Prediction of Students' Grades in Near Future Based on Markov Chains

Ank Phyu Win
Faculty of Computing, Computer University (Mandalay)
ankphyu08@gmail.com

Abstract

*Using the stochastic processes called Markov Chains are usually used in modelling many practical problems. In this paper, to predict the immediate future, the data sets for Computer University (Mandalay) are sought. The moving averages for the data are found and grouped them into five different states for the grades. Then Markov chains calculations are applied to the data to create a 5*5 transitional probability matrix. Using this transition matrix, a system of equations are solved by using Chapman-Kolmogorov Equations and found five steady states that will be the same in every row of the matrix. They are variables that represented the probability that a grade for a given year would fall into one of the five states. When this information is used, the actual data to these equations can be applied and the next grade for the near future be predicted. Near future grades using this method are able to be successfully predicted.*

Keywords: Chapman – Kolmogorov Equations, Markov Chains, Stochastic process, Steady State probabilities, Transition Probabilities.

1. Introduction

The realization of the importance of education has increased among the public and as a result, the traditional formal education has changed in recent years. The evaluation of students' progress is a very important part of any educational system. Computer University (Mandalay) follows a regular semester system. Each semester consists of approximately 16 weeks and each year includes two semesters. A typical student takes about 5 years to complete the required credit hours. In this paper, future year's worth of grades for Computer University (Mandalay) is analysed and moving averages are applied to predict the grades. Firstly, an approximate evaluation of data was created. To find the difference data set to apply Markov Chains was needed. These differences were going to be what Markov Chains was applied too. Using Markovian properties to our data, a system of equations with the unknown variables being the steady states that is aiming to obtain. These equations are sums of probabilities multiplied by unknown variables. Then, to solve the system of equations to find steady state probabilities is aimed.

2. Methodology

The grades are often used as a measure of transmission of knowledge in a course. A random sample of students was selected from Computer University (Mandalay). The data pertains to a period of 3 years from the academic years 2015-2016 to 2017-2018. The sample size for each year was determined in proportion to the total number of students in each year and the sample from each year was divided into 5 groups proportionally to the number of students in each level.

2.1. The Conceptual Framework: Markov Chains

Some background concerning the Markov Chains is presented. A Markov Chain is a stochastic process that has the Markovian property. Consider a finite discrete time homogeneous stochastic process with index set $Z^+ = \{0,1,2,....\}$; that is, a sequence $\{X_n : n \in Z^+\}$ of random variables. As usual the subscript n in X_n stands for the time and X_n denotes the state of the process at time n. If $X_n \in S$, then S is called state space of the stochastic process considered here satisfy the Markov property. Given the present state, the future of the process is independent of the past. That is, for $i, j;\ x_0,.., x_{n-1} \in S, P\big(X_{n+1} = j \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i \big) = P(X_{n+1} = j \mid X_n = i\big) = P_{ij}$

Conditional probabilities for Markov Chains are called transition probabilities. A stochastic process with this property is called a homogeneous Markov chain. The quantity P_{ij} stands for the probability of moving from state i to state j in just one transition and all these quantities define the matrix of one-step transition probabilities P:

$$P^{(n)} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \ldots & P_{1k} \\ P_{21} & P_{22} & P_{23} & \ldots & P_{2k} \\ P_{31} & P_{32} & P_{33} & \ldots & P_{3k} \\ & & \vdots & & \\ P_{k1} & P_{k2} & P_{k3} & \ldots & P_{kk} \end{bmatrix}$$

where the finite set $I_k = \{1,2,\ldots,k\}$ is the state space of the Markov chains. The entries P_{ij} of the matrix P must satisfy : (1) $P_{ij} \leq 0$, (2) $\sum_j P_{ij} = 1, i, j \in I_k$. The n-step transition probability $P_{ij}^{(n)}$is just the conditional probability that the system will be in state j after exactly n steps (time units) given that it starts in state i at any time t. When n=1, note that $P_{ij}^{(1)} = P_{ij}^{1}$.[1] Therefore, a Markov Chain is a stochastic process that states that the conditional probability of a future event relies on the present state of the process, rather than any past states or events.

A conventional way to note stationary transition probabilities that will be seen later in this paper is,

$$P_{ij} = P\big(X_{t+1} = j \mid X_t = i\big),$$
$$P_{ij}^{(n)} = P\big(X_{t+1} = j \mid X_t = i\big)$$

2.2. Chapman-Kolmogorov Equations

Chapman-Kolmogorov equations is used to provide a method to compute all of the n-step transition probabilities.

$$P_{ij}^{(n)} = \sum_{k=0}^{M} P_{ik}^{(m)} P_{kj}^{(n-m)}\ for\ all\ i, j = 0,1,..,M$$
$and\ any\ m = 1,2,\ldots,n-1, n = m+1, m+2, \ldots.$ [2]

These equations are used to point out that when from steady state to another in n steps is gone, the process will be in some other state after exactly m (m is less than n) states.

Thus, the summing up is just the conditional probability that, given a starting point in one state, the process goes to the other state after m steps and then to the next state in n-m steps. Therefore, by summing up these conditional probabilities over all the possible steady states must yield

$$P_{ij}^{(n)} = \sum_{k=0}^{M} P_{ik} P_{kj}^{(n-1)}$$

and $P_{ij}^{(n)} = \sum_{k=0}^{M} P_{ik}^{(n-1)} P_{kj}$

This means that these expressions allow to obtain the n steps probabilities from the one step transition probabilities recursively.

2.3. State Transition Diagrams

A convenient and useful method to visualize the state of Markov Chains when they have stationary transition probabilities and a finite number of states is through the use of a state transition diagram. In such diagram, each state of a Markov chains is drawn as a numbered node, and the conditional probability of moving from one state to another is drawn by connecting the nodes with an edge and labelling the edge with the numbered probability. For example [3], if the stock in a given day is recorded by the following stochastic process where:

$$\{X_t\}\ for\ t = 0,1,2,\ldots$$

$$X_t = \begin{cases} 1\ if\ stock\ t\ has\ gone\ up \\ 2\ if\ stock\ t\ has\ gone\ down \end{cases}$$

And $P\big(X_{t+1} = 1 \mid X_t = 1\big) = 0.7$
$P\big(X_{t+1} = 2 \mid X_t = 1\big) = 0.4$

Then the state transition diagram for this example is:

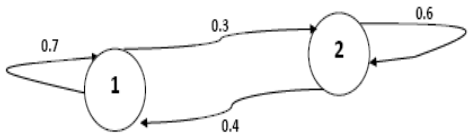


Figure 1. State transition diagram

2.4. Categorizing States of Markov Chains

Since Markov Chains are long run stochastic processes that include transitional probabilities which indicate the likelihood the process will move from one state to another, it is often necessary to categorize, or classify, the varying types of states.

2.5. Properties of Markov Chains in Long- Run

After n-step transition probabilities for a Markov chain have been calculated, the Markov chain will display the characteristic of a steady state. Meaning that, if the value of n is large enough, every row of the matrix will be the same, and such the probability that the process will be in each state k after a certain number of transitions is a limiting probability that exists independently of the initial state. This can be defined as:

For any irreducible ergodic Markov chains, $\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j > 0$ where, the π_j uniquely satisfy the following steady-state equations

$$\pi_j = \sum_{i=0}^{M} \pi_j \; P_{ij}\,, for\ j = 0,1,..,M$$

$$\sum_{j=0}^{M} \pi_j = 1$$

The steady state probabilities of the Markov chain, π_j indicate that after a large number of transition the probability of finding the process in a particular state.

$$P\{X_n = j\} = \pi_j$$

3. Procedural Overview

In this section, the effectiveness of the Markov chain model is demonstrated and it is applied to the students' grades for prediction near future in Computer University (Mandalay). The entries of grades of students in CU(Mandalay) for three years are used. The grades can be classified into five possible states (A,B,C,D,E) that are in the interval (0-100). The grades are expressed as A=very high (81-100), B=high(61-80),C=middle(41-60), D=low (21-40), E=very low (0-20). The data set of grades chosen for this project is three years' worth of grades for CU(Mandalay).

Once the grades were found, the first step towards applying Markov chains to the data set began with the calculation of moving averages. Moving averages provide a forecast for future grades and therefore are crucial to work here. Using the difference between the forecasts each year enables to make the predictions for the possibility of where future grades may lie.

3.1 Procedure with Data

In order to better understand the process of using Markov chains and how the data is reached, this section explains the procedure with a corresponding data set so as to better work. Once the data set is got, the corresponding frequencies are calculated[4].

Table 1 displays a (5×5) frequency matrix for grades in 2015-2016 academic year (1192 students). Another two frequency matrices for grades in 2016-2017 (1006 students) and 2017-2018 (866 students) academic year are displayed in Table 2 and Table 3 respectively.

Table 1. Frequency transition data (2015-16)

Grade	A	B	C	D	E
A	63	60	22	5	0
B	52	95	88	36	9
C	21	91	128	99	17
D	8	33	100	125	46
E	1	5	17	47	24

The numbers in the table show how many students moved from one state to another. For example, in the 1st row,1st column in Table 1, number 63 means the number students moved from Grade A to Grade A. For the purpose of computation, the transition matrix of Markov chain is required. The transition matrix, however, works with probabilities and not with frequencies as described in each Table.

Table 2. Frequency transition data (2016-17)

Grade	A	B	C	D	E
A	48	55	21	7	1
B	49	104	82	27	10
C	23	78	103	59	13
D	7	27	55	107	39
E	2	7	16	35	31

Table 3. Frequency transition data (2017-18)

Grade	A	B	C	D	E
A	90	56	29	13	1
B	56	76	62	21	2
C	27	61	79	51	9
D	14	15	43	73	29
E	3	2	14	21	17

3.2 Transition Probabilities

It is necessary to transform the matrix of frequencies into the proper transition matrix [5]. This can be done easily by normalizing each row of the matrix by

$$P_{ij} = \frac{f_{ij}}{\sum_j f_{ij}}$$

where f_{ij} acts as an element of the frequency matrix (Table 1,Table 2, Table 3) and p_{ij} is an element of the transition matrix. After the transitions were labelled with the appropriate intervals, the transition matrix was built and the ensuring linear system of equations was solved to find the steady states.

Table 4, Table 5 and Table 6 show the probability of transition matrix for the students' grades for three academic years.

Figure 2, Figure 3 and Figure 4 also show the probability of transition in state transition diagram for the students' grades for three academic years.

Table 4. Transition probabilities in grades (2015-16)

Grades	A	B	C	D	E
A	0.42	0.4	0.1467	0.0333	0
B	0.1857	0.3393	0.3143	0.1286	0.0321
C	0.0590	0.2556	0.3596	0.2781	0.0478
D	0.0256	0.1058	0.3205	0.4006	0.1474
E	0.0106	0.0532	0.1809	0.5000	0.2553

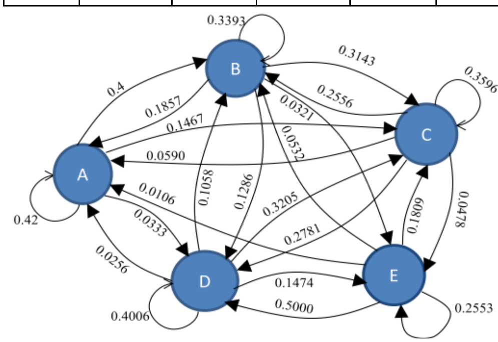


Figure 2. State transition diagram for students' grades in 2015-2016 academic year

Table 5. Transition probabilities in grades (2016-17)

Grades	A	B	C	D	E
A	0.3636	0.4167	0.1591	0.0530	0.0076
B	0.1801	0.3824	0.3015	0.0993	0.0368
C	0.0833	0.2826	0.3732	0.2138	0.0471
D	0.0298	0.1149	0.2340	0.4553	0.1660

E	0.0220	0.076 9	0.175 8	0.384 6	0.340 7

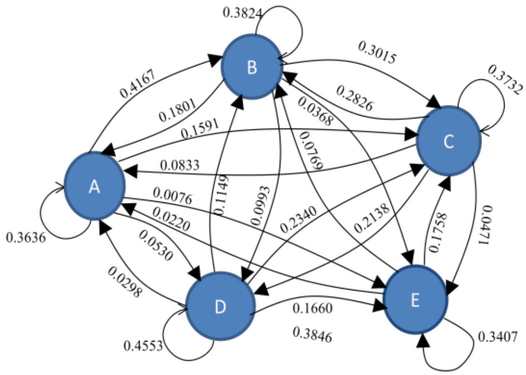


Figure 3. State transition diagram for students' grades in 2016-2017 academic year

Table 6. Transition probabilities in grades (2017-18)

Grad e	A	B	C	D	E
A	0.476 2	0.296 3	0.153 4	0.068 8	0.005 3
B	0.258 1	0.350 2	0.285 7	0.096 8	0.009 2
C	0.118 9	0.268 7	0.348 0	0.224 7	0.039 6
D	0.080 5	0.086 2	0.247 1	0.419 5	0.166 7
E	0.052 6	0.035 1	0.245 6	0.368 4	0.298 2

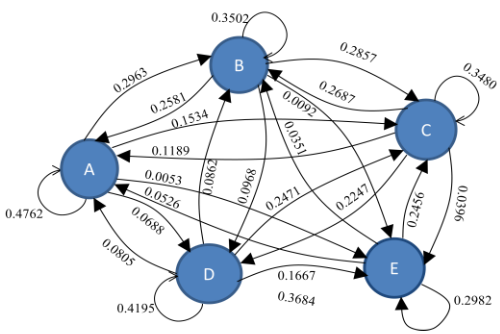


Figure 4. State transition diagram for students' grades in 2017-2018 academic year

4. Results

After the probability of transition matrix was built and the ensuring linear system of equations was solved to find the steady states. For example, for 2015-2016 academic year,

$$P^2 = \begin{bmatrix} 0.2602 & 0.3447 & 0.2508 & 0.1196 & 0.0248 \\ 0.1632 & 0.2851 & 0.2939 & 0.2048 & 0.0531 \\ 0.1011 & 0.2342 & 0.3161 & 0.2701 & 0.0786 \\ 0.0611 & 0.1783 & 0.3073 & 0.3378 & 0.1154 \\ 0.0405 & 0.1350 & 0.2898 & 0.3855 & 0.1492 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} 0.1193 & 0.2366 & 0.2986 & 0.2640 & 0.0816 \\ 0.1193 & 0.2366 & 0.2986 & 0.2640 & 0.0816 \\ 0.1193 & 0.2366 & 0.2986 & 0.2640 & 0.0816 \\ 0.1193 & 0.2366 & 0.2986 & 0.2640 & 0.0816 \\ 0.1193 & 0.2366 & 0.2986 & 0.2640 & 0.0816 \end{bmatrix}$$

Table 7. Steady- state probabilities in grades

Grade s	2015-16	2016-17	2017-18
A	0.1193	0.1259	0.2167
B	0.2366	0.2669	0.2390
C	0.2986	0.2754	0.2621
D	0.2640	0.2364	0.2113
E	0.0816	0.0956	0.0700

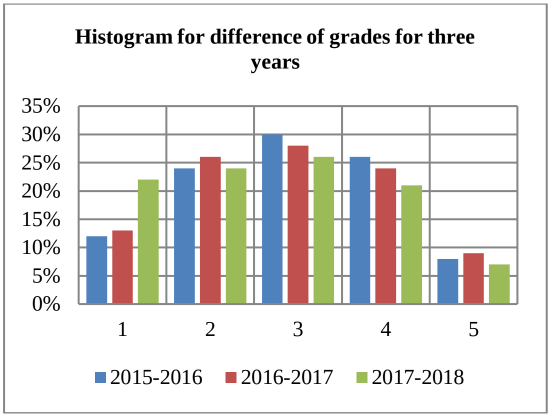
And then calculate until the steady state probability is reached. And finally, the steady state probabilities are resulted given in Table 7 for three academic years.

The steady states provide the probability that the difference will be in any of these academic year. Accordingly, the percentage of steady state in grades is shown in Table 8.

Table 8. Percentage in steady state probabilities in grades

Grade	2015-16	2016-17	2017-18
A	12%	13%	22%
B	24%	26%	24%
C	30%	28%	26%
D	26%	24%	21%
E	8%	9%	7%

After all, a histogram was created to provide the information got from the calculations.



5. Conclusion

The overall analysis of the results can be summarized as follows. 12%, 13%, and 22% chances that the grade will be in this interval (81-100), 24%, 26%,and 24% chances will be in this interval (61-80), 30%, 28%, and 26% chances will be in this interval (41-60), 26%, 24%, and 21% chances will be in this interval (21-40) and 8%, 9%, and 7% chances will be in this interval (0-20). Based on the previous result, the near future result will be: Grade A will be around the interval (10% - 20%), Grade B will be around the interval (20% - 25%), Grade C will be around the interval (25% - 30%), Grade D will be around the interval (20% - 25%) and Grade E will be around the interval (5% - 10%). Thus, the students in CU (Mandalay) will be more interested to study computing near future.

Acknowledgements

This paper is guided and supported by Dr. Thin Thin Naing, Head of Faculty of Computing, in Computer University (Mandalay).

References

[1] Frederick S. Hillier, *Introduction to Operations Research, Seventh Edition* (Stanford University), January 2000,

[2] Frederick S. Hillier and Gerald J. Lieberman, *Introduction to Operations Research, Eighth Edition*(New York: McGraw Hill,2005)

[3] Shuchi S.Mitra, Michael J.Riggieri, "*Predicting Stock Prices*", A Major Qualifying Project Report, January 2011.

[4] Fawaz S. Al-Anzi, Dia M. AbuZeina, "*A Survey of Markov Chain Models in Linguistics Applications*", November 2006.

[5]Robert Hlavaty and Ludmila Domeova, " *Students' Progress throughout Examination Process as Markov Chain",* ISSN 1913-9020 E-ISSN 1913-9039.