Poisson Modeling and Predicting English Premier League Goal Scoring

Quang Nguyen [1*]

[1] Loyola University Chicago - Chicago, IL

[*] Work completed at Wittenberg University - Springfield, OH - advised by *Dr. Douglas Andrews*

Abstract: 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 and the Poisson distribution, the time between goals and the exponential distribution, and the time location of goals in a match and the continuous uniform distribution. It turned out to be 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 a Premier League season's results, using different sets of season data and with a large number of simulations being involved.

Related links:

- bookdown.org/theqdata/honors_thesis
- github.com/qntkhvn/eplgoals

Background and Data

- <u>Poisson process</u> 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 goal scoring be modeled by a Poisson process?

- <u>Poisson regression</u> 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

- <u>Data</u>

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 EPL final scores from 1992-93 to 2018-19

Min	Matchweek	H1_stoppage	H2_stoppage	TimeBetween
3	1	2	5	0
83	1	2	5	82
34	2	5	6	46
95	2	5	6	66

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

#CMSAC2020

Goal Scoring and the Poisson Process

- <u>How well do our data fit the distributions?</u>
    - (1) Poisson and number of goals
    - (2) Exponential and time between goals
    - (3) Uniform and (re-scaled) goal times
- <u>Goodness-of-fit tests</u>
    - Chi-square for (1)
    - Kolmogorov-Smirnov for (2) and (3)

(2)

(1)

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

(3)

nminhquang.99@gmail.com - 3

#CMSAC2020
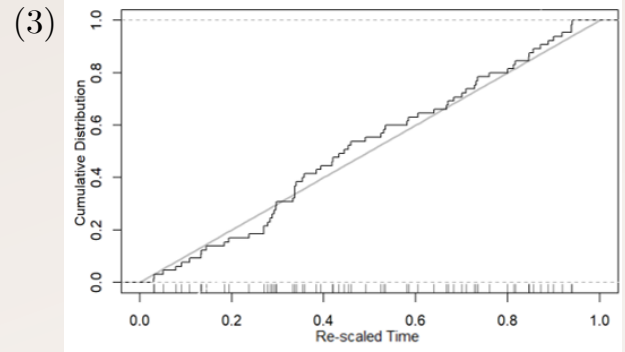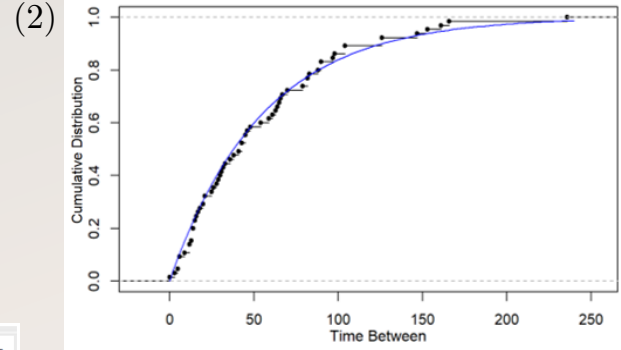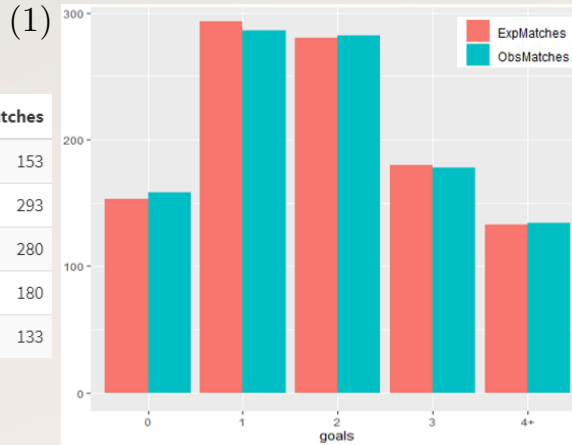
Using Poisson Regression to Predict Season Outcomes

- Fit <u>Poisson regression</u> models (Goals ~ Teams) to get expected scoring rate for home and away games from the models' coefficients

- <u>Predict "what would happen"</u> 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,...)

- <u>Future work</u>
  - 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

#CMSAC2020