Understanding Sports Participation through Markov Chain Analysis

Riley Morgan

29 April 2023

1 Introduction

Sports participation is an important aspect of many people's lives, starting from middle school on to high school and sometimes leading to professional careers. However, not all people continue to participate in sports throughout their academic and professional lives. To track sports participation, we will be using a Markov Chain model that considers participation in sports as a sequence of states. The absorbing state in this model is the option for people to quit at any time. Within the model, people will be able to continue playing sports at the next level or quit playing sports. Probabilities for these movements have been discovered in sources [1] and [2]. This project aims to analyze the probabilities of transitioning between different states, the probability of absorption (quitting sports), and the time taken to reach the absorbing state (quit sports). By using Markov Chain, we will be able to provide valuable information in regards to understanding sports participation overtime.

2 Methodology

Markov Chain is a mathematical model that is commonly used to simulate processes that involve a series of states. In this project, we will use Markov Chain to track sports participation of individuals over time. Each state in the Markov Chain represents a different level of participation, starting from middle school to 9th grade, 10th grade, 11th grade, 12th grade, college, professional and no participation. The absorbing state in this model is the decision of an athlete to quit sports at any point in time. The transition between these stages and their rates can be seen in figure 1.

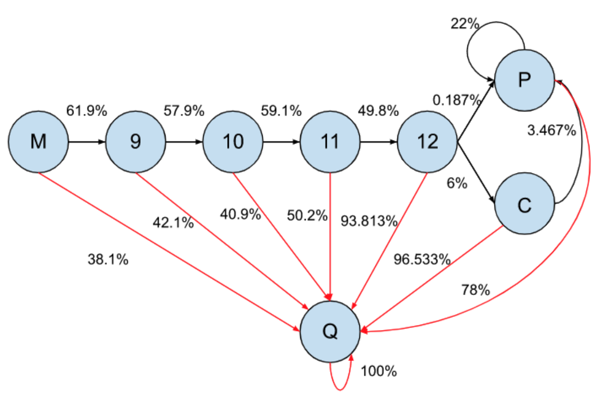

Figure 1: Sports Participation

2.1 Transition Matrix

To apply the Markov Chain method we need to first start by creating a transition matrix. The transition matrix is a key component in the Markov Chain method. The matrix defines the probabilities of transitioning between different states in a system over time. Each row of the matrix represents the current state, and each column represents the next possible state. The elements in the matrix represent the probabilities of transitioning from the current state to the next possible state. By analyzing the transition matrix, we can understand how a system evolves over time and predict future states. Using figure 1 and our given rates we are able to create our transition matrix for this model which is set equal to T and can be viewed below.

$$T = \begin{bmatrix} 0 & 0.619 & 0 & 0 & 0 & 0 & 0 & 0.381 \\ 0 & 0 & 0.579 & 0 & 0 & 0 & 0 & 0.421 \\ 0 & 0 & 0 & 0.591 & 0 & 0 & 0 & 0.409 \\ 0 & 0 & 0 & 0 & 0.498 & 0 & 0 & 0.502 \\ 0 & 0 & 0 & 0 & 0 & 0.06 & 0.00187 & 0.93813 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.03467 & 0.96533 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.22 & 0.78 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

In this matrix, each category is ordered from left to right and top to bottom in the following order: middle school, 9th grade, 10th grade, 11th grade, 12th grade, college, professional, and quit. In each state, an athlete can choose to quit or move on to the next stage. Once an athlete has chosen to quit they will remain in the quit compartment for the rest of the simulation as the quit compartment is an absorbing state. Additionally, once an athlete has become a professional they have the option to remain a professional or quit playing sports. We will continue using this transition matrix for further computation.

2.2 Initial State Matrix

The initial state matrix is an important piece of the Markov Chain method. This matrix defines the proportion of the population that is in each state at the start of the process. The initial state matrix can be multiplied by the transition matrix to calculate the probabilities of the system being in each state at different points in time. Therefore, the initial state matrix plays an important role in determining the future behavior of the system. In our model, the initial state matrix is set equal to *initial*.

$$initial = \begin{bmatrix} 0.22 & 0.20 & 0.19 & 0.18 & 0.16 & 0.04 & 0.01 \end{bmatrix}$$

2.3 Transient State Matrix

After forming our transition matrix we will aim to create our transient state matrix which includes the probabilities from our transient states. This matrix will eliminate the row and column from our quit state which is our only absorbing state. We will use this matrix in order to form the fundamental matrix of our T matrix. Our transient state matrix is equal to Q as seen below.

$$Q = \begin{bmatrix} 0 & 0.619 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.579 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.591 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.498 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.06 & 0.00187 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.03467 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.22 \end{bmatrix}$$

2.4 Identity Matrix

In order to form our fundamental matrix of matrix T, we will also need an identity matrix of the same dimensions of our Q matrix. An identity matrix is one filled with 1's along the diagonal and all other spots filled with 0. Our identity matrix is set equal to I and can be seen below.

$$
I = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

2.5 Fundamental Matrix of T

Now that we have gathered our identity matrix (I) and our transient state matrix (Q), we can now form our fundamental matrix of matrix T. This matrix will be equal to N and will be calculated using the following equation:

$$N = (I - Q)^{-1}$$

Applying this equation we get the following:

$$
N = \left(\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} - \begin{bmatrix}
0 & 0.619 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.579 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.591 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.498 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.06 & 0.00187 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.03467 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.22 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \right)^{-1}
$$

$$
N = \begin{bmatrix}
1 & 0.62 & 0.36 & 0.21 & 0.11 & 0.01 & 0 \\
0 & 1 & 0.58 & 0.34 & 0.17 & 0.01 & 0 \\
0 & 0 & 1 & 0.59 & 0.29 & 0.02 & 0 \\
0 & 0 & 0 & 1 & 0.5 & 0.03 & 0 \\
0 & 0 & 0 & 0 & 1 & 0.06 & 0.01 \\
0 & 0 & 0 & 0 & 0 & 1 & 0.04 \\
0 & 0 & 0 & 0 & 0 & 0 & 1.28
\end{bmatrix}
$$

The fundamental matrix plays a critical role in understanding the transient behavior of our system. The fundamental matrix provides information on how many times the system is likely to visit each non-absorbing state before it enters an absorbing state. The elements of the fundamental matrix are used to calculate the expected time until absorption and the probability of absorption.

2.6 Probability of Absorption

The probability of absorption matrix represents the probability that a system will reach the absorbing state starting from each of its non-absorbing states. To calculate the probability of absorption we will use the following equation:

$$B = N * R$$

In this equation, N represents our fundamental matrix of T, R represents the probability of reaching the quit state from an initial non-absorbing state, and B is our resulting probability of absorption matrix.

In our system, the R matrix is the remaining values from the transition matrix that were not added to our Q matrix. Thus, it holds the probabilities that given a current non-absorbing state, that the next step will be to the absorbing state. The R matrix for our model can be seen below.

$$R = \begin{bmatrix} 0.381 \\ 0.421 \\ 0.409 \\ 0.502 \\ 0.93813 \\ 0.96533 \\ 0.96 \end{bmatrix}$$

Using our probability of absorption equation, we get the following:

$$B = \begin{bmatrix} 1 & 0.62 & 0.36 & 0.21 & 0.11 & 0.01 & 0 \\ 0 & 1 & 0.58 & 0.34 & 0.17 & 0.01 & 0 \\ 0 & 0 & 1 & 0.59 & 0.29 & 0.02 & 0 \\ 0 & 0 & 0 & 1 & 0.5 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.06 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0.04 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.28 \end{bmatrix} * \begin{bmatrix} 0.381 \\ 0.421 \\ 0.409 \\ 0.502 \\ 0.93813 \\ 0.96533 \\ 0.96 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The resulting matrix B, represents the probability that given an initial state, when time goes on the group will eventually move to the absorbing state.

2.7 Expected Time Until Absorption

Using the matrices calculated in the previous subsections, we are able to compute the expected time until absorption. The expected time until absorption matrix provides valuable information about the system's behavior. It represents the expected time it takes for the system to reach an absorbing state from each non-absorbing state. To calculate the expected time until absorption we use the following equation:

$$t = Nc$$

In the given equation, c represents a column vector with all its elements equal to 1, and the resulting matrix is represented by t. When applied to our model, this equation gives the following outcome:

$$t = \begin{bmatrix} 1 & 0.62 & 0.36 & 0.21 & 0.11 & 0.01 & 0 \\ 0 & 1 & 0.58 & 0.34 & 0.17 & 0.01 & 0 \\ 0 & 0 & 1 & 0.59 & 0.29 & 0.02 & 0 \\ 0 & 0 & 0 & 1 & 0.5 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.06 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0.04 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.28 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$t = \begin{bmatrix} 2.31 \\ 2.1 \\ 1.9 \\ 1.53 \\ 1.07 \\ 1.04 \\ 1.28 \end{bmatrix}$$

The resulting matrix t, represents the average time it takes for each non-absorbing state in the model to reach the absorbing state.

The Markov Chain method allows us to calculate the probabilities of transitioning between different states, expected time until absorption, and the probability of absorption. These metrics provide valuable insights into how our system of states operates and the movement of values throughout the system over time. With this knowledge, we can understand how athletes participation in sports changes overtime and address our results.

3 Results

3.1 Probability of Absorption

The probability of absorption matrix provides insights into the likelihood that an athlete will end up in the absorbing state given time and an initial state. We can reference the equation and resulting matrix B from section 2.6. The probability of absorption matrix for our model can be seen below.

$$B = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Our probability of absorption matrix B, shows that given any initial state, each athlete will eventually quit and join the absorbing state. This is to be expected as no athlete will play a sport forever. Thus, given time every athlete in the model will eventually quit. This leads us to our next question, which is, how long will it take for our athletes to quit and join the absorbing state? This question can be answered using the expected time until absorption matrix in section 2.7.

3.2 Expected Time Until Absorption

The expected time until absorption matrix provides insights into the expected time duration that each group of athletes is likely to participate in sports before quitting. Using the methodology explained in section 2.7, we found the expected time until absorption matrix to be:

$$t = \begin{bmatrix} 2.31 \\ 2.1 \\ 1.9 \\ 1.53 \\ 1.07 \\ 1.04 \\ 1.28 \end{bmatrix}$$

The estimated time until absorption matrix provides valuable information about the average time duration it takes for athletes in each stage to quit and move to the absorbing state. By analyzing this matrix, we discovered that, on average, middle school students take 2.31 time steps to quit, while 9th graders take 2.1 time steps, and the duration decreases for higher levels of athletes except for professionals. This information is useful in identifying trends in sports participation and how time affects participation levels. By understanding these patterns, we can develop plans that can help prolong sports participation, leading to improved physical and mental health outcomes for athletes. Overall, the estimated time until absorption matrix is a powerful tool in analyzing the dynamics of sports participation and identifying ways to encourage longer participation.

4 Conclusion

In conclusion, this project aims to track the participation of individuals in sports using a Markov Chain model. The analysis of this model provides insights into the probabilities of transitioning between different levels of sports and the probability of quitting sports altogether. The probability of absorption matrix shows that every athlete will eventually quit playing sports, but our expected time until absorption matrix allows us to analyze how long it takes for athletes to quit given there participation level. The results of this analysis can be used to develop strategies to encourage and improve sports participation. Ultimately, this project can help us better understand how sports participation changes over time and the factors that influence sports participation.

References

[1] From High School to Pro Statistics - When You Can't Go pro... https://whenyoucantgopro.weebly.com/uploads/2/6/5/2/26529572/from$_{h}igh_{s}chool_{t}o_{p}ro_{s}tatistics.pdf$.

[2] Riser-Kositsky, Maya, and Holly Peele. "Statistics on School Sports: How Many Students Play Sports? Which Sports Do They Play?" Education Week, Education Week, 26 Oct. 2022, https://www.edweek.org/leadership/statistics-on-school-sports-how-many-students-play-sports-which-sports-do-they-play/2021/07.