MARKOVIAN Modeling of Students Performance by Cohorts in Secondary Schools

AMWAYI N. EVERLYNE[1], SIRENGO J. L.[2], KATI R. O.[3]
[1, 2, 3] *Kibabii University, Kenya*

Abstract- The evaluation of students' progress is a very important part of any educational system. Examination results is one of the indicators of a student's success and school performance. The assessment methods used in schools cannot be used to forecast with what certainty a student's level of performance will change in either direction. The Markov chain model can be sensitive to reflect the true assessment. In this paper Markovian modeling of student performance by cohorts in Secondary school was done. The paper sought to show how students' achievements in the immediate previous exams affect the final grade. With the help of Markovian chains, the probability of a student scoring a given grade at the end of the course was shown. The study was conducted through case studies, document analysis, Markovian chain theory, maximum likelihood estimator to estimate transition parameters and Matlab software. It is hoped that the findings of this study will yield information that will be useful to project planning and decision making by the candidates, teachers, parents, curriculum implementers and evaluators and other stakeholders in the education system.

Indexed Terms- Markov chain, Cohort, Transition and Fundamental matrices

I. INTRODUCTION

The problem of understanding and assessing the flow of students through educational systems has long been dealt with using finite absorbing Markov chains. Nicholls (2009) analyzed the flow of students at undergraduate and graduate level within a business faculty in a graduate school in Australia. He used the Markov model to forecast and determine the expected numbers in the DBA program, expected success rates for candidates, expected first passage times prior to absorption per see and also prior to withdrawal and graduation. Early work by Burke (1972), analyzed the flow of student-teachers at the undergraduate level by use of a Markov chain model to estimate a range of present values of the cost of training a cohort of university entrants and subsequent salary bill in teaching of the survivors.

In 2013 Imboga et. al. presented a paper on the construction of Markov Transition Matrices for cohorts of students pursuing a bachelor of science degree in Actuarial Science at JKUAT. He discussed a first order Markov chain with finite number of states and points to analyze the flow of students through the course in terms of transition ratios. Adeleke et al. (2014), applied Markov chain to the assessment of students' admission and academic performance in Ekiti state university in Nigeria. He constructed a transition matrix and obtained probabilities of absorption (graduating and withdrawal). He used the fundamental matrix to determine the expected length of students stay before graduating and made predictions on enrollment and academic performance of students.

Robert and Ludmila (2014), analyzed students' progress throughout examination process as a Markov chain in Czech university of life sciences, Praque, Czech Republic. In their paper, they showed how students' achievements during the semester affected the final result of the exam in terms of the final grade. They used the Markov chains to show the probability of success at the end of the course regarding students' behavior and diligence during the course.

II. MARKOV CHAIN THEORY

Markov Chain theory is one of the mathematical tools used to investigate dynamic behaviors of a system such as financial systems, health service system etc. A Markov chain is a special kind of stochastic process, where the outcome of the experiment depends only on the previous experiment. , in other words, the next state of the system depends only on the present state, not on preceding states. That is, the i^{th} result of a certain element of a system is only influenced by the $(i-1^{th})$ result in the transition, and has nothing to do with other results. Therefore

$$P(X_{t+1}=j\backslash X_0=x_0, X_1=x_1. X_2=x_{2,} \dots\dots X_t=i)=P(X_{t+1}=j|X_t=i) \quad (1)$$

$$P(X_{t+1}=j|X_t=i)=P_{ij} \quad (2)$$

$where\ i,j=1,2,\dots,m$

If $P_{ij}>0$, then state E_i can communicate with state E_j. A two way communication is possible if additionally $P_{ij}>0$. For fixed I, the list P_{ij} is a probability distribution. The states $E_1, E_2, \dots, E_m, m<\infty$ are exhaustive and mutually exclusive. The numbers P_{ij} are transition probabilities of the distribution. For $P_{ij}>0$, $\sum_{i=0}^{\infty} P_{ij}=1$, for each $i,j=1,2,\dots,m, m<\infty$.

Transition probabilities form an $m\ x\ m$ array that can be combined into a transition matrix T, where,

$T=[P_{IJ}]=\begin{bmatrix} P_{11} & \cdots & P_{1m} \\ \vdots & \ddots & \vdots \\ P_{m1} & \cdots & P_{mm} \end{bmatrix}$, m is the number of exclusive states of the system.

III. MODEL FORMULATION

Suppose a given cohort of students who were admitted in secondary school having scored particular grades in Kenya Certificate of Primary Education (K.C.P.E) exams, scored some grades in Kenya Certificate of Secondary Education (K.C.S.E) at the end of the four year course. We use the transition matrix to develop a prediction model for each state.

Let $T=P_{ij}$ be a transition matrix such that $P(X_{t+1}=j|X_t=i)\ where\ i,j=1,2,\dots,m$

We define the state i as the grade scored at K.C.P.E level, and state j, as the grade scored at K.C.S.E level, such that $P_{ij}\geq 0, P_{ji}=0, P_{ii}=0\ and\ P_{jj}=0$. P_{ij} denote the probability that a student in grade i at time $t-1$ will be in grade j at a time t. We define the transient states as Grade A, Grade A-, Grade B+, Grade B, Grade B-, Grade C+ Grade C. and absorbing states as Grade C-, Grade D+, Grade D, Withdrawal (W). The state withdrawal is the probability that a student who joined a form one class withdraws before completing her four year course. The absorbing states i represents those students who did not qualify to join form one class because they did not merit. A state S_i of a Markov chain is absorbing if it is impossible to leave it (i.e.$P_{ii}=1$). The transition matrix will have the form

$T=[P_{ij}]=\begin{bmatrix} Q & R \\ O & I \end{bmatrix}$ where Q represents a txt matrix in the upper left corner, expressing transitions between transient states. O is an rxt matrix of zeros on the lower left corner. R is a non-zero txr matrix on the upper right corner, expressing transient states to absorbing states. I is an rxr identity matrix. The transition matrix is filled with gathered data. (i.e. frequencies of students in a particular state). The computation is done using the fundamental matrix F given as

$$F=(I-Q)^{-1}. \quad (3)$$

By letting the initial probability vector to be $X_0=(i_1, i_2, \dots, i_n)$ and probability vector after n repetitions of the experiment as $P_{ij}^{(n)}=X_0\ x\ P^n$ where P is the transition matrix, which is the Maximum Likelihood estimator (MLE) of the transition probabilities.

IV. THE MAXIMUM LIKELIHOOD ESTIMATOR (MLE)

The MLE can be derived as follows;

The transition matrix , P, is unknown and we estimate it from data. The parameters we wish to infer are the $m\ xm$ matrix entries P_{ij}, which are defined as; $P_{ij}=\Pr(X_{t+1}=j|X_t=i)$, we observe a sample from a chain, $x_1^n=x_1, x_2, \dots, x_n$. This realization of the random variable X_1^n. The probability of this realization is

$$\Pr(X_1^n=x_1^n)=\Pr(X_1=x_1)\prod_{t=2}^{n}\Pr(X_t=x_t|X_1^{t-1}=x_1^{t-1})$$

$$\Pr(X_1^n=x_1^n)=\Pr(X_1=x_1)\prod_{t=2}^{n}\Pr(X_t=x_t|X_{t-1}=x_{t-1})$$

Re-writing in terms of the transition probabilities P_{ij}, to get the likelihood of the transition matrix:

$L(P)=\Pr(X_1=x_1)\prod_{t=2}^{n}\Pr(x_{t-1}|x_t)$. Define N_{ij} as the number of times i is followed by j in X_1^n, and re-write the likelihood in terms of them

$$L(P) = \Pr\,(X_1 = x_1)\prod_{i=1}^{k}\prod_{j=1}^{k} P_{ij}^{n_{ij}}$$

To maximize this function we take the logs and simplify for optimization

$l(P) = logL(P) = logPr(X_1 = x_1)\sum_{i,j} n_{ij} logP_{ij}$

and taking the derivative $\frac{\partial L}{\partial P_{ij}} = \frac{n_{ij}}{P_{ij}}$, we get the MLE $\widehat{P}$, as $\frac{n_{ij}}{\hat{p}_{ij}} = \frac{n_{ij}}{\hat{P}_{i1}}\frac{n_{j1}}{\hat{p}_{i1}} = \frac{\hat{p}_{ij}}{\hat{p}_{i1}}$, this holds for $j \neq 1$, we conclude that $\hat{p}_{ij} \propto n_{ij}$, and in fact

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}} \quad (4)$$

Clearly, the choice of p_{i1} as the transition probability to eliminate in favor of the others is totally arbitrary and we get the same result for any other choice.

V. DISCUSSION

We consider two cohorts to come up with the two transition matrices which we combine by means of weighted averages to get the transition matrix for prediction

In cohort 1 we considered 274 students enrolled in form one having obtained the following grades

K.C.P.E Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	Total
Number of students	6	23	91	114	32	6	2	0	0	0	274

Table1: Frequencies of K.C.P.E Grades

Table 1 gives initial probability vector as:

$p_0 =$
[0.021 0.084 0.332 0.416 0.117 0.022 0.007 0.00 0.00 0.00 0.00]

, f_{ij}, number of students in this cohort who had been in state i after K.C.P.E. and who were in state j after doing their K.C.S.E exams is shown in table 2.

	A	A-	B+	B	B-	C+	C	C-	D+	D	W	Total
A	0	2	3	0	0	0	0	0	0	0	1	6
A-	0	6	5	2	0	1	0	0	0	0	9	23
B+	0	7	27	22	10	5	0	0	0	0	20	91
B	0	1	20	26	22	11	7	0	1	0	26	114
B-	0	0	3	11	2	8	1	0	1	0	6	32
C+	0	0	0	0	0	1	0	0	1	1	3	6
C	0	0	0	0	1	0	0	1	0	0	0	2
C-	0	0	0	0	0	0	0	0	0	0	0	0
D+	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0
Total	0	16	58	61	35	26	8	1	3	1	65	274

Table 2: Transition frequencies

It can be seen from the table that, of the 6 students who scored grade A in K.C.P.E., none them scored A in K.C.S.E, but 2 of them scored grade A-, 3 scored B+, and 1 had withdrawn from that school before siting her K.C.S.E exams. Of the 23 who scored grade A- at K.C.P.E, 6 scored A- at K.C.S.E, 5 scored B+, 2 scored grade B, 1 scored a C+ and 9 had withdrawn already before sitting for final exams. This data can be displayed as an 11 x 11 matrix of transition probabilities as shown below:

$$T_1 = \begin{bmatrix} 0 & 0.333 & 0.500 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.167 \\ 0 & 0.260 & 0.217 & 0.087 & 0 & 0.043 & 0 & 0 & 0 & 0 & 0.391 \\ 0 & 0.078 & 0.297 & 0.242 & 0.11 & 0.055 & 0 & 0 & 0 & 0 & 0.22 \\ 0 & 0.009 & 0.175 & 0.228 & 0.193 & 0.096 & 0.061 & 0 & 0.087 & 0 & 0.228 \\ 0 & 0 & 0.094 & 0.343 & 0.193 & 0.25 & 0.031 & 0 & 0.031 & 0 & 0.188 \\ 0 & 0 & 0 & 0 & 0.063 & 0.167 & 0 & 0 & 0.167 & 0.167 & 0.50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.50 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In Cohort 2 we considered 326 students enrolled in form one in a certain year having obtained the following grades

K.C.P.E Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	Total
Number of students	12	65	110	77	52	8	2	0	0	0	326

Table 3: Frequencies of K.C.P.E Grades

This gives initial probability vector as:

$$p_0 = [0.037\ \ 0.199\ \ 0.337\ \ 0.236\ \ 0.159\ \ 0.025\ \ 0.006\ \ 0.00\ \ 0.00\ \ 0.00\ \ 0.00]$$

, f_{ij}, number of students in this cohort who had been in state i after K.C.P.E,. and who were in state j after doing their K.C.S.E exams is shown in table 4.

	A	A-	B+	B	B-	C+	C	C-	D+	D	W	Total
A	1	1	4	2	2	1	0	0	0	0	1	12
A-	4	5	13	14	8	5	2	0	0	0	14	65
B+	0	3	17	22	21	15	5	0	1	0	26	110
B	0	2	4	13	9	10	6	2	0	0	31	77
B-	0	0	3	2	5	11	3	4	2	0	22	52
C+	0	0	0	0	2	2	1	0	0	1	3	8
C	0	0	0	0	1	0	0	0	0	1	1	2
C-	0	0	0	0	0	0	0	0	0	0	0	0
D+	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0
Total	5	11	41	53	47	44	17	6	3	1	98	326

Table 4 Transition frequencies

It can be seen from the table that, of the 12 students who scored grade A in K.C.P.E., one them scored A in K.C.S.E, one of them scored grade A-, but 4 scored B+, 2 scored B, two scored B-, one scored C+ and 1 had withdrawn from that school before siting her K.C.S.E exams. Of the 65 who scored grade A- at K.C.P.E, 4 scored A at K.C.S.E, 5 scored A-, 13 scored B+, 14 scored a B, 8 scored B-, 5 scored a C+, 2 scored a C, and 14 had already withdrawn before sitting for the final exams. This data can be displayed as an 11x11 matrix of transition probabilities shown below:

$$T_2 = \begin{bmatrix} 0.083 & 0.083 & 0.333 & 0.167 & 0.083 & 0 & 0 & 0 & 0 & 0 & 0.083 \\ 0.062 & 0.077 & 0.200 & 0.215 & 0.123 & 0.076 & 0.031 & 0 & 0 & 0 & 0.215 \\ 0 & 0.027 & 0.155 & 0.200 & 0.191 & 0.136 & 0.450 & 0 & 0.009 & 0 & 0.236 \\ 0 & 0.026 & 0.052 & 0.169 & 0.117 & 0.130 & 0.078 & 0.026 & 0 & 0 & 0.400 \\ 0 & 0 & 0.058 & 0.038 & 0.096 & 0.251 & 0.058 & 0.077 & 0.038 & 0 & 0.423 \\ 0 & 0 & 0 & 0 & 0.25 & 0.25 & 0.125 & 0 & 0 & 0 & 0.375 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.500 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

VI. RESULTS

The probability transition matrix for this case is obtained by combining the above transition matrices by means of weighted average using the formula:

$$P(X) = PR(x) \times [1 - JN \div (RN + JN)] + pJ(x) \times [1 - RN \div (RN + BN)] \quad (5)$$

Where, $p(x)$ represents the probability in the given row that we are working out. $PR(x)$ is the probability value in the same row at cell x in the cohort R matrix. $PJ(x)$ is the probability value in the same row at cell x in the cohort J matrix. RN is the total number of counts for the same row in matrix R and JN is the total number of counts for the same row in matrix J. Using the above, we obtain the transition matrix as below:

$$T = \begin{bmatrix} 0.053 & 0.163 & 0.387 & 0.107 & 0.107 & 0.053 & 0 & 0 & 0 & 0 & 0.011 \\ 0.060 & 0.127 & 0.226 & 0.187 & 0.087 & 0.075 & 0.22 & 0 & 0 & 0 & 0.291 \\ 0 & 0.052 & 0.218 & 0.218 & 0.154 & 0.100 & 0.027 & 0 & 0.005 & 0 & 0.231 \\ 0 & 0.064 & 0.128 & 0.205 & 0.160 & 0.112 & 0.0067 & 0.010 & 0.054 & 0 & 0.376 \\ 0 & 0 & 0.070 & 0.154 & 0.082 & 0.226 & 0.047 & 0.048 & 0.035 & 0 & 0.334 \\ 0 & 0 & 0 & 0 & 0.182 & 0.228 & 0.091 & 0 & 0.046 & 0.046 & 0.409 \\ 0 & 0 & 0 & 0 & 0.250 & 0 & 0 & 0.250 & 0 & 0.250 & 0.250 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Using Matlab software and canonical form of the transition matrix, the following results were obtained

	A	A-	B+	B	B-	C+	C
A	0.053	0.163	0.387	0.107	0.107	0.053	0
A-	0.060	0.127	0.226	0.187	0.087	0.075	0.22
B+	0	0.052	0.218	0.218	0.154	0.100	0.027
B	0	0.064	0.128	0.205	0.160	0.112	0.0067
B-	0	0	0.070	0.154	0.082	0.226	0.047
C+	0	0	0	0	0.182	0.228	0.091
C	0	0	0	0	0.250	0	0

Table 5: Probabilities of transient to transient state

In the table the rows indicate the transient state I and columns transient state j. A student who scored grade A in K.C.P.E has a chance of scoring an A in K.C.S.E with probability 0.053.

	C-	D+	D	W
A	0.073	0.068	0.064	0.145
A-	0.125	0.068	0.115	0.123
B+	0.068	0.071	0.059	0.933
B	0.085	0.115	0.067	0.270
B-	0.100	0.084	0.055	0.822
C+	0.056	0.082	0.104	0.777
C	0.275	0.021	0.264	0.456

Table 6: Probabilities of absorption

Table 6 shows that a student who scored an A in K.C.P.E has 0.073 chance of scoring a C-, 0.068 chance of scoring a D+ and 0.143 chance of withdrawal.

Fundamental matrix

The elements of the fundamental matrix are given in the matrix below

$$F = \begin{bmatrix} 1.074 & 0.282 & 0.741 & 0.510 & 0.496 & 0.416 & 0.177 \\ 0.078 & 1.231 & 0.527 & 0.541 & 0.495 & 0.417 & 0.383 \\ 0.008 & 0.126 & 1.446 & 0.519 & 0.472 & 0.413 & 0.161 \\ 0.008 & 0.128 & 0.322 & 1.465 & 0.455 & 0.400 & 0.193 \\ 0.002 & 0.034 & 0.178 & 0.310 & 1.303 & 0.453 & 0.135 \\ 0.0006 & 0.009 & 0.047 & 0.082 & 0.346 & 1.415 & 0.154 \\ 0.0005 & 0.008 & 0.0045 & 0.077 & 0.326 & 0.113 & 1.034 \end{bmatrix}$$

The column vector is the expected number of steps before the chain is absorbed.

$$t = [3.6963 \quad 3.6711 \quad 3.1445 \quad 2.9702 \quad 2.4151 \quad 2.0537 \quad 1.6038]^T$$

It can be seen that a student who had scored an A has the highest number of steps, 3.69363 in order to move to an absorbing state. It is more difficult for bright students to perform poorly.

VII. PREDICTION OF PERFORMANCE

The objective of this paper was to forecast the expected future academic performance of a cohort of students in their examinations using the previous immediate performance. For instance having obtained the transition matrix, T, we now use it to predict performance of a cohort of students. The probability vector,

$$p_0 = [0.018 \quad 0.103 \quad 0.293 \quad 0.263 \quad 0.220 \quad 0.091 \quad 0.012]$$

. Pre multiplication of the probability vector on the transition matrix, T, yields the prediction vector as

$$p_{ij} - final = [0.071 \quad 0.048 \quad 0.143 \quad 0.173 \quad 0.136 \quad 0.138 \quad 0.067 \quad 0.016$$

The actual probability vector as seen from their K.C.S.E results is

$$p_{ij} - actual = [0.006 \quad 0.045 \quad 0.133 \quad 0.172 \quad 0.154 \quad 0.133 \quad 0.066 \quad 0.012 \quad 0.009 \quad 0.003 \quad 0.266]$$

This is gives table 7.

Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	W	Total
Prediction	2	15	44	57	51	44	22	4	3	1	88	331
Actual performance	2	16	47	57	50	45	22	5	9	3	103	358

Table7: Prediction versus actual performance

It can be seen that the prediction is approximately equal if not the same with the actual performance. There is a very small margin error in the number of grades predicted.

CONCLUSION

We conclude that the Markov model is the ideal model for predicting performance basing on the previous performance. By using this model, it is possible to award the deserving grade to students hence promote high quality, the model is also useful in projecting the number of dropouts in a given institution.

RECOMMENDATION

We recommend that examination bodies such as KNEC, Ministry of education and other stakeholders in education sector to consider using the Markov model in compiling the final grade of student scores especially when there is uncertainty on the credibility of the results.

REFERENCES

[1] Adeleke R. A., Oguntuase K. A. and Ogunsakin R. E. (2014). Application of Markov chain to the assessment of students' admission and academic performance in Ekiti State University, Nigeria. *International journal of scientific and technology research*. Vol 3, issue 7 349 – 357.

[2] Broadfoot, P. and Black P. (2004). *Redefining assessment. The first ten years of assessment in Education,* Assessment in Education.

[3] Burke G. (1872). A study in the economics of education with particular reference to the supply of secondary teachers for government schools in Victoria, *PhD thesis, Monash University.*

[4] Chan Cansong (2005). Application of homogenous Markov chain in the public teaching of impact assessment, *PhD thesis, Luoyang University*.

[5] Chen Xiangyang (2007). Application of Markov chain in educational assessment, *PhD thesis, University of mathematics*

[6] Imboga O. H., George O. O. and Humphreys H. M. (2013). Construction of Markov transition matrices for cohorts of students in BSc. Actuarial Science, *MSc. thesis, Jomo Kenyatta University, Kenya.*

[7] Nicolls M. G. (2009). The use of markov models as an aid to evaluation, planning and benchmarkingof Doctoral programmes, Journal of operational research society 60 1183 – 1190.