Lab 9: Paired comparison models

Michael Lopez, Skidmore College

Overview

In this lab, we are going to implement and analyze the results of Bradley-Terry models.

Recall some of our notation from lecture. Let players (or teams) be indicated by i and j, where $P(i$ beats $j)$ = $\frac{\alpha_i}{\alpha_i+\alpha_j}$, for some player **abilities** α_i and α_j, where $\alpha_i > 0$ and $\alpha_j > 0$.

The odds that i beats j are $\frac{\alpha_i}{\alpha_j}$; on the log scale, this simplifies to log(α_i) - log(α_j) = λ_i - λ_j, where $\lambda_i = log(\alpha_i)$ for all i.

The data

We are going to use a data set that comes with the `BradleyTerry2` package called **icehockey**.

Let's see what it looks like.

```
library(tidyverse)
library(BradleyTerry2)
head(icehockey)
summary(icehockey)
dim(icehockey)
```

The data contain 1083 games from the 2009-10 NCAA college hockey season. More of the variables are self-explanatory, but there are a few that are worth pointing out.

First, game results are coded as `0`, `0.5`, or `1`.

1. Look at the first ten rows of the `icehockey` data set. What do game results of `0`, `0.5`, and `1` correspond to?

Second, most games are played with `home.ice = TRUE`. However, that's not always the case.

```
icehockey %>% count(home.ice)
```

Roughly 70 games are played on neutral ice. In the NCAA, these are often nonconference games, or tournament games played at the beginning or ends of the year.

Fitting the standard Bradley-Terry model

First, let's start by fitting the standard BTM model on this data set.

```
standardBT <- BTm(outcome = result,
    player1 = visitor, player2 = opponent,
    id = "team", data = icehockey)
```

A few reminders on the code above, the first input, `outcome`, is the result of the game. The BTM has a different mechanisms for dealing with ties, the details of which are interesting but are best left for another day. For now, recognize that with ties stored as 0.5, the BTM assumes that's worth roughly half a win for each team.

The second two components are the participating team labels (`visitor`, `opponent`), and the next component is the identification name that we choose for each row of the output (in this case, `team`).

The output is lengthy, but worth looking at.

```r
library(broom)
tidy(standardBT)
```

2. Make a table of the visiting or home teams to identify which team is set as the reference team in the BTM. Given the estimates of team strength, does this team appear to be better than average, roughly about the league average, or worse than average?
3. Using the output, estimate the probability of Air Force beating Alaska.

Next, we look at the **abilities** for each team by exponentiating the player-level coefficients.

```r
abilities <- exp(BTabilities(standardBT))
abilities <- data.frame(abilities)
abilities$Team <- rownames(BTabilities(standardBT))
abilities %>%
  arrange(ability)
```

4. Which does the BTM rank as the league's best three teams? Use the internet to identify how these teams did in the 2010 NCAA men's hockey postseason.
5. Identify the biggest mismatch in teams, and what the probability is of the worse team winning that game.

BTM with home ice included.

Home ice effects are easily incorporated into the Bradley-Terry model, although it takes a bit of careful identification in the code.

```r
homeBT <- BTm(result,
              data.frame(team = visitor, home.ice = 0),
              data.frame(team = opponent, home.ice = home.ice),
              ~ team + home.ice,
              id = "team", data = icehockey)
```

Note that because not all of our games feature a team with home ice advantage, we don't assign 1 to every home ice for the second team. Instead, the `home.ice` input is our variable, also named `home.ice`, and in each game played without a home team, `home.ice` is stored as 0.

Let's check to see what home ice is worth in college hockey.

```r
tidy(homeBT)
tidy(homeBT) %>% tail()
```

6. Assuming team strength is held constant, what are the increased odds that the home team wins?

We can get a 95% confidence interval for the odds of a home win by exponentiating the R output.

```r
low.bound <- exp(0.402 - 1.96*0.071)
upp.bound <- exp(0.402 + 1.96*0.071)
low.bound; upp.bound
```

7. Interpret the interval above.

Let's return to our original equation, and add in a home advantage term.

Original: log(Odds (i beats j)) = λ_i - λ_j where $\lambda_i = log(\alpha_i)$ and $\lambda_j = log(\alpha_j)$

Home advantage: log(Odds (i beats j in a home game for i)) = λ_i - λ_j where $\lambda_i = log(\alpha_i + \omega_i)$ and $\lambda_j = log(\alpha_j)$, where ω_i is an indicator for the game being played at team i. *Note*: If the game is played at a neutral site, $\omega_i = 0$.

Returning to estimates of team-level terms, we have

```
head(BTabilities(homeBT))
```

8. Estimate the probability that Air Force beats Alaska in a game with no home-ice advantage (this should be similar to your answer to question 3).
9. Assume home ice advantage for Air Force, and estimate the probability that it beats Alaska.
10. Use your answers to the previous two questions to confirm that the odds of a home win in hockey are about 1.496 times higher than a road team win, assuming team strength is fixed.
11. Assume home ice advantage for Alaska, and estimate the probability that Air Force wins.