

Analysing spatial factors on football goal scoring by bivariate Poisson regression

Tolkyn Abdildina

June 12, 2020

Abstract

Soccer is a game that catches the attention of people all around the world. It makes an evaluation of goal scoring probability an interesting and challenging topic for research. There are various features and factors that affect the goal scoring such as the distance from where a shot was attempted or assist methods. We would like to see what are these factors and use them in calculating the probabilities. The first step would be to investigate the factors and analyze their significance in scoring a goal. Logistic regression will give us a clear image of importance of keeping specific features, describing how the predictor variables which is goal scoring depends on the other variables. After dropping all the insignificant factors, the next step is adding them to bivariate Poisson model to estimate the probabilities, which was an important part of this research. Most of the methods that are used to predict the results of football matches consider two independent

Poisson distribution for number of goals for each team. But the covariance between these variables gives more insights, that is why we have chosen this model. Moreover, it includes the effects of parameters such as attacking strength of home team, defensive strength of away team and home team advantage. As a result of the logistic regression we dropped completely 4 out of 27, while keeping 8 factors for all the teams. Finally, after calculating the goal scoring probabilities, we notice the home team advantage effect.