HW 10: Paired comparison models

Michael Lopez, Skidmore College

Preliminary notes for doing HW

1. All files should be knit and compiled using R Markdown. Knit early and often! I do not recommend waiting until the end of the HW to knit.
2. All questions should be answered completely, and, wherever applicable, code should be included.
3. If you work with a partner or group, please write the names of your teammates.
4. Copying and pasting of code is a violation of the Skidmore honor code

Overview

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

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

```
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`.

Exploratory analysis

1. Make a histogram identifying the total number of goals in each game (you'll have to make this variable), faceted by conference. Then calculate the mean total goals per game for each conference. Compare the centers, shapes, and spread of this variable (total number of goals scored) between the conferences.
2. Look only at the games not played at neutral sites (`home.ice == FALSE`). How many more goals per game does the home team score, on average, than the away team?

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)
library(broom)
tidy(standardBT)

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

3. Using the output, estimate the probability of Boston College beating Army.
4. Identify the biggest mismatch in teams, and what the probability is of the worse team winning that game.

BTM with home ice included.

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, $\omega_i = 0$.

Here's the home ice model

```
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)

tidy(homeBT)
tidy(homeBT) %>% tail()
head(BTabilities(homeBT), 10)
```

5. Assuming team strength is held constant, what are the increased odds that the home team wins?
6. Estimate the probability that Boston College beating Army in

- a game with no home-ice advantage (this should be similar to your answer to question 3).
- a game with Boston College having home advantage
- a game with Army having home advantage

7. Use your answers to the previous question 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.