# Bayesian Analysis of Transition Model for Longitudinal Ordinal Response Data: Application to Insomnia Data

S. Noorian[1] and M. Ganjali[2,*]

[1]Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran

[2]Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

**Abstract:** In this paper, we present a Bayesian framework for analyzing longitudinal ordinal response data. In analyzing longitudinal data, the possibility of correlations between responses given by the same individual needs to be taken into account. Various models can be used to handle such correlations such as marginal modeling, random effect modeling and transition (Markov) modeling. Here a transition modeling is used and a Bayesian approach is presented for analyzing longitudinal data. A cumulative logistic regression model and the Bayesian method, using MCMC, are implemented for obtaining the parameters estimates. Our approach is applied on a two-period longitudinal Insomnia data where the Bayesian estimate for measure of association, $\gamma$, between the initial and follow-up ordinal responses is obtained in each level of a treatment variable. Then, the sensitivity of posterior summaries to changes of prior hyperparameters is investigated. We also use Bayes factor criterion for testing some important hypotheses.

**Keywords:** Bayesian Analysis, Bayes Factor, Conditional Predictive Ordinate, Logistic Regression, Markov Model.

## 1. INTRODUCTION

In health-related and social science applications, we have to learn about longitudinal or panel studies where repeated ordinal response data commonly occur. For example, a physician might evaluate patients at baseline and at weekly follow-ups regarding whether a new drug treatment is successful. After collecting the data, the first and most important step in learning from the data about the process generating them is exploratory data analysis. This step may lead us to decide which statistical model is the most appropriate to use in order to answer the scientific questions of interest.

Many diseases are recorded with two types of scales: nominal and ordinal. When responses are observed in a longitudinal direction, there is a sequence of responses recorded on each individual. Values of these responses are called states. In practice, these states and the move between them are important. For example, transitions among the disease states may correspond to improvement, stabilization or deterioration of the disease.

In the current context, we have to take into account not only the fact that responses are ordinal in nature but also the possibility of dependency or correlation between responses given by the same individual. Agresti [1] and Lall *et al*. [2] conducted a comprehensive survey of models for ordered categorical data, in which the need for model interpretation is emphasized. Different models can be used to handle such dependency. One approach is the use of marginal modelling, which allow for inferences about parameters averaged over the whole population or trend over time (Ten Have *et al*. [3]; Kim [4]; Liang *et al*. [5]; Molenberghs and Lesaffre [6]). Another approach is making use of random effects modelling, which deliberately provide inferences about variability between respondents. In this approach, individual behavior is often of scientific interest (Harville and Mee [7]; Verbeke and Lesaffre [8]; Tutz and Hennevogl [9]; Verbeke and Molenberghs [10]; Diggle *et al*. [11] and Tutz [12]). However, both of these approaches are generally appropriate for longer sequences of measurements than those examined here. Another appropriate approach to investigate the reasons for the change of the responses is the use of Markov (transition) models (Garber [13]; Francom *et al*. [14]; Chung *et al*. [15]; Rezaee and Ganjali [16] and Rezaee *et al*. [17]) where we can consider the effect of previous response on current response. For reviews of transition and other models for longitudinal ordinal data, see McCullagh [18], Agresti [19], Diggle *et al*. [11] and Sung *et al*. [20].

Zeghnoun *et al*. [21] assumed proportional odds, and adopted a first-order Markov model in modeling the effect of ozone on the appearance of respiratory symptoms in school children. Chan and Wan [22] proposed a bivariate binary model with a separate model for informative dropout (ID). Their model incorporates mixture and random effects. Mansourian *et al*. [23] considered several flexible random effects

*Address corresponding to this author at the Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran;
Tel: +98-21-29902915; Fax: 00982122431649; E-mail: m-ganjali@sbu.ac.ir

E-ISSN: 1929-6029/12 

models and investigated their properties in the model fitting. They also adopted a proportional odds logistic regression model and incorporated the skewed version of the normal, Student's t and slash distributions for the random effects. Mandel *et al*. [24] utilized maximum partial likelihood estimation under a second-order Markov structure in modeling transitions on the expanded disability status scale.

Anderson and Goodman [25] show that the maximum likelihood estimator of the transition probabilities $\pi_{mabt}$, (the probability that $m^{\text{th}}$ individual moves from state $a$ at time $t-1$ to state $b$ at time $t$), when the Markov chain is homogeneous (see next section), is the fraction of the number of observed transitions from state $a$ to state $b$ over the total number of observations beginning in state $a$. A Bayesian analysis of the homogeneous Markov chains is presented by Lee *et al*. [26] using a Dirichlet prior distribution on transition probabilities $\pi_{mabt}$. Meshkani [27] presents an empirical Bayes approach for homogeneous chains and considers extensions to non-homogeneous Markov chains by viewing the problem as a parametric empirical Bayes problem. These earlier approaches have not considered the effects of covariates on transition probabilities. A Bayesian approach for analysis of longitudinal categorical data under the multinomial logistic model is proposed by Sung *et al*. [20]. Healy and Engler [28] propose the use of Bayesian variable selection in Markov models for obtaining estimation of subject-specific transition probabilities. Their approach simultaneously estimates the order of the Markov process and the transition-specific covariate effects.

In this paper, the use of a first order transition model and its Bayesian analysis for repeated ordinal responses will be presented. In a two-period longitudinal data the association between initial response and follow-up response will be measured by gamma ($\gamma$) ([29]). The Bayesian estimation of $\gamma$ will be presented and also the estimates of this measure in each level of a treatment variable will be obtained. Sung *et al*. [20] do not consider the use of gamma and its Bayesian analysis and also do not investigate the sensitivity of parameters estimates to change of the prior. We will discuss the use of gamma and we will present sensitivity to the choice of prior. We will also conduct the sensitivity of the results to eliminate some individuals. As test of hypothesis is very important to be done for answering to some scientific questions, we present the use of Bayes factor for testing some important hypotheses. We test some hypotheses about the regression coefficient parameters to know about the possible changing effects of some covariates over the level of the previous response.

For motivation, the Insomnia data are introduced in Section 2 and some Bayesian exploratory data analysis are performed for this data set. The cumulative logistic regression model is presented in Section 3. By setting up some prior distributions, the posterior estimates of parameters is obtained in Section 4 and then by using Bayes factor we do some Bayesian hypothesis testings about regression coefficients. The analyses based on frequentist and Bayesian views are presented in Section 5 and conclusions are presented in Section 6.

## 2. THE INSOMNIA DATA SET AND ITS EXPLORATORY ANALYSIS

Insomnia is a sleep disorder in which the patient does not get enough, or satisfactory sleep. Insomnia can vary as to how long it lasts and as to how often it occurs. Insomnia can be short-term (called acute insomnia) or last a long time (called chronic insomnia). Acute insomnia can last from one night to a few weeks. Chronic insomnia is present when a person has insomnia at least 3 nights a week for 1 month or longer. It can be caused by many things and often occurs along with other health problems. Common causes of chronic insomnia are depression, chronic stress, and pain or discomfort at night. Acute insomnia may not require treatment. Treatments for chronic insomnia include first treating any underlying conditions or health problems that may be causing the insomnia. If insomnia still continues, the health care provider may suggest either behavioral therapy or medication.

The data in Table **1** are extracted from results of a randomized, double-blind clinical trial comparing an active hypnotic drug with a placebo in 239 patients who have sleeplessness problems (Francom *et al*. [14]). The measure of interest is the patientís response to the question "How quickly did you fall asleep after going to bed?" The response was categorized as an ordinal scale: 'less than 20 minutes'; '20 to 30 minutes'; 'more than 30 and less than or equal to 60 minutes'; and 'greater than 60 minutes'. Patients were asked this question after a one week placebo washout period (baseline measurement) and following a two-week treatment period.

If there is no correlation between responses, one may fit separate marginal models to each response to examine the treatment effect. For the Insomnia data, we shall use the gamma association measure to calculate the association between two ordinal responses, (Goodman and Kruskal [29]). This measure is the difference between the concordant and the discordant pairs divided by the sum of the concordant and the discordant pairs and it takes values in the range $[-1,1]$.

Using the same notation as in Agresti [19], we define $\gamma$ as a measure of ordinal association and

**Table 1: Time to Falling Asleep Obtained from the Question, " How Quickly Did you Fall Asleep? " in Grouped Minutes (Follow-Up Response, $Y_2$, by Treatment and Initial Response, $Y_1$, Observed Counts and Row Percentages)**

| | | Follow-up ($Y_2$) | | | | |
|---|---|---|---|---|---|---|
| **Treatment** | **Initial ($Y_1$)** | **< 20** | **20 − 30** | **30 − 60** | **> 60** | **Total** |
| Active | < 20 | 7 | 4 | 1 | 0 | 12 |
| | | 58.3% | 33.3% | 8.3% | 0.0% | 100.0% |
| | 20 − 30 | 11 | 5 | 2 | 2 | 20 |
| | | 55.0% | 25.0% | 10.0% | 10.0% | 100.0% |
| | 30 − 60 | 13 | 23 | 3 | 1 | 40 |
| | | 32.5% | 57.5% | 7.5% | 2.5% | 100.0% |
| | > 60 | 9 | 17 | 13 | 8 | 47 |
| | | 19.1% | 36.2% | 27.7% | 17.0% | 100.0% |
| Placebo | < 20 | 7 | 4 | 2 | 1 | 14 |
| | | 50.0% | 28.6% | 14.3% | 7.1% | 100.0% |
| | 20 − 30 | 14 | 5 | 1 | 0 | 20 |
| | | 70.0% | 25.0% | 5.0% | 0.0% | 100.0% |
| | 30 − 60 | 6 | 9 | 18 | 2 | 35 |
| | | 17.1% | 25.7% | 51.4% | 5.7% | 100.0% |
| | > 60 | 4 | 11 | 14 | 22 | 51 |
| | | 7.8% | 21.6% | 27.5% | 43.1% | 100.0% |

illustrate it for $I \times J$ tables. Let $\{\pi_{ij}\}$ be the joint probability distribution in the table, in order to define $\gamma$ we should describe the probabilities of concordance and discordance in the table. For two independent observations from the table, the probabilities of concordance and discordance are denoted by $\Pi_C$ and $\Pi_D$, respectively, and are defined as:

$$\Pi_C = 2\sum_i\sum_j \pi_{ij}\left(\sum_{h>i}\sum_{t>j}\pi_{ht}\right),$$

and

$$\Pi_D = 2\sum_i\sum_j \pi_{ij}\left(\sum_{h>i}\sum_{t<j}\pi_{ht}\right).$$

The probabilities of concordance and discordance are given by $\frac{\Pi_C}{\Pi_C+\Pi_D}$ and $\frac{\Pi_D}{\Pi_C+\Pi_D}$, respectively. So, $\gamma$, as a measure of association, is the difference between these probabilities, i.e.,

$$\gamma = \frac{\Pi_C - \Pi_D}{\Pi_C + \Pi_D}.$$

If in addition to ordinal variables, there exists a control variable we can define $\gamma$ in each category of the controlling variable, $Z$. For this, we use $\{\pi_{ijk}\}$ for denoting the joint probability distribution in an $I \times J \times K$ table, where $K$ denotes the number of categories of the control variable. Within a fixed category $k$ of $Z$, we use $\{\pi_{ijk}\}$ instead of $\{\pi_{ij}\}$ in the above $\Pi_C$ and $\Pi_D$ formulas and we call them as $\Pi_{C(k)}$ and $\Pi_{D(k)}$, respectively, and then compute $\gamma$ (say $\gamma(k)$).

The frequentist estimate of gamma for the two responses is $0.546$ (S.E. $=0.063$, p-value=0.000) which shows a strong association between the two responses and consequently any statistical analysis of these data should take this association into account. Partial gamma (gamma for a specific treatment) is used to recognize any correlation between the two responses in each level of the control variable. The frequentist estimate is 0.461 (S.E. =0.105, p-value= 0.000) for the active drug and 0.635 (S.E. =0.075, p-value=0.000) for the placebo. As the association between the two responses is not the same for the two treatments, we need to choose a longitudinal approach which is able to take into account the fact that the covariance structure of the responses is dependent on the treatment.

For obtaining the Bayesian estimates of gamma and partial gamma, we can use the Dirichlet distribution. For this, consider a table having two ordinal responses,

**Table 2: Posterior Mean and Standard Error of Partial Gamma; γ(0) for Placebo Group and γ(1) for Active Drug Group**

| α | $10^{-3}$ | $10^{-2}$ | $10^{-1}$ | 1 | 10 |
|---|---|---|---|---|---|
| γ(0) | 0.632 | 0.629 | 0.623 | 0.542 | 0.251 |
| SE(γ(0)) | 0.077 | 0.073 | 0.075 | 0.071 | 0.670 |
| γ(1) | 0.456 | 0.451 | 0.445 | 0.348 | 0.081 |
| SE(γ(1)) | 0.076 | 0.075 | 0.075 | 0.071 | 0.068 |

$X$ and $Y$ with $I$ and $J$ categories, respectively, and a controlling variable $Z$ with $K$ categories. For each $k$, the number of subjects in all $I \times J$ cells are known. If $n_{ijk}$, $i = 1,\ldots,I$; $j = 1,\ldots,J$, be the number of events in $i^{\text{th}}$ row, $j^{\text{th}}$ column and $k^{\text{th}}$ level of the control variable, then we can assume that $n_k = (n_{11k},\ldots,n_{IJk})$ denotes multinomial random variables with parameters $N_k = \sum_i \sum_j n_{ijk}$, that is known, and probability vector $\Pi_{(k)} = (\pi_{11k},\ldots,\pi_{IJk})$ where the elements of $\Pi_{(k)}$ must satisfy $\sum_i \sum_j \pi_{ijk} = 1$.

Since the computation of posterior distribution of $\gamma(k)$ is intractable due to its complex form, we first obtain the posterior distribution of $\Pi_{(k)}$ and then by using the relationship between $\Pi_{(k)}$ and $\gamma(k)$ the posterior distribution and the posterior summaries of $\gamma(k)$ will be obtained. For prior distributions of $\Pi_{(k)}$ we consider Dirichlet distribution with parameters $\alpha_k = (\alpha_{11k},\ldots,\alpha_{IJk})$, (Lindley [30], Good [31]).

The conjugate density for multinomial distribution is the Dirichlet distribution and hence the posterior distribution of $\Pi_{(k)}$ is also Dirichlet distribution with parameters $\alpha_k^* = (\alpha_{11k}^*,\ldots,\alpha_{IJk}^*)$ where $\alpha_{ijk}^* = \alpha_{ijk} + n_{ijk}$.

In order to determine the posterior distribution of $\gamma(k)$, first we simulate $\Pi_{(k)}$ from Dirichlet distribution with known parameters $\alpha_k = (\alpha_{11k},\ldots,\alpha_{IJk})$. Then, we simulate $\Pi_{(k)}^*$ from Dirichlet distribution with vector of parameters $\alpha_k^* = (\alpha_{11k}^*,\ldots,\alpha_{IJk}^*)$. Finally, the posterior values of $\gamma(k)$ are computed by use of posterior values of $\Pi_{(k)}^*$.

For computing Bayesian estimates of $\gamma(k)$ in Insomnia data in Table **1**, we have chosen Dirichlet distribution as a prior distribution with several different vectors of parameters $\alpha_0 = (\alpha_{1,0},\ldots,\alpha_{16,0})$ for placebo group, and $\alpha_1 = (\alpha_{1,1},\ldots,\alpha_{16,1})$ for active drug group, in order to study sensitivity analysis. Also we consider $\alpha_{1,i} = \ldots = \alpha_{16,i} = \alpha$, for $i = 0,1$. For the values of $\alpha$ that are smaller than 1 (such as $\alpha = 10^{-3}, 10^{-2}$ and $10^{-1}$), the prior distributions are low informative, but the priors corresponding to values of $\alpha$ greater than 1 are highly informative. So, the posterior means of corresponding priors that are low informative, are similar to the frequentist estimate of $\gamma(k)$. But, for more informative priors ($\alpha = 10$), the Bayesian estimates have significant difference from those of frequentist ones. The results are given in Table **2**.

Table **3** displays marginal distributions for the initial and follow-up responses for the two treatments.

From Table **3**, we can conclude that, initially, the two groups have similar distributions, but at the follow-up, those patients on the active treatment tend to fall asleep more quickly.

For example, by Table **1**, the sample probability of a patient who initially took more than 60 minutes to fall asleep, having taken the active drug, took less than or equal to 30 minutes to fall asleep by the follow-up is

**Table 3: Marginal Distributions of Initial and Follow-Up Responses for Two Treatments**

| | | Response category | | | |
|---|---|---|---|---|---|
| **Response** | **Treatment** | **< 20** | **20-30** | **30-60** | **> 60** |
| Initial | Active | 0.101 | 0.168 | 0.336 | 0.395 |
| | Placebo | 0.117 | 0.167 | 0.292 | 0.425 |
| Follow-up | Active | 0.336 | 0.412 | 0.160 | 0.092 |
| | Placebo | 0.258 | 0.242 | 0.292 | 0.208 |

0.553. The same probability is just 0.294 for a patient on the placebo. This shows the level of improvement on using the active drug for an Insomnia patients who initially required more than 60 minutes falling asleep. An important question is whether or not this significant difference between the two treatments on follow-up response remains the same for all initial response levels.

However, as it can be seen, some cell counts in Table **1** are less than 5. For such a data set the use of a Bayesian approach has been suggested by so many researchers (for more details, see Agresti and Hitchcock [32]).

## 3. TRANSITION MODELS AND THE LIKELIHOOD

In this section we define homogeneous and non-homogeneous models and introduce longitudinal logistic cumulative regression model.

### 3.1. Homogeneous and Non-Homogeneous Markov Processes

Suppose that $\{Y_{m1}, Y_{m2}, ...\}$ be a set of random variables indexed by time where each element of this set can take finite values in $V = \{1, 2, ..., J\}$. Also, suppose that $\{Y_{m1}, Y_{m2}, ...\}$ forms a first-order Markov chain, (i.e. the conditional probability of $Y_{mt}$ given ($Y_{m1}$, $..., Y_{m,t-1}$) is equal to the conditional probability of $Y_{mt}$ given $Y_{m,t-1}$). Here $Y_{mt}$ represents the state of subject $m$ at time $t$. The transition probability matrix $\prod_{mt}$ for $m^{\text{th}}$ individual is

$$\Pi_{mt} = \begin{bmatrix} \pi_{m11t} & \cdots & \pi_{m1Jt} \\ \vdots & \ddots & \vdots \\ \pi_{mJ1t} & \cdots & \pi_{mJJt} \end{bmatrix}$$

where the $(a,b)^{\text{th}}$ element of $\Pi_{mt}$ is $\pi_{mabt} = P\left(Y_{mt} = b \mid Y_{m,t-1} = a\right)$ and represents the $m^{\text{th}}$ subject's probability of making transition from state $a$ at time $t-1$ to state $b$ at time $t$, that called transition probability from state $a$ to state $b$ at time $t$. If these transition probabilities are independent of time, i.e., $\Pi_{mt} = \Pi_m$ and hence $\pi_{mabt} = \pi_{mab}$, for all $t = 1, 2, ..., T$, then the Markov chain is called time homogeneous Markov chain. If the probabilities depend on time then we have a non-homogeneous Markov chain.

### 3.2. Logit Regression Model

Muenz and Rubinstein [33] presented a logistic regression setup for a binary Markov chains to incorporate covariate effects on the transition probabilities by using a logit transformation on the transition probabilities of chains. They also obtained the maximum likelihood estimates (MLE) for the transition probabilities. Also, the Markov logistic regression set-up for correlated longitudinal data and the maximum likelihood estimation for the model are discussed by [34]. In this subsection, the logit regression model is introduced. Here, we incorporate the ordinal nature of data by using a cumulative logit regression model. The cumulative logit model is specified in terms of cumulative transition probabilities

$$C_{m0b1} = P\left(Y_{m1} \le b\right) = \sum_{k=1}^{b} \pi_{m0k1},$$

$$C_{mabt} = P\left(Y_{mt} \le b \mid Y_{m,t-1} = a\right) = \sum_{k=1}^{b} \pi_{makt}, \text{ for } t > 1$$

where

$$\text{logit}\left(C_{m0b1}\right) = \log\left(\frac{C_{m0b1}}{1 - C_{m0b1}}\right) = \log\left(\frac{P\left(Y_{m1} \le b\right)}{P\left(Y_{m1} > b\right)}\right),$$

$$\text{logit}\left(C_{mabt}\right) = \log\left(\frac{C_{mabt}}{1 - C_{mabt}}\right) = \log\left(\frac{P\left(Y_{mt} \le b \mid Y_{m,t-1} = a\right)}{P\left(Y_{mt} > b \mid Y_{m,t-1} = a\right)}\right), \text{ for } t > 1$$

for $m = 1, 2, ..., M$, $a = 1, 2, ..., J$, $b = 1, 2, ..., J-1$ and $t = 1, 2, ..., T$. We write the cumulative logit model as

$$\text{logit}\left(C_{mabt}\right) = \alpha_{ab} + {}_{mt}\beta_t^a$$

where $\alpha_{ab}$ is the cut point parameter, ${}_{mt} = (X_{mt1}, ..., X_{mtK})$ is a $1 \times K$ vector of covariates for the $m^{\text{th}}$ individual and $\beta_t^a = (\beta_{t1}^a, ..., \beta_{tK}^a)'$ is a $K \times 1$ vector of regression coefficient parameters.

In ordinal data, we have a term called latent variable, $Z_{mat}$, where given $Y_{m,t-1} = a$, we will have $Y_{mt} = j$ if $\alpha_{a,j-1} < Z_{mat} < \alpha_{aj}$. The index $a$ in $Z_{mat}$ is used to show the first order Markov structure of data.

Francom *et al*. [14] used a log-linear model to analyze the data. Ganjali and Rezaee [35] used a kind of cumulative logistic regression model. We shall regard treatment as an explanatory and independent variable that affects the response or dependent variable of interest and analyze the parameters of cumulative regression model in a Bayesian framework.

For Insomnia data, an important question is whether or not this significant difference between the two treatments on follow-up response remains the same for all initial response levels. We answer this question by using a first-order transition model based on the

cumulative logistic regression framework. If we have an ordinal variable with $J$ levels, (in the Insomnia data, $J=4$), then the form of the transition model for T response variables (in the Insomnia data, $T=2$) is:

$$\text{logit}\left[P\left(Y_{m1}\le b;\alpha_b,\beta_1\right)\right]=\alpha_b+{}_{m1}\beta_1,$$

$$\text{logit}\left[P\left(Y_{mt}\le b\mid Y_{m,t-1}=a,\alpha_{ab},\beta_t^a\right)\right]=\alpha_{ab}+{}_{mt}\beta_t^a,\ \text{for}\ t>1,\quad (1)$$

where $\alpha_b$'s and $\alpha_{ab}$'s are the cut-point parameters in which $\alpha_1\le\alpha_2\ldots\le\alpha_{J-1}$ and for any $a$, $\alpha_{a1}\le\alpha_{a2}\ldots\le\alpha_{a,J-1}$, respectively. Further, the parameters $\beta_1=(\beta_1,\ldots,\beta_K)'$ and $\beta_t^a=(\beta_{t1}^a,\ldots,\beta_{tK}^a)'$ are $K\times1$ vectors of regression parameters.

Let $Y_m=(Y_{m1},Y_{m2},\ldots,Y_{mT})$ represents the observed transition of the $m^{\text{th}}$ individual over $t=1,2,\ldots,T$ time periods, and $Y=(Y_1,Y_2,\ldots,Y_M)$ be the observed transition of $M$ individuals over above time periods. For $T=2$ and our Insomnia data, let $\boldsymbol{\alpha_0}:=(\alpha_1,\alpha_2,\alpha_3)$, $\boldsymbol{\alpha_a}:=(\alpha_{a1},\alpha_{a2},\alpha_{a3})$ for $a=1,\ldots,4$, $\boldsymbol{\beta_1}=\boldsymbol{\beta_1}$, $\boldsymbol{\beta_t^a}=\boldsymbol{\beta_2^a}=(\beta_2^1,\beta_2^2,\beta_2^3,\beta_2^4)$, $\boldsymbol{\alpha}:=(\boldsymbol{\alpha_0},\boldsymbol{\alpha_1},\boldsymbol{\alpha_2},\boldsymbol{\alpha_3},\boldsymbol{\alpha_4})'$ and $\boldsymbol{\beta}:=(\beta_1,\beta_2^1,\beta_2^2,\beta_2^3,\beta_2^4)'$, using the transition model, the likelihood function for these data is:

$$L(\boldsymbol{Y};\alpha,\beta)=\prod_{m=1}^{M}P(Y_{m1}=y_{m1},Y_{m2}=y_{m2})\qquad(2)$$

$$=\prod_{m=1}^{M}P(Y_{m1}=y_{m1})\cdot P(Y_{m2}=y_{m2}\mid Y_{m1}=y_{m1})$$

$$=\prod_{m=1}^{M}P(Y_{m1}=y_{m1};\alpha,\beta)\prod_{m=1}^{M}P(Y_{m2}=y_{m2}\mid Y_{m1}=y_{m1};\alpha,\beta)$$

$$=\prod_{m=1}^{M}P(Y_{m1}=y_{m1};\alpha_0,\beta_1)\prod_{y_{m1}=1}^{J}\times$$

$$\prod_{[ml Y_{m1}=y_{m1}]}P(Y_{m2}=y_{m2}\mid Y_{m1}=y_{m1};\alpha_{y_{m1}},\beta_2^{y_{m1}}),$$

where

$$P(Y_{m1}=b;\alpha_0,\beta_1)=P(Y_{m1}\le b;\alpha_b,\beta_1)-P(Y_{m1}\le b-1;\alpha_{b-1},\beta_1),$$

and

$$P\left(Y_{m2}=b\mid Y_{m1}=a,\alpha_a,\beta_2^a\right)=P\left(Y_{m2}\le b\mid Y_{m1}=a,\alpha_{ab},\beta_2^a\right)$$
$$-P\left(Y_{m2}\le b-1\mid Y_{m1}=a,\alpha_{a,b-1},\beta_2^a\right).$$

The maximization of likelihood can be carried out using SPSS package, Ganjali and Rezaee [35], by modeling $Y_1$, then separately modeling the conditional probability of $Y_2$ given $Y_1$ in each level of previous response.

## 4. BAYESIAN INFERENCE

For Analysis of contingency table, where the sample size is small or some cells have frequencies less than 5, the frequentist approaches are so conservative and do not work properly. So, a Bayesian approach may be a better method for obtaining the parameter estimates and testing some hypotheses. In this section, we use the Bayesian paradigm to make inferences about parameters of the model (1) for Insomnia data.

Because of the ordering constraints in cut point parameters, the multivariate normal distribution is not suitable for setting as prior distribution. Thus we use independent truncated normal distributions for the cut point vectors $\boldsymbol{\alpha_0}=(\alpha_1,\ldots,\alpha_{J-1})$, (for initial time), and $\boldsymbol{\alpha_a}=(\alpha_{a1},\ldots,\alpha_{a,J-1})$, (for follow-up time), such that $\alpha_0\le\alpha_1\le\ldots\le\alpha_{J-1}\le\alpha_J$ and $\alpha_{a0}\le\alpha_{a1}\le\ \le\alpha_{a,J-1}\le\alpha_{aJ}$, respectively. We set $\alpha_0=-\infty$ and $\alpha_J=+\infty$ and also $\alpha_{a0}=-\infty$ and $\alpha_{aJ}=+\infty$ for $a=1,2\ldots,J$.

We set the first order Markov structure on $\beta_t^a$ by

$$\beta_t^a\mid\beta_{t-1}^a,\boldsymbol{\Sigma}\sim N(\beta_{t-1}^a,\boldsymbol{\Sigma}),\ \ \text{for}\ t\ge2$$

and for $t=1$

$$\beta_1^a\mid\boldsymbol{\Sigma}\sim N(0,\boldsymbol{\Sigma})$$

where $\Sigma$ is a $K$-dimensional diagonal covariance matrix defined as $\Sigma=\sigma_\beta^2 I$ and $\sigma_\beta^2=(\sigma_{\beta_1}^2,\ldots,\sigma_{\beta_K}^2)'$ is a known vector. For cut point parameters we set the following prior distributions:

$$\alpha_a\sim TN(\mu,\sigma^2),a=1,2,3.\ \text{where}\ \alpha_1\le\alpha_2\le\alpha_3,$$

$$\alpha_{ab}\sim TN(\mu_a,\sigma_a^2),b=1,2,3.\ \text{where}\ \alpha_{a1}\le\alpha_{a2}\le\alpha_{a3},$$

where $TN(\mu,\sigma^2)$ is a truncated normal distribution with parameters $\mu$ and $\sigma^2$. The transition probability $\pi_{mabt}$ is obtained as $\pi_{mabt}=C_{mabt}-C_{ma,b-1,t}$ for $b=2,\ldots,J$ and $\pi_{ma1t}=C_{ma1t}$. The joint posterior distribution given the transition data on $M$ individuals for $T=2$ time periods is proportional to

$$\left(\begin{array}{l}\prod_{m=1}^{M}P(Y_{m1}=y_{m1};\alpha_0,\beta_1)\prod_{y_{m1}=1}^{J}\prod_{[ml Y_{m1}=y_{m1}]}\\ P(Y_{m2}=y_{m2}\mid Y_{m1}=y_{m1};\alpha_{y_{m1}},\beta_2^{y_{m1}})\end{array}\right)\times\left(\prod_{t=1}^{2}\pi\left(\beta_t^a\mid\beta_{t-1}^a,\Sigma\right)\right)\left(\pi(\alpha)\right).$$

**Table 4: The Bayesian and Frequentist Estimates of Parameters by Marginal Modeling of Responses, without Considering the Relationship Between Two Responses. (Parameters Significant at the 5% Level are Highlighted in Bold)**

| Initial Response | | | | | Follow-up Response | | | |
|---|---|---|---|---|---|---|---|---|
| Bayesian | | | Frequentist | | Bayesian | | Frequentist | |
| Par. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. |
| $\alpha_1$ | **-2.155** | 0.242 | **-2.121** | 0.239 | **-1.320** | 0.198 | **-1.291** | 0.195 |
| $\alpha_2$ | **-0.989** | 0.189 | **-0.982** | 0.187 | 0.120 | 0.176 | 0.123 | 0.174 |
| $\alpha_3$ | **0.352** | 0.179 | **0.346** | 0.177 | **1.408** | 0.210 | **1.396** | 0.206 |
| $\beta$ | 0.033 | 0.237 | 0.035 | 0.237 | **0.725** | 0.241 | **0.761** | 0.238 |

For implementation, we use noninformative proper prior distributions. We used independent truncated normal distributions, $N(0,10000)$, for cut point parameters $\alpha_a$ and $\alpha_{ab}$. For the drug effect parameters $\beta_1$ and $\beta_2^a$ we use $N(0,10000)$ and $N(\beta_1,10000)$ priors, respectively. The results of frequentist and Bayesian analysis will be given in the next section. Here, we present the Bayesian analysis of the mentioned transition model by setting some prior distributions on the parameters of the logistic regression model. We use Metropolis-Hastings algorithm to generate samples from posterior distribution of $\alpha_a$, $a=1,2,3,$ and $\beta_1$. We consider the following prior distributions for our analysis;

$$\beta_1 \sim N(0,10000),$$

$$\alpha_a \sim TN(0,10000), a=1,2,3, \text{where} \quad \alpha_1 \le \alpha_2 \le \alpha_3.$$

We run the program in the WinBUGS for $10000$ iterations with a burn-in time 100 and the thin 100. The results are given in Table **4**.

When for $m^{\text{th}}$ individual, $Y_{m1}=a,\ a=1,\ldots,4$, i.e., the initial response is in the $a^{\text{th}}$ order category, we perform the Bayesian analysis by first detecting individuals who have initial response in $a^{\text{th}}$ category and then set up some prior distributions in parameters of logistic regression model. In the first step of Bayesian analysis we consider the following set-up for the parameters of logit model. For $m^{\text{th}}$ individual, if $Y_{m1}=a$, $a=1,\ldots,4$, we consider;

$$\beta_2^a \sim N(\beta_1,10000),$$

$$\alpha_{ab} \sim TN(0,10000),\ b=1,2,3, \text{where } \alpha_{a1} \le \alpha_{a2} \le \alpha_{a3}.$$

The posterior density of coefficient parameters are given in Figure **1** and the posterior summary for the parameters is given in Table **5**. Results will be interpreted in the next section.

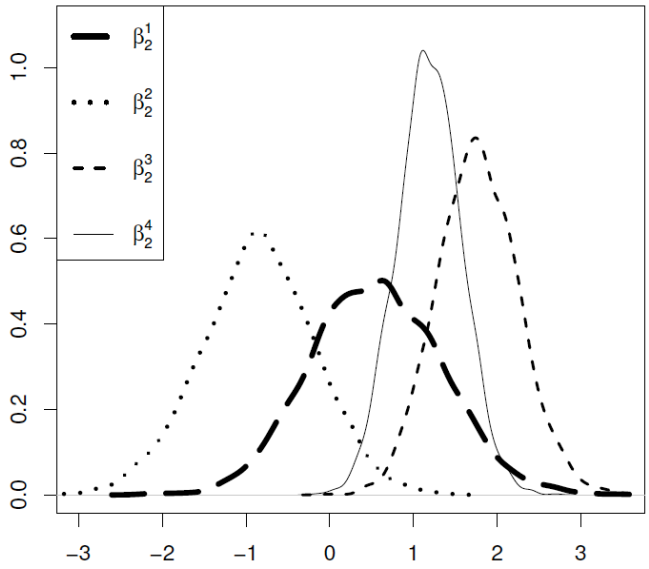


**Figure 1:** Posterior density functions of $\beta_2^1$, $\beta_2^2$, $\beta_2^3$ and $\beta_2^4$.

### 4.1. Hypothesis Testing

In this subsection, we would like to test some hypotheses about main parameters of the model for follow-up response. By these tests we shall find the best model that can be fitted to Insomnia data.

The first hypothesis that we want to test leads to a model which says that treatment is not significant for any given initial response, i.e.,

$$H_1: \beta_2^1=\beta_2^2=\beta_2^3=\beta_2^4=0.$$

The second hypothesis is that the treatment on follow-up response is significant but its performance is the same for all different levels of initial response, i.e.,

$$H_2: \beta_2^1=\beta_2^2=\beta_2^3=\beta_2^4=\beta^* \text{ for } \beta^* \neq 0.$$

For Insomnia data, by plot given in Figure **1**, one may get interest in testing $\beta_2^1=\beta_2^2$ and $\beta_2^3=\beta_2^4$. In other words, it seems that if we partition the patients in

**Table 5: The Bayesian and Frequentist Estimates of Parameters by Transition Modeling of Responses. (Parameters Significant at the 5% Level are Highlighted in Bold)**

| | $Y_2 \mid Y_1 < 20$ | | | | $Y_2 \mid 20 < Y_1 < 30$ | | | |
|---|---|---|---|---|---|---|---|---|
| | **Bayesian** | | **Frequentist** | | **Bayesian** | | **Frequentist** | |
| **Par.** | **Est.** | **S.E.** | **Est.** | **S.E.** | **Est.** | **S.E.** | **Est.** | **S.E.** |
| $\alpha_{a1}$ | -0.113 | 0.554 | -0.089 | 0.522 | 0.956 | 0.495 | 0.909 | 0.490 |
| $\alpha_{a2}$ | **1.635** | 0.675 | **1.478** | 0.621 | **2.561** | 0.642 | **2.377** | 0.623 |
| $\alpha_{a3}$ | **3.844** | 1.365 | **3.007** | 1.058 | **3.943** | 0.937 | **3.397** | 0.833 |
| $\beta_2^a$ | 0.566 | 0.798 | 0.507 | 0.765 | -0.848 | 0.671 | -0.792 | 0.651 |
| | $Y_2 \mid 30 < Y_1 < 60$ | | | | $Y_2 \mid Y_1 > 60$ | | | |
| | **Bayesian** | | **Frequentist** | | **Bayesian** | | **Frequentist** | |
| **Par.** | **Est.** | **S.E.** | **Est.** | **S.E.** | **Est.** | **S.E.** | **Est.** | **S.E.** |
| $\alpha_{a1}$ | **-2.313** | 0.452 | **-2.235** | 0.436 | **-2.639** | 0.391 | **-2.561** | 0.385 |
| $\alpha_{a2}$ | -0.089 | 0.345 | -0.083 | 0.330 | **-0.951** | 0.294 | **-0.926** | 0.288 |
| $\alpha_{a3}$ | **2.783** | 0.658 | **2.592** | 0.610 | 0.330 | 0.277 | 0.317 | 0.272 |
| $\beta_2^a$ | **1.760** | 0.489 | **1.705** | 0.477 | **1.200** | 0.385 | **1.161** | 0.381 |

two groups: acute and chronic patients; by their initial responses (where their times to asleep is less than 30 minutes and greater than 30 minutes, respectively), then the treatment effect on follow-up response is the same in each group, i.e., we want to also test

$$H_3 : \beta_2^1 = \beta_2^2 \text{ and } \beta_2^3 = \beta_2^4.$$

Finally, the last model that we consider is a general model where all coefficients are considered to be differently effective and so this model is the full model. We name this hypothesis as $H_4$, i.e.,

$$H_4 : \beta_2^1 \neq \beta_2^2 \neq \beta_2^3 \neq \beta_2^4.$$

In Bayesian paradigm, one approach to model selection is to use the Bayes factor. The comparison between two models $H_i$ versus $H_j$, $i, j \in \{1,2,3,4\}$ and $i \neq j$, is possible by computing the Bayes factor. If $\pi(H_i)$ and $\pi(H_j)$ are the prior probabilities of these models then the Bayes factor is obtained by

$$B_{ij} = \frac{f(Y \mid H_i)}{f(Y \mid H_j)} = \frac{\frac{\pi(H_i \mid Y)}{\pi(H_j \mid Y)}}{\frac{\pi(H_i)}{\pi(H_j)}}.$$

It is obvious that $B_{ij}$ is the ratio of posterior odds of $H_i$ versus $H_j$ and prior odds. For computing the Bayes factor, it is required that under each model the integral

$$f(Y \mid H) = \int f(Y \mid H, \beta_2^a) \pi(\beta_2^a \mid H) d\beta_2^a \quad (3)$$

can be computed. Because of computational complexity, the use of MCMC is very helpful. Due to lack of a closed form of full conditional of $\beta_2^a$, for estimating of $f(Y \mid H)$, we use a method proposed by Newton and Raftery [36]. For this, if $\{\beta_2^{a(i)}\}_{i=1}^m$ are samples from posterior distribution of $\pi(\beta_2^a \mid Y, H)$ in the $(j+1)^{\text{th}}$ iteration, the $f(Y \mid H)$ is estimated as :

$$f^{(j+1)}(Y \mid H) = \frac{\frac{\kappa m}{1-\kappa} + \sum_{i=1}^{m} \frac{f(Y \mid \beta_2^{a(i)}, H)}{\kappa f^{(j)}(Y \mid H) + (1-\kappa) f(Y \mid \beta_2^{a(i)}, H)}}{\frac{\kappa m}{(1-\kappa) f^{(j)}(Y \mid H)} + \sum_{i=1}^{m} \frac{1}{\kappa f^{(j)}(Y \mid H) + (1-\kappa) f(Y \mid \beta_2^{a(i)}, H)}}. \quad (4)$$

where $\kappa$ is small value in $(0,1)$.

The process is iterated till reaching to a convergence. In some small-scale numerical experiments, it has been shown that the above estimator was performed well for $\kappa$ as small as 0.01. The Bayes factor for each couple of models, given in this subsection, is given in Table **6**.

### 4.2. Conditional Predictive Ordinate

In this subsection, we use the conditional predictive ordinate (CPO), [37], which is a Bayesian diagnostic tool to detect observations discrepant from a given

**Table 6: Bayes Factor for Models in Subsection 4.1. In Each Cell the Number is the Bayes Factor of Column Model Against Row Model**

| Hypothesis | $H_1$ | $H_2$ | $H_3$ | $H_4$ |
|---|---|---|---|---|
| $H_1$ | 1 | 563.848 | $17.158\times10^{3}$ | $18.200\times10^{3}$ |
| $H_2$ | 0.002 | 1 | 30.432 | 32.278 |
| $H_3$ | $5\times10^{-05}$ | 0.033 | 1 | 1.061 |
| $H_4$ | $5\times10^{-05}$ | 0.031 | 0.943 | 1 |

model. It can be adapted to detect unusual patients in our data set.

Let $\boldsymbol{Y}=(y_1,y_2,\ldots,y_n)$ be the vector of conditionally independent observations with density $f(y_i\mid\theta)$ and $\boldsymbol{y}_{(i)}$ be the $\boldsymbol{y}$ after omitting $y_i$, i.e., $y_{(i)}=(y_1,y_2,\ldots,y_{i-1},y_{i+1},\ldots,y_n)$. For the $i^{\text{th}}$ observation, the CPO is defined as

$$\mathrm{CPO}_i=\boldsymbol{f}(y_i\mid\boldsymbol{y}_{(i)})=\int f(y_i\mid\theta,\boldsymbol{y}_{(i)})p(\theta\mid\boldsymbol{y}_{(i)})d\theta. \quad (5)$$

Small CPOs indicate observations that are not expected under the current model. A Monte Carlo estimate for $\mathrm{CPO}_i$ is given by

$$\widehat{\mathrm{CPO}}_i=\left(\frac{1}{T'}\sum_{t=1}^{T'}\frac{1}{f(y_i\mid\theta^{(t)})}\right)^{-1},$$

([38], p. 511) which is the harmonic mean of the probability distribution function evaluated at $y_i$ for each $\theta^{(t)}$ for $t=1,2,\ldots,T'$. For Insomnia data, where $Y_{m1}=a$, $a=1,\ldots,4$, we compute the CPOs for follow-up responses. After computing the CPOs we monitor them, then compare the values of CPOs to select the small values which detect abnormal responses.

According to Figure **2** $(i)$, we see that the $12^{\text{th}}$, $24^{\text{th}}$, $25^{\text{th}}$ and $26^{\text{th}}$ individuals have smaller values of CPO than others, hence they are odd observations. By a quick inspection, we can see that these are in fact odd individuals. For example, the $12^{\text{th}}$ individual uses active drug and her/his initial response is $1$, but moves to 3rd level of follow-up response. The $24^{\text{th}}$ and $25^{\text{th}}$ individuals who use placebo move to $3^{\text{rd}}$ level of follow-up response when their initial response were $1$. The $26^{\text{th}}$ individual also was in the first level of initial response and uses the placebo, but moves to $4^{\text{th}}$ level of follow-up response.

For people who are in the second level of initial response we see that the $17^{\text{th}}$ to $20^{\text{th}}$ and $40^{\text{th}}$ individuals have small values of CPO. Similar to the above individuals these individuals are also unusual.

For third level of initial response the $40^{\text{th}}$, $74^{\text{th}}$ and $75^{\text{th}}$ individuals have the smaller values of CPO than others. For fourth level of initial response we see that the $48^{\text{th}}$ to $51^{\text{th}}$ individuals have the small values of CPO. These are individuals who use placebo, but move to first level of follow-up response.

## 5. BAYESIAN AND FREQUENTIST APPROACHES FOR ANALYZING INSOMNIA DATA

In this section the numerical results in both frequentist and Bayesian approaches for Insomnia data are presented. The estimates of parameters of cumulative logit model from modeling the initial and the follow-up responses separately, are given in Table **4**. Results in Table **4** for initial response show that the effect of treatment is not significant on its cumulative probability. But results for follow-up response reveal the significant effect of treatment on its cumulative probability. This table says that the use of the drugs increases the probability of having a short time interval to get sleep in bed.

In Table **5**, the results of conditional components of the transition model are given by both frequentist and the Bayesian analysis. For different values of the initial response, $\alpha_a$, $a=1,2,3$, are intercepts which indicate the log-odds of lower, rather than higher time to falling asleep when we use the placebo. For example, when the initial time to falling asleep is greater than 60 minutes, for follow-up response log-odds of less than 60 rather than time more than 60 when we use the active drug is $-2.561+1.161=-1.400$ by frequentist approach and is $-2.639+1.200=-1.439$ by Bayesian approach. When we use the placebo, above log-odds are $-2.561$ and $-2.639$ by frequentist and Bayesian computations, respectively.

According to Table **5**, it is clear that for an initial value of 30-60 or more than 60, the positive effect of drug on the cumulative probability of low values of follow-up response is significant. But for an initial value of less than 20 or 20-30, there is no significant effect of drug. Both of frequentist and Bayesian results confirm the above inferences. These mean that the drug is more effective for patients with presleeping intervals greater than 30 minutes. Hence the knowledge of the

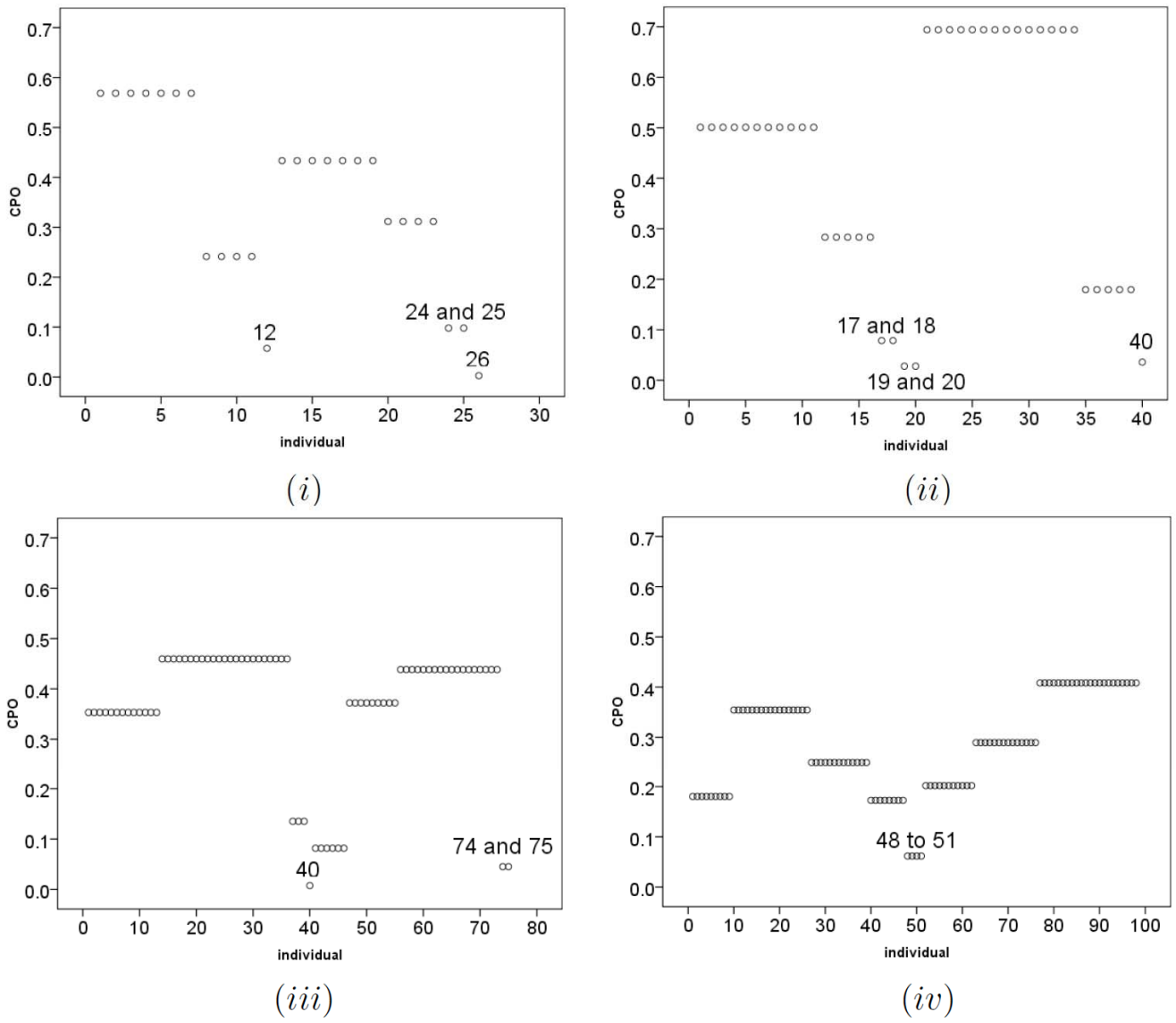


**Figure 2:** Conditional predictive ordinate for follow-up responses given their initial values, ( $i$ ) for $Y_{m1}=1$ , ( $ii$ ) for $Y_{m1}=2$ , ( $iii$ ) for $Y_{m1}=3$ and ( $iv$ ) for $Y_{m1}=4$ .

initial response may helps the practitioners to prescribe an effectual treatment, and also discourage them from prescribing an ineffectual treatment.

Bayes factors for comparing the models are given in Table **6**. According to this table, we see that between two models $H_1$ and $H_2$ we can approve the model $H_2$ because of high value of $B_{21}$. Then, between two models $H_2$ and $H_3$, the model $H_3$ is approved, because the value of $B_{32}$ is equal to $30.432$. Finally we want to decide which one of $H_3$ or $H_4$ should be chosen! The Bayes factor for comparing $H_3$ and $H_4$ is $1.061$. According to Jeffreys [39], we conclude that this value of $B_{43}$ didn't say that $H_3$ should be rejected against $H_4$. In fact model $H_4$ is a bit better than model $H_3$, but we have no strong evidence that $H_4$ dominates $H_3$. Hence, we can accept both of models $H_3$ and $H_4$, but the more parsimonious one is model $H_3$.

### 5.1. Sensitivity Analysis: Investigating Sensitivity of Posterior Inferences on Changes of Prior Variances

In section 4, we described a choice of prior structure on parameters of logit model. We now investigate the sensitivity of posterior results (with emphasis on coefficient and partial gamma parameters) to changes on the prior. In the next subsection, we also investigate the sensitivity to eliminate some unusual individuals.

**Table 7: Posterior Mean of Partial Gamma and Coefficients for Various Values of Prior Variance**

| | $10^{-6}$ | $10^{-4}$ | $10^{-2}$ | $10^{0}$ | $10^{2}$ | $10^{4}$ | $10^{6}$ |
|---|---|---|---|---|---|---|---|
| $\gamma(0)$ | 0.549 | 0.549 | 0.552 | 0.599 | 0.604 | 0.604 | 0.603 |
| $\gamma(1)$ | 0.549 | 0.549 | 0.545 | 0.509 | 0.507 | 0.507 | 0.507 |
| $\beta_1$ | $2.07\times10^{-6}$ | $7.13\times10^{-5}$ | 0.006 | 0.036 | 0.035 | 0.033 | 0.033 |
| $\beta_2^1$ | $1.35\times10^{-5}$ | $1.78\times10^{-4}$ | 0.009 | 0.335 | 0.558 | 0.566 | 0.570 |
| $\beta_2^2$ | $-6.94\times10^{-6}$ | $-1.46\times10^{-4}$ | -0.019 | -0.582 | -0.842 | -0.848 | -0.849 |
| $\beta_2^3$ | $1.31\times10^{-5}$ | $9.46\times10^{-4}$ | 0.080 | 1.430 | 1.760 | 1.760 | 1.760 |
| $\beta_2^4$ | $2.32\times10^{-6}$ | $8.83\times10^{-4}$ | 0.079 | 1.050 | 1.200 | 1.200 | 1.200 |

We conduct the sensitivity analysis with changing the variance of the prior distribution for coefficients of logit models ( $\beta$ ) and we want to know whether the posterior estimates will change with the change of the variance or they will be stable. For doing this, we set the values $10^{-6}$ , $10^{-4}$ , $10^{-2}$ , $10^{0}$ , $10^{2}$ , $10^{4}$ and $10^{6}$ for the variance of the prior distribution for different coefficients. The results are given in Table **7**. When we select a very small value for variance of the coefficients, such as $10^{-6}$ , the prior distributions are more informative than the prior distributions corresponding to high values for variance of the coefficients. The posterior means of the parameters corresponding to those obtained by noninformative priors, are similar to the frequentist ones. But, for more informative priors, the Bayesian estimates of parameters have significant difference from those of frequentist estimates. Figure **3** shows the posterior distributions of $\gamma(0)$ and $\gamma(1)$ for various values of prior variances of the coefficients.

### 5.2. Sensitivity Analysis: Investigating Sensitivity to Elimination of Some Odd Patients

In Insomnia data, some unexpected counts detected by using CPO. Now, we want to know how these individuals affect the posterior results.

For doing this, we select some individuals with low value of CPO and after deleting one or a group of them we obtain the posterior results. These results are given in Table **8**. According to Table **8**, when we omit some people from some cells, the parameters estimates will change and the size of change is dependent on the number and the position of deleted people. For example, when the initial response is 3, ('30 to 60 minutes'), the effect of the treatment, $\beta_2^3$, is very sensitive to deletion of the odd individual who uses active drug, (individual 40), where her/his follow-up response is high ('more than 60 minutes'). Another example is $\beta_2^2$ that has large change when we delete

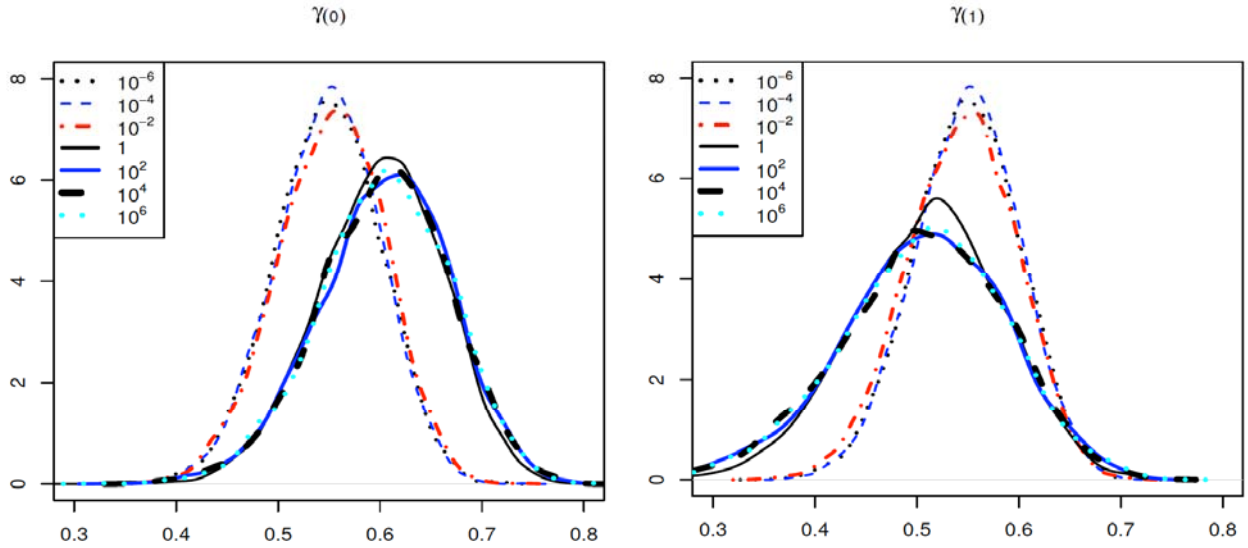


**Figure 3:** Posterior distribution of $\gamma(0)$ (left) and $\gamma(1)$ (right), for different values of prior variances of the coefficients.

**Table 8: Posterior Mean of Coefficients when we Omit Some Odd Individuals (ind.). (Parameters Significant at the 5% Level are Highlighted in Bold)**

| | **Full data** | **without 12th ind.** | **without 24th ind.** | **without 24th and 25th ind.s** | **without 26th ind.** | **without 17th ind.** | **without 17th and 18th ind.s** | **without 19th ind.** | **without 19th and 20th ind.s** |
|---|---|---|---|---|---|---|---|---|---|
| | | $Y_{m1}=1$ | $Y_{m1}=1$ | $Y_{m1}=1$ | $Y_{m1}=1$ | $Y_{m1}=2$ | $Y_{m1}=2$ | $Y_{m1}=2$ | $Y_{m1}=2$ |
| $\beta_1$ | 0.034 | 0.063 | 0.021 | 0.095 | 0.063 | 0.070 | 0.102 | 0.067 | 0.094 |
| $\beta_2^1$ | 0.566 | 1.220 | 0.338 | -0.004 | 0.309 | 0.564 | 0.551 | 0.563 | 0.558 |
| $\beta_2^2$ | -0.848 | -0.852 | -0.851 | -0.834 | -0.849 | -0.520 | -0.171 | -0.500 | -0..059 |
| $\beta_2^3$ | **1.760** | **1.770** | **1.770** | **1.750** | **1.770** | **1.770** | **1.760** | **1.770** | **1.710** |
| $\beta_2^4$ | **1.200** | **1.190** | **1.200** | **1.180** | **1.200** | **1.190** | **1.190** | **1.190** | **1.150** |
| | **without 40th ind.** | **without 40th ind.** | **without 74th ind.** | **without 74th and 75th ind.s** | **without 48th ind.** | **without 48th and 49th ind.s** | **without 48th to 50th ind.s** | **without 48th to 51th ind.s** | **without all ind.s with small CPO** |
| | $Y_{m1}=2$ | $Y_{m1}=3$ | $Y_{m1}=3$ | $Y_{m1}=3$ | $Y_{m1}=4$ | $Y_{m1}=4$ | $Y_{m1}=4$ | $Y_{m1}=4$ | |
| $\beta_1$ | 0.054 | 0.082 | 0.041 | 0.039 | 0.078 | 0.058 | 0.0381 | 0.019 | -0.036 |
| $\beta_2^1$ | 0.573 | 0.561 | 0.587 | 0.552 | 0.522 | 0.605 | 0.576 | 0.555 | 0.002 |
| $\beta_2^2$ | -1.090 | -0.855 | -0.836 | -0.839 | -0.856 | -0.832 | -0.829 | -0.849 | -0.258 |
| $\beta_2^3$ | **1.770** | **2.170** | **1.720** | **1.640** | **1.770** | **1.780** | **1.770** | **1.780** | **1.820** |
| $\beta_2^4$ | **1.180** | **1.190** | **1.180** | **1.200** | **1.250** | **1.390** | **1.460** | **1.590** | **1.610** |

people who use active drug in second level of initial response , (individuals 17, 18, 19 and 20).

## 6. CONCLUSION

For analysing contingency tables with small observed counts in some cells, frequentist inferences would not work properly and one may use the Bayesian statistics instead. We set a Bayesian logit regression model to the Insomnia data and estimate the parameters of the model. Then we compared the frequentist and Bayesian results and showed that by using some noninformative priors the Bayesian and frequentist approaches are given the same results.

For measure of association we used the gamma. In each level of control variable, the partial gamma was computed to measure the relationship between responses. Then the comparison between the Bayesian and frequentist results were presented by setting the Dirichlet prior distribution on cell probabilities. In our data set, there exist some observations that are abnormal under the current model. We illustrated how the CPO can be used to detect these individuals. We investigated the sensitivity of posterior results to changes on prior variances and to deletion of odd individuals. We also used the Bayes factor criterion to test some hypotheses about parameters of the model.

In Bayesian hypotheses testing, some hypotheses on regression coefficient parameters were tested and the best model was selected by using Bayes factor. The best model that was chosen, *via* using Bayes factor, was the model $H_3$ which considers that $\beta_2^1 = \beta_2^2$ and $\beta_2^3 = \beta_2^4$. This model says that if the patients are partitioned in two groups (acute and chronic), then the treatment effect on follow-up response is the same in each group. Our application was a two-period longitudinal study. Extension of our analysis on longitudinal data with more periods may make the model a more applicable one.

## ACKNOWLEDGEMENTS

The authors are grateful to the editor and referees for their valuable comments and helpful insights, which led to improvement of the paper.

## REFERENCES

[1] Agresti A. Modelling ordered categorical data: recent advances and future challenges. Stat Med 1999; 18: 2197-207.

http://dx.doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2191::AID-SIM249>3.0.CO;2-M

[2] Lall R, Campbell MJ, Walters SJ, Morgan K. A review of ordinal regression models applied on health-related quality of life assessments. Stat Methods Med Res 2002; 11: 49-67. http://dx.doi.org/10.1191/0962280202sm271ra

[3] Ten Have TR, Landis JR, Hartzel J. Population-averaged and cluster-specific models for clustered ordinal response data. Stat Med 1996; 15: 2573-88. http://dx.doi.org/10.1002/(SICI)1097-0258(19961215)15:23<2573::AID-SIM389>3.0.CO;2-O

[4] Kim K. A bivariate cumulative probit regression model for ordered categorical data. Stat Med 1995; 14: 1341-52. http://dx.doi.org/10.1002/sim.4780141207

[5] Liang KY, Zeger SL, Qaqish BF. Multivariate regression analyses for categorical data. J R Stat Soc Ser B Stat Methodol 1992; 54(1): 3-40.

[6] Molenberghs G, Lesaffre E. Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. J Am Stat Assoc 1994; 89: 633-44. http://dx.doi.org/10.1080/01621459.1994.10476788

[7] Harville DA, Mee RW. A mixed-model procedure for analyzing ordered categorical data. Biometrics 1984; 40: 393-408. http://dx.doi.org/10.2307/2531393

[8] Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random- effects population. J Am Stat Assoc 1996; 91: 217-21. http://dx.doi.org/10.1080/01621459.1996.10476679

[9] Tutz G, Hennevogl W. Random effects in ordinal regression models. Comput Stat Data Anal 1996; 22: 537-57. http://dx.doi.org/10.1016/0167-9473(96)00004-7

[10] Verbeke G, Molenberghs G. Linear mixed models in practice: a SAS-oriented approach. New York: Spinger 1997.

[11] Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. Analysis of longitudinal data. Oxford: Oxford University Press 2002.

[12] Tutz G. Modelling of repeated ordered measurements by isotonic sequential regression. Stat Model 2005; 5(4): 269-87. http://dx.doi.org/10.1191/1471082X05st101oa

[13] Garber AM. A discrete-time model of the acquisition of antibiotic-resistant infections in hospitalized patients. Biometrics 1989; 45: 797-16. http://dx.doi.org/10.2307/2531684

[14] Francom SF, Chuang-Stein C, Landis JR. A log-linear model for ordinal data to characterize differential change among treatments. Stat Med 1989; 8: 571-82. http://dx.doi.org/10.1002/sim.4780080506

[15] Chung H, Park YS, Lanza ST. Latent transition analysis with covariates: pubertal timing and substance use behaviours in adolescent females. Stat Med 2005; 24: 2895-10. http://dx.doi.org/10.1002/sim.2148

[16] Rezaei Ghahroodi Z, Ganjali M. Testing homogeneity in Markov models for analyzing longitudinal ordinal response data with random dropout. J Stat Theory Appl 2009; 8(2): 125-39. http://dx.doi.org/10.1080/10543400902964100

[17] Rezaei Ghahroodi Z, Ganjali M, Berridge D. A transition model for ordinal response data with random dropout: an application to the uvoxamine data. J Biopharm Statist 2009; 19(4): 658-71.

[18] McCullagh P. Regression models for ordinal data. J R Stat Soc Ser B Stat Methodol 1980; 42(2): 109-42.

[19] Agresti A. Analysis of ordinal categorical data. 2nd ed. New York: John Wiley & Sons 2010. http://dx.doi.org/10.1002/9780470594001

[20] Sung M, Soyer R, Nhan N. Bayesian analysis of non-homogeneous Markov chains: application to mental health data. Stat Med 2007; 26: 3000-17. http://dx.doi.org/10.1002/sim.2775

[21] Zeghnoun A, Czernichow P, Declercq C. Assessment of short-term association be- tween health outcomes and ozone concentrations using a Markov regression model. Environmetrics 2003; 14: 271-82. http://dx.doi.org/10.1002/env.585

[22] Chan JSK, Wan WY. Bayesian approach to analysing longitudinal bivariate binary data with informative dropout.Comput Stat 2011; 26: 121-44. http://dx.doi.org/10.1007/s00180-010-0213-5

[23] Mansouriana M, Kazemnejada A, Kazemi I, Zayeri F, Soheilian M. Bayesian analysis of longitudinal ordered data with flexible random effects using McMC: application to diabetic macular Edema data. J Appl Stat 2012; 39(5): 1087-100.

[24] Mandel M, Gauthier SA, Guttmann CRG, Weiner HL, Betensky RA. Estimating time to event from longitudinal categorical data: an analysis of multiple sclerosis progression. J Am Stat Assoc 2007; 102: 1254-66. http://dx.doi.org/10.1198/016214507000000059

[25] Anderson TW, Goodman LA. Statistical inference about Markov chains. Annals of mathematical statistics 1957; 28(1): 89-10. http://dx.doi.org/10.1214/aoms/1177707039

[26] Lee TC, Judge GG, Zellner A. Estimating the parameters of the Markov probability model from aggregate time series data. 2nd ed. Amsterdam: North-Holland Publication Company 1970.

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

[28] Healy BC, Engler D. Modeling disease-state transition heterogeneity through Bayesian variable selection. Stat Med 2009; 28: 1353-68. http://dx.doi.org/10.1002/sim.3545

[29] Goodman LA, Kruskal WH. Measures of association for cross-classification. J Am Stat Assoc 1954; 49: 732-804.

[30] Lindley DV. The Bayesian analysis of contingency tables. Annals of mathematical statistics 1964; 35: 1622-43. http://dx.doi.org/10.1214/aoms/1177700386

[31] Good I. The population frequencies of species and the estimation of population parameters. Biometrika 1953; 40: 237-64.

[32] Agresti A, Hitchcock DB. Bayesian inference for categorical data analysis. Stat Methods Appt 2005; 14(3): 297-30. http://dx.doi.org/10.1007/s10260-005-0121-y

[33] Muenz LR, Rubinstein LV. Markov models for covariate dependence of binary sequences. Biometrics 1985; 41(1): 91-101. http://dx.doi.org/10.2307/2530646

[34] Zeger S, Qaqish B. Markov regression models for time series: a quasi-likelihood approach. Biometrics 1988; 44: 1019-31. http://dx.doi.org/10.2307/2531732

[35] Ganjali M, Rezaee Z. A transition model for analysis of repeated measure ordinal response data to identify the effects of different treatments. Drug Inf J 2007; 41: 527-34.

[36] Newton MA, Raftery AE. Approximate Bayesian inference by the weighted likelihood bootstrap. J R Stat Soc Ser B Stat Methodol 1994; 56(1): 3-48.

[37] Geisser S, Eddy WF. A predictive approach to model selection. J Am Stat Assoc 1979; 74: 153-60. http://dx.doi.org/10.1080/01621459.1979.10481632

[38] Gelfand AE, Dey DK. Bayesian model choice, asymptotics and exact calculations. J R Stat Soc Ser B Stat Methodol 1994; 56(3): 50-514.

[39] Jeffreys H. Theory of probability. 3rd ed. Oxford: Oxford University Press 1961.

Received on 01-10-2012 Accepted on 18-11-2012 Published on 21-12-2012

http://dx.doi.org/10.6000/1929-6029.2012.01.02.08