A Random Walk on the Major Path Space: Examining Student Progression as a Random Process Using Markov Chains

George Ricco
Department of Electrical and Computer Engineering
University of Kentucky
Lexington, USA
ricco@uky.edu

James Ryan
Chubb Insurance
Philadelphia, USA
jfryan.1754@gmail.com

***Abstract*—Recent work done on student attrition and major switching has demonstrated that students who both succeed and fail to graduate tend to linger in the system the more said students have switched majors. An important phenomenon to examine from these trends would be the penalties of switching majors on the probability of graduation. Such exploration would benefit policymakers on all spectra of the academic system to better see what such penalties look like over large, incredibly varied groups of students. We employ the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) to develop a Markov chain describing the pathway a student would take through his or her terms in a university under some assumptions; such an approach treats major switching as a random event and allows effective modeling to occur at a high level. Representing nearly 13% of US engineering students across eleven partner institutions, MIDFIELD provides good coverage over what we wish to examine and similar data have been previously used to construct descriptive stochastic processes with respect to student progression. We create the Markov chain by allowing its state space to represent the number of times a student has switched majors with epochs denoted as terms spent in a university. We develop the state transition probabilities from 871,742 observations of first-time in college (FTIC) students in MIDFIELD and use this data-driven Random Walk to determine relevant, long-term properties of students who both succeed and fail to graduate in order to gain a clearer idea on what an optimal path through a major progression looks like. We present these results in conjunction with other work done on MIDFIELD to present a clearer profile of the probabilistic behavior of students with respect to switching and how it affects the education path space. We will also consider some of the assumptions that have gone into the construction of this analytical object and how higher levels of granularity can aid in their testing as needed.**

Keywords—data; MIDFIELD; policy; attrition; pathways; models; Markov.

I. Introduction

Markov chains are a widely utilized modeling technique for incorporating elements of uncertainty into arbitrary dynamic systems. In the broadest terms, these models map some sort of parameter domain, often temporal, to an arbitrary state space, the values of which the process can take. Often such models are constructed in order to understand the long term, probabilistic behavior of complicated systems which cannot be fully characterized by traditional descriptive statistics. We have seen the importance of Markov chain understanding increase as its utilization in research continues to expand.[1] We continue this trend by using this framework to model student attrition and its relationship with major switching.

Reviews of modeling theory in the educational space[2] suggest that such a mathematical construct could provide some relevant information about high level problems with regards to educational systems, hence eschewing the high variability present in building probabilistic models on an institutional level. Hence, to build our chain we employ the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) database to address student attrition on a multi-institutional level.

We consider studying the long term probabilities of attrition and graduation of students using both time and major switches as epochs of interest. Then, we divide the data into cohorts based on major to compare the differences in these probabilities for our groups of interest. By examining the data from these myriad vantage points we provide policy makers and researchers an analytical foundation from which further, more detailed work can be done.

II. Literature Review

Though Markov chains are a reasonably powerful tool for modeling dynamic behavior while incorporating randomness there has not been much done with respect to viewing student attrition. Work done by Ahmed and Al-Awadhi[2,3] employ both logistic regression and a Markov chain to examine student attrition by using terms as the epoch of interest and considering the number of students remaining at a term by term basis as the state space. Using their model they demonstrate that such models can be utilized to examine attrition in a novel way; they also demonstrate that such studies can provide policy makers with novel data to better make decisions about students most equipped to succeed while providing benchmarks for analysis. Due to limitations in the data they assumed attrition could be modeled by a binomial distribution of students; given our data we are free of such assumptions of underlying distributions and hence provide results based purely on the data observed.

More recently work done by Nicholls[4] looks at the development of an absorbing Markov chain to provide benchmark values for the analytical assessment of a so called "DBA" program in a graduate school in Australia. Such objects as expected student success rate and expected passage times prior to being absorbed into one of the final states—in this case a success and failure condition—are constructed to evaluate the program and subsequently pose potential changes to best optimize said program. Our model looks to provide the theoretical values that this paper works on to be used to address future questions on dealing with student attrition.

Purzer and Fila[5] also looked at constructing Markov chain frameworks to understand the concept of innovation in engineering. The paper attempted to utilize the Markov framework to delineate innovation ability. Despite providing a novel framework for such analysis, limitations in the data curtailed the robustness of their analysis.

The framework of initiating such models are often sound but limitations in data and/or scope have caused issues in model construction and the analysis thereof. A recent thesis completed by Ferko[6] aims at an institution-based study of retention and attrition rates using an absorbing Markov chain. The paper itself focuses on STEM and non-STEM students and looks at graduation rates over finite and long term time ranges, treating term classification as the states for each chain. Her results demonstrate the ability of the Markov model to find trends within the data such as the notion that even admitting a "better prepared" student body may not result in a nontrivial increase in retention rates.

When constructing our Markov chain we build our transition matrix in the spirit of Ferko and Nicholls[4, 6] by making transient states represent some sort of term in the university. We deviate from Ferko by looking at the number of terms enrolled for our first group of Markov chains. In addition we expand the scope of our analysis to look at three particular cohorts: all students in the database, students who stayed in the engineering group and those students who either left or graduated as engineers. This differs from the aforementioned model which accounts for "graduated non-[class]" as an absorbing state, hence giving us a broader lens from which to work. We also consider major switching as both a possible state and possible epoch which, to the best of our knowledge, has not been previously considered.

In addition, we take into account concerns echoed by Johnstone[2] and consider an examination of an entire education system. The question of variability within a university is something to be considered in future work; for now we focus on constructing generalized analyses for a system of educational institutions.

Though not mentioned in other work, Al-Awadhi considers a time-inhomogeneous Markov chain when doing his analysis.[3]. We cover the difference between time homogeneity and time inhomogeneity in the theory section but this is an important caveat to consider when constructing our chain. The other major papers in our field[4, 6] do not consider time inhomogeneity when constructing their respective chains so we proceed in the spirit of making such an assumption in order to provide the most general of results.

III. Theory

Before We look to define student progressing as a random walk; to do this we consider how to formulate the problem through the framework of a Markov chain. Formally, we define X_n as the set of possible states in time for a random variable $X, n \in \mathbb{N}$ describing the various epochs in which the aforementioned random variable is defined. Suppose X_n takes values in the state space $S = \{1, 2, \dots, M\}$ and that the epochs over which it is defined is a discrete time space. To define a Markov chain we are interested in defining:

$$P(X_n = i_n | X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \, , i_k \in S \; \forall \; k = 1, 2, \dots \quad (1)$$

In order to more readily perform computations we define the *Markov Property,*[7] which states that the above probability can be reduced in the following manner:

$$P(X_n = i_n | X_1 = i_1, \dots, X_{n-1} = i_{n-1}) = P(X_n = i_n | X_{n-1} = i_{n-1}) \, , i_k \in S \; \forall k = 1, 2, \dots \quad (2)$$

In English, this property states that *the past is conditionally independent from the future given the immediate present*. This simplifying assumptions allows for a tremendous ease in modeling physical processes as well as ease in computation.

Another simplifying assumption we rely upon is *time homogeneity*. This means that the probabilities that define state transitions do not change over time and we can subsequently define such probabilities in a natural manner:

$$P(X_n = i_n | X_{n-1} = i_{n-1}) = p(i_{n-1}, i_n) \quad (3)$$

Given this construction we can represent a Markov chain with an object known as a probability transition matrix. Denote this matrix by the symbol P. Intuitively, we allow the ij^{th} entry of the matrix to be the probability of transitioning from state i to state j for all $i, j \in S$. This also imposes the condition that the row sums of the matrix must add up to one since, by this construction, each row refers to the current state and the columns refer to a potential immediate future state. We formalize the delineated conditions below:

$$P_{ij} = p(i, j) \; ; i, j \in S \quad (4)$$

$$\textstyle\sum_j P_{ij} = 1 \; \forall i, j \in S \quad (5)$$

IV. Construction

We employ the aforementioned methodology regarding Markov chains to construct a random walk across student progressing. Our analyses define two state spaces of interest:

1. Terms spent in the university system.
2. Major changes, excluding first year engineering.

Both of these have naturally discrete time spaces—terms spent in the system—with the property that $p(i_l, i_k) = 0 \; \forall l \leq k \in S$ due to the inability of students to travel backward in time. We must also define a stopping point of interest for our analysis.

We subsequently define the stopping states of graduating from the university system or leaving without graduating. Denote these states respectively by G, F.

Let $k \in S(G, F)$ be a state in any of the above state spaces. We can say with certainty that k is a *transient state* meaning that the probability of returning to such a state is less than 1. In fact, by construction this probability is exactly 0. The states G, F are referred to as *absorbing states* since, once entered, they cannot be left.

Before construction we remove any student who has remained in the university system for longer than twenty-three terms due to the large variance associated with explaining small numbers of students. We excise any transfer student and those students who were in a university with a quarter system to prevent state space confounding. We consider three particular cohorts for our analysis:

1. All aggregated students in the MIDFIELD database; $N = 499{,}189$
2. Students who only had the group designation "engineer"; $N = 123{,}101$
3. Students with "engineering" their final group designation; $N = 103{,}148$

For major switching considered as the state space we allow for eight major switches in the aggregate cohort, five major switches for the always engineering cohort and six major switches for the left at engineering cohort. We will use the above numerical distinctions for shorthand in the remainder of this paper. The numbers were limited for similar reasons as the limitations imposed on terms.

We construct our probability transition matrices by using student proportions as provided by the MIDFIELD data set. Given the size of our data sets we employ said proportions with the understanding that they represent the maximum likelihood estimator of the true probability at each step.

Due to student attrition at each potential state we compensate for this proportional loss by taking the proportions with the current number of available students at each particular state. By assuming time homogeneity we can aggregate all possible students in a given state to perform this task. We hence end up with a probability transition matrix with the following form:

	1	2	3	4	…	F	G
1	0	p(1,2)	0	0	…	p(1,F)	p(1,G)
2	0	0	p(2,3)	0	…	p(2,F)	p(2,G)
3	0	0	0	p(3,4)	…	p(3,F)	p(3,G)
4	0	0	0	0	…	p(4,F)	p(4,G)
…	…	…	…	…	…	…	…
F	0	0	0	0	0	1	0
G	0	0	0	0	0	0	1

Fig. 1. Example probability transition matrix.

V. Result and Analysis

We shall break the analysis up into the three cohorts delineated above. In the first, most general cohort we provide detailed and rigorous explanation of the computations performed and their meanings. We do not provide the probability transition matrices and the resulting computed probabilities for the other two cohorts here, as they are too large to effectively display. All computations have been performed in the R programming environment[8] using its base packages for delineated computations and analyses.

A. The Fully Mixed Cohort, "I"

We consider first the set of 499,189 students in the fully mixed cohort for our analysis. Given this matrix we wish to find a natural way of computing graduation rates for the entire space of students with respect to the term in which they completed their path through the university space. Note that these probabilities only apply to the subset of students who completed their pathway at that given term, not with respect to the entire space.

We employ a tactic of transforming a probability transition matrix into its so called canonical form. Specifically, we convert the probability transition matrix into the following block form:

$$\begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \tag{6}$$

In this form, I is the identity matrix, 0 is a matrix of zeroes, R is the matrix containing the transition probabilities from transient to absorbing states, and Q contains the matrix of transition probabilities between transient states.

We wish to compute the probability of absorption given this form. Using a recurrence relationship[9] we note that the absorption probabilities, denoted here as U, can be written as:

$$U = R + QU \tag{7}$$

$$R = (I - Q)U \tag{8}$$

$$U = (I - Q)^{-1}R \tag{9}$$

Hence we simply need to conform our probability transition matrix into the canonical form and solve a simple matrix equation to produce our desired absorption probabilities as the number of epochs considered becomes very large.[10] We list the absorbing probabilities computed in this way in the matrix presented below:

	F	G
1	0.51695	0.483051
2	0.487777	0.512223
3	0.423145	0.576855
4	0.382194	0.617806
5	0.325124	0.674876
6	0.281505	0.718495
7	0.232748	0.767252

8	0.19666	0.80334
9	0.181696	0.818304
10	0.17308	0.82692
11	0.177466	0.822534
12	0.175993	0.824007
13	0.174536	0.825465
14	0.172855	0.827145
15	0.171094	0.828906
16	0.168726	0.831274
17	0.165368	0.834633
18	0.158862	0.841138
19	0.158615	0.841385
20	0.164236	0.835764
21	0.169903	0.830097
22	0.175042	0.824958
23	0.16269	0.83731

Fig. 2. Absorbing probabilities

We perform a similar analysis when considering major switches as our state space of interest with the caveat that the first state is the state of zero major switches. The constructed probability transition matrix for all students as well as the probabilities of, respectively, leaving or graduating the system are respectively provided below.

	F	G
0	0.51695	0.483051
1	0.387187	0.612813
2	0.330237	0.669763
3	0.304362	0.695638
4	0.29814	0.70186
5	0.288854	0.711146
6	0.289474	0.710526
7	0.333333	0.666667
8	0.666667	0.333333

Fig. 3. Switching probabilities to eight switches.

Note that the probabilities for eight major switches appear rather deviant when compared to the general trend which appears to be present when considering long time absorption probabilities. This is due to the number of students left at this epoch being rather small, hence increasing the variance and making analyzing the students remaining in this method somewhat difficult.

VI. Caveats and Discussion

Looking at all students, Figure 2 provides us a baseline understanding of student attrition. The probability of a student failing to graduate (state F) verses a student successfully graduating (state G) are roughly even within the first four terms before stabilizing at the eight term to a probability of graduation of approximately 0.8. This statistic seems stable over the remainder of the terms. Similarly, examining the absorbing probabilities in Figure 3, using major switches as a state space, shows us that the probability of graduating after more than two major switches seems to level out at about 0.7. Major switching seems to reduce rates of graduation while not affecting terms spent in the system. These results hold for the other two cohorts of study.

There are two primary caveats with the formulation of these models. We do not take into account the matriculation of students on a year to year basis and the potential noise that could carry into the model. This would seem, on some level, to violate the assumption of time homogeneity under which we are operating. Other work in the field[4, 6] has not necessarily addressed this factor and for the time being we will continue on with our analyses of the Markov chains constructed under the parameters delineated in the previous section. Future work will aim to explore this potential pitfall in the analysis and see if anything changes by taking these potentially more complex dynamical considerations into account.

In a similar vein we are also making an assumption that the Markov property holds true for student progression with respect to both potential state spaces. As we do not consider variables such as the order under which certain combinations of relevant coursework are performed, let alone analyze the performance therein, this could be a potential lurking issue within the reality of the phenomena that proves problematic when assuming the past and immediate future are conditionally independent given past work.

As it is difficult to analyze the above under a series of models in which this assumption is nearly a centerpiece other work must be done in myriad modeling spaces in order to more reasonably evaluate how much leeway we have with regards to making such assumptions about student behavior. Such work will focus on providing more evidence that such a stochastic treatment of student progression is appropriate beyond assumptions made in the current body of literature within this universe of study.

We must also note that the construction of the Markov chain involving major switching is rather odd in comparison to the others. Instead of using terms as epoch we consider major switching as both a state and an epoch in and of itself. This is a function of the method in which we derived our model form the data itself. Such a model could provide interesting findings, such as the expected number of switches before either graduation or attrition and how these relate over time, but some computations make more sense when considering a more natural temporal epoch such as term. It also may cause issues with time homogeneity as we only assume that major switches have the same dynamics with respect to the transition matrices. These caveats highlight the care that must be taken into construction of the Markov chain model itself.

Acknowledgment

The authors would like to thank Russell Long and Matthew Ohland of Purdue University, and Peter Fox of RPI.

REFERENCES

[1] Kaplan, K. and J. Kaplan. Markov Chains: Reintroducing Lost Knowledge Back into a Modeling and Simulation Course in Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition 2005.

[2] Johnstone, J.N. and H. Philp, The application of a Markov Chain in educational planning. Socio-Economic Planning Sciences, 1973. 7(3): p. 283-294.

[3] Al-Awadhi, S.A. and M.A. Ahmed, Logistic models and a Markovian analysis for student attrition. Kuwait Journal of Science & Engineering, 2002. 29(2): p. 25-40.

[4] Nicholls, M., The Use of Markov models as an aid to the evaluation, planning and benchmarking of Doctoral Programs. Journal of the Operational Research Society, 2009. 60(9): p. 1183-1190.

[5] Purzer, S. and N. Fila. Indicators of Creative and Entrepreneurial Thinking Among Engineering and Technology Students. in American Society for Engineering Education. 2013.

[6] Ferko, S., Using a Markov Model to Analyze Retention and Graduation Rates in Mathematics. 2014, University of Akron: Akron.

[7] Lawler, G., Introduction to Stochastic Processes. 1995, New York City: Chapman and Hall.

[8] Team, R.C.D., R: A Language and Environment for Statistical Computing. 2011, Vienna: R Foundation for Statistical Computing.

[9] Resnick, S., Adventures in Stochastic Processes. 1992, Boston: Birkhauser.

[10] 10. Grinstead, C. and J. Snell, Introduction to Probability. 1997, Providence, Rhode Island: American Mathematical Society.