1 Model explanation

To predict corner counts in the test set, we assume corner counts are drawn from a Geometric-Poisson distribution. To account for the lack of information in the training set, we determine the mean of the Geometric-Poisson distribution for each game using a Glicko-like rating system for how "corner-centric" the play of each team is. This is similar to the well-known ELO rating system as used in games such as chess, but ratings become more susceptible to change when there is a lack of recent information about the play of the team. Additionally, we use a Monte Carlo method to account for the future changes in ELO during the time period in which the matches in the test set take place.

To motivate the Geometric-Poisson distribution, consider a Markov chain modelling a football game with two states: open play and corner. It is reasonable to assume that the rate of transitioning from open play to a corner is proportional to the rate of goal scoring, since both of these tend to occur due to good attacking play. The variance-to-mean ratio of total goals per game is $1.06 \approx 1$, suggesting total goals can be modelled well by a Poisson distribution. However, it is well-known that corners off come in clusters, with defenders often putting the ball out for another corner to defend the initial corner. This is supported by the variance-to-mean ratio of total corners per game being $1.48 > 1$, e.g. overdispersed and thus not well-modelled by a Poisson distribution. Since the occurrence of a cluster represents a transition from open play to a corner, then we expect to be able to model clusters using a Poisson distribution. If we assume there is a fixed probability θ of returning to open play (either by clearance or conceding a goal), then the number of corners in each cluster can be drawn from a geometric distribution. Formally, the total corners in a match i where we expect λ_i corner clusters is given by the random variable

$$Y_i = \sum_{j=1}^{N_i} X_j \tag{1}$$

where $N_i \sim \text{Pois}(\lambda_i)$ and $X_j \sim G(\theta)$ for all $j = 1, \ldots, N$. It is well-known that the Geometric-Poisson distribution has mean $\mu_i = \frac{\lambda_i}{\theta}$.

We estimate the rate of corner clusters as follows. Let i be a match in the league l between the home team h and the away team a, μ_l the mean number of corners per game in league l, and R_h and R_a the corner-centric ratings of the teams h and a, respectively. For some weights $w_h, w_a \in \mathbb{R}$, we estimate the corner count in the match i by

$$\hat{\mu}_i(\hat{\mu}_l, \hat{R}_h, \hat{R}_a; w_1, w_2) = \hat{\mu}_l * \exp(w_1 \hat{R}_h + w_2 \hat{R}_a)\,. \tag{2}$$

We calculate $\hat{\mu}_l$ by averaging over the training set. It is trickier to estimate the ratings R_h and R_a. Though it would take considerable computational power, we could train them as team-specific parameters. However, the training set takes place over a 5 year period, so it is more appropriate to treat these ratings as a function of time. Initially, one would like to Taylor expand the ratings as a function of time and estimate the coefficients, but this would be even more computationally expensive. Instead, we will use some of the training set to make predictions with naive corner-centric ratings and thus make subsequent corrections to them. This is reminiscent of the "burn-in" period in a Markov Chain Monte Carlo method. In particular, we outline the following method for a team b:

Corner-centric rating update scheme

1. Before the first match in the training set, initialise the estimated corner-centric rating of b as $\hat{R}_b = 0$.

2. For each match i involving b, compute the expected corner count $\hat{\mu}_i$ according to (2). In particular, if b is at home, then $\hat{R}_h = \hat{R}_b$ and $\hat{R}_a$ is determined by the opponent. If b is the away team, then vice versa.

3. Compute the prediction error $\Delta_i = y_i - \hat{\mu}_i$, where y_i is the actual corner count.

4. Update $\hat{R}_b$ by the rule $\hat{R}_b = \hat{R}_b + k(t; k_{\text{min}}, c, k_{\text{max}})\Delta_i$ where t is the time in days since team b's last match and

$$k(t; k_{\text{min}}, c, k_{\text{max}}) = \begin{cases} k_{\text{max}}\,, & i \text{ is one of team } b\text{'s first 10 matches in either dataset} \\ \min(\sqrt{k_{\text{min}}^2 + c^2 t}, 0)\,, & \text{otherwise} \end{cases} \tag{3}$$

   Here, k_{min}, c and k_{max} are all parameters which we will train. The idea is that it is harder to estimate the corner-centric ratings of teams who do not have enough total data or not enough recent data, so these estimates are less robust against prediction error.

We train the parameters $\theta, w_1, w_2, k_{\text{min}}, c$ and k_{max} by minimising the log-likelihood function of matches in which both teams have at least 10 prior matches in the training set and have both played at least one match in the last 50 days. For clarity, the log-likelihood function is

$$L(\theta, w_1, w_2, k_{\text{min}}, c, k_{\text{max}}) = \sum_{i \in I} \log[\mathbb{P}(\hat{Y}_i = y_i | \theta, w_1, w_2, k_{\text{min}}, c, k_{\text{max}})] \tag{4}$$

where $\hat{Y}_i$ is the random variable which follows the Geometric-Poisson distribution with mean $\lambda_i = \hat{\mu}_i \theta$. We use the Optim package in Julia to minimise this function.

Having determined the optimal parameters, we simulated the test set 10^6 times to estimate the distribution of corner counts for each match. By taking a simulation approach, we account for the fact that the corner-centric ratings are dynamic and so will change throughout the test period, as well as over the first three months of 2011 where we have no data. In other words, we continue to update the corner-centric ratings through each prediction. Since our parameters have been optimised for matches where both teams have at least 10 prior matches in the training set and have both played at least one match in the last 50 days in the training, we will not place bets on any games which fall under this category. Notice that means that we don't place any bets on each team's first match in the test set. While a different prediction model could deal better with this case, it is safe to assume that the bookmaker will have more information than us, since they are not limited to the training set. Finally, we use full Kelly betting sizing to determine our bet sizes as fractions of our total bankroll.