USING MARKOV CHAIN TO DESCRIBE THE PROGRESSION OF CHRONIC DISEASE

by

SIJIA DAVIS

B.A., Kansas State University, 2012

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Abigail Jager

Abstract

A discrete-time Markov chain with stationary transition probabilities is often used for the purpose of investigating treatment programs and health care protocols for chronic disease. Suppose the patients of a certain chronic disease are observed over equally spaced time intervals. If we classify the chronic disease into n distinct health states, the movement through these health states over time then represents a patient's disease history. We can use a discrete-time Markov chain to describe such movement using the transition probabilities between the health states.

The purpose of this study was to investigate the case when the observation interval coincided with the cycle length of the Markov chain as well as the case when the observational interval and the cycle length did not coincide. In particular, we are interested in how the estimated transition matrix behaves as the ratio of observation interval and cycle length changes.

Our results suggest that more estimation problems arose for small sample sizes as the length of observational interval increased, and that the deviation from the known transition probability matrix got larger as the length of observational interval increased. With increasing sample size, there were fewer estimation problems and the deviation from the known transition probability matrix was reduced.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Markov Chains

A stochastic process is a collection of random variables defined on a common probability space indexed by the index set T, $\{Y_t, t \in T\}$, which describes how some system evolves over time (Resnick, 1992). T can be continuous, discrete, or even a collection of regions in some cases. A stochastic process is said to be stationary when the statistical evolution of the process over an interval is the same as that of the process over a translated interval. That is to say, a stochastic process $\{Y_n, n \geq 0\}$ is stationary if for any integers $m \geq 0, k > 0$, $(Y_0, \dots, Y_m)$ has the same distribution with $(Y_k, \dots, Y_{m+k})$.

A Markov chain is often a realistic stochastic process for real life situations. When constructing a stochastic process, a challenge is to have dependencies among the random variables that allow for sufficient realism but also are mathematically tractable. One of the main advantages of a Markov chain process is that it balances these two demands nicely (Resnick, 1992). To define a Markov chain, let $\{X_n, n = 0, 1, 2, \dots\}$ be a stochastic process that takes on a finite or countable number of values. The set consisting of all possible values is called the state space, which is denoted by S. For the expression $X_n = i$, we say that the process is in state i at time n or after the n^{th} step. It is assumed that every time the process is in state i, there exists a fixed probability p_{ij} that the process will move to state j in the next step. That is,

$$p_{ij} = P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\}$$

for all states $i_0, i_1, \dots, i_{n-1}, i, j \in S$ and for all $n \geq 0$. The process described above is known as a Markov chain process.

One important characteristic of Markov chain process is that the conditional distribution of any future state X_{n+1}, given the past states $X_0, X_1, \dots, X_{n-1}$ and the present state X_n, is independent of all the past states and only depends on the present state X_n. That is,

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\} = p_{ij}$$

for $n \geq 0$. Since p_{ij} indicates the probability that the process will move from state i to state j in the next step, it has to be a nonnegative value. Also,

$$\sum_{j=0}^{\infty} p_{ij} = 1 \text{ for } i = 0, 1, \dots$$

Let P denotes the matrix consisting of all one-step transition probabilities p_{ij}, so that $P = [p_{ij}]$, where p_{ij} is the element of P in the i^{th} row and j^{th} column. P is called the transition probability matrix of the Markov chain. Since p_{ij} add to 1 across all possible values of j, each row of the transition probability matrix P sums to 1 as well. If P does not depend on the number of steps n, we then say that the p_{ij} are stationary transition probabilities, and the Markov chain $\{X_n, n = 0, 1, 2, ...\}$ is said to be homogeneous (Resnick, 1992).

We can derive higher-ordered transition probabilities using simple matrix multiplication as long as the one-step transition probabilities p_{ij} are known. Let us define the n-step transition probability p_{ij}^n as the probability that a process ends up in state j after n steps given that the process starts in state i. That is, $p_{ij}^n = P\{X_{k+n} = j | X_k = i\}$ for all $n \geq 0$, $k \geq 1$ and $i, j \geq 0$. The Chapman-Kolmogorov equations can be used to derive these n-step transition probabilities (Resnick, 1992). According to the Chapman-Kolmogorov equations,

$$p_{ij}^{n+m} = \sum_{k=0}^{\infty} p_{ik}^n \, p_{kj}^m \text{ for all } n, m \geq 0$$

The term $p_{ik}^n p_{kj}^m$ represents the probability that starting in state i, the process will enter state j in $n + m$ steps through a path which takes it into state k at the n^{th} transition. Let P^n denote the matrix of all n-step transition probability where p_{ij}^n is the element in the i^{th} row and j^{th} column. P^n is the n-step transition probability matrix. The Chapman-Kolmogorov equations then are

$$P^{n+m} = P^n P^m$$

where P^n is taking the one-step transition probability matrix to the n^{th} power. In particular, if the probability that a process going from state i to state k in n steps does not depend on the time at which the process is initiated, then the $p_{ij}^n s$ are stationary n-step transition probabilities.

Let $\{X_n : n \geq 0\}$ be a Markov chain with state space S, and let i and j be two states in S. State j is said to be accessible from state i (written $i \to j$) if $p_{ij}^n > 0$ for some $n \geq 0$. In other words, if it is possible for a process to enter state j in a finite number of steps given that the process starts at state i, then j is accessible from i. Furthermore, state i and j are said to be communicating (written $i \leftrightarrow j$) if they are accessible from each other. Note that any state communicates with itself. Communication is an equivalence relation on the state space S since it satisfies the following three properties (Resnick, 1992):

Reflective property: $i \leftrightarrow i$ for all $i \in S$

Symmetry property: $i \leftrightarrow j$ iff $j \leftrightarrow i$

Transitive property: if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$

Equivalence classes are defined to be disjoint subsets of the state space S. Specifically, the union of all equivalence classes makes up the entire state space. Two states that communicate with each other belong in the same equivalence class. A Markov chain is said to be irreducible if the only equivalent class of the state space S is S itself. So, for any two states $i, j \in S$ in an irreducible Markov chain, i communicates with j. A subset C of the state space S is closed if the process starting at any state $i \in C$ never leaves C. Note that C is closed if and only if $p_{ij} = 0$ for all $i \in C$ and $j \in C^c$. If a closed set only contains one single state j, then j is called an absorbing state. Note that j is absorbing if and only if $p_{jj} = 1$, in other words, a process that enters state j never leaves j.

A state i is recurrent if the Markov chain returns to i with probability 1 in a finite number of steps. In particular, a state i is said to be positive recurrent if the expected value of the number of steps it takes for the process to return to i is finite, and it is called null recurrent if i is recurrent but the expected value of the number of steps it takes for the process to return to i is infinite. In a finite-state Markov chain, all recurrent states are in fact positive recurrent. On the other hand, a state i is transient if the probability that the process will return to i at some point is less than 1. In other words, for a transient state i, there is a positive probability that the process will never return to i. Note that if i is the initial state, say $X_0 = i$, then state i is recurrent if and only if the expected number of visits by the Markov chain to i is infinite. The state i is transient if and only if the expected number of visits by the Markov chain to state i is finite. Also note that if the state space of a Markov chain is finite, then not all states are transient. Thus, at least one state must be recurrent for a finite-state Markov chain. Suppose that state $i \in S$ is a recurrent state, and state i communicates with state j. In this case, state j is also recurrent. If C is a recurrent equivalence class in the state space S, then C is closed. Also, if C is a finite closed equivalence class, then C is recurrent (Resnick, 1992).

The period d of a state i is the greatest common divisor of $\{n \geq 1: p_{ii}^n > 0\}$. That is to say, for a process that starts at state i, its returns to state i are only possible via paths whose lengths are multiples of d. If $d = 1$, then state i is said to be aperiodic. On the other hand, if $d > 1$, then state i is said to be periodic.

Let C be an equivalence class of the state space S and suppose that whenever $i \in C$ has a particular property, it follows that the property also applies to every other state $j \in C$. Such a property is called a solidarity property or class property. It turns out that recurrence, transience, and periodicity are all solidarity properties (Resnick, 1992). For example, if $i \in C$ is recurrent, then every $j \in C$ is also recurrent. And if $i \in C$ has period d, then $j \in C$ has the same period d. If a state is both positive recurrent and aperiodic, then this state is said to be ergodic.

As $n \to \infty$, p_{ij}^n converges to some value that is the same for all i. This value is called the limiting distribution. For an irreducible ergodic Markov chain, the limiting probability $lim_{n\to\infty} p_{ij}^n$ exists and is independent of the initial state i. If we denote the limiting probability by $\pi_j = lim_{n\to\infty} p_{ij}^n$ for $j \geq 0$, then π_j is the unique nonnegative solution of the equation $\pi_j = \sum_{i=0}^{\infty} \pi_i \, p_{ij}$ for $j \geq 0$. Also, $\sum_{j=0}^{\infty} \pi_j = 1$. The limiting probability that the process will be in state j after n steps also equals the long-run proportion of time that the process will be in state j (Ross, 2009). The limiting probability is often called stationary probability.

For further information on Markov chains, please refer to Resnick's *Adventures in Stochastic Processes* and Ross' *Introduction to Probability Model.*

Chapter 2 - Our Problem

A discrete-time Markov chain with stationary transition probabilities is often used for the purpose of investigating treatment programs and health care protocols for chronic diseases. A Markov chain model is appropriate in such a situation for two reasons. First, the progression of chronic disease is often expressed in terms of different health states. The Markov chain is a simple but effective model to describe such a progression. Second, a Markov chain can be constructed in a simple way, and we can investigate its properties through matrix analysis and simulation (Craig & Sendi, 1998).

Suppose patients with a certain chronic disease are observed over equally spaced time intervals (Craig & Sendi, 1998). These intervals are called the observation intervals. If we classify the chronic disease into n distinct health states, the movement through these health states over time can then represent a patient's disease history. We can use a discrete-time Markov chain to describe such movement using the transition probabilities between the health states. In the ideal situation, the observation intervals coincide with the cycle length of the Markov chain. However, this does not happen very often in real situations. One thing to note here is that the Markov chain process simply models the health state at the end of each cycle, it does not consider the progression between cycles (Craig & Sendi, 1998).

The Markov chain model we will use for the purpose of describing chronic disease progression has state space $S = \{1,2, \dots, n\}$ representing distinct health states. The transition probability matrix P consists of transition probabilities $\{p_{ij}: i, j = 1,2, \dots, n\}$, where p_{ij} indicates the probability of a movement from health state i to health state j by the end of a cycle. According to a property of transition probabilities, $\sum_{j=1}^{n} p_{ij} = 1$ for all $i \in S$. In addition, we assume a common cycle length of the Markov chain.

Depending on the relationship between observation intervals and the cycle length of the Markov chain, different methods can be used to obtain the maximum likelihood estimate of the transition matrix.

Section 2.1 - First Case

Let us first consider the case when the common observation interval coincides with the cycle length of the Markov chain. Suppose we have a chronic disease with n distinct health states, and we would like to estimate a one-year transition matrix where the data comes from a cohort with one-year observation intervals. We first obtain the one-year observed count matrix

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$$

where c_{ij} is the number of patients moving from health state i to state j in an one-year cycle.

The maximum likelihood estimate of the transition matrix given the observed count matrix is simply the row proportions of C (Craig & Sendi, 1998). If we denote the unknown transition matrix by P, then the elements of the maximum likelihood estimate $\hat{P}$ can be expressed as

$$\hat{p}_{ij} = c_{ij} / \sum_{j=1}^{n} c_{ij}.$$

Section 2.2 – Second Case

Let us next consider the case when the common observation interval does not coincide with the cycle length. Let L_0 denote the common observation interval and L_d the cycle length of the Markov chain. $\hat{P}_0$, the maximum likelihood estimate of the transition matrix associated with L_0 , is obtained using the method described in section 2.1 for when the observation interval and cycle length coincide.

The maximum likelihood estimate of the transition matrix associated with L_d, denoted by $\hat{P}_d$, can then be expressed as:

$$\hat{P}_d = \hat{P}_0^{1/k} \text{ , where } k = \frac{L_0}{L_d}.$$

For example, supposed the common observation interval is 3 years, and the desired cycle length is 1 year. Then $k = 3$ and $\hat{P}_d = \hat{P}_0^{1/3}$. In other words, one would take the cubic root of the estimated three-year transition matrix in order to obtain the estimated one-year transition matrix.

In order to estimate P_d, we will need to compute powers of the matrix $\hat{P}_0$, so we decompose the matrix $\hat{P}_0$ into its eigenvalues and eigenvectors (Gilbert & Gilbert, 2004). Based on the decomposition, the $n \times n$ transition matrix $\hat{P}_0$ can be expressed as

$$\hat{P}_0 = BDB^{-1}$$

where B is the n by n matrix of eigenvectors and

$$D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

where λ_i is the i^{th} eigenvalue. It then follows that

$$\hat{P}_0^{1/k} = BD^{1/k}B^{-1}$$

where

$$D^{1/k} = \begin{pmatrix} \lambda_1^{1/k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^{1/k} \end{pmatrix}$$

In our simulation study, we set the cycle length of the Markov chain equal to one year, and considered the case when the observation interval coincided with the cycle length as well as the case when the observational interval and the cycle length did not coincide. In particular, we are interested in how the estimated transition matrix behaves as the ratio of observation interval and cycle length changes.

Chapter 3 - Methodology for Simulation Study

Section 3.1 – Method

Our objective for this study is to estimate the transition matrix for a Markov chain with a cycle length of one year in both the case when the observation interval coincides with the cycle length as well as the case when the observational interval and the cycle length do not coincide. In particular, we are interested in how the estimated transition matrix behaves as the ratio of observation interval and cycle length changes.

First, we created a Markov function in R to generate a Markov chain from a known $t \times t$ transition probability matrix P. This function took in three parameters: The transition probability matrix, the number of steps of the Markov chain to simulate, and the initial state of the Markov chain. We then modified the Markov function to create another function called Markovk, which generated every k^{th} step of a Markov chain from a known $t \times t$ transition matrix P. Please refer to the Appendix for R functions and R code.

Second, we generated a dataset containing information about patients' chronic disease progression. Each row in this dataset represented one patient's disease progression, which was a sequence generated from a Markov chain with known parameters. The initial state was generated randomly. We also created additional datasets with each row being a sequence generated from a Markov chain with known parameters, but taking each k^{th} observation.

The transition matrix P_0 was estimated from each dataset using the method described in Section 2.1. This is the case when the common observation interval coincides with the cycle length of the Markov chain. In order to obtain the estimated transition matrix $\hat{P}_0$, we first obtained the observed count matrix of the dataset:

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$$

where c_{ij} was the number of observed movements from state i to state j in an one-period cycle. The maximum likelihood estimate of the transition matrix given the observed count matrix is

simply the row proportions of C (Craig & Sendi, 1998). In other words, the entries of the maximum likelihood estimate $\hat{P}_0$ could be expressed as

$$\hat{p}_{ij} = c_{ij} / \sum_{j=1}^{n} c_{ij}.$$

The k^{th} step estimated transition matrix P_d was then estimated from the dataset assuming we observed every k^{th} state. This is the case when the common observation interval does not coincide with the cycle length of the Markov chain, and is described in Section 2.2.

Several problems arose in the decomposition of $\hat{P}_0$ when we were conducting the simulation. Please refer to section 3.2 for further discussion.

In our study, we investigated 12 combinations of conditions. We used two different transition probability matrices: an irreducible matrix P_1 and a reducible matrix P_2, where

$$P_1 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix} \text{ and } P_2 = \begin{pmatrix} 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

two different sample sizes: 50 and 200, and three values of observation interval: 1-year, 3-year, and 5-year. The number of observations for each subject was fixed at 5. For each of the 12 combinations, we repeated the process of generating the dataset using the known transition probability matrix and estimating the transition probability matrix 1000 times.

Finally, we compared the estimated transition matrices with the known transition matrix by looking at both their element-wise deviations and total deviations. For the element-wise deviations, we computed the average of the differences between each of the 1000 estimated transition probabilities and the known transition probabilities, along with their standard deviations. This was done for each of the p_{ij} in P. For the total deviation, we computed the sum of the absolute values of the differences between p_{ij} and $\hat{p}_{ij}$ for each of the 1000 estimated transition matrix. And then we took the average of the sums along with the standard deviation of the sums.

Section 3.2 - Problems with Estimation

Several problems arose in the decomposition of the estimated transition matrix when we were conducting the simulation. First, for even values of k, the eigenvalues of P_0 have multiple roots. For example, when $k = 2$, the 2^{nd} step estimated transition matrix $\hat{P}_d$ is expressed as

$$\hat{P}_d = BD^{1/2}B^{-1}$$

where B was the $n \times n$ matrix of the eigenvectors of $\hat{P}_0$, and

$$D^{1/2} = \begin{pmatrix} \lambda_1^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^{1/2} \end{pmatrix}$$

where λ_i was the i^{th} eigenvalue of $\hat{P}_0$. In this example, we would end up with multiple results for the 2^{nd} step estimated transition matrix $\hat{P}_d$ due to the multiple roots of $D^{1/2}$. In order to avoid this problem, we decided to only choose odd values of k.

Second, we ran into several issues with matrix decomposition of $\hat{P}_0$ while we were trying to estimate the k^{th} step estimated transition matrix $\hat{P}_d$. The first problem was that occasionally the matrix of eigenvectors of $\hat{P}_0$was singular. In other words, the $n \times n$ matrix of the eigenvectors of $\hat{P}_0$, which was denoted by B in the expression $\hat{P}_d = BD^{1/k}B^{-1}$, was singular. This was not a large issue as it only happened in $1 - 2\%$ of simulations.

A more problematic issue was when the estimated transition matrix had complex eigenvalues or eigenvectors. This can occur even when the true transition matrix has only real eigenvalues and eigenvectors. A square matrix P has characteristic polynomial $f(x) = \det(x \cdot I - P)$. The roots of the characteristic polynomial are the eigenvalues of the matrix. When we estimate the transition matrix, we essentially obtain the characteristic polynomial of the estimated transition matrix $\hat{P}$ by moving the characteristic polynomial of the true transition matrix P slightly. If the characteristic polynomial of P barely crosses the x-axis (Figure 3-1), it is likely that the characteristic polynomial of the estimated transition matrix will no longer have an intersection with the x-axis (Figure 3-2). In this case, $\hat{P}$ may have complex eigenvalues and eigenvectors, even though P has real eigenvalues and eigenvectors.

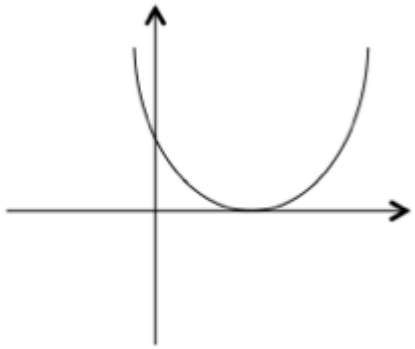

Figure 3-1. Possible characteristic polynomial for $\widehat{P}$

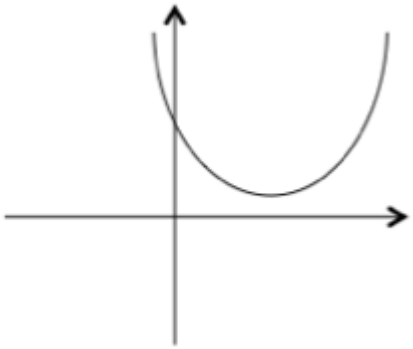

Figure 3-2. Characteristic polynomial for P

As an example, consider the transition matrix

$$P = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

The characteristic polynomial for P is $f(x) = x^4 - x^3$, which has roots 0 and 1. Figure 3-3 shows the plot of the characteristic polynomial for P.

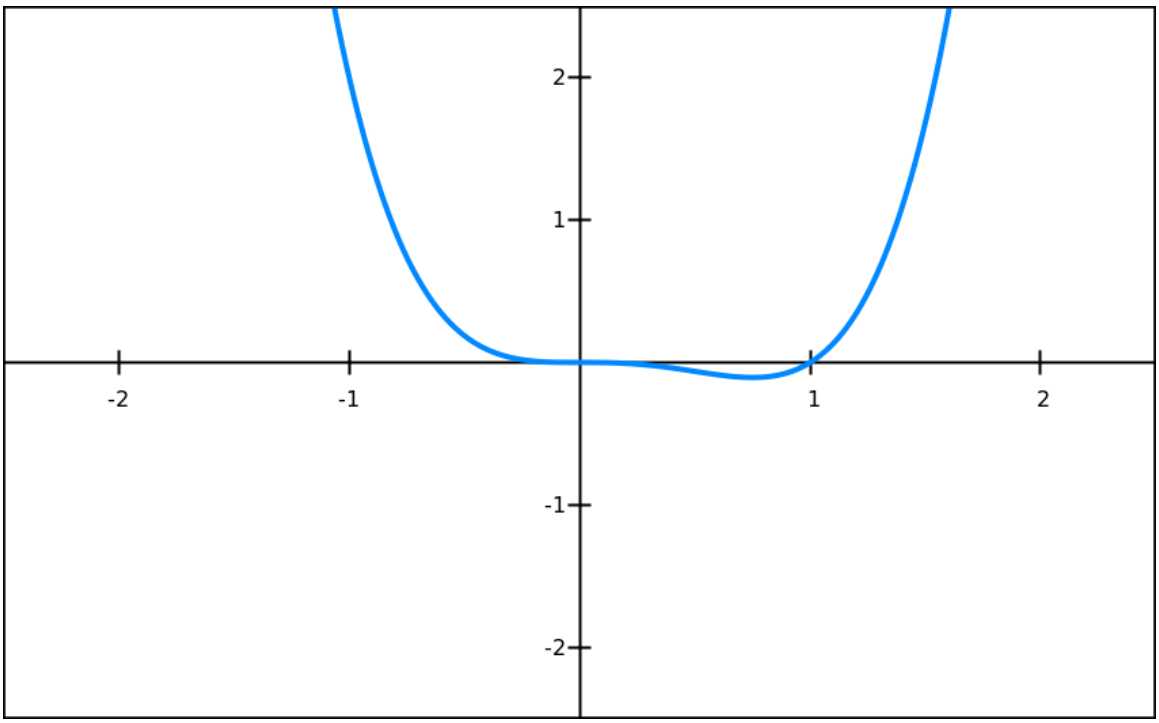

Figure 3-3. Characteristic polynomial for *P*

Since the characteristic polynomial of P barely intercepts the x-axis, when we estimate the transition matrix, the characteristic polynomial of $\hat{P}$ frequently did not intersect the x-axis. Hence we often ran into issues with $\hat{P}$ having complex eigenvalues or eigenvectors.

To avoid this issue as much as possible, we tried to pick the transition matrix P carefully. When the issue did occur in simulations, we dropped the unestimable results.

Chapter 4 - Results

In our study, 12 combinations of conditions were investigated. We used two different transition probability matrices: one 4×4 matrix and one 5×5 matrix.

$$P_1 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix} \text{ and } P_2 = \begin{pmatrix} 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

two different sample sizes: 50 and 200, and three values of observation interval: 1-year, 3-year, and 5-year. For each of the 12 combinations, we repeated the process of generating the dataset using the known transition probability matrix and estimating the transition probability matrix 1000 times. Finally, we compared each estimated transition matrix with the known transition matrix by looking at both element-wise deviations and total deviations.

Section 4.1 - Results for P_1

Below are the results for the first six combinations using the transition probability matrix P_1. Inside the parentheses we included the standard error of the deviations. Table 4-1 contains the total deviations for P_1 and Table 4-2 contains the element-wise deviations for P_1.

	K=1	**K=3**	**K=5**
N=50	Tot Dev= 0.454 (SE=0.174) unestimable=0	Tot Dev = 1.701 (SE=0.368) unestimable=7	Tot Dev = 2.9 (SE=0.927) unestimable=252
N=200	Tot Dev = 0.228 (SE=0.084) unestimable=0	Tot Dev = 1.191 (SE=0.29) unestimable=0	Tot Dev = 2.015 (SE=0.335) unestimable=11

Table 4-1. Total deviations for P_1

	N=50	N=200
K=1	$\begin{pmatrix} 0.001 & -0.001 & 0 & 0 \\ (0.074) & (0.074) & (0) & (0) \\ 0.004 & 0 & -0.004 & 0 \\ (0.07) & (0) & (0.07) & (0) \\ 0 & 0.001 & 0 & -0.001 \\ (0) & (0.071) & (0) & (0.071) \\ 0 & 0 & -0.003 & 0.003 \\ (0) & (0) & (0.071) & (0.071) \end{pmatrix}$ unestimable=0	$\begin{pmatrix} 0.002 & -0.002 & 0 & 0 \\ (0.036) & (0.036) & (0) & (0) \\ 0.0002 & 0 & -0.0002 & 0 \\ (0.035) & (0) & (0.035) & (0) \\ 0 & 0.001 & 0 & -0.001 \\ (0) & (0.036) & (0) & (0.036) \\ 0 & 0 & -0.002 & 0.002 \\ (0) & (0) & (0.036) & (0.036) \end{pmatrix}$ unestimable=0
K=3	$\begin{pmatrix} 0.012 & -0.012 & -0.006 & 0.005 \\ (0.141) & (0.152) & (0.119) & (0.111) \\ -0.007 & 0.003 & 0.012 & -0.009 \\ (0.148) & (0.114) & (0.141) & (0.116) \\ -0.006 & 0.006 & 0.007 & -0.006 \\ (0.115) & (0.143) & (0.111) & (0.152) \\ 0.005 & -0.005 & -0.011 & 0.011 \\ (0.109) & (0.118) & (0.151) & (0.142) \end{pmatrix}$ unestimable=7	$\begin{pmatrix} 0.002 & -0.001 & -0.002 & 0.001 \\ (0.096) & (0.099) & (0.079) & (0.074) \\ 0.001 & 0.0003 & 0.002 & -0.004 \\ (0.098) & (0.076) & (0.094) & (0.076) \\ -0.0003 & -0.002 & 0.0005 & 0.002 \\ (0.077) & (0.095) & (0.078) & (0.097) \\ 0 & 0.0002 & -0.001 & 0.001 \\ (0.075) & (0.079) & (0.098) & (0.094) \end{pmatrix}$ unestimable=0
K=5	$\begin{pmatrix} 0.008 & 0.0003 & -0.006 & -0.0022 \\ (0.239) & (0.241) & (0.239) & (0.22) \\ 0.008 & 0.003 & 0.013 & -0.024 \\ (0.252) & (0.222) & (0.238) & (0.235) \\ -0.023 & 0.011 & 0.004 & 0.009 \\ (0.224) & (0.204) & (0.209) & (0.249) \\ 0.014 & -0.017 & -0.025 & 0.028 \\ (0.231) & (0.231) & (0.265) & (0.242) \end{pmatrix}$ unestimable=252	$\begin{pmatrix} 0.001 & 0.005 & -0.007 & 0.001 \\ (0.158) & (0.165) & (0.15) & (0.14) \\ 0.006 & -0.007 & -0.004 & 0.006 \\ (0.161) & (0.143) & (0.151) & (0.147) \\ 0.001 & -0.003 & 0.003 & -0.001 \\ (0.148) & (0.153) & (0.138) & (0.156) \\ -0.003 & -0.0002 & 0.006 & -0.003 \\ (0.135) & (0.147) & (0.157) & (0.151) \end{pmatrix}$ unestimable=11

Table 4-2. Element-wise deviations for P_1

The results of the 6 combinations using P_1 show that when we estimated the transition probability matrix, more estimation problems arose when the sample size was 50 as the length of observational interval increased. When k=1, all 1000 transition matrices were estimable. When k=3, 0.7% were unestimable, and when k=5, 25.2% were unestimable. This makes sense because when the time between each visit of the patients gets longer, we begin to have more and more missing data. When k=1, there is no missing data at all (i.e. the cycle length and observation interval coincide), but when k=3, two thirds of the data are missing. And when k=5, four fifths of the data are missing. Hence more errors are going to occur as the length of observational interval increases, and more missing data will give a less accurate $\hat{P}_0$, and so $\hat{P}_0$ is more likely to have complex eigenvalues and eigenvectors.

We also observed that the deviation from the known transition probability matrix got larger as the length of observational interval increased. In the case when sample size was 50, the total deviation was 0.454 with standard deviation 0.174 for k=1. When k increased to 5, the total deviation increased to 2.9 with a larger standard deviation 0.927. We believe that this is also due to the missing data caused by longer time between each visit of the patients.

Moreover, with increasing sample size, there were fewer estimation problems, and the deviation from the known transition probability matrix was reduced. This is as expected, since as the sample size gets bigger, we will obtain a more accurate estimate of the known transition probability matrix.

Section 4.2 - Results for P_2

Below are the results for the other six combinations using the transition probability matrix P_2. Inside the parenthesis we included the standard error of the deviations. Table 4-3 contains the total deviations for P_2 and Table 4-4 contains the element-wise deviations for P_2.

The results of the 6 combinations using P_2 indicated that when estimating the transition probability matrix, we encountered more estimation problems when the sample size was 50 as

the length of observational interval increased. When k=1, 2.5% of the transition matrices were estimable, when k=3, 1.6% were unestimable, and when k=5, the unestimable amount of transition matrices increased to 19.6%. The reason for more estimation problems is the same as reason explained in Section 4.1 for P_1. Again it makes sense that more errors are going to occur as the length of observational interval increases.

The deviation from the known transition probability matrix also increased as the length of observational interval increased. In the case when sample size was 200, the total deviation was 0.296 with standard deviation 0.11 for k=1. When k increased to 5, the total deviation increased to 1.43 with a larger standard deviation 0.58. This is also because of the missing data caused by longer time between each visit of the patients.

Furthermore, with increasing sample size, fewer estimation problems appeared, and the deviation from the known transition probability matrix decreased. This is reasonable, since as the sample size gets bigger, we will obtain a more accurate estimate of the known transition probability matrix.

	K=1	**K=3**	**K=5**
N=50	Tot Dev=0.587 (SE=0.228) unestimable=25	Tot Dev=1.685 (SE=0.681) unestimable=16	Tot Dev=2.9 (SE=1.27) unestimable=196
N=200	Tot Dev=0.296 (SE=0.11) unestimable=37	Tot Dev=0.81 (SE=0.29) unestimable=33	Tot Dev=1.43 (SE=0.58) unestimable=26

Table 4-3. Total deviations for P_2

	N=50	N=200
K=1	$\begin{pmatrix} 0 & 0 & 0.004 & 0 & -0.004 \\ (0) & (0) & (0.121) & (0) & (0.121) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \\ -0.0002 & 0.002 & 0 & -0.0005 & -0.001 \\ (0.063) & (0.066) & (0) & (0.066) & (0.064) \\ 0 & 0 & 0.002 & 0 & -0.002 \\ (0) & (0) & (0.118) & (0) & (0.118) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \end{pmatrix}$ unestimable=25	$\begin{pmatrix} 0 & 0 & -0.0005 & 0 & 0.0005 \\ (0) & (0) & (0.059) & (0) & (0.059) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \\ 0.0002 & 0.002 & 0 & 0 & -0.002 \\ (0.032) & (0.032) & (0) & (0.032) & (0.034) \\ 0 & 0 & -0.0004 & 0 & 0.0004 \\ (0) & (0) & (0.061) & (0) & (0.061) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \end{pmatrix}$ unestimable=37
K=3	$\begin{pmatrix} 0 & 0 & -0.009 & 0 & 0.009 \\ (0) & (0) & (0.2) & (0) & (0.2) \\ 0 & 0 & -0.04 & 0 & 0.036 \\ (0) & (0) & (0.3) & (0) & (0.3) \\ -0.01 & 0.01 & 0 & -0.01 & 0.01 \\ (0.1) & (0.1) & (0) & (0.1) & (0.2) \\ 0 & 0 & -0.004 & 0 & 0.004 \\ (0) & (0) & (0.3) & (0) & (0.3) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \end{pmatrix}$ unestimable=16	$\begin{pmatrix} 0 & 0 & 0.004 & 0 & -0.004 \\ (0) & (0) & (0.12) & (0) & (0.12) \\ 0 & 0 & -0.01 & 0 & 0.01 \\ (0) & (0) & (0.13) & (0) & (0.13) \\ -0.001 & 0.001 & 0 & -0.0003 & 0.001 \\ (0.07) & (0.06) & (0) & (0.07) & (0.08) \\ 0 & 0 & 0.004 & 0 & -0.004 \\ (0) & (0) & (0.1) & (0) & (0.1) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \end{pmatrix}$ unestimable=33
K=5	$\begin{pmatrix} 0 & 0 & -0.02 & 0 & 0.02 \\ (0.0001) & (0.0004) & (0.4) & (0.001) & (0.4) \\ 0 & 0 & -0.07 & 0 & 0.07 \\ (0.002) & (0.0001) & (0.6) & (0.0008) & (0.6) \\ -0.01 & -0.01 & 0 & -0.005 & 0.03 \\ (0.3) & (0.2) & (0) & (0.3) & (0.3) \\ 0 & 0 & -0.04 & 0 & 0.04 \\ (0.005) & (0.0008) & (0.4) & (0.0001) & (0.4) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \end{pmatrix}$ unestimable=196	$\begin{pmatrix} 0 & 0 & -0.002 & 0 & 0.002 \\ (0.0006) & (0.0004) & (0.2) & (0.0008) & (0.2) \\ -0.0003 & 0 & -0.04 & 0.0003 & 0.04 \\ (0.006) & (0.0001) & (0.3) & (0.007) & (0.3) \\ -0.003 & 0.007 & 0 & -0.005 & 0.0006 \\ (0.1) & (0.1) & (0) & (0.1) & (0.1) \\ 0 & 0 & -0.02 & 0 & 0.02 \\ (0.0007) & (0.0001) & (0.2) & (0.0006) & (0.2) \\ 0 & 0 & 0 & 0 & 0 \\ (0) & (0) & (0) & (0) & (0) \end{pmatrix}$ unestimable=26

Table 4-4. Element-wise deviations for P_2

Bibliography

Craig, B. A., & Sendi, P. P. "Estimating the Transition Matrix of a Homogeneous Markov Chain." *Technical Report* # 98-12 (1998). Print.

Gilbert, J. & Gilbert, L. *Linear Algebra and Matrix Theory*. Cengage Learning, 2004. Print.

Resnick. S. I. *Adventures in Stochastic Processes*. Birkhauser, 1992. Print.

Ross. S. M. *Introduction to probability models.* Academic Press, 2009. Print.

Appendix – R Code

Markov function for generating a data set:

```
markov<-function(P,n,i) {
        m<-seq_len(n)
        m[1]<-i
        for (i in 2:n) {
                m[i]<-sample(1:ncol(P),1,prob=P[m[i-1],])
        }
m
}
```

Markovk function for generating a data set:

```
markovk<-function(P,k,n,i) {
        m<-markov(P,1+(n-1)*k,i)
        s<-m[seq(1,length(m),k)]
                s
}
```

Estimating transition matrix:

```
library(expm)
library(matrixcalc)
est.transk<-function(data,n,k){
        x<-factor(c(data[,-ncol(data)]),levels=1:n)
        y<-factor(c(data[,-1]),levels=1:n)
        trans<-table(x,y)
        trans<-trans[1:n,1:n]
        trans<-trans/rowSums(trans)
        for (i in 1:n){
                for(j in 1:n){
                        if(trans[i,j]=="NaN"){
```

```
                                trans[i,j]=1/n
                        }
                }
        }
        t.eig<-eigen(trans)
        imagval<-is.numeric(t.eig$values)
        imagvec<-is.numeric(t.eig$vectors)
        if (imagval=="FALSE" | imagvec=="FALSE"){
                mistake<-1
                transk<-matrix(0,nrow=n,ncol=n,byrow=TRUE)
        }
        else{
        sin<-is.singular.matrix(t.eig$vectors)
        if (sin=="TRUE") {
                mistake<-1
                transk<-matrix(0,nrow=n,ncol=n,byrow=TRUE)
        }
        else {
        mistake<-0
        transk<-
t.eig$vectors%*%diag(sign(t.eig$values)*abs(t.eig$values)^(1/k))%*%solve(t.eig$vectors)}
        }
        return(c(mistake,transk))
}
```

First combination, P is 4 by 4, n=50, m=5, k=1:

```
mistakes<-0
d11<-vector(,1000)
d12<-vector(,1000)
d13<-vector(,1000)
```

```
d14<-vector(,1000)
d21<-vector(,1000)
d22<-vector(,1000)
d23<-vector(,1000)
d24<-vector(,1000)
d31<-vector(,1000)
d32<-vector(,1000)
d33<-vector(,1000)
d34<-vector(,1000)
d41<-vector(,1000)
d42<-vector(,1000)
d43<-vector(,1000)
d44<-vector(,1000)
Tot<-vector(,1000)
for (r in 1:1000) {
P<-matrix(c(1/2,1/2,0,0,1/2,0,1/2,0,0,1/2,0,1/2,0,0,1/2,1/2),ncol=4,byrow=TRUE)
#### create dataset ###
data<-matrix(nrow=50,ncol=5,byrow=TRUE)
for (c in 1:50){
        i<-sample(1:4,1)
        data[c,]<-markovk(P,1,5,i)
}
### estimate transition matrix ###
est<-est.transk(data,4,1)
if (est[1]==1){
mistakes<-mistakes+1
d11[r]<-NA
d12[r]<-NA
d13[r]<-NA
d14[r]<-NA
d21[r]<-NA
```

```
d22[r]<-NA
d23[r]<-NA
d24[r]<-NA
d31[r]<-NA
d32[r]<-NA
d33[r]<-NA
d34[r]<-NA
d41[r]<-NA
d42[r]<-NA
d43[r]<-NA
d44[r]<-NA
Tot[r]<-NA
}
else {
#create estmatrix from est[2]-est[17]#
estmatrix<-matrix(est[2:17],ncol=4,byrow=FALSE)
### get the difference and total ###
d<-P-estmatrix
d11[r]<-d[1,1]
d12[r]<-d[1,2]
d13[r]<-d[1,3]
d14[r]<-d[1,4]
d21[r]<-d[2,1]
d22[r]<-d[2,2]
d23[r]<-d[2,3]
d24[r]<-d[2,4]
d31[r]<-d[3,1]
d32[r]<-d[3,2]
d33[r]<-d[3,3]
d34[r]<-d[3,4]
d41[r]<-d[4,1]
```

```
d42[r]<-d[4,2]
d43[r]<-d[4,3]
d44[r]<-d[4,4]
Tot[r]<-sum(abs(d))
}
}
### Elementwise Deviations ###
(avg<-
matrix(c(mean(d11,na.rm=T),mean(d12,na.rm=T),mean(d13,na.rm=T),mean(d14,na.rm=T),mean(d21,na.rm=T),mean(d22,na.rm=T),mean(d23,na.rm=T),mean(d24,na.rm=T),mean(d31,na.rm=T),mean(d32,na.rm=T),mean(d33,na.rm=T),mean(d34,na.rm=T),mean(d41,na.rm=T),mean(d42,na.rm=T),mean(d43,na.rm=T),mean(d44,na.rm=T)),ncol=4,byrow=TRUE))
(stddev<-
matrix(c(sd(d11,na.rm=T),sd(d12,na.rm=T),sd(d13,na.rm=T),sd(d14,na.rm=T),sd(d21,na.rm=T),sd(d22,na.rm=T),sd(d23,na.rm=T),sd(d24,na.rm=T),sd(d31,na.rm=T),sd(d32,na.rm=T),sd(d33,na.rm=T),sd(d34,na.rm=T),sd(d41,na.rm=T),sd(d42,na.rm=T),sd(d43,na.rm=T),sd(d44,na.rm=T)),ncol=4,byrow=TRUE))
### Total Deviation ###
mean(Tot,na.rm=T)
sd(Tot,na.rm=T)
```

Eighteenth combination, P is 5 by 5, n=200, m=5, k=1:

```
mistakes<-0
d11<-vector(,1000)
d12<-vector(,1000)
d13<-vector(,1000)
d14<-vector(,1000)
d15<-vector(,1000)
d21<-vector(,1000)
d22<-vector(,1000)
```

```
d23<-vector(,1000)
d24<-vector(,1000)
d25<-vector(,1000)
d31<-vector(,1000)
d32<-vector(,1000)
d33<-vector(,1000)
d34<-vector(,1000)
d35<-vector(,1000)
d41<-vector(,1000)
d42<-vector(,1000)
d43<-vector(,1000)
d44<-vector(,1000)
d45<-vector(,1000)
d51<-vector(,1000)
d52<-vector(,1000)
d53<-vector(,1000)
d54<-vector(,1000)
d55<-vector(,1000)
Tot<-vector(,1000)
for (r in 1:1000) {
P<-matrix(c(0.5,0.5,0,0,0,0.5,0,0.5,0,0,0,0.5,0,0.5,0,0,0,0.5,0,0.5,0,0,0,0.5,0.5
),ncol=5,byrow=TRUE)
### create dataset ####
data<-matrix(nrow=200,ncol=5,byrow=TRUE)
for (c in 1:200){
        i<-sample(1:5,1)
        data[c,]<-markovk(P,1,5,i)
}
### estimate transition matrix ###
est<-est.transk(data,5,1)
if (est[1]==1){
```

```
mistakes<-mistakes+1
d11[r]<-NA
d12[r]<-NA
d13[r]<-NA
d14[r]<-NA
d15[r]<-NA
d21[r]<-NA
d22[r]<-NA
d23[r]<-NA
d24[r]<-NA
d25[r]<-NA
d31[r]<-NA
d32[r]<-NA
d33[r]<-NA
d34[r]<-NA
d35[r]<-NA
d41[r]<-NA
d42[r]<-NA
d43[r]<-NA
d44[r]<-NA
d45[r]<-NA
d51[r]<-NA
d52[r]<-NA
d53[r]<-NA
d54[r]<-NA
d55[r]<-NA
Tot[r]<-NA
}
else {
#create estmatrix from est[2]-est[26]#
estmatrix<-matrix(est[2:26],ncol=5,byrow=FALSE)
```

```
### get the difference and total ###
d<-P-estmatrix
d11[r]<-d[1,1]
d12[r]<-d[1,2]
d13[r]<-d[1,3]
d14[r]<-d[1,4]
d15[r]<-d[1,5]
d21[r]<-d[2,1]
d22[r]<-d[2,2]
d23[r]<-d[2,3]
d24[r]<-d[2,4]
d25[r]<-d[2,5]
d31[r]<-d[3,1]
d32[r]<-d[3,2]
d33[r]<-d[3,3]
d34[r]<-d[3,4]
d35[r]<-d[3,5]
d41[r]<-d[4,1]
d42[r]<-d[4,2]
d43[r]<-d[4,3]
d44[r]<-d[4,4]
d45[r]<-d[4,5]
d51[r]<-d[5,1]
d52[r]<-d[5,2]
d53[r]<-d[5,3]
d54[r]<-d[5,4]
d55[r]<-d[5,5]
Tot[r]<-sum(abs(d))
}
}
### Elementwise Deviations ###
```

```
(avg<-
matrix(c(mean(d11,na.rm=T),mean(d12,na.rm=T),mean(d13,na.rm=T),mean(d14,na.rm=T),mean(d15,na.rm=T),mean(d21,na.rm=T),mean(d22,na.rm=T),mean(d23,na.rm=T),mean(d24,na.rm=T),mean(d25,na.rm=T),mean(d31,na.rm=T),mean(d32,na.rm=T),mean(d33,na.rm=T),mean(d34,na.rm=T),mean(d35,na.rm=T),mean(d41,na.rm=T),mean(d42,na.rm=T),mean(d43,na.rm=T),mean(d44,na.rm=T),mean(d45,na.rm=T),mean(d51,na.rm=T),mean(d52,na.rm=T),mean(d53,na.rm=T),mean(d54,na.rm=T),mean(d55,na.rm=T)),ncol=5,byrow=TRUE))
(stddev<-
matrix(c(sd(d11,na.rm=T),sd(d12,na.rm=T),sd(d13,na.rm=T),sd(d14,na.rm=T),sd(d15,na.rm=T),sd(d21,na.rm=T),sd(d22,na.rm=T),sd(d23,na.rm=T),sd(d24,na.rm=T),sd(d25,na.rm=T),sd(d31,na.rm=T),sd(d32,na.rm=T),sd(d33,na.rm=T),sd(d34,na.rm=T),sd(d35,na.rm=T),sd(d41,na.rm=T),sd(d42,na.rm=T),sd(d43,na.rm=T),sd(d44,na.rm=T),sd(d45,na.rm=T),sd(d51,na.rm=T),sd(d52,na.rm=T),sd(d53,na.rm=T),sd(d54,na.rm=T),sd(d55,na.rm=T)),ncol=5,byrow=TRUE))
### Total Deviation ###
mean(Tot,na.rm=T)
sd(Tot,na.rm=T)
```