Corners Challenge

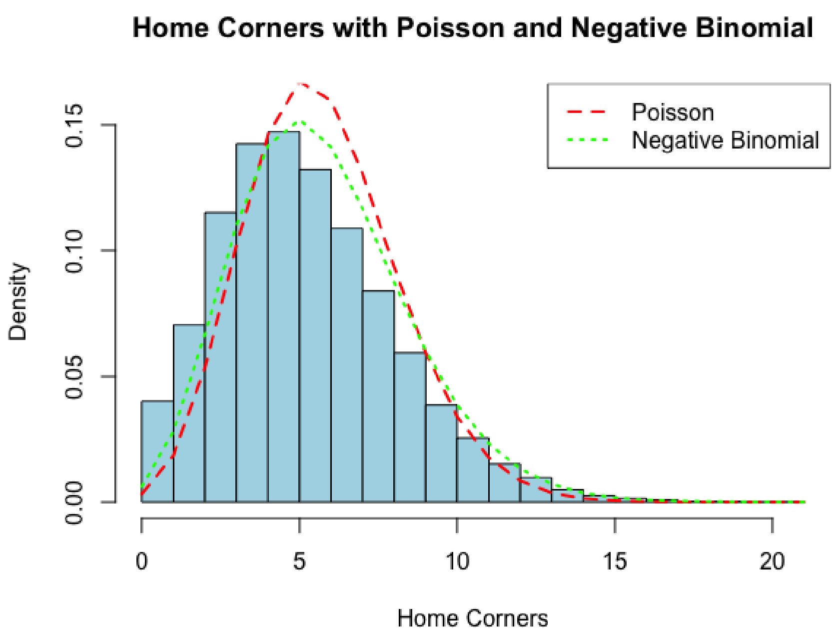
Rishik Kumar

Model selection – Poisson & Negative Binomial

Initially, I decided to test a Poisson model since corners are count data, and there is substantial literature supporting this model. However, I had some concerns because corners could violate the conditions of a Poisson distribution. One of these conditions is the independence of events: "The occurrence of one event does not affect the probability of another event occurring." This is violated in a football game, as corners can often lead to subsequent corners. Another property of the Poisson distribution is that the mean and variance should be the same, which is not the case in the provided data:

	Home Corners	Away Corners
Mean	5.73	4.49
Variance	8.09	6.28

It is clear from the data that there is over-dispersion, indicating that a Negative Binomial Model is much more appropriate due to its additional parameter.



From the histogram, it appears that the negative binomial model fits the training data better, as indicated by its lower AIC. However, the difference in predicted probabilities is not extreme, as shown below;

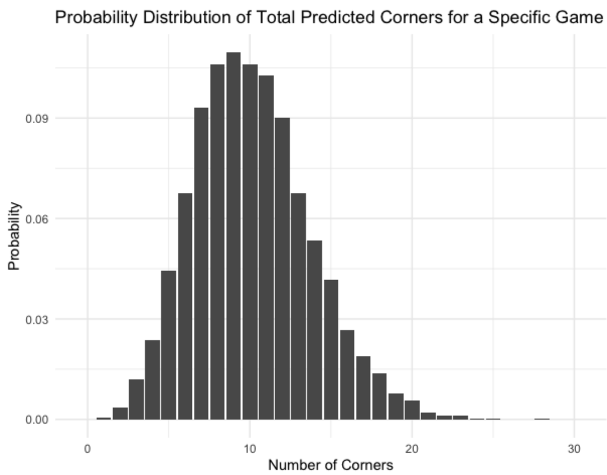
	Pr(Under)	Pr(At)	Pr(Above)
Average difference	0.014	0.003	0.011

Therefore, I also investigated the Pearson Chi Squared and Deviance results indicating considerable improvement over the Poisson model.

I therefore developed two Negative Binomial models to separately predict Home corners and Away corners, both using 'HomeTeamID' and 'AwayTeamID' as variables, which were converted into factors first. This approach inherently accounts for the home/away variable. For instance, the same team (e.g. team 410) is treated differently depending on whether they are playing at home or away, acknowledging the differences in corners given in each scenario. For example, team 410 averages 6.36 corners at home but only 5.15 corners when playing away, indicating a significant difference.

Probability Calculation

To calculate the probabilities, I created a function that simulates a match between the two teams applying both Negative Binomial models (Home and Away), returning a probability matrix containing every possible number of corners from 0 to 30 (as 25 being the highest total in the dataset). Here is a plot demonstrating the first match in the test data;



To handle the line, I considered two scenarios: when the line is an integer and when it is a decimal. When the line is non-integer value there can only be a result of Over or Under as you cannot have a fraction of a corner. For these matches, the floor and ceiling values of the line were used accordingly.

The function then returns the appropriate probabilities in each scenario to arrive at an accurate probability measure.

Here is an example output:

Pr(Under)	Pr(At)	Pr(Above)
0.596	0.000	0.404
0.572	0.000	0.428
0.521	0.000	0.480
0.582	0.094	0.324
0.571	0.089	0.340

Clearly the probability of getting a decimal number of corners is zero as it is not possible in a football match.

Betting Strategy

To determine how to stake the 341 units available, I calculated the expected value (EV) of each bet (Over or Under) and chose the bet with the higher EV. I then used the Kelly criterion to determine the stake amount.

EV:

$$EV_U = \big(P(U) \times (odds - 1)\big) - P(O)$$

$$EV_O = \big(P(O) \times (odds - 1)\big) - P(U)$$

This method accounts for the probability of the stake coming back to you in a push bet.

Kelly Criterion formula:

$$K\,\% = \frac{bp - q}{b}$$

Where b = odds – 1, p = probability of success, q = 1 – p

Negative values indicate a negative EV, so I removed these bets to ensure we avoid betting on games with unfavourable outcomes.
I then normalised the stakes to ensure that the total units add up to 341. This approach spreads the total stakes across many matches, reducing the risk of losing large amounts on a few events, therefore leveraging diversification.

A detailed breakdown of all probabilities and workings can be found in ‘Test_with_workings_RK’.csv, while just the required can be found in ‘Test_RK’.csv. My code can also be found in ‘Corners_RK’.R

In total, the model indicates to only bet on **221** specific matches with **120** proving to be unfavourable bets.