IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 30, 2021, accepted July 11, 2021, date of publication July 15, 2021, date of current version July 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3097363

Detection of Factors Affecting State Transition Based on Non-Homogeneous Markov Chain Model

SHEN XIUJUAN, LIU HEFEI, LI YONG, AND YUAN BIN
School of Mathematics and Statistics, Qujing Normal University, Qujing 655011, China

Corresponding author: Li Yong (qjsfxyly@163.com)

This work was supported in part by the Yunnan Provincial Science and Technology Department of China under Grant 2018FH001-109 and Grant 2019FH001-108; and in part by the Yunnan Provincial Department of Education of China under Grant 2019J0602, Grant 2021J0499, and Grant 2021J0503.

ABSTRACT The dependent and independent variables in traditional linear regression models are continuous numerical variables. When the dependent variable or independent variable is a discrete variable, the traditional linear regression model can no longer be used to analyze. To solve this problem, this article introduces the non-homogeneous Markov chain model. It introduces the mathematical definition of the non-homogeneous Markov chain model. And then this article uses Bayesian estimation method to derive posterior distribution of model parameters. Through the MCMC algorithm, we simulate an experiment, posterior means value of the parameters is estimated, and the estimation effect is found to be better. Finally, we analyze the impact of learning state transition about college students on the non-homogeneous Markov chain model. Influencing factors include whether to receive a scholarship and whether to serve as a class leader. In this paper, non-homogeneous Markov chain model is used to analyze and detect the impact of discrete variables on dependent variables. This is the major innovation in this article.

INDEX TERMS Non-homogeneous Markov chain model, Bayesian estimation, MCMC algorithm, state transition.

I. INTRODUCTION

The correlation and influence relationship between variables is important research content of statistics. The linear regression model is the most traditional and classic statistical such as, the regression equation, parameter estimation, application range and other contents of the linear regression model introduced in detail in the literature [1]. However, the linear regression model requires that both the independent variable and the dependent variable are continuous numerical variables. Therefore, when the independent variable or the dependent variable is a discrete classified variable, it is impossible to establish a linear regression model to analyze the influence of the independent variable on the dependent variable. Many statistical models have been improved and developed to meet the needs of practical problems. For example, when the independent variable is a continuous numerical variable and the dependent variable is a discrete categorical variable 0 and 1, the logistic regression model is derived [2]. However, when the independent variables are discrete categorical variables or the independent variables and dependent variables are discrete categorical variables, most of them are analyzed by simple and rough descriptive statistical methods such as statistical charts and tables. At present, the better method for qualitative variable analysis is the contingency table test [3]. However, contingency table analysis can only test whether qualitative variables are related to each other, and cannot further explore who influences who and how. For example, if the university students are class cadres or not, and whether the award-winning grants affect their learning status, we cannot directly model and analyze them by using the above-mentioned methods. Under this background, this paper introduces and improves the hidden Markov model to analyze the interaction between discrete classification variables.

The hidden Markov model is a new statistical model. The model cannot directly observe the state sequence in practical application, that is, the state sequence of the model is hidden in the observation sequence, so it is called ‘‘hidden’’ Markov model [4]. In the nineties of last century, hidden

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

Markov models had achieved great success in the field of speech recognition. With the development of science and technology, hidden Markov models have a good application in the frontier fields of artificial intelligence, machine learning and so on. Therefore, it has received extensive attention and research. Song et al introduced latent variables into hidden Markov model and analyzed the model with Bayesian method [5]. And they studied the semiparametric hidden Markov model with latent variables by Bayesian method [6]. Xia *et al.* analyzed the hidden Markov factor analysis model with the semi parametric Bayesian method [7]. Liu *et al.* introduced linear regression model to hidden Markov model and used the maximum likelihood estimation method for inference analysis [8]. Wang et al proposed the hidden Markov structural equation model and estimated the corresponding parameters with Bayesian method [9]. On the basis of these studies, this paper improves the state transition mode of Markov chain and proposes a non-homogeneous Markov chain model, which is used to analyze the influencing factors of discrete classification variables.

The rest of this article is organized as follows: firstly, this paper introduces the mathematical definition of the non-homogeneous Markov chain model; secondly, it introduces the Bayesian inference method of the model; thirdly, it verifies the reliability of the estimation method through simulation experiments; finally, it uses this method to analyze the influencing factors of College Students' learning state transition.

II. NON-HOMOGENEOUS MARKOV CHAIN MODEL

Homogeneous Markov chain is a Markov process which indicates that time and state sequence is discrete. The transition process of homogeneous Markov state sequence satisfies the following conditions.

$$\begin{aligned} P(Z_{i,t} &= s|Z_{i,1}, Z_{i,2}, \ldots, Z_{i,t-1} = u) \\ &= P(Z_{i,t} = s|Z_{i,t-1} = u) = a_{us} \end{aligned} \tag{1}$$

where, $Z_{i,t}$ is the state of the ith observation object at the t-th time, a_{us} is the transition probability from state u at occasion $t-1$ to state S at occasion t In this transition mode, the current state is only related to the state of the previous moment, that is, it satisfies the homogeneous hypothesis. We call this state transition mode homogeneous Markov state transition [10].

However, in many state sequences, the state of the current observation time is not only related to the state of the previous time, but also related to some characteristics of the current time. We call this state transition non-homogeneous Markov state transition in the literature [11], this non-homogeneous Markov state transition as shown in formula 2.

$$\begin{aligned} P(Z_{i,t} &= s|Z_{i,1}, Z_{i,2}, \ldots, Z_{i,t-1} = u) \\ &= P(Z_{i,t} = s|Z_{i,t-1} = u) = p_{i,t,u,s} \\ i &= 1, 2, \ldots, N, \quad t = 2, 3, \ldots, T, \\ u &= 1, 2, \ldots, S, \quad s = 1, 2 \ldots, S \end{aligned} \tag{2}$$

where, $Z_{i,t}$ is the state of the ith observation object at the $S-th$ time, $Z_{i,t} = 1, 2, \ldots, S$, S is the total number of states in the model. $P_{i,t,u,s}$ is the transition probability from state $Z_{i,t-1}$ at occasion $t-1$ to state $Z_{i,t} = s$ at occasion t for the i th individual.

To model the transition probabilities, we assume that the hidden states $\{1, 2, \ldots, S\}$ are ordered. Instead of modeling $P_{i,t,u,s}$ directly, we consider a continuation-ratio logistic model for $\eta_{i,t,u,s} = P(Z_{i,t} = s|Z_{i,t} \geq s, Z_{i,t-1} = u)$ with $s \leq S-1$ [11]. So, we assume that.

$$\begin{aligned} \log(\eta_{i,t,u,s}) &= \log \frac{P(Z_{i,t} = s|Z_{i,t-1} = u)}{P(Z_{i,t} > s|Z_{i,t-1} = u)} \\ &= \log \frac{p_{i,t,u,s}}{p_{i,t,u,s+1} + \ldots + p_{i,t,u,S}} \\ &= \zeta_{u,s} + \boldsymbol{\alpha}'\boldsymbol{b}_{i,t}, \\ i &= 1, 2, \ldots, N, \quad t = 2, 3, \ldots, T, \\ u &= 1, 2, \ldots, S, \quad s = 1, 2 \ldots, S. \end{aligned} \tag{3}$$

where, $\zeta_{u,s}$ are the state transition parameters of the model $\boldsymbol{b}_{i,t}$ is the covariate vector that that may influence the transition probabilities, $\boldsymbol{\alpha}$ is an m-dimensional coefficient vector $\boldsymbol{b}_{i,t}$. Such parameterization only is intended to facilitate the interpretation of transition to a state. Based on (3), we can obtain the followings.

$$\begin{cases} p_{i,t,u,s} = \dfrac{\exp(\zeta_{u,s} + \boldsymbol{\alpha}'\boldsymbol{b}_{i,t})}{\prod_{k=1}^{s}\left\{1 + \exp(\zeta_{u,k} + \boldsymbol{\alpha}'\boldsymbol{b}_{i,t})\right\}} \\ p_{i,t,u,S} = \dfrac{1}{\prod_{k=1}^{S-1}\left\{1 + \exp(\zeta_{u,k} + \boldsymbol{\alpha}'\boldsymbol{b}_{i,t})\right\}} \end{cases}, \quad s = 1, 2, \ldots, S-1 \tag{4}$$

We assume that the initial time of the state sequence, that is, the probability of each initial state is: $(p_1, p_2, \ldots, p_S) = \boldsymbol{p}$. Then the whole state sequence can be determined by the initial probability $\boldsymbol{p}$ and the continuation-ratio logistic transfer model [6].

III. BAYESIAN INFERENCES

A. INFERENCE PRINCIPLE

The Bayesian inference problem of non-homogeneous Markov chain model is $\boldsymbol{P}(\boldsymbol{p}, \boldsymbol{\zeta}, \boldsymbol{\alpha}|\boldsymbol{Z})$. Where, $\boldsymbol{p}, \boldsymbol{\zeta}, \boldsymbol{\alpha}$ are the parameters to be estimated in the model. The posterior distribution of the Bayesian inference contains high-dimensional integral, so the calculation is very complex, and the specific form of the marginal posterior distribution of each parameter cannot be obtained directly. In this paper, we use MCMC algorithm to simulate the posterior distribution, and the posterior means of the parameters is used as the Bayesian estimation value of model parameters [12]. The specific execution process of MCMC algorithm is as follows.

(1) Update initial probability $\boldsymbol{p}$;
(2) Update the transition parameter $\boldsymbol{\zeta}$;
(3) Update the transfer coefficient vector $\boldsymbol{\alpha}$.

Each update step is based on the Gibbs sampling and MH algorithm [13]. We require setting prior distribution for each parameter and deriving posterior distribution.

B. PRIOR DISTRIBUTION OF PARAMETERS

In Bayesian theory, unknown parameters are regarded as random variables [14]. Before Bayesian inference, it is necessary to set prior distribution for each unknown parameter. Then, posterior distribution of parameters is obtained by using prior distribution and sample information. Based on existing research experience [15], we assume that the prior distribution for each parameter is as follows.

$$(p_1, p_2, \ldots, p_S) \sim Dirichlet(a, a, \ldots, a)$$
$$\zeta_{u,s} \sim (\zeta^0_{u,s}, \delta^2_\zeta)$$
$$\boldsymbol{\alpha} \sim N(\boldsymbol{\alpha}^0, \textstyle\sum_\alpha)$$

where, a, $\zeta^0_{u,s}$, δ^2_ζ, $\boldsymbol{\alpha}^0$, $\sum_\alpha$ is a super parameter in prior distribution. $\boldsymbol{Z}$ is the state in the model, and it is a known quantity.

C. POSTERIOR DISTRIBUTION OF PARAMETERS

Firstly, we give the likelihood function of the model, multiply the likelihood function by prior distribution of each parameter, and then derive full conditional posterior distribution of each parameter. The likelihood function is.

$$Lik = \prod_{u=1}^{S} \prod_{s=1}^{S} \prod_{i=1}^{N} \prod_{t=2}^{T} p_{i,t,u,s}^{I\{Z_{i,t-1}(u)Z_{i,t}(s)\}} \cdot \prod_{s=1}^{S} p_s^{n_s} \tag{5}$$

where, $I\left\{Z_{i,t-1}(u)Z_{i,t}(s)\right\}$ is a characteristic function, n_s is the number of samples on $Z_{i,1} = s$ [16].

(1) Full conditional distribution of $(p_1, p_2, \ldots, p_S) = \boldsymbol{p}$.

$$\boldsymbol{P}(\boldsymbol{p}|\boldsymbol{Z}, \boldsymbol{\zeta}, \boldsymbol{\alpha}) \propto \prod_{s=1}^{S} p_s^{a-1} \prod_{s=1}^{S} p_s^{n_s} = \prod_{s=1}^{S} p_s^{a+n_s-1}$$

So, the full conditional distribution of $\boldsymbol{p}$ is.

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{p}|\boldsymbol{Z}, \boldsymbol{\zeta}, \boldsymbol{\alpha}) &= (p_1, p_2, \ldots, p_S) \\ &= \boldsymbol{p} \sim Dirichlet(a + n_1, a + n_2, \ldots, a + n_s)\end{aligned} \tag{6}$$

(2) Full conditional distribution of $\zeta_{u,s}$.

$$\begin{aligned}P(\zeta_{u,s}\,|.) \propto\ & \frac{1}{\sqrt{2\pi\delta^2_\zeta}} \exp\left\{-\frac{(\zeta_{u,s} - \zeta^0_{u,s})^2}{2\delta^2_\zeta}\right\} \\ & \times \exp\left\{\log \prod_{k=s}^{S} \prod_{i=1}^{N} \prod_{t=2}^{T} p_{i,t,u,s}^{I\{Z_{i,t-1}(u)Z_{i,t}(k)\}}\right\} \\ \propto\ & \exp\left\{-\frac{(\zeta_{u,s} - \zeta^0_{u,s})^2}{2\delta^2_\zeta}\right\} \\ & \times \exp\left\{\sum_{k=s}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T} I\left\{Z_{i,t-1}(u)Z_{i,t}(k)\right\} \times \log(p_{i,t,u,s})\right\} \\ =\ & \exp\left\{\sum_{k=s}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T} I\left\{Z_{i,t-1}(u)Z_{i,t}(k)\right\} \times \log(p_{i,t,u,s}) - \frac{(\zeta_{u,s} - \zeta^0_{u,s})^2}{2\delta^2_\zeta}\right\}\end{aligned}$$

So, full conditional distribution of $\zeta_{u,s}$

$$P(\zeta_{u,s}\,|.) \propto \exp\left\{\sum_{k=s}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T} I\left\{Z_{i,t-1}(u)Z_{i,t}(k)\right\} \times \log(p_{i,t,u,s}) - \frac{(\zeta_{u,s} - \zeta^0_{u,s})^2}{2\delta^2_\zeta}\right\} \tag{7}$$

(3) Full conditional distribution of $\boldsymbol{\alpha}$.

$$\begin{aligned}& P(\boldsymbol{\alpha}\,|.) \\ & \propto \frac{1}{\sqrt{(2\pi)^m \left|\sum_\alpha\right|}} \\ & \quad \times \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)' \textstyle\sum_\alpha^{-1} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right)\right\} \\ & \quad \times \exp\left\{\log \prod_{u=1}^{S}\prod_{s=1}^{S}\prod_{i=1}^{N}\prod_{t=2}^{T} p_{i,t,u,s}^{I\{Z_{i,t-1}(u)Z_{i,t}(s)\}}\right\} \\ & \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)' \textstyle\sum_\alpha^{-1} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right)\right\} \\ & \quad \times \exp\left\{\sum_{u=1}^{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T} I\left\{Z_{i,t-1}(u)Z_{i,t}(s)\right\} \log(p_{i,t,u,s})\right\} \\ & = \exp\left\{\sum_{u=1}^{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T} I\left\{Z_{i,t-1}(u)Z_{i,t}(s)\right\} \log(p_{i,t,u,s}) - \tfrac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)' \textstyle\sum_\alpha^{-1} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right)\right\}\end{aligned}$$

So, full conditional distribution of $\boldsymbol{\alpha}$.

$$P(\boldsymbol{\alpha}\,|.) \propto \exp\left\{\sum_{u=1}^{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{t=2}^{T} I\left\{Z_{i,t-1}(u)Z_{i,t}(s)\right\} \log(p_{i,t,u,s}) - \tfrac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)' \textstyle\sum_\alpha^{-1} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right)\right\} \tag{8}$$

D. MCMC PROCESS

(1) Extract a vector from $Dirichlet(a+n_1, a+n_2, \ldots, a+n_s)$, and update the initial probability vector $\boldsymbol{p}$;

(2) Update the transition parameter $\zeta_{u,s}$ by MH algorithm according to $P(\zeta_{u,s}\,|.)$;

(3) Update the transfer coefficient vector $\boldsymbol{\alpha}$ by MH algorithm according to $P(\boldsymbol{\alpha}\,|.)$;

IV. EMPIRICAL SIMULATION

In order to test the Bayesian estimation method of the model parameters, it is necessary to set the real values of the parameters in the model in advance. Then, according to posterior distribution derived from Bayesian inference method, MCMC algorithm is used to simulate the parameters. Posterior mean value is taken as Bayesian estimation value of the parameter, and then compared with real value to observe the estimation effect, so as to measure the reliability of the Bayesian inference method.

We consider an HMM with the hidden state number $S = 2$. Where $\boldsymbol{b}_{i,t} = (b_{i,t,1}, b_{i,t,2})'$, in which $b_{i,t,1}$ and $b_{i,t,2}$ are independently generated from Bernoulli(0.5) [17];

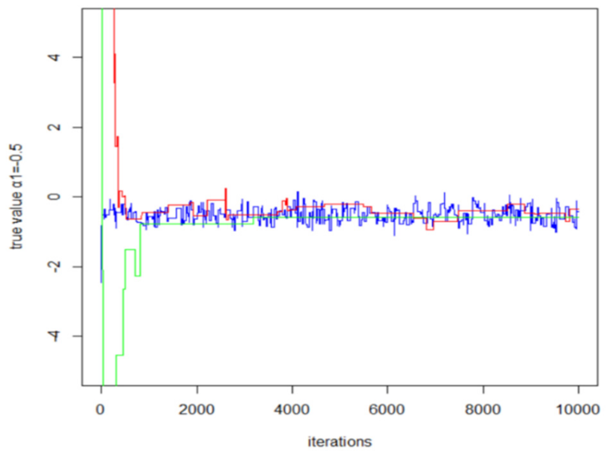


FIGURE 1. **Renderings of α_1 estimated for different initial values.**

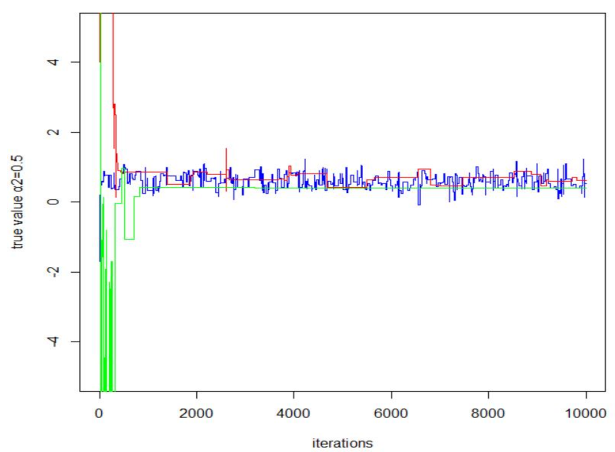


FIGURE 2. **Renderings of α_2 estimated for different initial values.**

$\boldsymbol{p} = (p_1, p_2)' = (0.5, 0.5)', \zeta_{1,1} = \zeta_{2,1} = -0.1$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (-0.5, 0.5)'$. The value of super parameter in prior is as follows: $a = 1, \zeta^0_{u,s} = 0, \delta^2_\zeta = 1.0$, $\boldsymbol{\alpha}^0 = (0, 0)', \sum_\alpha = \boldsymbol{I}$.

In the simulation, the number of sample observation objects is $n = 50$, and the number of observation times is $t = 10$. In MCMC algorithm, the total number of iterations is 10,000 times. For the parameters to be estimated, three groups of greatly different initial values are taken to explore the influence of different initial values on the convergence of the Markov chain. The experimental results are shown in Figure 1, Figure 2, Figure 3 and Figure 4.

Fig. 1 and Fig. 2 show Markov chains with transfer coefficients α_1 and α_2 at three different initial values. Fig. 3 and Fig. 4 respectively show the Markov chains of transfer parameters $\zeta_{1,1}$ and $\zeta_{2,1}$ under three different initial values. As can be seen from the above figures, under three groups of different initial values, the Markov chain of each parameter to be estimated all converges in a relatively short time. Therefore, different initial values have no effect on the convergence of the Markov chain [18]. In view of this, in MCMC algorithm, the total number of iterations is 10,000 times, and it is

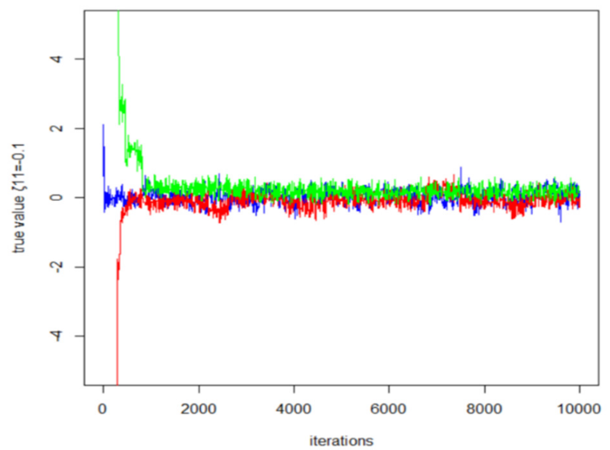


FIGURE 3. **Renderings of $\zeta_{1,1}$ estimated for different initial values.**

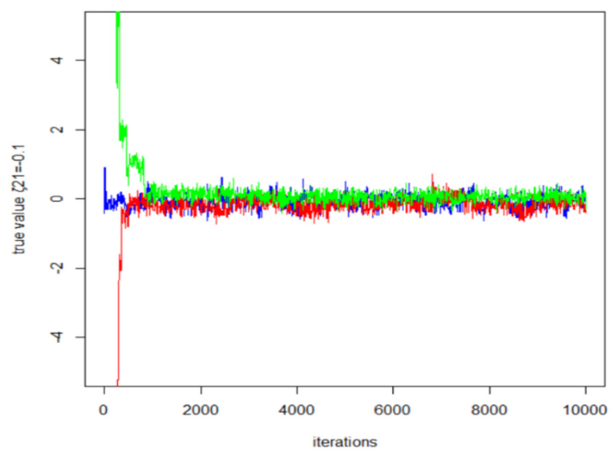


FIGURE 4. **Renderings of $\zeta_{2,1}$ estimated for different initial values.**

TABLE 1. **Parameter estimate results.**

parameter	Bias	Rmse
p_1	-0.033	0.075
p_2	0.033	0.075
$\zeta_{1,1}$	0.002	0.201
$\zeta_{2,1}$	0.029	0.169
α_1	-0.017	0.173
α_2	-0.005	0.173

reasonable to take the posterior mean of the last 5,000 times as Bayesian estimate value after removing the first 5,000 times.

In this paper, a total of 100 simulations are carried out, and the parameter estimate results in 100 experiments are calculated. The results are shown in Table 1.

In the simulation, the real values of each parameter are set in advance, and the state set is generated by simulation.

Then the parameters in the model are deduced by Bayesian estimation method, and the estimated results are compared with real values set in advance. The experimental results show that the Bias between the posterior mean and the real value of all parameters is within the acceptable range, which indicates that the inference method in this paper can estimate the parameters of the model correctly. The Rmse of the parameters shows that the estimation result of the inference method is stable.

V. STUDY ON THE TRANSFER OF COLLEGE STUDENTS' LEARNING STATE

In the study of College Students' learning state transition, the influencing factors are discrete categorical variables, which cannot be directly introduced into regression equation as continuous numerical variables and analyzed by linear regression model. In order to solve this problem, we use the non-homogeneous Markov chain model, regards whether to be a class leader or not and whether to receive scholarships as the covariates in the transfer model, to study their influence on the transfer of College Students' learning state.

A. DATA SELECTION

The influencing factor of College Students' learning state is an important research issue in pedagogy, and many scholars have conducted relevant research. For example, Xing Wenya used factor analysis model and logistic model to analyze and study the influencing factors of College Students' learning state [19]. Based on the curriculum reform; Lu Fang and others analyzed and evaluated the learning status of college students in the context of big data [20].

This paper selects 49 students from class 20171111 of Qujing Normal University for six semesters' academic performance, whether they are awarded grants and whether they are class leaders. There are two kinds of learning states: qualified and unqualified. If the score of all courses in a semester is above 65, it is regarded as "qualified"; if the score of one or more courses is below 65, it is regarded as "unqualified"; the "qualified" is recorded as state 1, and the "unqualified" is recorded as state 2. Finally, the state sequence matrix of 49 × 6 is obtained. At the same time, two co variables are selected: whether to win the scholarship or not and whether to serve as the class cadres. The scholarship for a semester is recorded as: 1, the non-award-winning scholarship is recorded as: 0, the number of classes cadres in a semester is:1 and that of the non-class cadres is: 0. The final observation data is a three-dimensional array of variables.

B. RESULTS ANALYSIS

The results of the example are shown in Table 2 and Table 3.

The results of the parameters in Table 2 are brought into equation (6) to obtain the transition probability values in Table 3. Furthermore, the state transition probability matrix is obtained when two covariates are 0 or 1.

Transfer probability of learning status of students who have not won the scholarships or grants and did not serve as a class leader.

TABLE 2. Bayesian estimation results of each parameter of the example.

parameter	Estimate	SD
$\zeta_{1,1}$	1.442	0.257
$\zeta_{2,1}$	-0.271	0.235
α_1	1.992	0.534
α_2	-0.218	0.447

TABLE 3. Transition probability.

probability	result	probability	result
$P_{i,\ t,1,1(0,0)}$	0.809	$P_{i,\ t,1,2(0,0)}$	0.191
$P_{i,\ t,2,1(0,0)}$	0.433	$P_{i,\ t,2,2(0,0)}$	0.567
$P_{i,\ t,1,1(0,1)}$	0.773	$P_{i,\ t,1,2(0,1)}$	0.227
$P_{i,\ t,2,1(0,1)}$	0.380	$P_{i,\ t,2,2(0,1)}$	0.620
$P_{i,\ t,1,1(1,0)}$	0.969	$P_{i,\ t,1,2(1,0)}$	0.031
$P_{i,\ t,2,1(1,0)}$	0.848	$P_{i,\ t,2,2(1,0)}$	0.152
$P_{i,\ t,1,1(1,1)}$	0.961	$P_{i,\ t,1,2(1,1)}$	0.039
$P_{i,\ t,2,1(1,1)}$	0.818	$P_{i,\ t,2,2(1,1)}$	0.182

$$P_{(F_{00})} = \begin{pmatrix} P_{i,t,1,1(0,0)} & P_{i,t,1,2(0,0)} \\ P_{i,t,2,1(0,0)} & P_{i,t,2,2(0,0)} \end{pmatrix} = \begin{pmatrix} 0.809 & 0.191 \\ 0.433 & 0.567 \end{pmatrix}$$

Transition probability of learning status of class leaders who have not won scholarships or grants.

$$P_{(F_{01})} = \begin{pmatrix} P_{i,t,1,1(0,1)} & P_{i,t,1,2(0,1)} \\ P_{i,t,2,1(0,1)} & P_{i,t,2,2(0,0)} \end{pmatrix} = \begin{pmatrix} 0.773 & 0.227 \\ 0.380 & 0.620 \end{pmatrix}$$

Transfer probability of learning status of students who won the scholarships or grants and did not serve as a class leader.

$$P_{(F_{10})} = \begin{pmatrix} P_{i,t,1,1(1,0)} & P_{i,t,1,2(1,0)} \\ P_{i,t,2,1(1,0)} & P_{i,t,2,2(1,0)} \end{pmatrix} = \begin{pmatrix} 0.969 & 0.031 \\ 0.848 & 0.152 \end{pmatrix}$$

Transition probability of learning status of class leaders who won scholarships or grants:

$$P_{(F_{11})} = \begin{pmatrix} P_{i,t,1,1(1,1)} & P_{i,t,1,2(1,1)} \\ P_{i,t,2,1(1,1)} & P_{i,t,2,2(1,1)} \end{pmatrix} = \begin{pmatrix} 0.961 & 0.039 \\ 0.818 & 0.182 \end{pmatrix}$$

where, $p_{i,t,u,s(m,n)}$ denotes the probability of the ith observation object changing from state u at time $t-1$ to state s at time t when the covariates are m and n respectively, $m = 0, 1$, $n = 0, 1, i = 1, 2, \ldots, N$, $t = 2, 3, \ldots, T$, $u = 1, 2, \ldots, S, s = 1, 2 \ldots, S$. Table 3 shows that, when students are not awarded scholarships or grants and do not serve as class leaders, the probabilities of students' learning status from qualified in the previous semester to qualified and unqualified in the next semester are 0.809 and 0.191 respectively; The probabilities of students' learning status from the unqualified in the previous semester to the qualified and unqualified in the next semester are 0.433 and 0.567 respectively; When the students were not awarded the scholarship and served as class leaders, the probabilities of the students' learning status from qualified in the previous semester to qualified and unqualified in the next semester are 0.773 and 0.227 respectively; the probabilities of the students' learning status from unqualified in the previous semester to qualified and unqualified in the next semester are 0.380 and 0.620 respectively; When the students won the scholarship and did not serve as class leaders, the probabilities of the students' learning status from qualified in the previous semester to qualified and unqualified in the next semester are 0.969 and 0.031, respectively; the probabilities of the students' learning status from unqualified in the previous semester to qualified and unqualified in the next semester are 0.848 and 0.152, respectively; When students are awarded the scholarship and serve as class cadres, the probability of students' learning status from the previous semester to the qualified and unqualified two states in the next semester is 0.961 and 0.031 respectively; the probability of students' learning status from the former semester to the latter semester is 0.818 and 0.182 respectively.

The results show that the probability of state transfer of whether the students are class cadres does not change significantly when they do not receive the scholarship. Similarly, the probability of state transfer of whether the students are class cadres does not change significantly when they are awarded the grant. It shows that whether to be a class leader or not has no significant effect on the transfer of students' learning state. When students are not class leaders, the probability of state transition from state 1 to state 1 increase significantly, and the probability of state transition from state 2 to state 1 decrease significantly. Similarly, when students are class leaders, the probability of state transition from state 1 to state 1 increases significantly, and the probability of state transition from state 2 to state 1 increases significantly The probability of state 1 is also significantly reduced, which indicates that whether or not the award-winning grant has a significant impact on the transfer of College Students' learning state.

In order to further explain the influence of covariates on state transition, this paper makes a Bayesian significance test on the covariate coefficient vector, and we made a significant test on the transfer coefficient α_1.

$$\left|\frac{1.992}{0.534}\right| = 3.73 > Z_{0.025} = 1.96$$

It shows that whether or not the scholarship has a significant impact on the transfer of students' learning status.

Test transfer coefficient α_2.

$$\left|\frac{-0.218}{0.447}\right| = 0.49 < Z_{0.025} = 1.96$$

It shows that whether to be a class leader or not has no significant effect on the transfer of students' learning state.

Through the comprehensive analysis, it is found that whether the scholarship has significant impact on the transfer of students' learning status. The findings of the study are positive and can be adjusted in determination of scholarships in order to increase students' motivation to learn.

VI. CONCLUSION

The linear regression model is the most classic statistical model to analyze the relationship between variables. However, the traditional linear regression model requires that the independent variables and dependent variables are continuous numerical variables. When the independent variables are classified discrete variables or the independent variables and dependent variables are classified discrete variables, the traditional linear regression model is no longer applicable. In this paper, a non-homogeneous Markov chain model was introduced to solve this problem. We introduced the non-homogeneous Markov chain model and its Bayesian inference method. Through the simulation, it is found that Bias and Rmse of the model parameter estimation are relatively small, which indicates that the parameter estimation results are correct and reliable. Finally, this paper analyzes the influence of whether college students are awarded grants and whether they are class leaders on the learning state transition by using the non-homogeneous Markov chain model, and draws the conclusion that whether they are awarded grants has a significant impact on the learning state of college students, which has a more positive practical significance.

The most important contribution of this paper is to analyze the influence of discrete categorical variables on dependent variables by using non-homogeneous Markov chain model, which cannot be described by traditional linear regression models. It provides a new method to study the influence of discrete categorical variables on dependent variables in the future.

REFERENCES

[1] Y.-Z. Li and C.-X. He, "Application and discrimination of linear regression model," South China Univ. Technol. Press, Guangzhou, China, Tech. Rep., 2016, pp. 113–154.

[2] B.-C. Yang, Y. Zhang, and W. Guo, "Research on enterprise financial risk diagnosis based on logistic regression model," *J. Syst. Sci.*, vol. 20, no. 3, pp. 29–35, 2021.

[3] T.-Y. Cao and M.-Q. Chen, "Independence test of two dimensional contingency table based on distance covariance," *Syst. Sci. Math.*, vol. 40, no. 9, pp. 56–60, 2019.

[4] H.-F. Liu, K. Wang, and C.-F. Jiang, "Bayesian inference of hidden Markov multivariate normal distribution with unknown number of hidden states," *Stat. Res.*, vol. 46, no. 34, pp. 124–134, 2017.

[5] X.-Y. Song, Y.-M. Xia, and H.-T. Zhu, "Hidden Markov latent variable models with multivariate longitudinal data," *Biometrics*, vol. 73, no. 1, pp. 221–243, 2017.

[6] X. Song, K. Kang, M. Ouyang, X. Jiang, and J. Cai, "Bayesian analysis of semiparametric hidden Markov models with latent variables," *Struct. Equation Model., A Multidisciplinary J.*, vol. 25, no. 1, pp. 1–20, Jan. 2018.

[7] Y.-M. Xia, J.-W. Gou, and Y.-A. Liu, "Semi parametric Bayesian analysis of hidden Markov factor analysis model," *J. Appl. Math.*, vol. 30, no. 1, pp. 78–86, 2015.

[8] H.-F. Liu, K. Wang, and Y. Li, "Hidden Markov linear regression modeland its parameter estimation," *IEEE Access*, vol. 8, pp. 187037–187042, 2020.

[9] K. Wang, H.-F. Liu, and C.-F. Jiang, "Hidden Markov structural equation model and its Bayesian estimation," *Math. Statist. Manage.*, vol. 37, no. 2, pp. 46–61, 2018.

[10] P. Tiawongsombat, M.-H. Jeong, A. Pirayawaraporn, J.-J. Lee, and J.-S. Yun, "Vision-based attentiveness determination using scalable HMM based on relevance theory," *Sensors*, vol. 19, no. 13, pp. 5331–5351, 2019.

[11] Y. Xu, "Maximum likelihood estimation of parameters of non-homogeneous hidden Markov model and its application," Capital Univ. Econ. Trade, Beijing, China, 2020, pp. 67–89.

[12] H.-F. Liu, X.-Y. Song, Y.-L. Tang, and B.-X. Zhang, "Bayesian quantile nonhomogenous hidden Markov models," *Stat. Methods Med. Res.*, vol. 30, no. 1, pp. 112–128, 2021.

[13] H.-F. Liu, X.-Y. Song, Y.-L. Tang, and B.-X. Zhang, "Bayesian analysis of mixture structural equation models with an unknown number of components," *Struct. Equation Model., A Multidisciplinary J.*, vol. 25, no. 12, pp. 41–55, Jan. 2018.

[14] H.-F. Liu, K. Wang, and F.-X. Zong, "Bayesian inference for multivariate normal mixed models with unknown number of mixtures," *Statist. Decis.*, vol. 29, no. 2, pp. 112–128, 2019.

[15] N. Metropolis, A.-W. Rosenbluth, and M.-N. Rosenbluth, "Equations of state calculations by fast computing machine," *J. Chem. Phys.*, vol. 89, no. 21, pp. 671–734, 1953.

[16] H. Liu and X. Song, "Bayesian analysis of hidden Markov structural equation models with an unknown number of hidden states," *Econometrics Statist.*, vol. 18, pp. 29–43, Apr. 2021.

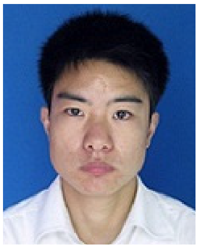
[17] Y.-Z. Jia, M.-Q. Xu, and R.-X. Wang, "Symbolic important point perceptually and hidden Markov model based hydraulic pump fault diagnosis method," *Sensors*, vol. 18, no. 13, pp. 4460–4479, 2018.

[18] W. Qi, G. Zong, and H. R. Karimi, "Sliding mode control for nonlinear stochastic singular semi-Markov jump systems," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 361–368, Jan. 2020.

[19] W.-Y. Xing and Q. Zhang, "Empirical analysis on the influence of multinomial logit model on college students' learning state," *J. Higher Educ.*, vol. 66, no. 22, pp. 89–103, 2015.

[20] F. Lu and L.-T. Wei, "Research methods of college students' learning state under the background of big data," *Heilongjiang Higher Educ. Res.*, vol. 37, no. 12, pp. 176–187, 2019.

SHEN XIUJUAN received the M.S. degree from the School of Probability Theory and Mathematical Statistics, Yunnan University, Kunming, China, in 2012. She is currently a Teaching Assistant with the School of Mathematics and Statistics, Qujing Normal University, China. Her research interests include mathematical statistics and data analysis.



LIU HEFEI received the M.S. degree from the School of Probability Theory and Mathematical Statistics, Yunnan University, Kunming, China, in 2011. He is currently an Assistant Professor with the School of Mathematics and Statistics, Qujing Normal University, China. His research interests include mathematical statistics and Bayesian inference.

LI YONG received the M.S. degree from the School of Computer Science, Wuyi University, Jiangmen, China, in 2010. He is currently an Assistant Professor with the School of Information Engineering, Qujing Normal University, China. His research interests include computer networks, information security, and cloud computing.

YUAN BIN received the M.S. degree from the School of Computer Technology, Yunnan University, Kunming, China, in 2009. He is currently a Lecturer with the School of Mathematics and Statistics, Qujing Normal University, China. His research interests include cloud computing and data mining.