Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 67(2) 2013 209-213

Predicting Course of Smoking Cessation in a Transtheoretical Model: A Markovian Approach

Asha Seth Kapadia
University of Texas School of Public Health, POB 20186, Houston, Texas 77025, USA

Received 17 September 2012; Accepted 12 March 2013

SUMMARY

This article is based on a study by Carbonari *et al.* (1999) in which they used the transtheoretical stages of change model to study the smoking cessation process in smokers. They introduced Markov matrices which are constructed for each six month interval to study the movement from one stage of cessation to the next. In this paper the matrices derived by them are used to obtain the probability distribution of the amount of time an individual spends in the various stages during the 30 month period of the study.

Keywords: Markov chains, Duration of stay.

1. INTRODUCTION

Markov Chain Processes have been extensively used in agricultural research. Harrison and Alexander (1969) of the Department of Business and Agricultural Economics at Queens University in Belfast used Markov chain methodology to project farm numbers and farm sizes structure to examine the nature of the change in the number of viable farms in Ireland. Matis *et al.* (1989) used this methodology to forecast cotton yields. Andres Kuusk (1995) developed a Markov chain model of Canopy reflectance and its angular distribution. They also studied the correlation of leaf positions in adjacent layers and how they influence the gap probability in a stand. Singh and Ibrahim (1996) used wheat production. Ramasubramanian and Jain (1999) studied use of growth indices in Markov chain models for crop yield forecasting. Reddy *et al.* (2008) used this methodology for estimating the probability of dry and wet weeks and statistical analysis of weekly rainfall as well as spell distribution for the sequence of dry and wet periods for proper management of soil and water conservation. Banik *et al.* (2009) analyzed weekly rainfall in determining drought proneness using Markov chains. More recently, Selvis and Selvraj (2011) studied variation in annual rainfall in Tamilnadu using this methodology.

Markov Methods have been used extensively in health services research and are of particular relevance to insurance companies as well as federal and local agencies who provide and control funding for programs like smoking cessation and treatment for substance addiction. Markov chains methods help clarify the path or course of treatment followed by individuals enrolled in such programs and the outcomes from such investments.

A study by Cabonari *et al.* (1999) examined the application of Markov chain analysis to model the behavior of participants in a smoking cessation program. Three hundred and eight subjects were queried at six month intervals to study where each was located on a five state scale consisting of Pre-contemplation, Contemplation, Maintenance, Preparation and Action stages associated with a typical

E-mail address: *asha.s.kapadia@uth.tmc.edu*

Transtheoretical "Stages of Change" model. The individuals were followed for a period of 30 months (2.5 years). Five separate transition probability matrices associated with a five state Markov process were developed to model the probabilities of movement in each six month period. It was assumed that flow of subjects from one stage to the other was a function of the time in the system and governed by these transition matrices. The emphasis in their paper was to evaluate and see if stage movement behavior could be modeled as a Markov chain. The authors calculated an "average" or "composite" of the five matrices and obtained a sixth matrix which was assumed to predict average movement probabilities between states in any six month period. In their study of the actual transitions between the proposed stages, they sought a system that could demonstrate orderly movement between stages.

In the Carbonari *et al.* (1999) study two matrices were derived. The first matrix represents the probabilities of movement from beginning to end of the study using Markov chain modeling for the entire 2.5 years. The second matrix represents the average time spent (in six month units) in each stage over the period of the study (2.5 years) as a function of the initial probability vector.

The purpose of the present paper is to use Markov chain methods developed by Howard (1971) to study the dynamics of the movement of the 308 individuals in the Carbonari *et al*. (1999) study. Specifically, using the transition probability matrices derived by Carbonari *et al.* (1999) we have incorporated the fluctuations in the behavior of the subjects over the course of observation. In particular, the probability distribution of the amount of time an individual spends in each stage during the 30 months in the study with or without consideration of his or her starting state is derived.

2. MARKOV CHAIN MODEL

In the Carbonari *et al.* (1999) study, initially, participants were classified into one of five entry stages. A matrix of this type was valuable in our previous study (Kapadia *et al.* 1985) on predicting the course of treatment in a rehabilitation hospital as well as our study dealing with the duration of stay in a pediatric care unit (Kapadia *et al.* 2000).

Carbonari *et al*. (1999) assumed that the probabilities that individuals occupied different states at time six months was 0.162, 0.382, 0.222, 0.112, 0.122 respectively. The transition matrices associated with each of the five time intervals 6 to 12 months, 12 to 18 months, 18 to 24 months, 24 to 30 months and 30 to 36 months are obtained. A considerable amount of effort was expended in establishing the state an individual occupied initially and the stationarity of the Markov process.

Using the results obtained by Carbonari *et al*. (1999) the initial set of probabilities associated with being in the Pre contemplations, Contemplation, Preparation, Action and the Maintenance stages group will be defined by the vector I. We will refer to this as the initial state probability vector. The elements of this vector are

$$I = (0.162 \quad 0.382 \quad 0.222 \quad 0.112 \quad 0.122)$$

For purposes of this paper we will represent the five stages of change by using the following notation.

S_1 = Pre contemplation state

S_2 = Contemplation state

S_3 = Preparation stage

S_4 = Action stage

S_5 = Maintenance stage.

According to Carbonari *et al.* (1999) the elements of the transition probability matrix $\mathbf{P}_1$ below representsthe probabilities of a transition from S_i to S_j (i, j = 1, 2 ..., 5) in the first time period (6 months to time period 12 months). We will refer to this as time period one (t = 1) and the corresponding matrix will be represented by $\mathbf{P}_1$ shown below

$$\mathbf{P}_1 = \begin{pmatrix} .540 & .280 & .080 & .100 & .000 \\ .102 & .661 & .144 & .093 & .000 \\ .088 & .235 & .559 & .118 & .000 \\ .029 & .059 & .294 & .206 & .412 \\ .000 & .000 & .053 & .105 & .842 \end{pmatrix}$$

Similarly, the four transition probability matrices representing the probabilities of transition from 12 to 18 months, 18 months to 24 months, 24 months to 30 months and from 30 months to 36 months obtained by

Carbonari *et al.* (1999) are represented by $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$ and $\mathbf{P}_5$ as follows

$$\mathbf{P}_2 = \begin{pmatrix} .630 & .196 & .044 & .130 & .000 \\ .145 & .582 & .200 & .064 & .009 \\ .056 & .394 & .480 & .070 & .000 \\ .029 & .057 & .171 & .314 & .429 \\ .000 & .000 & .065 & .044 & .891 \end{pmatrix},$$

$$\mathbf{P}_3 = \begin{pmatrix} .640 & .320 & .020 & .020 & .020 \\ .146 & .660 & .136 & .058 & .000 \\ .015 & .209 & .552 & .164 & .060 \\ .000 & .065 & .097 & .290 & .548 \\ .018 & .017 & .035 & .018 & .912 \end{pmatrix},$$

$$\mathbf{P}_4 = \begin{pmatrix} .612 & .367 & .021 & .000 & .000 \\ .198 & .574 & .109 & .119 & .000 \\ .053 & .316 & .526 & .105 & .000 \\ .000 & .036 & .143 & .321 & .500 \\ .014 & .000 & .068 & .055 & .863 \end{pmatrix},$$

$$\mathbf{P}_5 = \begin{pmatrix} .630 & .204 & .092 & .074 & .000 \\ .116 & .558 & .221 & .084 & .021 \\ .058 & .216 & .569 & .118 & .039 \\ .064 & .161 & .194 & .129 & .452 \\ .039 & .000 & .052 & .117 & .795 \end{pmatrix}$$

The element (i, j) in each of these matrices represents the probability of a transition in one time period. For the sake of simplicity we will designate the five time periods (6-12months, 12-18 months, 18-24 months, 24-30 months and 30-36 months) as time periods 1, 2, 3, 4 and 5 respectively and represent them by the variable t (= 1, 2, ...5).

If we were to multiply all the above matrices together we will obtain a matrix whose $(i, j)^{th}$ element represents the probability of moving from stage i to stage j in 5 transactions or 2.5 years of time. This resulting probability matrix is

$$\mathbf{M}_{15} = \begin{pmatrix} 0.227174 & 0.301271 & 0.209260 & 0.095873 & 0.174185 \\ 0.195360 & 0.304757 & 0.229252 & 0.097737 & 0.176310 \\ 0.175869 & 0.297569 & 0.237991 & 0.099846 & 0.197210 \\ 0.114513 & 0.173572 & 0.182228 & 0.108140 & 0.422533 \\ 0.073433 & 0.086840 & .0.136838 & 0.113826 & 0.589183 \end{pmatrix}$$

It may be observed that the sum of elements of each row in the above matrix is unity. Furthermore, the above matrix states that if an individual is initially in the Pre contemplation stage there is greater than 50% chance that he/she will either stay in the Pre Contemplation stage or move to Contemplation stage in 2.5 years. Similarly an individual who is either in the Contemplation or the Preparation stage at time six months has more than 50% likelihood of staying in these two stages. Additionally, individuals who were in the Action stage have a probability of 42% of ending up in the Maintenance stage whereas those in the Maintenance stage have almost a 60% likelihood of staying in that stage at the end of 2.5 years.

Using Markov chain theory developed by Howard (1971), the probability of an individual being in each of the five stages at time 30 months irrespective of the starting state is obtained by pre multiplying the 5 step matrix $\mathbf{M}_{15}$ obtained above by the initial state vector to obtain the probabilities

$$IM_{15} = (0.162 \quad 0.382 \quad 0.222 \quad 0.112 \quad 0.122)M_{15}$$
$$= (0.171 \quad 0.259 \quad 0.211 \quad 0.101 \quad 0.258)$$

Note that in the above vector the sum of the elements is one and after 30 months of observation an individual is most likely to be in the Contemplation or the Action stages irrespective of where the individual was initially (at time 6 months).

Just as in Howard's book (1971) and subsequently used in Kapadia *et al*. (1985, 2000) we define matrix $\phi(m)$ consisiting of elements $\phi_{ij}(m)$ such that $\phi_{ij}(m)$ = Probability an individual who was in S_i initially (at time six months) will be in S_j at $t = m$ (i, j = 1, 2, ..., 5, m = 1, 2, ..., 5) and $V_{ij}(m)$ a variable representing the number of times transitions from state i to state j were made in m time periods.

In order to be able to calculate the average time spent by an individual in the various states in five time periods we will also define $x_{ij}(m)$ an indicator variable such that

$$x_{ij}(m) = \begin{cases} 1 & \text{if an individual participant in the study who is in } S_i \text{ at a paritcular point in time will be in } S_j, m \text{ time units later} \\ 0 & \text{otherwise} \end{cases}$$

Then

$$P\,[x_{ij}\,(m) = 1] = \phi_{ij}\,(m)$$

and

$$\mathrm{E}\,[x_{ij}\,(m) = 1 \times \phi_{ij}\,(m) + 0 \times [1 - \phi_{ij}\,(m)] = \phi_{ij}(m)$$

Hence

$$\mathrm{E}[V_{ij}\,(5)] = \sum_{m=1}^{5} \mathrm{E}[x_{ij}\,(m)]$$

$$= \sum_{m=1}^{5} \phi_{ij}\,(m) = \phi_{ij}\,(1) + \phi_{ij}\,(2) \ldots \phi_{ij} \quad (5)$$

The $\phi(m)$ matrix is easily derived from the following relationship

$$\phi(m) = P_1 \ldots P_m,\ m = 1, 2, \ldots, 5$$

and the expected number of times each state is visited by individuals in five time periods is then given by

$$\mathbf{M} = \phi\,(1) + \phi\,(2) + \phi\,(3) + \phi\,(4) + \phi\,(5)$$

$$= P_1 + P_1P_2 + P_1P_2P_3 + P_1P_2P_3P_4 + P_1P_2P_3P_4P_5$$

$$\equiv \begin{pmatrix} 1.7130 & 1.6142 & 0.7000 & 0.4963 & 0.4765 \\ 0.8521 & 2.2213 & 0.9708 & 0.4881 & 0.4677 \\ 0.7040 & 1.6253 & 1.5905 & 0.5442 & 0.5360 \\ 0.3542 & 0.7753 & 1.0189 & 0.6063 & 2.2453 \\ 0.1447 & 0.2457 & 0.4967 & 0.4224 & 3.6905 \end{pmatrix}$$

= Matrix of average duration of stay in different states in 5 units of time (5 time periods).

In the above matrix, observe that the row sums add up to five for the five time units of observations. Here each time unit is of six months duration and the movement of patients is observed for five time units totaling to 2.5 years of observations. The average stay in the five states independent of the starting state (*i.e* the state an individual is in at time 6 months) may be obtained by pre multiplying the above **M** matrix by the initial starting vector, obtaining:

$$\mathbf{I} \times \mathbf{M} = (0.81662 \quad 1.58766 \quad 1.01204 \quad 0.50713 \quad .07655).$$

Elements of the above vector represent the average amount of time spent by individuals in the program in the five states irrespective of their initial state. Notice that the sum of all the elements in the above vector is 5, as it should be. Unfortunately data were not available on the actual duration of stay in the various stages to make a comparison between theoretical and actual results.

Note that here again, on average the subjects are likely to spend maximum amount of time in the Contemplation and Action stages. Since we do not have data on the history of smoking prior to joining the study, it is difficult to predict the behavior of individuals who barely started smoking prior to the observation period or individuals who had a long history of smoking.

The above results are useful in predicting behavior of individuals in the smoking cessation program observed by Carbonari *et al.* (1999). The methodology presented in this paper can be extended to study other addictive behaviors such as drug and alcohol consumption as well as gambling, internet, sex and food. The information obtained using the model presented in this paper may have important implications for the development of utilization review strategies allowing variation in the movement of patients from one stage to another. Estimation of the duration of stay in the various stages may provide healthcare providers with an estimate of cost for a rehabilitation program associated with smoking cessation or other kinds of addictions. This model may also be adopted for studying the behavior of farmers in adopting technology which may be useful and efficient for their purposes.

REFERENCES

Banik, P., Mandal, A. and Rahman, M.S. (2009). Markov chain analysis of weekly rainfall data in determining drought proneness. *Discreet Dynamics Nature Society*, **7**, 451-561.

Carbonari, Joseph P., DiClementi Carlo, C. and Sewell, Kenneth, B. (1999). Stage transitions and the trans theoretical "stages of change" model of smoking cessation. *Swiss J. Psychology,* **58(2)**, 134-139.

Harrison, H. and Alexander, D.J. (1969). Projection of farm numbers and farm size structure with Markov chains. *J. Agric. Eco. Rural Sociology*, **2(1)**, 65-103.

Howard, Ronald (1971). *Dynamic Probabilistic Systems* Vol I and II. Wiley, New York

Kapadia, A.S., Vineberg, S.E. and Rossi, C.D. (1985). Predicting course of treatment in a rehabilitation hospital: A Markovian model. *Computers Operations Res.,* **12**, 459-469.

Kapadia, A.S., Chan, W., Sachdeva, R., Moye, L.A. and Jefferson, L.S. (1999). Predicting duration of stay in a pediatric care unit: A Markovian approach. *European J. Operations Res.*, **124**, 353-359.

Kuusk, Andres (1995). A Markov chain model of canopy reflectance. *Agril. Forest Meteorology,* **76**, 221-236.

Matis, J.H., Birkett, T. and Boudreaux, D. (1989). An application of the Markov chain approach to forecasting cotton yields from surveys. *Agril. Sys.*, **29**, 357-370.

Ramasubramanian,V. and Jain, R.C. (1999). Use of growth indices in Markov chain models for crop yield forecasting. *Biometrical J.*, **41(1)**, 99-109.

Singh, Randir and Ibrahim, A.E. (1996). Use of spectral data in Markov chain model for crop forecasting. *J. Indian Soc. Remote Sensing,* **24(3)**, 145-152.

Sirinivasa, G.V., Reddy, S.R., Bhaskar, R.C., Purohit, J. and Chittora, A.K. (2008). Markov chain model probability of dry, wet weeks and statistical analysis of weekly rainfall for agricultural planning at Bangalore. *Karnataka J. Agril. Sci.*, **21(1)**, 12-16.

Tamil, Selvis and Samuel Selvaraj, R. (2011). Stochastic modeling of annual rainfall inTamilnadu. *Universal J. Environ. Res. Tech.*, **1(4)**, 566-570.