# An application of Markov chain method applied to study the smoking cessation of U.S.A adults 

A.R.C. De Vas Gunasekara, A.A.C.A. Jayathilake and A.A.I. Perera<br>Department of Mathematics, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.

## ARTICLE INFO

## Article history:

Received: 18 April 2016;
Received in revised form: 13 May 2016;
Accepted: 18 May 2016;

## Keywords

Markov Chain,
Mathematical modelling, Matrices,
Probability,
Statistics.


#### Abstract

Modern probability theory studies several processes for which the knowledge of previous outcomes influences predictions for future experiments. When we observe a sequence of chance experiments, all of the past outcomes could influence our predictions for the next experiment. In this work, our aim is to discuss the properties of the Markov Chain model applied to the data set which includes the details on smoking cessation of U.S.A. adults. In this set of data, the selected possible outcomes are, an adult being a non - smoker (A),a smoker who is interested in quitting (B),a smoker who is not interested in quitting (C). All the data was taken from CDC Morbidity and Mortality Weekly reports 2011 and 2009. Using the information given in the data set, the transition probabilities of matrix $P$ were calculated and they are $\mathrm{P}_{\mathrm{AA}}=0.951, \mathrm{P}_{\mathrm{BA}}=0.062, \mathrm{P}_{\mathrm{CA}}=0, \mathrm{P}_{\mathrm{AB}}=0, \mathrm{P}_{\mathrm{BB}}=0.524, \mathrm{P}_{\mathrm{CB}}=$ $0.879, \mathrm{P}_{\mathrm{AC}}=0.049, \mathrm{P}_{\mathrm{BC}}=0.414, \mathrm{P}_{\mathrm{CC}}=0.121$. Since column entries of matrix P add up to 1 this is a stochastic matrix (a Transition matrix). Then, the probability vector for this study was obtained and named as $\mathrm{X}_{0}, \mathrm{X}_{0}=[0.794 ; 0.0933 ; 0.1127]$; which explains the probability of non- smokers in $2008=0.794$, probability of smokers who are interested in quitting $2008=0.0933$, probability of smokers who are not interested in quitting $2008=$ 0.1127 . Furthermore, the properties of P were analyzed and regularity was determined and the equilibrium approach was calculated. Using this method smoking behavior of US adults was predicted. Our choice of transition probabilities for each outcome, lead to a regular transition matrix P. Hence, after 92 steps, the system converged to a steady state vector $\mathrm{V}=[0.4508 ; 0.3562 ; 0.1929]$. This can be seen after 184 years. Therefore, mandatory actions can be taken to prevent tobacco smoking. All the matrix calculations were implemented by MATLAB software.


© 2016 Elixir All rights reserved.

## 1.Introduction

In a physical or mathematical system that has $n$ possible states, for which the system is in one and only one of its $n$ states. As well, assume that at a given observation period, say $k^{\text {th }}$ period, the probability of the system being in a particular state depends only on its status at the $k$-1st period. Such a system is called Markov Chain or Markov process. We describe a Markov chain as follows: We have a set of states, $S=\left\{s_{1}, s_{2}, \ldots, s_{r}\right\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state $s_{\mathrm{i}}$, then it moves to state $s_{j}$ at the next step with a probability denoted by $p_{i j}$, and this probability does not depend upon which state the chain was in before the current state. The probabilities $p_{i j}$ are called transition probabilities. The process can remain in the state it is in, and this occurs with probability $p_{\text {ii }}$. An initial probability distribution, defined on $S$, specifies the starting state. Usually this is done by specifying a particular state as the starting state.

A Stochastic matrix or a Transition matrix ( $P$ ) of order $n$ is a $n \times n$ square matrix with real non negative entries, for which each column adds up to 1. (George Nakos, 1998) Transition probability $P_{i j}$ is the probability of transition from state $i$ to state $j$. In this study $i, j=A, B, C$. Probability
vector $\left(X_{0}\right)$ is a column vector with non-negative real entries that add up to 1 . Markov chain is defined as a sequence of probability vectors $X_{0}, X_{1} \ldots$ together with the stochastic matrix $P$ such that, $X_{1}=P X_{0}, \quad X_{2}=P X_{1} \ldots ; X_{k+1}=P X_{k}$ (Lay, 2003)

Markov chain is a special type of a stochastic process; where outcome of an experiment depends only on the outcome of the previous experiment. That is next state of the system depends only on the present state. This process can be used as a future predicting model.
Transition matrix is said to be regular if some powers of the transition matrix has only positive elements. (Charles Miller Grinstead)

A Markov chain is a regular Markov chain; if its transition matrix is regular. If Markov chain with transition matrix $P$ is regular, then there exist a unique vector $V$ such that for any probability vector $v$ and for large values of $n$, $v . P^{n} \approx V$. As $n\left(n^{\text {th }}\right.$ state) gets larger and larger the product $v . P^{n}$ approaches a unique vector $V$ for any initial probability vector $v$ is called equilibrium vector or fixed vector. (Friedberg, 2008) A Markov chain is called an Ergodic or Irreducible if it is possible to go from every state to each and every state

[^0]In this work, a transition matrix $(P)$ was constructed using U.S. smoking data and properties of that matrix was studied throughout this piece of work.

## 2. Materials and Methods

Our aim of this study is to analyze the properties of Transition matrix $P$ using Markov process. Entries of $P$ consist of transition probabilities of U.S.A adult smoker cessation which have been calculated using the data available in the data set. All the data was collected from CDC- Centers for Disease Control and Prevention, Morbidity and Mortality Weekly Report November 11, 2011 and November 13, 2009.
Here adult is defined as a person whose age is 18 or greater than 18. 21781 US citizens were surveyed. Using these data three possible states (outcomes) can be identified as a person who belongs to the state A is a non-smoker, person who belongs to the state $B$ is a smoker who is interested in quitting and finally a person who belongs to the state C is a smoker who is not interested in quitting. Since we have only three outcomes, the transition matrix takes only 9 entries which can be defined as follows:

1. $\mathrm{P}_{\mathrm{AA}}=$ probability of non-smoker in 2008 being a nonsmoker in 2010 as well,
2. $\mathrm{P}_{\mathrm{AB}}=$ probability of non-smoker in 2008 being a smoker who is interested in quitting in 2010,
3. $\mathrm{P}_{\mathrm{AC}}=$ probability of non-smoker in 2008 being a smoker who is not interested in quitting.,
4. $\mathrm{P}_{\mathrm{BA}}=$ probability of smoker who is interested in quitting in 2008 being a non-smoker in 2010,
5. $\mathrm{P}_{\mathrm{BB}}=$ probability of smoker who is interested in quitting in 2008 being a smoker who is interested in quitting in 2010 as well,
6. $\mathrm{P}_{\mathrm{BC}}=$ probability of smoker who is interested in quitting in 2008 being a smoker who is not interested in quitting in 2010,
7. $\mathrm{P}_{\mathrm{CA}}=$ probability of smoker who is not interested in quitting in 2008 being a non- smoker,
8. $\mathrm{P}_{\mathrm{CB}}=$ probability of smoker who is not interested in quitting in 2008 being a smoker who is interested in quitting in 2010,
9. $\mathrm{P}_{\mathrm{CC}}=$ probability of smoker who is not interested in quitting in 2008 also being a smoker who is not interested in quitting in 2010,

From the data set, we were able to identify the following values as our transition probabilities for all the transition states mentioned above.
Then, the matrix $P$ can be written as;

2010

$$
P=\left[\begin{array}{ccc}
A & 2008 \\
0.951 & 0.062 & C \\
0 & 0.524 & 0.879 \\
0.049 & 0.414 & 0.121
\end{array}\right]_{C}^{A} B
$$

Moreover, the initial probability vector $X_{0}$ contains the probability values recorded in the previous stage and in this work, $X_{0}$ takes the probability values of non-smokers (A), the smokers who interested in quitting (B) and smokers who are not interested in quitting (C) found in 2008 and it can be written as $X_{0}=[0.794,0.0933,0.1127]^{T}$.

Also, the computing the highest powers of the transition matrix $P$ shows that $P$ is a regular matrix as its second power results all entries of the transition matrix with non-zero positive numbers. These calculations were performed by MATLAB.

Consider


As $P^{2}$ has strictly positive entries $P$ is a regular transition matrix. Thus, we can obtain a steady state vector/equilibrium vector for this $P$.

As $\mathrm{P}^{2}$ has strictly positive entries P is a regular transition matrix. Thus, we can obtain a steady state vector/equilibrium vector for this $P$.

Let $P$ be the transition matrix, $X_{0}$ be the initial probability vector, then $X_{1}$ is the probability vector after $1^{\text {st }}$ step and $X_{0}=X_{1}$.
$\left[\begin{array}{ccc}0.951 & 0.062 & 0 \\ 0 & 0.524 & 0.879 \\ 0.049 & 0.414 & 0.121\end{array}\right] \times\left[\begin{array}{c}0.7940 \\ 0.0933 \\ 0.1127\end{array}\right]=\left[\begin{array}{c}0.7609 \\ 0.1480 \\ 0.0912\end{array}\right]$
${ }^{\text {Then, }} X_{1}=\left[\begin{array}{l}0.7609 \\ 0.1480 \\ 0.0912\end{array}\right]$.That is after one stage probabilities of non-smokers, smokers who are interested in quitting and smokers who are not interested in quitting are respectively $0.7609,0.1480$ and 0.0912 . The continuous iterations results the following probability values for each outcome $A, B$ and $C$.

Table 1. Iterative scheme of probability values for each outcome.

| $\boldsymbol{k}$ | Probability value of each outcome |  |  |
| :--- | :--- | :--- | :--- |
|  | $\boldsymbol{X}_{\boldsymbol{k}}(\boldsymbol{A})$ | $\boldsymbol{X}_{\boldsymbol{k}}(\boldsymbol{B})$ | $\boldsymbol{X}_{\boldsymbol{k}}(\boldsymbol{C})$ |
| 1 | 0.7609 | 0.148 | 0.0912 |
| 5 | 0.6617 | 0.2096 | 0.1288 |
| 10 | 0.5808 | 0.2658 | 0.1534 |
| 15 | 0.531 | 0.3005 | 0.1686 |
| 20 | 0.5003 | 0.3218 | 0.1779 |
| 25 | 0.4813 | 0.335 | 0.1837 |
| 30 | 0.4696 | 0.3432 | 0.1872 |
| 35 | 0.4624 | 0.3482 | 0.1894 |
| 40 | 0.4579 | 0.3513 | 0.1908 |
| 45 | 0.4552 | 0.3532 | 0.1916 |
| 50 | 0.4535 | 0.3544 | 0.1921 |
| 55 | 0.4525 | 0.3551 | 0.1924 |
| 60 | 0.4518 | 0.3556 | 0.1926 |
| 65 | 0.4514 | 0.3558 | 0.1927 |
| 70 | 0.4512 | 0.356 | 0.1928 |
| 75 | 0.451 | 0.3561 | 0.1929 |
| 80 | 0.4509 | 0.3562 | 0.1929 |
| 85 | 0.4509 | 0.3562 | 0.1929 |
| 90 | 0.4509 | 0.3562 | 0.1929 |
| 91 | 0.4508 | 0.3562 | 0.1929 |
| 92 | 0.4508 | 0.3562 | 0.1929 |
| 93 | 0.4508 | 0.3562 | 0.1929 |
| 95 | 0.4508 | 0.3562 | 0.1929 |
| 97 | 0.4508 | 0.3562 | 0.1929 |
| 100 | 0.4508 | 0.3562 | 0.1929 |

## 3. Results and Discussion

After observing transition matrix $P$ it can be deduced that the transition probabilities of state C to state A and state A to state B are 0 . Hence, this Markov chain is not irreducible. Therefore the Markov chain applied for $P$ is not an ergodic. Since $P$ is a regular transition matrix, the complete Markov chain regarding $P$ is also regular. Thus a steady state vector can be obtained by considering the iterative calculations for the probability vectors after $k$ steps;

$$
X_{1}=P X_{0}, X_{2}=P X_{1} \ldots ; X_{k+1}=P X_{k}, \ldots \text { and hence, }
$$ we have $X_{k}=P^{k} X_{0}$. The following graph shows the variation of these probability vectors.



Figure 1. Graph of the probability vectors with respect to iterative scheme obtained in Table 01.
From the above graph, we can see that the probability values for non-smokers (A) have reduced and became to an equilibrium value after $91^{\text {st }}$ iteration resulting a $4^{\text {th }}$ order polynomial. Moreover, the probability values for both the smokers who interested in quitting (B) and smokers who are not interested in quitting (C) have increased and reached to an equilibrium state after $80^{\text {th }}$ and $75^{\text {th }}$ iteration and resulted $4^{\text {th }}$ order and $6^{\text {th }}$ order polynomials respectively. Hence, there is an equilibrium state $(0.4508,0.3562,0.1929)$ for the system after $92^{\text {nd }}$ step of the iteration.

## 4. Conclusion

Since our aim is to determine whether the system has a steady state, we need to determine whether the transition matrix is a regular or not. By the above choice of transition probabilities for each outcome, we could obtain a regular transition matrix $P$. Hence, further calculations were
proceeded. It can be seen that after 92 steps system converged to a steady state vector $V=[0.4508 ; 0.3562 ; 0.1929]$. Since, the data points are analyzed once in two years, after $\mathbf{9 2} * \mathbf{2}=\mathbf{1 8 4}$ years, the percentages of smokers who are interested in quitting (B), smokers who are not interested in quitting (C) and non-smokers (A) are a constant, which is equals to ${ }^{V}$. According to these results authorities can take actions to prevent tobacco smoking among US adults.
This method can be extended to study smoking behavior of any country or specific groups within a country such as gender base, educational level, ethnicity etc.

## References

Charles Miller Grinstead, J. L. Introduction to probability.
Friedberg, S. I. (2008). Elementary Linear Algebra, A matrix approach. Pearson international edition.
George Nakos, D. J. (1998). Linear Algebra with applications.
Lay, D. C. (2003). Linear Algebra and its applications. Addison Wesley .
Malarcher, A. $(2009,2001)$. Morbidity and Mortality Weekly Report USA: CDC - Center for disease control and prevention. Malarcher, A. $(2009,2011)$. Morbidity and Mortality Weekly Report. USA: CDC - Centers for Disease Control and Prevention.
Malarcher, A. (2011,2009). Morbidity and Mortality Weekly Report. USA: CDC - Centers for Disease Control and Prevention.


[^0]:    Tele:
    E-mail address: ajani.gunasekara@gmail.com

