A Markov Chain Model for Unskilled Workers and the Highly Mobile

Author(s): Michael Sampson
Source: *Journal of the American Statistical Association*, Vol. 85, No. 409 (Mar., 1990), pp. 177-180
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2289541
Accessed: 09-03-2015 09:18 UTC

A Markov Chain Model for Unskilled Workers and the Highly Mobile

MICHAEL SAMPSON*

A parametric class of Markov chains that can be used to model such phenomena as job transitions for unskilled workers is presented in this article. Maximum likelihood estimation for the Markov chain model and the mover–stayer model is discussed and is shown to be easy to carry out. An empirical example, using industry-of-occupation data, is presented at the end of the article.

KEY WORDS: Maximum likelihood estimation; Mover–stayer model.

1. INTRODUCTION

Markov chains, or mixtures of Markov chains such as the mover–stayer model, are commonly used in the social sciences to model various forms of dynamic behavior such as occupational mobility, consumers' brand preferences, geographic migration, and income dynamics (for examples and references, see Frydman 1984; Geweke, Marshall, and Zarkin 1986; Singer and Spilerman 1976, 1977).

Consider a discrete time finite-state Markov chain with an $s \times s$ transition matrix $Q = [p_{ij}]$, where p_{ij} is the probability that the stochastic process is in state j this period, given that it was in state i in the previous period. Let ι be an $s \times 1$ vector of ones, $p^* = [p_j^*]$ be a $1 \times s$ vector of probabilities (so that $p^*\iota = 1$), and $\theta = \text{diag}[\theta_i]$ be an $s \times s$ diagonal matrix with $0 \leq \theta_i \leq 1$. The class of transition matrices I examine is

$$Q = \theta + (I - \theta)\iota p^* \quad \text{or} \quad p_{ij} = \theta_i\delta_{ij} + (1 - \theta_i)p_j^*, \quad (1)$$

where I is the identity matrix and δ_{ij} is the Kronecker delta.

To get some idea where such transistion probabilities could occur, consider the case where the states are job categories and workers are unskilled. By unskilled I mean that if such a worker were to go on the labor market, then he would have no characteristics (i.e., skills) that would distinguish him from other unskilled workers (alternatively, what skills he has are specific to his current job). Suppose that such a worker who is in job i has a probability θ_i of keeping his job and a probability $(1 - \theta_i)$ of being laid off. Since the worker is unskilled, if he is laid off his previous job category plays no role in determining his next job category. Therefore, a laid-off worker will have probability p_j^* (which is independent of i) of being in job category j, and p_{ij} takes the form in (1).

This kind of interpretation can be extended to other situations where Markov chains are used, and as such it can form a meaningful hypothesis about the nature of the underlying behavior. For example, $1 - \theta_i$ could represent the probability that a consumer would decide to look for a new product, or that a son would look for a different job from his father, or that an individual would look for another geographic area to live in; once this decision is made, the previous product, or the father's job, or the old area, would play no role in determining where the agent will be in the future. In other words, the past history of individuals who have decided to (or are forced to) move has no influence on where they will end up in the next period.

If all of the θ_i's are identical, one then obtains the persistence class of Markov chains considered by Barton, David, and Fix (1962) and Goodman (1964). (If they are all equal to 0 the process exhibits intertemporal independence.) The parameter θ_i determines the extent to which being in state i influences the next period's state. If $\theta_i = 0$, then state i has no influence, whereas if $\theta_i = 1$, then $p_{ii} = 1$, and i is an absorbing state (thus an agent in that state remains there forever).

Let the $1 \times s$ probability vector $p = [p_j]$ be the equilibrium distribution of Q so that $pQ = p$ (p determines the proportion of visits to each state over the long run). In general, $p^* \neq p$, except when the θ_i's are identical. Nevertheless, premultiplying (1) by p, one finds

$$p^* = p(I - \theta)/p(I - \theta)\iota \quad \text{or} \quad p_j^* = p_j(1 - \theta_j)\Big/\sum_k p_k(1 - \theta_k), \quad (2)$$

so

$$Q = \theta + \frac{(I - \theta)\iota p(I - \theta)}{p(I - \theta)\iota} \quad \text{or} \quad p_{ij} = \delta_{ij}\theta_i + \frac{(1 - \theta_i)(1 - \theta_j)p_j}{\sum_k p_k(1 - \theta_k)}. \quad (3)$$

Therefore, from p and θ one can determine p^*. [The representation in (3) may be useful if one requires a parametric form that includes the equilibrium distribution as a direct argument.] From p^* and θ one can determine p to be

$$p = p^*(I - \theta)^{-1}/p^*(I - \theta)^{-1}\iota \quad \text{or} \quad p_j = p_j^*(1 - \theta_j)^{-1}\Big/\sum_k p_k^*(1 - \theta_k)^{-1}, \quad (4)$$

which can be verified from (1). The vectors p and ι are

* Michael Sampson is Assistant Professor, Department of Economics, Concordia University, Montreal, Quebec, H3W 1M8, Canada. The author thanks Allan Gregory, an associate editor, and a referee for helpful comments.

Journal of the American Statistical Association
March 1990, Vol. 85, No. 409, Theory and Methods

the left and right eigenvectors associated with the eigenvalue 1. In Sampson (1987) it is shown that the remaining eigenvalues are the roots of the function $f(\lambda) = p^*(\theta - \lambda I)^{-1}\iota$, where λ is a scaler, with only one eigenvalue lying between θ_i and θ_{i+1} when $\theta_1 \le \theta_2 \le \cdots \le \theta_s$. The left and right eigenvectors corresponding to the eigenvalue λ are then given by $p^*(\lambda I - \theta)^{-1}$ and $(\lambda I - \theta)^{-1}(I - \theta)\iota$, respectively.

In the remainder of this article I deal with the problem of maximum likelihood (ML) estimation of p^* and θ. Section 2 addresses ML estimation for the Markov chain model, and Section 3 covers ML estimation for the mover–stayer model. An empirical example is considered in Section 4.

2. MAXIMUM LIKELIHOOD ESTIMATION FOR THE MARKOV CHAIN MODEL

I now consider ML estimation for the Markov chain model, given N independent realizations of length $T + 1$ periods for $t = 0, 1, \ldots, T$. The log-likelihood is given by

$$l = \sum_{ij} n_{ij} \ln(p_{ij}), \qquad (5)$$

where n_{ij} is the number of transitions in the sample from state i to state j. The unrestricted ML estimates are then $\hat{p}_{ij} = n_{ij}/n_i$, where n_i is the number of occurrences of i for the periods $t = 0, 1, \ldots, T - 1$. Let $\tilde{n}_i$ be the number of occurrences of i for $t = 1, 2, \ldots, T$, and let $n = N \times T$. Note that $n_i = \sum_k n_{ik}$, $\tilde{n}_i = \sum_k n_{ki}$, and $n = \sum_k n_k = \sum_k \tilde{n}_k$, so the $1 \times s$ probability vectors $\bar{p} = [n_i/n]$ and $\tilde{p} = [\tilde{n}_i/n]$ satisfy

$$\tilde{p} = \bar{p}\hat{Q}_u, \qquad (6)$$

where $\hat{Q}_u = [\hat{p}_{ij}]$ is the unrestricted ML transition matrix.

Substituting (1) into (5) gives the restricted likelihood

$$l = \sum_i \{n_{ii} \ln(\theta_i + (1 - \theta_i)p_i^*) + (\tilde{n}_i - n_{ii})\ln(p_i^*) + (n_i - n_{ii})\ln(1 - \theta_i)\}. \qquad (7)$$

Maximizing l subject to $p^*\iota = 1$ with Lagrange multiplier λ results in the first-order conditions

$$\frac{\partial L}{\partial \theta_i} = 0 = \frac{n_{ii}(1 - \hat{p}_i^*)}{\hat{\theta}_i + (1 - \hat{\theta}_i)\hat{p}_i^*} - \frac{(n_i - n_{ii})}{(1 - \hat{\theta}_i)}, \quad i = 1, 2, \ldots, s, \qquad (8)$$

and

$$\frac{\partial L}{\partial p_i^*} = 0 = \frac{n_{ii}(1 - \hat{\theta}_i)}{\hat{\theta}_i + (1 - \hat{\theta}_i)\hat{p}_i^*} + \frac{(\tilde{n}_i - n_{ii})}{\hat{p}_i^*} - \hat{\lambda}, \quad i = 1, 2, \ldots, s. \qquad (9)$$

From (8), it follows that

$$\hat{\theta}_i + (1 - \hat{\theta}_i)\hat{p}_i^* = n_{ii}/n_i = \hat{p}_{ii}, \qquad (10)$$

so $\hat{Q}_r = \hat{\theta} + (I - \hat{\theta})\iota\hat{p}^*$, the restricted ML transition matrix, has the same diagonal elements as $\hat{Q}_u$, the unrestricted ML transition matrix. Combining (8) and (10) and using $\hat{p}^*\iota = 1$ yields $\hat{\lambda}/n = \bar{p}(I - \hat{\theta})\iota$, so

$$\tilde{p} = \bar{p}(\hat{\theta} + (I - \hat{\theta})\iota\hat{p}^*) = \bar{p}\hat{Q}_r; \qquad (11)$$

hence both $\hat{Q}_r$ and $\hat{Q}_u$ share the property of transforming $\bar{p}$ into $\tilde{p}$ [compare (6) and (11)].

The ML estimates are easy to compute. From (10) and (11), it follows that $\hat{\theta}$ will be a root of $F_i(\theta)$, where

$$F_i(\theta) = \theta_i + \frac{(1 - \theta_i)(\tilde{p}_i - \bar{p}_i\theta_i)}{\bar{p}(I - \theta)\iota} - \frac{n_{ii}}{n_i}, \quad i = 1, 2, \ldots, s, \qquad (12)$$

which must be solved iteratively for $\hat{\theta}$. Newton's method is easy to implement and takes the form

$$\hat{\theta}_{it} = \hat{\theta}_{it-1} - F_i a_i + \frac{a_i \sum_k \bar{p}_k F_k a_k b_k}{1 + \sum_k \bar{p}_k a_k b_k}, \qquad (13)$$

where $a_i(\theta)$ and $b_i(\theta)$ are

$$a_i(\theta) = (1 + (2\theta_i\bar{p}_i - \tilde{p}_i - \bar{p}_i)/\bar{p}(I - \theta)\iota)^{-1}$$

$$b_i(\theta) = (1 - \theta_i)(\tilde{p}_i - \bar{p}_i\theta_i)/(\bar{p}(I - \theta)\iota)^2, \qquad (14)$$

and where all functions are evaluated at $\hat{\theta}_{t-1}$. Once $\hat{\theta}_i$ is calculated, $\hat{p}_i^*$ is given by $\hat{p}_i^* = 1 - (1 - n_{ii}/n_i)/(1 - \hat{\theta}_i)$. The ML estimate of $\hat{p}$, the equilibrium distribution, can be calculated from (4).

3. MAXIMUM LIKELIHOOD ESTIMATION FOR THE MOVER–STAYER MODEL

The mover–stayer model was first introduced by Blumen, Kogan, and McCarthy (1955) as a generalization of the Markov chain model. The basic idea is that there are two populations in the sample: stayers, who always remain in their initial state, and movers, whose state transitions are governed by a Markov chain process. Thus if s_i is the proportion of stayers in state i, then the probability of moving from state i to j in m periods is $s_i\delta_{ij} + (1 - s_i)p_{ij}(m)$, where $p_{ij}(m)$ is the i, j element of Q^m. Frydman (1984) treats ML estimation when Q is unrestricted.

I consider restricted ML estimation when $Q = \theta + (I - \theta)\iota p^*$. This makes for an appealing hypothesis, since with occupational mobility, for example, it is not unreasonable to suppose that workers are stayers because they have job-specific skills, whereas workers who are movers would not have these skills, and their transition probabilities would have the form $\theta + (I - \theta)\iota p^*$.

Frydman (1984) showed that the log-likelihood is given by

$$l = \sum_i [r_i \ln(s_i + (1 - s_i)p_{ii}^T) + (n_i(0) - r_i)\ln(1 - s_i) + (n_i - Tr_i)\ln(p_{ii}) + \sum_{k\neq i} n_{ik}\ln(p_{ik})], \qquad (15)$$

where r_i is the number of individuals who remain in state i for $t = 0, 1, \ldots, T$, and $n_i(0)$ is the number of individuals in state i at $t = 0$. Frydman (1984) shows that the ML estimate of s_i is

$$1 - \hat{s}_i = (1 - r_i/n_i(0))/(1 - \hat{p}_{ii}^T), \quad (16)$$

where $\hat{p}_{ii}$ is the ML estimate of p_{ii}. Using (1) and (16) in (15), we obtain, apart from some constants, the concentrated likelihood:

$$l = \sum_i [-(n_i(0) - r_i)\ln(1 - p_{ii}^T) + (n_{ii} - Tr_i)\ln(p_{ii}) + (\tilde{n}_i - n_{ii})\ln(p_i^*) + (n_i - n_{ii})\ln(1 - \theta_i)], \quad (17)$$

where I write p_{ii} for $\theta_i + (1 - \theta_i)p_i^*$. Maximizing l subject to $p^*\iota = 1$ with Lagrange multiplier λ then yields

$$\frac{\partial L}{\partial \theta_i} = 0 = \frac{(n_i(0) - r_i)T\hat{p}_{ii}^{T-1}(1 - \hat{p}_i^*)}{(1 - \hat{p}_{ii}^T)} + \frac{(n_{ii} - Tr_i)(1 - \hat{p}_i^*)}{\hat{p}_{ii}} - \frac{(n_i - n_{ii})}{(1 - \hat{\theta}_i)}, \quad i = 1, 2, \ldots, s, \quad (18)$$

and

$$\frac{\partial L}{\partial p_i^*} = 0 = \frac{(n_i(0) - r_i)T\hat{p}_{ii}^{T-1}(1 - \hat{\theta}_i)}{(1 - \hat{p}_{ii}^T)} + \frac{(n_{ii} - Tr_i)(1 - \hat{\theta}_i)}{\hat{p}_{ii}} + \frac{(\tilde{n}_i - n_{ii})}{\hat{p}_i^*} - \hat{\lambda}, \quad i = 1, 2, \ldots, s. \quad (19)$$

From (18), using $(1 - \hat{p}_i^*)(1 - \hat{\theta}_i) = 1 - \hat{p}_{ii}$ and some algebra gives

$$(n_i - Tn_i(0))\hat{p}_{ii}^{T+1} + (Tn_i(0) - n_{ii})\hat{p}_{ii}^T + (Tr_i - n_i)\hat{p}_{ii} + (n_{ii} - Tr_i) = 0. \quad (20)$$

This is identical to the polynomial that determines $\hat{p}_{ii}$ for the unrestricted mover–stayer model, however. [See Frydman (1984), eq. (4), where it is shown that (20) has a unique solution in the (0, 1) interval.] Thus, just as with the Markov chain model, the diagonal elements of the ML restricted and unrestricted transition matrices are identical. In addition, from (16) the ML estimates of s_i are identical for the restricted and unrestricted models.

Now, from (18) and (19) it follows that

$$\frac{(n_i - n_{ii})}{(1 - \hat{p}_i^*)} + \frac{(\tilde{n}_i - n_{ii})}{\hat{p}_i^*} = \hat{\lambda}, \quad i = 1, 2, \ldots, s, \quad (21)$$

which must be solved iteratively for $\hat{p}^*$. Newton's method is easy to implement and takes the form

$$\hat{p}_{it}^* = \hat{p}_{it-1}^* + a_i\left(\sum_k a_k b_k \Big/ \sum_k a_k\right) - a_i b_i, \quad i = 1, \ldots, 1, 2, \ldots, s, \quad (22)$$

where $a_i(p^*)$ and $b_i(p^*)$ are evaluated at $\hat{p}_{t-1}^*$ and are given by

$$a_i(p^*) = [(n_i - n_{ii})/(1 - p_i^*)^2 - (\tilde{n}_i - n_{ii})/(p_i^*)^2]^{-1}$$
$$b_i(p^*) = (n_i - n_{ii})/(1 - p_i^*) + (\tilde{n}_i - n_{ii})/p_i^*. \quad (23)$$

Once $\hat{p}_{ii}$ is calculated from (20) and once $\hat{p}^*$ is calculated from (22), then $\hat{\theta}$ is given by

$$1 - \hat{\theta}_i = (1 - \hat{p}_{ii})/(1 - \hat{p}_i^*), \quad i = 1, 2, \ldots, s. \quad (24)$$

4. AN EMPIRICAL EXAMPLE

The data I use in this section come from the National Longitudinal Survey of Young Men and consist of the industry of employment of 1,344 men between the ages of 14 and 24 for a six-year period (1966–1971). Since the data set consists of young men, it is reasonable to hypothesize that the men in the sample are unskilled, so the transition matrix would take the form $\theta + (I - \theta)\iota p$. The state definitions are the following:

1. Agriculture, forestry, fishing, and mining
2. Construction
3. Manufacturing
4. Transportation, communications, and public utilities
5. Wholesale and retail trade
6. Finance, insurance, real estate, business and repair services, and entertainment and recreation services
7. Professional and related services, and public administration

I first consider estimating the Markov chain model using the data for 1966 and 1967. The transition counts are

$$[n_{ij}] = \begin{bmatrix} 110 & 10 & 12 & 3 & 13 & 5 & 3 \\ 2 & 69 & 17 & 6 & 7 & 2 & 6 \\ 11 & 14 & 369 & 9 & 29 & 13 & 11 \\ 0 & 5 & 7 & 50 & 5 & 1 & 1 \\ 2 & 8 & 33 & 10 & 198 & 23 & 17 \\ 1 & 6 & 14 & 4 & 21 & 59 & 12 \\ 1 & 4 & 9 & 2 & 6 & 5 & 119 \end{bmatrix},$$

so the unrestricted ML estimate of the transition matrix is

$$\hat{Q}_u = \begin{bmatrix} .705 & .064 & .077 & .019 & .083 & .032 & .019 \\ .018 & .633 & .156 & .055 & .064 & .018 & .055 \\ .024 & .031 & .809 & .020 & .064 & .029 & .024 \\ .000 & .073 & .101 & .725 & .073 & .015 & .015 \\ .007 & .028 & .113 & .034 & .680 & .079 & .058 \\ .009 & .051 & .120 & .034 & .180 & .504 & .103 \\ .007 & .027 & .062 & .014 & .041 & .034 & .815 \end{bmatrix},$$

where the log-likelihood is −1,350.31.

The restricted ML estimates are given by

$$\hat{p}^* = [.042 \quad .115 \quad .275 \quad .078 \quad .245 \quad .126 \quad .118]$$

and

$$\hat{\theta} = \text{diag}[.692 \quad .585 \quad .737 \quad .701 \quad .577 \quad .433 \quad .790],$$

where the log-likelihood is −1,371.59. Thus a worker in construction, for example, has an estimated 1 − .585 = .415 probability of being laid off (or quitting), whereas a laid-off worker has a .245 probability of getting a job in wholesale and retail trade.

The likelihood ratio test statistic for $H_0: Q = \theta + (I -$

$\theta)\iota p^*$ is 42.54, which compares with a critical value (at the 5% level) of $\chi^2_{.05}(29) = 42.56$, so the null hypothesis is (just barely) accepted. Given the large number of observations and the fact that 29 restrictions are being tested, this suggests that the model fits the data reasonably well.

Now, consider estimating the mover–stayer model using the entire sample from 1966–1971 (so that $T = 5$). The transition counts, r_i and $n_i(0)$, are

$$[n_{ij}] = \begin{bmatrix} 435 & 33 & 53 & 10 & 33 & 12 & 10 \\ 26 & 456 & 59 & 19 & 24 & 14 & 22 \\ 41 & 73 & 1{,}989 & 37 & 127 & 50 & 45 \\ 9 & 23 & 28 & 343 & 17 & 7 & 9 \\ 16 & 33 & 139 & 40 & 979 & 79 & 59 \\ 8 & 15 & 52 & 12 & 79 & 344 & 37 \\ 7 & 15 & 45 & 11 & 30 & 31 & 685 \end{bmatrix},$$

$$[r_i] = [50 \quad 44 \quad 250 \quad 34 \quad 91 \quad 26 \quad 77],$$

and

$$[n_i(0)] = [156 \quad 109 \quad 456 \quad 69 \quad 291 \quad 117 \quad 146],$$

from which the unrestricted ML estimates of the mover's transition matrix and the proportions of stayers s_i are

$$\hat{Q}_u = \begin{bmatrix} .606 & .086 & .138 & .026 & .086 & .031 & .026 \\ .060 & .621 & .136 & .044 & .055 & .032 & .051 \\ .030 & .053 & .730 & .027 & .092 & .036 & .033 \\ .030 & .077 & .094 & .688 & .057 & .024 & .030 \\ .016 & .033 & .138 & .040 & .636 & .079 & .059 \\ .018 & .034 & .119 & .027 & .180 & .537 & .084 \\ .013 & .028 & .083 & .020 & .056 & .058 & .742 \end{bmatrix}$$

and

$$[\hat{s}_i] = [.260 \quad .343 \quad .430 \quad .401 \quad .233 \quad .186 \quad .390],$$

where the log-likelihood is $-5{,}748.14$. Thus an estimated 34.3% of construction workers in the sample are stayers and 65.7% are movers for example.

The restricted ML estimates of p^* and θ are

$$\hat{p}^* = [.064 \quad .117 \quad .287 \quad .075 \quad .228 \quad .121 \quad .109]$$

and

$$\hat{\theta} = \text{diag}[.579 \quad .571 \quad .621 \quad .663 \quad .529 \quad .473 \quad .711],$$

where the log-likelihood is $-5{,}817.12$. Thus a worker (who is a mover) in construction, for example, has a $1 - .571 = .429$ estimated probability of being laid off (or quitting), whereas a worker who is laid off has a .287 probability of ending up in manufacturing.

The likelihood ratio test statistic of $H_0: s_i = 0$ $(i = 1, 2, \ldots, s)$, that is, of the null hypothesis that the Markov chain model is the true model, is 267.39, which is greater than the critical value $\chi^2_{.05}(7) = 14.07$, and hence the Markov chain model is rejected in favor of the mover–stayer model.

The likelihood ratio test statistic for $H_0: Q = \theta + (I - \theta)\iota p^*$ is 137.97, which is greater than the critical value $\chi^2_{.05}(29) = 42.56$, so the null hypothesis is rejected for this data set.

[*Received July 1986. Revised August 1989.*]

REFERENCES

Barton, E. E., David, F. N., and Fix, E. (1962), "Persistence in a Chain of Multiple Events When There Is Simple Dependence," *Biometrika,* 49, 351–357.

Blumen, I., Kogan, M., and McCarthy, P. (1955), *The Industrial Mobility of Labor as a Probability Process* (Cornell Study of Industrial and Labor Relations), Ithaca, NY: Cornell University Press.

Frydman, H. (1984), "Maximum Likelihood Estimation in the Mover–Stayer Model," *Journal of the American Statistical Association,* 79, 632–638.

Geweke, J., Marshall, R., and Zarkin, G. (1986), "Mobility Indices in Continuous Time Markov Chains," *Econometrica,* 54, 1407–1423.

Goodman, L. A. (1964), "The Analysis of Persistence in a Chain of Multiple Events," *Biometrika,* 46, 405–411.

Sampson, M. (1987), "General Persistence: A Parametric Class of Markov Chains With an Application to Modelling Job Transitions for Unskilled Workers," unpublished manuscript, University of Western Ontario, Dept. of Economics.

Singer, B., and Spilerman, S. (1976), "Some Methodological Issues in the Analysis of Longitudinal Surveys," *Annals of Economic and Social Measurement,* 5, 447–474.

——— (1977), "Trace Inequalities for Mixtures of Markov Chains," *Advances in Applied Probability,* 9, 747–764.