Iranian Journal of Science & Technology, Transaction A, Vol. 30, No. A3
Printed in The Islamic Republic of Iran, 2006

"Research Note"

REGRESSION ANALYSIS IN MARKOV CHAIN*

A. Y. ALAMUTI AND M. R. MESHKANI**

Department of Statistics, Shahid Beheshti University, Evin, Tehran, 19838, I. R. of Iran
Email: ay.alamuti@gmail.com

Abstract – In a finite stationary Markov chain, transition probabilities may depend on some explanatory variables. A similar problem has been considered here. The corresponding posteriors are derived and inferences are done using these posteriors. Finally, the procedure is illustrated with a real example.

Keywords – Bayes, Empirical Bayes, estimation, Markov chain, maximum likelihood, regression, transition probability matrix

1. INTRODUCTION

We aim to estimate the transition probability matrix (tpm) of a discrete-time stationary Markov chain using explanatory variables. It will be done by methods of Maximum Likelihood (ML), Bayes (B) and Empirical Bayes (EB). First, we define some terms and give preliminary results in section 2. Then in section 3, 4 and 5 we provide the estimators of regression parameters by ML, B and EB methods, respectively. In section 6 we provide the estimates of tpm. Finally, an example is provided in section 7.

2. SOME PRELIMINARIES

Suppose $\{X_t, t \in \tau\}$ is a Markov chain with values in the finite state space S={1,...,s}, where $\tau = \{1,...,T\}$. We assume that the chain is simple; that is, its order of dependency is 1, and it has a stationary and irreducible tpm $\Lambda(\underline{z})$ with elements $0 \le \lambda_{jk}(\underline{z}) \le 1, j,k \in S$ and $\Sigma_{k=1}^{s}\lambda_{jk}(\underline{z}) = 1$, where $\underline{z}$ is the vector of explanatory variables.

The data are outcomes of n identically and independently repeated experiments called a "panel". In each experiment, we observe and record the states visited by the chain during a fixed period, $T > 1$. Let a realization of an experiment be $x_i = (x_1,...,x_T)$, i=1,...,n. The subscripts refer to the order in which the observations were taken and not to their values. For example, x_1 is a state that the chain chooses in $t = 1$. For each fixed $T > 1$, let the frequency count matrix (fcm) of the chain be denoted by $\mathbf{F} = (F_{jk})$, where F_{jk} is the number of times the event $\{X_{t-1} = j, X_t = k; \forall t \in T\}$ has occurred. The distribution of F under various assumptions has been discussed in Billard and Meshkani [1]. Here, we consider a simple case known as 'panel study'. We define

$$\underline{F}_j = [F_{j1},...,F_{js}],\ \ F_{j+} = \sum_{k\in S} F_{jk},\ \ F_{+k} = \sum_{j\in S} F_{jk}.$$

Given F_{j+}, each row of fcm has a multinomial distribution, $M(F_{j+}, \underline{\lambda}_j(\underline{z}))$ with $\underline{\lambda}_j(\underline{z}) = (\lambda_{j1}(\underline{z}),...,\underline{\lambda}_{js}(\underline{z}))$. From Anderson and Goodman [2], the pdf of jth row of fcm for a given $\underline{\lambda}_j(\underline{z})$ is

*Received by the editor May 15, 2005 and in final revised form January 27, 2007
**Corresponding author

$$P(\underline{F}_j \mid \underline{\lambda}_j) = \frac{F_j + !}{\prod_{k=1}^{s-1} F_{jk}!}[\prod_{k=1}^{s-1} \lambda_{jk}^{F_{jk}}(\underline{z})][1 - \sum_{k=1}^{s-1} \lambda_{jk}(\underline{z})]^{F_{js}}. \tag{1}$$

Let $\underline{z} = (z_0, ..., z_p), z_0 \equiv 1$, be a vector of explanatory variables with elements of tpm as

$$\lambda_{jk}(\underline{z}) = exp(-\underline{z}\underline{\beta}_{jk}) \tag{2}$$

where $\underline{\beta}_{jk} = (\beta_{jk0}, ..., \beta_{jkp})$.

The particular parametric form in (2), a log-linear model for tp, is chosen primarily for analytical convenience; other parameterizations may be more appropriate in particular applications. This model does have, however, the attractive feature of yielding nonnegative tp whenever $\underline{z}\underline{\beta}_{jk} > 0$, and has been suggested by several authors.

$$\lambda_{hj1} = exp(-\beta_{j10} - z_{h1}\beta_{j11} - z_{h2}\beta_{j12} - z_{h1}z_{h2}\beta_{j13}),\ j = 1, 2,\ h = 1, 2, 3, 4.$$

3. ESTIMATION OF REGRESSION PARAMETERS BY ML METHOD

Considering different levels of explanatory variables $\underline{z}_h$, we can categorize the observations into r distinct groups with fixed values of explanatory variables.

From (1) and (2), the likelihood function for the jth row, i.e., for $\underline{\beta}_j = [\underline{\beta}_{j1}, ..., \underline{\beta}_{j(s-1)}]$ and the group identified with $\underline{z}_h$, h=1,...,r is

$$L_j = [\prod_{k=1}^{s-1} exp(-F_{jk}\underline{z}_h\underline{\beta}_{jk})][1 - \sum_{k=1}^{s-1} exp(-\underline{z}_h\underline{\beta}_{jk})]^{F_{js}} \quad j, k = 1, ..., s \tag{3}$$

This function has to be maximized with respect to $\underline{\beta}_j$, subject to $\underline{z}_h\underline{\beta}_{jk} > 0$, h=1,...,r. We can see that the normal equations are not solvable analytically. Thus (3) is maximized in $\underline{\beta}_j$ by numerical methods to provide the MLE $\underline{\beta}_{j,ML}$ whose covariance matrix can be estimated from the Fisher information matrix, Alamuti and Meshkani [3].

We note that (3) is the likelihood for a member of the multiparameter exponential family. Thus, we can claim that estimates extracted from this likelihood share the good properties of MLE such as asymptotic normality and consistency, Cox and Hinkley [4].

4. ESTIMATION OF REGRESSION PARAMETERS BY BAYES METHOD

The Bayes estimator relative to the squared error loss function is the posterior expectation of the $\underline{\beta}_j$, Meshkani and Billard [5].

The posterior function of $\underline{\beta}_j$ relative to the prior $\pi(\underline{\beta}_j)$ will be

$$\pi(\underline{\beta}_j \mid \underline{z}_h, \underline{F}_j) = \frac{L(\underline{F}_j \mid \underline{\beta}_j, \underline{z}_h)\pi(\underline{\beta}_j)}{m(\underline{F}_j)} \tag{4}$$

where

$$m(\underline{F}_j) = \int L(\underline{F}_j \mid \underline{\beta}_j, \underline{z}_h)\pi(\underline{\beta}_j)d\underline{\beta}_j.$$

We assume exponential prior distribution. First, we suppose the prior distributions of regression parameters are independent exponentials, i.e., for j=1,...,s; k=1,...,s-1 and m=0,...,p,

$$\pi(\beta_{jkm}) = \gamma_{jkm} exp(-\gamma_{jkm}\beta_{jkm}), \qquad \gamma_{jkm}, \beta_{jkm} > 0 \tag{5}$$

and

$$\pi(\underline{\beta}_j) = \prod_{k=1}^{s-1}\prod_{m=0}^{p} \gamma_{jkm} exp(-\gamma_{jkm}\beta_{jkm}) = \eta_j e^{-\sum_{k=1}^{s-1}\underline{\gamma}_{jk}\underline{\beta}_{jk}}. \tag{6}$$

where

$$\eta_j = \prod_{k=1}^{s-1}\prod_{m=0}^{p} \gamma_{jkm} \; and \; \underline{\gamma}_{jk} = [\gamma_{jk0},...,\gamma_{jkp}]. \tag{7}$$

The merits of this prior are its mathematical convenience, properness, and having short tails, thus being more inline with realistic situations.
We use the polynomial expansion and transform $(1-e^{-\underline{z}_h\underline{\beta}_{j1}}-...-e^{-\underline{z}_h\underline{\beta}_{j(s-1)}})^{F_{js}}$ into the following summation:

$$(1-e^{-\underline{z}_h\underline{\beta}_{j1}}-...-e^{-\underline{z}_h\underline{\beta}_{j(s-1)}})^{F_{js}} = \sum_C \binom{F_{js}}{\alpha_o,\alpha_1,..,\alpha_{s-1}}(-1)^{F_{js}-\alpha_0} e^{-\sum_{k=1}^{s-1}\alpha_k \underline{z}_h\underline{\beta}_{jk}}$$

where

$$C = \{(\alpha_o,\alpha_1,...,\alpha_{s-1}) : 0 \le \alpha_k \le F_{js}, \sum_{k=0}^{s-1}\alpha_k = F_{js}\}.$$

Also the regression parameters are assumed to be independent, leading to (s-1)(p+1) separate integrals. Thus $m(\underline{F}_j)$ is computed as

$$m(\underline{F}_j) = \eta_j \sum_C \binom{F_{js}}{\alpha_o,\alpha_1,..,\alpha_{s-1}}(-1)^{F_{js}-\alpha_0}\prod_{k=1}^{s-1}\prod_{m=0}^{p}\frac{1}{[(F_{jk}+\alpha_k)z_{hm}+\gamma_{jkm}]}. \tag{8}$$

Thus the marginal posterior density of β_{jkl} is obtained as

$$\pi(\beta_{jkl} \mid \underline{F}_j, \underline{z}_h) = \sum_C B_j \frac{exp(-[(F_{jk}+\alpha_k)z_{hl}+\gamma_{jkl}]\beta_{jkl})}{\prod_k^{*}\prod_m^{*}[(F_{jk}+\alpha_k)z_{hm}+\gamma_{jkm}]} \tag{9}$$

where

$$B_j = \frac{1}{m(\underline{F}_j)}\eta_j \binom{F_{js}}{\alpha_o,\alpha_1,..,\alpha_{s-1}}(-1)^{F_{js}-\alpha_0}. \tag{10}$$

The $*$'s in products exclude kl combination. Thus, the Bayes estimate of β_{jkl} respective to the exponential prior distribution for the h th group is given by

$$\beta_{jkl;B} = E(\beta_{jkl} \mid \underline{F}_j, \underline{z}_h)$$

$$= \sum_C \frac{B_j}{[(F_{jk}+\alpha_k)z_{hl}+\gamma_{jkl}]\prod_{k=1}^{s-1}\prod_{m=0}^{p}[(F_{jk}+\alpha_k)z_{hm}+\gamma_{jkm}]}. \tag{11}$$

The posterior variance is obtained by subtraction of the square of $\underline{\beta}_{jkl;B}$ from

$$E(\beta_{jkl;B}^2 \mid \underline{F}_j, \underline{z}_h) = \sum_C \frac{2B_j}{[(F_{jk} + \alpha_k)z_{hl} + \gamma_{jkl}]^2 \prod_{k=1}^{s-1} \prod_{m=0}^{p} [(F_{jk} + \alpha_k)z_{hm} + \gamma_{jkm}]}. \tag{12}$$

5. ESTIMATION OF REGRESSION PARAMETERS BY EMPIRICAL BAYES METHOD

In the empirical Bayes approach, one needs to estimate the hyper-parameters from the data itself and follow the Bayes rule, using the estimated prior. To this end, we maximize $m(\underline{F}_j)$ with respect to the hyper-parameters and obtain their ML estimates Suppose numerical maximization of $m(\underline{F}_j)$ leads to $\hat{\gamma}_{jkm}$; by replacing $\hat{\gamma}_{jkm}$'s in (11) and (12) we obtain the empirical Bayes estimates.

6. ESTIMATION OF TRANSITION PROBABILITY MATRIX

Let the regression parameter estimates for (j,k) th entry of tpm be estimated by one of the three methods. Denote this estimate by $\hat{\underline{\beta}}_{jk}$. Replacing $\underline{\beta}_{jk}$ in (2) we obtain an estimate of tpm

$$\hat{\lambda}_{jk} = exp(-\underline{z}_h \hat{\underline{\beta}}_{jk}). \tag{13}$$

The variance of $\lambda_{jk;B}$ is estimated by

$$var(\hat{\lambda}_{jk}) = h'(\hat{\underline{\beta}}_{jk}) var(\hat{\underline{\beta}}_{jk}) h(\hat{\underline{\beta}}_{jk}) \tag{14}$$

where

$$h(\hat{\underline{\beta}}_{jk}) = \left[\frac{\partial \lambda_{jk}}{\partial \beta_{jk0}}, ..., \frac{\partial \lambda_{jk}}{\partial \beta_{jkp}}\right]_{\underline{\beta}_{jk} = \hat{\underline{\beta}}_{jk}}$$

and

$$var(\hat{\underline{\beta}}_{jk}) = diag\{var(\beta_{jkp}), ..., var(\beta_{jkp})\}.$$

7. AN EXAMPLE

We present an example having data we have gathered ourselves. We would like to test the effects of gender and place of residence on students' academic performance.

To test these hypotheses, the educational records of a sample of 38 students during 6 consecutive academic semesters are examined. At the end of each semester, each student's performance is classified as *unsatisfactory* or *satisfactory*. We model the students' status as a 2-state Markov chain. We use the following notations:

State space: S = $\{1, 2\}$, $1 \equiv unsatisfactory$, $2 \equiv satisfactory.$

Explanatory variables: Z_1 = place of residence and Z_2 = gender.

$$Z_1 = \begin{cases} 3 & dormitory \\ 2 & o.w. \end{cases}, \quad Z_2 = \begin{cases} 1 & female \\ 0 & male \end{cases}$$

Thus, there are 4 groups corresponding to $(z_1, z_2) = \{(3,1), (3,0), (2,1), (2,0)\}$ numbered by

$$h=1,\ldots,4, \quad z_h = [1, z_{h1}, z_{h2}, z_{h1}z_{h2}],$$

with

$$z_1 = [1,3,1,3],\ z_2 = [1,3,0,0],\ z_3 = [1,2,1,2],\ z_4 = [1,2,0,0].$$

Their total fcm are

$$F_1 = \begin{bmatrix} 0 & 2 \\ 2 & 1 \end{bmatrix}, F_2 = \begin{bmatrix} 2 & 5 \\ 4 & 19 \end{bmatrix}, F_3 = \begin{bmatrix} 2 & 7 \\ 5 & 76 \end{bmatrix}, F_4 = \begin{bmatrix} 4 & 9 \\ 9 & 43 \end{bmatrix}.$$

Here, we have assumed a stationary Markov chain for each group whose tpm is a realization from a prior distribution. Before embarking on the estimation procedure, we check the underlying assumptions by tests suggested in [2], [6] and [7].

Having confirmed the required assumptions, we now apply our proposed method to estimate the regression coefficients and tpm.

One drawback of the ML method is that it gives unreasonable estimates when the sample size is small. We could see that by the EB method, we have reasonable estimates with good precision, even for very small sample sizes.

Although 6 periods of observations for a Markov chain are not large enough to reach equilibrium, we use equation $\underline{\pi}\lambda = \underline{\pi}$ to find $\underline{\pi}$, the stationary probability vector. This determines what percentage of students will eventually be in state 1 or 2, (see Table 1). This fact can be very useful for administrators.

Table 1. The equilibrium distribution for 4 groups

h	π_1	π_2
(3,1)	0.33	0.67
(3,0)	0.22	0.78
(2,1)	0.16	0.84
(2,0)	0.22	0.78

For example, in the long run, 33 percent of female students living in dormitories will be on probation which is quite alarming.

To compare the rates of dwelling of various groups *on probation* for 2 consecutive semesters, we compute the transition probability ratios:

$$\frac{\lambda_{11}(3,1)}{\lambda_{11}(3,0)} \simeq 1.08, \frac{\lambda_{11}(2,1)}{\lambda_{11}(2,0)} \simeq 2.90$$

These odds show that the adverse effect of living out of the dormitory is almost thrice for females than males.

To choose the best fitting model, we compared the following models using posterior odds [8], which led to

$$\lambda_{hj1} = exp(-\beta_{j10} - z_{h1}\beta_{j11} - z_{h2}\beta_{j12} - z_{h1}z_{h2}\beta_{j13}),\ j = 1,2,\ h = 1,2,3,4.$$

REFERENCES

1. Billard, L. & Meshkani, M. R. (1992). *Empirical Bayes estimation for a finite Markov chain. Biometrika, 79,*

185-193.

2. Anderson, T. W. & Goodman, L. A. (1957). Statistical Inference about Markov chains. *Annal: of Statist, 28,* 89-110.
3. Alamuti, A. Y. & Meshkani, M. R. (2005). Exponential Regression for Odds in a Stationary Markov Chain. Tentatively accepted by Metrika.
4. Cox, D. R. & Hinkley, D. (1974). *Theoretical statistics.* London, Chapman and Hall.
5. Meshkani, M. R. & Billard, L. (1995). Estimation of a stationary Markov chain. *J. Amer. statist. Assoc, 90*, 234-243.
6. Berger, J. O. (1985). *Statistical Desition Theory and Bayesian Analysis.* NewYork, Spiringer Verlag.
7. Schwarz, G. (1978) Estimating the dimension of a model. *Ann. of statist, 6*, 461-464.
8. Robert, C. P. (2001). *The Bayesian choice.* New York, Springer.