Modeling football match results using goal scoring probabilities

Raymond Chew - 1001254910
Lorenzo Kogler Anele - 1003795529

1 Short Description

We plan on building a mathematical model of goal-scoring in soccer matches. By looking at goals as counts during a given time-frame, it is natural describe goals scored by a team during a match as a Poisson distribution. After optimizing the parameters, we plan on running simulations to see if they match actual results in terms of wins/draws/losses. If time permits, we plan on running our model's probabilities against that of betting websites, for a possible application.

2 Background

In soccer, the scoring of a goal is considered a rare and unpredictable event. As such there is significant interest among academics and fans alike in predicting the outcome of soccer matches. To date, there have been numerous modelling approaches attempting to predict team performance for a given match. A popular approach is to assume that the number of goals scored will follow a Poisson distribution. This has been accomplished with varying degrees of effectiveness. For example, Inan (2020) considered the average number of goals scored and conceded during a game as metrics for offensive and defensive strength respectively. Using these as parameters, this model yielded accuracies ranging from 52% to 62% across different leagues. Other models such as the one constructed by Azhari, Widyanginsih and Lestari (2018), used a Poisson regression model which accounted for not only attack and defense strength but also the home team advantage. This model yielded a prediction accuracy of 61% accross 100 games.

3 Previous work and resources

- Inan (2021)
- Azari et al (2018)
- Betting example
- Datasets

4 Our contributions and analysis

There is plenty of work using Poisson models to predict goal scoring in soccer matches, specially with the financial incentive in gambling. However, most

use a more naive approach to estimating parameters, by using only averaged quantities as estimators for the Poisson mean. For a given match between teams i and j, with i playing at home, if we seek to model the number of goals scored by i, we will use a Poisson distribution with mean given by the product of i's offensive strength at home and j's defensive strength away. Note that this assumes that the number of goals scored by i is independent of the number of goals scored by j.

With the parameters as above, the Poisson distribution's pdf is given by

$$P(X = k) = p(k) = \frac{(O_{H,i} \cdot D_{A,j})^k \cdot \exp(-O_{H,i} \cdot D_{A,j})}{k!}$$

where $O_{H,i}$ represents i's offensive strength at home and $D_{A,j}$ is j's defensive strength away. $p(k)$ is the probability that i scores k goals in the match. To estimate these parameters, we will use gradient descent to maximise the log-likelihood of each match. This treats each match independently of each other, with the only relationship being the teams' properties in the Poisson parameter. The gradient descent update is given by

$$O_{H,i} \leftarrow O_{H,i} + \alpha * \left(\frac{k}{O_{H,i}} - D_{A,j} \right)$$
$$D_{A,j} \leftarrow D_{A,j} + \alpha * \left(\frac{k}{D_{A,j}} - O_{H,i} \right)$$

for learning rate α.

We also plan on implementing a momentum factor, which will be common to all teams for the sake of simplicity. This momentum factor will multiply the Poisson parameter: $O_{H,i} \cdot D_{A,j} \cdot M^{s_i}$ where s is the number of wins in a row that team i has. This term will be common for all teams because a season only consists of 38 games, so winning streaks are rare and the data is sparse.

5 Expectations

We expect our model to be able to predict match results to a certain extent, but more importantly wins/draws/losses, and overall league standings. We will run simulations with the fitted parameters to be able to compare the our results with the actual season, and validate the model's predictions. We will also look for whether the our parameters have a significance in predicting a team's performance (a good offense should theoretically score more often). Lastly, we will hopefully be able to produce win/draw/loss probabilities for a given match, which can be compared to betting house odds. If there is a discrepancy between the two, and we trust a model, we would expect that it produces a profitable betting strategy.

6 References

1. Inan, T. (2020). Using poisson model for goal prediction in European football. Journal of Human Sport and Exercise, in press. doi:https://doi.org/10.14198/jhse.2021.164.16

2. Azhari, H., Widyaningsih, Y., amp; Lestari, D. (2018). Predicting Final Result of Football Match Using Poisson Regression Model. Journal of Physics. doi:10.1088/1742-6596/1108/1/012066