Statistical inference for the mixed Bradley-Terry model

Mohand Feddag
University of Warwick (UK)

Ongoing joint work: D. Firth (Warwick) & C. Varin (Padova, Italy)

ESRC National Centre for Research Methods

NCRM Working Paper Series

9/06
Statistical inference for the mixed Bradley-Terry model

Mohand Feddag

University of Warwick (UK)

Ongoing joint work: D. Firth (Warwick) & C. Varin (Padova, Italy)
Plan

1. Introduction
   1.1. Model
   1.2. Motivation

2. Estimation by the pairwise likelihood approach

3. Illustrations
   3.1. Simulation study
   3.2. Real data

4. Concluding remarks
1- Introduction

Bradley-Terry models

Data structure

- $N$ players
- Compare “player” $i$ with “player” $j$ in contest $ijt$, $t = 1, \ldots, T_{ij}$

Simplest version: Binary

$$y_{ijt} = \begin{cases} 
1 & \text{if } i \text{ beats } j \\
0 & \text{if } j \text{ beats } i 
\end{cases}$$

Elaborations: ties; margin of victory (ordinal or continuous)
1.1- Bradley-Terry models

**Linear predictor:** \( \eta_{ijt} = g[Pr(y_{ijt} = 1)] \)

- Pure B.T: \( \eta_{ijt} = \alpha_i - \alpha_j \)
- General: \( \eta_{ijt} = \alpha_i - \alpha_j + z_{ijt}^T \gamma \)

\[ e.g. \quad z_{ijt} = \begin{cases} 
1 & i \text{ is at home} \\
-1 & j \text{ is at home} \\
0 & \text{otherwise} 
\end{cases} \]

**Some typical aims**

1. *Rank* the players according to their ability score \( \alpha_i \)

2. *Explain* ability in terms of player-specific covariates \( x_i \):

\[ \alpha_i = x_i^T \beta + u_i, \quad u_i \text{ iid } \sim F_u(\sigma) \]
1.1- Bradley-Terry models

Random effects

- **Aim 1**: \( u_i = \alpha_i \)
  ensures appropriate shrinkage of ability estimates to take account of *imprecision* of estimation (e.g., Efron and Morris, JASA, 1975)

- **Aim 2**: \( u_i = \alpha_i - x_i^T \beta \)
  to represent *unexplained* variation in ability; often not recognised in the literature (e.g., Springall, 1973)

**General model**: The linear predictor has the form

\[
\eta_{ij} = (x_i^T - x_j^T) \beta + u_i - u_j + z_{ij}^T \gamma
\]
1.2- Motivation: Lizards data

Lizards data: Whiting et al. (2006)
1.2- Motivation: Lizards data
1.2- Lizards data: description

- 189 male lizards are captured and explanatory variables are made
- Then released, and *contests* (fights) observed
- 100 contests (winner, loser) were observed involving 77 lizards

- Explanatory variables
  - PC1throat, PC2throat, PC3throat: first 3 PCs of throat spectrum
  - SVL: snout-vent length
  - HL.res, HW.res, HH.res: residual of head length, width, height on SVL
Lizard tournament
1.2- Motivation: Lizards data

**Previous work** [Whiting et al. (2006)] & [D. Firth (2005)]

\[ g[Pr(y_{ij} = 1)] = \sum_{r=1}^{4}(x_{ir} - x_{jr})\beta_r, \]

\[ x_1 = \text{PC1throat}, \quad x_2 = \text{PC3throat}, \quad x_3 = \text{SVL}, \quad x_4 = \text{HL.res} \]

**Present work**

\[ g[Pr(y_{ij} = 1)] = \sum_{r=1}^{4}(x_{ir} - x_{jr})\beta_r + u_i - u_j, \]

\[ u_i \sim N(0, \sigma^2) \]

**Marginal likelihood**

\[ L(\beta, \sigma; y) = \int_{\mathbb{R}^{77}} \prod_{i=1}^{77} \prod_{j=i+1}^{77} Pr(y_{ij} \mid u_i, u_j)\varphi_2(u)du \]
2- Pairwise likelihood

- *Belongs to the composite likelihood class* defined by Lindsay (1988)

**Definition** A rich class of pseudo likelihoods based on the composition likelihood type objects

**Idea** Choose a set of events, write the likelihood for each of them and then take the weighted product

**Examples**

1. Besag’s pseudolikelihood (Besag, JRSSb, 1974)
2. Partial likelihood (Azzalini, BKA, 1983)
3. Composite marginal likelihood
   - Pairwise (Cox and Reid, 2004; Bellio and Varin, 2005)
   - Triplewise likelihood
   - Combination of both them
Motivation

1. Make likelihood type inference in complex models for dependent data
2. Reduce the computational effort and difficulties
3. Gain in statistical robustness with respect to full likelihood
4. In many applications, the cost in efficiency reduction relatively to the full likelihood is moderate
5. Under suitable regularity conditions, the MCLE is consistent and asymptotically normal
Based on all the observed pairs of contests with a common player: 
\((y_{ij}, y_{ij}^*)\)

\[
L_2(\beta, \sigma; y) = \prod_{i=1}^{N} \prod_{\{j<j^*; j,j^* \neq i\}} Pr(y_{ij}, y_{ij}^*; \beta, \sigma)
\]

\[
Pr(y_{ij}, y_{ij}^*; \beta, \sigma) = \int_{\mathbb{R}^3} Pr(y_{ij} | u_i, u_j)Pr(y_{ij}^* | u_i, u_j^*) \varphi(u_i, u_j, u_j^*) du_i du_j du_j^*
\]

1. **Probit link:**

\[
Pr(y_{ij} = 1, y_{ij}^* = 1) = \Phi_2\left( \frac{(x_i-x_j)^T \beta}{\sqrt{1+2\sigma^2}}, \frac{(x_i-x_j^*)^T \beta}{\sqrt{1+2\sigma^2}}; \frac{\sigma^2}{\sqrt{1+2\sigma^2}} \right)
\]

2. **logit link:** Using scale mixture of Drum and McCullagh (1993)

\[
F(t) = \frac{e^t}{1 + e^t} \sim \sum_{i=1}^{k} p_{k,i} \Phi(t s_{k,i}),
\]

where \((s_{k,i}, p_{k,i})\) are known for \(k = 1, \ldots, 8\).
3.1- Illustrations: Simulation study

Simulation

- based on 300 data sets
- 2 sizes: $N = 20, 30$
- $\beta = (\beta_1, \beta_2) = (-1, 2)$
- $\sigma = 0.5, 1, 2$
- $x_1, x_2 \sim N(0, 1)$

**Interest:** mean and standard deviation (sd) of estimators of $\beta$ and $\sigma$
3.1- Simulation study

<table>
<thead>
<tr>
<th>parameter</th>
<th>true</th>
<th>N=20</th>
<th>N=30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>sd</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-1</td>
<td>-1.03</td>
<td>0.27</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>2</td>
<td>2.09</td>
<td>0.39</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.5</td>
<td>0.48</td>
<td>0.19</td>
</tr>
</tbody>
</table>

| $\beta_1$| -1   | -1.04| 0.37 | -1.02| 0.29 |
| $\beta_2$| 2    | 2.10 | 0.55 | 2.08 | 0.41 |
| $\sigma$ | 1    | 1.01 | 0.38 | 0.97 | 0.29 |

| $\beta_1$| -1   | -1.19| 0.64 | -1.12| 0.39 |
| $\beta_2$| 2    | 2.31 | 0.95 | 2.10 | 0.57 |
| $\sigma$ | 2    | 2.09 | 0.81 | 2.08 | 0.45 |
3.2- Real data: Lizards data

1. Structure
   - *Sparse data*: 77 lizards with only 100 contests in total
   - Two lizards with *missing values* in the covariates: liz096, liz099
     (removed from the analysis)
   - Two lizards which *always win* (7 victories for each): liz040, liz073

2. Analysis with liz040 & liz073
   - *Problem* in the optimization due to *infinite values* (very frequent in binary data)
   - liz040, liz073 are the cause

3. Analysis without liz040 & liz073
   **Probit link:**
   \[
   \hat{\beta} = (-0.12, 0.27, 0.25, -0.67), \quad \hat{\sigma} = 0.59, \quad l_2(\hat{\beta}, \hat{\sigma}) = -113.31
   \]
   **Logit link:**
   \[
   \hat{\beta} = (-0.22, 0.49, 0.47, -1.27), \quad \hat{\sigma} = 1.14, \quad l_2(\hat{\beta}, \hat{\sigma}) = -113.20
   \]
4- Concluding remarks

Work in progress on:

1. Solve the problem of the infinite values
   - Optimality of the approach
     - WPL: Weighted pairwise likelihood
     - mixture of $L_2$ and $L_1$
   - Penalized pairwise likelihood

2. Bayesian version of the model using McMC


