HEALTH ECONOMICS
Health Econ. **11**: 33–42 (2002)
DOI:10.1002/hec.654

Estimation of the transition matrix of a discrete-time Markov chain

Bruce A. Craig[a,*] and Peter P. Sendi[b]
[a] *Department of Statistics, Purdue University, West Lafayette, USA*
[b] *Internal Medicine Outpatient Department, University of Basel, Switzerland*

Summary

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. Estimation of the matrix, however, is often complicated by the complex relationship among transition probabilities. This paper summarizes 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. In addition, the bootstrap is discussed as a method to assess the uncertainty of the maximum likelihood estimate and to construct confidence intervals for functions of the transition matrix such as expected survival. Copyright © 2002 John Wiley & Sons, Ltd.

Keywords transition matrix; disease progression; bootstrap; EM algorithm; cost-effectiveness

Introduction

Economic evaluations of health care technologies are becoming ever more important for guiding policy makers. To assess the cost effectiveness of a new technology, modeling techniques are unavoidable if resources limit a formal study and data from different sources must be combined [1]. The discrete-time Markov chain is currently the most popular model used for evaluating interventions aimed at treating chronic diseases [2–4]. 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 through these states. Second, this model is easy to construct and study through matrix analysis and/or simulation.

The typical discrete-time Markov chain limits the description of each subject's history to equally spaced time points. In other words, instead of modeling the possibility of progression at every instant in time (i.e. a continuous-time model), the disease state is only classified at distinct-time points (e.g. day, month, or year). The interval between these time points is known as the cycle length. A homogeneous Markov chain is one where the state-to-state transition probabilities remain constant over time (i.e. cycles) and across subjects. A non-homogeneous Markov chain is one where these probabilities vary over time and/or across subjects. For example, a simple yet common non-homogeneous model over time is

*Correspondence to: Department of Statistics, 1399 Mathematical Sciences, Purdue University, West Lafayette, IN 47907-1399, USA. Tel.: +765-494-6043; fax: +765-494-0558; e-mail: bacraig@stat.purdue.edu

Received 2 June 2000
Accepted 27 April 2001

one where it is a product of homogeneous Markov chains [5]. In other words, the first a cycles of the chain follow one transition matrix and then the next b cycles follow another and so on.

In disease modeling, the cycle length is often set to an interval associated with medical follow-ups (e.g. six months) 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 [6,7]. While a Bayesian approach to this problem in the context of a non-homogeneous model has been suggested [8], there has been surprisingly little written on estimating the probability of a homogeneous model or a non-homogeneous model over time which is a product of homogeneous chains [9]. To aid those using this discrete-time model, methods are described to obtain the maximum likelihood estimate of the homogeneous transition matrix.

In addition to the matrix estimate itself, one is often interested in summaries that are functions of this matrix. To assess the uncertainty and construct confidence intervals of these summaries, the bootstrap is suggested. While sensitivity analysis [10], which involves varying single or multiple transition parameters, is frequently used to investigate the behavior of the transition matrix, it should not be used to form confidence intervals because it does not adequately account for the complex relationship among the transition probabilities. The bootstrap, on the other hand, varies the entire set of transition probabilities by generating other possible cohort data sets. Using these other data sets to form a set of possible transition matrices overcomes this complicated dependency among probabilities.

The organization of this paper is as follows: In 'Discrete-time homogeneous Markov model' the discrete-time Markov model is described in terms of its probability transition matrix. In Estimation, the estimation techniques are described for 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 'Confidence intervals using the bootstrap' the Bootstrap is used to assess uncertainty and construct confidence intervals and in 'Examples', these techniques are illustrated with two examples. This is followed by a discussion.

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.

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 $\sum_{j=1}^{h} \theta_{rj} = 1$ for all r. A common cycle length is assumed 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, if it exists, is not simply obtained by converting each probability in the single-cycle matrix to a rate and recomputing the probability for half the time [7].

This approach is only appropriate if restricted to a single probability. With a Markov chain, a transition across two cycles involves a complex combination of transition probabilities (through matrix multiplication). Appropriate methods need to take into account this dependent structure of the transition probabilities over cycles.

Estimation

In this section, it is assumed that the transition matrix will be estimated from longitudinal cohort data with observation intervals common to all subjects. Attention is restricted 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. 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.

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 two-year observation intervals. The three health states for individual i are labeled 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 the 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:

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

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 N,

$$\widehat{M} = \left\{\widehat{\theta}\right\} \quad \text{where } \widehat{\theta}_{rc} = n_{rc} \Big/ \sum_{j=1}^{h} n_{rj}$$

This estimation technique is commonly used and is presented here as a reference for the other two situations [9].

Observation intervals do not coincide

Let L_o be the common observation interval and L_d 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 'Observation intervals coincide'. By the invariance property, the maximum likelihood estimate of the transition matrix associated with cycle length L_d is

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

where $t = L_d/L_o$. 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 estimated two-year transition matrix ($t = 0.5$).

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

$$\widehat{M}_o = PDP^{-1} \quad \text{where } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_h \end{bmatrix}$$

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

$$\widehat{M}_o^t = PD^tP^{-1} \quad \text{where } D^t = \begin{bmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_h^t \end{bmatrix}$$

The eigenvalues are raised to the power t but the eigenvectors do not change. Many software

packages, such as Splus, have matrix decomposition functions, so these calculations can be done very quickly.

This method is very similar to one method used to obtain the MLE estimate of the continuous-time transition matrix [11]. However, while this always works in the continuous-time case, it does not always work in the discrete-time case. 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. Provided the estimated transition matrix $\widehat{M}$ 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 $\widehat{M}$ is not positive semidefinite, the method described in the following section can be used.

Unequal observation intervals

In many situations, the observation intervals may be unequal in length [6,8]. 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. This can be done using the EM algorithm [12]. The E-step imputes the missing data by computing the expected number of single-cycle transitions. The M-step treats the expected number of single-cycle transitions as the true data set and maximizes the likelihood. This is repeated until the transition probabilities stabilize.

Recall 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

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

and the method of 'Observation intervals coincide' provides the MLE of $M = \{\theta\}$.

Consider that there are T observation intervals which are integer 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 these states at the unobserved cycles, tallying the expected number of transitions, and then using the methods of 'Observation intervals coincide' 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. Convergence to the MLE is not guaranteed (may converge to local maximum) so several initial transition matrices are recommended.

For the E-step, the estimated transition matrix is used 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 $M \times M$. Labeling 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).

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 and then tallying the number of single-cycle transitions. In the above 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

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

so the expected number of subjects 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 each subject could have followed 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.

Confidence intervals using the bootstrap

As one can see from the product of $M \times M$, model summaries, such as the probability of entering the absorbing state by cycle 5 or the expected survival, are a complex function of single-cycle probabilities. Methods which vary only a subset of the transition parameters (e.g. sensitivity analysis) do not properly address this complex relationship. While they can still be very helpful in understanding the behavior of the model, other methods must be used to assess uncertainty and construct confidence intervals.

For this purpose, Efron's bootstrap is recommended [13]. With this method, other possible data sets, the same size as the original, are formed by sampling with replacement from the original data set. This is done by addressing each row of the transition count matrix N separately. Letting $n_{r.}$ denote the total number of transitions for row r, bootstrapping row r simply involves sampling $n_{r.}$ transitions with replacement from the observed $n_{r.}$ transitions. In other words, $n_{r.}$ draws are taken from a Multinomial distribution with probabilities $\{\widehat{\theta}_{rc}\}$ to generate a new set of transition counts for row r. Combining the results of each row forms a new transition count matrix N^* and thus another possible transition probability matrix $\widehat{M}^*$. If the desired cycle length and observation interval do not coincide then spectral decomposition or the EM algorithm would be used on each bootstrap sample to obtain a new transition matrix [14,15].

The collection of bootstrapped transition matrices approximates the sampling distribution. From this distribution, one could assess the uncertainty of each probability in the transition matrix as well as any function of the transition matrix. For example, if one were interested in the expected survival of an individual starting in state s_1. One could compute this expected value for each matrix in the bootstrap set thereby creating a sampling distribution for expected survival. An example of this bootstrap approach is found in Sendi *et al.* [5].

Examples

To illustrate these methods, two examples are presented. The first example utilizes the Swiss HIV cohort study (SHCS) database to estimate one-month CD4-cell count state transition probabilities from six-month follow-ups. The second example, in order to describe the EM calculations in detail, is based on a simulated data set. In addition, for the second data set, Bootstrap procedures are used to form a 95% CI for the expected number of cycles until entering the absorbing state.

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 [16]. 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 SHCS. This is a multi-center,

observational study where HIV-infected patients have fairly regular six month follow-up visits.

The six-month transition matrix, estimated from Table 1, is as follows:

$$\widehat{M}_6 = \begin{pmatrix} 0.9216 & 0.0446 & 0.0338 \\ 0.5811 & 0.2415 & 0.1774 \\ 0.2346 & 0.2346 & 0.5309 \end{pmatrix}$$

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

$$P = \begin{bmatrix} -1.0000 & -0.2507 & -0.0447 \\ -1.0000 & 0.5630 & 1.3847 \\ -1.0000 & 1.8646 & -0.7722 \end{bmatrix},$$

$$D = \begin{bmatrix} 1.0000 & 0 & 0 \\ 0 & 0.5702 & 0 \\ 0 & 0 & 0.1238 \end{bmatrix}$$

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

$$\widehat{M}_1 = \begin{pmatrix} 0.9819 & 0.0122 & 0.0059 \\ 0.1766 & 0.7517 & 0.0717 \\ 0.0177 & 0.0933 & 0.8830 \end{pmatrix}$$

If this matrix is then multiplied together six times, the result will be the matrix $\widehat{M}_6$ as expected. This procedure is much faster and simpler than any approach suggested in the literature [6]. As a comparison, if one were to apply the methods of Miller and Homan [7] to each individual transition probability and then standardize each set of row probabilities so they sum to one, the estimated matrix would be

$$\widehat{M}_1^{\mathrm{MH}} = \begin{pmatrix} 0.9630 & 0.0211 & 0.0159 \\ 0.6367 & 0.2123 & 0.1510 \\ 0.2119 & 0.2119 & 0.5762 \end{pmatrix}$$

If this matrix is multiplied together six times, the results are not close to $\widehat{M}_6$ as shown below:

$$\widehat{M}_6^{\mathrm{MH}} = \begin{pmatrix} 0.9193 & 0.0360 & 0.0447 \\ 0.8966 & 0.0430 & 0.0604 \\ 0.8471 & 0.0582 & 0.0947 \end{pmatrix}$$

This matrix suggests there will be far too many patients in state 1 after six cycles.

Table 1. Observed six-month transitions – CD4 cell count (1993–1995)

Initial CD4 cell count	Six month CD4 cell count		
	0–49	50–74	75–UP
0–49	682	33	25
50–74	154	64	47
75–UP	19	19	43

EM example

Because the number of potential paths a subject can follow grows rapidly as h and k_t increase, this example consists of a simulated data set from a three state model (state 3 is an absorbing state) with observations at the second and third cycles (Table 2). A one-month transition matrix is desired. This is a small enough problem that the EM algorithm can be done using a spreadsheet package.

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 $(r, c \leqslant 2)$ in the two-cycle transition matrix is a sum of two-path probabilities, $\theta_{r1}\theta_{1c} + \theta_{r2}\theta_{2c}$, while the last cell in the first two rows is a sum of three-path probabilities. The third

Table 2. Summary of observed transitions between health states

Initial state	Final state		
	1	2	3
One-month transitions			
1	227	22	21
2	20	70	17
3	0	0	138
Two-month transitions			
1	214	45	41
2	56	62	82
3	0	0	0

row can be ignored since state 3 is an absorbing state.

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

$$\left(\widehat{M}_2\right)^{0.5} = \begin{pmatrix} 0.8312 & 0.1097 & 0.0591 \\ 0.2048 & 0.5362 & 0.2590 \\ 0.0000 & 0.0000 & 1.0000 \end{pmatrix},$$

$$\widehat{M}_1 = \begin{pmatrix} 0.8407 & 0.0815 & 0.0778 \\ 0.1869 & 0.6542 & 0.1589 \\ 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$$

are used. Convergence occurred in 8–10 iterations for both cases. The MLE is

$$\widehat{M}_1^{\mathrm{EM}} = \begin{pmatrix} 0.8363 & 0.0952 & 0.0685 \\ 0.1964 & 0.5754 & 0.2282 \\ 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$$

For each of the following steps, the hat notation is dropped from the equations simply for readability. It should be noted that the θ and n will be changing each iteration.

E-Step:

$$n_{11} = 227 + 2(214)\left(\frac{\theta_{11}\theta_{11}}{\theta_{11}\theta_{11} + \theta_{12}\theta_{21}}\right) + 45\left(\frac{\theta_{11}\theta_{12}}{\theta_{11}\theta_{12} + \theta_{12}\theta_{22}}\right) + 41\left(\frac{\theta_{11}\theta_{13}}{\theta_{11}\theta_{13} + \theta_{12}\theta_{23} + \theta_{13}}\right) + 56\left(\frac{\theta_{21}\theta_{11}}{\theta_{21}\theta_{11} + \theta_{22}\theta_{21}}\right)$$

$$n_{12} = 22 + 214\left(\frac{\theta_{12}\theta_{21}}{\theta_{11}\theta_{11} + \theta_{12}\theta_{21}}\right) + 45 + 41\left(\frac{\theta_{12}\theta_{23}}{\theta_{11}\theta_{13} + \theta_{12}\theta_{23} + \theta_{13}}\right) + 62\left(\frac{\theta_{21}\theta_{12}}{\theta_{21}\theta_{12} + \theta_{22}\theta_{22}}\right)$$

$$n_{13} = 21 + 41\left(\frac{\theta_{13} + \theta_{11}\theta_{13}}{\theta_{11}\theta_{13} + \theta_{12}\theta_{23} + \theta_{13}}\right) + 82\left(\frac{\theta_{21}\theta_{13}}{\theta_{21}\theta_{13} + \theta_{22}\theta_{23} + \theta_{23}}\right)$$

$$n_{21} = 20 + 214\left(\frac{\theta_{12}\theta_{21}}{\theta_{11}\theta_{11} + \theta_{12}\theta_{21}}\right) + 56 + 62\left(\frac{\theta_{21}\theta_{12}}{\theta_{21}\theta_{12} + \theta_{22}\theta_{22}}\right) + 82\left(\frac{\theta_{21}\theta_{13}}{\theta_{21}\theta_{13} + \theta_{22}\theta_{23} + \theta_{23}}\right)$$

$$n_{22} = 70 + 45\left(\frac{\theta_{12}\theta_{22}}{\theta_{11}\theta_{12} + \theta_{12}\theta_{22}}\right) + 56\left(\frac{\theta_{22}\theta_{21}}{\theta_{21}\theta_{11} + \theta_{22}\theta_{21}}\right) + 2(62)\left(\frac{\theta_{22}\theta_{22}}{\theta_{21}\theta_{12} + \theta_{22}\theta_{22}}\right) + 82\left(\frac{\theta_{22}\theta_{23}}{\theta_{21}\theta_{13} + \theta_{22}\theta_{23} + \theta_{23}}\right)$$

$$n_{23} = 17 + 41\left(\frac{\theta_{12}\theta_{23}}{\theta_{11}\theta_{13} + \theta_{12}\theta_{23} + \theta_{13}}\right) + 82\left(\frac{\theta_{22}\theta_{23} + \theta_{23}}{\theta_{21}\theta_{13} + \theta_{22}\theta_{23} + \theta_{23}}\right)$$

M-Step:

$$\theta_{11} = \frac{n_{11}}{n_{11} + n_{12} + n_{13}} \qquad \theta_{12} = \frac{n_{12}}{n_{11} + n_{12} + n_{13}}$$

$$\theta_{13} = \frac{n_{13}}{n_{11} + n_{12} + n_{13}}$$

$$\theta_{21} = \frac{n_{21}}{n_{21} + n_{22} + n_{23}} \qquad \theta_{22} = \frac{n_{22}}{n_{21} + n_{22} + n_{23}}$$

$$\theta_{23} = \frac{n_{23}}{n_{21} + n_{22} + n_{23}}$$

Bootstrap example

A function of the transition matrix that is usually of great interest is the time until a subject reaches the absorbing state. For example, when the absorbing state is death, this time is the life-expectancy of the subject. Consider the EM example with death as the absorbing state and suppose there was interest in estimating the life expectancy of someone in state 1. Using the fundamental matrix solution [2], the expected number of cycles is estimated to be 10.23 cycles. Since each cycle represents one month, the

life-expectancy is 10.23 months.

$$(I - Q) = \begin{pmatrix} 1 - 0.8363 & -0.0952 \\ -0.1964 & 1 - 0.5754 \end{pmatrix}$$

$$(I - Q)^{-1} = \begin{pmatrix} 8.357 & 1.874 \\ 3.865 & 3.222 \end{pmatrix}$$

To assess the precision of this estimate, a 95% CI is constructed using Efron's bootstrap. In this case, this involves resampling both the single- and double-cycle transition count matrices and then using the EM algorithm to estimate the transition matrix. For example, to bootstrap the first row of the two-cycle transition count matrix, one would sample 300 transitions with replacement from a collection of 214 $1 \rightarrow 1$ transitions, 45 $1 \rightarrow 2$ transitions and 41 $1 \rightarrow 3$ transitions.

For each one-month matrix that is generated, the fundamental matrix solution is calculated. Fig. 1 displays the sampling distribution of the expected number of cycles based on 100 bootstrap samples. Using an equi-tailed confidence interval, the 95% CI is (8.98,11.73) which is a little smaller than three months in length. It should be noted that while the spectral decomposition may be possible with the observed data set, it may not work for all bootstrapped data sets. In these cases, the EM estimation approach should be used.

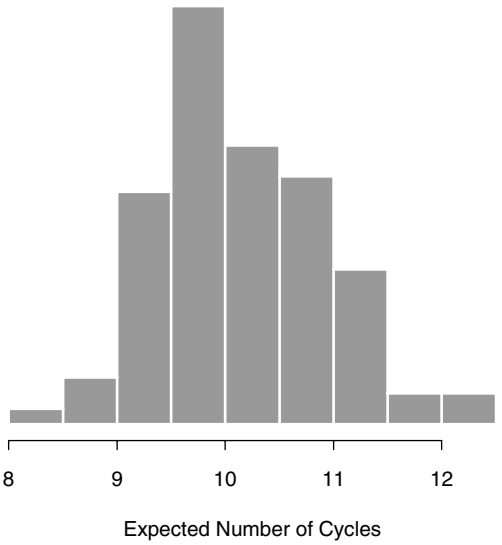

Figure 1. The sampling distribution of the expected number of cycles until entering state 3 starting in state 1 based on 100 bootstrap samples. If state 3 were death, this summary is the life expectancy (cycle is a month here)

Discussion

With the growing popularity of discrete-time Markov chains and decision tree analyses which incorporate a Markov chain, 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. These estimation techniques not only provide the researcher with a method to obtain the MLE estimate but they can also be combined with Efron's bootstrap to assess the uncertainty of the matrix or function of the matrix, such as life-expectancy or a cost-effectiveness statistic. Because of the relative simplicity of the techniques, a computer algorithm can perform these assessments in very little time.

While sensitivity analysis is a very helpful technique to investigate the behavior of the Markov model, it should not be used to construct a confidence interval because it does not take into account the model restrictions (e.g. row sums must be one) and complex dependency of all the transition probabilities. As shown in the first example, when a probability is altered and the row 'standardized', this distorts the relationship among the probabilities giving vastly different results. The bootstrap, on the other hand, is a very natural way to satisfy the row sum restriction as well as the complex relationship among the probabilities. Thus, while we have only addressed a Markov chain model, the bootstrap would be helpful with any model that incorporates a complicated parameter structure.

More recently, stochastic versions of this model have been used to allow for within (e.g. time difference) and between patient variability in the transition probabilities. This involves describing the transition probabilities in terms of a joint distribution and sampling from these distributions to determine the transition probabilities for a specific cycle/subject. In most cases, however, this joint distribution is centered at MLE estimates so these estimation techniques can still be useful in this situation. It also should be noted that while

these stochastic models provide more realism, this also means more uncertainty which can mask treatment protocol differences. As a result, it is important that researchers assess which model better addresses the primary questions of their research and not just select the most realistic model.

Although not discussed, model fit is also very important to consider and is often overlooked. If one is considering the model for prediction purposes, predicting the outcomes of an independent yet similar cohort is one valuable check of the model [16]. Also a likelihood ratio, or asymptotically equivalent χ^2 test statistic, as described in Anderson and Goodman [9] provides a measure of fit to one's own data. If the cohort data set is large enough, one could also use a cross validation technique.

Acknowledgements

Contract grant/sponsor: NEI Small Research; number: EY12254-01.

Appendix

In this problem, the key to the EM algorithm is an efficient E-step. Recall, 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. In this section, a matrix-oriented approach to keep track of all the potential paths is described. Shown below is an example of such a matrix which contains all potential paths for $k = 3$ cycles when there are only $h = 2$ states:

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

Notice the organization of these paths. The first column contains those paths that end in state 1 and the second column contains those paths that end in state 2. The first four rows contain paths which start in state 1 and the last four rows contain paths which start in state 2. Finally, the three single cycle transitions that make up a path have a particular pattern as you go down the paths within a column. This type of organization makes the accounting of the E-step very easy.

To construct such a matrix, 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. Since the matrix construction is similar for each cycle length, assume a single interval equal to k cycles. The matrix, P_k (a $h^k \times h$ matrix) is constructed using the following iterative matrix multiplication equation, $P_1 = M$ and

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

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_k has all potential 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 $\widehat{N}(i,j)$ represent the expected number of subjects to follow the path described in row i and column j of P_k. The exp-ected number single-cycle transitions from r to c is

$$\widehat{n}_{r,c} = \sum_{l=1}^{k-1} \sum_{i=1}^{h^{l-1}} \sum_{j=1}^{h} \sum_{m=1}^{h^{k-1-l}} \widehat{N}(s(c,h,k,l) + h^{k+1-l}(i-1) + m - 1,\ j) + \sum_{i=1}^{h} \widehat{N}(r + h(i-1), c)$$

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. This equation is simply summing together each $\widehat{N}(i,j)$ whose path 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 column c.

References

1. Buxton MJ, Drummond MF, van Hout BA, *et al*. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997; **6**: 217–227.
2. Beck JR, Pauker SG. The Markov process in medical prognosis. *Med Decision Making* 1983; **3**: 419–458.3. Briggs A, Sculper M. An introduction to Markov modeling for economic evaluation. *Pharmacoeconomics* 1998; **13**: 397–409.
4. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decision Making* 1993; **13**: 322–338.
5. Sendi PP, Bucher HC, Craig BA, Pfluger D, Battegay M. Estimating AIDS-free survival in a severely immunosuppressed asymptomatic HIV-infected population in the era of antiretroviral triple combination therapy. *J Acquir Immune Defic Syndr Hum Retrovirol* 1999; **20**: 376–381.
6. Dasbach EJ, Fryback DG, Newcomb PA, Klein R, Klein BEK. Cost-effectiveness of strategies for detecting diabetic retinopathy. *Med Care* 1991; **29**: 20–39.
7. Miller DK, Homan SM. Determining transition probabilities: confusions and suggestions. *Med Decision Making* 1994; **14**: 52–58.
8. Craig BA, Fryback DG, Klein R, Klein BEK. A Bayesian approach to modelling the natural history of a chronic condition from observations with intervention. *Stat Med* 1999; **18**: 1355–1371.
9. Anderson TW, Goodman LA. Statistical inference about Markov chains. *Ann Math Stat* 1957; **28**: 89–110.
10. Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In *Cost-Effectiveness in Health and Medicine*, Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Oxford University Press: New York, 1996; 247–275.
11. Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *J Amer Stat Assoc* 1985; **80**: 863–871.
12. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc* 1977; **39**: 1–38.
13. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall: New York, 1993.
14. LePage R, Billard L. *Exploring the Limits of the Bootstrap*. Wiley: New York, 1992.
15. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. Wiley: New York, 1997.
16. Sendi PP, Craig BA, Pfluger D, Gafni A, Bucher HC. Systematic validation of disease models for pharmacoeconomic evaluations. *J Eval Clin Pract* 1999; **5**: 283–295.

