Poisson Model for Live Football Probability Prediction

Pietro Bogani
linkedin.com/pietrobogani · GitHub Repository

February 5, 2026

Abstract

This document presents a statistical model for predicting real-time probability changes in football matches using Poisson and Skellam distributions. The model calibrates expected goal rates from public live betting market quotes and computes how match outcome probabilities shift when goals are scored.

Contents

1 Introduction

During a live football match, the probability of each outcome (Home Win, Draw, Away Win) changes as goals are scored. This model provides a mathematical framework to predict these probability shifts using:

- **Poisson distribution** for modeling goal-scoring rates
- **Skellam distribution** for the difference between two Poisson variables
- **Dual calibration** from live market data

2 Mathematical Foundation

2.1 Poisson Distribution for Goals

Football goals are modeled as independent Poisson processes. If a team has an expected number of remaining goals λ, the probability of scoring exactly k goals is:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots \tag{1}$$

We define:

- λ_H = expected remaining goals for home team
- λ_A = expected remaining goals for away team

2.2 Skellam Distribution

The difference between two independent Poisson random variables follows a **Skellam distribution**. If $X \sim \text{Poisson}(\lambda_H)$ and $Y \sim \text{Poisson}(\lambda_A)$, then:

$$\Delta = X - Y \sim \text{Skellam}(\lambda_H, \lambda_A) \tag{2}$$

The probability mass function is:

$$P(\Delta = k) = e^{-(\lambda_H + \lambda_A)} \left(\frac{\lambda_H}{\lambda_A}\right)^{k/2} I_{|k|}(2\sqrt{\lambda_H \lambda_A}) \tag{3}$$

where $I_n(x)$ is the modified Bessel function of the first kind.

2.3 Match Outcome Probabilities

Let d be the current goal difference (home goals − away goals). The final goal difference will be $d + \Delta$ where Δ is the Skellam-distributed remaining goal difference.

$$P(\text{Home Win}) = P(d + \Delta > 0) = \sum_{k > -d} P(\Delta = k) \tag{4}$$

$$P(\text{Draw}) = P(d + \Delta = 0) = P(\Delta = -d) \tag{5}$$

$$P(\text{Away Win}) = P(d + \Delta < 0) = \sum_{k < -d} P(\Delta = k) \tag{6}$$

3 Calibration from Market Data

3.1 The Calibration Problem

Given observed market probabilities, we need to find λ_H and λ_A. This is an inverse problem: we observe the output (probabilities) and must infer the inputs (Poisson parameters).

3.2 Dual Calibration Method

Using only 1X2 (match winner) prices leaves the problem underdetermined—many combinations of (λ_H, λ_A) can produce similar 1X2 probabilities. For example, $(\lambda_H, \lambda_A) = (2, 1)$ and $(4, 2)$ yield a similar home-win edge (both have $\lambda_H/\lambda_A \approx 2$), but the second pair implies far more total goals. The Over/Under market distinguishes between these cases. We solve this using **dual calibration**:

3.2.1 Step 1: Total Goals from Over/Under Markets

Define:

- $m = \lambda_H + \lambda_A$ (total expected remaining goals)
- g = current total goals scored in the match
- $n.5$ = the Over/Under line (e.g., 2.5)

The sum of two independent Poisson variables is also Poisson:

$$X + Y \sim \text{Poisson}(m) \tag{7}$$

The probability of going Over $n.5$ goals is the probability that remaining goals plus current goals exceed $n.5$:

$$P(\text{Over } n.5) = P(\text{remaining goals} \geq \lceil n.5 \rceil - g) \tag{8}$$

Using the Poisson CDF $F_{\text{Poisson}}(k; m) = P(X \leq k)$:

$$P(\text{Over } n.5) = 1 - F_{\text{Poisson}}(\lceil n.5 \rceil - g - 1; m) = f(m) \tag{9}$$

Calibration: Given from real market data an observed market probability $P_{\text{over}}^{\text{market}}$, we seek $m = f^{-1}(P_{\text{over}}^{\text{market}})$. Since $f(m)$ has no closed-form inverse but is monotonically increasing in m, we use binary search until convergence.

Note that the match minute does not enter the model explicitly. The market prices already encode the remaining time: as the match progresses, Over/Under probabilities naturally decrease, yielding lower m values through calibration. The model delegates time modeling to the market.

3.2.2 Step 2: Goal Share from 1X2 Markets

Define $q = \lambda_H/m$ (home team's share of remaining goals).

Given m from Step 1:

$$\lambda_H = m \cdot q \tag{10}$$

$$\lambda_A = m \cdot (1 - q) \tag{11}$$

We find q by minimizing the squared error between model and observed 1X2 probabilities:

$$q^* = \arg\min_q \sum_{i \in \{H,D,A\}} \left(P_i^{\text{model}}(q) - P_i^{\text{market}} \right)^2 \tag{12}$$

Define $L(q) = \sum_i (P_i^{\text{model}}(q) - P_i^{\text{market}})^2$ as the loss function. Given m from Step 1, we compute $P_i^{\text{model}}(q)$ via the Skellam distribution with $\lambda_H = mq$ and $\lambda_A = m(1-q)$.

Since $L(q)$ has no closed-form minimum, we use numerical optimization (e.g., Nelder-Mead) over $q \in [0,1]$ to find q^*.

4 Goal Impact Prediction

4.1 Computing New Probabilities

When a goal is scored, we:

1. Update the goal difference: $d' = d \pm 1$
2. Keep λ_H, λ_A unchanged (calibrated before the goal)
3. Recompute probabilities using the Skellam distribution with new d'

4.2 Example Calculation

Consider a match at 0-0 in the 30th minute with:

- Market: $P(H) = 50\%$, $P(D) = 28\%$, $P(A) = 22\%$
- Calibrated: $\lambda_H = 0.85$, $\lambda_A = 0.65$

If home scores (new score 1-0, $d' = 1$):

$$P'(\text{Home Win}) = P(\Delta > -1) = \sum_{k=0}^{\infty} P(\Delta = k) = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} P(X = j+k) \cdot P(Y = j) \approx 71\%$$

where $\Delta \sim \text{Skellam}(\lambda_H, \lambda_A)$ is the random variable representing the home minus away goals yet to be scored. $X \sim \text{Poisson}(0.85)$ and $Y \sim \text{Poisson}(0.65)$. Similarly:

$$P'(\text{Draw}) = P(\Delta = -1) = \sum_{j=0}^{\infty} P(X = j) \cdot P(Y = j+1) \approx 19\% \tag{13}$$

$$P'(\text{Away Win}) = P(\Delta \le -2) = 1 - P'(\text{Home}) - P'(\text{Draw}) \approx 10\% \tag{14}$$

The home win probability increases by approximately +21%.

5 Over/Under Markets

5.1 Over/Under Probability

For Over/Under $n.5$ goals, with current total g and expected remaining goals $m = \lambda_H + \lambda_A$:

$$P(\text{Over } n.5) = 1 - F_{\text{Poisson}}(\lceil n.5 \rceil - g - 1; m) \tag{15}$$

Using the same match state ($\lambda_H = 0.85$, $\lambda_A = 0.65$, so $m = 1.5$), at 0-0 ($g = 0$) for Over 2.5:

$$P(\text{Over } 2.5) = 1 - F_{\text{Poisson}}(2; 1.5) \approx 19\%$$

After home scores ($g' = 1$), we now need only 2 more goals to exceed 2.5:

$$P'(\text{Over } 2.5) = 1 - F_{\text{Poisson}}(1; 1.5) \approx 44\%$$

The Over 2.5 probability increases by +25% with the goal.

6 Model Validation

The model was validated on 20 goals across 10 matches in several football leagues:

Metric	Value
Mean Absolute Error	4.4%
Median Absolute Error	3.6%
Predictions within 5%	65%
Predictions within 10%	100%

Table 1: Validation results comparing predicted vs actual market probabilities

6.1 Visual Examples

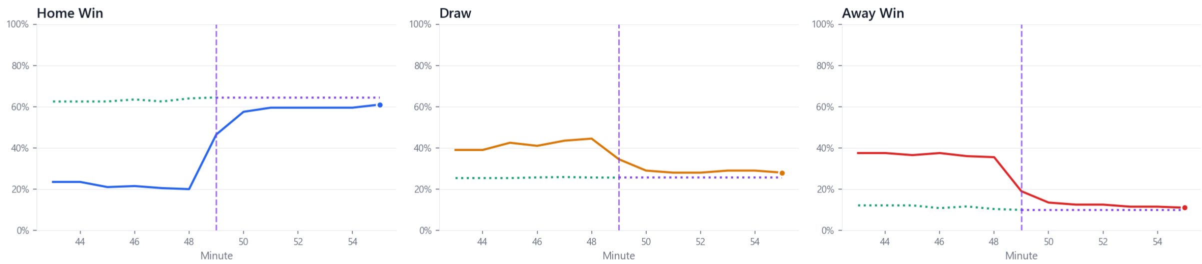

Figure 1: Udinese vs Roma – [H] 49’ Ekkelenkamp (1-0). MAE: 3.3%

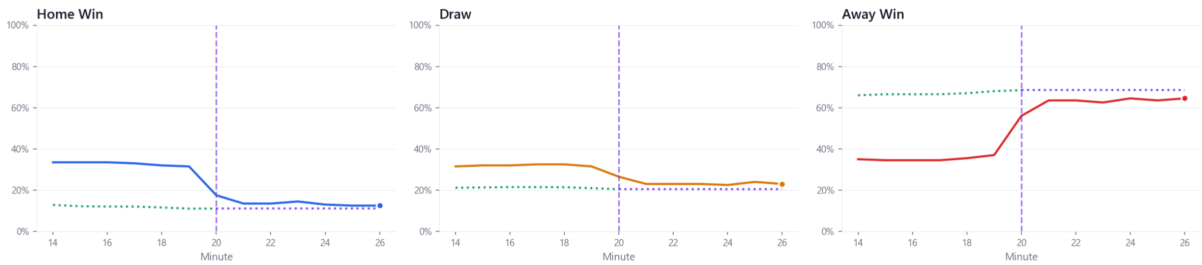

Figure 2: Bologna vs Milan – [A] 20’ Loftus-Cheek (0-1). MAE: 2.5%

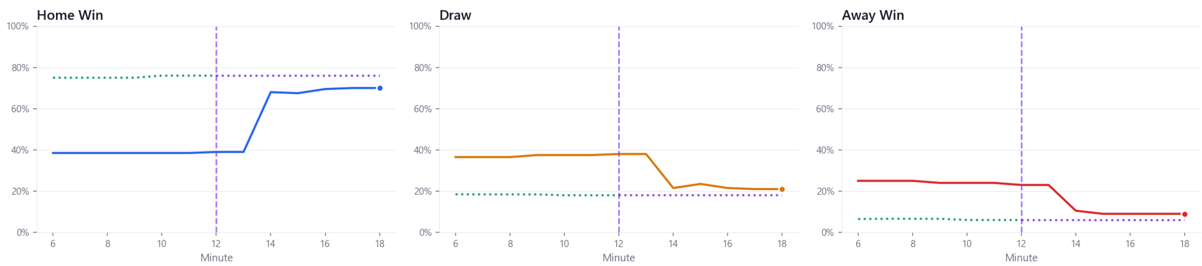

Figure 3: Universitatea Cluj vs FC Arges – [H] 12’ Macalou (1-0). MAE: 3.9%

7 Limitations

- **Independence assumption**: Goals are assumed independent, ignoring momentum effects
- **Static rates**: λ values don't change with match context (red cards, substitutions)
- **Late-game accuracy**: Predictions less reliable when little time remains
- **Market efficiency**: Assumes market prices reflect true probabilities

8 Conclusion

The Poisson/Skellam model provides a principled statistical framework for predicting live football probabilities. The dual calibration approach – using Over/Under markets to pin down total expected goals m, and 1X2 markets to split them into λ_H and λ_A – resolves the underdetermined inverse problem without requiring historical data or match-specific priors, yielding accurate predictions.