim Multinomial(\boldsymbol{\pi}_{mi}, 1),$$
$$\eta_{mij} = logit(\pi_{mij}) = \boldsymbol{F}_m\,\boldsymbol{\theta}^{ij},$$
$$\boldsymbol{\theta}_q^i|\boldsymbol{\mu}_q^i, \boldsymbol{W}_q \sim MVN(\boldsymbol{\mu}_q^i, \boldsymbol{W}_q),$$
$$\boldsymbol{W}_q^{-1}|\boldsymbol{R}, k \sim Wish(\boldsymbol{R}, k). \tag{10}$$

The hierarchical setup (10) associated with the ith row of the transition matrix $\boldsymbol{\Pi}_m$ is generalized to include $i = 1, \ldots, J$, that is, at the first level of the hierarchy, $\boldsymbol{x}_{mi}$'s are independent given $\boldsymbol{\pi}_{mi}$'s for $i \neq j$. At the second level, $\boldsymbol{\pi}_{mi}$'s are conditionally independent for $i \neq j$. The unknown quantities $\boldsymbol{W}_q$**,** 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. 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 hierarchical Bayesian representation given by (10) to the nonhomogeneous Markov chains. We note that the time nonhomogeneity of transition probabilities can be 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

$$\boldsymbol{\Theta}_t^i = \begin{bmatrix} \theta_{i11t} & \cdots & \theta_{iJ1t} \\ \vdots & \ddots & \vdots \\ \theta_{i1Qt} & \cdots & \theta_{iJQt} \end{bmatrix}. \tag{11}$$

We assume a Markov structure on the q-th row of $\boldsymbol{\Theta}_t^i$, that is, on $\boldsymbol{\theta}_{qt}^i = (\theta_{i1qt} \dots \theta_{iJqt})$. 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 $\boldsymbol{\theta}_{qt}^i$ follows a random walk model as

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

where $\boldsymbol{\omega}_{qt}^i$ is a $1 \times J$ vector of uncorrelated error terms for the parameter vector $\boldsymbol{\theta}_{qt}^i$. We assume that $\boldsymbol{\omega}_{qt}^i$'s are normally distributed with mean vector **0** and unknown covariance matrix $\boldsymbol{W}_q$ where $\boldsymbol{W}_q^{-1}|\boldsymbol{R}, k \sim Wish(\boldsymbol{R}, k)$ as in (9).

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

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

where $\boldsymbol{\eta}_{mit} = (\eta_{mi1t} \dots \eta_{miJt})$. Thus, the logit transform of time dependent transition probability π_{mijt} is defined as

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

where $\boldsymbol{\theta}_t^{ij}$ is the time dependent version of the $Q \times 1$ vector of regression parameters in (4), for $i = 1, \dots, J$, $j = 1, \dots, J-1$, and $t = 1, \dots, T$. 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 $\boldsymbol{\theta}_{qt}^i$'s, the rows of (11) are independent for $q = 1, \dots, Q$. As in section 2.2, $\boldsymbol{W}_q$'s are independent of each other for $q = 1, \dots Q$ and at time t, $\boldsymbol{\Theta}_t^i$'s are conditionally independent for $i = 1, \dots, J$. It follows from (12) that

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

and for $t = 0$ we assume that $(\boldsymbol{\theta}_{q0}^i | \boldsymbol{W}_q) \sim N(\mathbf{0}, \boldsymbol{W}_q)$.

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

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

$$\eta_{mijt} = logit(\pi_{mijt}) = \boldsymbol{F}_m\,\boldsymbol{\theta}^i_{jt},$$

$$\boldsymbol{\theta}^i_{qt}|\boldsymbol{\theta}^i_{q,t-1}, \boldsymbol{W}_q \sim N(\boldsymbol{\theta}^i_{q,t-1}, \boldsymbol{W}_q),$$

$$\boldsymbol{W}_q^{-1}|\boldsymbol{R}, k \sim Wish(\boldsymbol{R}, k) \text{ and } \boldsymbol{\theta}^i_{q0}|\boldsymbol{W}_q \sim N(\boldsymbol{0}, \boldsymbol{W}_q). \tag{15}$$

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 for $i \neq j$. As in the homogeneous case, (15) represents the hierarchical setup for individual m.

3. Posterior Analysis of Markov Chain Models

We consider the hierarchical Bayesian representations given by (10) and (15) for homogeneous and nonhomogeneous Markov chain models. We note that the hierarchical Bayesian setups 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 $\boldsymbol{\pi}_{mi}$'s and $\boldsymbol{\pi}_{mit}$'s. In what follows, we will present the Bayesian analyses of both homogeneous and nonhomogeneous Markov chain models.

3.1. Posterior analysis for homogeneous 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

$$p(\Pi_1,\ldots,\Pi_M,\boldsymbol{\Theta}^1,\ldots,\boldsymbol{\Theta}^J,\boldsymbol{W}_1,\ldots,\boldsymbol{W}_Q|\,\boldsymbol{S}_1,...,\boldsymbol{S}_M)$$

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

where the components of Π_m 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:

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

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 by $p(\phi|\boldsymbol{S},\phi^{(-)})$ where $\phi^{(-)}$ includes all random quantities except ϕ.

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

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

which can be rewritten as proportional to

$$\prod_{m=1}^{M}\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\frac{exp(\boldsymbol{F}_m\boldsymbol{\theta}_j^i)}{\sum_{j=1}^{J}exp(\boldsymbol{F}_m\,\boldsymbol{\theta}_j^i)}\right)^{x_{mijt}}exp\{-\frac{1}{2}\sum_{q=1}^{Q}(\boldsymbol{\theta}_q^i-\boldsymbol{\mu}_q^i)'\boldsymbol{W}_q^{-1}(\boldsymbol{\theta}_q^i-\boldsymbol{\mu}_q^i)\}\tag{19}$$

which is not a known density form. However, it can be shown that (19) is log-concave in $\boldsymbol{\Theta}^i$; see Appendix A for details and we can use the adaptive rejection sampling algorithm of Gilks and Wild [16].

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

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

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

3.2. Posterior analysis for 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

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

$$\propto \prod_{m=1}^{M}\prod_{i=1}^{J}\prod_{t=1}^{T} p(\boldsymbol{x}_{mit}|\boldsymbol{\Theta}_t^i) \prod_{q=1}^{Q} p(\boldsymbol{\theta}_{qt}^i|\boldsymbol{\theta}_{q,t-1}^i, \boldsymbol{W}_q) p(\boldsymbol{W}_q). \tag{21}$$

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

$$p(\boldsymbol{\Theta}_t^i|\boldsymbol{S}, \boldsymbol{\Theta}_t^{i(-)}) \propto \prod_{m=1}^{M} p(\boldsymbol{x}_{mit}|\boldsymbol{\Theta}_t^i) \prod_{q=1}^{Q} p(\boldsymbol{\theta}_{qt}^i|\boldsymbol{\theta}_{q,t-1}^i, \boldsymbol{W}_q) p(\boldsymbol{\theta}_{q,t+1}^i|\boldsymbol{\theta}_{qt}^i, \boldsymbol{W}_q), \tag{22}$$

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

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

$$+ (\boldsymbol{\theta}_{q,t+1}^i - \boldsymbol{\theta}_{qt}^i)' \boldsymbol{W}_q^{-1} (\boldsymbol{\theta}_{q,t+1}^i - \boldsymbol{\theta}_{qt}^i)\Big)\Big]. \tag{23}$$

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. Thus, it can be shown that (23) is a log concave density and we can use the adaptive rejection sampling algorithm to draw $\boldsymbol{\Theta}_t^i$'s .

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

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

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

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 the period some patients are discharged from the treatment program and understanding the reasons of 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 $\boldsymbol{F}_m=(1,\,1,z_m)$, and $\boldsymbol{\theta}^{ij}=(\gamma_j,\,\gamma_{ij},\,\beta_{ij})'$ for the homogeneous chains and $\boldsymbol{\theta}_{\boldsymbol{t}}^{ij}=(\gamma_{jt},\,\gamma_{ijt},\,\beta_{ij})'$ for the nonhomogeneous chains, where $z_m=Age_m$ is the age of the m-th patient at time 0. Thus, we can write

$$\begin{pmatrix}\eta_{mi1}\\ \vdots \\ \eta_{miJ}\end{pmatrix}=\begin{pmatrix}\gamma_1\\ \vdots \\ \gamma_J\end{pmatrix}+\begin{pmatrix}\gamma_{i1}\\ \vdots \\ \gamma_{iJ}\end{pmatrix}+Age_m\begin{pmatrix}\beta_{i1}\\ \vdots \\ \beta_{iJ}\end{pmatrix}, \tag{26}$$

for the homogeneous case and

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

for the nonhomogeneous 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. As the $(J+1)$-th state, the exit state, is used as the baseline category, $\gamma_{J+1}=\gamma_{i,J+1}=\beta_{i,J+1}=0$ 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 homogeneous Markov chain 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 nonhomogeneous Markov chain model, for time homogeneous parameters we used the same priors as given above for the homogeneous 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 homogeneous 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 homogeneous 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 covariate. The models were implemented using WinBugs 1.4 [17]. The posterior simulated samples of transition probabilities did not show any convergence problems. The modified Gelman-Rubin convergence statistics [18], calculated in WinBUGS, quickly approached to 1 after 1,500 monitored iterations in all cases, which indicates convergence of both the pooled and within interval widths to stability.

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 homogeneous and nonhomogeneous models. Table 1 shows that the DIC is in favor of the 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 homogeneous case, but it is considerably small in the 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 nonhomogeneity as observed by Nhan [1] and as indicated by the DIC criterion in Table 1. Thus, in the remainder of this section, the results from the 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

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

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

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

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

TABLE 2 ABOUT HERE

In Table 2, we present the posterior means and standard deviations of $(\gamma_{jt} + \gamma_{1jt})$ for transitions from state 1. Each posterior distribution represents the values of log odds with respect to the state 5. We note that as we move from right to the left in a given row of the table, that is, when we move to better states, the mean of the posterior distribution decreases. This implies that when we control the age effect, as we move to better states, the log probability difference between that state and the exit 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. At time 4, for example, 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 1 shows how π_{2jt}'s for $j = 1, \ldots, 5$ differ over a range of time periods, $t = 1, \ldots, 7$, for age group 10. From state 2, the transition probabilities to state 3 or 4 slightly increase with time, but transition probabilities to state 1 or 2 decrease with time, implying that the subjects are more likely to make progress in the treatment program. The likelihood of discharge rapidly increases with time, 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 1 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 making improvement over time, we can compare two age groups, say, 14 and 10. We can examine posterior probabilities of the quantity

$$D_{14-10} = \{\pi_{ijt}|Age = 14\} - \{\pi_{ijt}|Age = 10\},$$

for $i < j$, that is, we can infer differences in transition probabilities (for improvement) between the two age groups. Figure 2 shows the mean and 95% credible interval of posterior distribution of D_{14-10} obtained for π_{23t} and π_{24t}. We note that while the probability differences are positive and therefore implying more likely improvement for older children, the differences decrease with time. Furthermore, differences for π_{23t} seem to decrease more rapidly than for π_{24t}.

FIGURE 2 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 exit probabilities from each state are illustrated in Figure 3 for time periods $t = 1, \ldots, 7$. The distributions are presented for subjects in the age group of 10. We note in each frame of Figure 3 that the exit probability increases with time regardless of the prior state. The exit probabilities do not seem to differ much from one state to the other upto period $t = 3$. After period $t = 4$, increase in exit probability seems to be accelerated from states 1 and 4. This implies that as time passes patients will exit the program either because they get better or because they do not show much improvement. This can also be seen from Table 3 by comparing the exit probabilities from different states over time. While the overall exit behavior is similar for other age groups, older patients (age 16) show overall higher exit probabilities and increasing rates of exit over time than younger patients (age 10) as compared in Table 3.

FIGURE 3 ABOUT HERE

TABLE 3 ABOUT HERE

5. Conclusions

In this paper, we presented Bayesian methods for modeling and analyses of nonhomogeneous Markov chains, and 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.

APPENDIX A

Log concavity of $p(\boldsymbol{\Theta^i}|\boldsymbol{S}, \boldsymbol{\Theta^{i(-)}})$

The log of (19) can be written as proportional to

$$\sum_{m,t,j} x_{mijt}\Big(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}} - log\big[\sum_{j=1}^{J} exp(\boldsymbol{F}_{mt}\,\boldsymbol{\theta}_j^{\boldsymbol{i}})\big]\Big) - \frac{1}{2}\sum_{q=1}^{Q}(\boldsymbol{\theta}_q^{\boldsymbol{i}} - \boldsymbol{\mu}_q^{\boldsymbol{i}})'\boldsymbol{W}_q^{-1}(\boldsymbol{\theta}_q^{\boldsymbol{i}} - \boldsymbol{\mu}_q^{\boldsymbol{i}}),$$

where the expression $\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}}$ is linear in $\boldsymbol{\theta}_j^{\boldsymbol{i}}$, and thus the second derivative will be zero. The last term consists of a negative of a quadratic form which is concave. We next consider the second term in the bracket in the term. First derivative of the log term, $log\sum_{j=1}^{J} exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})$ is given by

$$\frac{exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})}{\sum_{j=1}^{J} exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})}$$

which is defined as π_{mij} in (5), and derivative of this quantity, is

$$\frac{exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})}{\sum_{j=1}^{J} exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})}\left(1 - \frac{exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})}{\sum_{j=1}^{J} exp(\boldsymbol{F}_{mt}\boldsymbol{\theta}_j^{\boldsymbol{i}})}\right).$$

which is always positive and thus implies the log concavity of (19) in $\boldsymbol{\Theta^i}$.

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, Müller 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. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; 7:434-455.

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.

Table 1 DIC comparison between two classes of models

	D-bar	D-hat	p_D	DIC
Homogeneous model	5093.91	5073.13	22.77	5116.68
Nonhomogeneous model	4871.50	4826.77	44.73	4916.22

Table 2. Posterior means and standard deviations (SD) 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	SD	Mean	SD	Mean	SD	Mean	SD
$t = 1$	7.53	0.91	6.17	0.84	4.22	0.84	2.24	1.14
$t = 2$	5.60	0.71	4.46	0.59	2.54	0.45	0.60	0.51
$t = 3$	4.61	0.62	3.51	0.62	1.28	0.51	− 1.06	0.59
$t = 4$	3.98	0.58	2.92	0.53	0.79	0.56	− 1.50	0.67
$t = 5$	3.51	0.51	2.46	0.48	0.30	0.54	− 2.09	0.78
$t = 6$	2.54	0.60	1.55	0.59	− 0.63	0.53	− 3.14	0.81
$t = 7$	2.18	0.54	1.19	0.59	− 1.00	0.55	− 3.54	0.82

Table 3. Comparison of posterior means of exit probability (π_{i5t}'s, $i = 1, ..., 4, t = 1, ..., 7$) between ages 10 and 16.

	π_{15t}		π_{25t}		π_{35t}		π_{45t}	
	Age10	Age16	Age10	Age16	Age10	Age16	Age10	Age16
$t = 1$	0.00	0.01	0.00	0.02	0.00	0.01	0.01	0.04
$t = 2$	0.01	0.03	0.02	0.08	0.03	0.10	0.03	0.10
$t = 3$	0.04	0.09	0.04	0.13	0.03	0.10	0.07	0.20
$t = 4$	0.06	0.14	0.06	0.19	0.06	0.16	0.20	0.47
$t = 5$	0.10	0.21	0.08	0.24	0.10	0.27	0.20	0.45
$t = 6$	0.22	0.40	0.11	0.31	0.21	0.46	0.32	0.60
$t = 7$	0.29	0.49	0.13	0.33	0.21	0.45	0.47	0.75

Figure 1. Posterior transition probabilities from state 2 at different time points for age 10 (Mean: solid line; 95% credible interval: dashed line)

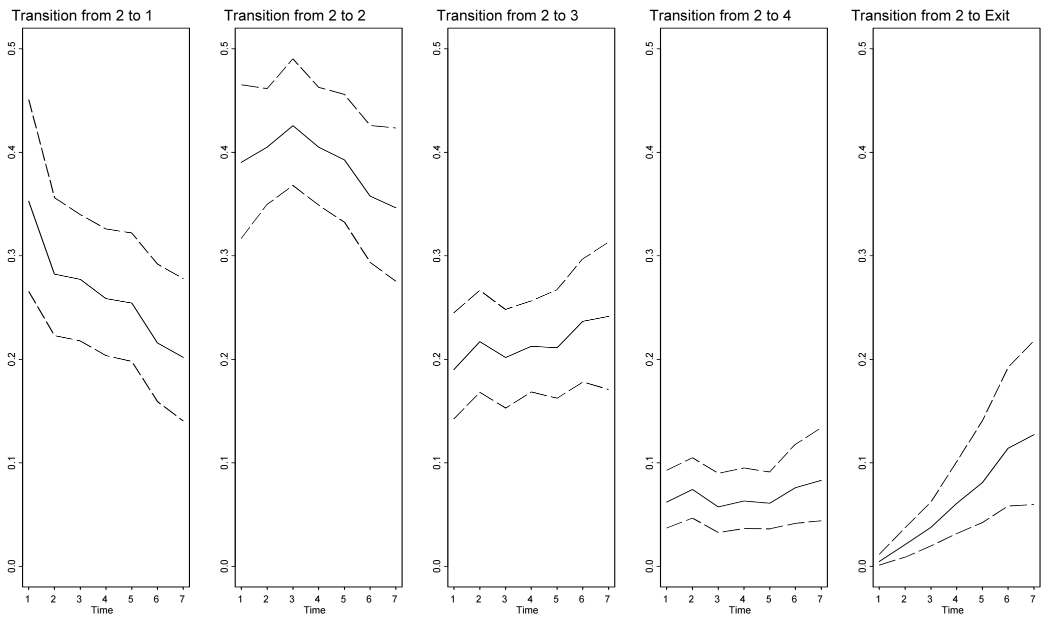

Figure 2. Posterior distributions of differences in π_{23t} and π_{24t} between age groups 14 and 10 (Mean: solid line; 95% credible interval: dashed line).

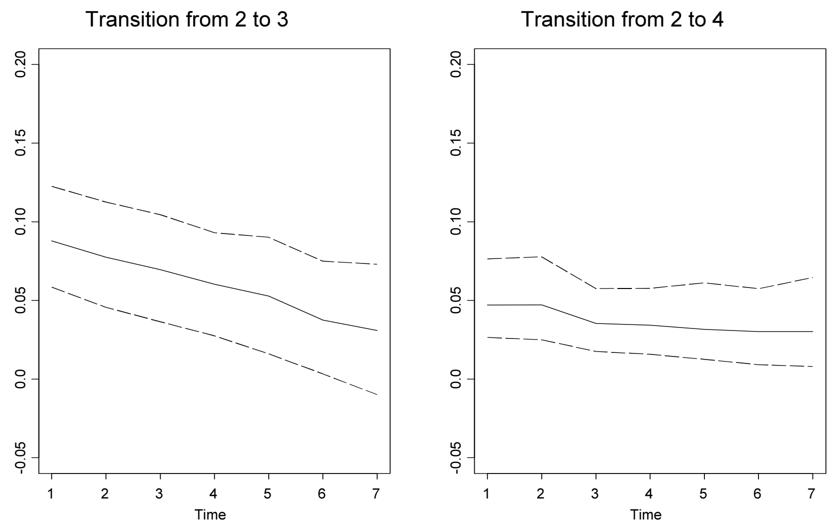

Figure 3. Posterior distributions of exit probability (π_{i5t}'s) from each state at $t = 1, ..., 7$: Age=10 (Mean: solid line; 95% credible interval: dashed line)