PREDICTING THE OUTCOMES OF FIFA WORLD CUP TOURNAMENTS

Derek Miller
Chicago Booth School of Business
dgmxm7@gmail.com

July 31, 2019

ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

*K*eywords First keyword · Second keyword · More

1 Introduction

One way to predict match outcomes from tournament-style sporting events like the FIFA World Cup is to set up the problem as a matrix completion problem.

For the purposes of this case study, I will be using custom built utilities. All code can be found at github.com/dgmiller/fifa-world-cup.

The first thing we need to do is to set up the problem. Part of this is understanding the tournament structure. In particular, the tournament structure for a FIFA World Cup lends itself nicely to Bayesian analysis. Here's how the tournament is structured for the 2019 Women's World Cup.

The first stage of the tournament is called the group phase. Each of 24 teams are assigned to a group with 3 other teams. Each team will compete in at least 3 matches, one for every other team in their group. The match duration is 90 minutes. The possible outcomes for a team are a win, a loss, or a draw. Points are awarded for each outcome: 3 points for a win, 1 point for a draw, 0 points for a loss. The top 16 teams with the most points after the group stage advance to the knockout phase of the tournament. Each round after that eliminates half of the teams from the tournament. The naming conventions for each round are the Round of 16, the quarter finals, the semi-finals, and the final.

The second stage of the tournament is the knockout phase, where a team must win to advance. No draws are allowed in this stage. When two teams are tied at the end of 90 minutes, the match continues for two additional 15-minute periods. If the teams are still tied at the end of extra time, the match proceeds to penalty kicks to determine the winner.

1.1 Problem Setup

Now that we have reviewed the tournament structure, let's set up the problem. The heart of the problem is that we know that teams will be eliminated over the course of the tournament but we don't know which ones (except a few at the end of the group stage). So in some sense, we want to be able to compute counterfactuals, the what-if scenarios where every

team were to play every other team in the tournament. To do this, let's create a matrix where each team is indexed from 1-24, ordered first by their group (A-F) then by their rank within their group. This list looks like the following:

Now let's create a 24x24 matrix with entries corresponding to the number of goals team i scored against team j in the group stage. The color corresponds to the numeric value of the goals scored with grey corresponding to empty values in the matrix. The empty values are exactly the values we want to fill. This is a matrix completion problem; our goal is to fill in this matrix with some estimate of the number of goals we expect each team to score on average. That is, we want to compute $E[g_{i,j}]$ where g is the number of goals scored by team i against team j.

1.2 Previous Research

2 A Poisson Model of Soccer Goals

3 Estimation Procedure

4 Comparison of Estimation Strategies

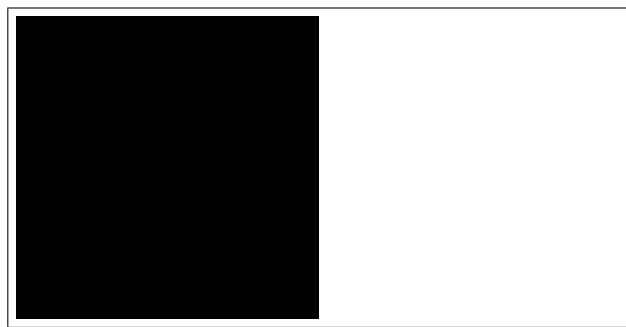## 5 Data and Estimation Results

6 Conclusion

Figure 1: Sample figure caption.

6.1 Tables

References

[1] Gianluca Baio and Marta A. Blangiardo. Bayesian hierarchical model for the prediction of football results year unknown. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.8659&rep=rep1&type=pdf

[2] Rasmus Baath. Modeling Match Results in Soccer using a Hierarchical Bayesian Poisson Model. 2015. http://sumsar.net/papers/baath_2015_modeling_match_resluts_in_soccer.pdf

[3] http://danielweitzenfeld.github.io/passtheroc/blog/2014/10/28/bayes-premier-league/

[4] Giovanni Angelini and Luca De Angelis. PARX model for football match predictions. 11 April 2017 in the Journal of Forecasting, Vol 36, Issue 7. https://doi.org/10.1002/for.2471