Eli Zelle

Dr. Penn

19 December 2018

Using the Banzhaf Index to Play Daily Fantasy Basketball

**Introduction**

In some preference aggregation systems, not all people have the same amount of voting power. Consider a board meeting for a company where each person’s voting power is proportional to the number of shares they own. People who have more shares will have stronger votes than people with fewer shares and, are more influential in the final outcome. To better understand the voting strengths of people in a situation like the company board, John Banzhaf created the Banzhaf Power Index. According to Banzhaf, “voting power is not always proportional to the number of votes a legislator may cast” (318). The Banzhaf index looks to better quantify the voting power of each voter by examining their effects on decisive coalitions. A decisive coalition is defined as a set of voters that control the outcome of a vote. For instance, a shareholder that owns 60% of a company’s stock would constitute a decisive coalition in the board meeting scenario. A voter’s Banzhaf index is the proportion of coalitions for which they are decisive to the total number of decisive coalitions. Then it follows that individuals with more voting weight will have more voting power than those will less voting weight but, the relationship between voting weight and voting power need not be linear. This paper explores the use of the Banzhaf index to determine the strength of fantasy basketball players.

Daily fantasy basketball is an online game where people compete by drafting players to fill a lineup. Lineups score points based on real world performance. Game statistics like points, rebounds, assists, steals and blocks are worth a certain amount of points in fantasy basketball. Fantasy players also have an associated salary that managers must pay to draft them for their team. Each manager has the same total spending power to use on players and compete by drafting the team they think will produce the most fantasy points. The inspiration for this project came from looking at a fantasy basketball lineup like a coalition of voters. Each player brings a certain amount of power to a lineup based on the amount of fantasy points they score. Because a fantasy lineup can be modeled as a coalition of voters, a Banzhaf index can be computed for each player, which reflects their contribution to the lineup.

**Part 1: Decisive Coalitions of Players**

**Methodology**

For this project, the fantasy baseball game was modeled after the daily fantasy basketball competitions hosted by Draftkings.com. A standard competition on Draftkings requires a lineup of 8 players with multiple utility positions. However, to compute the indices of fantasy players using that format would have been much more computationally expensive. So, for this project, all lineups had the following positions: Point Guard, Shooting Guard, Small Forward, Power Forward, Center. With five possible positions, the problem space shrinks to a much more computationally manageable size. Draftkings defines a winning lineup as a lineup that scores in the top 15% of all lineups. This measure is subjective as it doesn’t guarantee that the best possible lineup is produced by any entries. Despite that, the same definition of a winning lineup was used in this project. Draftkings pays out more money to the highest scoring lineup than those that merely win. However, this analysis treats winning as a binary variable and does not asses a player’s power with respect to payout.

Data collection occurred from November 7, 2018 to November 18, 2018. Each day, a list of NBA players who played that day and, their associated DraftKings salary and position eligibility was pulled from Draftking.com. Then the following day, each player’s statistics from the night before was pulled from stats.nba.com. These two datasets were then cleaned and joined on the “Player Name” attribute to produce a dataset consisting of one player per row, with their associated statics, position, and Draftkings salary for a specific date. Each dataset had between 100 and 400 players. Using the single day statistics, each player’s fantasy score was computed using Drafkings’ scoring rules. The score attribute was appended as a column to the dataset and saved for further analysis.

Using the dataset with from the previous step, 50000 eligible fantasy lineups were constructed for each day. To do this, players were grouped by position and mapped to a new data frame. These data frames contained columns for player, salary, position and points. Players that are eligible for more than one position would appear multiple data frames. The Python package Pandas has a useful sampling function that returns one row from a data frame. To construct one lineup, this sampling function would be used on each of the five positional sets of players. The player’s respective salaries were aggregated and if the it exceeded the threshold of 31500$, the lineup was discarded. Additionally, if a player managed to be picked for more than one position, the lineup was discarded. If the salary and personnel constraints were met, the scores of each player was added to get a final score for the lineup. The five players, the total price and the total score was saved in a new data frame. This process was repeated until 50000 valid lineups were generated and saved in the data frame. Then using the Pandas Drop\_Duplicates function, any repeated lineups in the final data frame were removed to ensure that each lineup was unique.

By sampling 50000 lineups from the set of possible eligible lineups, of the distribution of fantasy points should approximate the true distribution for all possible lineups. The subset of lineups generated for this project should have the same characteristics as the entire set. Therefore, studying the 50000 lineups in this project will give valid insight into the behavior of fantasy lineups in general. The number 50000 was chosen by trying to capture about one percent of all possible lineups. For example, there were 25 players eligible for each position, then there would be 25^5 possible lineups to analyze. Doing computations with that lineups would be time consuming and unfeasible. The number 50000 was chosen to minimize this computation cost while still having enough samples to produce valid results. Because the 50000 lineups are randomly sampled from the set of all possible lineups, they should approximate the true distribution of possible lineups. This is, the average amount of points should be close to the true average for the set of all lineups. Then, the 15%-point threshold used to define winning lineups should be very close to the numerical value of the 15% threshold for the set of all possible lineups.

Once the 50000 lineups were generated for each day, the top fifteen percent of lineups were identified by their point total. A variable called min\_score was used to mark the minimum value that a lineup needed to exceed to be winning. A binary variable was appended to this data frame that represented the winning status of a lineup. If a lineup scored above threshold, it would receive a 1 in the Win column. Otherwise, it received a 0. Thinking of these lineups as coalitions of voters, a decisive coalition of players would be a lineup with a 1 in the Win column. From here, the Banzhaf indices were computed for each player for all the days in the data set. A dictionary data structure was used to keep track of tuples containing (Player name, Count), where the Count variable counts the number of times a player is decisive for a winning coalition. These results were aggregated into a separate excel documents, by date, that contain the attributes for player, score, salary and index. These excel documents were used to conduct data analysis portion for this project.

**Results**

The dataset produced by computing the Banzhaf indices for each fantasy player shows a very strong linear trend between fantasy points and Banzhaf index. This is somewhat expected: in a board meeting, the more shares someone has, the more influential their voting power should be. However, the correlation between fantasy points and index was R=0.7685, as visualized in the figure below.

Figure 1:

For reference, the plot of Salary vs Fantasy points has a strong, positive correlation but weaker than the plot above. As expected, a player’s salary should explain some variance in the points they score. Players that cost more to draft usually score more fantasy points.

Figure 2:

A notable observation; in the first plot, the player that scored 80 points had a lower index than the player that scored 60. From the model that Banzhaf proposed, even through voting power isn’t necessarily proportion to voting weight, someone with less voting weight should not have more voting power than someone with more voting weight. Clearly, the data contradicts this assumption. The second half of this paper explores the reason behind this deviation.

**Part 2: Positional Analysis**

One key difference between modeling fantasy lineup as voting coalitions and real voting coalitions is the positional constraint on fantasy players. Each player has a maximum of two eligible positions, and therefore can only appear in two places in a lineup. On the other hand, voters could fill in anywhere in a coalition. This introduces a notion of positional scarcity that

this project further explores. In Figure 1, the players with 80 points is Lebron James, who is eligible for the PF and SF positions. The player with 60 points is Blake Griffin, who is eligible for the C and PF positions. So even though Lebron scored more points than Griffin, there must be a way to explain the discrepancy between their indices. Then how does a player’s position impact their Banzhaf index?

To get at the quantitative influence of a player’s positional eligibility, the problem of computing a player’s Banzhaf index was broken down by position. For a player that plays multiple positions, their index was computed separately for each position. For instance, Chris Paul has fantasy eligibility as a point guard and a shooting guard. Then his single row in the dataset was decomposed into Chris Paul with point guard eligibility and Chris Paul with shooting guard eligibility. By treating each player as a player-position tuple, it becomes realistic to analyze the influence each position has in a fantasy lineup.

**Methodology**

Using the 50000 lineups generated in the first part of the project, the Banzhaf indices were computed for each player with respect to position. For this process, 5 dictionary data structures were created to track the (Player name, Counts) tuple for each position. Then Chris Paul had a count of his decisive appearances as a point guard and as a shooting guard. By looking over each winning lineup in the 50000 sample lineups, each player’s Banzhaf index was computed for their eligible positions. Finally, a dataset was saved as a excel document for further analytics. This dataset had the following columns: Player Name, Position, Salary, Points, Index.

**Results**

As hypothesized, a player’s position had a noticeable effect on their Banzhaf index and thus, their voting power in a fantasy lineup. When comparing the indices for players eligible for more than one position, it became clear that some positions are inherently more valuable in lineups than other. As Figure 3 points out, players eligible for center and power forward have a higher index as a center.

Figure 3:

In every case highlighted in the chart above, theses players are almost twice as valuable to a lineup when selected as a center rather than a power forward. This figure only displays the results from one day of data but, the trend continues for all other days tested. This understanding between the value of centers and power forwards explains the discrepancy between Lebron James and Blake Griffin in Figure 1. The same analysis was conducted for all other positional pairs. The general trend continues to show that each position has an inherent value in a fantasy lineup. Chats analyzing this finding are included in the appendix.

**Conclusion**

The Banzhaf index gives valuable insight into constructing winning daily fantasy lineups. By highlighting the positional value of the five positions, a winning strategy can be built by allocating more funds to the more valuable positions. The difference between the indices of players eligible for more than one position probably comes from positional scarcity. That is, more players have eligibility at power forward than center. Where there might be 10 power forwards who score more than 40 fantasy points on a given day, there might only be 2 centers that score 40 points. That makes selecting one of those two centers more influential in a lineup than selecting one of the ten power forwards. For other positions, the difference between indices was not as large, but still prevalent.

Going forward, the results presented in this project could be used to created winning lineups for daily fantasy basketball. One approach to this problem would be to use machine learning algorithms to learn from the data collected, as well as the Banzhaf indices of each player, to produce lineups. This would consider the positional scarcity aspects discussed in Part 2 and, could prove valuable to winning the game. Additionally, it would be interesting to look at players that have high Banzhaf indices for loosing lineups. If a player appears in very few winning lineups, they would have a high index with respect to loosing lineups. Then using a similar positional analysis, one could deduce the types of players to avoid when creating fantasy lineups.

**Works Cited:**

Banzhaf, John F. III. "Weighted Voting Doesn't Work: A Mathematical Analysis," *Rutgers Law Review* vol. 19, no. 2 (Winter 1965): p. 317-344. HeinOnline, <https://heinonline.org/HOL/P?h=hein.journals/rutlr19&i=324>.

**Appendix:**

All datasets and code for this project are available at

Drafkings Fantasy Scoring Rules:

![A screenshot of a cell phone

Description automatically generated]()

![A screenshot of a cell phone

Description automatically generated]()

Sample Lineup Data:

![A screenshot of a cell phone screen with text

Description automatically generated]()

Sample Player Index Data:

![A screenshot of a cell phone

Description automatically generated]()

Positional Analysis Results:

Fantasy Points Vs Banzhaf Index as a General Trend: