MS&E 220 PROJECT

THE ENGLISH PREMIER LEAGUE
A PROBABILISTIC ANALYSIS

Ankur Prasad, Rajiv Kurian, Raman Bhatia, and Xiaolong Tan
BACKGROUND

• The top 20 soccer teams in England compete yearly in the English Premier League. Each of the 20 teams plays against all the other teams in the championship, once at home and once at the home of that opponent.
• For each win a team is awarded 3 points, 1 point for a drawn game and 0 points for a lost game. At the end of the season, the team with the maximum number of points is declared champions of the English Premier League.
• At the beginning of the season, which team amongst the 20 participants is most likely to be champion and with what probability?
FORMULATION

Step 1: Determination of the probability of win, loss or draw in a matchup between any 2 teams.

For each matchup we model the possible outcomes as follows:
- Each team is represented by two urns (home and away)
- The number of balls of each kind is determined by the losses, draws and wins (at home or away) of that team in the past 2 years
- The balls can be of: type 0 (one for each loss), type 1 (one for each tie), type 2 (one for each win). The win, loss, tie statistics are derived from the weighted averages of the team’s performance over the past 2 years.
- We calculate the probabilities of win, loss for tie for team A as follows
  - Event (Ball type x is picked from urn for team T) is represented by: \( P_b(x,T) \)
    - \( P_b(A \text{ wins}) = P_b(2,A) \times P_b(1,B) + P_b(2,A) \times P_b(0,B) + P_b(1,A) \times P_b(0,B) \)
  - Thus A wins in the following situations:
    - We pick a ‘win’ ball for A and either a ‘loss’ or a ‘tie’ ball for B.
    - We pick a ‘tie’ ball for A and a ‘loss’ ball for B.
  - We calculate \( P_b(A \text{ loses}) \) and \( A(\text{ ties}) \) similarly.

![Diagram showing urns with balls representing wins, losses, and draws](image-url)
FORMULATION

Step 2: Determination of probability of team scoring certain number of points at the end of the season

<table>
<thead>
<tr>
<th>Match no</th>
<th>Opponent</th>
<th>Points</th>
<th>Probability of scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Middlesbrough</td>
<td>0</td>
<td>Pb(W2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.31911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.25052</td>
</tr>
<tr>
<td>2</td>
<td>Sunderland</td>
<td>0</td>
<td>Pb(D2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.16303</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.097236823</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.281252</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.286321</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Probability of scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tottenham beats Sunderland</td>
<td>Pb(W2)</td>
</tr>
<tr>
<td>Tottenham loses to Sunderland</td>
<td>Pb(L2)</td>
</tr>
<tr>
<td>Tottenham ties Sunderland</td>
<td>Pb(D2)</td>
</tr>
<tr>
<td>3 points after the first game</td>
<td>Pb(3)</td>
</tr>
<tr>
<td>0 points after the first game</td>
<td>Pb(0)</td>
</tr>
<tr>
<td>1 point after the first game</td>
<td>Pb(1)</td>
</tr>
<tr>
<td>Four points after the first game</td>
<td>Pb(4)</td>
</tr>
</tbody>
</table>

Four points after the second game: Pb(1)*Pb(W2)+Pb(3)*Pb(D2)+Pb(4)*Pb(L2)
FORMULATION

Step 3: Determination of the probability of a team being champions at the end of the season

Number of points required to win the league is between: 40 (37 draws and 1 win) and 114 (all wins)
A team wins the league if its number of points at the end of the season is the highest

P (Chelsea Wins)

\[ P(\text{Chelsea Wins}) = \sum_{n=0}^{114} P(\text{Chelsea} > n | \text{All other teams score} < n) \]

\[ = \sum_{n=60}^{114} P(\text{Chelsea} > n) \times P(\text{ManU} < n | \text{Chelsea} > n) \times P(\text{Arsenal} < n | \text{ManU} < n \& \text{Chelsea} > n) \ldots \]
ANALYSIS

Assumption of weak dependency:

Total # of games played by all the teams: 380
Total # of points accumulated by all the teams: between 380*2 =760(all draws) and 380*3=1140(no draws).
Total # of points won by the champion: between 40 and 114
ANALYSIS

PMF of total number of scores
ANALYSIS

Analysis around Central Limit Theorem
CONCLUSION

Normalized probability of winning the English Premier League

- Chelsea: 37.538%
- Manu: 31.832%
- Arsenal: 19.337%
- Liverpool: 9.034%
- Everton: 0.973%
- Aston Villa: 0.872%
- Blackburn: 0.317%