Forecasting football results with the Bradley-Terry model

Matthew Barrett
School of Mathematics, University of Bristol, UK

University of Bristol
School of Mathematics

Bradley-Terry model

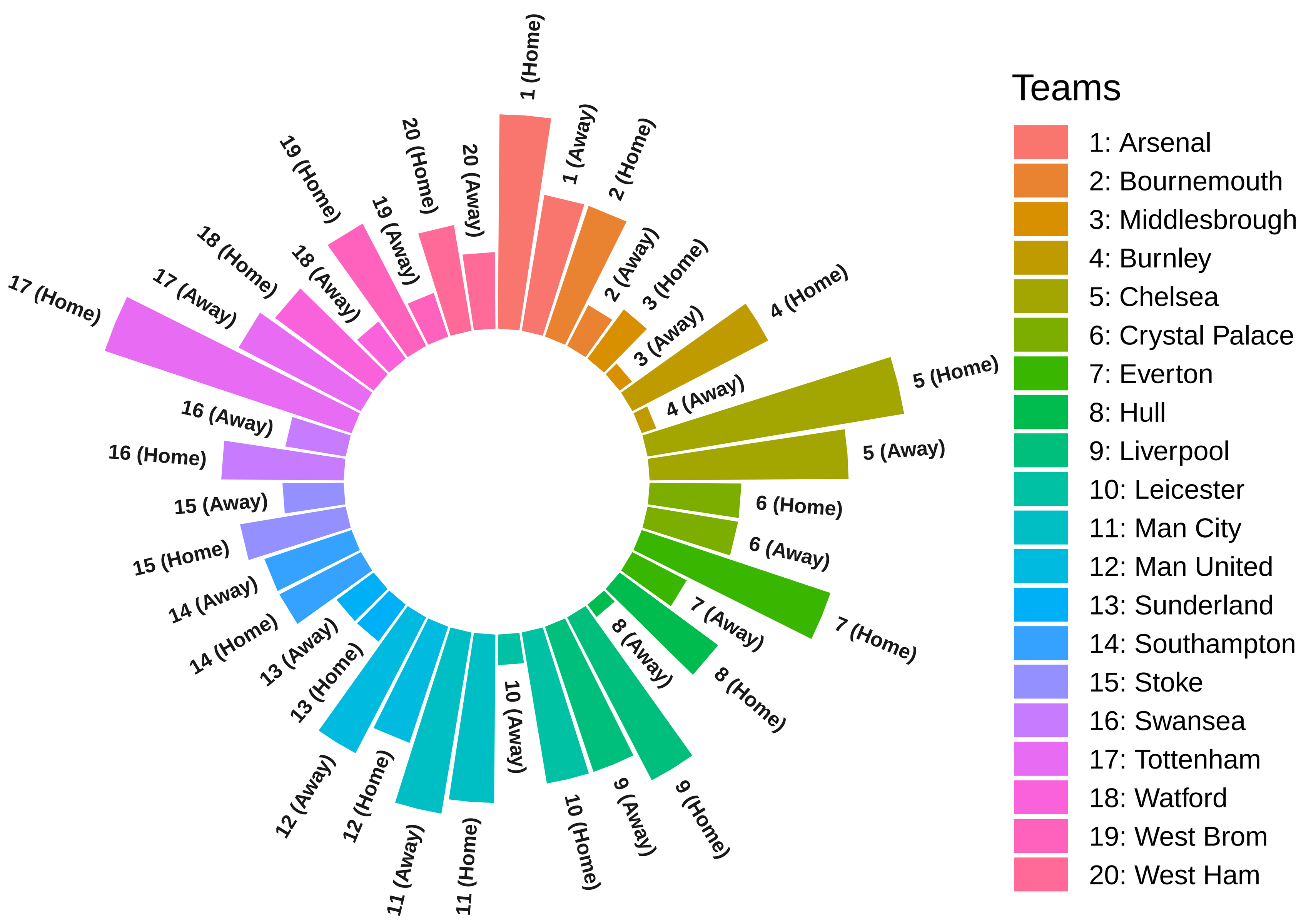
The Bradley-Terry model [2] tells us that in a paired comparison of two objects, i and j, the probability that i is preferred to j is given by

$$\frac{\pi_i}{\pi_i+\pi_j}$$

where π_i is a positive score assigned to object i. We will look at applying the Bradley-Terry model to the context of football in the Premier League. Hence the objects being compared are the 20 different teams in the Premier League, and the π_i terms can be thought of as the strength of team i. Additionally in this context i being preferred to j means that team i beats team j.

Model extensions

The original Bradley-Terry model falls short when it comes to modelling the outcomes of football matches. This is because in football draws occur fairly often, so we need to be able to assign a probability to the outcome of a draw, as well as either of the teams winning. Also, the following plot shows how many wins each team got when playing at home and away, in the 2016/17 Premier League season.



We can see that for the majority of the teams, they win more games at home than they do away. Hence we want to incorporate a home advantage effect into the model as well. We propose that if i is the home team, then

$$\mathbb{P}(\text{i beats j}) = \frac{\theta\pi_i}{\theta\pi_i+\pi_j+\nu\sqrt{\theta\pi_i\pi_j}}$$

$$\mathbb{P}(\text{i draws j}) = \frac{\nu\sqrt{\theta\pi_i\pi_j}}{\theta\pi_i+\pi_j+\nu\sqrt{\theta\pi_i\pi_j}}$$

$$\mathbb{P}(\text{j beats i}) = \frac{\pi_j}{\theta\pi_i+\pi_j+\nu\sqrt{\theta\pi_i\pi_j}}$$

where $\theta \geq 0$ is the magnitude of the home advantage (or disadvantage) [1], and $\nu \geq 0$ controls the prevalence of draws [3]. To offer some brief justification of why the probability of a draw presented may be appropriate, note that $\sqrt{\theta\pi_i\pi_j}$ attains its maximum value when $\theta\pi_i = \pi_j$. In other words, the probability of a draw is maximal when the strengths of the two teams are equal, which aligns with our intuitions about the game. Of course we also add the $\nu\sqrt{\theta\pi_i\pi_j}$ term into the denominator to ensure all outcome probabilities sum to 1.

Inference

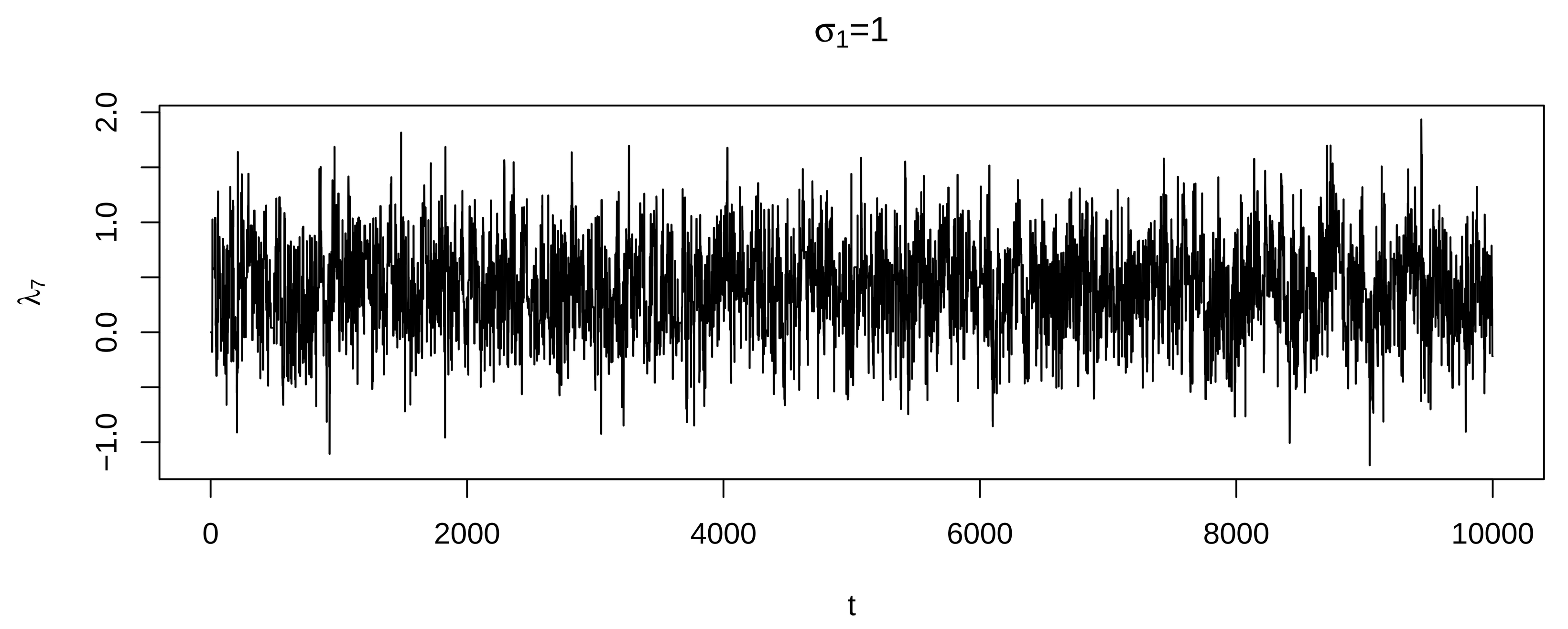
Now that we have specified our model, we want to be able to infer the values of the parameters. The first method we will use is maximum likelihood estimation. In particular, we construct the log likelihood function using some observed outcomes, and we find the values of the parameters which maximise this function.

Alternatively, we use Markov chain Monte Carlo methods to infer the posterior means of the parameters. To do this we first reparameterise such that $\pi_i = e^{\lambda_i}$, $\nu = e^{\phi}$ and $\theta = e^{\alpha}$. We do this to impose the positivity constraint that we have on all of our original parameters. Hence we substitute these reparameterisations into our probability terms, and we instead try to infer the values of the log parameters λ_i, ϕ and α.

We don't have much prior knowledge about the values of the log parameters, and we reflect this in our relatively uninformative prior distributions. In particular, we set all the priors to be independent standard normals. Now having set priors, we want to find the mean of the posterior distribution. We will do this by obtaining samples from the posterior using the random walk Metropolis-within-Gibbs algorithm, and approximating the mean with the average of these samples.

As the name suggests, the algorithm performs random walks on the log parameters. More specifically, the next value in the random walk we choose to be normally distributed with mean equal to the current value and the standard deviation is to be tuned. Each new value is either accepted or rejected, according to how likely the new value is under the posterior. This accept/reject rule ensures that the random walk generates a Markov chain which has the posterior distribution as invariant. This means that once the Markov chain has converged to its invariant distribution, then all subsequent values will be distributed according to the posterior.

As an example, using all the outcomes from the 16/17 season we perform the Metropolis-within-Gibbs algorithm and we arbitrarily show the random walk on the log parameter λ_7.



In this case a standard deviation of 1 seems to create a Markov chain that mixes well, but this isn't necessarily the same for the other log parameters. Also, we can see that the Markov chain is fluctuating around the same set of values, and this means that the chain has indeed converged to its invariant distribution. Hence, we could take the average of the 10000 values in this chain to be our estimate of the posterior mean of the parameter λ_7.

Results

We consider predicting the outcomes of all games from the 16/17 season from matchweek 11 onward. In order to predict the outcomes in week 11, we use all the results from weeks 1-10 to estimate the parameters of our model. With these estimates we can then predict the outcome probabilities of the games in week 11, and we take the outcome with the highest predicted probability as the predicted outcome. Then to predict the games in week 12 we update our parameter estimates using all of the results from weeks 1-11. Continuing this procedure until the end of the season we have an outcome prediction for every game from week 11 until the end of the season.

Performing this procedure with the parameter estimates obtained by maximum likelihood estimation and comparing the predictions to the actual results, we have that 60.7% of the predictions are correct. The same procedure with the parameter estimates given by the posterior means results in predictions which are correct 61.8% of the time.

To gauge how successful the model is at forecasting the results, we look at the null classification. This is simply how often the most frequent outcome occurs. Naturally, since the home team has an advantage the most frequent outcome is a home win. Of the matches we attempted to predict, 51.8% of them ended in a home win. It follows that a trivial model which predicts a home win for every game is correct 51.8% of the time, and so this is the baseline accuracy. It seems then that our model is relatively successful at forecasting results, with an accuracy of around 10% higher than this baseline.

References

[1] Alan Agresti. *Categorical data analysis*. Vol. 482. John Wiley & Sons, 2003.

[2] Ralph Allan Bradley and Milton E. Terry. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". In: *Biometrika* 39.3/4 (1952), pp. 324–345. ISSN: 00063444. URL: http://www.jstor.org/stable/2334029.

[3] Roger R. Davidson. "On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments". In: *Journal of the American Statistical Association* 65.329 (1970), pp. 317–328. ISSN: 01621459. URL: http://www.jstor.org/stable/2283595.