Bayesian Forecasting of UEFA Champions League under alternative seeding schemes[1]

M. Mishra[1] R. Vashistha[1] S. S. Banerjee[1] S. Grover[2]

[1]Department of Mathematics and Statistics
[2]Department of Aerospace Engineering
Indian Institute of Technology, Kanpur

April 21, 2022

[1]Main paper: Corona et al. (2019)

Table of Contents

Problem Statement

- Can we forecast results of Champions League using Bayesian Poisson model?

- Is there an effect of different seeding rule on chances of qualifying for teams?

UEFA Champions League

- A football tournament played in Europe every year.
- 32 teams from various countries play in the tournament.
  - Teams are divided in 8 groups.
  - Top 2 teams in each group qualify for knock-out stage.
- Participation in UCL is mostly determined by teams' respective position in their domestic league.

Seeding Rules

- Seeding rules determine division of teams into different groups.
- Old seeding rule
  - Teams are divided into 4 pots on basis of "UEFA coefficients".
  - Teams from same countries and within same pot can't be in the same group.
- New seeding rule
  - Pot 1 has reigning champion of UCL and the champion clubs from the seven strongest national leagues.
  - All of the remaining clubs are distributed across pots 2–4 according to the rank order of their coefficient rankings.
- Random Seeding Rule: Randomly divide teams into groups

Poisson Model

- Poisson regression is often used to model the counts.
- In case of a univariate data, the model is given by-

$$log(\lambda_i) = \beta_0 + \beta_1 x_i$$

, where the observed values $Y_i \sim Poisson(\lambda_i)$

- λ is estimated by the maximum likelihood estimation method in the frequentist framework
- In Bayesian approach, we estimate the distribution of λ to account for parameter uncertainty .

We apply a Poisson regression model (BP) for the goals scored by each team in a match as follows:

$$Y_{t,k} \sim Poisson(\lambda_{T,k})$$

where $Y_{t,k}$ represents the number of goals scored by team T in a match at time k and

$$log(Y_{t,k}) = \beta_{A_T} X_{A_{T,K}} + \beta_{A_O} X_{A_{O,K}} + \beta_{H_T} X_{H_{T,K}} + \beta_{F_T} X_{F_{T,K}} + \beta_{Aw_T} X_{Aw_{T,K}}$$

where

- $X_{A_{T,K}}$ represents the strength of team T.
- $X_{A_{O,K}}$ is the strength of the opposing club.
- $X_{H_{T,K}}$ indicates that team T is playing at home.
- $X_{Aw_{T,K}}$ indicates that team T is playing away.
- $X_{F_{T,K}}$ indicates that the match is the final of the whole competition.

Data Description

Home_coeff	Away_coeff	goals_scored	Home	Away	Final
22.5	1.2	2	1	0	0
1.2	22.5	1	0	1	0
9.75	18	0	1	0	0
18	9.75	0	0	1	0
3.75	4.75	5	1	0	0
4.75	3.75	2	0	1	0
1.2	1	5	1	0	0
1	1.2	0	0	1	0
1.28	2.85	0	1	0	0
2.85	1.28	0	0	1	0
0.8	2.75	2	1	0	0
2.75	0.8	3	0	1	0
8.5	3	1	1	0	0
3	8.5	1	0	1	0
5.5	1.05	2	1	0	0
1.05	5.5	2	0	1	0
7.25	0.775	3	1	0	0
0.775	7.25	0	0	1	0

Bayesian Framework

- We choose non-informative prior for each component of $\beta = (\beta_{A_T},\ \beta_{A_0},\ \beta_{H_T},\ \beta_{Aw_T},\ \beta_{F_T})'$.
- We choose the non-informative priors as Uniform(-2,2) distribution.
- We implement Metropolis algorithm to generate samples for successive elements of β.
- We calculate the win, draw and loss probabilities as follows:

$$p_{W,k} = \sum_{i_T=0}^{\infty} \sum_{i_0=1}^{i_T-1} P(y_{T,k} = i_T) P(y_{O,k} = i_O)$$

$$p_{D,k} = \sum_{i_T=0}^{\infty} P(y_{T,k} = i_T) P(y_{O,k} = i_T)$$

$$p_{L,k} = \sum_{i_O=0}^{\infty} \sum_{i_T=1}^{i_O-1} P(y_{T,k} = i_T) P(y_{O,k} = i_O)$$

Entropy

We use Entropy as a measure of uncertainty as to which team will win the tournament.
The entropy is defined as follows:

$$e_0 = -\sum_{j=1}^{32} p_{jW} \log_2 p_{jW}$$

Depending on whether the entropy under particular seeding scheme is more or less, we can get an idea if teams are more or less being benefited or are made to lose out on chances to win the tournament under that particular scheme.

Results

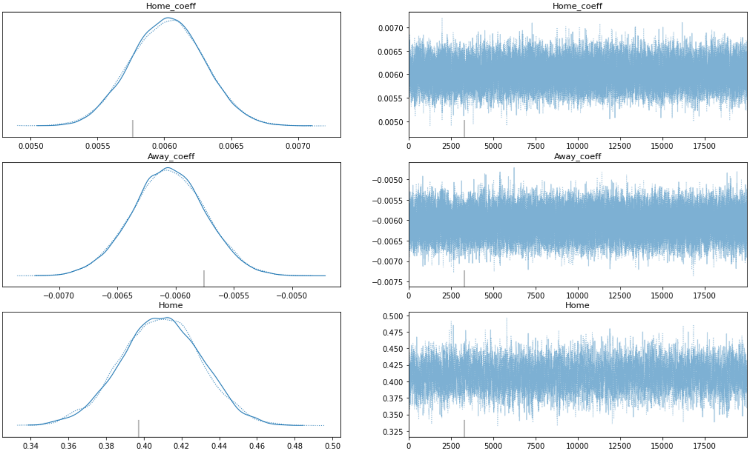


Figure: Density and Traceplot of posterior parameter distribution of a Poisson regression model

Results

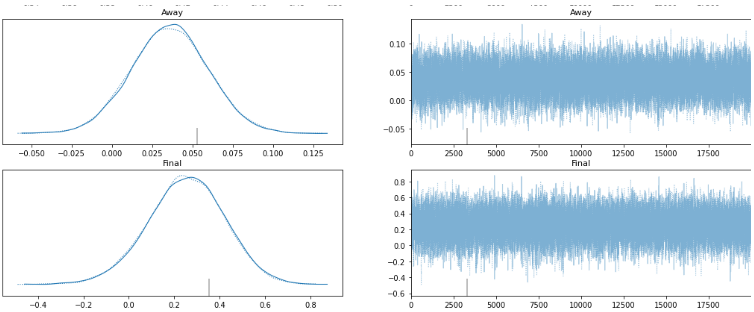


Figure: Density and Traceplot of posterior parameter distribution of a Poisson regression model

Results

	mean	sd
Home_coeff	0.0060	0.0003
Away_coeff	-0.0061	0.0003
Home	0.4091	0.0215
Away	0.0359	0.0239
Final	0.2525	0.1698

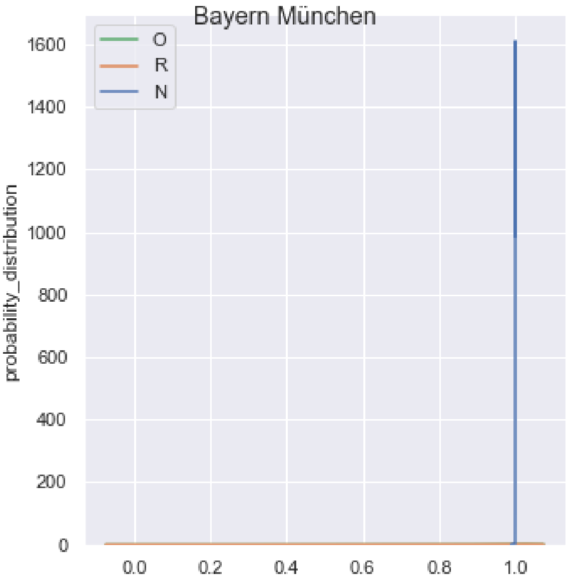
Figure: Summary of Poisson Regression Model

Tournament Simulation

The idea is as follows:

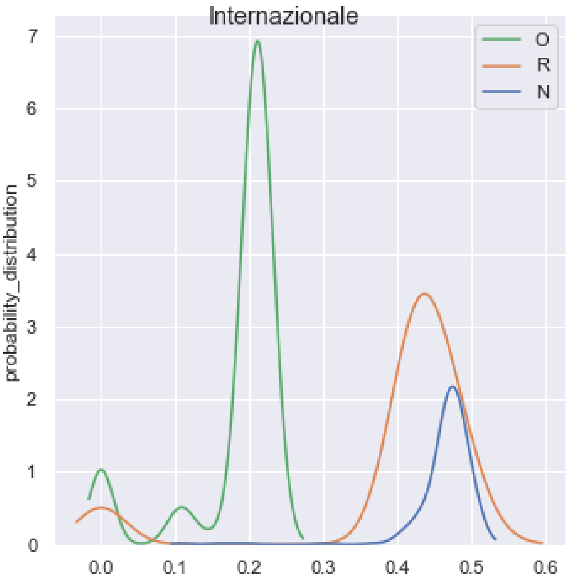
- Through MCMC we obtain 5000 samples for each parameter.
- For each parameter we simulate the tournament 1000 times and calculate the probability of winning and qualifying from he group stage.
- We repeat the step 2 for each 4999 values of the parameters obtained in step 1

Results

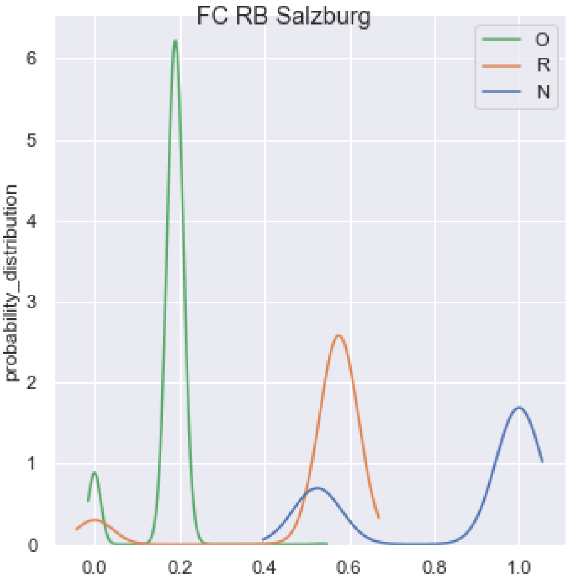


Bayern München
O
R
N
probability_distribution
1600
1400
1200
1000
800
600
400
200
0
0.0
0.2
0.4
0.6
0.8
1.0

Results


Internazionale
probability_distribution
O
R
N
0
1
2
3
4
5
6
7
0.0
0.1
0.2
0.3
0.4
0.5
0.6

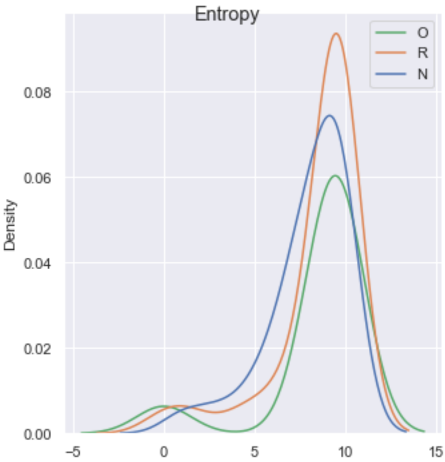
Results



FC RB Salzburg
probability_distribution
6
5
4
3
2
1
0
0.0
0.2
0.4
0.6
0.8
1.0
O
R
N

Entropy evaluation

We plot the densities of the Entropies calculated for each seeding scheme.



Final Remarks

- We have successfully reproduced the paper 'Bayesian Forecasting of UEFA Champions League under alternative seeding schemes'.
- We observe that 'weak' teams are benefited the most by the random seeding rule while strong teams tend to benefit from the new seeding rule.
- Bayern München team has the highest posterior probability of winning the tournament (0.3423).

References I

Corona, F., Forrest, D., Tena, J., and Wiper, M. (2019). Bayesian forecasting of uefa champions league under alternative seeding regimes. *International Journal of Forecasting*, 35(2):722–732.