

Bayesian analysis of non-homogeneous Markov chains: Application to mental health data

Minje Sung^{1,*}, Refik Soyer² and Nguyen Nhan^{3,†}

¹*School of Business Administration, Ajou University, Suwon, Korea*

²*Department of Decision Sciences, The George Washington University, Washington, DC 20052, U.S.A.*

³*Graydon Manor Psychiatric Center, Leesburg, VA, U.S.A.*

SUMMARY

In this paper we present a formal treatment of non-homogeneous 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 non-homogeneity of transition patterns. In doing so, we introduce a logistic regression set-up 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. Copyright © 2006 John Wiley & Sons, Ltd.

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 of time. Modelling the states of the subjects over time, understanding the changing behaviour 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 modelling 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.

*Correspondence to: Minje Sung, School of Business Administration, Ajou University, Suwon, Korea.

†E-mail: sungmj@ajou.ac.kr

‡Research Director (now retired).

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 non-homogeneous transition patterns for patients. For example, in his study Nhan [1] observed strong evidence in favour of non-homogeneity in transition probabilities for patients.

In this paper we present a formal treatment of non-homogeneous 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 is focused 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 modelling transition matrices for our analysis of non-homogeneous 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 homogeneous chains who considered extensions to non-homogeneous 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 set-up for a binary Markov chain, and obtained the maximum likelihood estimates for the transition probabilities. Zeger and Qaqish [13] presented the Markov logistic regression set-up for correlated longitudinal data and discussed maximum likelihood estimation (MLE) for the model. This set-up 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 set-up 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 non-homogeneity is *via* inclusion of time-dependent deterministic covariates.

In this paper we present Bayesian methods for modelling and analyses of non-homogeneous Markov chains, and develop inference procedures to be able to address issues encountered in the analyses of data from psychiatric treatment programs. In doing so, we introduce a class of models for describing non-homogeneity in the transition probabilities. Our modelling strategy is based on the logistic regression set-up 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 set-up for non-homogeneous Markov chains using time-dependent covariates. We extend our set-up by introducing a first-order Markov structure for describing the time dependence of transition probabilities of the non-homogeneous 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 set-up are considered in Section 5 and conclusions are presented in Section 6.

2. MODELS FOR NON-HOMOGENEOUS MARKOV CHAINS

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

2.1. Notation and preliminaries

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

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

where $1(A)$ takes the value 1 if event A occurs and 0 otherwise. Then, the vector $\mathbf{x}_{mit} = (x_{mi1t}, \dots, x_{miJt})$ is a multinomial random variable with probability vector $\boldsymbol{\pi}_{mit} = (\pi_{mi1t}, \dots, \pi_{miJt})$ 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

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

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

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

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 $t = 1, \dots, T$, then the Markov chain is called a time homogeneous Markov chain whereas the case with time-dependent transition probabilities is referred to as a non-homogeneous Markov chain.

2.2. Static logistic regression set-up for non-homogeneous Markov chains

The logistic regression set-up 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 set-up 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 $\boldsymbol{\pi}_{mit} = (\pi_{mi1t} \dots \pi_{miJt})'$ for the non-homogeneous Markov chain by using time-dependent covariates while model parameters are static. In what follows we will present the static Bayesian logistic regression set-up for the J -dimensional Markov chain.

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

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

for $i = 1, \dots, J, j = 1, \dots, J - 1, t = 1, \dots, T$ where \mathbf{F}_{mt} is a $1 \times K$ covariate vector for the m th individual, and $\boldsymbol{\theta}^{ij} = (\theta_{ij1}, \dots, \theta_{ijK})'$ is a $K \times 1$ vector of regression parameters. We note that the multinomial logit model in (4) is for nominal categorical response and the cumulative logit model for ordinal categorical response is discussed in Section 5. We use the J th category as a baseline category in (4). Thus, the transition probability π_{mijt} is given by

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

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

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

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

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

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 $\boldsymbol{\theta}_k^i = (\theta_{i1k} \dots \theta_{i,J-1,k})$ and assume that each row of (7) is a multivariate normal vector defined as

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

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

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

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

In summary, the static logistic regression set-up for non-homogeneous Markov chains can be represented as a hierarchical Bayesian model as

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

The hierarchical set-up (10) is associated with the i th row of the transition matrix $\boldsymbol{\Pi}_{mt}$. The same set-up is applied to all rows of the transition matrix $\boldsymbol{\Pi}_{mt}$ for $i = 1, \dots, J$, that is, at the first level of the hierarchy, \mathbf{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 \mathbf{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 non-homogeneous Markov chains

The logistic regression set-up 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 non-homogeneous Markov chains. The time non-homogeneity of transition probabilities was incorporated into the model by using time-dependent covariates \mathbf{F}_{mt} in (4). However, in what follows, we consider a formal treatment of non-homogeneity 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

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

We assume a Markov structure on the k th row of $\boldsymbol{\Theta}_t^i$, that is, on $\boldsymbol{\theta}_{kt}^i = (\theta_{i1kt} \dots \theta_{i,J-1,kt})$. 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

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

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 \mathbf{W}_k where $\mathbf{W}_k^{-1} | \mathbf{R}, v \sim \text{Wish}(\mathbf{R}, v)$ as in (9). The choice of the non-stationary 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 Θ_{kt}^i .

Thus, the multivariate logit transformation for the non-homogeneous chain is given by

$$\eta_{mit} = \mathbf{F}_{mt} \Theta_t^i$$

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

$$\eta_{mijt} = \text{logit}(\pi_{mijt}) = \log \left(\frac{\pi_{mijt}}{\pi_{miJt}} \right) = \mathbf{F}_{mt} \theta_t^{ij} \tag{13}$$

where θ_t^{ij} is the time-dependent version of the $K \times 1$ 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 J th 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 Θ_{kt}^i 's, the rows of (11) are independent for $k = 1, \dots, K$. As in Section 2.2, \mathbf{W}_k 's are independent of each other for $k = 1, \dots, K$ and at time t , Θ_t^i 's are conditionally independent given $\theta_{k,t-1}^i$'s and \mathbf{W}_k 's for $i = 1, \dots, J$. It follows from (12) that

$$(\theta_{kt}^i | \theta_{k,t-1}^i, \mathbf{W}_k) \sim \text{MVN}(\theta_{k,t-1}^i, \mathbf{W}_k) \quad \text{if } t > 0 \tag{14}$$

where $\theta_{k,t-1}^i$ is a $1 \times (J - 1)$ mean vector and \mathbf{W}_k is $(J - 1) \times (J - 1)$ covariance matrix. For $t = 0$ we assume that $(\theta_{k0}^i | \mathbf{W}_k) \sim \text{MVN}(\mathbf{0}, \mathbf{W}_k)$.

Thus, the dynamic logistic regression set-up for non-homogeneous Markov chains can be represented as a hierarchical Bayesian model as

$$\begin{aligned} \mathbf{x}_{mit} | \boldsymbol{\pi}_{mit} &\sim \text{Multinomial}(\boldsymbol{\pi}_{mit}, 1) \\ \eta_{mijt} = \text{logit}(\pi_{mijt}) &= \mathbf{F}_{mt} \theta_{jt}^i \\ \theta_{kt}^i | \theta_{k,t-1}^i, \mathbf{W}_k &\sim \text{N}(\theta_{k,t-1}^i, \mathbf{W}_k) \\ \mathbf{W}_k^{-1} | \mathbf{R}, v &\sim \text{Wish}(\mathbf{R}, v) \quad \text{and} \quad \theta_{k0}^i | \mathbf{W}_k \sim \text{N}(\mathbf{0}, \mathbf{W}_k) \end{aligned} \tag{15}$$

The hierarchical Bayes set-up (15) is associated with the i th row of the transition matrix $\boldsymbol{\Pi}_{mt}$ in (3). It can be applied to all rows of the transition matrix for $i = 1, \dots, J$, that is, at the first level of the hierarchy, \mathbf{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 θ_{kt}^i 's for $i \neq j$. As in the static case, (15) represents the hierarchical set-up for individual m .

3. POSTERIOR ANALYSIS OF MARKOV CHAIN MODELS

We note that the hierarchical Bayesian set-ups (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 set-up to all states, $i = 1, \dots, J$, for M individuals, $m = 1, \dots, M$, is straightforward due to the conditional independence of \mathbf{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 non-homogeneous Markov chain models.

3.1. Posterior analysis for static non-homogeneous 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

$$p(\Pi_{11}, \dots, \Pi_{MT}, \Theta^1, \dots, \Theta^J, \mathbf{W}_1, \dots, \mathbf{W}_K | \mathbf{S}_1, \dots, \mathbf{S}_M) \\ \propto \prod_{m=1}^M \prod_{i=1}^J \left[\prod_{t=1}^T p(\mathbf{x}_{mit} | \Theta^i) \right] \prod_{k=1}^K p(\theta_k^i | \mu_k^i, \mathbf{W}_k) p(\mathbf{W}_k) \quad (16)$$

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

$$[\Theta^i | \mathbf{S}, \Theta^{i(-)}], \quad [\mathbf{W}_k | \mathbf{S}, \mathbf{W}_k^{(-)}] \quad (17)$$

where $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_M)$ and for notational convenience, we denote the full conditional posterior distribution of a random quantity ϕ_i by $[\phi_i | \mathbf{S}, \phi_i^{(-)}]$ where $\phi_i^{(-)}$ includes all random quantities except ϕ_i . For example, in (17), we have $\Theta^{i(-)} = (\Theta^1, \dots, \Theta^{i-1}, \Theta^{i+1}, \dots, \Theta^J, \mathbf{W}_1, \dots, \mathbf{W}_K)$.

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

$$[\Theta^i | \mathbf{S}, \Theta^{i(-)}] \propto \prod_{m=1}^M \left[\prod_{t=1}^T p(\mathbf{x}_{mit} | \Theta^i) \right] \prod_{k=1}^K p(\theta_k^i | \mu_k^i, \mathbf{W}_k) \quad (18)$$

which can be rewritten as proportional to

$$\prod_{m=1}^M \prod_{t=1}^T \prod_{j=1}^J \left(\frac{\exp(\mathbf{F}_{mt} \theta_j^i)}{\sum_{j=1}^J \exp(\mathbf{F}_{mt} \theta_j^i)} \right)^{x_{mijt}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^K (\theta_k^i - \mu_k^i)' \mathbf{W}_k^{-1} (\theta_k^i - \mu_k^i) \right\} \quad (19)$$

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

$$\propto |\mathbf{W}_k^{-1}|^{v/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\left(\mathbf{R} + \sum_{i=1}^J (\theta_k^i - \mu_k^i) (\theta_k^i - \mu_k^i)' \right) \mathbf{W}_k^{-1} \right] \right\} \quad (20)$$

which is a Wishart density with degrees of freedom, $v + J + 1$, and scale matrix $\frac{1}{2} (\mathbf{R} + \sum_{i=1}^J (\theta_k^i - \mu_k^i) (\theta_k^i - \mu_k^i)')$.

3.2. Posterior analysis for dynamic non-homogeneous chains

Given the transition data on M individuals for T time periods, for the non-homogeneous Markov chains set-up, we need to obtain the joint posterior distribution

$$[\Pi_{11}, \dots, \Pi_{MT}, \Theta_1^1, \dots, \Theta_T^J, \mathbf{W}_1, \dots, \mathbf{W}_K | \mathbf{S}_1, \dots, \mathbf{S}_M] \\ \propto \prod_{m=1}^M \prod_{i=1}^J \prod_{t=1}^T p(\mathbf{x}_{mit} | \Theta_t^i) \prod_{k=1}^K p(\theta_{kt}^i | \theta_{k,t-1}^i, \mathbf{W}_k) p(\mathbf{W}_k) \quad (21)$$

For simulating Θ_t^i , we can use the Markov property as implied by (12) and write

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

implying that $[\Theta_t^i | \mathbf{S}, \Theta_t^{i(-)}]$ is

$$\begin{aligned} &\propto \prod_{m=1}^M \left[\prod_{j=1}^J \pi_{mij}^{x_{mijt}} \right] \exp \left[-\frac{1}{2} \sum_{k=1}^K ((\theta_{kt}^i - \theta_{k,t-1}^i)' \mathbf{W}_k^{-1} (\theta_{kt}^i - \theta_{k,t-1}^i) \right. \\ &\quad \left. + (\theta_{k,t+1}^i - \theta_{kt}^i)' \mathbf{W}_k^{-1} (\theta_{k,t+1}^i - \theta_{kt}^i)) \right] \end{aligned} \tag{23}$$

Note that the conditional posterior distribution of Θ_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 $[\mathbf{W}_k | \mathbf{S}, \mathbf{W}_k^{(-)}]$, we note that the full conditional of \mathbf{W}_k^{-1} can be written as proportional to

$$|\mathbf{W}_k^{-1}|^{(v+T)/2} \exp \left[-\frac{1}{2} \text{tr} \left\{ \left(\mathbf{R} + \sum_{i=1}^J \sum_{t=1}^T (\theta_{kt}^i - \theta_{k,t-1}^i)(\theta_{kt}^i - \theta_{k,t-1}^i)' \right) \mathbf{W}_k^{-1} \right\} \right] \tag{24}$$

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

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 by 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 in 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. Data collection started 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 set-up 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

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

where $\pi_{m,J+1,j,t} = 0$ for $j \neq J + 1$. It is important to note that the last row of the transition matrix (25) is not a stochastic row and thus it does not enter the Bayesian inference. The first J rows of (25) represent the transition probabilities and the likelihood function associated with these row probabilities is a multinomial likelihood as before, but with dimension $J + 1$.

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

$$\begin{pmatrix} \eta_{mi1t} \\ \vdots \\ \eta_{miJt} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_J \end{pmatrix} + \begin{pmatrix} \gamma_{i1} \\ \vdots \\ \gamma_{iJ} \end{pmatrix} + \text{Age}_{mt} \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iJ} \end{pmatrix} \quad (26)$$

for the static case and

$$\begin{pmatrix} \eta_{mi1t} \\ \vdots \\ \eta_{miJt} \end{pmatrix} = \begin{pmatrix} \gamma_{1t} \\ \vdots \\ \gamma_{Jt} \end{pmatrix} + \begin{pmatrix} \gamma_{i1t} \\ \vdots \\ \gamma_{iJt} \end{pmatrix} + \text{Age}_{mt} \begin{pmatrix} \beta_{i1} \\ \vdots \\ \beta_{iJ} \end{pmatrix} \quad (27)$$

for the dynamic case.

In (26) vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_J)'$ represents factors common across the rows, whereas the vector $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{iJ})'$ 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 $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{iJ})'$ 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 diffuse 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 \mathbf{R} of the Wishart was assumed to be $\text{diag}(0.01, 0.01, 0.01, 0.01)$ implying a high degree of uncertainty.

In the dynamic non-homogeneous 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 $(\gamma_0|\mathbf{W}_1) \sim \text{MVN}(\mathbf{0}, \mathbf{W}_1)$ for $t=0$ and $(\gamma_t|\gamma_{t-1}, \mathbf{W}_1) \sim \text{MVN}(\gamma_{t-1}, \mathbf{W}_1)$ for $t>0$. In this case, \mathbf{W}_1^{-1} has the same Wishart prior as specified in the above for the static case. For the row-specific vector γ_{it} , we assume that $(\gamma_{i0}|\mathbf{W}_2) \sim \text{MVN}(\mathbf{0}, \mathbf{W}_2)$ for $t=0$ and $(\gamma_{it}|\gamma_{i,t-1}, \mathbf{W}_2) \sim \text{MVN}(\gamma_{i,t-1}, \mathbf{W}_2)$ for $t>0$, where \mathbf{W}_2^{-1} has the same Wishart prior as in the static case.

4.2. Analysis and results

Analysis was developed using WinBugs 1.4 [16]. In implementation of the Gibbs sampler in WinBUGS, if a particular full posterior conditional distribution is of a known form such as the Wishart distribution, then WinBUGS samples from that distribution. However, for cases where distribution is not a known form, Metropolis-within-Gibbs algorithm is used for updating. Thus, in our case the Metropolis-within-Gibbs algorithm was used for simulating Θ_t^i in (22) and \mathbf{W}_k was simulated from Wishart distribution as shown in (24).

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 2000 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 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 γ_{jt} , $j = 1, 2$, $t = 1$ and β_{ij} , $i = 1$, $j = 1, \dots, 4$ of non-homogeneous model in (27). The initial values of three independent runs were generated in WinBUGS using different seed values. The Gelman–Rubin convergence statistics [17], approached to 1 after 1500 monitored iterations in all cases, which indicates convergence of both the pooled and within interval widths to stability. Other parameters also exhibit similar patterns.

We use the deviance information criterion (DIC), a generalization of AIC, developed by Spiegelhalter *et al.* [18] as a measure of goodness-of-fit when we compare the static and dynamic non-homogeneous models. Table I shows that the DIC is in favour of the dynamic non-homogeneous model as implied by the lower DIC value. In the table, \bar{D} is the posterior mean of the deviance, \hat{D} 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 $\text{DIC} = \bar{D} + p_D$. 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 non-homogeneous model indicating that not all time-dependent parameters effectively contribute to explaining the transition behaviour of the subjects.

Analysis of the data shows strong evidence in favour of dynamic non-homogeneity as indicated by the DIC criterion in Table I. Thus, in the remainder of this section, the results from the dynamic non-homogeneous Markov chain model will be presented.

In modelling 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

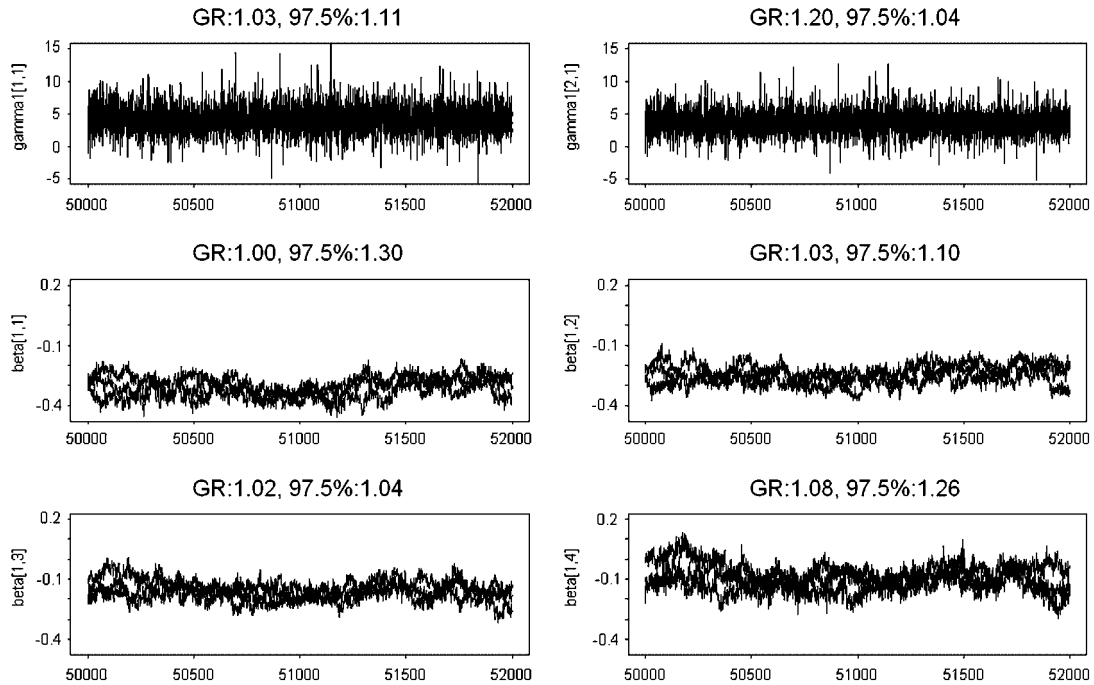


Figure 1. Convergence monitoring plots of 3 independent chains for γ_{jt} , $j = 1, 2$, $t = 1$, β_{ij} , $i = 1$, $j = 1, \dots, 4$ in dynamic model: GR=Gelman Rubin's [17] scale reduction factor, 97.5%=97.5 percentile of GR.

Table I. 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

transition to the j th state from a given row i at time t as

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

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.

In Table II, we present the posterior means and 95 per cent 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

Table II. Posterior means and 95% credible intervals of fixed effects for transitions from state 1 to other states.

<i>t</i>	$\gamma_{1t} + \gamma_{11t}$		$\gamma_{2t} + \gamma_{12t}$		$\gamma_{3t} + \gamma_{13t}$		$\gamma_{4t} + \gamma_{14t}$	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
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
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
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
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
5	4.73	3.11, 6.21	3.38	1.80, 4.84	0.95	-0.59, 2.49	-0.11	-3.24, 0.63
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
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

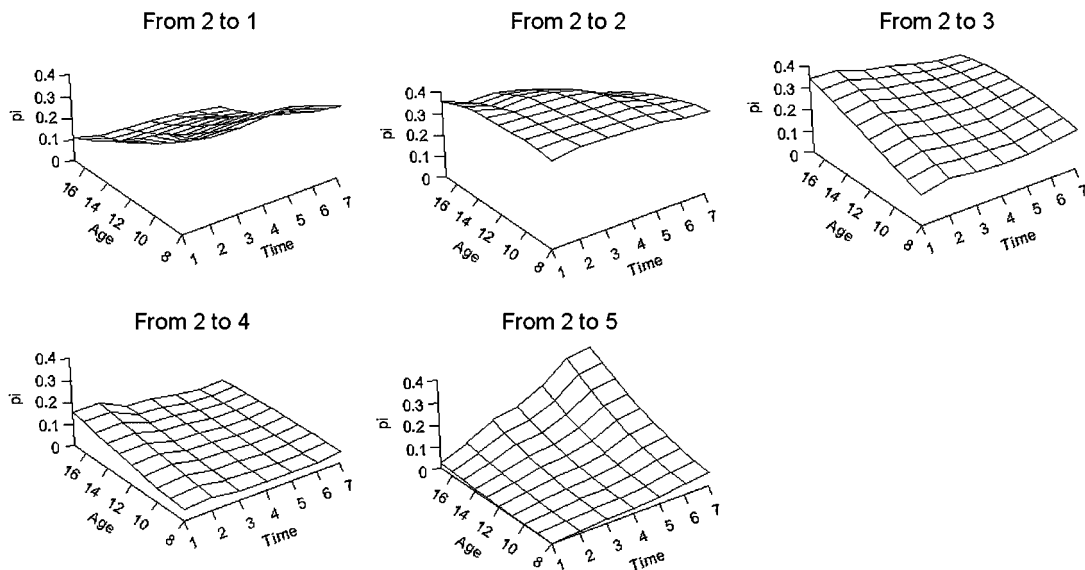


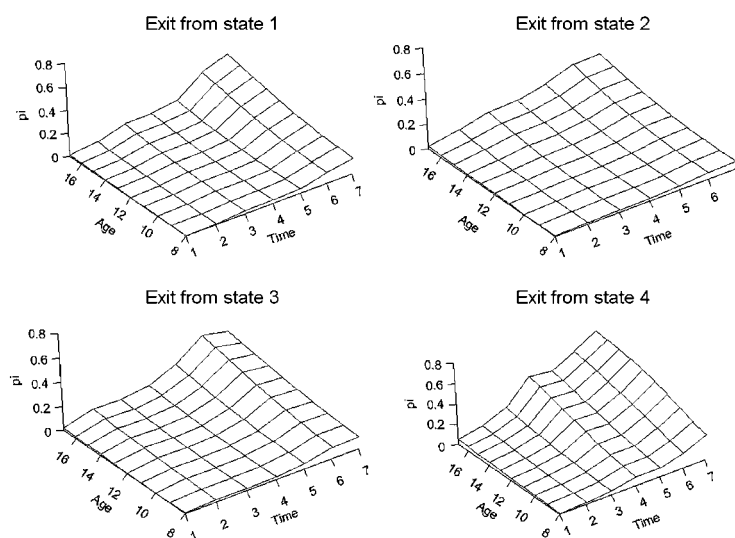
Figure 2. Posterior means of transition probabilities from state 2 at different time points and ages.

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 discharge 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 favour 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 $j = 1, \dots, 5$ differ by time ($t = 1, \dots, 7$), 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

Table III. Comparison of posterior means and 95% credible intervals for age effect β_{it} .

i	β_{i1}		β_{i2}		β_{i3}		β_{i4}	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
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
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
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
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 3. Posterior distributions of exit probabilities (π_{i5t} 's) from each state by age and time.

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.

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 III. The table shows posterior means and 95 per cent 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.

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 discharges. 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 $t = 1, \dots, 7$ 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.

In our experience we have found that the posterior results presented in this section were robust. We examined the posterior sensitivity to the prior specification by varying specified values of hyperparameters of the prior distributions and observed that the use of different specifications 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 non-homogeneous Markov chains. In this section, we incorporate the ordinal nature of the data and consider a cumulative logit model for the non-homogeneous Markov chains. The cumulative logit model can be specified in terms of cumulative transition probabilities

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

as

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

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

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

where δ_{ij} is cut point parameter and $\boldsymbol{\theta}_t^i = (\theta_{i1t}, \dots, \theta_{iKt})'$ 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 $G(\delta_{ij} - \mathbf{F}_{mt} \boldsymbol{\theta}_t^i)$ and $s_{mt} = j$ if $\delta_{i,j-1} < Y_{mit} < \delta_{ij}$ [19]. As Johnson and Albert [20] 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 $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{iJ})$ so that $\delta_{i0} \leq \delta_{i1} \leq \dots \leq \delta_{iJ}$ for $i = 1, \dots, J$. 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

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

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

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 $j = 2, \dots, J + 1$, and $\pi_{mit} = Q_{mit}$. The full conditional posterior distribution of the cutpoint δ_{ij} can be obtained as

$$\begin{aligned} p(\delta_{ij} | \mathbf{S}, \delta_{ij}^{(-)}) &\propto \prod_{m=1}^M \prod_{j=1}^J \left[\prod_{t=1}^T p(\mathbf{x}_{mit} | \delta_{ij}, \boldsymbol{\theta}_t^i) \right] p(\delta_{ij} | \tau_\delta) \\ &\propto \prod_{m=1}^M \prod_{j=1}^J \left[\prod_{t=1}^T (Q(\delta_{ij} - \mathbf{F}_{mt} \boldsymbol{\theta}_t^i) - Q(\delta_{i,j-1} - \mathbf{F}_{mt} \boldsymbol{\theta}_t^i))^{x_{mijt}} \exp\left(-\frac{\tau_\delta}{2} (\delta_{ij} - \mu_\delta)^2\right) \right] \end{aligned}$$

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

$$\begin{aligned} p(\boldsymbol{\theta}_t^i | \mathbf{S}, \boldsymbol{\theta}_t^{i(-)}) &\propto \prod_{m=1}^M p(\mathbf{x}_{mit} | \delta_{ij}, \boldsymbol{\theta}_t^i) \prod_{k=1}^K p(\theta_{ikt} | \theta_{ik,t-1}, \tau_{\theta_k}) p(\theta_{ik,t+1} | \theta_{ikt}, \tau_{\theta_k}) \\ &\propto \prod_{m=1}^M \left[\prod_{j=1}^J (Q(\delta_{ij} - \mathbf{F}_{mt} \boldsymbol{\theta}_t^i) - Q(\delta_{i,j-1} - \mathbf{F}_{mt} \boldsymbol{\theta}_t^i))^{x_{mijt}} \right] \\ &\quad \times \exp\left[-\frac{1}{2} \sum_{k=1}^K ((\theta_{ikt} - \theta_{ik,t-1})\tau_{\theta_k} + (\theta_{ik,t+1} - \theta_{ikt})\tau_{\theta_k})\right] \end{aligned}$$

The full conditional of τ_δ can be written as proportional to

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

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 $\delta_{i0} \leq \delta_{i1} \leq \dots \leq \delta_{i,J-1}$ for $i = 1, \dots, J$ is described by independently and normally distributed increments $N(0, \tau_\delta)$. 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 2000 simulated posterior samples after 50 000 burn-in iterations. While the transition patterns in Figure 4 are generally similar to those in Figure 2, the monotonicity observed in Figure 2 is not fully preserved in Figure 4 especially for the older age group. This difference is possibly due to the different modelling of age effect parameters in (27) and (28). In (27), the age effect parameter β is static, whereas the age effect parameter is dynamic in (28).

An alternative modelling strategy for non-homogeneous 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 set-up similar to Albert and Chib's [21] treatment of probit models. More specifically, we define a continuous latent variable Z_{mit} such that

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

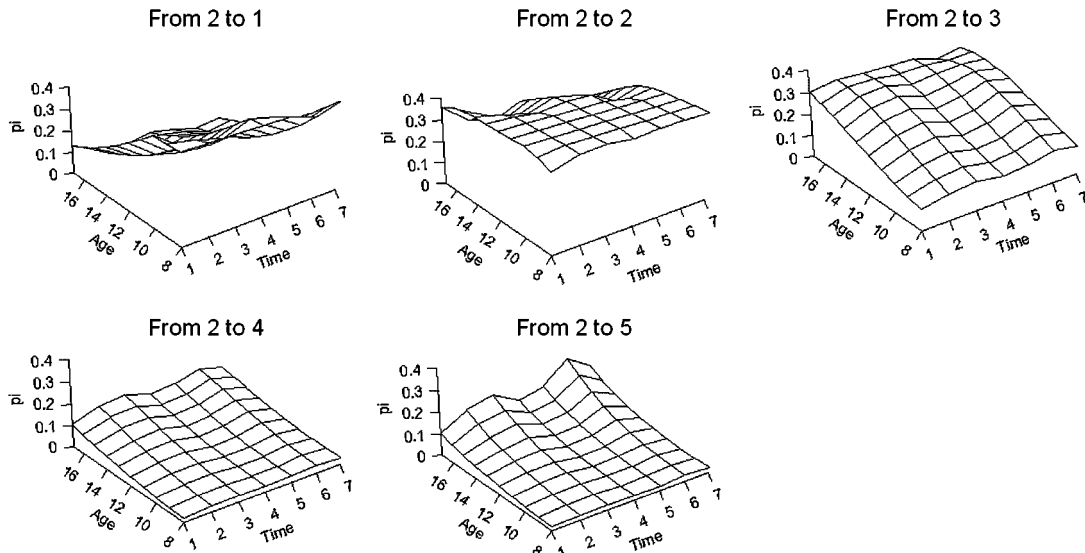


Figure 4. Posterior means of transition probabilities obtained from the cumulative logit model: transitions from state 2 at different time points and ages.

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

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

where \mathbf{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 \mathbf{W}_θ has an inverse Wishart prior with scale matrix \mathbf{R} and degrees of freedom $r > K$.

The above set-up can be represented as a DLM in the sense of West and Harrison [22] as

$$\begin{aligned} \mathbf{Z}_{it} &= \mathbf{F}_t\boldsymbol{\theta}_t^i + \mathbf{u}_{it} \\ \boldsymbol{\theta}_t^i &= \boldsymbol{\theta}_{t-1}^i + \mathbf{w}_{it} \end{aligned}$$

where \mathbf{Z}_{it} is a $M \times 1$ vector of observations, \mathbf{F}_t is an $M \times K$ matrix of covariates, \mathbf{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 \mathbf{I}_M , and \mathbf{w}_{it} 's are normal with mean $\mathbf{0}$ and variance-covariance matrix \mathbf{W}_θ . Given the DLM structure it is possible to develop a Gibbs sampler using ideas of Albert and Chib [21] and posterior simulation methods presented in Chapter 15 of West and Harrison [22] 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 and can directly draw from the joint posterior distribution of $\boldsymbol{\theta}_t^i$'s using the forward filtering backward sampling algorithm in West and Harrison [22]. In this case the full conditional of \mathbf{W}_θ 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 cannot be obtained in any known form

when we use independently and normally distributed increments as in the logistic set-up case. However, one can use the type of priors considered by Albert and Chib [21] and implement the Gibbs sampler.

The attractive feature of the probit set-up 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 approach requires further study and consideration of other methods for state space models such as the simulation smoother by de Jong and Shephard [23] and these will be considered elsewhere.

6. CONCLUSIONS

In this paper, we presented Bayesian methods for modelling and analyses of non-homogeneous Markov chains. In doing so, 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. We note that 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.

Our models exhibit considerable complexity due to the hierarchical multivariate structures of the parameters. In implementation of the Gibbs sampler in such complex structures, use of reparametrization *via* centring or hierarchical centring is suggested to accelerate the convergence [24, 25]. However, these reparametrizations have been used for univariate distributions and our experience in using these methods with multivariate distributions has shown no significant effect. Improvement of convergence in the case of multivariate distributions is not a trivial task due to the high dimensionality and correlation structure of the involved components. Roberts and Sahu [26] compared blocking and parametrization methods for multivariate target distributions and pointed out that block updating would require more computational effort thus can make the convergence slower. As discussed by Brooks and Gelman [27], it is possible to develop multivariate versions of Gelman–Rubin-type convergence statistics, but these are ‘... computationally expensive in even moderate dimensions ...’.

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. *Journal of the Royal Statistical Society, Series B* 1976; **38**:205–247 (with discussion).
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: 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: Oxford, 1994.
15. Grunwald G, Raftery A, Guttorp P. Time series for continuous proportions. *Journal of the Royal Statistical Society, Series B* 1993; **55**:103–116.
16. Spiegelhalter D, Thomas A, Best N, Gilks W. *Bayesian Inference Using Gibbs Sampling Manual (Version ii)*. MRC Biostatistics Unit, Cambridge University, 1996.
17. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**:457–511.
18. Spiegelhalter D, Best N, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 2002; **64**:583–616.
19. Agresti A. *Categorical Data Analysis*. Wiley: New York, 1990.
20. Johnson VE, Albert JH. *Ordinal Data Modelling*. Springer: New York, 1999.
21. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 1993; **88**:669–679.
22. West M, Harrison J. *Bayesian Forecasting and Dynamic Models*. Springer: New York, 1997.
23. de Jong P, Shephard N. The simulation smoother for time series models. *Biometrika* 1995; **62**:339–350.
24. Qiu Z, Song P, Tan M. Bayesian hierarchical models for multi-level repeated ordinal data using WinBUGS. *Journal of Biopharmaceutical Statistics* 2002; **12**:121–135.
25. Vines SK, Gilks WR, Wild P. Fitting Bayesian multiple random effects models. *Statistics and Computing* 1996; **6**:337–346.
26. Roberts GO, Sahu SK. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society, Series B* 1997; **59**:291–317.
27. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**:434–455.