

Application of Markov chain in students' assessment and performance: a case study of School of Mathematical Sciences, one of the public university in Malaysia

Khairun Husna Yahaya^{1*}, and Husna Hasan¹

¹School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia

Abstract. The quality education is an essential element in economic, political and social development of any country. Therefore, enrollment forecasting is needed in higher education to assist the universities in the preparation of their educational frameworks including budgeting, provide all necessary facilities and planning the overall short and long term goals. This research study the pattern of students' assessment and their academic performance in School of Mathematical Sciences, Universiti Sains Malaysia. The target population is all undergraduate enrollment from 2016/2017 until 2018/2019 sessions. An absorbing Markov chain model is applied to study the absorption, retention and repetitive rates of the students by the academic programs and gender. The fundamental matrix is constructed to determine the expected duration of schooling before graduating. The enrollment projection is also estimated to study the probability of the performances of the students in the long run. In summary, this research addresses on the use of Markov chain model to describe the stochastic pattern of the enrollment and assessment of the students.

1 Introduction

Education plays an important role to create better future in our life. Nowadays, the student's academic performance and graduation rates are the greatest concern in higher education. Over the years, mathematical models had been developed by various institutes to help in educational planning. Educational planning is a crucial element in improving the quality of the higher education system. Enrollment forecasting is one the greatest significant in educational planning. According to [1], the tertiary education will keep increasing from years to years. Data from the World Bank shows that the gross enrollment in tertiary education increased from 19% in 2000 to 38% in 2017 with the female enrollment ratio to the male by 4%.

*Corresponding author: khusnayahaya@gmail.com

1.1 Background of the study

A Markov chain is a special type of stochastic model. It is a conditional distribution of the future events depend solely on the present event. In addition, the Markov chain model is capable in dealing with more than one possible reality of how the process might evolve across time. The Markovian property is characterized as memoryless property which means that a certain event happened regardless how much time has elapsed. The sequence data of the students' performance state is suitable to be fitted to this Markov chain model. Later, it will be used to forecast the number of students and their performance in the future.

1.2 Research objective

The research objectives of this study are as follows. The first objective of this paper is to implement of the Markov chain model in describing the stochastic pattern of enrollments and assessment of students. Next, the model is used to study the retention rate, repetitive rate, absorbing (graduate and withdrawn) rates of the students and to determine the expected length of students' stay before graduating according to different academic programs and gender. Finally, this study is to measure the probabilities of the future performances of the students by the academic programs.

2 Literature review

Shah and Burke [2] proposed the Markov chain model of the movement of undergraduates through the higher education system in Australia according to age and the gender of the students and the fields of study undertaken. Nyandwaki et al. [3] showed the Markov chain techniques in studying progression of secondary school student from the time of enrollment/entry in form one to graduation after the expected four years in Kenya's school level of education. Adeleke et al. [4] presented the pattern of students' enrollment and their academic performances in the Department of Mathematical Sciences at Ekiti State University. Nicholls [5] explored the use of absorbing Markov chain model to facilitate planning in the Doctor of Business Administration Program in a Graduate School in Australia. Bairagi and Kakaty [6] presented a stochastic process approach to analyze students' performance of Nagaon District in Gauhati University, Assam in India. Egbo et al. [7] used Markov chain approach for projection of secondary enrollment and projection of teacher. They obtained estimation of expected wastage, expected length of stay, the variance and standard deviation of length of stays and calculation of the probabilities of attaining higher grades. Walde [8] modeled the triple absorbing Markov chain in which the students who withdrawn from the system and academic dismissal were separately and grouped into three absorbing states. In conclusion, the Markov chain model has been widely used in education field. All these journals focus on the use of absorbing finite Markov chain techniques in predicting the enrollments for an education system and study the progression of performances and admission of students.

3 Methodology

The data consists of repeated, promoted and withdrawn status of the students enrolled in the first year, second year, third year and final year for 3 academic sessions (2016/2017) - (2018/2019) were obtained from the Student and Record, Academic Manage Division, USM. Based on these data, the states of the process are defined. The data are then

categorized according to the states and later transformed into the transition probability matrix.

3.1 Markov chain

According to certain probabilistic theory, a Markov chain is a mathematical system that undergoes transition process from one state to subsequent state. Specifically, Markov chain can also be defined as the process that at any given time t , the probabilities of all future state X_{t+1} only depend on the current state X_t and do not depend on any past state X_1, X_2, \dots, X_{t-1} . For any positive integer t , $\{X_t : t \in Z^+\}$ where $Z^+ = \{0, 1, 2, \dots\}$ and possible state of the random variable

$$p_{ij} = P(X_{t+1} = j | X_1 = x_1, \dots, x_t = i) = P(X_{t+1} = j | X_t = i). \tag{1}$$

The value p_{ij} represents the probability that the process will, when in state i , next make a transition into state j .

Since the behavior of a Markov chain at time $t+1$ depends only on its current behaviour at time t and does not depend on the past behaviour of time $t-1$, the chain can be described completely by its one step transition probability and this is called the Markovian property. These transition probabilities can be grouped together into a matrix known as transition probability matrix. These one step transition probabilities does not depend on time. A stochastic process with this property is called a discrete-time Markov chain (DTMC)[9].

3.2 States in a system

A Markov chain is an absorbing chain if it has at least one absorbing state and from any non-absorbing state, it is possible to go to non-absorbing state in a finite number of steps. In an absorbing Markov chain, the non-absorbing state is called transient. In this study, the states are defined in the Table 1.

Table 1. Model notation.

States	Description of States
1	The students in first year level
2	The students in second year level
3	The students in third year level
4	The students in fourth year level
W	The students withdrawn due to dropouts or academic dismissal before attaining the maximum qualification
G	The student graduating after attaining the maximum qualification

3.3 Transition probability matrix

Assuming the multinomial distribution, the transition probabilities are estimated by:

$$p_{ij}^{t,t+1} = \frac{n_{ij}^{t,t+1}}{n_i^t}, \quad i, j = 1, 2, \dots, N \tag{2}$$

where $n_{ij}^{t,t+1}$ represent the number of student in class i at time t who later will be in class j at time $t+1$ and n_i^t represent the number of student in class i at time t . Thus, the one step transition matrix with element p_{ij} is

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & 0 & 0 & 0 & p_{1W} \\ 0 & p_{22} & p_{23} & 0 & 0 & p_{2W} \\ 0 & 0 & p_{33} & p_{34} & 0 & p_{3W} \\ 0 & 0 & 0 & p_{44} & p_{4G} & p_{4W} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

Assuming time homogeneity, the transition probability matrix \mathbf{P} with the transient and absorbing states are organized into the following canonical form:

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{4}$$

where \mathbf{Q} is a transition probability matrix between non-absorbing state, \mathbf{R} is a transition probability matrix from non-absorbing states to absorbing state, $\mathbf{0}$ is a transition probability matrix from an absorbing state to non-absorbing state and \mathbf{I} is an identity matrix which gives transition probabilities between non-absorbing states.

3.4 The n -step transition probability matrix and absorbing rates

Using the Chapman- Kolmogorov equations, the iterated multiplication of the \mathbf{P} yields

$$\mathbf{P}^n = \mathbf{P}^1 * \mathbf{P}^{(n-1)}, \tag{5}$$

where \mathbf{P}^n is the n -step the transition probability matrix which consists of predicted future performance after n years later, $\mathbf{P}^1 = \mathbf{P}$ denote the initial transition probability matrix which contain the current enrollment of students and $\mathbf{P}^{(n-1)}$ denote the transition probability matrix which consists of predicted future performance after $(n-1)$ years later. Hence, by induction [10],

$$\mathbf{P}^n = \begin{pmatrix} \mathbf{Q}^n & (\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots + \mathbf{Q}^{n-1})\mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{Q}^n & \mathbf{R}^n \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \tag{6}$$

In the long run, as $n \rightarrow \infty$, the matrix \mathbf{P}^n can be defined as

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{0} & (\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots)\mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{7}$$

where the matrix $(\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{F}$ is called the fundamental matrix of the absorbing Markov chain [11]. In this study, the fundamental matrix represents the average lifetimes of students in each transient states. Given the fundamental matrix, the expected length of study before graduating, $\boldsymbol{\mu}$ is:

$$\boldsymbol{\mu} = (\mathbf{I} - \mathbf{Q})^{-1}(\mathbf{1} \ \mathbf{1} \ \mathbf{1} \ \mathbf{1})^T. \tag{8}$$

Under the double absorbing states, the absorbing rate is given by

$$(\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R} = \mathbf{FR}, \tag{9}$$

where **FR** is the probability matrix of the students that starting in the transient state will end up in the absorbing (graduating and withdrawn). The retention and repetitive rates can be determined using the initial transition probability matrix.

4 Results and discussion

As mentioned, the data used for this study are the students of School of Mathematical Sciences, Universiti Sains Malaysia. It consists of 540 pure sciences program students and 1024 applied sciences program students.

4.1 The initial transition probability matrix

The data are grouped according to their academic programs and gender. Using the frequency table shown in Table 2 and the equation (1), the one-step transition probability matrix is developed by calculating the dropout, repetitive, promotion and graduation proportions.

Table 2. The frequency tables of the data.

	First year				Second year				Third year				Fourth year			
	Pure		Applied		Pure		Applied		Pure		Applied		Pure		Applied	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Promoted	33	104	61	164	31	104	62	178	26	111	67	191	0	0	0	0
Repeated	2	1	0	2	1	1	1	0	0	1	0	1	0	1	2	6
Withdrawn	4	3	13	22	1	2	1	2	1	0	0	2	0	2	0	3
Graduate	0	0	0	0	0	0	0	0	0	0	0	0	17	94	68	178
Total	39	108	74	188	33	107	64	180	27	112	67	194	17	97	70	187

* M = male, F = female

Then, this one-step transition probability matrix for the pure and applied sciences programs are constructed and organized into the canonical form as shown in Table 3 and Table 4, respectively.

Table 3. The initial transition probability matrix P_{pure} .

State	1	2	3	4	W	G
1	0.2041	0.9320	0	0	0.04762	0
2	0	0.0143	0.9643	0	0.02142	0
3	0	0	0.0072	0.9856	0.0072	0
4	0	0	0	0.0088	0.0175	0.9737
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 4. The initial transition probability matrix $\mathbf{P}_{applied}$.

State	1	2	3	4	W	G
1	0.0076	0.8588	0	0	0.1336	0
2	0	0.0041	0.9836	0	0.0123	0
3	0	0	0.0038	0.9885	0.0077	0
4	0	0	0	0.0311	0.0117	0.9572
W	0	0	0	0	1	0
G	0	0	0	0	0	1

The initial transition probability matrix are also further disaggregated by the gender. From Table 2, the canonical form of the transition probability matrix for the male and female students from the pure sciences are as shown in Table 5 and Table 6.

Table 5. The initial transition probability matrix $\mathbf{P}_{pure(male)}$.

State	1	2	3	4	W	G
1	0.0513	0.8462	0	0	0.1026	0
2	0	0.0303	0.9394	0	0.0303	0
3	0	0	0	0.9630	0.0370	0
4	0	0	0	0	0	1
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 6. The initial transition probability matrix $\mathbf{P}_{pure(female)}$.

State	1	2	3	4	W	G
1	0.0093	0.9630	0	0	0.0278	0
2	0	0.0093	0.9720	0	0.0187	0
3	0	0	0.0089	0.9911	0	0
4	0	0	0	0.0103	0.0206	0.9691
W	0	0	0	0	1	0
G	0	0	0	0	0	1

The canonical form of the transition probability matrix of the male and female students from applied sciences program are as shown in Table 7 and Table 8.

Table 7. The initial transition probability matrix $\mathbf{P}_{applied(male)}$.

State	1	2	3	4	W	G
1	0	0.8243	0	0	0.1757	0
2	0	0.0156	0.9688	0	0.0156	0
3	0	0	0	1	0	0
4	0	0	0	0.0286	0	0.9714
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 8. The initial transition probability matrix $P_{applied(female)}$.

State	1	2	3	4	W	G
1	0.1064	0.8723	0	0	0.1170	0
2	0	0	0.9889	0	0.0111	0
3	0	0	0.0052	0.9845	0.0103	0
4	0	0	0	0.0321	0.0160	0.9519
W	0	0	0	0	1	0
G	0	0	0	0	0	1

4.2 The expected duration of study

The fundamental matrix, F and expected length of study, μ were constructed using the SAS software. The total expected length of study until graduation (completion) is calculated using the equation (3). The result from Table 9 gives the total expected length of study until graduation (completion) for both of the pure and applied sciences program.

Table 9. Expected duration of study before graduating (years).

	Pure Sciences	Applied Sciences
1	3.8556	3.6100
2	2.9796	3.0071
3	2.0088	2.0280
4	1.0088	1.0321

It is found that the expected duration of study of the 1st, 2nd, 3rd and 4th year students from the pure science program are 3.8556, 2.9796, 2.0088 and 1.0088 years, respectively, while for the students in the applied sciences program, the expected duration of study for the 1st, 2nd, 3rd and 4th year students are 3.6100, 3.0071, 2.0280 and 1.0321 years, respectively. It is clear that the expected number of years required before graduating for the applied sciences program is higher than the pure sciences program except for the first years students.

Table 10 gives the total expected length of study until graduation for both of the male and female students for both programs.

Table 10. Expected duration of study before graduating (years).

	Pure Science		Applied Science	
	Male	Female	Male	Female
1	3.6699	3.9163	3.4837	4.0526
2	2.9329	2.9901	3.0131	3.0051
3	1.9630	2.0194	2.0294	2.0276
4	1	1.0104	1.0294	1.0331

From these results, it is found that the expected duration of study of the male students in the pure sciences program for the 1st, 2nd, 3rd and 4th year students are 3.6699, 2.9329, 1.9620 and 1 year, respectively. The expected duration of study of the female students for the 1st, 2nd, 3rd and 4th year students are 3.9163, 2.9901, 2.0194 and 1.0104 years,

respectively. For the pure science program, it shows that the female students spend more time of study before graduating than the male students.

It can be seen that the expected duration of study for the male students in 1st, 2nd, 3rd and 4th years are 3.4837, 3.01331, 2.0294 and 1.0294 years, respectively while the expected duration of study for the female students of the 1st, 2nd, 3rd and 4th year students are 4.0526, 3.0051, 2.0276 and 1.0331, respectively. It is clear that the average number of years required before graduating of the applied sciences program for the male students in second and third years levels are higher than the female students. However, in the first and fourth years levels, the average number of years required before graduating are higher for the female students.

4.3 Absorbing rates

The probability of absorption that is the probability of graduating and withdrawn are calculated using the equation (4).

4.3.1 Absorbing rates by academic program

For both of the program, the probabilities of reaching the graduating and withdrawn states are as shown in Table 11.

Table 11. Probability of reaching the graduating and withdrawn states by academic program.

	Graduating		Withdrawn	
	Pure	Applied	Pure	Applied
1	0.9076	0.8380	0.0924	0.1621
2	0.9540	0.9682	0.0461	0.0318
3	0.9752	0.9804	0.0248	0.0196
4	0.9823	0.9880	0.0177	0.0120

It can be said that in the long run, the percentages of the pure science students who are in the 1st, 2nd, 3rd and 4th years that end up successfully graduated from the system are 90.76%, 95.40%, 97.52% and 98.23%, respectively. The percentage of the applied science students who are in the 1st, 2nd, 3rd and 4th years that end up successfully graduated from the system are 83.80%, 96.82%, 98.04% and 98.80%, respectively. The probability of graduation is higher for the applied science students that the pure science students except for the first year.

On the other hand, the 1st, 2nd, 3rd and 4th year students from the pure sciences program that withdrawn from the system without attaining maximum qualification are 9.24%, 4.61%, 2.48% and 1.77%, respectively while the 1st, 2nd, 3rd and 4th year students from the applied sciences program that withdrawn from the system without attaining maximum qualification are 16.21%, 3.18%, 1.96% and 1.20%, respectively. Different from the probability of graduation, the tendency to drop out from the system is higher for the pure science students compared to the applied sciences student except for the first year level.

4.3.2 Absorbing rates by program and gender

The probabilities of reaching the graduating and withdrawn states according to the programs and gender are as in Table 12.

Table 12. Probability of reaching the graduating and withdrawn states by programs and gender.

	Pure Sciences				Applied Sciences			
	Graduating		Withdrawn		Graduating		Withdrawn	
	Male	Female	Male	Female	Male	Female	Male	Female
1	0.8320	0.9338	0.1680	0.0662	0.8112	0.9395	0.1888	0.1676
2	0.9239	0.9607	0.0671	0.0399	0.9841	0.9624	0.0159	0.0376
3	0.9630	0.9792	0.0370	0.0208	1	0.9732	0.0370	0.0027
4	1	0.9792	0	0.0208	1	0.9834	0	0.0167

The percentage of the 1st, 2nd, 3rd and 4th years of the male students in the pure science program that successfully graduated from this system are 83.20%, 92.39%, 96.30% and 100%, respectively while the percentage of the 1st, 2nd, 3rd and 4th years of the male students in the pure science program that successfully graduated from this system are 93.38%, 96.07%, 97.92% and 97.92%, respectively. The probability of graduation of the female students in the first year, second year and third year are higher than that of the male students.

The 1st, 2nd, 3rd and 4th years of the male students in the pure science courses that will end up dropped from the system are 16.79%, 6.71%, 3.70% and 0% respectively while for the 1st, 2nd, 3rd and 4th year of the female students in the pure science courses that will end up dropped from the system are 6.62%, 3.93%, 2.08% and 2.08%, respectively. In pure sciences program, the male student has higher tendency to drop out from the system than the female students except for the fourth year students. However, in the fourth year, for the probability of dropout, the male ratio is exceeding than the female ratios.

The 1st, 2nd, 3rd and 4th years of the male students from the applied sciences program that successfully graduated from the system are 81.12%, 98.41%, 100% and 100 %, respectively while the female students from the applied sciences program that successfully graduated from the system are 93.95%, 96.24%, 97.32% and 98.34%, respectively. The probability of reaching the graduated state for the male student are higher than the female students except in the first year level. In the first year level, the female students exceeding the male students.

The percentage of the male students of the 1st, 2nd, 3rd and 4th years in the applied sciences program that end up withdrawn from the system without attaining maximum qualification are 18.88%, 1.59%, 3.70% and 0%, respectively while the percentage of the female students of the 1st years, 2nd years, 3rd years and 4th years in the applied sciences program that end up withdrawn from the system without attaining maximum qualification are 16.76%, 3.76%, 0.27% and 1.67%, respectively. The probability of withdrawn from the system is higher for the male students in the first and third year level. However, in the second year and fourth years, the tendency to drop from the system were higher for the female students.

4.4 Retention rates

Retention rate or the percentage of progressing can be determined using the transition probability matrix. Retention rate is important to maintain the enrollment of the students. Besides, it can be an indicator of well-being for the community as a whole. Therefore, a high retention rate gives a greater understanding of the performance of students for higher institution.

4.4.1 Retention rates by academic programme

It can be seen from Table 3 that the retention rate of the first, second, third and fourth years of the pure sciences program are 93.20%, 96.43%, 98.56% and 97.37%, respectively. Similarly, from Table 4, the retention rate of the first, second, third and fourth years of the applied sciences program are 85.88%, 98.36%, 98.85% and 95.72%, respectively.

It can be concluded that the retention rate of pure sciences group in the first and fourth years are higher than that of the applied sciences group. However, in the second and third year, the retention rates are higher for applied sciences group as compared to the pure sciences group.

4.4.2 Retention rates of pure sciences program by gender

Considering the same analysis, the result are further differentiated between the males and females. From the Table 5, the retention rate for male students of the first, second, third and fourth years are 84.62%, 93.94%, 96.30% and 100%, respectively. Similarly, based on the Table 6, the retention rate for the female students of the first, second, third and fourth years are 96.30%, 97.20%, 99.11%, and 96.91%, respectively.

It can be said that the retention rate for the female students in the first, second and third years are higher than the male students. However, in the fourth year, the retention rate of the female students is exceeding the male students by 3.09%.

4.4.3 Retention rates of applied sciences program by gender

The retention rate for applied science program based on the gender is also constructed. From the transition probability matrix in Table 7, the retention rate for the male students of first, second, third and fourth years are 82.43%, 96.88%, 100%, and 97.14%, respectively. From the Table 8, for the female student, the percentage of progressing of the first, second, third and fourth years are 87.23%, 98.89%, 98.45% and 95.19%, respectively.

It can be said that the retention rate for the female students in the first and second year are higher than the male students. However, in the third and fourth years, the retention rates are higher for the male students.

4.5 Repetitive rate

Grade repetition defines as those students that are held in the same level in the following years. Repetition rate was also determined from the transition probability matrix.

4.5.1 Repetitive rates by academic program

Based on the Table 3, it can be seen that, the percentage of the first, second, third and fourth year students from the pure sciences programs that remain in their state are 2.04%,

1.43%, 0.72% and 0.88%, respectively. Based on the Table 4, the percentage of the first, second, third and fourth year students from the applied sciences programs that remain in their state are 0.76%, 0.41%, 0.38% and 3.11%, respectively.

It can be seen that the repetitive rate for the pure sciences program is exceeding the applied sciences program except for the fourth year level. In the fourth year level, the repetition rate is higher in the applied sciences program than the pure sciences program.

4.5.2 Repetitive rates of pure sciences program by gender

As before, the result is also disaggregated by the gender. From transition probability matrix in Table 5, the percentage that the first, second, third and fourth year of the male students remain in their state are 5.13%, 3.03%, 0% and 0%, respectively, while for the female students, as shown in Table 6, the probability of repetitive rate are 0.93%, 0.93%, 0.89% and 1.03%, respectively.

It is clear that the male students in the pure sciences program is higher than the females in first and second year levels. However, the female students have higher repetition rate in the third and fourth year levels.

4.5.3 Repetitive rates of applied sciences program by gender

Based on the Table 7, the repetitive rates of the male students from the applied sciences program for the first, second, third and fourth year levels are 0%, 1.56%, 0% and 2.86%, respectively while the repetitive rates of the female students from Table 8 for first, second, third and fourth years are 10.64%, 0%, 0.52% and 3.21%, respectively.

The repetitive rate for the female students in the first, third and fourth year levels are higher than of the male students. However, in the second year, the rate are higher for the male students. Large number of students are repeating in the first year level, particularly the male students.

4.6 Enrollment projection

Using the initial transition matrix, it is possible to forecast the number of student in the future though in a short period. The enrollment projection is also developed using SAS software.

4.6.1 Forecasting the future performances of the pure sciences program

Given the initial transition probability matrix, the vector for future performances can be calculated using equation (2). The results are as shown in Table 13, Table 14 and Table 15.

Table 13. The estimated transition probabilities matrix of Pure Sciences program for $n = 2$, \mathbf{P}_{pure}^2 .

State	1	2	3	4	W	G
1	0.0004	0.0323	0.8987	0	0.0686	0
2	0	0.0002	0.0207	0.9504	0.0287	0
3	0	0	0	0.0157	0.0245	0.9597
4	0	0	0	0	0.0177	0.9822
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 14. The estimated transition probabilities matrix of Pure Sciences program for $n = 3$, \mathbf{P}_{pure}^3 .

State	1	2	3	4	W	G
1	0	0.0009	0.0376	0.8858	0.0757	0
2	0	0	0.0003	0.0288	0.0455	0.9254
3	0	0	0	0.0001	0.0248	0.9750
4	0	0	0	0	0.0177	0.9823
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 15. The estimated transition probabilities matrix of Pure Sciences program for $n = 4$, \mathbf{P}_{pure}^4 .

State	1	2	3	4	W	G
1	0	0	0.0011	0.0449	0.0916	0.8624
2	0	0	0	0.0006	0.0460	0.9534
3	0	0	0	0	0.0248	0.9752
4	0	0	0	0	0.0177	0.9823
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Considering that for those students in the first year level, only 93.20 % of the students are expected to proceed to the second year, 89.87% of the students are expected to proceed to the third year, 88.58% of the students are expected to proceed to the fourth year and finally only 86.24% of the students are expected to graduate.

Similarly, for those students in the second year level, only 96.43% of the students are expected to proceed to the third year, 95.04 % of the students are expected to proceed to the fourth year, and finally only 92.54% of the students are expected to graduate.

For those students that are currently in the third year level, only 98.56 % of the students are expected to proceed to the fourth year and only 95.97% of the students finally graduate from the system. Lastly, considering those students in the fourth year level, only 97.37% are expected to graduate from the system.

4.6.2 Forecasting the future performances of the applied sciences program

Similarly, the vector of future performance for applied sciences program is calculated. The results are as shown in Table 16, Table 17 and Table 18.

Table 16. The estimated transition probabilities matrix of applied sciences program for $n = 2$,

$$P_{applied}^2$$

State	1	2	3	4	W	G
1	0	0.0101	0.8447	0	0.1452	0
2	0	0	0	0.9723	0.0199	0
3	0	0	0	0.0346	0.0192	0.9462
4	0	0	0	0.0009	0.0120	0.9870
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 17. The estimated transition probabilities matrix of applied sciences program for $n = 3$,

$$P_{applied}^3$$

State	1	2	3	4	W	G
1	0	0	0.0131	0.8350	0.1518	0
2	0	0	0	0.0380	0.0313	0.9307
3	0	0	0	0.0011	0.0196	0.9793
4	0	0	0	0	0.0120	0.9879
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Table 18. The estimated transition probabilities matrix of applied sciences program for $n = 4$,

$$P_{applied}^4$$

State	1	2	3	4	W	G
1	0	0	0.0001	0.0390	0.1616	0.7993
2	0	0	0	0.0012	0.0317	0.9670
3	0	0	0	0	0.0196	0.9803
4	0	0	0	0	0.0120	0.9880
W	0	0	0	0	1	0
G	0	0	0	0	0	1

Considering those students in the first year level, only 85.88% of the students are expected to proceed to the second year, 84.47% of the students are expected to proceed to the third year, 83.50% of the students are expected to proceed to the fourth year and finally only 79.93% of the students are expected to graduate.

As before, for those students that are currently in the second year level, only 98.36% of students are expected to proceed to the third year, 97.23% of the students are expected to proceed to the fourth year, and finally only 93.07% of students are expected to graduate.

In the same way, for those students that are currently in the third year level, only 98.85 % of the students are expected to proceed to the fourth year and only 94.62% of the students finally graduate from the system. Last but not least, considering that for those students in the fourth year level, only 95.72% are expected to graduate from the system.

5 Conclusion

The overall analysis from the results indicate that the students of the first, second, third and fourth year levels for both of the pure and applied sciences programs spend their study approximately 4 years, 3 years, 2 years and 1 year, respectively, to complete their study and attain their graduate level. It is seen that the probability of reaching graduation and the retention rates are increasing over time as the students move to higher levels. On the other hand, the dropout rate is decreasing over time as the students move to the higher levels, with the highest is in the first year level (16.21%) which are dominated mostly by male students and the lowest is in the final year level (1.20%). The repetition rates, on average, are estimated in between 0.7% to 2.04% in the pure science program and between 0.38% to 3.11% in the applied sciences program. The repetition rate of this school is relatively low. The future performance of the pure sciences students are higher than the applied sciences students.

Even though the first year students have higher dropout rate and lower retention rate, the performance of these students improve over time as they move from one level to another because they understand the system better as they pass form one level to another. The results of this study indicate that the Markov chain model is very useful for describing the probabilistic behavior of students' enrollment data. The formulated Markov chain model had shown its flexibility to deal with students' yearly data.

5.1 Recommendations

From the above conclusions, the following recommendations are made. The faculty members can play an active role by providing the university guidance programs such as mentor mentee programs and promote social life or institutional adaption especially to the first year students in order to assist them to adjust to the campus life such as Freshman Year Program. The higher institution should also organizes the career center program to empower the students in applying their study toward succeeding in their career path such as Career Carnival. The performance of the students can also be improved by creating a positive environment using innovative technology such as collaborative and cooperative learning in the classroom.

We would like to acknowledge the Universiti Sains Malaysia for providing the data and the support.

References

1. S. Giannini, #CommitToEducation, Publisher UNESCO, France (2019)
2. C. Shah, G. Burke, Higher Education **37**, 359 (1999)
3. M.J. Nyandwaki, O.E. Akelo, O.O. Samson, O. Fredrick, Math. Theory and Modelling **4**, 73 (2014)

4. R.A. Adeleke, K.A. Oguntuase, R.E. Ogunsakin, *Int. J. Sci. Technol. Res.* **3**, 349 (2014)
5. M.G. Nicholls, *J. Oper. Res. Soc.* **60**, 1183 (2009)
6. A. Bairagi, S.C. Kakaty, *Int. J. of Stat. Syst.* **12**, 323 (2017)
7. M.N. Egbo, D.C. Bartholomew, J.U. Okeke, E.N. Okeke, *Open J. Stat.* **8**, 533 (2018)
8. G.S. Walde, *J. Phy. Conf. Ser.* **1176**, 1 (2019)
9. S.M. Ross, J.J. Kelly, R.J. Sullivan, W.J. Perry, D. Mercer, R.M. Davis, T.D. Washburn, E.V. Sager, J.B. Boyce, V.L. Bristow, *Stochastic processes*. 2nd ed. (New York: Wiley, 1996)
10. C. Grinstead, L. Snell, *Introduction to probability* (American Mathematical Society, 1997)
11. Wikipedia contributors. Absorbing Markov chain [Online], Available from the World Wide Web: https://en.wikipedia.org/wiki/Absorbing_Markov_chain [Accessed 10th June 2020].