Application of Markov Chains in Sports Rankings

By: Jason Katz, Tobenna Okwara, Benjamin Caraballo, Melissa Heath

Abstract

There are many different ranking methods that can be used to predict the outcome of different sporting events. This project will focus specifically on the use of Markov chains for ranking teams in the NBA playoffs. Our goal is to use the Markov method to make more accurate predictions of the outcomes of the playoffs based off of regular season performances than the NBA's general rankings. This will involve applying multiple topics that our group studied. These topics consist of basic Markov Chains, the Markov method for rankings, voting matrices, accounting for sensitivity, and solving for stationary distributions. Our investigation concluded that the Markov Method for rankings outperformed the NBA's rankings with 1.79% better accuracy.

Section 1: Introduction

There have been many different methods and algorithms developed to determine the rankings of different sports teams. Our project used the application of Markov chains to create a ranking system. The basic form of a Markov chain links events so that the probability of future events depends only on the current state that system is in. Specifically, the main model that we used was the Markov Method. The basics of this algorithm is to produce a square matrix that represents head to head competitions between teams. Every entry in the matrix is the result of each individual competition where the losing team votes for the winning team. This matrix is then converted into a stochastic matrix in order to then find the dominant Eigenvector or probability vector. This will allow us to obtain the ranking results of our model.

Markov chains were originally introduced by Andrey Markov Jr. [2] and used in a variety of fields. The first algorithms that were developed for ranking systems in sports where the

Massey method as well as the Colley method [1]. However, these methods used a system of linear equations to develop its ratings and did not utilize Markov chains. Markov chains became utilized as they took into account individual results of competitions, unlike the other methods previously developed. The method used in this paper, the Markov Method, was first developed by Kvam and Sokol [1]. The method originally used point scores and home court advantages to rank men's basketball teams in division 1. This method was then used by Govan in 2008 for the purpose of ranking sports teams [1].

The application of these algorithms have been seen in many other contexts too. In addition to ranking sports teams the Markov method has also been utilized for recruiting purposes in ranking individual players [2]. This method also ranges across a wide spectrum of fields. Besides athletics, a couple examples would be Google using these algorithms to rank their webpages [2]. In science, this can be used to identify genes in DNA. As well as in physics, this method has been used in systems of interacting particles [2].

For our project, our team wanted to investigate the sports ranking system that the NBA uses and compare it to our ranking system using the Markov Method. Our goal is to use the Markov method to provide more accurate rankings for the playoffs, while also taking into account the possibility of upsets. After running our simulation we want to show that our method will rank more teams accurately then the NBA system by comparing each method to the final results

Section 2: Mathematical Background

Before introducing the concept of Markov chains, we must first define a stochastic process, which we will only be using in discrete time. A discrete-time stochastic process is a sequence of random variables $\{X_n, n = 0, 1, \ldots, \}$ that occur in the same probability space (Ω, F, $\mathbf{P}$) and take on values within the same state space, S. Ω is sample space, which contains all possible outcomes. F is the set of events, which each contain zero or more outcomes from Ω. **P** assigns a probability to each event in F. The state space is the range of possible values for each of the random variables [3].

A Markov chain is a discrete-time stochastic process where, for all n ϵ **N** and all $i_0,\ldots,i_{n-1},i,j \epsilon S$,

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1},\ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i) = P_{ij} \text{ [3].}$$

In other words, the probability of being in state j at step n+1 only depends on the state at step n, where the chain was in the previous step. All states prior to state n are irrelevant to P_{ij}, which is the probability of moving from state i to state j. Our model utilizes the homogenous process, where

$$\mathbf{P}(X_{n+1} = j \mid X_n = i) = \mathbf{P}(X_1 = j \mid X_0 = i)$$

for all n ϵ **N** and $i,j \epsilon S$ [3]. In simpler terms, the probability of moving from state i to state j is the same at every step of the process, rather than changing at each step. To model a Markov chain, we use a transition probability matrix $P = [P_{ij}]_{i,j \epsilon S}$, which will always be square. On the state space S, we must have $P_{ij} \geq 0$ and $\sum_{j \epsilon S} P_{ij} = 1$ for all $i \epsilon S$ [3]. Equivalently, the probability of moving from state i to state j cannot be negative, and, while in state i, the chain must take another step to some state j in the state space. The probability of the chain not taking another step is 0.

Another feature of Markov chains is the stationary distribution. This is a probability distribution on the state space S of a Markov chain that remains unchanged by the transition probability matrix as the chain progresses through steps. For a n x n transition probability matrix P, the stationary distribution π is a 1 x n vector where $\pi = \pi P$. The elements of π sum to one and are all between 0 and 1. Stationary distributions can model the long-time behavior of a Markov chain, which is of relevance to our ranking model. If a Markov chain is irreducible, aperiodic, and positive recurrent, then, as the number of steps, t, gets very large, $(P_{ij})^t = \pi_j$. The probability of going from i to j in t steps converges to the jth element in π, and therefore each row of P converges to π. Irreducibility requires that, through some sequence of steps, any state i can reach any state j. Aperiodicity means that the greatest common denominator of t is 1, where $(P_{ii})^t > 0$. In other words, the greatest common denominator of the various numbers of steps the chain can take to get from state i back to state i is 1. A Markov chain is positive recurrent if the expected number of steps it takes to get from state i back to state i is less than infinity. With these three attributes satisfied, the stationary distribution, in the long run, will represent the proportion of time spent in each state in S [3].

The Markov method is a technique of ranking teams utilizing Markov chains. Essentially, in each individual game, the losing team casts a vote for the winning team, the (0,1) method. The number of votes that one team has cast for each of the others is then recorded as an element in a matrix representing the head-to-head competitions. Each team has exactly one row and one column in this matrix. It is then normalized, making the sum of all elements in a row equal to 1, $\sum_{j \epsilon S} P_{ij} = 1$ for all $i \epsilon S$ [3]. This is now our transition probability matrix, which we will use to find the stationary distribution. Recall that the stationary distribution, π, is a vector where $\pi = \pi P$. Solving for π, we use the elements of this vector to rank teams. Teams are ranked, first to last, according to the value of their element in π; the team with the greatest value comes first [1].

There are three axioms which all ranking methods should fulfill to be both fair and complete. Under Axiom I, victories over better teams should carry greater weight, which is a major advantage of the Markov method. Since the rating vector, the stationary distribution, originates from the head-to-head voting matrix, quality of wins matters. For team j, its rating is $\pi_j = \sum p_{ij} * \pi_i$, which will favor victories over better teams. Axiom II states that teams must always have incentive to win. Losing should never improve a team's ranking and winning must never lower a team's ranking, which the Markov method can violate in certain cases due to its sensitivity to small changes in the data. The ratings are very reactive to upsets. In certain cases, team A, with one loss, can be ranked equally with team B, who beat them, even if that was team B's only win. Further, team A can improve both its rating and ranking by losing another game, violating Axiom II. Axiom III requires that order of matches should have no effect on rankings, which holds for the Markov method, since it uses a static matrix of total wins [1].

To fix the Markov method and satisfy Axiom II, Vaziri et al developed the (1,α) method. In each game, the winner votes one time for the loser, while the loser votes α times for the winner, where α > 1. This mitigates the Markov method's sensitivity by creating a fairer vote ratio. Smaller values of α will remove the incentive to lose, satisfying Axiom II. As α gets very large, though, it converges to the (0,1) method and Axiom II will be violated again. The (1, α) method is an extension of the Markov method, so it will still fulfill Axioms I and III [1].

Section 3: Computational Output

Real world applications of Markov Chains can be found in ranking methods. This paper looked into ranking methods in sports, particular basketball. Data from all NBA seasons were downloaded. To obtain this data, heavy web scraping was performed using Python. First, results from all regular season games were extracted from the site basketball reference [4]. Next, playoff results and final regular season standings were extracted. This process was conducted for all seasons dating back to the NBA/ABA merger in 1976. It was decided to generate rankings based on regular season results (one set of rankings for each season). To access how good a ranking method is, the rankings for each season were used to predict the playoff series that were played each season. The accuracy of a ranking method is what percent of the time the higher ranked team in the method's rankings won the series. In total, there were 587 different playoff series tested on. The Markov Method for rankings was compared to the standard NBA rankings (seeds), based on final regular season standings. To generate the Markov Rankings for each season, a voting matrix was calculated using the head to head results of all regular season games. The process for generating a voting matrix is defined in section 2 above. Figure 1 below shows a sub-sample of an initial voting matrix generated from regular season results.

	Atlanta Hawks	Boston Celtics	Brooklyn Nets	Charlotte Hornets	Chicago Bulls	Cleveland Cavaliers	Dallas Mavericks	Denver Nuggets	Detroit Pistons	Golden State Warriors
Atlanta Hawks	0	1	2	3	1	1	0	0	2	2
Boston Celtics	2	0	0	0	2	3	0	2	1	1
Brooklyn Nets	2	4	0	3	3	3	2	1	1	2
Charlotte Hornets	1	4	1	0	2	4	0	0	3	2
Chicago Bulls	3	2	1	1	0	0	2	2	2	1
Cleveland Cavaliers	3	1	0	0	4	0	1	1	2	1
Dallas Mavericks	2	2	0	2	0	1	0	3	2	3
Denver Nuggets	2	0	1	2	0	1	1	0	2	2
Detroit Pistons	1	3	2	1	2	2	0	0	0	2
Golden State Warriors	0	1	0	0	1	1	0	1	0	0

Figure 1: A sub-sample (10 of 30 teams) of a voting matrix before normalization. The full matrix is 30x30.

After the initial voting matrix is calculated using the chosen value of alpha, the rows are normalized to sum to one. Figure 2 shows the matrix changes after normalization.

	Atlanta Hawks	Boston Celtics	Brooklyn Nets	Charlotte Hornets	Chicago Bulls	Cleveland Cavaliers	Dallas Mavericks	Denver Nuggets	Detroit Pistons	Golden State Warriors
Atlanta Hawks	0.000000	0.025641	0.051282	0.076923	0.025641	0.025641	0.000000	0.000000	0.051282	0.051282
Boston Celtics	0.068966	0.000000	0.000000	0.000000	0.068966	0.103448	0.000000	0.068966	0.034483	0.034483
Brooklyn Nets	0.032258	0.064516	0.000000	0.048387	0.048387	0.048387	0.032258	0.016129	0.016129	0.032258
Charlotte Hornets	0.021739	0.086957	0.021739	0.000000	0.043478	0.086957	0.000000	0.000000	0.065217	0.043478
Chicago Bulls	0.073171	0.048780	0.024390	0.024390	0.000000	0.000000	0.048780	0.048780	0.048780	0.024390
Cleveland Cavaliers	0.096774	0.032258	0.000000	0.000000	0.129032	0.000000	0.032258	0.032258	0.064516	0.032258
Dallas Mavericks	0.040816	0.040816	0.000000	0.040816	0.000000	0.020408	0.000000	0.061224	0.040816	0.061224
Denver Nuggets	0.047619	0.000000	0.023810	0.047619	0.000000	0.023810	0.023810	0.000000	0.047619	0.047619
Detroit Pistons	0.022222	0.066667	0.044444	0.022222	0.044444	0.044444	0.000000	0.000000	0.000000	0.044444
Golden State Warriors	0.000000	0.066667	0.000000	0.000000	0.066667	0.066667	0.000000	0.066667	0.000000	0.000000

Figure 2: The same matrix as the one in Figure 1, except the rows have been normalized to sum to one.

This matrix now represents the transition probability matrix of the Markov Chain for the regular season results for all teams. Next, the stationary distribution for the Markov Chain is calculated by deriving the dominant eigenvector for the matrix. This stationary distribution represents the ratings of each team. Figure 3 below shows the ratings of the top 10 teams in 2017, derived from the stationary distribution of the matrix in Figure 2.

	1	2	3	4	5	6	7	8	9	10
Team	Golden State Warriors	San Antonio Spurs	Los Angeles Clippers	Houston Rockets	Memphis Grizzlies	Utah Jazz	Boston Celtics	Cleveland Cavaliers	Oklahoma City Thunder	Toronto Raptors
Rating	0.0555	0.0548	0.045	0.044	0.0429	0.0416	0.0405	0.0397	0.039	0.0385

Figure 3: These are the ratings of the top 10 teams for 2017 according the Markov Method for ranking

Once the ranking for each team is generated using each team's rating, predictions for the playoff series were made and compared to the predictions made by the official NBA seeding. As described in section 2 above, different alpha values when generating the voting matrix lead to different sensitivity to upsets for the Markov Method. To find the optimal alpha value, rankings were generated for a span of alpha values and those rankings were used to make predictions. Figure 4 below shows the prediction results the Markov Method, testing different alpha values compared to the seed rankings.

Alpha	0.00	1.01	1.1	1.25	1.50	2.5	5.00	10.00	25.00	50.00	100.00	250.00
Markov	70.95	73.24	73.3	73.79	73.69	73.3	72.58	71.54	71.44	71.28	71.12	70.95
Seed	72.00	72.00	72.0	72.00	72.00	72.0	72.00	72.00	72.00	72.00	72.00	72.00

Figure 4: The prediction accuracy of the Markov Method using different values of alpha, compared to the accuracy of using the seed rankings. As alpha grows large, the predictions approach that of an alpha of 0.

The peak for accuracy occurs at an alpha of 1.25, where the accuracy of the Markov Method is 1.79% higher than that of the seed rankings. Given the size of the sample, this is a significant difference, and it can be concluded that the Markov Rankings are better than using NBA seeding. Figure 5 below shows a graphs of accuracies across different alpha values.

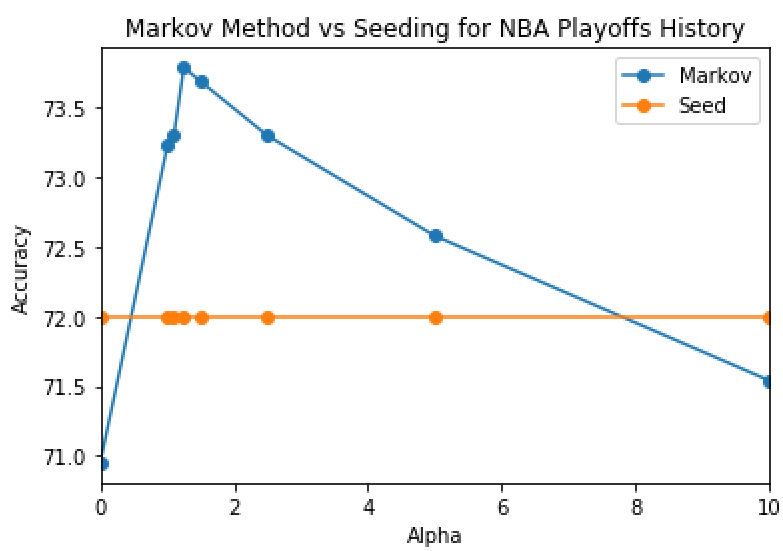

Figure 5: Graph showing the accuracy of the Markov Method vs seed rankings for different levels of alpha.

Section 4: Relation to Other Methods

Since our project topic revolved around using the Markov ranking method to evaluate basketball teams, the main relation to class material was through the use of Markov Chains and their properties. The topic of probability matrices as covered in class was fundamental to the ranking method. Each team's head to head performance against other teams was used to create a voting matrix for that team with each win counting as a "vote" for that team. Once normalized this voting matrix functioned as the transition probability matrix for that team's regular season [1]. The Markov ranking method also makes use of another topic covered in class, the stationary distribution. Each team's stationary distribution was derived from their probability matrix, allowing a rating to be calculated for each team [1].

In the process of reading through our project paper, we were introduced to several mathematical concepts not covered in class. One of these was Logistic Regression Markov Chain method (LRMC) which uses Logistic Regression to factor point scores and home court advantage into the transition probability matrix [1]. Another method discussed was the PageRank algorithm. Originally developed to rank web pages, it works by modeling each webpage as a node, and each hyperlink between them as an edge, then constructing an adjacency matrix of 1's and 0's depending on if a link exists between nodes, and finally calculating a rating vector from this matrix [1]. The Park-Newman method was also discussed as a unique variation of Markov chain rankings. This method counted for "indirect wins" by labeling as scenario where team X beats team Y, and team Y beats team Z, as a second degree win for team X over team Z [1]. Learning about these many different methods provided us with a broad perspective on the different ways that Markov Chain ranking methods can be used and

Section 6: Future Work

A couple things that our team discussed but were unable to fully explore is how the accuracy of the Markov Method and our simulation was for other sports as well. An interest in seeing how the rankings would turn out for other professional sports teams like baseball or football. Another area that we spoke about was how to get player injuries and other factors into our ranking system instead of just head to head wins and losses and using the alpha to account for the sensitivity of upsets. However, our team did not get to explore these areas further as our team did not have the time to analyze other professional sports by our project deadline. Taking into consideration player injuries and other factors that affects wins vs. losses would have needed much more involved coding that our team also did not have the time to do. Given more time our team would have liked to explore these areas further as this was a topic that we all found very interesting and enjoyed working on.

Section 7: Bibliography

[1] Vaziri, Baback. “Markov-based ranking methods” (2016). *Open Access Dissertations*. 721. http://docs.lib.purdue.edu/open_access_dissertations/721

[2] Hayes, Brian. “First Links in the Markov Chain” (2017). *American Scientist*. https://www.americanscientist.org/article/first-links-in-the-markov-chain

[3] Lecture Notes, Kavita Ramanan, APMA 1200, Brown University, Spring 2018

[4] “Basketball Statistics and History.” Basketball-Reference.com, www.basketball-reference.com/.