

Poisson Modeling and Predicting English Premier League Goal Scoring

QUANG NGUYEN

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 accurately describe 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 an English club's chances of being the champions, finishing in the top four and bottom three, and relegation points.

KEYWORDS AND PHRASES: English Premier League, sports statistics, Poisson process, Poisson regression.

1. INTRODUCTION

Association football, also commonly known as soccer in America, is undoubtedly the most widely-played sport in the world. Often referred to as the “king of sport,” football can be played almost anywhere, from grass fields to indoor gyms, streets, parks, or beaches, due to the simplicity in its principal rules and essential equipment. Europe is known to be the birthplace of modern football [6], and the European soccer culture is unlike any other. The Old Continent is home to numerous top-level professional football leagues, and the English Premier League (EPL) distinguishes itself because of its competition quality, overall balance, and popularity. Some of the best football coaches and players in the world come together to compete for the prestigious Premier League trophy.

The EPL was founded in 1992, and over the last three decades, we have witnessed numerous memorable matches and countless outstanding performances by clubs and their players. The EPL is currently a competition of twenty English football clubs. At the end of each season, the bottom three teams get relegated to the second-highest division of English football, in exchange for three promoted teams. A Premier League season usually takes place from mid-August to mid-May. Each team gets to play every other team twice, once at home and once on the road, hence there are a total of thirty-eight fixtures in a season for each team [8].

The most important aspect of the game of football is indisputably scoring goals. Despite the significance of other factors like ball possessing or disciplined defending, we have

to admit that the main reason we pay to watch soccer is to see the ball being put in the back of the net. The rule is very simple: in order to win, you must score more than your opponent. In the Premier League, each match happens within the span of ninety minutes (plus stoppage time), and the match consists of two 45-minute halves. Each team can get one of these three results after each match: a win, a draw, or a loss. If there is a draw, the two clubs receive a point apiece, and for non-drawing matches, the winner is rewarded with three points and the losing team gets punished with zero points. Thus the club with the most points at the end of the year will have their hands on the exquisite EPL trophy, and the total points also determines the fates of teams in the relegation zone [8]. This makes every single match so critical, as losing one single point could end up costing a team's chance of winning a title or remaining in the top tier football league in England.

In this paper, we attempt to use statistical methods to model and predict goal scoring and match results in the Premier League. We will first determine whether notable aspects of goal scoring, namely, the number of goals scored, the time between goals, and time location of goals in a match, fit the characteristics of a Poisson process. We will then use Poisson regression to predict what would happen in the 2018–19 EPL season, for instance, which clubs are more likely to win the title or get relegated, using different subsets of data from prior seasons. The paper is outlined as follows: We first introduce the data that we used for our analyses in Section 2. Next, our methodologies are described in Section 3. We then spend section 4 on our two main topics of

this research – using the Poisson process to model goal scoring, and utilizing Poisson regression to predict the 2018–19 season outcomes. Lastly, in Section 5, we give a quick summary of our results as well as discuss possible future work related to this research.

2. DATA

The first dataset for our investigation simply consists of match final scores of all Premier League games from its inaugural competition, the 1992–93 season, to the last fixture of 2018–19 season. The main attributes of this dataset are the season, the home and away teams, and the number of goals scored by each team. We rely on Football-Data.co.uk’s data [4], which contains all Premier League match final scores from 1993 to 2019. Each season has its own data file, and we read in and then join the individual datasets together to get our desired data table. We utilize this data to model the number of goals scored and then to make predictions of the 2018–19 season, using three different subset of seasons: 1) data from all seasons prior to 2018–19, 2) data from only the 2010s, and 3) data from all seasons, but assigning more weight to more recent competitions.

To obtain the data for the first two season subsets, we simply filter out the seasons that don’t belong to the year ranges from the initial table. For the assigning weight simulation method, our weight allocation approach is very simple, as we let the weight number be equivalent to the number of times the data for a particular season is duplicated. We have decided that the previous five years before 2018–19 are almost all that matter. Thus, every season from 1992–93 to 2012–13 are given weight 1, then the weight increases by 1 for each one of 2013–14, 2014–15, and 2015–16. After that, we have the 2 most recent years left and we multiply the weight by 2. Our weight values are depicted in Figure 1. Finally, since we are also interested in examining goal scoring time and the time between goals in the EPL, data on these two topics for Manchester United, a well-known Premier League club, are collected and stored in a spreadsheet by ourselves. This data set contains five columns: minutes, which is the point of time during a match at which a goal is scored; matchweek, which is the fixture number of each game; the stoppage time in minutes for both halves of each game, and finally, the time between goals, which is the difference in minutes between the scoring events, taking into account stoppage time. We collected these five variables for all Manchester United goals during their 2018–19 Premier League campaign [9]. In addition to the initial attributes, we created a new variable for normalizing the goal scoring minutes by dividing each one of them by the total minutes of their respective game. The reason for re-scaling the minutes is because the match total time varies, since we also take into account stoppage time of football games, which means some matches take place for a longer time than others.

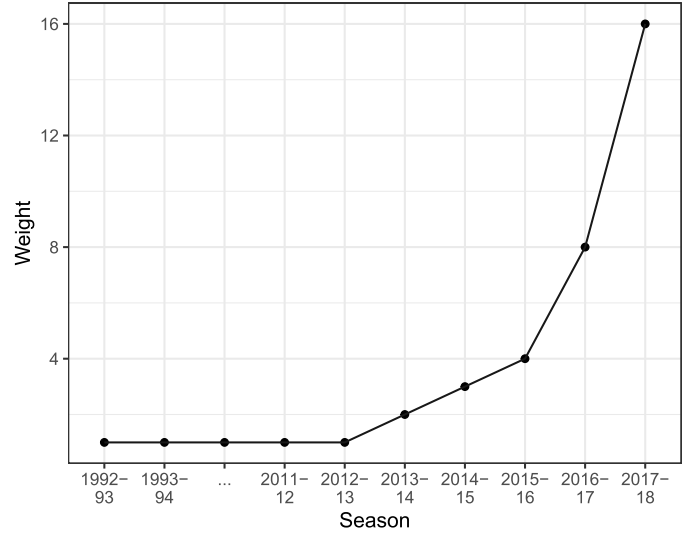


Figure 1: Weight values across seasons from 1992–93 to 2017–18 for predicting 2018–19 season outcomes.

3. METHODOLOGY

3.1 Goal Scoring and the Poisson Process

The Poisson process [3] is a stochastic process used to model the occurrence of phenomena over a continuous interval, which in most cases represents time. There are several characteristics of the Poisson process that can be observed, including, the number of events happening in a given time period; the time between those events; and when (at what point of time) the events occur. Playing a huge role in the Poisson process is the Poisson distribution [1], which deals with the number of occurrences of an event in a fixed period of time, with a rate of occurrence parameter λ . Named for French mathematician Siméon Denis Poisson, the Poisson distribution is a discrete probability distribution that expresses the number of occurrences of an event over a given period of time. The probability density function of a Poisson random variable X with parameter λ is given by

$$p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}; \quad x = 0, 1, 2, \dots \text{ and } \lambda > 0, \quad (3.1)$$

where X represents the number of occurrences of an event in a given unit time period, and λ is the constant rate of occurrence per time period. The mean and variance of our Poisson random variable X , denoted by μ_X and σ_X^2 respectively, are

$$\mu_X = \lambda \quad \text{and} \quad \sigma_X^2 = \lambda. \quad (3.2)$$

Another key distribution in this process is the exponential distribution [1], which has a strong connection with the Poisson distribution, in that if the number of occurrences per interval of time are illustrated by Poisson, then the description of the length of time between occurrences are provided

by the exponential distribution. If we have a non-negative random variable X that is the time until the next occurrence in a Poisson process, then X follows an exponential distribution with probability density function

$$f_X(x) = \lambda e^{-\lambda x} = \frac{1}{\beta} e^{-\frac{1}{\beta}x}; \quad x \geq 0, \quad (3.3)$$

where λ represents the average rate of occurrence and β is the average time between occurrences. The mean and variance of an exponentially distributed random variable X are

$$\mu_X = \frac{1}{\lambda} = \beta \quad \text{and} \quad \sigma_X^2 = \frac{1}{\lambda^2} = \beta^2. \quad (3.4)$$

Furthermore, there is a connection between Poisson and another famous probability distribution – the continuous uniform distribution [1]. If a Poisson process contains a finite number of events in a given time interval, then the unordered times, or locations, or positions, or points of time at which those events happen are uniformly distributed on that continuous interval. The continuous uniform distribution is a probability distribution with equally likely outcomes, meaning that its probability density is the same at each point in an interval $[A, B]$. A continuous random variable X is uniformly distributed on $[A, B]$ if its probability density function is defined by

$$f_X(x) = \frac{1}{B - A}; \quad A \leq x \leq B. \quad (3.5)$$

In addition, X has mean and variance

$$\mu_X = \frac{A + B}{2} \quad \text{and} \quad \sigma_X^2 = \frac{(B - A)^2}{12}. \quad (3.6)$$

We postulate that goal scoring in football can be modeled by a Poisson process. According to the characteristics described above, if goal scoring for a club happens at a certain rate in a given time period, then a Poisson distribution can be used to model the number of goals scored. Additionally, the waiting time (in minutes) between successive goals can be described using an exponential distribution. Moreover, the time positions (or “minute marks”) in a game at which scoring events transpire may be uniformly distributed. We will explore these relationships in more detail in Section 4.

3.2 Simulating and Predicting Season Outcomes Using Poisson Regression

Our second goal of this research is to use the method of Poisson regression to predict the outcomes for EPL matches. Poisson regression is a member of a broad class of models known as the Generalized Linear Models (GLM) [5]. A generalized linear model has the general form

$$E(Y_i) = \mu_i = g^{-1}(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}). \quad (3.7)$$

There are three main components to a generalized linear model:

1. A random component, indicating the conditional distribution of the response variable Y_i (for the i th of n independently sampled observations), given the values of the explanatory variables. Y_i 's distribution must be a member of an exponential family, such as Gaussian, Binomial, Poisson, or Gamma.

2. A linear predictor ($\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$), which is a linear combination of the predictors (the X 's), with the β 's as the regression coefficients to be estimated.

3. A canonical link function $g(\cdot)$, which transforms the expected value of the response variable, $E(Y_i) = \mu_i$, to the linear predictor.

Poisson regression models are generalized linear models with the natural logarithm as the link function. It is used when our response's data type is a count, which is appropriate for our case since our count variable is the number of goals scored. The model assumes that the observed outcome variable follows a Poisson distribution and attempts to fit the mean parameter to a linear model of explanatory variables. The general form of a Poisson regression model is

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}. \quad (3.8)$$

To make predictions for Premier League matches and to determine what would happen in the 2018–19 season using Poisson regression, we fitted two models to get the scoring rates for every EPL team, 1) at home, and 2) away from home. Here we are interested in evaluating the model equation at different values of the explanatory variables. Since the link function for Poisson regression is the natural log function, we would back-transform the equation with the corresponding exponential function. This will then give us the home and away mean (expected) scoring rates for every EPL club, aggregated across all opponents.

After that, we executed a large number of simulations, to get the hypothetical 2018–19 season results and then analyzed and compared what we got for each of the three subsets of season mentioned in the previous section. For each subset of data, we performed 10000 simulations, and this was accomplished by randomly generating the match final score for every team matchup, using the clubs' average scoring rates that we got from fitting the Poisson regression models, which returns a random integer for each team's number of goals scored. In addition, the number of points for every match outcome based on the teams' number of goals scored were also calculated (see Table 1), as a side gets 3 points if they score more than their opponent, 1 point if the final score is a tie, and 0 points if the opposing roster has more goals. For each simulated season (out of 10000 total for each method), we tallied up the points, calculated the goal differentials, and obtained the final standings for EPL clubs (see Table 2). From this information, we kept track of various metrics for EPL clubs and utilized them to evaluate and compare the models and their predictions, which will be discussed in the next section.

Table 1. Simulation table of 2018–19 EPL matches. The two columns *HomeScore* and *AwayScore* indicate the number of goals scored by the home and away clubs, randomly generated from their average scoring rates. The points each team receives versus their opponent (3 for a win, 1 for a draw, 0 for a loss) are determined from the match final score.

HomeTeam	HomeRate	AwayTeam	AwayRate	HomeScore	AwayScore	HomePoints	AwayPoints
Newcastle	1.667	Arsenal	1.510	1	3	0	3
Bournemouth	1.474	Southampton	0.989	0	0	1	1
West Ham	1.429	Brighton	0.526	1	0	3	0
Fulham	1.413	Tottenham	1.256	2	3	0	3
Cardiff	1.053	Leicester	1.148	0	2	0	3

Table 2. Final standings for a simulated 2018–19 season. The team ranks are arranged by total points, followed by goal differential (goal scored minus goal conceded).

Rank	Team	Played	Points	GD
1	Man United	38	77	30
2	Chelsea	38	73	27
3	Liverpool	38	69	20
4	Arsenal	38	68	26
5	Leicester	38	62	9
6	Man City	38	61	10
7	Newcastle	38	56	5
8	Everton	38	56	2
9	Bournemouth	38	55	-1
10	Southampton	38	53	0
11	West Ham	38	53	-3
12	Tottenham	38	51	-2
13	Wolves	38	48	-5
14	Burnley	38	48	-6
15	Cardiff	38	47	-10
16	Fulham	38	40	-9
17	Huddersfield	38	39	-33
18	Watford	38	38	-14
19	Crystal Palace	38	36	-19
20	Brighton	38	24	-27

Table 3. Descriptive statistics of Manchester United's number of goals scored.

min	Q1	median	Q3	max	mean	sd	n
0	1	2	3	9	1.916	1.405	1038

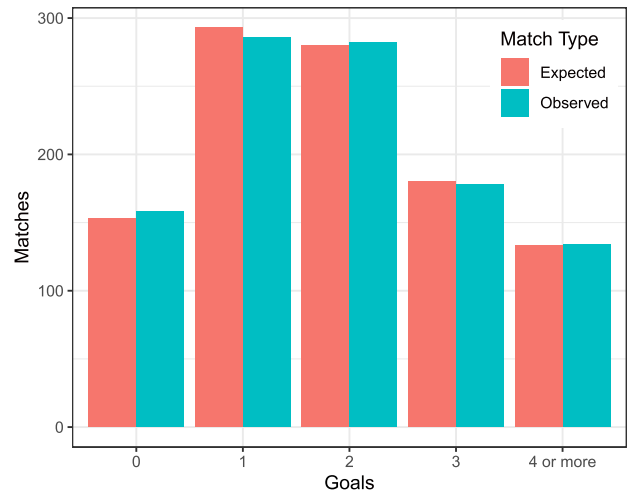


Figure 2: Histogram of Manchester United's goals scored.

4. RESULTS

4.1 Goal Scoring and the Poisson Process

For our first analysis of the relationship between the number of goals scored and the Poisson distribution, we used Manchester United (MU) as our case of inspection. Our question here was “Does MU’s distribution of number of goals scored follow a Poisson distribution?” Table 3 and Figure 2 are numerical and visual summaries of Manchester United number of goals scored in every EPL season until the final fixture of their 2018–19 campaign. To examine MU’s number of goals scored closer, we divided the goals variable into four levels: 0, 1, 2, 3, and 4 or more goals; and also tally up the number of matches in real life from 1992 to 2018 having those goal values. From there, we obtained the probability of Manchester United scoring each individual goal value, using the probability density function of a Poisson distribution with mean $\lambda = 1.916$ (see Table 3);

and then multiplied the probabilities with the total number of matches of $n = 1038$ (see Table 3), to get the expected number of matches for each number of goals category. The counts for the observed and expected number of matches are very close to one another, as illustrated by Table 4 and Figure 3, which is a significant validation for us. The results from a Chi-square goodness-of-fit test ($\chi^2 = 0.381$, p -value = 0.984) further confirm that there is no significant difference between the data’s and the empirical distribution, hence the number of goals scored by Manchester United is consistent with a Poisson distribution.

We were also interested in verifying the connections between the time between goals in a season and the exponential distribution, and the re-scaled goal scoring minutes in a match and the standard uniform distribution. We continued to use Manchester United to investigate these topics and explore their goal scoring time data described in Section 2. We first plotted the cumulative distribution curve of our data

Table 4. Observed and expected frequencies of the number of matches for each goal value, alongside their Poisson probabilities.

Goals	Probability	Observed	Expected
0	0.147	158	153
1	0.282	286	293
2	0.270	282	280
3	0.173	178	180
4 or more	0.128	134	133

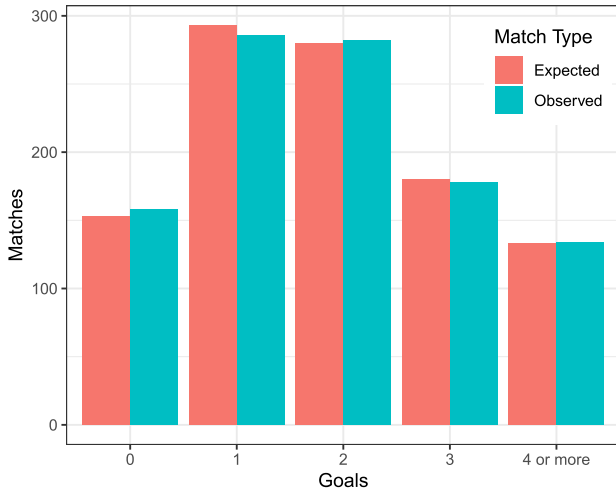


Figure 3: Side-by-side bar graph comparing the observed and expected matches.

and compare them with the reference distributions (exponential and uniform), as shown in Figures 4 and 5. As we can see, the differences between the curves in each figure are marginal, which is a validation and gives us more reason to believe in our claim that our goal scoring data are consistent with the specified distributions. After that, we conducted a Kolmogorov-Smirnov goodness-of-fit test to confirm the fit between our data and the hypothetical distributions – exponential and uniform. Based on the test statistics and p -values ($D = 0.089$, p -value = 0.679 for Time between goals vs. Exponential and $D = 0.085$, p -value = 0.731 for Time location of goals vs. Uniform), there is insufficient evidence to support a conclusion that our data for the time between goals and minutes of scoring are not consistent with the exponential and uniform distributions. Therefore, the time between goals and re-scaled scoring time for Manchester United in the 2018–19 EPL season are exponentially and uniformly distributed, respectively.

4.2 Simulating and Predicting Season Outcomes Using Poisson Regression

In this section, we discuss the results of our Poisson regression models and simulations described in Section 3. After fitting the models and conducting the simulations,

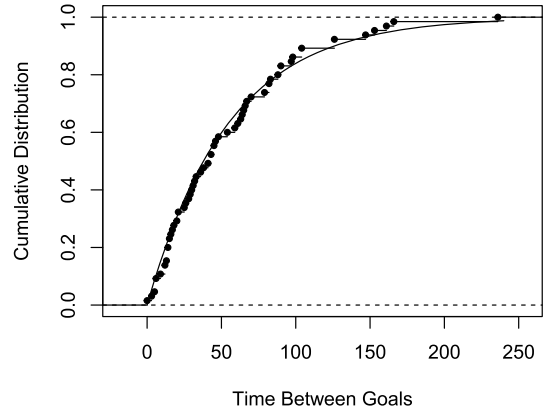


Figure 4: Cumulative distribution curves of time between goals and the exponential distribution.

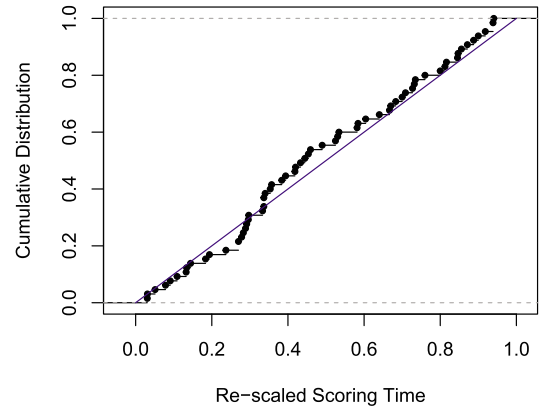


Figure 5: Cumulative distribution curves of the re-scaled goal scoring minutes and the standard uniform distribution.

there are various analyses we could do, from comparing how teams’ chances of winning differ across the three subsets of data used for simulation, to analyzing which method gives us the most consistent predictions compared to the actual outcomes. In this paper, we focus on the title race and the relegation battle in the Premier League, namely, the chances of winning the league title and getting relegated for EPL teams.

We first look at the chances of finishing first in the EPL table at the end of the 2018–19 season for the “Big 6” in English football, which includes Manchester United, Liverpool, Arsenal, Chelsea, Manchester City, and Tottenham (see Table 5). Overall, it is clear that MU’s chance of having higher ranks decreases drastically if we prioritize recent data over just using data from all seasons. This is consistent with the club’s situation in real life, as MU has not been doing well lately, but was winning a lot prior to the 2010s. Their in-town rivals, Manchester City, on the other hand, have much higher percentages of winning the league in 2018–19 if we focus on data from recent years. In fact, they were the champions of the 2018–19 EPL season. Similar to Manchester

Table 5. Chances of winning the 2018–19 Premier League title for the Big 6.

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

Table 6. Chances of getting relegated after the 2018–19 season for Premier League teams.

Team	All Seasons	2010s	Assign Weight
Huddersfield	69.15	72.91	71.03
Cardiff	49.99	54.38	53.47
Brighton	41.65	45.03	44.03
Burnley	31.60	43.38	33.34
Watford	26.28	17.24	15.82
Wolves	22.04	14.22	22.32
Crystal Palace	17.54	14.42	10.65
Fulham	9.87	7.91	11.74
West Ham	7.46	6.73	6.39
Southampton	6.49	5.23	11.65
Leicester	5.53	1.75	2.39
Bournemouth	4.76	5.46	5.80
Everton	3.48	2.17	3.63
Newcastle	2.40	8.78	7.41
Tottenham	1.02	0.16	0.08
Man City	0.38	0	0
Chelsea	0.15	0.09	0.13
Arsenal	0.10	0.04	0.06
Liverpool	0.10	0.05	0.02
Man United	0.01	0.05	0.04

United, our results for Manchester City also make sense, due to the fact that City has emerged into an EPL title contender over the last decade, but they were just an average team dating back to the old 1990s and 2000s days. It is also notable that both assigning weight to more recent years and using only data from the 2010s lower Chelsea and Arsenal’s likelihood of finishing first, whereas these two methods increase Tottenham’s chance of winning the league, though their winning chance is still the smallest in each category.

Next, we investigate the likelihood of getting relegated after the 2018–19 season for EPL clubs. The relegation zone, or the last three places in the final rankings, is where no teams in the Premier League want to end up at the end of the season, because after each season, the bottom three clubs get relegated to the second highest division of English football. The results from our three subsets of season data (see Table 6) are pretty consistent with each other, with Huddersfield and Cardiff essentially being the two “locks” to play in the English football division below the EPL in the following season. Brighton and Burnley also have high

Table 7. 40-point safety rule comparison between the three subsets of season data. We tallied up the total number of teams as well as distinct simulated seasons with teams being relegated while having at least 40 points for each simulation method.

Subset	Seasons	Teams
All Seasons	3434	4424
2010s	2066	2470
Assign Weight	2346	2889

chances of being in the bottom three, although in reality, both of these teams successfully remained in the league for another year. The third team that got dismissed in 2019, Fulham, does not have high chances of relegation in any of the three simulation methods. Unsurprisingly, Big 6 teams have the smallest chances of getting related from the EPL at the completion point of the 2018–19 season.

On a related note, the 40-point safety rule [2] is an interesting myth associated with the EPL’s relegation zone. Since the league’s reduction in the number of teams to twenty clubs before the start of the 1995–96 season, there have been only three times that a squad got relegated despite hitting the 40-point mark. They are West Ham in 2002–03 with 42 points, and Bolton and Sunderland both with 40 points at the end of 1997–98 and 1996–97 seasons respectively. This mythical 40-point mark has been crucial for the relegation battle for many years, as subpar teams often view getting there as their “security blanket” for remaining in the top division of English football. From Table 7, using data from 2010s and assigning weight to recent years give us significantly less number of both teams and seasons that violate the 40-point rule than using data from every season prior to 2018–19. Therefore, the 40-point safety rule seems to hold much better for the two simulations that focus on recent data than the one with data from all previous seasons. This actually makes sense, since in the past two decades in reality, teams with 40 or more points at the end of EPL seasons all survived from relegation, as the last time this rule did not happen was the 1997–98 season.

5. CONCLUSION AND DISCUSSION

Overall, we have found that Premier League goal scoring fits the characteristics of a Poisson process. Our first result was that a Poisson distribution can be used to predict the number of matches with each number of goals scored. Additionally, the time between each individual goal in a season can be described by an exponential distribution. We also have evidence that the normalized goal scoring time positions after are uniformly distributed. We also used different sets of data prior to the 2018–19 Premier League season, namely, data from all seasons before, data from only the 2010s, and data from all previous years but assigning more weight to recent competitions, to predict what would happen in the 2018–19 season. We got each team’s goal scoring

rate at home and away from home by doing Poisson regression, and then performed simulations using those rate parameters. Different team metrics like how many points each team got and what place each team finished were being kept track of from the simulations, and then we make use of those variables to analyze and compare our models of different season data subsets.

In the future, there are additional topics we could explore, including:

1. Besides the number of goals scored, there are many other factors that can be used to determine outcomes of football matches. In future research, we could use various factors to predict goal scoring and find out if they will be as helpful as using just number of goals. We could look into variables that are likely to contribute to the outcomes of Premier League football matches such as clean sheets, possession time, pass accuracy, shots on target, and numerous other soccer statistics. On top of that, we could compare different models with different predictors and evaluate them to find out which set of variables best predicts league outcomes, and then use them to simulate and predict match results.

2. In football and many other sports, team performance tends to vary throughout a season and across seasons. Some Premier League teams have the tendency of getting hot in early months, some clubs reach their peak during the middle period of the season, and a few others are more likely to do better at the season's home stretch. Winning and losing streaks are also important factors in sports, as some clubs are streaky, while others tend to be more consistent. Thus, in future research, we could apply match results of EPL teams from past games within the season, and maybe find a way to emphasize winning and losing streaks, to predict the outcome of later matches. As a follow up, we could investigate a model's performance throughout the season. Some models may work better and predict more accurate results at certain times in the year than others.

3. In addition to predicting match results, another popular application of statistical modeling in sports analytics is determining betting odds. We could use the probabilities from our Poisson regression models and simulations to calculate the odds of possible game outcomes for different team matchups. We could also look into and compare different types of bets such as over and under, money line wager, or point spread; determining if it is a good idea to bet on a match, and if so, how much profit we could win.

ACKNOWLEDGMENTS

This work was completed as the author's senior honors thesis, in partial fulfillment of the requirements for earning Departmental Honors in Mathematics at Wittenberg University in Springfield, Ohio. The author would like to express his special gratitude and thanks to his advisor, Professor Douglas M. Andrews, for his many ideas, suggestions and guidance throughout the research process. The author would also like to thank the Department of Mathematics and Computer Science at Wittenberg University for providing him the valuable knowledge during his undergraduate career, and for giving him the opportunity to participate in the Departmental Honors Program.

SUPPLEMENTARY MATERIAL

All of the materials related to this research are available on GitHub at <https://github.com/qntkhvn/eplgoals>.

Accepted 12 July 2021

REFERENCES

- [1] ASIMOW, L. A. and MAXWELL, M. M. (2015) *Probability and Statistics with Applications: A Problem Solving Text*, 2nd ed. AC-TEX Publications. MR0044951
- [2] DAVIS, M. (2018). *Do Premier League teams really need 40 points to avoid relegation?* <https://www.bbc.com/sport/football/43049564>.
- [3] DOBROW, R. P. (2016) *Introduction to Stochastic Processes with R*. Wiley. <https://doi.org/10.1002/9781118740712>. MR3495869
- [4] FOOTBALL-DATA (2020). *Football-Data.co.uk*. <https://www.football-data.co.uk/englandm.php>.
- [5] FOX, J. (2015) *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications.
- [6] JOHNSON, B. (2015). *Association Football or Soccer*. <https://www.historic-uk.com/CultureUK/Association-Football-or-Soccer>.
- [7] NGUYEN, Q. <https://github.com/qntkhvn/eplgoals>.
- [8] PREMIER LEAGUE (2020). *Premier League explained*. <https://www.premierleague.com/premier-league-explained>.
- [9] WIKIPEDIA CONTRIBUTORS (2019). *2018–19 Manchester United F.C. season*. https://en.wikipedia.org/wiki/2018-19_Manchester_United_F.C._season.

Quang Nguyen. Department of Mathematics and Statistics, Loyola University Chicago, Chicago, Illinois, USA.
E-mail address: nnguyen22@luc.edu