Analysis of Conventional Betting Strategies in Sports Betting

Douglas Silverman, Calvin Lok, Jason Le *University of Amherst*

4/29/2020

Over the course of the past 10 years fantasy sports have seen an explosion in popularity and interest. People participate in various forms of fantasy sports gambling ranging from a single $10 league with friends, to multiple money leagues, all the way to multiple wagers on draftkings and other one-day-Sunday fantasy teams. The recent popularity of websites like draftkings and fanduel are a way for Americans to feel the thrill of gambling. Sports gambling in the United States was banned in 1992 by the PASPA (Professional and Amateur Sports Protection Act of 1992). This made it difficult for states, particularly New Jersey, to have some sort of sports betting. However, in May of 2018, the Supreme Court overturned PASPA allowing states to pass their own legislation for sports betting. Since the decision, 17 states have fully legalized sports betting and 48 states have some sort of legislation. As more states trend to legalize sports betting, the analysis of different models or betting strategies is interesting to look at.

Construction of Simulation

In order to analyze different strategies, models will be built based on the strategies. The models will then be tested using a simulation. A simulation will consist of a starting amount of $1000 and the requirement to place a bet on every game that was played during a given season. Each bet will be a money-line wager on a head to head matchup between 2 teams. The model will decide which team to pick as their favorite and place a bet on them. A money-line bet is a wager that one team will beat another team in a discrete match. If the team the wager was placed on, won the match, then the model will profit according to the historic las vegas odds. The profit is stored at many points during the season to track the progress of the models. Any negative profit values indicate that the model is losing money, and a positive value indicates the model is making money. The purpose of simulating this way is to understand different ways people place bets on sports.

Data sets for the Analysis

Our Datasets consist of data for the past 3 seasons of MLB, NFL, and NBA professional play. The benefit of using multiple sports for analysis is that there is variation. Each sport varies in the length of the game, the length of the season, and the likelihood for an upset (i.e. a team with a better record losing to another team). Each game includes: the date of the game, the 2 teams playing, the final score, and the American odds[1] for each team winning. Bookmakers in Las Vegas put out payouts information in the form of American odds prior to every sports contest.

Traditional Betting Models

For our project, we started by looking at conventional betting models. These models are strategies many casual gamblers use to minimize losses. The concept behind this is known as the gambler's fallacy. An example of this is tossing a fair coin. Each toss is statistically independent and the result of the toss does not depend on the previous tosses. Let's say we toss a coin 10 times. If the first 9 tosses are heads, many will believe the next toss must land tails. This is known as the gambler's fallacy. This belief is wrong however. Let the event H_i represent the ith toss lands hands. The probability of 10 heads in a row is what most people look at:

$$P(H_1, H_2, ...H_9, H_{10}) = \prod_{i=1}^{9} P(H_i) \cdot P(H_{10}) = \frac{1}{2^9} \cdot \frac{1}{2} = \frac{1}{2^{10}}$$

This probability is equal to $1_{\overline{2^{10}}}$ which is a really small value. However, this is the same as the probability of 9 heads in a row and then 1 tails:

$$P(H_1, H_2, ...H_9, H_{10}) = \prod_{i=1}^{9} P(H_i) \cdot P(T_{10}) = \frac{1}{2^9} \cdot \frac{1}{2} = \frac{1}{2^{10}}$$

Given individual wagers are independent, the gambler's fallacy is incorrect. The most infamous case of the gambler's fallacy occurred in August of 1913 during a roulette game at the Monte Carlo Casino. The roulette ball landed on black 26 times in a row. The probability of this happening was : $(\frac{18}{37})^{26-1}$ or around 1 in 66 million

[1] American odds refer to the payout for winning the bet. A positive number like 250 means that on a bet of \$100 you will win \$250 if you win. A negative number refers to the amount needed to place on a bet in order to win \$100. For example, if the odds are -115, then a bet of \$115 would yield \$100 if won.

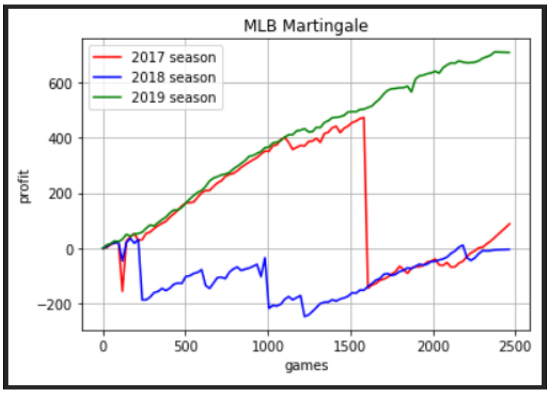
Martingale Betting Algorithm

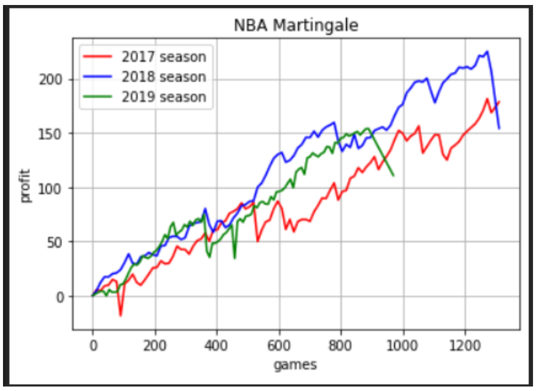
The first model is based on a martingale of the initial capital. A martingale is a sequence of random variables such that the next value in the sequence has an expected value equal to the present value. The formal definition of discrete-time martingale (i.e. a discrete-time stochastic process, meaning a sequence of random variable $X_1, X_2, X_3, ...$) that satisfies:

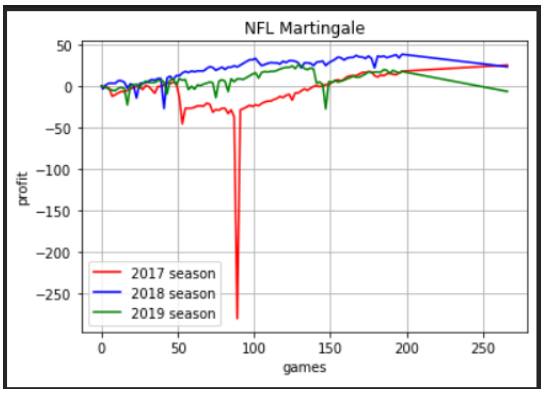
$$\forall n : E[|X_n|] < \infty \land E[X_{n+1} \mid X_1, X_2, ..., X_n] = X_n$$

Our model uses a martingale based on the initial loss value, which starts at 0. We start with an initial bet size so that losing 10 bets in a row would bankrupt us (this value can be adjusted). After a win, the bet size stays the same and the loss remains the same. After a loss, the next bet size will be equal to the amount needed to wager so that the entire loss will be made up. If loss is 0 then the bet size will be the initial bet size. As consecutive losses occur, the bet size grows exponentially. This shows how gamblers can amass debt very quickly.

Martingale Performance







Expected Value for a single round of gambling:

A single round can be defined as a stochastic process (a sequence of random variables) such that all discrete events are losses. After a set of losses, the first win happens the round is over because the gambler has "reset". The next round of gambling occurs and the process continues. A continuous sequence of martingale bets can be split into independent rounds. let

q = the probability of losing a given bet (For American roulette, $q = \frac{20}{38}$ for a bet on red or black)

b = the initial bet size

n = number of bets the gambler can afford to lose before being bankrupt. n must be a positive integer ($n \in +Z$).

The probability that all n bets will be lost is q^n

$$P(n\ losses\ in\ a\ row) = \prod_{i=1}^{n} P(loss) = \prod_{i=1}^{n} q = q^n$$

The total amount lost in this case is:

$$\sum_{i=1}^{n} b \cdot 2^{i-1} = b \cdot (2^n - 1)$$

The probability of a not losing all bets is:

$$P(not\ losing\ all\ bets) = 1 - q^n$$

Therefore the expected profit is:

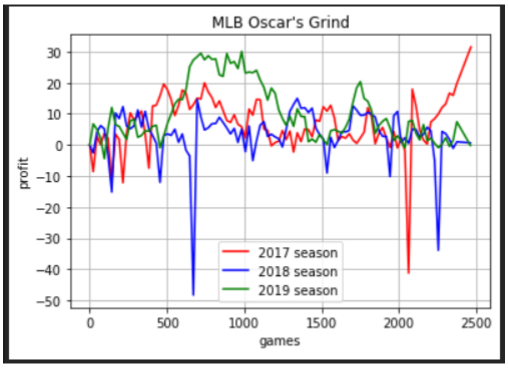
$$b \cdot (1 - q^n) - q^n \cdot b(2^n - 1) = b(1 - (2q)^n)$$

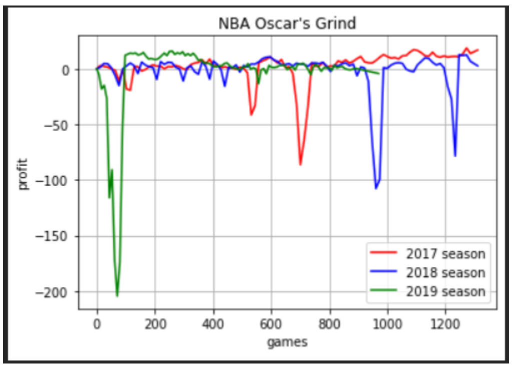
Notice that whenever $q > \frac{1}{2}$ the expected value of the round will be negative, for all $n > 0$. This is the reason why the house always wins for any casino game. This means that for any wager where the gambler is more likely to lose than win, the gambler is expected to lose money using a martingale strategy. Increasing the wager per round only increases the average loss rather than making up the difference.

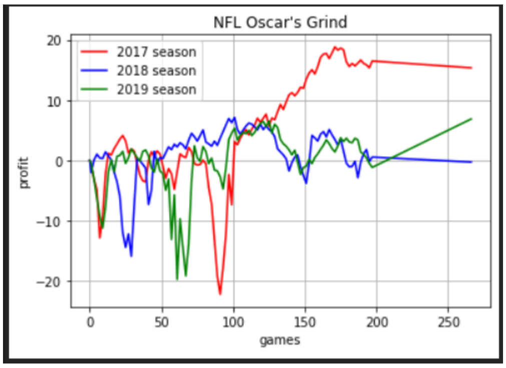
Oscar's Grind

Oscar's Grind is a betting strategy first documented in 1965 in The Casino Gambler's Guide, and is designed to minimize risks for steady profits. It applies to all even money bets, and the only goal is to win one unit of profit. Each bet is considered a series that is continued until one unit is won, then reset to begin a new cycle. Initially, one unit is bet. If this is a win, the series ends and a new one begins. If it is a loss, the bet stays at the same size and you continue betting. When the bet is a win and the profit threshold has not been met, the successive bet is increased by one unit. These two rules are repeated until the series ends in one unit of profit, then reset to a bet of one unit for the next series. Given infinite time and infinite wagers, Oscar's Grind always return a profit.

Oscar's Grind Performance







Kelly Criterion

The Kelly Criterion is a simple formula to assist gamblers in deciding how much money each bet should receive. Each bet is a fraction of the current amount of assets, scaling after each bet. The equation is :

$$K\% = \frac{ap - q}{a}$$

Where

K = the fraction of the bankroll to bet

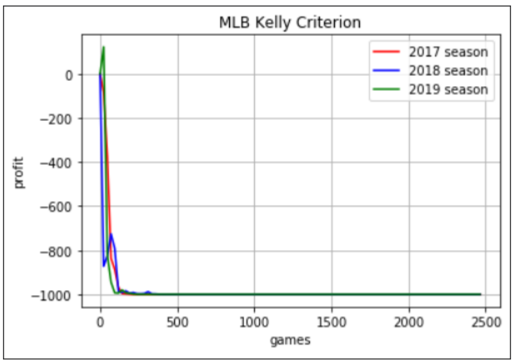
a = the decimal odds of a bet – 1

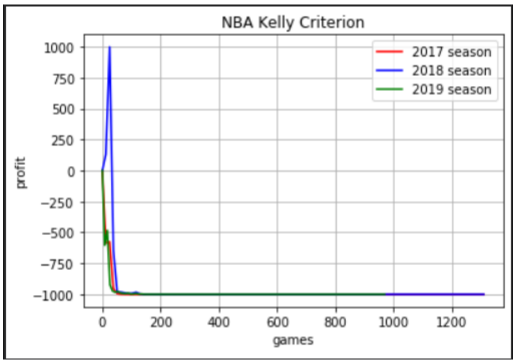
p = the probability of winning

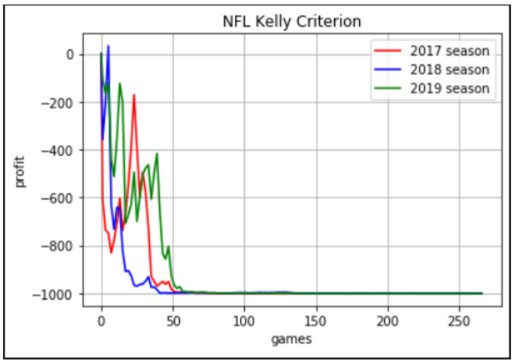
q = the probability of losing

The result is a percentage of your capital that is recommended to bet on the gamble. If this is negative then it is a sign to avoid the bet and maybe consider betting on the other option.

Kelly Criterion Performance







Poisson Based Prediction

The Poisson Distribution is a discrete probability distribution that can be used to express the probability of certain events from happening when known how often the event has occurred. In terms of sports, it can be used to determine the probability of the number of points a team scores in a game. With this method, we can calculate the probability that a favored team wins against a certain team. Below is the Poisson distribution formula:

$$P(X) = \frac{\lambda^x e^{-\lambda}}{X!}$$

For X is the desired number of points a team will score and λ is the expected number of points a team scores. Before we can use the Poisson distribution to calculate winning odds, we need to find the value of λ. We can find this value by calculating the attack and defence strengths of a home team and an away team. Attack and defence strengths are based on the number of points a team has scored and the number of points they have conceded in a given season respectively. We must also know the average number of points scored per game for both at home and away teams. To find the averages per season, we will use these variables for the following equations:

- AH = Average points scored at home
- PH = Total points scored by home

- HG = Total home games
- AA = Average points scored away
- TA = Total points scored by away
- AG = Total away games

$$\text{AH} = \frac{\text{PH}}{\text{HG}}$$

$$\text{AA} = \frac{\text{TA}}{\text{AG}}$$

We will also need to find the average number of points conceded by both the home and away team. However, these are just the opposite to the average points scored per game. The average number of points conceded by the home team would be equal to the average points scored away and the average number of points conceded by the away team would be equal to the average points scored at home.

Next we will calculate the home team's attack strength which is needed to calculate the expected value of points for the home team. We will use these two equations below:

- HAP = Home team's average points per home game
- HS = Home team's score at home
- NHG = Number of home games played
- HAS = Home team's attack strength

$$\text{HAP} = \frac{\text{HS}}{\text{NHG}}$$

$$\text{HAS} = \frac{\text{HAP}}{\text{AH}}$$

With this, we will find the home team's attack strength for the given season. The same can almost be done when calculating the away team's defensive strength:

- $AAPC$ = Away team's average score conceded per away game
- PCA = Points conceded away by away team
- NAG = Number of away games played
- ADS = Away team's defense strength
- $APCA$ = Average points conceded by away team

$$\text{AAPC} = \frac{\text{PCA}}{\text{NAG}}$$

$$\text{ADS} = \frac{\text{AAPC}}{\text{APCA}}$$

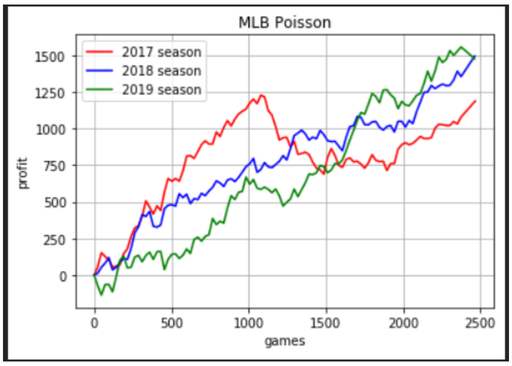
With this, we now have everything needed to calculate the expected points a team will score. To find the expected value of the home team, multiply the home team's attack strength with the away team's defense strength and the average number of home games. Now we plug this value into λ and we can find the probability that the home team will score X points against the away team. We can also find the away team's expected points by simply flipping all the home and away variables.

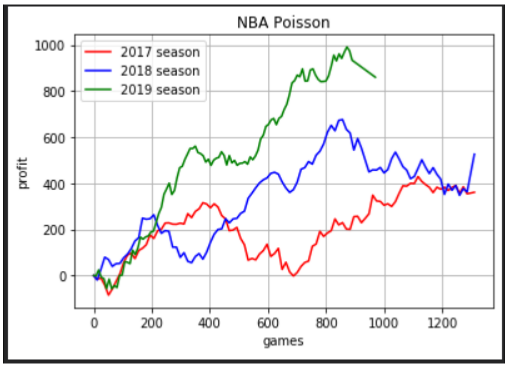
Based on the sport, X should be a reasonable value that an average team could achieve. For example, in soccer, there is no reason to find the probability that a team will score above 5 goals because it is highly unlikely to score that many points in the sport. Once X has been established, We should get a list of probabilities for the desired range of scores for both teams. If we want to find the probability of a specific final score of a game for example, an outcome of 2-0 home vs away, simply multiply the respective probabilities together to obtain thef probability of the outcome. To find the probability that the home team wins however, we will have to find the probability of every possible outcome where the home team wins and add them all together. This can be simulated with the formula below:

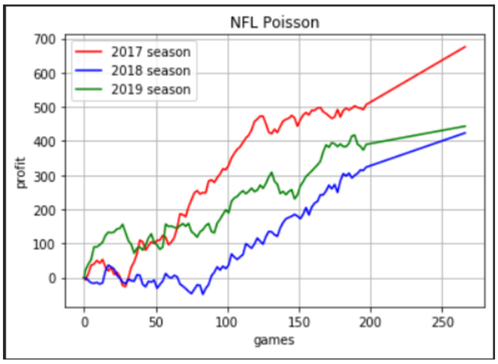
$\sum_{i=1}^{M} \sum_{j=0}^{i}$ P(home scores i * P(visitor scores j)

Where M is the upper bound of the score. As you can see, the visitor's score will always be smaller than the home team's score. With this formula, we can determine who to bet on based on how likely the favored team will score more points than the opposing team.

Poisson Performance







Conclusion

Traditional betting models created from the Martingale performed fairly accurately, because the expected value of a round of betting must be positive for sports bets. By having the models bet on the favorites, it creates a scenario where the probability of losing a bet is less than fifty percent, which is the threshold for the model. The Martingale was more effective than the Oscar's Grind model, because it recoups all losses with a single win, rather than slowly trying to rebuild lost funds.

The Kelly Criterion fared poorly, because it is centered on the assumption that sports bettings odds are both accurate and fair. Oddsmakers and bookkeepers model their own assumptions that betting will be split among the two sides of the bet, and build overhead accordingly.

The Poisson betting model performed exceptionally well, due to its success at analyzing past data and game history. The model focused on gathering past statistics and averages to leverage assumptions about future games, while also taking into account the stochastic nature of the future by breaking down games into smaller more manageable segments.

The unpredictability and turmoil of sports makes short term models difficult to profit from, but long term trends and careful analysis help even the odds. Models that leverage a certain strategy tend to profit, while those that make incorrect assumptions quickly perish.

References

[1] Cronin, Benjamin. “Poisson Distribution: Predict the Score in Soccer Betting.” Pinnacle, 27 Apr. 2017, www.pinnacle.com/en/betting-articles/Soccer/how-to-calculate-poisson-distribution/MD62MLXUMKMXZ6A8.

[2] Kelly, J. L. (1956). ”A New Interpretation of Information Rate” (PDF). Bell System Technical Journal. 35 (4): 917–926. doi:10.1002/j.1538-7305.1956.tb03809.x.

[3] Lester E. Dubins; Leonard J. Savage (1965), How to gamble if you must: inequalities for stochastic processes, McGraw Hill

[4] Mason Malmuth; Lynne Loomis (1993), Fundamentals of Craps, Two Plus Two Publishing, p. 46, ISBN 1-880685-30-2

[5] ”Why we gamble like monkeys”. BBC.com. 2015-01-02.

Dataset reference

https://www.sportsbookreviewsonline.com/