USING MARKOV CHAIN TO MODEL THE LEARNING PROCESS OF STUDENTS FROM THE FACULTY OF EDUCATION AT THU DAU MOT UNIVERSITY

Le Thi Thu [(1)]

(1) Faculty of Education, Thu Dau Mot University

Corresponding author: thult.khtn@tdmu.edu.vn

DOI: 10.37550/tdmu.EJS/2025.04.684

Article Info

Volume: 7
Issue: 4
Dec: 2025
Received: Nov. 17th, 2025
Accepted: Dec. 15th, 2025
Page No: 879-886

Abstract

This study applies a first-order Markov chain to analyze and model the academic progression of 317 students from the Faculty of Education at Thu Dau Mot University, utilizing their semester Grade Point Averages (GPA) as the core data. Students' GPAs were methodologically classified into four distinct academic performance states: Weak (0–4.99), Average (5.0–6.99), Good (7.0–7.99), and Excellent (8.0–10.0). Transition matrices were constructed to capture the movements between these performance states across consecutive semesters. Descriptive analysis reveals a positive performance trend, specifically a frequent transition from the Average to the Good group, and a high level of stability observed within the Excellent group, particularly in the later stages of the program. A crucial Chi-square test for homogeneity revealed statistically significant differences, indicating that the learning process is non-homogeneous over time, reflecting fluctuations in student learning behavior. However, to fulfill the objective of forecasting the expected distribution of student performance in the subsequent semester, a weighted average transition matrix was computed, giving greater emphasis to the influence of more recent academic data. Forecasting results suggest that approximately 90% of students are expected to concentrate within the Good and Excellent categories, confirming a high standard of academic performance and providing valuable empirical evidence for targeted student support and curriculum management within the Faculty of Education.

Keywords: Academic performance; Educational modeling; GPA; Markov chain; Student learning progression; Transition matrix.

1. Introduction

In the current context of higher education, assessing, monitoring, and predicting students' academic performance have become essential tasks. Students' learning outcomes not only reflect individual capabilities but are also influenced by multiple dynamic factors that evolve over time, such as teaching methods, learning environments, learning motivation, and adaptability. Among mathematical tools, the Markov chain model has been recognized as an appropriate approach for describing stochastic processes whose evolution depends on the preceding state (Thanh, 2005; Privault, 2018).

In both Vietnam and other countries, numerous studies have explored the application of the Markov chain for analysis and forecasting across different domains. Within the field of education, Alawadhi and Konsowa (2010) applied a Markov chain to analyze learning progression, estimate students' mean lifetimes in different levels of study, as well as the percentage of dropping out of the system, and predict graduation rates among students at Kuwait University. Hlavatý and Dömeová (2014) used an absorbing Markov chain to model students' progress through the examination process in a Mathematical Methods in Economics course, estimate transition probabilities between evaluation stages, and show how continuous assessment and bonus points significantly increase the probability of passing and achieving higher final grades, while identifying critical stages that influence students' overall success. Adam (2015) applied a Markov chain to examine students' flow in Sudanese universities, estimating the number of graduates, delayed students, and dropouts over the batches. Similarly, Baggia et al. (2017) utilized a discrete-time homogeneous absorbing Markov chain to model students' academic performance at Maribor University, Slovenia, and to forecast students' enrolment for the next three academic years. Kibiya et al. (2020) investigated student transitions across academic levels at Mewar University, India, using a Markov chain model, and found that the probability of graduation increases while the rate of withdrawal decreases as the student progress through higher years of study. Moody and DuCloux (2014) used a discrete Markov-chain model and historical NAEP data to estimate transition patterns in mathematics achievement gaps between African American and White American students across ages 9, 13, and 17, showing that the gap is likely to close within 10–15 years depending on age level, thus demonstrating the predictive value of Markov processes for long-term educational inequality analysis. Moreover, Mallak et al. (2023) integrated Markov chains with data mining techniques to predict academic outcomes for E-commerce students at Kadoorie University, Palestine. Their model, constructed from students' semester GPAs, predicted probabilities of achieving grades A, B, C, and D as 3.43%, 13.88%, 31.04%, and 23.78%, respectively, with 27.87% of students at risk of not graduating.

Besides educational applications, Markov chains have been widely used in various fields such as weather forecasting and political election modeling (Jing Xun, 2021), predicting natural forest dynamics (Quy et al., 2017), predicting the behavior of gold price (Mallak & Abdoh, 2022), and estimating labor supply and workforce size (Dung & Ha, 2015). However, to the best of our knowledge, no study in Vietnam has applied the Markov chain model to represent the entire learning process across multiple semesters based on students' GPA data.

In this paper, we use the first-order Markov chain model to describe changes in academic performance among Mathematics majors using GPA data across four semesters. The objectives are to identify the long-term stationary distribution and to forecast trends in students' academic progression over time.

2. Materials

2.1. Markov chain

A Markov chain is a stochastic process $\{X_n\}$ in which the state space S is a finite or countably infinite set satisfying the memoryless property $P(X_{n+1} = j | X_n = i,\ X_{n-1} =$

$i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j | X_n = i) = P_{ij}$, for all $n \geq 0$ and all $i_0, i_1, \dots, i_{n-1}, i, j \in S$ (Thanh, 2005; Privault, 2018).

The Markov matrix or transition probability matrix $P = (P_{ij})$ of the process satisfies

$0 \leq P_{ij} \leq 1, \forall\, i, j \in S$ and $\sum_i P_{ij} = 1$.

If the transition probability P_{ij} depends on n, then the Markov chain is said to be heterogeneous. Conversely, if the transition probability P_{ij} does not depend on n, then the Markov chain is said to be homogeneous.

Let $\pi^{(0)}$ be the probability distribution vector at time $n = 0$. The probability distribution at time $n = k$ is calculated by the formula $\pi^{(k)} = \pi^{(k-1)}P$. The distribution π^* that satisfies $\pi^* = \pi^* P$ and $\sum_i \pi_i^* = 1$ is called a stationary (or invariant) distribution.

2.2 Data

The dataset used in this study consists of the semester grade point averages (GPA) of 317 students of the 2023 cohort from the Faculty of Education at Thu Dau Mot University. The semesters included in the analysis in this study are the first and second semesters of both the first and second academic years. The third-semester GPA of each academic year was excluded from the analysis because this semester is primarily designated for students retaking courses they had previously failed. Only data from students who remain enrolled were analyzed, while information from those who have discontinued their studies was omitted. Students were categorized into four academic performance groups based on their semester GPA in each semester as follows: Weak (0–4.99), Average (5.0–6.99), Good (7.0–7.99), and Excellent (8.0–10.0).

In the next sections, the semesters included in the analysis are denoted sequentially as Semester 1 through Semester 4. The matrices of transitional probabilities of a Markov chain with states of these four categories were formed, where the transition probability P_{ij} represents the probability of reaching state j from state i in one step (one semester). In addition, a homogeneity test for the transition matrices was conducted. Finally, a forecast of the academic performance distribution for the subsequent semester was also provided.

3. Results and discussions

3.1. Distribution of grade point averages

Descriptive statistics of the semester GPAs for 317 students indicate noticeable variation across different stages of their academic progression. Specifically, the mean GPAs for the four semesters were 7.73 (Semester 1), 7.49 (Semester 2), 7.59 (Semester 3), and 7.80 (Semester 4). Although the trend does not show continuous growth, the overall GPA range remains between 7.4 and 7.8, suggesting a generally stable performance at a good level.

A more detailed analysis of transitions among the academic performance categories (Weak–Average–Good–Excellent) will be presented in the next sections.

3.2. Transition probability matrix

Let P_{ij} denote the transition probability matrix representing the students' academic performance transitions from semester i to semester j. The symbols W, A, G, and E correspond to the performance states Weak, Average, Good, and Excellent, respectively. For instance, P_{WA} denotes the probability of transitioning from state W in semester i to state A in semester j. Each element of P_{ij} is computed from the corresponding transition count matrix by dividing the relevant entry by the sum of its row.

$$P_{ij} = \begin{bmatrix} P_{WW} & P_{WA} & P_{WG} & P_{WE} \\ P_{AW} & P_{AA} & P_{AG} & P_{AE} \\ P_{GW} & P_{GA} & P_{GG} & P_{GE} \\ P_{EW} & P_{EA} & P_{EG} & P_{EE} \end{bmatrix}$$

Based on the semester GPA data of 317 students from the faculty of Education across four academic terms, the progression of academic performance was characterized using transition count matrices constructed between consecutive semesters. These matrices are presented in Tables 1, 2, and 3.

Table 1. Transition count matrix from semester 1 to semester 2

Semester 1 \ Semester 2	Weak	Average	Good	Excellent
Weak	0	0	0	0
Average	3	13	6	0
Good	2	48	123	22
Excellent	0	2	43	55

(Source: Authors' compilation from primary data, 2025)

The elements of each row of Table 1 were then divided by the total sum of that row, producing the transition probability matrix P_{12} from semester 1 to semester 2 as below.

$$P_{12} = \begin{bmatrix} - & - & - & - \\ 0.136 & 0.591 & 0.273 & 0.000 \\ 0.010 & 0.246 & 0.631 & 0.113 \\ 0.000 & 0.020 & 0.430 & 0.550 \end{bmatrix}$$

The transition matrix P_{12} indicates that no students were recorded in the Weak category in Semester 1. Students tended to maintain or improve their performance within the Average and Good groups. For the Average group, the probability of remaining in the same state is relatively high (59.1%), while the probability of improving to the Good level is also considerable (27.3%); however, there remains a possibility of declining to Weak (13.6%). Similarly, students in the Good group in Semester 1 exhibit a high probability of maintaining their status (63.1%), though transitions to Average (24.6%) and improvements to Excellent (11.3%) are also observed. The Excellent group demonstrates moderate stability, with 55% remaining Excellent, while 43% transitioned to Good and 2% moved to Average.

Table 2. Transition count matrix from semester 2 to semester 3

Semester 2 \ Semester 3	Weak	Average	Good	Excellent
Weak	1	3	1	0
Average	1	21	41	0
Good	0	13	136	23
Excellent	0	0	22	55

(Source: Authors' compilation from primary data, 2025)

The elements of each row of Table 2 were then divided by the total sum of that row, producing the transition probability matrix P_{23} from semester 2 to semester 3 as below.

$$P_{23} = \begin{bmatrix} 0.200 & 0.600 & 0.200 & 0.000 \\ 0.016 & 0.333 & 0.651 & 0.000 \\ 0.000 & 0.076 & 0.791 & 0.134 \\ 0.000 & 0.000 & 0.286 & 0.714 \end{bmatrix}$$

The transition matrix P_{23} indicates a marked increase in the stability of academic performance compared with the previous stage. The Good group demonstrates particularly strong persistence, with 79.1% of students maintaining their status, while 7.6% fall to the Average and 13.4% improve to the Excellent category. The Excellent group also sustains a high level of stability, as 71.4% remain in the same category and 28.6% transition to the Good group. Within the Average group, 1.6% decline to the Weak group, 33.3% remain their status, and a substantial proportion (65.1%) improve to the Good category, reflecting substantial improvement during this period. The Weak group shows encouraging movement, with the majority (60%) advancing to the Average level, suggesting positive academic development among lower-performing students.

Table 3. Transition count matrix from semester 3 to semester 4

Semester 3 \ Semester 4	Weak	Average	Good	Excellent
Weak	1	1	0	0
Average	0	15	21	1
Good	1	20	103	76
Excellent	0	0	5	73

(Source: Authors' compilation from primary data, 2025)

The elements of each row of Table 3 were then divided by the total sum of that row, producing the transition probability matrix P_{34} from semester 3 to semester 4 as below.

$$P_{34} = \begin{bmatrix} 0.500 & 0.500 & 0.000 & 0.000 \\ 0.000 & 0.405 & 0.568 & 0.027 \\ 0.005 & 0.100 & 0.515 & 0.380 \\ 0.000 & 0.000 & 0.064 & 0.936 \end{bmatrix}$$

The transition matrix P_{34} indicates a substantial increase in performance stability as students move from Semester 3 to Semester 4. The Excellent group displays strong persistence, with 93.6% of students remaining in this category— the highest stability observed across all stages. The Good group also demonstrates a moderate degree of stability, with 51.5% maintaining their status, while 38.0% progress to the Excellent level, indicating meaningful upward mobility among higher-performing students. Students in the Average group continue the improvement pattern seen in earlier stages, with 56.8% transition to the Good category, suggesting sustained academic advancement. Although the Weak group shows a balanced split between remaining Weak and moving to Average, their small sample size limits further interpretation. Overall, this matrix illustrates a learning environment in which stability becomes increasingly pronounced at the upper performance levels, while opportunities for improvement remain evident for students in the mid-performance range.

3.3 Test of homogeneity

In this section, we applied the Chi-square test to evaluate whether the transition process of students' academic performance can be regarded as temporally homogeneous. The objective of this analysis is to determine whether the transition probabilities may be represented by a single, time-invariant transition matrix or whether they exhibit statistically meaningful variation across different stages. The null hypothesis H_0 states that "the transition process of students' academic performance is homogeneous over time." Because the Weak category contains no observations in the transition from Semester 1 to Semester 2 (resulting in an entire row of zeros), the assessment is restricted to the three states for which complete data are available: Average, Good, and Excellent.

We used the chi2 and chi2_contingency functions in the Python software SciPy library to perform this test. The test result yields a Chi-square statistic of $\chi^2 = 129.037$ with a p-value of 2.2×10^{-16}. Because of the extremely small p-value, the null hypothesis H_0 is rejected. This indicates that the transition matrices are not homogeneous over time. In other words, the structure of academic performance transitions differs significantly across semesters. This pattern suggests that, as students progressively adapt to the university learning environment, their academic performance becomes notably more stable, particularly within the Good and Excellent groups.

3.4 Forecasting the academic performance distribution for the subsequent semester using a time-weighted average transition matrix

Since the test of homogeneity indicates that the transition process of academic performance is not temporally homogeneous, the study does not employ the aggregated frequency matrix typically used as the standard estimator in a homogeneous Markov model.

Instead, to generate forecasts for academic performance in the next semester, we adopt a time-weighted average transition matrix constructed in the form $P_{avg} = w_1 \times P_{12} + w_2 \times P_{23} + w_3 \times P_{34}$, where the weights (w_1, w_2, w_3) satisfy $w_1 + w_2 + w_3 = 1$ (Berchtold and Raftery, 2002). The values w_1, w_2, and w_3 are chosen to reflect the relative relevance and reliability of each stage. In this study, we choose $(w_1, w_2, w_3) = (0.2, 0.3, 0.5)$, so $P_{avg} = 0.2 \times P_{12} + 0.3 \times P_{23} + 0.5 \times P_{34}$. The assigned weights reflect the greater importance of more recent stages in shaping future transitions. The transition matrices P_{12}, P_{23}, and P_{34} estimated from the empirical data, were presented earlier in Section 3.2. Based on these components, the resulting time−weighted average transition matrix is obtained as follows:

$$P_{avg} = \begin{bmatrix} 0.310 & 0.430 & 0.060 & 0.000 \\ 0.032 & 0.421 & 0.534 & 0.014 \\ 0.005 & 0.122 & 0.621 & 0.253 \\ 0.000 & 0.004 & 0.204 & 0.792 \end{bmatrix}$$

Based on the academic performance distribution in Semester 4, given by $\pi^{(4)}$ =(0.006, 0.114, 0.407, 0.473), the forecast for Semester 5 is obtained using $\pi^{(5)} = \pi^{(4)} \times P_{avg} = (0.007, 0.102, 0.410, 0.479)$.

The forecasting results indicate a positively stable distribution, with approximately 90% of students expected to fall within the Good and Excellent categories. The upward movement from the Average group toward the Good and Excellent levels continues, as reflected in a slight decrease in the proportion of Average students and a corresponding increase in the higher-performance categories. The Weak group remains extremely small, suggesting minimal risk of academic decline at the lower end of the distribution.

4. Conclusion

This study applied a first-order Markov chain model to investigate the transitions in academic performance among students of the 2023 cohort in the Faculty of Education at Thu Dau Mot University, based on their semester GPA data across four academic terms. By classifying student performance into four states —Weak, Average, Good, and Excellent —the transition matrices for the three periods—Semester 1 to Semester 2, Semester 2 to Semester 3, and Semester 3 to Semester 4 — demonstrated a clear tendency toward upward mobility. Students in the Average category showed a strong tendency to transition upward to the Good group, while the Good group became increasingly stable over time. Meanwhile, the Excellent group exhibited a persistently high probability of remaining in the same state, especially in the later stages. These patterns suggest a general convergence toward the upper performance states, whereas lower states occur less frequently and are rarely sustained across semesters. The Chi-square homogeneity tests confirmed statistically significant differences among the three transition matrices, indicating that the academic evolution observed across semesters cannot be represented by a single, time-invariant transition matrix. Instead, the transition structure varies meaningfully over time, reflecting the distinct academic dynamics of each stage.

To support forecasting, a weighted average transition matrix emphasizing recent data was constructed. The resulting predictions suggested that the distribution of academic performance in the subsequent semester remains highly stable, with nearly 90% of students expected to fall within the Good and Excellent categories, accompanied by a slight rise in the Excellent group and a minor decline in the Average group. These findings suggest a generally positive and stable pattern of academic development among students.

References

Adam, R. Y. (2015). An Application of Markov Modeling to the Student Flow in Higher Education in Sudan. *International Journal of Science and Research*, *4*(2), 49–54.

Alawadhi, S., & Konsowa, M. (2010). Markov Chain Analysis and Student Academic Progress: An Empirical Comparative Study. *Journal of Modern Applied Statistical Methods*, *9*(2), 26. https://doi.org/10.22237/JMASM/1288585500

Berchtold, A., & Raftery, A. E. (2002). The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. *Statistical Science, 17*(3), 328–356.

Brezavšček, A., Bach, M. P., & Baggia, A. (2017). Markov analysis of students' performance and academic progress in Higher Education. *Organizacija*, *50*(2), 83–95. https://doi.org/10.1515/orga-2017-0006

Dung, D. G., & Thao, H. T. P. (2015). Some applications of Markov chain. In *Proceedings of the CITA 2015 Scientific Conference: Information Technology and Applications in Various Fields* (pp. 1–5)

Hlavatý, R., & Dömeová, L. (2014). Students' Progress throughout Examination Process as a Markov Chain. *International Educations Studies*, *7*(12), 20–29. https://doi.org/10.5539/ies.v7n12p20.

I.KIBIYA, Y., A. SABO, S., MUSA, I. Z., & IBRAHIM, J. M. (2020). Modeling of student's academic performance and progression in Indian private universities using Markov chain. *International Journal of Research in Engineering & Science*, *4*(6). https://doi.org/10.26808/rs.re.v4i6.01

Mallak, S., Kanan, M., Al-Ramahi, N., Qedan, A., Khalilia, H., Khassati, A., Wannan, R., Mara'beh, M., Alsadi, S., & Al-Sartawi, A. (2023). Using Markov chains and data mining techniques to predict students' academic performance. *Information Sciences Letters, 12*(9), 2073–2083. https://doi.org/10.18576/isl/120914

Mallak, S., Abdoh, D. (2022). Predicting the Behavior of Gold Price Using Markov Chains and Markov Chains of the Fuzzy States. *Mathematical Statistician and Engineering Applications*, *71*(4), 2906–2920.

Privault, N. (2018). *Understanding Markov Chains: Examples and Applications*. Springer.

Quy, K. V., Bao, T. Q., Dien, P. V., Hai, V. D., Duy, N. N., Ha, T. T. T., Mung, H. T., & Tuyen, N. N. (2017). Applying the Markov chain model approach in forecasting changes in evergreen forests in North Central Vietnam. *Agriculture and Rural Development*, *1*(7), 1–11.

Thanh, N. H. (2005). *Applied mathematics.* Hanoi National University of Education Publishing House.

Moody, V. R., & DuCloux, K. K. (2014). Application of Markov chains to analyze and predict the mathematical achievement gap between African American and White American students. *Journal of Applied & Computational Mathematics, 3*(3), https://doi.org/10.4172/2168-9679.1000161.

Xun, J. (2021). The research of Markov chain application under two common reals world examples. *Journal of Physics: Conference Series, 1936*(1), 012004. https://doi.org/10.1088/1742-6596/1936/1/012004.