Markovian Models of Basketball

Chris Harding, Tim Najar
University of Michigan
CMPLXSYS 511 - Final Project
J.Dunworth
4/27/24

Abstract

This project presents 2 methods of simulating outcomes of professional basketball (NBA) games in a Markovian framework. An approach using aggregate team statistics as transition probabilities between team possessions, and an approach using player tracking passing statistics as transition probabilities of player passing. These models are used to simulate the 2023-24 NBA season.

1 Introduction

Basketball is a team sport where two teams send five players at a time to try to score as many points as they can before time runs out (National Basketball Association, 2023). A 2023 study found that 46% of American adults have placed at least one sports bet in the past year and basketball is the most preferred sport to bet on (Rodgers, 2023). Predicting the outcome of basketball games is an interesting and contested question. A complex systems approach to basketball modeling seems obvious because of the complexity of human nature. At any time, a basketball game depends on at least 10 individuals who are all experts at the game, and therefore there is no way to model a game without randomness and complexity being deeply ingrained in the model. The simple insight for the models presented in this project is that, when a player has the ball there are two options for their next action. They can either pass to their teammate or attempt to shoot the ball. In the parlance of a Markov chain, the transition probabilities of their next actions sum to 1. A basic chain for this is shown in Figure 1.

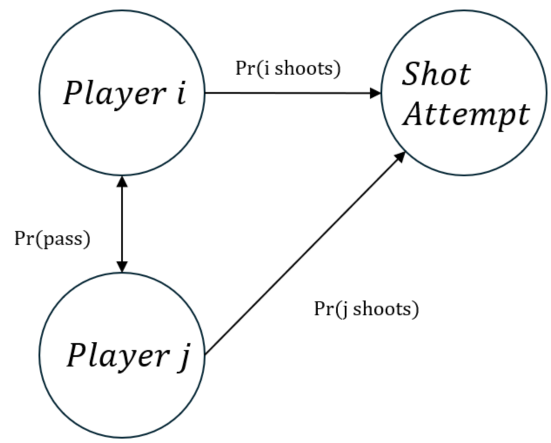

Figure 1. Basic Markov chain of basketball

There have been other approaches to modeling basketball using Markovian frameworks, e.g., (Štrumbelj 2012), but many of these approaches are prior to the advent of player tracking statistics, which the NBA implemented starting in 2014 (NBA Communications 2016).

2 Methods and Models for Simulation and Prediction

2.1 Team Game Markov Chain

The first approach extends the idea shown in Figure 1 to a game between two basketball teams. When Team A has possession of the ball, their next action is to attempt a shot or turn the ball over to Team B. The probability of a team turning the ball over can be calculated from box score data (as a percentage), and so, the probability of a shot attempt is just the complement of the turnover probability. After the shot is attempted, possession of the ball can return to either team via a rebound. Offensive rebound probability can also be calculated from box score data, and defensive rebound probability is implied to be the complement of offensive rebound probability. Note that the calculations for turnover percentage and offensive rebounding percentage require estimates of possessions, Kubatko et al. (2007) show how this can be estimated. This Markov chain is shown in Figure 2, and its transition matrix in Figure 3.

The stationary distribution of this Markov chain can be interpreted simply as the team with the higher probability of being in the scoring opportunity states has exactly that - more opportunities to score. However, having more opportunities to score does not equate to winning, those opportunities need to convert into actual scoring, and so, multiplying each of these states by respective teams' *true shooting percentage* (TS) will give a measure of points scored. TS is used as it is an encompassing scoring efficiency statistic that includes free throws and 3 point shooting. The team with the higher measure is the predicted winner of the game. The difference between these two final measures is called the *margin of victory* (MOV). A negative value of MOV means Team A lost and a positive value of MOV means Team A won.

$$MOV = ScOpp_A TS_A - ScOpp_B TS_B$$

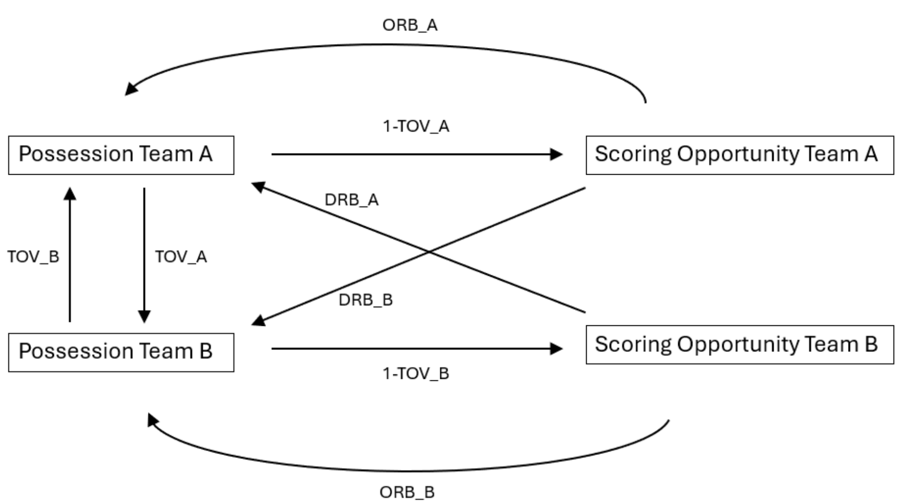

Figure 2. Markov chain of game between two teams

$$\begin{pmatrix} 0 & \mathrm{TOV}_B & \mathrm{ORB}_A & 1-\mathrm{ORB}_B \\ \mathrm{TOV}_A & 0 & 1-\mathrm{ORB}_A & \mathrm{ORB}_B \\ 1-\mathrm{TOV}_A & 0 & 0 & 0 \\ 0 & 1-\mathrm{TOV}_B & 0 & 0 \end{pmatrix}$$

Figure 3. Transition matrix of game between two teams

Each of the transition probabilities in this model are themselves random variables. Meaning that in any given game between two teams, their performance varies game to game. Colloquially, some nights teams shoot the ball well, and some nights they struggle to score. The same is true for offensive rebounding and turnovers, they vary from game to game. To account for this, these statistics were taken for every game and every team from the 2023-24 NBA season and a beta distribution was fit to each stat for each team. See Figure 4 for two examples of such histograms and fit distributions. Then, when constructing the transition matrix of a game simulation between two teams, each of the transition probabilities and TS are sampled from the fit probability distributions of the respective teams.

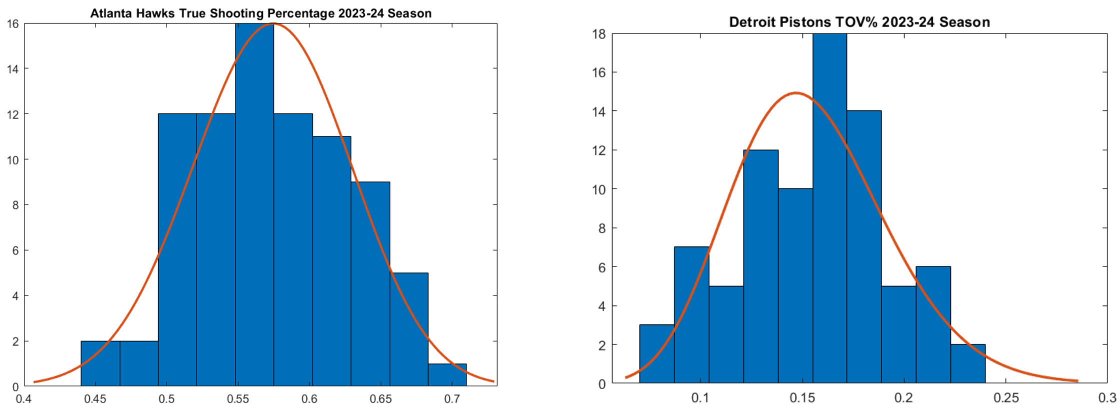

Figure 4. Atlanta Hawks TS and Detroit Pistons TOV for each game of 2023-24 season with fit beta distribution

2.2 Player Passing Markov Chain

The second approach takes a deeper look into the dynamics of a team's possession. Public data from NBA.com gives the probability of a player passing to any other player on his team. Using this information, a transition matrix can be created for each team. A game state can be defined by two random variables, Team possession, which belongs to either team A or team B, and ball possession, which can belong to any player i on team j where team j is the team with possession. At each change in ball possession, the model takes the following steps for player i. Use player i's average time per possession to pull a poisson random number with that mean to determine how much time runs off during this possession. If the time run off will lead to team j's total time of possession reaching 24 seconds (the length of the shot clock), player i shoots. If not, player i has some probability of shooting. If player i shoots, use his data to determine if he shoots a 2 or 3 pointer, and then determine if he makes or misses the shot. Additionally, every shot has some probability of the player then taking a free throw.

If player i is not shooting, then see the transition matrix for where the possession goes to next. There is some prob_TOV that player i turns the ball over and team possession changes teams. If not, pull from the transition matrix to see where ball possession will go next, and repeat all the past steps. An example transition matrix can be seen in figure 5.

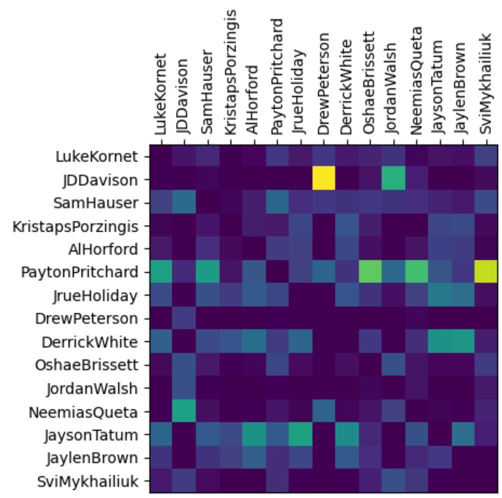

Figure 5: transition matrix for the Boston Celtics.
Columns sum to 1, brighter colors represent larger probabilities.

3 Results

3.1 Team Model

To verify the validity of the team model, the MOV calculation was made for every game of the 2023-24 season (n=1230) using the actual data of that game as the transition probabilities and TS. This is compared with the actual MOV (expressed in points) of the results of each game. A scatter plot is shown in Figure 6. The correlation coefficient of the model prediction is 0.9756. It is reasonable to believe that the model is well correlated with actual game data because the metrics used to build the model are generally referred to as the *Four Factors* (Oliver 2004), and are the essence of success in basketball games.

Then, the 2023-24 season is simulated for each NBA team (82 games each), now using the sampling method for transitions and shooting. The season is simulated 4 times and the average number of wins for each team are plotted with the actual number of wins the team had that season. The average win differential is 7.2 games. The results are shown in Figure 9a.

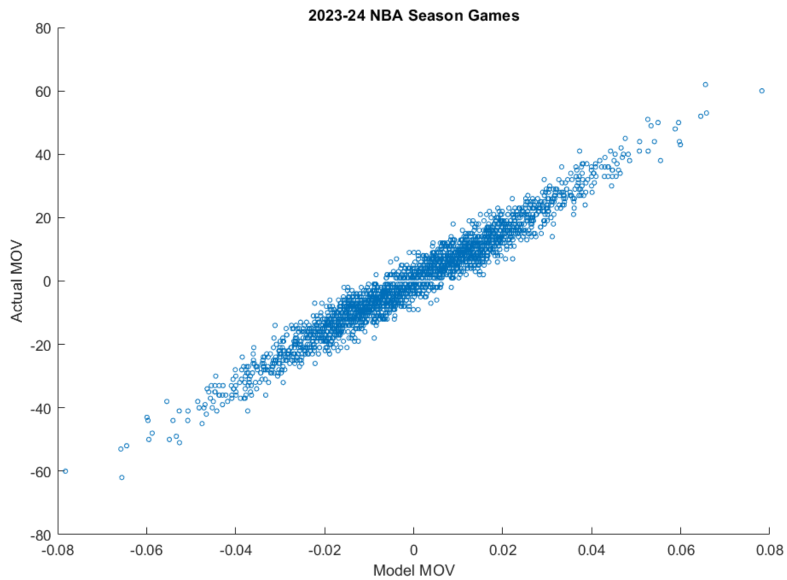

Figure 6. Team model MOV with actual game MOV outcomes

3.2 Player Model

To verify the transition matrix, the stationary distribution for each team was compared to the actual time of possession per player. An example comparison can be seen in figure 7a. For each team, a RMS similarity score was calculated by taking the square root of the sum of squared differences and dividing by the number of players on that team. Figure 7b shows the similarity scores for each team.

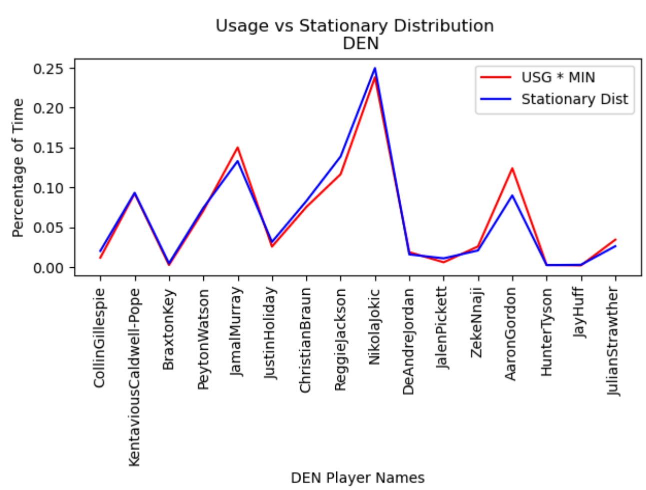

Fig 7a: Stationary distribution of transition matrix (blue) versus actual percentage of ball possessions (red) for the Denver Nuggets.

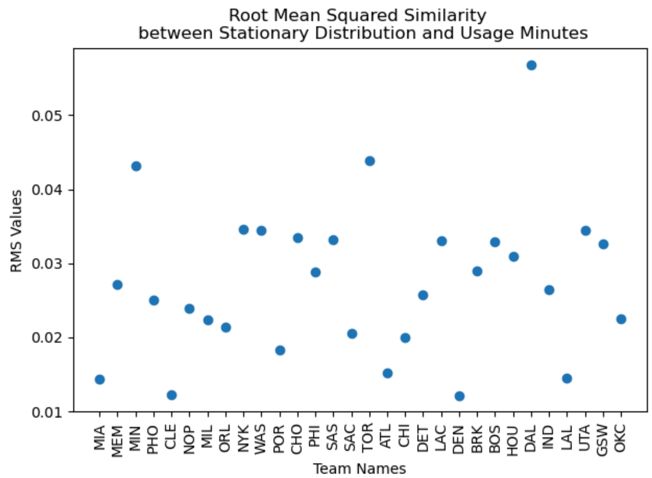

Fig 7b: Similarity scores for each team

Additionally, individual player stats are able to be retrieved from the model. Figure 8 shows points per game and shot attempts per game in the model compared to the real season data.

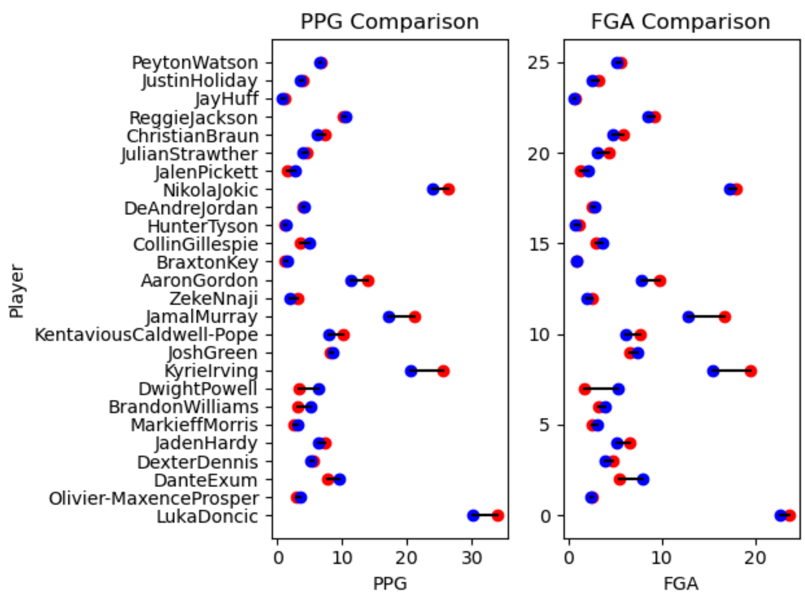

Figure 8: PPG and FGA per game from the model (blue) versus real data (red)
Averaged 40 games between Denver and Dallas

As above, the model was compared against the 2023-2024 NBA season. Again, it was averaged over 4 seasons. The average win differential between the model and reality was 6.6 wins. Figure 9b shows the results for each team.

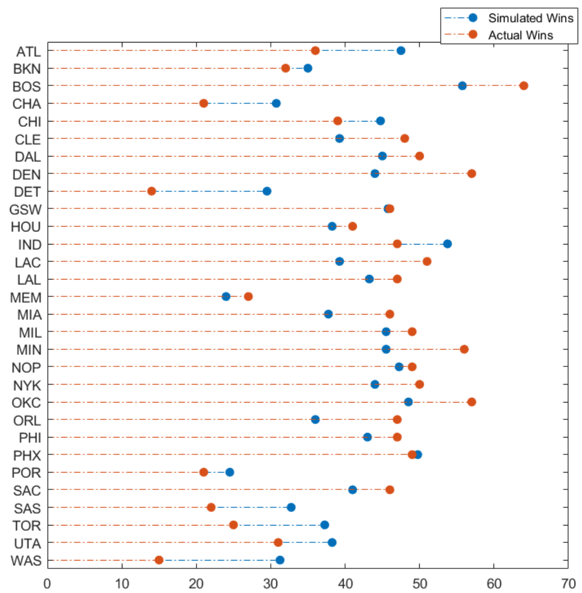

Figure 9a. Simulated wins of team model and actual wins of each team from 2023-24 NBA season for team model

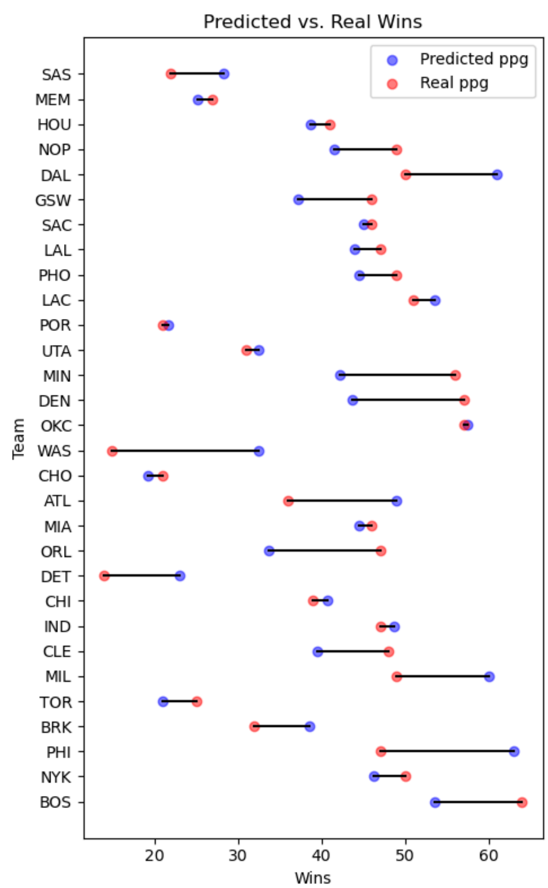

Figure 9b. Simulated wins of team model and actual wins of each team from 2023-24 NBA season for player model

4 Discussion

The predicted (simulated) vs. actual wins sheds light on the limitations of the models in this project. For example, the team model (Figure 9a) overestimates the number of wins of the Atlanta Hawks. This is partially explained by the lack of defensive metrics in the model. In 23-24 the Atlanta Hawks were ranked 27th of 30 in defensive rating (points allowed per 100 possessions). Also, consider the Detroit Pistons, the team model over overestimates their number of wins. The Detroit Pistons had a historic losing season in 23-24, setting the franchise record for fewest wins (14), and NBA record for most consecutive losses in a single season (28). This can be explained by player mentality and effort. In the midst of the Pistons historic losing streak, opposing teams did not want to be the team to lose to the Pistons to end their streak. As a player, why would you be okay with losing to the worst team? So, effort increased when playing against the Pistons in the midst of their losing streak, resulting in more losses for the Pistons than their skill or statistics should actually reflect.

In the player model (Figure 9b), the Philadelphia 76ers are predicted as one of the best teams, but their actual win total was less. Joel Embiid, their best player, spent a majority of the season injured. The player model assumes all players will play in every game of the season, whereas the absence of a star player in the team model is borne out in the aggregate team statistics.

References

Kubatko, Justin, et al. “A Starting Point for Analyzing Basketball Statistics.” *Journal of Quantitative Analysis in Sports*, vol. 3, no. 3, Jan. 2007. *DOI.org (Crossref)*, https://doi.org/10.2202/1559-0410.1070.

“NBA Announces Multiyear Partnership with Sportradar and Second Spectrum.” *NBA.Com: NBA Communications*, 22 Sept. 2016, https://pr.nba.com/nba-announces-multiyear-partnership-sportradar-second-spectrum/.

Oliver, Dean. *Basketball on Paper: Rules and Tools for Performance Analysis*. 1st ed., Brassey’s, 2004.

Štrumbelj, Erik, and Petar Vračar. “Simulating a Basketball Match with a Homogeneous Markov Model and Forecasting the Outcome.” *International Journal of Forecasting*, vol. 28, no. 2, Apr. 2012, pp. 532–42. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ijforecast.2011.01.004.

National Basketball Association. (2023). 2023-2024 NBA Rulebook. Retrieved from https://official.nba.com/rulebook/

Rodgers, Emily. (2023, August 2). Sports Betting Statistics, Trends, & Predictions. Retrieved from https://www.driveresearch.com/market-research-company-blog/sports-betting-statistics/