Emirates FA Cup Giant-Killings

The 10 Least Probable Scorelines

January 2022

This report has been commissioned by the Football Association

Report compiled by Dr. Adwaye Rambojun & Prof. Andreas Kyprianou

Disclaimer: We believe that for the purposes of exploratory data analysis in the context of advertising campaigns and public engagement, the model we proposed in this report is appropriate. We advise against using the model for the purposes of modern match predictions.

CONTENTS

1 Introduction

The Emirates FA Cup is the oldest domestic knockout football competition in the world, and men's teams across 10 levels of the English football pyramid are eligible to compete on an annual basis.

The purpose of this report is to give a technical explanation of how the '10 least probable scorelines' over a select number of seasons were identified following consultation with The FA and using a bespoke mathematical model. One of its key objectives is to ensure reproducibility of results.

In the competition's modern format, there are 12 randomly drawn knockout rounds followed by its Semi Finals and Final. Six of these rounds are classified as preliminaries and qualifiers, where teams from levels 8 (highlighted in Table 1) and below of the English football pyramid play for the chance to progress to the First Round Proper. There is also a system of exemptions through which teams from higher-ranked leagues join at later stages of the competition.

The unique aspect of the Emirates FA Cup is that it is genuinely possible for a lower league team to beat teams from higher or even the highest leagues due to its format. This regularly happens in Third Round, when teams from the top two leagues in England enter the competition (also highlighted in Table 1). The Third Round has seen many unusual pairings over the years where players from lower leagues get the chance to play against their idols from higher leagues.

In this report, we analyse 24 famous Third Round games where lower league teams beat those from the top two divisions. We use a Poisson model to compute the probability of the scoreline observed, taking into account the relative strengths of each team as well as the timeline of the goals scored. The parameters of our model are estimated from match data spanning the period of 1959-60 to 2009-10 (8152 matches). All Third Round match probabilities have been calculated using the modern format of the Emirates FA Cup competition and, for consistency of the mathematical modelling, based on equivalent league levels of clubs within the current English football league system.

Round	New entrants	Nr. matches
Qualifying		
Extra Preliminary Round	Level 9 and 10 teams	184
Preliminary Round	Level 8 clubs	160
First Round Qualifying	Level 7 clubs	116
Second Round Qualifying	Level 6 clubs	80
Third Round Qualifying	None	40
Fourth Round Qualifying	Level 5 clubs	32
Competition Proper		
First Round	Level 3 and 4 clubs	40
Second Round	None	20
Third Round	Level 1 and 2 clubs	32
Fourth Round	None	16
Fifth Round	None	8
Quarter-finals	None	4
Semi-finals	None	2
Final	None	1

Table 1: The table shows the format of the Emirates FA Cup . This was taken from The FA web pages. We are interested in the Third Round, when so called Giant Killings occur.

We present our main findings in Section 2. This includes a list of the 10 most unlikely results bench marked against everyday life events of approximately the same odds in Section 2.2, and a set of equations describing our probability model that are fit for use across The FA's corresponding campaign in Section 2.3. In Section

2.4, we investigate how the model predicts the different probabilities, and an in-depth mathematical description of how the model is provided in Section 3.

2 Main Findings

In this section we present the main deliverables of this project, which is a list of most unlikely results and how the probability of this happening compares with some benign events. The probabilities presented here have been obtained from the model that we describe in Section 3. The driving equations are shown in Section 2.3.

2.1 The Ten Least Probable Scorelines

We find that the following 10 matches have the lowest probability of occurring. The top two matches in this list involve teams from league level 7, based on the modern league structure, which both have a probability of 0.0065 of qualifying for Third Round. This low probability along with the fact that goals were scored in quick succession are what drive the rarity of these two events.

SEASON	Home Team (League)	Away Team (League)	SCORE SEQUENCE	One in Chance
1990-1991	Second Division (2)	Isthmian League (7)	**WBA 2-4 Woking** WBA [34], Woking [59], [65], [74], [89], WBA [90]	15,959,312
1971-1972	Southern League (7)	First Division (1)	**Hereford 2-1 Newcastle** Newcastle [82], Hereford [85], [103*] *Extra Time	32,449
2010-2011	League Two (4)	Premier League (1)	**Stevenage 3-1 Newcastle** Stevenage [50], [55], Newcastle, [92], Stevenage [93]	7,712
1985-1986	First Division (1)	Alliance Premier League (5)	**Birmingham City 1-2 Altrincham** Birmingham City [63], Altrincham [65], [75]	4,376
2015-2016	League Two (4)	Premier League (1)	**Oxford Utd 3-2 Swansea** Swansea [23], Oxford [45], [49], [59], Swansea [66]	3,487
1988-1989	Conference (5)	First Division (1)	**Sutton Utd 2-1 Coventry** Sutton [42], Coventry [52], Sutton [59]	3,260
1974-1975	First Division (1)	Southern League (7)	**Burnley 0-1 Wimbledon** Wimbledon [49]	2,515
1979-1980	Isthmian League (7)	Second Division (2)	**Harlow Town 1-0 Leicester City** Harlow [42]	1,800
2001-2002	Premier League (1)	League Two (4)	**Derby County 1-3 Bristol Rovers** Bristol Rovers [14], [40], [62], Derby [88]	397
2018-2019	League Two (4)	Premier League (1)	**Newport 2-1 Leicester City** Newport [10], Leicester [82], Newport [85]	337

2.2 Comparison to Everyday Life Event

The table below compares the probability of a benign real life event of a similar probability to the observed scorelines.

Match	Odds 1 in	Everyday Life Events	Odds 1 in
WBA 2-4 Woking	15,959,312	Conceiving identical quadruplets	15,000,000
Hereford 2-1 Newcastle	32,449	Chance of growing to over seven foot tall	26,315
Stevenage 3-1 Newcastle	7,712	Rolling five consecutive sixes with a dice	7,776
Birmingham City 1-2 Altrincham	4,376	Being dealt a four-of-a-kind poker hand	4,165
Oxford United 3-2 Swansea	3,487	Scoring a hat-trick in the Final and winning the Emirates FA Cup	2993
Sutton United 2-1 Coventry	3,260	Chance of the Emirates FA Cup trophy landing in the UK if dropped from space	2103
Burnley 0-1 Wimbledon	2,515	Becoming a NASA astronaut	1,525
Harlow Town 1-0 Leicester City	1,800	Being born on a leap day	1,461
Derby County 1-3 Bristol Rovers	397	Coin tossing 8 heads-in-a-row	256
Newport 2-1 Leicester City	337	Conceiving identical twins	250

2.3 Equations

The numbers we have computed come from the full equation below:

$$p_{\text{scoreline}} = 0.5^{N(\ell_1)+N(\ell_2)} p_{1,2}^{(\ell_1)} p_{2,3}^{(\ell_1)} p_{1,2}^{(\ell_2)} p_{2,3}^{(\ell_2)} \prod_{i=1}^{g_2+1} \frac{\left(\frac{\lambda_1(i)}{t_i}\right)^{k_i}}{k_i!} e^{-\frac{\lambda_1(i)}{t_i}} \prod_{j=1}^{g_1+1} \frac{\left(\frac{\lambda_2(j)}{u_j}\right)^{m_j}}{m_j!} e^{-\frac{\lambda_2(j)}{u_j}}, \quad (1)$$

which can be simplified to

$$p_{\text{scoreline}} = 0.5^{N(\ell_1)+N(\ell_2)} p_{1,2}^{(\ell_1)} p_{2,3}^{(\ell_1)} p_{1,2}^{(\ell_2)} p_{2,3}^{(\ell_2)} \prod_{i=1}^{g_2+1} \text{Pois}\left(k_i; \frac{\lambda_1(i)}{t_i}\right) \prod_{j=1}^{g_1+1} \text{Pois}\left(m_j; \frac{\lambda_2(j)}{u_j}\right). \quad (2)$$

However, it might be a good idea to show the equation for the probability of the team from the lower league scoring the number of goals they have, which is given by

$$p = 0.5^N p_{1,2} p_{2,3} \prod_{i=1}^{g+1} \frac{\left(\frac{\lambda(i)}{t_i}\right)^{k_i}}{k_i!} e^{-\frac{\lambda(i)}{t_i}}$$

where we have removed a lot of the subscripts and superscripts for simplicity. This can be further simplified to

$$p = 0.5^N p_{1,2} p_{2,3} \prod_{i=1}^{g+1} \text{Pois}\left(k_i; \frac{\lambda(i)}{t_i}\right)$$

2.4 Features of the Model

Let us discuss some of the key features of the model we have built, how it factors in various aspects of the game and where the role of historical data comes in to estimate the respective probabilities of the top 10 least likely scorelines.

Reaching the Third Round vs in-match performance: The model takes into account both the number of games each team has to play before reaching the Third Round of the Emirates FA Cup as well as what happens during the knockout match. On inspecting the formula (1), the first half of the formula, i.e.

$$0.5^{N(\ell_1)+N(\ell_2)} p_{1,2}^{(\ell_1)} p_{2,3}^{(\ell_1)} p_{1,2}^{(\ell_2)} p_{2,3}^{(\ell_2)},$$

deals with the probability of reaching the Third Round, where as the second half, i.e.

$$\prod_{i=1}^{g_2+1} \frac{\left(\frac{\lambda_1(i)}{t_i}\right)^{k_i}}{k_i!} e^{-\frac{\lambda_1(i)}{t_i}} \prod_{j=1}^{g_1+1} \frac{\left(\frac{\lambda_2(j)}{u_j}\right)^{m_j}}{m_j!} e^{-\frac{\lambda_2(j)}{u_j}},$$

deals with the likelihood of observing the particular outcome during the match.

The in-game model looks at the goals scored by each team and how this impacts the goal scoring rate of their opponent. The more goals a team has conceded, historical data suggests the less goals it is likely to score. This is controlled by the parameters $\lambda_1(i)$ and $\lambda_2(j)$, which are drawn from Table 3, where one can see that, in general, goal scoring rates decreases with increasing number of goals conceded. Our model also accounts for the clustering of goals in short time periods. The shorter the time period in which a number of goals is scored, the more unlikely such an event will be.

The influence of a team's current league: In the data analysis on matches from 1959-1960 to 2009-2010 seasons, it was revealed that teams from lower leagues have a lower goal scoring rate than teams from higher leagues. This leads to a lower probability of a win by a lower league team over a higher league team. This phenomenon is built into the parameters $\lambda_1(i)$ and $\lambda_2(j)$; see again Table 3. These parameters also take an averaged view of whether a game is played at home or away.

A team's current league also affects the number of games that it has to play before reaching the Third Round. Teams from lower leagues have to play more games, increasing the chance of being defeated before reaching the Third Round, and thus reducing the probability of them beating a team from a higher league.

The influence of goal scoring profiles: Our model divides the game into epochs in which a team has to score. These are periods of time between successive goals. For example, in the case of only one goal being conceded, the first epoch would be the start of the game to the time the goal was conceded and the second epoch would be the time that the goal was conceded to the end of the game. If no goals were conceded, then there is just one epoch. Each epoch would be characterised by one goal scoring rate for each team, which, on average, goes down with the epoch number. If a team scores many goals in a short epoch, then this will drive down the probability due to the nature of the probabilistic model that we are using.

The role of historical data: The principal role of all the historical data provided is to use it to calibrate the model described by (1). That is to say, the data can be used to estimate the relevant parameters e.g. $\lambda_1(i)$, $p_{1,2}^{(\ell_1)}$ and so on, in formula (1).

3 Detailed Technical Model Description

In order that our computations lead to a meaningful comparison of odds, all Third Round match probabilities have been calculated on the basis of the outcome occurring under current format of the Emirates FA Cup. Moreover, the league positioning of teams has been mapped to equivalent league levels within the current English football league system.

We want to compute the probability that a team from a lower league beats a team from a higher league in the Third Round by a certain scoreline. Here by scoreline, we mean the number of goals and when these were distributed within the game.

To get the total probability of the event in question, we use Bayes Formula for conditional probability for two events A and B as follows

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \tag{3}$$

For our purposes, A represents the observed scoreline between two teams facing off in a match and B represents the event that the match is played in Third Round. Hence, in the above, $P(A|B)$ is the probability of observing this scoreline in the Third Round and $\mathbb{P}(B)$ is the probability of both teams entering the Third Round. For simplicity, we assume that once they enter the Third Round, the configuration of the games (who plays who) is pre-determined, i.e. we do not take account of random draws. We do this because what we want to model is the progression of the weak team based on their strength and not based on luck. As such we develop two probability models as described in the next two sections, one for $\mathbb{P}(A|B)$ and one for $\mathbb{P}(B)$.

3.1 Third Round Qualification Model

The probability that a team from league ℓ qualifying for the First Round is given by $0.5^{N(\ell)}$ where

$$N(\ell) = \begin{cases} 0 & \text{if} \quad \ell < 5, \\ 1 & \text{if} \quad \ell = 5, \\ 3 & \text{if} \quad \ell = 6, \\ 4 & \text{if} \quad \ell = 7, \\ 5 & \text{if} \quad \ell = 8, \\ 6 & \text{if} \quad \ell \geq 9. \end{cases} \tag{4}$$

Once a team has qualified for the First Round of the Emirates FA Cup, they advance to the Second Round with probability $p_{1,2}$, and once they are in the Second Round, they advance to Third Round with probability $p_{2,3}$.

Note that, due to lack of data from the Preliminary Rounds, we assumed that it is equally likely for any team playing in the Preliminary Round to get qualified and disqualified. We then build the above function based on Table 1.

3.2 In-Match Score Model

Consider two teams playing each other in the Third Round of the Emirates FA Cup coming from leagues ℓ_1 and ℓ_2 respectively with $\ell_i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ for $i = 1, 2$. In this formulation, the league is higher when ℓ_i is small. Team 1 scores g_1 goals at times given by $[s_1, s_2, ..., s_{g_1}]$ and the times Team 2 scores g_2 goals at times given by $[r_1, r_2, ..., r_{g_2}]$.

We assume the number of goals scored follows a conditional Poisson distribution. Consider the goals scored by Team 1. We assume that the rate at which they score depends on the number of goals they have conceded so far and is given by $\lambda_1(n)$ where n is the number of goals they have conceded so far. Hence, we can split the match into intervals of lengths $[t_1^*, t_2^*, ..., t_{g_2+1}^*]$

$$t_i = \begin{cases} s_{i+1} - s_i & \text{if } 1 < i < g_2, \\ s_1 & \text{if } i = 1, \\ T - s_i & \text{if } i = g_2, \end{cases} \tag{5}$$

where T is the duration of the game. The probability that Team 1 scores at the times $[s_1, s_2, ..., s_{g_1}]$ is given by

$$p^{(1)}_{score} = \prod_{i=1}^{g_2+1} \frac{\left(\frac{\lambda_1(i)}{t_i}\right)^{k_i}}{k_i!} e^{-\frac{\lambda_1(i)}{t_i}}, \tag{6}$$

where k_i is the number of goals scores by Team 1 in the i-th time interval and where we define the inverse time interval lengths as

$$t_i = \frac{T}{t_i^*}.$$

Note that this probability implicitly builds in the difference in strengths between two teams by the conditional match goal rates $\lambda(i)$. The data shows that this is a valid way to capture the relationship between two teams as shown in Figure 1.

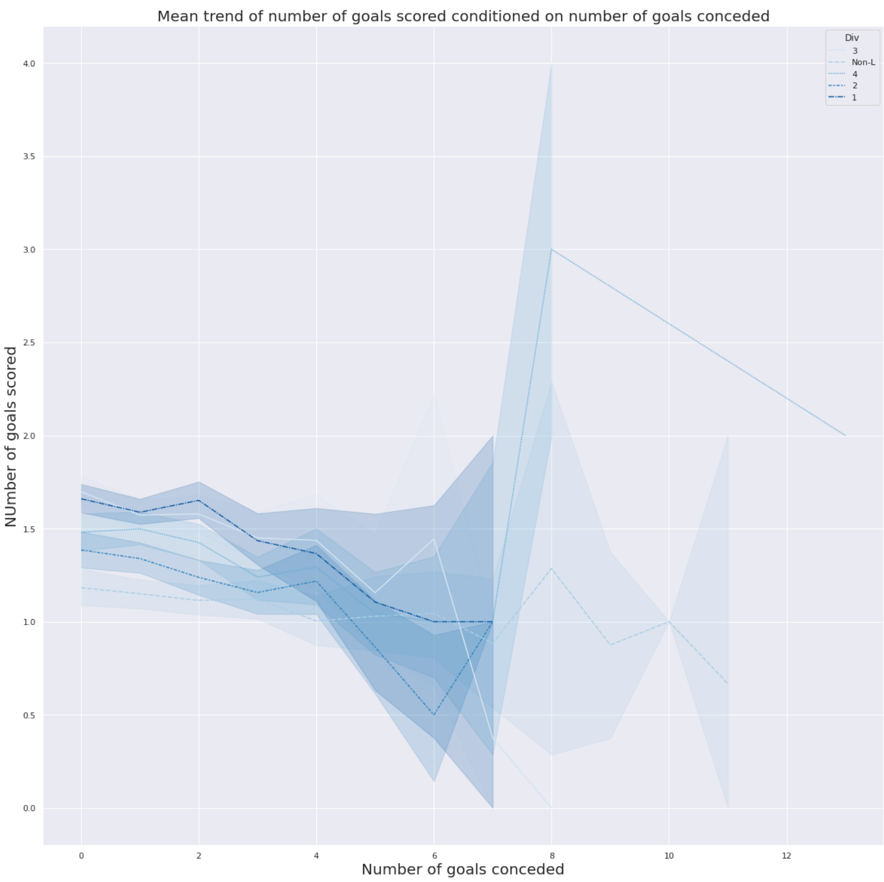

Figure 1: The Figure shows the mean number of goals scored by teams against the number of goals conceded. The lines show the population mean along with the standard deviation.

3.3 Full Model

Let $p^{(\ell)}_{1,2}$ and $p^{(\ell)}_{2,3}$ be the probabilities that a team playing in league ℓ qualifies from the First Round to the Second Round and from the Second Round to the Third Round respectively. By independence, we have that

$$\mathbb{P}(B) = 0.5^{N(\ell_1)+N(\ell_2)} p^{(\ell_1)}_{1,2} p^{(\ell_1)}_{2,3} p^{(\ell_2)}_{1,2} p^{(\ell_2)}_{2,3}, \tag{7}$$

which is the product of the probability of Team 1 qualifying for Third Round of the Emirates FA Cup and the probability of Team 2 qualifying for the Emirates FA Cup.

By symmetry and conditional independence, we have that

$$\mathbb{P}(A|B) = \prod_{i=1}^{g_2+1} \frac{\left(\frac{\lambda_1(i)}{t_i}\right)^{k_i}}{k_i!} e^{-\frac{\lambda_1(i)}{t_i}} \prod_{j=1}^{g_1+1} \frac{\left(\frac{\lambda_2(j)}{u_j}\right)^{m_j}}{m_j!} e^{-\frac{\lambda_1(j)}{u_j}}, \tag{8}$$

where, for team 2, u_i and m_i play the same role as t_i and k_i do for Team 1.

Hence, for the match in question, the probability of Team 1 scoring at times $[s_1, s_2, ..., s_{g_1}]$, when Team 2 scores at times $[r_1, r_2, ..., r_{g_2}]$ in Third Round of the Emirates FA Cup , is given by

$$\mathbb{P}(A \cap B) = 0.5^{N(\ell_1)+N(\ell_2)} p_{1,2}^{(\ell_1)} p_{2,3}^{(\ell_1)} p_{1,2}^{(\ell_2)} p_{2,3}^{(\ell_2)} \prod_{i=1}^{g_2+1} \frac{\left(\frac{\lambda_1(i)}{t_i}\right)^{k_i}}{k_i!} e^{-\frac{\lambda_1(i)}{t_i}} \prod_{j=1}^{g_1+1} \frac{\left(\frac{\lambda_2(j)}{u_j}\right)^{m_j}}{m_j!} e^{-\frac{\lambda_1(j)}{u_j}}. \tag{9}$$

3.4 Parameter Estimation

Data was provided by Opta, who are the data partner of the Football Association. The data covers matches from all Rounds of the Emirates FA Cup for the period 1959-1960 to 2009-2010. The data contains the final scores of all of the games, the league that the team played in each season covered, whether the teams playing each match qualified for the next round and the home and away status of each team.

Equation (8) is parameterised by the probabilities of Team 1 and 2 qualifying for the Second Round from the Third Round and for the Third Round from the Second Round given by $p_{1,2}^{(\ell_1)}, p_{2,3}^{(\ell_1)}, p_{1,2}^{(\ell_2)}, p_{2,3}^{(\ell_2)}$, respectively; and the conditional goal scoring rate $\lambda(i)_1, \lambda(i)_2$ for teams 1 and 2 respectively.

For simplicity, consider $\ell_1 = \ell$, then we estimate $p_{1,2}^{(\ell)}$ by finding the historical proportion of teams from league l that qualify from the First Round to the Second Round. Similarly, we estimate $p_{1,2}^{(\ell)}$ by finding the historical proportion of teams from league l that qualify from the Second Round to the Third Round.

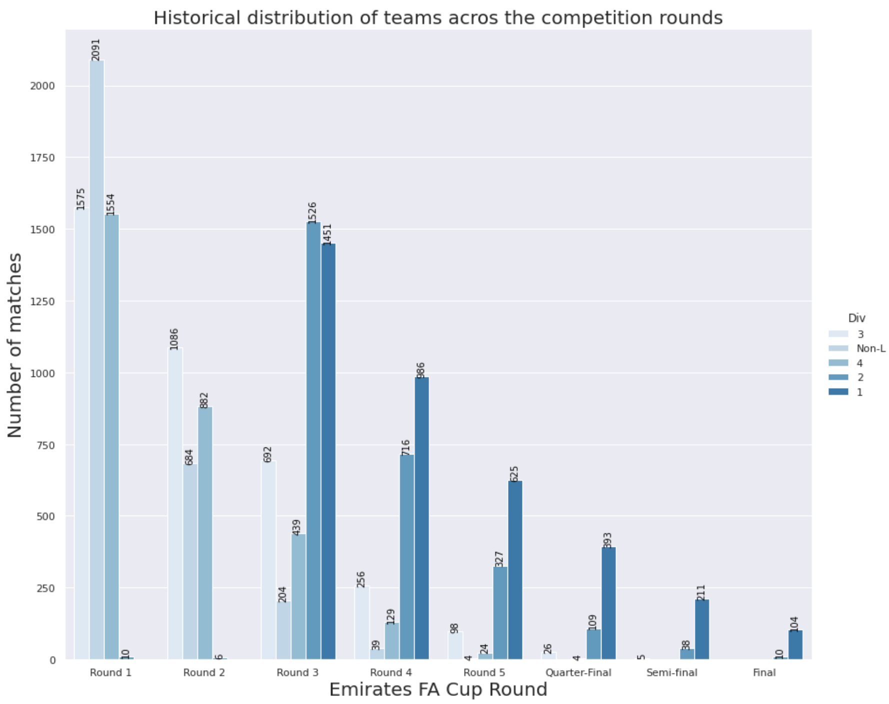

Figure 2: The Figure shows the number of teams that have played in each round of the Emirates FA Cup filtered by leagues since the 1959-1960 season.

For a team playing in League ℓ, we estimate $\lambda(i)$ by computing the Poisson rate of the distribution of goals scored by teams from league ℓ in a match where they have conceded i goals. The actual estimation uses binned regression on the histogram of goals.

Let y_k be the proportion of matches that has seen a teams from a certain league score k goals. Assuming that these goals come from a Poisson distribution of unkown rate λ, the predicted probability (proportion) of this happening is $z_k = \frac{\lambda^k}{k!} e^{-\lambda}$. We thus have a data-prediction pair $Y = [y_0, ..., y_{N-1}]$, $Z = [z_0, ..., z_{N-1}]$ on which we can perform non linear regression to find the parameter λ by minimising $\sum_{k=0}^{N-1} |y_k - z_k|^2$.

Goals Conceded i	League Level				
	1	2	3	4	Non-L
0	1.5592	1.3130	1.5412	1.3295	1.0475
1	1.5625	1.2691	1.4785	1.3949	1.0973
2	1.5807	1.1921	1.4099	1.3379	1.1041
3	1.3731	1.1152	1.3662	1.2413	1.0785
4	1.3710	1.2370	1.2108	1.2854	0.9828
5	0.9688	0.9329	1.1391	1.2141	1.0200
6	1.1559	0.4271	1.5661	1.1826	1.0655
7	1.0195	1.2240	0.2486	0.8645	0.9063
8	nan	nan	nan	3.1144	0.9480
9	nan	nan	nan	nan	1.0063
10	nan	nan	nan	nan	1.2240
11	nan	nan	nan	nan	0.2931
12	nan	nan	nan	2.1979	nan

Table 3: The Figure shows the consitional goal scoring Poisson rate $\lambda(i)$ across the different rates. For cases where this is no available (due to lack of data for example), we set it equal to 0.01

In binned regression, y_k and z_k are the observed and predicted proportion of observing an an interval of goals. As such, z_k would then be a sum of Poisson probability mass functions over the interval it represents.

Appendix

A Full List of Probabilities

All Third Round match probabilities have been calculated using the modern format of the Emirates FA Cup competition and, for consistency of the mathematical modelling, based on equivalent league levels of clubs within the current English football league system.

SEASON	Home Team League	Away Team League	SCORE SEQUENCE	One in Chance
1990-1991	Second Division (2)	Isthmian League (7)	**WBA 2-4 Woking** WBA [34], Woking [59], [65], [74], [89], WBA [90]	15,959,312
1971-1972	Southern League (7)	First Division (1)	**Hereford 2-1 Newcastle** Newcastle [82], Hereford [85], [103*] *Extra Time	32,449
2010-2011	League Two (4)	Premier League (1)	**Stevenage 3-1 Newcastle** Stevenage [50], [55], Newcastle, [92], Stevenage [93]	7,712
1985-1986	First Division (1)	Alliance Premier League (5)	**Birmingham City 1-2 Altrincham** Birmingham City [63], Altrincham [65], [75]	4,376
2015-2016	League Two (4)	Premier League (1)	**Oxford Utd 3-2 Swansea** Swansea [23], Oxford [45], [49], [59], Swansea [66]	3,487
1988-1989	Conference (5)	First Division (1)	**Sutton Utd 2-1 Coventry** Sutton [42], Coventry [52], Sutton [59]	3,260
1974-1975	First Division (1)	Southern League (7)	**Burnley 0-1 Wimbledon** Wimbledon [49]	2,515
1979-1980	Isthmian League (7)	Second Division (2)	**Harlow Town 1-0 Leicester City** Harlow [42]	1,800
2001-2002	Premier League (1)	League Two (4)	**Derby County 1-3 Bristol Rovers** Bristol Rovers [14], [40], [62], Derby [88]	397
2018-2019	League Two (4)	Premier League (1)	**Newport 2-1 Leicester City** Newport [10], Leicester [82], Newport [85]	337
2005-2006	Premier League (1)	League Two (4)	**Fulham 1-2 Leyton Orient** Leyton Orient [17], [44], Fulham [50]	308
2020-2021	League Two (4)	Premier League (1)	**Crawley Town 3-0 Leeds Utd** Crawley [50], [53], [70]	302
2018-2019	Premier League (1)	League Two (4)	**Fulham 1-2 Oldham Athletic** Fulham [52], Oldham [76], [88]	294

2011-2012	League Two (4)	Premier League (1)	**Swindon Town 2-1 Wigan Athletic** Wigan [35], Swindon [40], [76]	271
2018-2019	Championship (2)	National League (5)	**Sheffield Utd 0-1 Barnet** Barnet [21]	225
1988-1989	First Division (1)	Fourth Division (4)	**Middlesbrough 1-2 Grimsby** Middlesbrough [42], Grimsby [70], [88]	201
2017-2018	League Two (4)	Premier League (1)	**Coventry 2-1 Stoke City** Coventry [24], Stoke [54], Coventry [68]	198
2002-2003	League Two (4)	Premier League (1)	**Shrewsbury Town 2-1 Everton** Shrewsbury [37], Everton [60], Shrewsbury [88]	183
1991-1992	Fourth Division (4)	First Division (1)	**Wrexham 2-1 Arsenal** Arsenal [43], Wrexham [82], [84]	179
1983-1984	Third Division (3)	First Division (1)	**AFC Bournemouth 2-0 Manchester Utd** Bournemouth [60], [62]	71
1979-1980	Fourth Division (4)	First Division (1)	**Halifax Town 1-0 Manchester City** Halifax Town [75]	52
1998-1999	League Two (4)	Premier League (1)	**Swansea City 1-0 West Ham Utd** Swansea [29]	52
2008-2009	Premier League (1)	League One (3)	**Everton 0-1 Oldham Athletic** Oldham [45]	35
2009-2010	Premier League (1)	League One (3)	**Manchester Utd 0-1 Leeds Utd** Leeds Utd [19]	35

B Probabilities of Comparative Events

B.1 Dice Rolls

Assuming independence of rolls and that the dice is fair, the chance of getting n 6s is simply given by $1/6^n$ so that:

Number of consecutive 6s	4	5	6	7	8	9
Probability (one in chance)	1,296	7,776	46656	279,936	1,679,616	10,077,696

B.2 Coin Tosses

Assuming independence of tosses and fairness of a coin, the chance of getting n heads in a row is given by $1/2^n$

Number of consecutive heads	8	9	10	11	12
Probability (one in chance)	256	512	1,024	2,048	4,096

B.3 Poker Moves

The computation of the probabilities here requires us to know how to count the number of ways to choose 5 cards from a pack of 52. This uses the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

which roughly translates to the number of ways one can choose k items from a list of n distinct items.

The total number of possible hands that one can be dealt in poker equals

$$\binom{52}{5} = \frac{52!}{(52-5)!5!}$$

which is equivalent to the number of ways one can get choose 5 cards from a pack of 52 cards. The table below shows the number of ways one can get various poker hands. The probability of each of these hands is the frequency expression divided by the total number of possible hands.

Moves	Frequency Expression	one in chance	Brief Explanation
Royal Flush	$\binom{4}{1}$	649,740	Choose which suit
Straight Flush	$\binom{10}{1}\binom{4}{1} - \binom{4}{1}$	72,193	Choose a card out of 10 and choose the next 4 ones but remove the number of royal flushes
Four of a kind	$\binom{13}{1}\binom{4}{4}\binom{12}{1}\binom{4}{1}$	4,165	Choose a card and get the same one across 4 suits and choose any other card and choose a suit for it
Flush	$\binom{13}{5}\binom{4}{1} - \binom{10}{1}\binom{4}{1}$	509	Choose a suit and choose 5 cards from this suit, and subtract the number of straight flushes and royal flushes
Straight	$\binom{10}{1}\binom{4}{1}^5 - \binom{10}{1}\binom{4}{1}$	255	Choose a card from 10 and choose the next 4 cards in a row and remove the number of straight flushes

B.4 Twins, Triplets, Quadruplets, Quintuplets

Medical data is reasonably clear about the statistics for this. However, for identical siblings beyond twins, there is very little data available because it is such a rare event and hence opinion varies. The following are approximate but capture the order of magnitude with decreasing accuracy beyond twins:

Number of siblings	Odds, one in
Twins	60
Identical Twins	250
Non-identical triplets	10,000
Identical Triplets	100,000 (cursory)
Non-identical quadruplets	500,000 (cursory)
Identical quadruplets	15,000,000 (cursory)
Non-identical quintuplets	55,000,000 (cursory)

B.5 Human Height

It has been found that around 2800 people in the world are seven foot tall or taller. With a global population of 7.4 billion people, this means that the percentage of seven-footers is 0.000038% That's about 1 in 26,315.

B.6 Becoming an Astronaut

In 2017, NASA held an open application process to recruit new astronauts. It is well publicised that over 18,300 applicants applied from which 12 NASA astronauts were selected. This gives a 1 in 1,525 chance of becoming a NASA astronaut.

B.7 Born on the 29th of February

In a 4 year cycle, the extra leap day occurs just once. There are a total of 1461 days over the four year cycle, so the odds are 1 in 1461.

B.8 Winning the Emirates FA Final with a Hat-trick

Let X be the number of goals scored by the winning team and let Y be the number of goals scored by the losing team. We assume that X, Y both follow a poisson distribution with rate λ. Let X_1 be the number of goals scored by the hatrick scorer. By uniformity and independence, this is poisson with rate $\lambda_1 = \frac{\lambda}{10}$. Let $X_2 = X - X_1$ be the number of goals scored by the rest of the winning team. Then this is poisson with rate $\lambda_2 = \lambda - \frac{\lambda}{10}$.

The probability we are after is

$$\mathbb{P}\left(X_1 + X_2 > Y, X_1 \geq 3\right)$$

.
This is given by:

$$\begin{aligned}
\mathbb{P}\left(X_1 + X_2 > Y, X_1 \geq 3\right) &= \sum_{k=3}^{\infty} \mathbb{P}\left(Y < k + X_2 | X_1 = k\right) \mathbb{P}(X_1 = k) \\
&= \sum_{k=3}^{\infty} \mathbb{P}\left(Y < k + X_2\right) \mathbb{P}(X_1 = k) \quad \text{Assuming independence} \\
&= \sum_{k=3}^{\infty} \sum_{m=0}^{\infty} \mathbb{P}\left(Y < k + m | X_2 = m\right) \mathbb{P}(X_2 = m) \mathbb{P}(X_1 = k) \\
&= \sum_{k=3}^{\infty} \sum_{m=0}^{\infty} \mathbb{P}\left(Y < j | j < k + m\right) \mathbb{P}(X_2 = m) \mathbb{P}(X_1 = k) \\
&= \sum_{k=3}^{\infty} \sum_{m=0}^{\infty} \sum_{j=0}^{k+m-1} \mathbb{P}\left(Y = j\right) \mathbb{P}(X_2 = m) \mathbb{P}(X_1 = k)
\end{aligned} \tag{10}$$

Using the data provided by The FA, we estimated the goal scoring rate in the Final to be 1.3321. Estimating the above sum with a truncated one yields

$$\mathbb{P}\left(X_1 + X_2 > Y, X_1 \geq 3\right) = 0.0003341$$

which is roughly one in 2993 chance of happening.

B.9 Throwing Things From Space

The land area of of the UK is 242,495 km^2. The total surface area of the Earth is 510,000,000 km^2. Hence, the chance of throwing the Emirates FA Cup from the space station and it landing in the UK is

$$\frac{242,495}{510,000,000} \approx 0.000475$$

which is the same as 1 in 2103.