ESTIMATING THE TRANSITION MATRIX OF A HOMOGENEOUS MARKOV CHAIN

by

Bruce A. Craig
Purdue University

and

Peter P. Sendi
Medizinische Universitäts-Polikinik
Basel, Switzerland

Technical Report # 98-12

Department of Statistics
Purdue University

June 1998

△

Estimating the Transition Matrix of a Homogeneous Markov Chain

Bruce A. Craig
Department of Statistics Purdue University
Peter P. Sendi
Medizinische Universitäts-Polikinik
Basel, Switzerland

Abstract

Discrete-time Markov chains have been successfully used to investigate treatment programs and health care protocols for chronic diseases. In these situations, the transition matrix, which describes the natural progression of the disease, is often estimated from a cohort observed at common intervals. We summarize methods to obtain the maximum likelihood estimate of the transition matrix when the cycle length of the model coincides with the observation interval, the cycle length does not coincide with the observation interval, and when the observation intervals are unequal in length.

1. Introduction

The discrete-time homogeneous Markov chain is a popular and often effective model to describe the progression of a chronic disease[1–3]. As a result, researchers have used this model to assess the ramifications of different treatment programs and investigate the cost-effectiveness of different health care protocols[4–10]. The popularity of this model is due primarily to two factors. First, chronic diseases can often be described in terms of distinct health states and the Markov chain is a simple yet powerful model to describe progression. Second, this model is easy to construct and study through matrix analysis and/or simulation.

The usual discrete-time Markov chain limits the description of each subject's history to equally spaced time points. The interval between these time points is known as the cycle length. In disease modeling, this length is often set to an interval associated with subject follow-ups and inference of the transition matrix is drawn from observational cohort data where each subject is observed at common intervals. Difficulties in estimation have been noted when the observation intervals are of varying length and/or do not coincide with the cycle length[4,11,12]. While a Bayesian approach to this problem in the context of a non-homogeneous model has been suggested[11], there has been surprisingly little written on estimating the probability transition matrix for this situation or in general[13]. This is most likely due to the

fact that the discrete-time model is often considered a special case of the continuous-time model for which various numerical algorithms have been proposed[14,15].

There is, however, a subtle difference between the two models that can alter the maximum likelihood estimate of the probability transition matrix. For the continuous-time model, the process is defined to be Markov at any discrete cycle length. For the discrete-time process, this is not the case. The process is Markov for any cycle length that is a multiple of the original length but it is not necessarily Markov for other cycle lengths. This is a variation of the embeddability problem of homogeneous Markov processes[15,16]. While the continuous-time model restricts the parameters to guarantee the Markov condition for all cycle lengths, the discrete-time model does not have this restriction and thus may better explain the data with the potential loss of not being Markov for all cycle lengths.

We do not want to argue the merits of continuous versus discrete-time models but simply point out that each model is being used in practice and deserves attention. To aid those using the discrete-time model, we describe methods to obtain the maximum likelihood estimate of the transition matrix. The organization of this paper is as follows: In Section 2, we describe the discrete-time Markov model in terms of its probability transition matrix. We then describe estimation techniques in relation to three common situations; 1) when the observation intervals coincide with the cycle length, 2) when the observation intervals do not coincide with the cycle length, and 3) when the observation intervals are unequal in length. In Section 4, we address two examples using these techniques followed by a discussion.

2. Discrete-Time Homogeneous Markov Model

Suppose a chronic disease can be classified into h distinct, non-overlapping health states. A subject's disease history can then be described by the movement through these states over time. The discrete-time Markov model describes this movement by modeling the states at distinct times termed cycles. This model does not concern itself with the progression between cycles and simply models the health state at the end of each cycle.

2.1 Transition Matrix

The key to the Markov model is the Markov property. This states that given the entire past history of the subject, the present state depends only on the most recent past state. This memoryless property allows the model to be described solely in terms of a single-cycle transition matrix. The transition matrix contains the probabilities, $\{\theta_{rc}; r, c = 1, 2, \ldots, h\}$, where θ_{rc} represents the probability of moving from state r to state c by the end of a cycle and $\sum_{c=1}^{h} \theta_{rc} = 1$ for all r. We assume a common cycle length so these probabilities are the same for each cycle.

As an example, consider a progressive disease with five health states ordered from least to most severe. A progressive disease means that the health state of an individual can never improve ($\theta_{rc} = 0$ for $c < r$) and is represented by the transition matrix M.

$$M = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} \\ 0 & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} \\ 0 & 0 & \theta_{33} & \theta_{34} & \theta_{35} \\ 0 & 0 & 0 & \theta_{44} & \theta_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

States 1-4 are called transitional states while State 5 is an absorbing state because once a subject is in this state, the subject remains in this state.

Since the transition matrix describes the progression, any model summary is a function of the single-cycle matrix. For example, the transition matrix for a cycle double in length would involve multiplying the single-cycle matrix with itself. Likewise, the transition matrix for a cycle half the original length would involve finding a half-cycle matrix such that the square of this matrix is the single-cycle matrix. This solution is not simply obtained by converting each probability in the single-cycle matrix to a rate and recomputing the probability for half the time. This is only appropriate if interest is restricted to a single probability. Appropriate methods need to take into account the dependent structure of the transition probabilities over cycles.

3. Estimation

In this section, we discuss three specific situations where inference of the transition matrix is drawn from longitudinal cohort data with observation intervals common to all subjects. We restrict our attention to obtaining the maximum likelihood estimate of the transition matrix for three specific cases increasing in complexity. The first case is when the observation intervals are constant and coincide with the cycle length. This represents the ideal situation. The second case is when the observation intervals are constant but do not coincide with the cycle length. The method discussed in this section can only be used in certain situations. When it cannot, the method discussed for the third case is possible. The third case represents the most common situation when the observation intervals are not equal in length. The cycle length may or may not coincide with one of these intervals.

3.1 Observation Intervals Coincide

Suppose you have a disease with h distinct health states. You want to estimate a two-year transition matrix and the data is from a cohort that was followed for four years with two year observation intervals. We will label the three health states for individual i as s_{i0}, s_{i2} and s_{i4}.

In this case, the observed two-year intervals coincide with the desired two-year transition matrix. Because our model is homogeneous, the observed transitions between the first two years can be pooled with the transitions between the second two years to form an observed two-year transition count matrix

$$T = \begin{pmatrix} n_{11} & n_{12} & \dots & n_{1h} \\ n_{21} & n_{22} & \dots & n_{2h} \\ \vdots & \vdots & \vdots & \vdots \\ n_{h1} & n_{h2} & \dots & n_{hh} \end{pmatrix},$$

where n_{rc} is the number of occurrences where $s_{i0} = r$ and $s_{i2} = c$ or $s_{i2} = r$ and $s_{i4} = c$.

Given the observed count matrix, the maximum likelihood estimate of the transition matrix is simply the row proportions of T,

$$\begin{aligned} \widehat{M} &= \left\{\widehat{\theta}\right\} \\ &\text{where} \\ \widehat{\theta}_{rc} &= n_{rc} / \sum_{j=1}^{h} n_{rj}. \end{aligned}$$

This estimation technique is commonly used and is presented here as a reference for the other two situations[13].

3.2 Observation Intervals Do Not Coincide

Let L_o be the common observation interval and L_c the desired cycle length. The maximum likelihood estimate of the transition matrix $\widehat{M}_o$, associated with the cycle length L_o, is obtained using the methods of Section 3.1. By the invariance property, the maximum likelihood estimate of the transition matrix associated with cycle length L_d is

$$\widehat{M}_d = \widehat{M}_o^k$$

where $k = L_d/L_o$. For example, if in the previous example a one-year rather than a two-year transition matrix were desired ($L_o = 2$ and $L_d = 1$), one would take the square root of the the estimated two-year transition matrix ($k = .5$).

Computation of this matrix is straightforward from the decomposition of $\widehat{M}_o$ into its eigenvalues and eigenvectors (spectral decomposition). Based on this decomposition, the $h \times h$ matrix $\widehat{M}_o$ can be expressed as

$$\widehat{M}_o = PDP^{-1}$$

where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_h \end{bmatrix}$$

and λ_i is the ith eigenvalue and its associated eigenvector is the ith column of P. It then follows that

$$\widehat{M}_o^k = PD^kP^{-1}$$

where

$$D^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_h^k \end{bmatrix}.$$

The eigenvalues are raised to the power k but the eigenvectors do not change.

This method is very similar to one method used to obtain the MLE estimate of the continuous-time transition matrix[14]. However, while this always works in the continuous-time case, it does not always work in the discrete-time case. Recall that a discrete-time model is not necessarily Markov at all cycle lengths. This is comparable to saying the eigenvalues of the transition matrix can be negative. While the continuous-time model guarantees these eigenvalues to be non-negative, the discrete-time model does not. Provided the estimated transition M is positive semidefinite (all the eigenvalues are non-negative), this method will allow you to compute the MLE directly. In situations where L_o is even and M is not positive semidefinite, the method described in the following section can be used.

3.3 Unequal Observation Intervals

In many situations, the observation intervals may be unequal in length[11]. As an example, suppose a one-year transition matrix is desired but the cohort was observed at year two and three. In this situation, the one-year transition matrix could be estimated using only the year two to three information but this throws away half of the observed data. Ideally one would

like to use all the information. We use the EM algorithm, an iterative method consisting of two steps, to compute the maximum likelihood estimate[17]. The E-step imputes the missing data by computing their expected value. The M-step pretends the imputed data is the true data and maximizes the likelihood. This is repeated until the results stabilize.

Consider the situation where the observation intervals and cycle length coincide. If n_{rc} represents the number of individuals that move from state r to state c in one cycle, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{r=1}^{h} \prod_{c=1}^{h} \theta_{rc}^{n_{rc}}$$

and the method of Section 3.1 provides the MLE of $\boldsymbol{\theta}$.

For this situation, there are T observation intervals which we assume are multiples $(k_1, k_2, \ldots, k_T)$ of the cycle length. The missing data are the health states for each individual at the unobserved cycles. Thus the EM algorithm involves imputing the data using their expected values and then using the methods of Section 3.1 to obtain a new estimate of the transition matrix. This is repeated until the transition matrix stabilizes. An initial transition matrix is needed to start the algorithm. We recommend starting values based on ignoring some of the data and using the methods described in Sections 3.1 or 3.2 to estimate the matrix. Convergence to the MLE is not guaranteed (may converge to local maximum) so several initial transition matrices are recommended.

For the E-step, we use the estimated single-cycle transition matrix to compute the probability of each path a subject could have followed to end up where he/she did after k_t cycles. For example, given a one-year transition matrix M, the two-year transition probabilities are given by computing $M \times M$. If we label the one-year transition matrix

$$M = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1h} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2h} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{h1} & \theta_{h2} & \cdots & \theta_{hh} \end{pmatrix},$$

this product can be expressed in terms of the one-year transition probabilities as

$$M \times M = \begin{pmatrix} \sum_{j=1}^{h} \theta_{1j}\theta_{j1} & \sum_{j=1}^{h} \theta_{1j}\theta_{j2} & \cdots & \sum_{j=1}^{h} \theta_{1j}\theta_{jh} \\ \sum_{j=1}^{h} \theta_{2j}\theta_{j1} & \sum_{j=1}^{h} \theta_{2j}\theta_{j2} & \cdots & \sum_{j=1}^{h} \theta_{2j}\theta_{jh} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^{h} \theta_{hj}\theta_{j1} & \sum_{j=1}^{h} \theta_{hj}\theta_{j2} & \cdots & \sum_{j=1}^{h} \theta_{hj}\theta_{jh} \end{pmatrix},$$

where each probability product $(\theta_{rj}\theta_{jc})$ represents one possible path from the initial state r to state c after two cycles (years).

Let us denote the number of observed subjects moving between state r and state c after k_t cycles as $n_{rc}^{k_t}$. Given the probability of each possible path, the remainder of this step involves estimating the number of subjects who follow each of these paths, given the observed data, and then tallying the number of single-cycle transitions. In our h state model, there are h paths in each cell of the two-year transition matrix. Each one of the individuals in that cell must have followed one of the h paths. The expected number of individuals to follow each path is based on the relative probability of each path (multinomial distribution). For example, in the upper left cell, the probability of an individual following the path $(1 \rightarrow 1 \rightarrow 1)$ is

$$P(1 \rightarrow 1 \rightarrow 1 \mid 1 \rightarrow ? \rightarrow 1) = \frac{\theta_{11}\theta_{11}}{\sum_{\mathrm{j=1}}^{\mathrm{h}} \theta_{1\mathrm{j}}\theta_{\mathrm{j1}}}.$$

So the expected number to have followed this path is

$$n_{11}^{2} \frac{\theta_{11}\theta_{11}}{\sum_{j=1}^{h} \theta_{1j}\theta_{j1}}.$$

The expected number of one-cycle transitions is then a bookkeeping exercise. For example, the path $1 \rightarrow 1 \rightarrow 1$ involves two $1 \rightarrow 1$ transitions and the path $1 \rightarrow 2 \rightarrow 1$ involves a $1 \rightarrow 2$ transition and a $2 \rightarrow 1$ transition. A single-cycle transition count matrix is generated and the M-step estimates a new transition matrix. This matrix is used to redefine the probability of each path in the next iteration.

The number of paths depends on the number of health states h, the number of cycles between observations k_t, and any restrictions imposed on the transition matrix (e.g. progressive disease). While the number of paths can be quite large, it is easy for a computer to handle. The appendix contains a description of one possible algorithm for the E-step.

4. Examples

4.1 Swiss HIV Cohort Study

Researchers constructed a homogeneous Markov chain to describe the monthly progression of HIV-infected subjects at the greatest risk of developing Mycobacterium avium complex (MAC) infection[9]. This progression included the possibility of movement between three distinct CD4-cell count ranges (with and without AIDS). Estimates of the monthly transitional probabilities are based on data from the Swiss HIV cohort study (SHCS). This is a multi-center, observational study where HIV-infected patients have fairly regular six month

follow-up visits[7,9]. To protect the authenticity of the SHSC data set, the following table of transitional counts is based on only a fraction of the entire SHSC data set.

Observed Six-Month Transitions - CD4 cell count(1993-1995)

Cell Count	Cell Count 0-49	50-74	75-UP
0-49	189	8	3
50-74	93	97	20
75-UP	37	70	293

Computing the row percentages, the estimated six-month transition matrix is

$$\widehat{M}_6 = \begin{pmatrix} .9450 & .0400 & .0150 \\ .4429 & .4619 & .0952 \\ .0925 & .1750 & .7325 \end{pmatrix}.$$

For this analysis, the desired cycle length is one month. To estimate the transition matrix for this interval, we decompose $\widehat{M}_6$. Using the **eigen** function in Splus, the matrices P and D are

$$P = \begin{bmatrix} -1.0000 & -0.1300 & -0.0623 \\ -1.0000 & 0.1898 & 1.0624 \\ -1.0000 & 1.1795 & -0.5244 \end{bmatrix} \qquad D = \begin{bmatrix} 1.0000 & 0 & 0 \\ 0 & 0.7505 & 0 \\ 0 & 0 & 0.3889 \end{bmatrix}.$$

Taking the sixth root of D and remultiplying the matrices, the estimated one month transition matrix is

$$\widehat{M}_1 = \begin{pmatrix} .9885 & .0091 & .0024 \\ .1032 & .8724 & .0244 \\ .0073 & .0460 & .9467 \end{pmatrix}.$$

4.2 EM Example

Because the number of potential paths grows rapidly, this example consists of a simulated data set from a two state model with observations at the second and third cycles. A one month transition matrix is desired. This is a small enough problem that the EM algorithm could be done using a spreadsheet.

Observed Two-Month Transitions

Health State	Health State 1	Health State 2
1	323	30
2	148	65

Observed One-Month Transitions

Health State	Health State 1	Health State 2
1	205	21
2	88	42

The E and M steps are detailed below. The E-step equations combine the observed one cycle transitions (first total in each equation) with the imputed number of one cycle transitions based on the observed two cycle transitions. In this case, each rc cell in the two cycle transition matrix is a sum of two path probabilities $(\theta_{r1}\theta_{1c} + \theta_{r2}\theta_{2c})$.

As initial values, we used the estimated one-year transition matrix $\widehat{M}_1$ and the square root of $\widehat{M}_2$ which are

$$\left(\widehat{M}_2\right)^{.5} = \begin{pmatrix} .9421 & .0579 \\ .4729 & .5271 \end{pmatrix} \qquad \widehat{M}_1 = \begin{pmatrix} .9071 & .0929 \\ .6769 & .3231 \end{pmatrix}.$$

Convergence occurred in about 20 iterations for both cases. The final matrix is

$$\widehat{M}_{\text{EM}} = \begin{pmatrix} .9228 & .0772 \\ .5623 & .4377 \end{pmatrix}.$$

E-Step

$$\begin{aligned}
\widehat{n}_{11} &= 205 + 2(323)\left(\frac{\widehat{\theta}_{11}\widehat{\theta}_{11}}{\widehat{\theta}_{11}\widehat{\theta}_{11} + \widehat{\theta}_{12}\widehat{\theta}_{21}}\right) + 30\left(\frac{\widehat{\theta}_{11}\widehat{\theta}_{12}}{\widehat{\theta}_{11}\widehat{\theta}_{12} + \widehat{\theta}_{12}\widehat{\theta}_{22}}\right) + 148\left(\frac{\widehat{\theta}_{21}\widehat{\theta}_{11}}{\widehat{\theta}_{21}\widehat{\theta}_{11} + \widehat{\theta}_{22}\widehat{\theta}_{21}}\right) \\
\widehat{n}_{12} &= 21 + 323\left(\frac{\widehat{\theta}_{12}\widehat{\theta}_{21}}{\widehat{\theta}_{11}\widehat{\theta}_{11} + \widehat{\theta}_{12}\widehat{\theta}_{21}}\right) + 30\left(\frac{\widehat{\theta}_{11}\widehat{\theta}_{12} + \widehat{\theta}_{12}\widehat{\theta}_{22}}{\widehat{\theta}_{11}\widehat{\theta}_{12} + \widehat{\theta}_{12}\widehat{\theta}_{22}}\right) + 65\left(\frac{\widehat{\theta}_{21}\widehat{\theta}_{12}}{\widehat{\theta}_{21}\widehat{\theta}_{12} + \widehat{\theta}_{22}\widehat{\theta}_{22}}\right) \\
\widehat{n}_{21} &= 88 + 323\left(\frac{\widehat{\theta}_{12}\widehat{\theta}_{21}}{\widehat{\theta}_{11}\widehat{\theta}_{11} + \widehat{\theta}_{12}\widehat{\theta}_{21}}\right) + 148\left(\frac{\widehat{\theta}_{21}\widehat{\theta}_{11} + \widehat{\theta}_{22}\widehat{\theta}_{21}}{\widehat{\theta}_{21}\widehat{\theta}_{11} + \widehat{\theta}_{22}\widehat{\theta}_{21}}\right) + 65\left(\frac{\widehat{\theta}_{21}\widehat{\theta}_{12}}{\widehat{\theta}_{21}\widehat{\theta}_{12} + \widehat{\theta}_{22}\widehat{\theta}_{22}}\right) \\
\widehat{n}_{22} &= 42 + 30\left(\frac{\widehat{\theta}_{12}\widehat{\theta}_{22}}{\widehat{\theta}_{11}\widehat{\theta}_{12} + \widehat{\theta}_{12}\widehat{\theta}_{22}}\right) + 148\left(\frac{\widehat{\theta}_{22}\widehat{\theta}_{21}}{\widehat{\theta}_{21}\widehat{\theta}_{11} + \widehat{\theta}_{22}\widehat{\theta}_{21}}\right) + 2(65)\left(\frac{\widehat{\theta}_{22}\widehat{\theta}_{22}}{\widehat{\theta}_{21}\widehat{\theta}_{12} + \widehat{\theta}_{22}\widehat{\theta}_{22}}\right)
\end{aligned}$$

M-Step

$$\widehat{\theta}_{11} = \frac{\widehat{n}_{11}}{\widehat{n}_{11} + \widehat{n}_{12}} \qquad \widehat{\theta}_{12} = \frac{\widehat{n}_{12}}{\widehat{n}_{11} + \widehat{n}_{12}}$$

$$\widehat{\theta}_{21} = \frac{\widehat{n}_{21}}{\widehat{n}_{21} + \widehat{n}_{22}} \qquad \widehat{\theta}_{11} = \frac{\widehat{n}_{22}}{\widehat{n}_{21} + \widehat{n}_{22}}$$

5. Conclusions

Although not discussed, model fit and accounting for uncertainty in the estimated matrix are also very important since most model summaries, such as life expectancy, are a function of this matrix. For fit, a likelihood ratio, or asymptotically equivalent chi-squared test statistic is described in Anderson and Goodman[13]. Confidence intervals based on bootstrap techniques[18] are implemented in Sendi et al.[7] and based on large sample theory in Anderson and Goodman[13]. A Bayesian approach to estimation (and uncertainty) has been described in Craig and Newton[11] for a non-homogeneous Markov chain.

With the growing popularity of discrete-time Markov chains, we feel it is important to describe appropriate techniques to estimate the transition matrix. While methods for the continuous-time Markov chain have been available in the literature for some time, we are unaware of any sources which summarize the techniques available for the homogeneous discrete-time chain.

REFERENCES

1. Beck, J.R. and Pauker, S.G. 'The Markov process in medical prognosis', *Medical Decision Making* 3, 419–458 (1983).

2. Sonnenberg F.A. and Beck, J.R. 'Markov models in medical decision making: A practical guide', *Medical Decision Making* 13, 322–338 (1993).

3. Briggs, A. and Sculper, M. 'An introduction to Markov modeling for economic evaluation', *Pharmacoeconomics* 13, 397–409 (1998).

4. Dasbach, E.J., Fryback, D.G., Newcomb, P.A., Klein, R. and Klein, B.E.K. 'Cost-effectiveness of strategies for detecting diabetic retinopathy', *Medical Care* 29(1), 20–39 (1991).

5. McCarthy, B.D., Wong, J.B., Munoz, A., Sonnenberg, F.A. 'Who should be screened for HIV infection?', *Archives of Internal Medicine* 153, 1107–1116 (1993).

6. Chancellor, J.V., Hill, A.M., Sabin, C. A., Simpson, K.N., and Youle, M. 'Modelling the cost effectiveness of Lamivudine/Zidovudine combination in HIV infection', *Pharmacoeconomics* 12(1), 54 –66 (1997).

7. Sendi P.P., Bucher H.C., Craig B.A., Pfluger D., and Battegay M. 'Modeling disease progression in HIV-infected patients without AIDS in the era of antiretroviral combination therapy', 12th World AIDS Conference, June 28 - July 3, 1998, Geneva, Switzerland (abstract, poster #43474).

8. Garg, S.K., Marshall G., Chase, H.P., Jackson, W.E., Archer, P., and Crews, M. 'The use of the Markov process in describing the natural course of diabetic retinopathy', *Archives of Ophthalmology* 108, 1245–1247 (1990).

9. Sendi P.P., Craig, B.A., Pfluger D., Gafni, A., and Bucher, H.C. 'Stepwise model validation of disease models with application to Mycobacterium avium complex infection in HIV disease', under review.

10. Freedberg, K.A., Scharfstein, J.A., Seage III, G.R., Losina, E., Weinstein, M.C., Craven, D.E., and Paltiel, D.A. 'The cost-effectiveness of preventing AIDS-related Opportunistic Infections', *Journal of the American Medical Association*, 279, 130–136 (1998).

11. Craig B.A. and Newton, M.A. 'Modeling the history of diabetic retinopathy' *in Case Studies in Bayesian Statistics III*, C. Gatsonis et al. eds, New York: Springer-Verlag, 305–323 (1997).

12. Miller, D.K. and Homan, S.M. 'Determining transition probabilities: Confusions and suggestions', *Medical Decision Making*, 14: 52–58 (1994).

13. Anderson, T.W. and Goodman, L.A. 'Statistical inference about Markov chains', *Annals of Mathematical Statistics*, 28, 89–110 (1957).

14. MAP Workshop, 'Markov models' ,*Statistics in Medicine*, **12**, 2127 – 2130 (1993).

15. Kalbfleisch, J.D. and Lawless, J.F. 'The analysis of panel data under a Markov assumption', *Journal of the American Statistical Association*, **80**, 863–871 (1985).

16. Kingman, J.F.C. 'The imbedding problem problem for finite Markov chains', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 1, 14–24 (1962).

17. Dempster, A.P., Laird, N.M., and Rubin, D.B. 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society* 39, 1–38 (1977).

18. Efron, B. and Tibshirani, R.J. 'An introduction to the bootstrap', Chapman & Hall. New York, 168-199 (1993).

APPENDIX

We briefly describe a matrix-oriented method to do the E-step. Consider a $h \times h$ single-cycle transition matrix M and data observed at T unique interval lengths equal to $k_t : t = 1, 2, ..., T$ cycles. For each of these intervals, the E-step involves 1) calculating the probability of each possible path, 2) obtaining the expected number of subjects to follow each path, and 3) tallying the number of single-cycle transitions.

Since the procedure is similar for each cycle, we focus on a single interval equal to k cycles. Consider constructing the following $h^k \times h$ matrix P_k using the following iterative procedure starting with $P_1 = M$,

$$P_k(h(r-1)+1,\ j) = P_{k-1}(r,c) \times M(c,j) \begin{cases} r = 1, 2, \ldots, h^{k-1} \\ c = 1, 2, \ldots, h \\ j = 1, 2, \ldots, h \end{cases}$$

In other words, the first row of P_k is the Kronecker product of the first element in P_{k-1} and the first row of M. The second row is the Kronecker product of the second element $P_{k-1}(1,2)$ and the second row of M and so on. The matrix P contains all possible paths of length k and has these paths arranged such that each column c contains all paths that end in state c with the first h^{k-1} rows containing the paths that start in state 1, the next h^{k-1} rows containing the paths that start in state 2, and so on. This allows easy computation of the expected number of subjects to follow each path since it arranges all the possible paths in adjacent rows and a single column.

In the construction of each of the probabilities in P_k, k single elements of M were multiplied together. We use the multiplication pattern to tally the single-cycle transitions. Let $\widehat{N}_k(r,c)$ represent the expected number of subjects to follow the path described in row r and column c of the P_k. The expected number of single-cycle transitions is

$$\begin{aligned} \widehat{n}_{r,c} &= \sum_{l=1}^{k-1} \sum_{i=1}^{h^{l-1}} \sum_{j=1}^{h} \sum_{m=1}^{h^{k-1-l}} \widehat{N}\left(s(c,h,k,l) + h^{k+1-l}(i-1) + m - 1,\ j\right) \\ &+ \sum_{i=1}^{h} \widehat{N}_k\left(r + h(i-1), c\right) \end{aligned}$$

where $s(c,h,k,l) = h^{k-1-l}(h(r-1)+c-1)$. The variable l represents the lth ordered single-cycle transition of a path. We're summing together each $\widehat{N}_k$ that contains a rc transition in the lth position. The second sum represents this process for the last single-cycle transition in each path. It is separate because the rc transition is only possible in one column.

We recommend the reader construct some P matrices in terms of single-cycle transitions to get an idea of the ordering. For example, consider the situation where subjects are observed after $k = 3$ cycles. If this were a $h = 2$ state model,

$$P_3 = \begin{pmatrix} \theta_{11}\theta_{11}\theta_{11} & \theta_{11}\theta_{11}\theta_{12} \\ \theta_{11}\theta_{12}\theta_{21} & \theta_{11}\theta_{12}\theta_{22} \\ \theta_{12}\theta_{21}\theta_{11} & \theta_{12}\theta_{21}\theta_{12} \\ \theta_{12}\theta_{22}\theta_{21} & \theta_{12}\theta_{22}\theta_{22} \\ \theta_{21}\theta_{11}\theta_{11} & \theta_{21}\theta_{11}\theta_{12} \\ \theta_{21}\theta_{12}\theta_{21} & \theta_{21}\theta_{12}\theta_{22} \\ \theta_{22}\theta_{21}\theta_{11} & \theta_{22}\theta_{21}\theta_{12} \\ \theta_{22}\theta_{22}\theta_{21} & \theta_{22}\theta_{22}\theta_{22} \end{pmatrix}.$$