

CAUTION: These lecture notes are under construction. You may find parts that are incomplete.

### 3 THE BRADLEY-TERRY MODEL

We learned in Chapter ?? that a team's past score differential is generally a stronger predictor of its future winning percentage than is its past winning percentage. Missing from a team's score differential is any information about the quality of its opposition (i.e. strength of schedule). That is the focus of this chapter.

We start by introducing some mathematical notation to describe score differential. Suppose we have a dataset of  $n$  games between  $p$  teams. For each game  $i \in \{1, 2, \dots, n\}$ , we observe the home team  $h_i \in \{1, \dots, p\}$ ; the away team  $a_i \in \{1, \dots, p\}$ ; and the score differential  $y_i$  (home score minus away score). We use  $x_{ij} = \{\mathbb{I}(h_i = j) - \mathbb{I}(a_i = j)\}$  to indicate whether team  $j$  is home/away/inactive (+1/-1/0) in game  $i$ . From these definitions, we calculate the average score differential per game for team  $j$  as:

$$s_j = \frac{\sum_{i=1}^n x_{ij} \cdot y_i}{\sum_{i=1}^n x_{ij}^2}.$$

EXAMPLE: 1995-1996 Hogwarts Inter-House Quidditch Cup

Hogwarts quidditch data are woefully incomplete. The Inter-House Quidditch Cup is a round-robin (each team plays each other team once) tournament in which the winner is determined by total number of points scored. We know the results of four of the six matches from the 1996-96 cup:

Gryffindor def. Slytherin, 200-20  
Hufflepuff def. Gryffindor, 240-230<sup>†</sup>  
Hufflepuff def. Ravenclaw, 230-210\*  
Gryffindor def. Ravenclaw, 190-40\*

There is no concept of "home team" in the cup, so let's treat the first team alphabetically as the home team in each match. Hufflepuff's average score differential is:

$$(0 \cdot 180 + -1 \cdot -10 + 1 \cdot 20 + 0 \cdot 150) / (0 + 1 + 1 + 0) = 30/2 = 15$$

<sup>†</sup> Miraculously, this match only lasted 22 minutes!

\* Result known but exact score unknown. Most likely score inferred from Rowling JK (2003) *Harry Potter and the Order of the Phoenix*. See [harrypotter.fandom.com/wiki/Inter-House\\_Quidditch\\_Cup](http://harrypotter.fandom.com/wiki/Inter-House_Quidditch_Cup)

We may interpret  $s_j$  as an estimate of the strength of team  $j$ . We could use these estimated strengths to calculate a strength-of-schedule-adjusted score differential for each game. If  $o_{ij}$  is the opponent of team  $j$  in game  $i$ , then we would achieve this with  $x_{ij} \cdot y_i + s_{o_{ij}}$ . This leads to a strength-of-schedule-adjusted average score differential for team  $j$ :

$$s_j^* = \frac{\sum_{i=1}^n x_{ij} \cdot y_i + s_{o_{ij}}}{\sum_{i=1}^n x_{ij}^2}. \quad (1)$$

Perhaps  $s_j^*$  is a better estimate of team strength than  $s_j$  because it accounts for strength of schedule. So maybe we can do better by replacing  $s_{o_{ij}}$  with  $s_{o_{ij}}^*$  in the equation above. This would yield a better strength estimate  $s_j^{**}$  to replace  $s_{o_{ij}}^*$ , and so on. Instead of repeating this operation indefinitely, let us jump straight to the end with a cogent statistical model.

#### 3.1 MODELING SCORE DIFFERENTIAL

In its simplest form, the Bradley-Terry model describes the distribution of the random variable  $Y_i$ , representing the score differential of game  $i$ :

$$\begin{aligned} \eta_i &= \beta_0 + \beta_{h_i} - \beta_{a_i} \\ Y_i &\sim \text{Normal}(\eta_i, \sigma^2). \end{aligned} \quad (2)$$

This model has  $p + 1$  regression coefficients to estimate: one  $\beta$  for each team (interpretable as the team's strength), as well as  $\beta_0$  (interpretable as home-field advantage).<sup>1</sup> Unfortunately, this model is not identifiable, meaning that the parameters cannot be uniquely estimated from the data. To see this, note that if we add any constant  $c$  to all team strengths, the predictions do not change because

$$\beta_0 + \beta_{h_i} - \beta_{a_i} = \beta_0 + (\beta_{h_i} + c) - (\beta_{a_i} + c)$$

for any  $c \in \mathbb{R}$ . To make the model identifiable, we must introduce an additional constraint. The simplest, most common constraint is to set  $\beta_1 = 0$ , meaning that the first team is the reference team against which all other teams are measured. Having established this constraint, we proceed with  $p$  regression coefficients to estimate.

You may recognize this as a linear regression model. The simplest method for estimating the unknown parameters is ordinary least squares (OLS). Using  $\boldsymbol{\beta}$  to denote the  $p \times 1$  vector of regression coefficients, we choose  $\boldsymbol{\beta}$  to minimize the sum of squared residuals (equivalent to maximizing likelihood of the data):

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - (\beta_0 + \beta_{h_i} - \beta_{a_i}))^2 \quad \text{s.t. } \beta_1 = 0. \quad (3)$$

This optimization problem has a closed-form solution, which can be derived by finding where the gradient is zero<sup>2</sup> (because the objective function is convex and differentiable). To express the solution, it is helpful to introduce matrix notation for the model. We use  $\mathbf{y}$  to denote the  $n \times 1$  vector of observed score differentials  $(y_1, y_2, \dots, y_n)^T$ , and we use  $\mathbf{X}$  to denote the  $n \times p$  matrix encoding the intercept (+1), home team (+1) and away team (-1) of each game. Namely:

$$(\mathbf{X})_{ij} = \begin{cases} +1 & \text{if } j = 1 \text{ (intercept column)} \\ +1 & \text{if } h_i \text{ is team } j \\ -1 & \text{if } a_i \text{ is team } j \\ 0 & \text{otherwise} \end{cases}.$$

Note that every row of  $\mathbf{X}$  has at most three nonzero entries: one +1 for the intercept, one +1 for the home team and one -1 for the away team. Recall that team 1 is not represented in  $\mathbf{X}$  (because of the constraint that  $\beta_1 = 0$ ), so rows corresponding to games in which team 1 is active will have only two nonzero entries. Because the vast majority of  $\mathbf{X}$  is zero, we call it a *sparse* matrix.<sup>3</sup> Using this notation, the OLS solution is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

EXAMPLE: 1995-1996 Hogwarts Inter-House Quidditch Cup

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 180 \\ -10 \\ 20 \\ 150 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{\text{Hufflepuff}} \\ \beta_{\text{Ravenclaw}} \\ \beta_{\text{Slytherin}} \end{bmatrix}$$

Because there is no home team, let's drop the home-field advantage coefficient  $\beta_0$  from the model.

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 180 \\ -10 \\ 20 \\ 150 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_{\text{Hufflepuff}} \\ \beta_{\text{Ravenclaw}} \\ \beta_{\text{Slytherin}} \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \left( \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & -1 \\ -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & -1 \\ -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 180 \\ -10 \\ 20 \\ 150 \end{bmatrix} = \begin{bmatrix} -36.7 \\ -103.3 \\ -180.0 \end{bmatrix}$$

<sup>1</sup>The variance parameter  $\sigma^2$  is also a fixed, unknown parameter. It is considered a *nuisance parameter* because its estimation has no bearing on the regression coefficients, which are of primary interest.

<sup>2</sup>We omit the details of this calculation. You can learn this in STAT 315 or DSCI 301.

<sup>3</sup>The most notable aspect of sparse matrices for our needs is that we can leverage them for computational speedups.

We interpret  $\hat{\beta}_j$ , the element of  $\hat{\beta}$  corresponding to team  $j$ , to be the estimated strength of team  $j$ . A satisfying property of these estimated strengths is that they are a steady-state solution to equation (1), meaning:

$$\hat{\beta}_j = \frac{\sum_{i=1}^n x_{ij} \cdot y_i + \hat{\beta}_{o_{ij}}}{\sum_{i=1}^n x_{ij}^2}. \quad (4)$$

In other words,  $\hat{\beta}_j$  is equivalent to the score differential for team  $j$  after adjusting for strength of schedule, where team strengths are defined by  $\hat{\beta}$ .

### 3.2 MODELING WIN-LOSS OUTCOME

Although we prefer to evaluate teams based on score differential for most applications, the Bradley-Terry model can be generalized to model the win-loss outcome as well. The random variable  $Z_i = \mathbb{I}(Y_i > 0)$  represents an indicator of whether the home team wins game  $i$ .

$$\begin{aligned} \eta_i &= \beta_0 + \beta_{h_i} - \beta_{a_i} \\ Z_i &\sim \text{Bernoulli}\left(\frac{e^{\eta_i}}{1 + e^{\eta_i}}\right). \end{aligned} \quad (5)$$

We require the same constraint as in equation (2), that  $\beta_1 = 0$ . You may recognize this as a logistic regression model. As with linear regression, we choose  $\beta$  to maximize the log-likelihood of the observed data  $z_1, z_2, \dots, z_n$ , equivalent to minimizing the negative log-likelihood:

$$\hat{\beta} = \arg \min_{\beta} - \sum_{i=1}^n \left( z_i \log \left( \frac{e^{\eta_i}}{1 + e^{\eta_i}} \right) + (1 - z_i) \log \left( \frac{1}{1 + e^{\eta_i}} \right) \right) \quad \text{s.t. } \beta_1 = 0.$$

Unlike equation (3), this optimization problem does not have a closed-form solution, and it is solved using an iterative algorithm.<sup>4</sup>

### 3.3 PREDICTING FUTURE OUTCOMES

For both the linear regression model in equation (2) and the logistic regression model in equation (5), we calculate the  $n \times 1$  vector  $\hat{\eta}$  of fitted values from the regression the same way:

$$\hat{\eta} = \mathbf{X}\hat{\beta}.$$

For the linear regression,  $\hat{\eta}_i$  is directly interpretable as the predicted score differential for game  $i$ , based on the estimated strengths of the home and away teams. For the logistic regression,  $\hat{\eta}_i$  is the log-odds of the home team winning game  $i$ . We can convert this to a probability by applying the logistic function:

$$\hat{z}_i = \frac{e^{\hat{\eta}_i}}{1 + e^{\hat{\eta}_i}}.$$

We can also produce predictions for unseen games by encoding the teams involved in a new matrix  $\tilde{\mathbf{X}}$ .

EXAMPLE: 1995-1996 Hogwarts Inter-House Quidditch Cup

We never got to see Ravenclaw play against Slytherin in the 1995-96 cup. According to the fitted Bradley-Terry model above, what would be the expected score differential?

$$\tilde{\mathbf{X}}\hat{\beta} = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} -103.3 \\ -180.0 \end{bmatrix} = \begin{bmatrix} 76.7 \end{bmatrix}$$

Or, to put it more simply,  $\hat{\beta}_{\text{Ravenclaw}} - \hat{\beta}_{\text{Slytherin}} = -103.3 - (-180.0) = 76.7$ .

<sup>4</sup>We omit the details of this calculation. You can learn this in STAT 410.

One interpretable insight we can draw from the estimated regression coefficients is an estimate for the home-field advantage. In the linear regression,  $\hat{\beta}_0$  is the estimated effect on the point spread in favor of the home team. In the logistic regression,  $\hat{\beta}_0$  is the estimated effect on the log-odds in favor of the home team.

Predicting the future outcomes is an important application for sports teams (because of the strategic implications) and for sports bettors. The design matrix  $\mathbf{X}$  does not encode future games, but the formula for extracting predictions from the model is the same as for past games. For a future game between home team  $h$  and away team  $a$ , the predicted score differential using the linear regression is  $\hat{\beta}_0 + \hat{\beta}_h - \hat{\beta}_a$ ; and the predicted probability of the home team winning using the logistic regression is  $e^{\hat{\beta}_0 + \hat{\beta}_h - \hat{\beta}_a} / (1 + e^{\hat{\beta}_0 + \hat{\beta}_h - \hat{\beta}_a})$ . To make a large number of predictions in a computationally efficient way, one could encode the future matchups in a new design matrix  $\tilde{\mathbf{X}}$  and extract the predictions using  $\tilde{\mathbf{X}}\hat{\beta}$ .

### 3.4 DISCUSSION QUESTIONS

1. Look for a website to answer the question: Which NFL team had the toughest strength of schedule last year? What method was used to come up with an answer?
2. When might we prefer to use win-loss outcome in the Bradley-Terry model instead of score differential?

### 3.5 EXERCISES

1. Why do we need to force  $\beta_1 = 0$  in the Bradley-Terry model?
2. Suppose we estimate the Bradley-Terry model and find  $\hat{\beta}_2 > 0$ . What does this tell us about team 2?
3. Suppose we estimate the win-loss outcome model and find  $\hat{\beta}_0 = 0.5$ ;  $\hat{\beta}_2 = 1$ ; and  $\hat{\beta}_p = 0.75$ . What is the estimated probability that team  $p$  beats team 2 at home?