Simulation of national football leagues and determination of match importance

Tristan MARTIN and Joachim COLLIN - Supervisor : Julien GUYON

Main Objectives

1. **Simulate football matches** in order to be able to determine the most likely **end-of-season ranking** given the matchs played.
2. Create and implement a new indicator: **the importance of a match**, that is to say how important the outcome of a match is for a team with respect to an end-of-season objective.

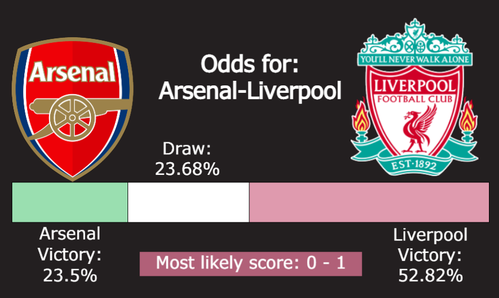
Model

- The number of goals scored by a team in a match is modeled by a **Poisson variable** in **the Dixon-Coles model** because, each time a team has the ball, it has the opportunity to attack and score. The probability that an attack will result in a goal is, of course, small, but the number of times a team has possession during a match is very large. If this probability is constant and attacks are independent, the number of goals will be Binomial and the Poisson approximation will apply very well.
- If a team i is playing at home against a team j and the observed score is (x_{ij}, y_{ij}), we shall assume that X_{ij} and Y_{ij} are independent modified Poisson variable as followed: $P(X_{i,j} = x, Y_{j,i} = y) = \tau_{\lambda,\mu}(x,y)\frac{e^{-\lambda}\lambda^x}{x!}\frac{e^{-\mu}\mu^y}{y!}$

$$\text{where } \tau_{\lambda,\mu}(x,y) = \begin{cases} 1-\lambda\mu\rho & \text{if } x=y=0 \\ 1+\mu\lambda & \text{if } x=0, y=1 \\ 1+\rho\mu & \text{if } x=0, y=1 \\ 1-\rho & \text{if } x=y=1 \\ 1 & \text{otherwise} \end{cases} \quad \text{and } \lambda = \alpha_i\beta_j k \ \mu = \alpha_j\beta_i$$

α_i and βi represent respectively the strength of the attack and the weakness of the defense of team i and k represents the advantage of hosting the match. The greater the attack parameter, the more goals the team scores and the highest the defense parameter is, the more goals they will concede. We modified the Poisson distribution to better reflect reality because small scores were poorly estimated.

- **Simulating a match consists** in **simulating two scores**, the home team attack against the away team defense and the away team attack against the home team defense.
- If you are a football fan, you will probably see Arsenal - Liverpool tomorrow. If you aren't, we highly recommand you to see it! If you want to bet to spice it up, here are the odds we calculated:



Learning a team's parameters

We **learned those parameters** for all the teams of a football league **maximizing a log-likelihood**. We focused on the English Premiere League for now but our work can be applied to every league present in the API we are using.

- Our data set consists of **all the matchs played in the English Premiere League and in the English Championship** (english second league) **for at least the last three years**. To evalute the parameters of the promoted teams, we need a way to compare those teams to current Premier League teams through matches. That is why we used a three year dataset : the year of their promotion, the future promoted teams met with former Premier League that had been relegated and had met with current Premier League clubs the year before.
- Futhermore, we want **recent matches to have more weight** than matches played a long time ago. To do so, we transform a bit the likelihood by putting in exponent a function of time over each match.

Parameters for the English Premiere League

Team's parameters		
Teams	α_H	β_H
Arsenal	1.68	0.78
Chelsea	1.87	0.50
Liverpool	2.23	0.52
Everton	1.02	1.08
Leeds	1.17	1.50
Aston V.	1.37	0.91
Shared Parameters		
k	1.21	
ϵ	0.0065	

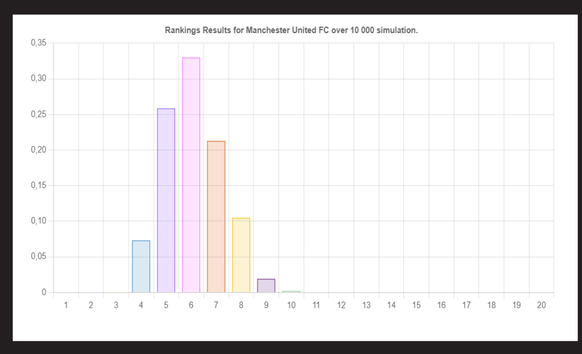
We see here two groups of teams:
The first three teams attack well and defend well. Through the parameters, we see that Aston V. is better than Everton and Leeds both in attack and defense.
$k = 1.21$ means that the home effect is a factor of 1.21. $\epsilon = 0.0065$ means that the influence of a match is in $e^{-0.0065t}$ (t = time difference in days).

End-of-season Ranking Prediction

Predicted Rank	Current Rank	Team	Points	Wins	Draws	Losses	GF	GA
1	1	Man City	92.2	29.15	4.74	4.11	91.22	24.9
2	2	Liverpool	87.79	26.61	7.96	3.43	95.15	28.41
3	3	Chelsea	80.94	23.6	10.15	4.25	79.1	25.74
4	4	Arsenal	69.73	21.3	5.82	10.88	63.47	45.26
5	8	Tottenham	64.33	19.55	5.69	12.76	54.69	46.9
6	5	Man United	61.75	17.2	10.16	10.64	60.45	51.92
7	6	West Ham	60.51	17.38	8.36	12.25	60.33	46.96
8	7	Wolverhampton	55.29	16	7.3	14.71	34.35	32.96
9	12	Leicester	53.18	14.86	8.6	14.54	63.15	63.15
10	9	Aston Villa	49.14	14.42	5.88	17.7	53.18	53.88
11	10	Southampton	48.96	11.82	13.5	12.68	48.93	580.5
12	11	Crystal Palace	47.16	10.95	14.31	12.74	54.19	52.67
13	13	Brighton	44.52	9.92	14.77	13.32	35.93	47.49
14	14	Newcastle	40.18	9.19	12.61	16.2	42.98	66.82
15	15	Brentford	37.7	9.79	8.33	19.88	41.14	61.3
16	19	Burnley	34.68	6.39	15.5	16.11	32.38	50.22
17	17	Everton	34.42	9.14	7.01	21.86	40.81	63.84
18	16	Leeds	32.91	7.59	10.14	20.27	41.63	83.38
19	18	Watford	28.34	7.37	6.22	24.4	36.71	70.87
20	20	Norwich	24.15	5.64	7.22	25.14	24.23	79.32

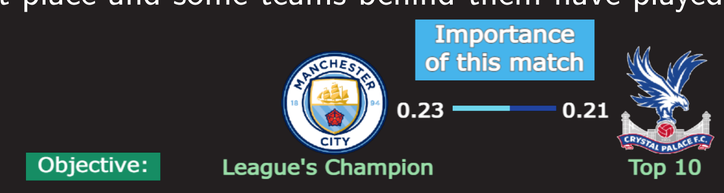
Using multiprocessing tools, we made in 16 minutes those predictions simulating 10 000 ends of seasons.

- Recently, Manchester United's performances has been poor and they have had a lot of internal issues whereas Arsenal has been walking on water with five consecutive wins and have a ligther calendar that Manchester. So it's logical that our method find Arsenal ahead of Manchester even if, right now, Manchester are ahead in the rankings.
- Leeds United FC have had such terrible resuts recently that they've just fired their coach. Even if they are not currently in the relegating zone, it seems more than legitimate to think that they will be relegated at the end of the season.
- To make the prediction table, we simulated 10 000 end of seasons possible scenarios and then calculated for each team their probability to finish at each rank of the final ranking. For **Manchester United**, this was the results:



New Indicator: Importance of a Match

- We define the importance of a match for a team and for a given end-of-season objective by how much a win or a loss affects the probability of that objective. For simplicity, we consider only 1-0 wins and losses. The larger the change in probability, the more important the game. We used two methods, one where we didn't re-learn the parameters after setting the score and the other,closer to the truth, where we did.
- For Manchester City-Crystal Palace yesterday, the match importance for Manchester was 0.23. It is coherent because Liverpool are really close to them and if they lose this match (that they are supposed to win), Liverpool will get ahead of them. Crystal Palace needs to win because they are in the 11st place and some teams behind them have played less matches.



References

M. J. MAHER: Statistica Neerlandica, 36(3):109-118,1982.
Modelling association football scores.

Mark J. DIXON and Stuart G. COLES: Applied Statistics, 46(2):265-280,1997.
Modelling association football scores and Inefficiencies in the Football Market Data.

Julien GUYON: Le Monde, 2020.
Football : comment décider du classement final de la Ligue 1 si elle devait s'arrêter ici ?