Predicting Outcomes of games

Lee Przybylski

Iowa State University

September 28, 2020

Problem Overview

Sports books present often present 3 main ways to bet on the outcomes of games. Take this example from sportsbook.draftkings.com:

TOMORROW	POINT SPREAD		TOTAL POINTS		MONEYLINE
7:20PM **MIA Dolphins**	+3	-110	O 48	-110	+138
JAX Jaguars	-3	-110	U 48	-110	-157

- Point Spread (Miami receives a 3 point handicap.)

- Point Spread (Miami receives a 3 point handicap.)
- Total Points (Will the teams score more or less than 48 points combined?)

- Point Spread (Miami receives a 3 point handicap.)
- Total Points (Will the teams score more or less than 48 points combined?)
- Moneyline (Risking $100 on Miami will give a net gain of $138 if successful.)

- Point Spread (Miami receives a 3 point handicap.)
- Total Points (Will the teams score more or less than 48 points combined?)
- Moneyline (Risking $100 on Miami will give a net gain of $138 if successful.)

We focus here on the moneyline and the point spread.

Understanding the Moneyline

The money line corresponds to odds which corresponds to winning probabilities. We can check these winning probabilities, estimate our own, and wager accordingly.

According to the moneyline $L = \text{MIA } +138$ suggests MIA has a win probability of

$$\mathbb{P}(\text{MIA}) = \frac{100}{100 + L} = \frac{100}{100 + 138} \approx 0.4202$$

Understanding the Moneyline

The money line $L =$ JAX -157 implies that Jacksonville has a win probablity of

$$\mathbb{P}(\text{JAX}) = \frac{L}{100 + L} = \frac{157}{100 + 157} \approx 0.6109.$$

Understanding the Moneyline

The money line $L =$ JAX -157 implies that Jacksonville has a win probablity of

$$\mathbb{P}(\text{JAX}) = \frac{L}{100 + L} = \frac{157}{100 + 157} \approx 0.6109.$$

Wait a minute!

$$\mathbb{P}(\text{JAX}) + \mathbb{P}(\text{MIA}) \approx 1.0311 > 1$$

Understanding the Moneyline

We can normalize these probabilities to get what the book considers as fair probabilities without vigorish:

$$\mathbb{P}(\mathsf{MIA}) = \frac{\mathbb{P}_{\mathsf{vig}}(\mathsf{MIA})}{\mathbb{P}_{\mathsf{vig}}(\mathsf{MIA}) + \mathbb{P}_{\mathsf{vig}}(\mathsf{JAX})} \approx \frac{0.4202}{1.0311} = 0.4075.$$

$$\mathbb{P}(\mathsf{JAX}) = \frac{\mathbb{P}_{\mathsf{vig}}(\mathsf{JAX})}{\mathbb{P}_{\mathsf{vig}}(\mathsf{MIA}) + \mathbb{P}_{\mathsf{vig}}(\mathsf{JAX})} \approx \frac{0.6109}{1.0311} = 0.5925.$$

Understanding the Moneyline

According to these fair probabilities, our expected winnings for each bet would be

Understanding the Moneyline

According to these fair probabilities, our expected winnings for each bet would be

$$\mathbb{E}[W_{\mathsf{MIA}}] = 0.4075 \cdot \$138 + 0.5925 \cdot (-\$100) = -\$3.02$$

Understanding the Moneyline

According to these fair probabilities, our expected winnings for each bet would be

$$\mathbb{E}[W_{\text{MIA}}] = 0.4075 \cdot \$138 + 0.5925 \cdot (-\$100) = -\$3.02$$

Similarly, betting on \$100 on JAX gives us expected winnings of

$$\mathbb{E}[W_{\text{JAX}}] = 0.5925 \cdot \$63.69 + 0.4075 \cdot (-\$100) = -\$3.01$$

Understanding the Moneyline

According to these fair probabilities, our expected winnings for each bet would be

$$\mathbb{E}[W_{\text{MIA}}] = 0.4075 \cdot \$138 + 0.5925 \cdot (-\$100) = -\$3.02$$

Similarly, betting on \$100 on JAX gives us expected winnings of

$$\mathbb{E}[W_{\text{JAX}}] = 0.5925 \cdot \$63.69 + 0.4075 \cdot (-\$100) = -\$3.01$$

According to the book's probabilities, either way is a losing bet. In fact, the loss for each bet might be the same if not for a rounding error. The probabilities estimated by the book might not be accurate, since they are set for the purposes of avoiding risk. If more accurate win probabilities can be estimated, then we can exploit market inefficiencies.

Vigourish

What does the book do with the extra 3.11%?

Vigourish

What does the book do with the extra 3.11%?

The extra probability is called overround and lets us compute the vigourish.

$$V = \frac{O}{1 + O}.$$

In our case,

$$V = \frac{0.0311}{1.0311} = 0.0302.$$

This means the sports book expects to make \$3.02 for every \$100 wagered on this particular money line.

We will now consider a couple of simple models to determine our own win probabilities.

Monte Carlo Simulation

We can simulate the outcome of the game using random numbers. Using Excel we do the following:

1. Compute the average points scored and average points allowed for each team.

Monte Carlo Simulation

We can simulate the outcome of the game using random numbers. Using Excel we do the following:

1. Compute the average points scored and average points allowed for each team.
2. Compute the standard deviation for points scored by each team.

Monte Carlo Simulation

We can simulate the outcome of the game using random numbers. Using Excel we do the following:

1. Compute the average points scored and average points allowed for each team.
2. Compute the standard deviation for points scored by each team.
3. Sample a random score for each team using a normal distribution with the appropriate mean and variance and determine a winner.

Monte Carlo Simulation

We can simulate the outcome of the game using random numbers. Using Excel we do the following:

1. Compute the average points scored and average points allowed for each team.
2. Compute the standard deviation for points scored by each team.
3. Sample a random score for each team using a normal distribution with the appropriate mean and variance and determine a winner.
4. Repeat 10,000 times.

Monte Carlo Results

In Excel, we found $\mathbb{P}(\text{MIA}) = 0.2701$ and $\mathbb{P}(\text{JAX}) = 0.7299$. If we bet \$100 on MIA, we get \$138 for winning, so according to our model, our expected winnings are

Monte Carlo Results

In Excel, we found $\mathbb{P}(\text{MIA}) = 0.2701$ and $\mathbb{P}(\text{JAX}) = 0.7299$. If we bet \$100 on MIA, we get \$138 for winning, so according to our model, our expected winnings are

$$\mathbb{E}[W_{\text{MIA}}] = 0.2701 \cdot \$138 + 0.7299 \cdot (-\$100) = -\$35.72$$

Monte Carlo Results

In Excel, we found $\mathbb{P}(\text{MIA}) = 0.2701$ and $\mathbb{P}(\text{JAX}) = 0.7299$. If we bet \$100 on MIA, we get \$138 for winning, so according to our model, our expected winnings are

$$\mathbb{E}[W_{\text{MIA}}] = 0.2701 \cdot \$138 + 0.7299 \cdot (-\$100) = -\$35.72$$

Similarly, betting on \$100 on JAX gives us expected winnings of

$$\mathbb{E}[W_{\text{JAX}}] = 0.7299 \cdot \$63.69 + 0.2701 \cdot (-\$100) = \$19.48$$

Closing Thoughts on Monte Carlo

- Betting on the underdog is not worth it here.

Closing Thoughts on Monte Carlo

- Betting on the underdog is not worth it here.
- The normal distribution might not always be best.

Closing Thoughts on Monte Carlo

- Betting on the underdog is not worth it here.
- The normal distribution might not always be best.
- Might want to consider home field advantage.

Closing Thoughts on Monte Carlo

- Betting on the underdog is not worth it here.
- The normal distribution might not always be best.
- Might want to consider home field advantage.
- Might want to consider data from games that do not involve the two teams.

Closing Thoughts on Monte Carlo

- Betting on the underdog is not worth it here.
- The normal distribution might not always be best.
- Might want to consider home field advantage.
- Might want to consider data from games that do not involve the two teams.
- Simulations would be easier in R or Python.

Linear Regression Model

Instead of simulations, we could use the data from all games played in the league.

$$Y_i = \beta_{i_h} - \beta_{i_a} + \beta_1 + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2).$$

Linear Regression Model

Instead of simulations, we could use the data from all games played in the league.

$$Y_i = \beta_{i_h} - \beta_{i_a} + \beta_1 + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2).$$

- Y_i represents the home team's margin of victory.
- β_1 represents the effect of home field advantage.

Linear Regression Model

Instead of simulations, we could use the data from all games played in the league.

$$Y_i = \beta_{i_h} - \beta_{i_a} + \beta_1 + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma^2).$$

- Y_i represents the home team's margin of victory.
- β_1 represents the effect of home field advantage.
- β_{i_h} represents the team effect of the home team in the ith game.

Linear Regression Model

Instead of simulations, we could use the data from all games played in the league.

$$Y_i = \beta_{i_h} - \beta_{i_a} + \beta_1 + \epsilon_i, \ \ \epsilon_i \sim N(0, \sigma^2).$$

- Y_i represents the home team's margin of victory.
- β_1 represents the effect of home field advantage.
- β_{i_h} represents the team effect of the home team in the ith game.
- β_{i_a} represents the team effect of the home team in the ith game.

Linear Regression Model

Using R-studio, we can collect data from NFL games to create a design matrix and fit the linear model.

Linear Regression Model

Using R-studio, we can collect data from NFL games to create a design matrix and fit the linear model.

Caution: We cannot estimate individual team effects using lm() since the design matrix is not full column rank.

Linear Regression Results

From our regression model we found that

$$\mathbb{E}[Y] = \hat{\beta}_{\text{JAX}} - \hat{\beta}_{\text{MIA}} + \hat{\beta}_1 = 5.3749.$$

$\mathbb{P}(\text{MIA}) = 0.3293$ and $\mathbb{P}(\text{JAX}) = 0.6706$.

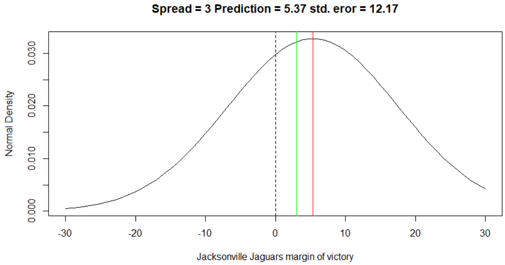

Figure: Probability density for JAX margin of victory according to the regression model.

Linear Regression Results

In R, we found $\mathbb{P}(\text{MIA}) = 0.3293$ and $\mathbb{P}(\text{JAX}) = 0.6706$. If we bet \$100 on MIA, we get \$138 for winning.

Linear Regression Results

In R, we found $\mathbb{P}(\text{MIA}) = 0.3293$ and $\mathbb{P}(\text{JAX}) = 0.6706$. If we bet \$100 on MIA, we get \$138 for winning. According to our model, our expected winnings for betting \$100 on MIA are

$$\mathbb{E}[W_{\text{MIA}}] = 0.3293 \cdot \$138 + 0.6706 \cdot (-\$100) = -\$21.62$$

Linear Regression Results

In R, we found $\mathbb{P}(\text{MIA}) = 0.3293$ and $\mathbb{P}(\text{JAX}) = 0.6706$. If we bet \$100 on MIA, we get \$138 for winning. According to our model, our expected winnings for betting \$100 on MIA are

$$\mathbb{E}[W_{\text{MIA}}] = 0.3293 \cdot \$138 + 0.6706 \cdot (-\$100) = -\$21.62$$

Similarly, betting on \$100 on JAX gives us expected winnings of

$$\mathbb{E}[W_{\text{JAX}}] = 0.6706 \cdot \$63.69 + 3293 \cdot (-\$100) = \$9.78$$

Closing Thoughts on Linear Regression

- This model thought the game was closer, so betting on the underdog was not as bad, but still not worth it.

Closing Thoughts on Linear Regression

- This model thought the game was closer, so betting on the underdog was not as bad, but still not worth it.
- Our model assumes the same random error for each game. This could be improved.
- Estimates for differences in team effects will improve as we see more games.

Closing Thoughts on Linear Regression

- This model thought the game was closer, so betting on the underdog was not as bad, but still not worth it.
- Our model assumes the same random error for each game. This could be improved.
- Estimates for differences in team effects will improve as we see more games.
- Might consider a combination of the Monte Carlo and Regression Model.

Closing Thoughts on Linear Regression

- This model thought the game was closer, so betting on the underdog was not as bad, but still not worth it.
- Our model assumes the same random error for each game. This could be improved.
- Estimates for differences in team effects will improve as we see more games.
- Might consider a combination of the Monte Carlo and Regression Model.
- The regression model should also give some insight on the point spread.

Point Spread

We already determined from the regression model that $\mathbb{E}[W_{\text{JAX}}] = \$9.78$ if we bet on the moneyline. What if we bet on the spread?

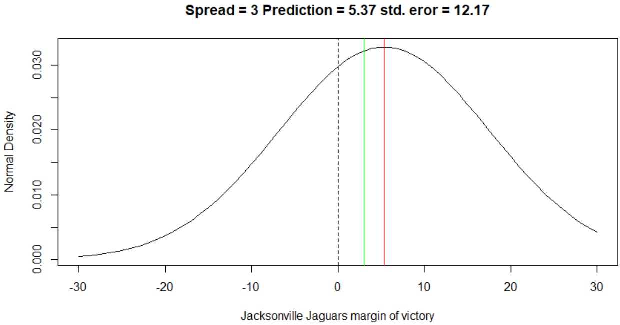

Figure: Probability density for JAX margin of victory.

Point Spread

According to the regression model, the probability of JAX beating the spread is 0.5774.

Point Spread

According to the regression model, the probability of JAX beating the spread is 0.5774.

If we bet \$100 on JAX beating the spread, our expected winnings according to the regression model is

$$\mathbb{E}[W_{\text{JAX}}] = 0.5774(\$90) + 0.4226(-\$100) = \$9.71$$

Point Spread

According to the regression model, the probability of JAX beating the spread is 0.5774.

If we bet \$100 on JAX beating the spread, our expected winnings according to the regression model is

$$\mathbb{E}[W_{\mathsf{JAX}}] = 0.5774(\$90) + 0.4226(-\$100) = \$9.71$$

We very close to \$9.78, but still less. Bet on the money line.

Questions?

[1] Lloyd Danzig. An Intro to Quantitative Modeling for Sports Bettors (in Excel).
https://medium.com/@lloyddanzig/quantitative-sports-betting-6976e1ceaf0f

[2] Lloyd Danzig. Bookmaking Economics: Vigorish & Overround.
https://medium.com/analytics-vidhya/bookmaking-economics-8710d25a42a5

[3]Hal Stern. On the Probability of Winning a Football Game. *The American Statistician*. (1991): 179–183.