MECH&AE 298 Mini-Project 2 Report

William Hsieh

February 13, 2025

1 Introduction

This mini-project infers the result of championship games between two top teams. It does so by analyzing regular season games, and then simulating 500 two-leg finals. A Bayesian statistical model is used in order to calculate the probabilities of each team winning, losing, or drawing a match. A multi-level model is used, which means the parameters vary at more than one level.

2 Our Model

2.1 Multi-Level Model

In this case, there are two levels to this model. First, each team has their own attack and defense statistics (mean μ and variance σ^2). They also have another statistic for "home-field advantage." All of these parameters make up the first level of this model. Using these parameters, the probability of a win as the home team and a win as the away team can be calculated. This makes up the second level, which is used for the results of our model.

In summary, the parameters from the first level of the model (team statistics) are fed into the second level of the model to calculate the scores and win probabilities.

2.2 Description

The mean attacking power, mean defending power, variance of attacking power, variance of defending power, and home advantage are defined as follows:

$$
\begin{aligned}
\mu_{\text{att}} &\sim \mathcal{N}(0, 0.1) \\
\mu_{\text{def}} &\sim \mathcal{N}(0, 0.1) \\
\sigma_{\text{att}} &\sim \exp(1) \\
\sigma_{\text{def}} &\sim \exp(1) \\
\text{home} &\sim \mathcal{N}(0, 1)
\end{aligned}
$$

These are then used to calculate the win probabilities for each team:

$$
\begin{aligned}
\text{win home} &= \text{home advantage} + \text{attack[home]} + \text{defense[away]} - \text{offset} \\
\text{win away} &= \text{attack[away]} + \text{defense[home]} - \text{offset}
\end{aligned}
$$

Note that offsets (calculated via $\mu_{\text{att}} + \mu_{\text{def}}$) are added to center the distribution. Finally, the scores can be calculated using the win probabilities:

$$
\text{score} \sim \text{Poisson}\big(\exp(\text{win})\big)
$$

2.3 Championship Games

After 3000 samples of matchups between the 20 teams, the best two teams are selected. Now, we can sample 500 "finals" between the two teams. Team A and team B take turns being the home team; the win probabilities and scores are calculated using the same methods as above.

However, there is one key difference when calculating the scores. Instead of updating the parameters for each team after each game, the parameters remain unchanged. The scores are instead sampled from the Poisson distribution 500 times for each team, to simulate 500 matches without changes in parameters:

$$\text{score} = \text{rand}\Big(\text{Poisson}\big(\exp(\text{win})\big), 500\Big)$$

The result is a list of scores, one for each team for the 150,000 (3000 $\times$ 500) championship games. (The 3000 comes from the previous step with the 20 teams.)

3 Results

Plotting the scores of team A (Manchester City) and team B (Manchester United) gives the following:

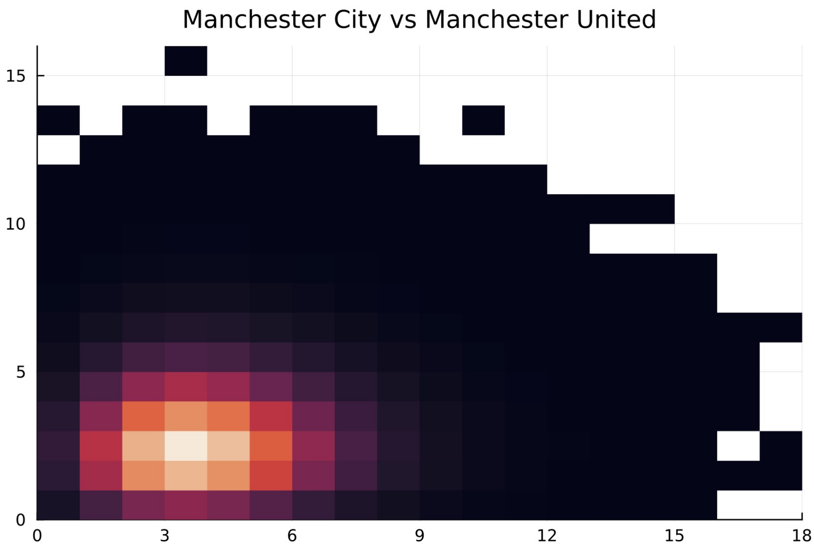

Figure 1: Heatmap of Scores

Each cell represents the number of points each team scored in a game. For example, the whitest spot is the most common, while the darker and non-populated spots show a low frequency. The one brightest spot represents a game where team A scored 3 points and team B scored 2 points, resulting in a win for team A.

By counting the number of occurrences where team A scored more than team B, we can figure out the win ratio for team A, and vice versa for team B. If they scored the same amount of points (represented by the diagonal starting at $(0, 0$, it results in a draw.

- Win probability for team A: 61.504%
- Win probability for team B: 24.090%
- Draw probability: 14.405%

As a sanity check, all of these probabilities add up to 100%.

4 Discussion and Conclusion

Based on the heatmap, we can see that most of the games had around 2-4 points, but in some rare cases could go up to 17 points. The games were also relatively close, with most of the probability mass around the slope $m = 1$. However, this small difference was enough to correspond to a win probability of 61.504% for team A and a much lower win probability of 24.090% for team B.

Overall, this mini-project analyzed games between 20 teams, picked the best 2 teams out from there, and simulated many championship games between these two teams. Bayesian statistics is used to calculate the win probabilities and scores for each game, and a multi-level model was used to convert raw team statistics into a quantifiable score.

5 Code

The Jupyter notebook used is appended to this report.