BradleyTerry2: Flexible Models for Paired Comparisons

Heather Turner and David Firth

Department of Statistics
University of Warwick, UK

21 July 2010
Pair Comparisons

In situations where one object is pitted against another, e.g.

- players/teams in sport
- consumer products in market research
- images in psychology experiments
- plants in pest-resistance trials
- alleles in transmission from parent to child
Bradley-Terry Models

In a contest between two players \( i \) and \( j \), the basic Bradley-Terry Model is given by

\[
\text{odds}(i \text{ beats } j) = \frac{\text{pr}(i \text{ beats } j)}{\text{pr}(j \text{ beats } i)} = \frac{\alpha_i}{\alpha_i + \alpha_j} = \frac{\alpha_i}{\alpha_j}
\]

where \( \alpha_i, \alpha_j > 0 \) are the player abilities.

The abilities can be estimated via maximum likelihood by re-framing the model as a logistic model:

\[
\text{logit}(\text{pr}(i \text{ beats } j)) = \lambda_i - \lambda_j
\]
Structured Bradley-Terry Models

Contest-specific effects

\[ \lambda_{ik} = \alpha_i + \sum_r \beta_r x_{ikr} \]

Ability modelled by player attributes

\[ \lambda_i = \sum_r \beta_r x_{ir} + e_i \]

The prediction error \( e_i \)

- allows for variability between players with equal covariate values
- induces correlation between comparisons with a common player
The *BradleyTerry2* package

New features to accommodate general Bradley-Terry model

- **flexible formula interface** to modelling fitting function \( \text{BTm()} \): allows player-specific, judge-specific, contest-specific variables and random effects
- **PQL algorithm** for estimation of GLMMs
- **efficient data management** of multiple data frames

Best of original *BradleyTerry* package

- translation of ability formula to design matrix
- methods for fitted model object, e.g. \( \text{anova}, \text{BTabilities} \)
- handling of missing data in player covariates
College Ice Hockey: Men’s Division I

1083 games from the 2009-10 composite schedule

58 teams from 6 conferences

Results in data frame `icehockey`
- `visitor` visiting team
- `opponent` usually home team
- `result` 1, 0 or 0.5
KRACH Ratings

Ken’s Ratings for American College Hockey are obtained from a standard Bradley-Terry model with a separate ability for each team:

```r
> standardBT <- BTm(outcome = result,
+    player1 = data.frame(team = visitor),
+    player2 = data.frame(team = opponent),
+    id = "team", formula = ~ team, data = icehockey)
```

The default behaviour provides some simplification

```r
> standardBT <- BTm(outcome = result,
+    player1 = visitor, player2 = opponent,
+    id = "team", data = icehockey)
```
Converting BTm Results to KRACH

The BTabilities function returns the log-abilities and s.e.

```r
> head(BTabilities(standardBT), 4)

                   ability      s.e.
Alaska-Anchorage  0.0000000 0.0000000
Air Force        -1.4135091 0.6560509
Alabama-Huntsville -0.6825408 0.6052347
American Int’l    -2.9316119 0.7121359
```

KRACH rescales the abilities so that \( \text{KRACH} = 100 \Rightarrow \text{RRWP} = 0.5 \)

```r
> KRACH <- exp(BTabilities(standardBT)[,1])*scale
> head(sort(round(KRACH, 1), decr = TRUE))

    Denver    Miami   Wisconsin
543.0     488.2     481.3
North Dakota Boston College St. Cloud State
434.3     346.2     345.3
```
Home Ice Advantage

The official NCAA ranking gives more credit for neutral-site/road wins. Equivalently we can adjust for home ice advantage.

> levelBT <- BTm(result,
+     data.frame(team = visitor, home.ice = 0),
+     data.frame(team = opponent, home.ice = home.ice),
+     ~ team + home.ice, id = "team", data = icehockey)
Effect on Selection?

The 6 regional winners automatically qualify for the NCAA tournament, whilst another 10 are selected by ranking.

<table>
<thead>
<tr>
<th>KRACH</th>
<th>level KRACH</th>
<th>NCAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denver</td>
<td>Denver</td>
<td>Miami</td>
</tr>
<tr>
<td>Miami</td>
<td>Miami</td>
<td>Denver</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Wisconsin</td>
<td>Wisconsin</td>
</tr>
<tr>
<td>St. Cloud State</td>
<td>St. Cloud State</td>
<td>St. Cloud State</td>
</tr>
<tr>
<td>Minnesota Duluth</td>
<td>Bemidji State</td>
<td>Bemidji State</td>
</tr>
<tr>
<td>Northern Michigan</td>
<td>Northern Michigan</td>
<td>Northern Michigan</td>
</tr>
<tr>
<td>Colorado College</td>
<td>New Hampshire</td>
<td>Yale</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>Minnesota Duluth</td>
<td>New Hampshire</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Colorado College</td>
<td>Alaska</td>
</tr>
<tr>
<td>Bemidji State</td>
<td>Vermont</td>
<td>Vermont</td>
</tr>
</tbody>
</table>

(H. Turner and D. Firth (Warwick, UK))
Lizards Data Revisited
Data Structure in R

> str(flatlizards)
List of 2
$ contests : ’data.frame’: 100 obs. of 2 variables:
 ..$ winner: Factor w/ 77 levels "lizard003","lizard005",..: 27 33 12 18 50 49 1 73 30 26 ...
 ..$ loser : Factor w/ 77 levels "lizard003","lizard005",..: 3 6 7 7 7 8 10 10 13 15 ...
$ predictors: ’data.frame’: 77 obs. of 18 variables:
 ..$ id : Factor w/ 77 levels "3","5","6","9",..: 1 2 3 4 ...
 ..$ throat.PC1 : num [1:77] -1.16 -13.19 -12.47 4.75 -13.47 ... 
 ..$ throat.PC2 : num [1:77] 1.066 2.127 -0.771 8.399 -1.968 ... 
 ..$ throat.PC3 : num [1:77] 3.2152 0.8776 -1.6139 0.0786 0.4982 ... 
...

R code:

```r
> liz <- BTm(1, winner, loser, ~throat.PC1[..] + throat.PC3[..] + head.length[..] + SVL[..] + (1|..), data = flatlizards)
```

<table>
<thead>
<tr>
<th></th>
<th>No random effects</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td>lizard096</td>
<td>16.42</td>
<td>36.68</td>
</tr>
<tr>
<td>lizard099</td>
<td>0.84</td>
<td>1.16</td>
</tr>
<tr>
<td>throat.PC1</td>
<td>-0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>throat.PC3</td>
<td>0.34</td>
<td>0.11</td>
</tr>
<tr>
<td>head.length</td>
<td>-1.13</td>
<td>0.49</td>
</tr>
<tr>
<td>SVL</td>
<td>0.19</td>
<td>0.10</td>
</tr>
<tr>
<td>(\sigma_e)</td>
<td></td>
<td>1.1</td>
</tr>
</tbody>
</table>
Main conclusions from original study:

- Overall brightness (PC1) and UV intensity (PC3) of the throat are clearly significant predictors of fight-winning ability.
- PC3 has by far the largest effect: in a contest between lizards at \( \pm 2 \) standard deviations the odds are estimated as \( \approx 30 \) in favour of the lizard with greater UV reflectance on the throat.

Thankfully unaffected by allowing for variation between lizards with the same covariate values!
Summary

BradleyTerry2 can be downloaded from http://cran.r-project.org/package=BradleyTerry2

Package vignette gives further examples.

Further development planned, e.g.

- better handling of ties
- random effects for judges