Modeling Football Match Outcomes with Poisson Attack–Defence Ratings

Thomas Gourlay

February 11, 2026

1 Mathematical Model of Football Match Outcomes

We model football match scores using a **Poisson attack–defence model** that accounts for the strength of each team in a given season and includes a global home advantage. To capture changes in team performance over time, we allow ratings to evolve between seasons following a random walk. This is essentially a prior; however, it is not proper. We just use the likelihood of changes in ratings, not the original distribution.

1.1 Notation

Let:

- $i = 1, \ldots, N$ index teams,
- $s = 1, \ldots, S$ index seasons,
- m index matches,
- H_m, A_m denote the home and away teams in match m,
- $Y_m^{(h)}$, $Y_m^{(a)}$ denote the goals scored by the home and away teams.

Each team i in season s has an **attack rating** $\alpha_{i,s}$ and a **defence rating** $\delta_{i,s}$. There is also a **global home advantage parameter** γ that applies across all seasons.

1.2 Poisson Model for Match Goals

We assume that, given the team ratings, goals scored in a match are independent and follow a Poisson distribution:

$$Y_m^{(h)} \sim \text{Poisson}(\lambda_m^{(h)}), \quad Y_m^{(a)} \sim \text{Poisson}(\lambda_m^{(a)})$$

with expected goals:

$$\lambda_m^{(h)} = \exp\left(\gamma + \alpha_{H_m,s(m)} - \delta_{A_m,s(m)}\right), \quad \lambda_m^{(a)} = \exp\left(-\gamma + \alpha_{A_m,s(m)} - \delta_{H_m,s(m)}\right)$$

where $s(m)$ indicates the season of match m. The exponential ensures all expected goal values are positive.

1.3 Identifiability Constraint

Attack and defence ratings are not uniquely identifiable if left unconstrained. To resolve this, we enforce that the attack ratings in each season sum to zero:

$$\sum_{i=1}^{N} \alpha_{i,s} = 0, \quad \forall s$$

1.4 Likelihood of Observed Matches

The likelihood for a single match is given by the Poisson probabilities for the home and away goals. Taking logs and summing over all matches gives the total log-likelihood:

$$\mathcal{L}_{\text{matches}}(\theta) = \sum_{s=1}^{S} \sum_{m \in s} \log \mathbb{P}(Y_m^{(h)} \mid \lambda_m^{(h)}) + \log \mathbb{P}(Y_m^{(a)} \mid \lambda_m^{(a)})$$

1.5 Modeling Rating Changes Between Seasons

Team performance can change from one season to the next. To model this, we assume ratings evolve smoothly following a random walk. Specifically, for each team i and season $s > 1$:

$$\alpha_{i,s} = \alpha_{i,s-1} + \varepsilon_{i,s}^{(\alpha)}, \quad \delta_{i,s} = \delta_{i,s-1} + \varepsilon_{i,s}^{(\delta)}, \quad \varepsilon_{i,s}^{(\alpha)}, \varepsilon_{i,s}^{(\delta)} \sim \mathcal{N}(0, \sigma^2)$$

Here, σ controls how much ratings are allowed to change from season to season. This is the improper prior distribution. Moreover, this is equivalent to adding a Gaussian penalty to differences in ratings between consecutive seasons:

$$\mathcal{L}_{\text{RW}}(\theta) = \sum_{s=2}^{S} \sum_{i=1}^{N} \left[-\frac{(\alpha_{i,s} - \alpha_{i,s-1})^2}{2\sigma^2} - \frac{(\delta_{i,s} - \delta_{i,s-1})^2}{2\sigma^2} \right] + \text{const}$$

1.6 Total Likelihood and Parameter Estimation

The total log-likelihood combines the match likelihood and the inter-season random walk:

$$\mathcal{L}_{\text{total}}(\theta) = \mathcal{L}_{\text{matches}}(\theta) + \mathcal{L}_{\text{RW}}(\theta)$$

The parameter vector θ contains all attack and defence ratings for every season, plus the home advantage:

$$\theta = (\alpha_{1,1}, \ldots, \alpha_{N,1}, \delta_{1,1}, \ldots, \delta_{N,1}, \ldots, \alpha_{1,S}, \ldots, \delta_{N,S}, \gamma)$$

We estimate parameters using **maximum likelihood**, typically via numerical optimisation such as L-BFGS-B.

1.7 Using the Fitted Model

Once fitted, the model can produce probabilities for derived events. For example, the probability that the total goals in a match exceed k is:

$$\mathbb{P}(Y_m^{(h)} + Y_m^{(a)} > k) = 1 - F_{\text{Poisson}}(k; \lambda_m^{(h)} + \lambda_m^{(a)})$$

This allows direct computation of over/under probabilities and other useful statistics.