Poisson Modeling and Predicting English Premier League Goal Scoring

Quang Nguyen
nnguyen22@luc.edu

Follow this and additional works at: https://ecommons.luc.edu/grs

Nguyen, Quang, "Poisson Modeling and Predicting English Premier League Goal Scoring" (2021). Graduate Research Symposium. 1.
https://ecommons.luc.edu/grs/2021/Posters/1

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Poisson Modeling and Predicting English Premier League Goal Scoring

Quang Nguyen – MS Student in Applied Statistics

Abstract: The English Premier League is well-known for being not only one of the most popular professional sports leagues in the world, but also one of the toughest competitions to predict. The first purpose of this research was to verify the consistency between goal scoring in the English Premier League and the Poisson process; specifically, the relationships between the number of goals scored in a match and the Poisson distribution, the time between goals throughout the course of a season and the exponential distribution, and the time location of goals during football games and the continuous uniform distribution. We found that the Poisson process and the three probability distributions do perform a great job of describing Premier League goal scoring. In addition, Poisson regression was utilized to predict outcomes for a Premier League season, using different sets of season data and with a large number of simulations being involved. We examined and compared various soccer metrics from our simulation results, including English club's chances of being the champions, finishing in the top four and bottom three, and relegation points.

Related link: github.com/qntkhvn/eplgoals

nnguyen22@luc.edu
Background and Data

• Poisson process describes the occurrences of events over a continuous interval, with key distributions:
  - Poisson (number of events)
  - Exponential (time between events)
  - Uniform (time locations of events)
→ Can soccer goal scoring be modeled by a Poisson process?

• Poisson regression is used to model responses that are counts and follow a Poisson distribution (like number of goals)
→ Determine teams’ expected scoring rate
→ Simulate and examine results

Data

All EPL final scores from 1992-93 to 2018-19

| Season    | HomeTeam | AwayTeam  | Home.Goals | Away.Goals |
|-----------|----------|-----------|------------|------------|
| 2018-2019 | Man United | Cardiff  | 0          | 2          |
| 2018-2019 | Southampton | Huddersfield  | 1          | 1          |
| 2018-2019 | Tottenham | Everton  | 2          | 2          |
| 2018-2019 | Watford  | West Ham | 1          | 4          |

All Manchester United's goal scoring times in 2018-19

| Min | Matchweek | H1_stoppage | H2_stoppage | TimeBetween |
|-----|-----------|-------------|-------------|-------------|
| 3   | 1         | 1           | 2           | 5           | 0           |
| 83  | 1         | 2           | 5           | 0           |
| 34  | 2         | 2           | 5           | 6           | 46          |
| 95  | 2         | 5           | 6           | 6           | 66          |
Goal Scoring and the Poisson Process

- How well do our data fit the distributions?
  1. Poisson and number of goals
  2. Exponential and time between goals
  3. Uniform and (re-scaled) goal times

- Goodness-of-fit tests
  - Chi-square for (1)
  - Kolmogorov-Smirnov for (2) and (3)

Chi-square for number of goals:

| Goals | ObsMatches | PoisProb | ExpMatches |
|-------|------------|----------|------------|
| 0     | 158        | 0.147    | 153        |
| 1     | 286        | 0.282    | 293        |
| 2     | 282        | 0.270    | 280        |
| 3     | 178        | 0.173    | 180        |
| 4+    | 134        | 0.128    | 133        |
Using Poisson Regression to Predict Season Outcomes

• Fit Poisson regression models (Goals ~ Teams) to get expected scoring rate for home and away games from the models’ coefficients

• Predict “what would happen” in 2018-19
  - 10000 simulations
  - 3 subsets of season data
    (all seasons, the 2010s, all seasons but put more weight on recent years)
  - Compare different metrics
    (champions, top 4, bottom 3,…)

• Future work
  - Factor in more soccer statistics
  - Organizational changes
  - Win probabilities → betting odds

| HomeTeam   | HomeRate | AwayTeam  | AwayRate | HomeScore | AwayScore |
|------------|----------|-----------|----------|-----------|-----------|
| Southampton| 1.440    | Brighton  | 0.684    | 3         | 1         |
| West Ham   | 1.440    | Cardiff   | 0.658    | 0         | 0         |
| Newcastle  | 1.650    | Tottenham | 1.274    | 2         | 1         |
| Huddersfield| 0.684  | Brighton  | 0.684    | 0         | 1         |

Scoring rates and randomly generated scores

| Rank | Team        | FinalPoints | GD |
|------|-------------|-------------|----|
| 1    | Man United  | 86          | 43 |
| 2    | Arsenal     | 81          | 33 |
| 3    | Tottenham   | 67          | 15 |
| 4    | Man City    | 63          | 12 |
| 5    | West Ham    | 59          | 11 |
| 6    | Leicester   | 58          | 0  |
| 7    | Liverpool   | 57          | 13 |

Sample simulated season

| Team     | All Seasons | 2010s | Assign Weight |
|----------|-------------|-------|---------------|
| Arsenal  | 19.68       | 15.05 | 14.07         |
| Chelsea  | 14.28       | 11.70 | 9.03          |
| Liverpool| 12.50       | 10.96 | 17.61         |
| Man City | 7.00        | 41.71 | 38.09         |
| Man United| 36.99      | 10.47 | 10.53         |
| Tottenham| 3.91        | 7.38  | 8.34          |

Chances (%) of winning 2018-19 season

nnguyen22@luc.edu - 4