SPARSE PAIRED COMPARISONS IN THE BRADLEY-TERRY MODEL

Ting Yan\(^1\), Yaning Yang\(^1\) and Jinfeng Xu\(^2\)

\(^1\)University of Science and Technology of China and \(^2\)National University of Singapore

Abstract: When a group of subjects is repeatedly compared in pairs, the Bradley-Terry model is often used. By assuming that each pair has the same fixed number of comparison(s), Simons and Yao (1999) proved that the maximum likelihood estimates of the parameters retain good asymptotic properties when the number of subjects goes to infinity. In many applications, however, paired comparisons may be sparse and only exist in some pairs. In this note, we show that under a simple condition that controls sparsity, asymptotic results similar to Simons and Yao’s continue to hold. Simulation studies and an application are further provided for the illustration of the sparsity condition and the asymptotic results.

Key words and phrases: Bradley-Terry model, maximum likelihood estimate, paired comparisons, sparsity.

1. Introduction

In a broad range of problems, e.g., the quantification of the influence of statistical journals (Stigler (1994)) or the transmission/disequilibrium test in genetics (Sham and Curtis (1995)), it occurs that a group of subjects are repeatedly compared in pairs. To describe the probabilities of the possible outcomes in paired comparisons, Bradley and Terry (1952) suggested that the probability subject \(i\) beats subject \(j\) be specified by

\[ p_{ij} = \frac{u_i}{u_i + u_j}, \quad (1.1) \]

where the “merit vector” \((u_0, u_1, \ldots, u_t)\) represents the merit parameters of \(t + 1\) subjects.

As discussed in Colonius (1980) and David (1988), the Bradley-Terry model is the only one among the many paired comparison models that satisfies certain desirable properties and hence is often used in practice. It has been generalized in several directions (e.g., Luce (1959); Rao and Kupper (1963); Huang, Weng, and Lin (2006)). For a wide class of generalizations, Hunter (2004) proposed the
iterative minorization-maximization (MM) algorithms for maximum likelihood estimation and established their good convergence properties.

When the total number of subjects, \( t + 1 \), is assumed fixed and the number of comparisons in each pair goes to infinity, the consistency and asymptotic normality of the maximum likelihood estimators are standard. However, in most applications, \( t + 1 \) is quite large while the number of comparisons in each pair is relatively small. Then asymptotics considering the number of merit parameters \( t + 1 \) as going to infinity is more appealing. In contrast to the well-known Neyman-Scott problem [Neyman and Scott (1948)] where the maximum likelihood estimate fails to attain consistency, Simons and Yao (1999) proved that the maximum likelihood estimates of the merit parameters retain good asymptotic properties when each pair has the same fixed number of comparison(s) and the number of subjects \( t + 1 \) goes to infinity. The assumption of the same number of comparisons in each pair used in their paper is stringent but realistic. For example, many basketball conferences under the purview of the NCAA have each pair of teams play each other exactly twice. In some other applications, however, comparisons are applied among a small portion of all possible pairs. As a motivating example, we consider 32 teams in two conferences of the National Football League (NFL). There are eight divisions each consisting of four teams. In the regular season, each team plays 16 matches, 6 within the division and 10 between the divisions. In total, there are \( 32 \times 31/2 = 496 \) pairs but comparisons only exist in \( 32 \times 13/2 = 208 \) pairs. It is therefore interesting to extend Simons and Yao’s results to the situation where paired comparisons are unequal and sparse.

We study the sparsity condition under which the likelihood-based inferences have favorable asymptotic properties when the total number of subjects \( t + 1 \) goes to infinity. The rest of the paper is organized as follows. The main results are given in Section 2. Numerical results are presented in Section 3. Some discussion is given in Section 4. All proofs are relegated to the appendix.

2. Main Results

Let \( a_{ij} \) be the number of times that subject \( i \) beats subject \( j \) as a result of \( n_{ij} \) comparisons. For convenience, \( a_{ii} = 0 \) and \( n_{ii} = 0 \). Then \( a_{ij} + a_{ji} = n_{ij} = n_{ji}, \ i, j = 0, \ldots, t \). In what follows, \( n_{ij} \leq N \) for all \( i, j \), where \( N \) is a fixed positive integer. The likelihood function can be written as

\[
L(u) = \prod_{i, j=0}^{t} n_{ij}^{a_{ij}} \frac{\prod_{i=0}^{t} u_i^{a_{ii}}}{\prod_{0 \leq i < j \leq t} (u_i + u_j)^{n_{ij}}},
\]  

(2.1)
where \( a_i = \sum_{j=0}^{t} a_{ij} \) is the total number of wins of subject \( i \). For identifiability, we set \( u_0 = 1 \) as in Simons and Yao (1999). Denote the maximum likelihood estimates (MLEs) of \( u_1, \ldots, u_t \) by \( \hat{u}_1, \ldots, \hat{u}_t \) and take \( \hat{u}_0 = 1 \). The likelihood equation is
\[
a_i = \sum_{j=0}^{t} \frac{n_{ij} \hat{u}_i}{u_i + \hat{u}_j} \quad i = 1, \ldots, t. \tag{2.2}
\]

By Zermelo (1929) and Ford (1957), the following condition guarantees the existence and uniqueness of the MLEs in (2.1).

**Condition A.** For every partition of the subjects into two nonempty sets, a subject in the second set has beaten a subject in the first at least once.

The matrix \((a_{ij})_{(t+1)\times(t+1)}\) can be regarded as the adjacency matrix of a directed graph and Condition A is equivalent to the strong connectivity of the directed graph. Similarly, the matrix \((n_{ij})_{(t+1)\times(t+1)}\) can be regarded as the adjacency matrix of an undirected graph and, under Condition A, we know that the undirected graph is strongly connected, i.e., for any two subjects \( i \neq j \), there exits a connected path: \( n_{i,l_1}, n_{l_1,l_2}, \ldots, n_{l_m,j} > 0 \).

To ensure that Condition A is satisfied asymptotically under sparse comparisons, we need some conditions for \( n_{ij} \) as well as the merit parameters \( u_{ij} \) that motivate the sparsity condition. Let
\[
n_i = \sum_{j=0}^{t} n_{ij}, \tag{2.3}
\]
\[
C_{ij} = \#\{k : n_{ik} > 0, n_{jk} > 0\} + I(n_{ij} > 0), \quad D_t = \min_{0 \leq i < j \leq t} C_{ij}, \tag{2.4}
\]
where \( I(\cdot) \) is the indicator function, \( C_{ij} \) is the total number of paths between \( i \) and \( j \) with length 2 or 3, and \( D_t \) is the smallest \( C_{ij}, 0 \leq i < j \leq t \). Note that \( D_t \) measures the sparsity level and \( D_t > 0 \) is a sufficient condition for the strong connectivity of the undirected graph. In Simons and Yao (1999), \( D_t = t \) since there exists direct comparisons for each pair. Let
\[
M_t = \max_{i,j=0,\ldots,t} \frac{u_i}{u_j}, \quad \delta_t = \frac{8M_t}{(D_t/t)} \sqrt{\frac{N \log(t + 1)}{t}}, \tag{2.5}
\]
\[
\Delta u_i = \frac{\hat{u}_i - u_i}{u_i}. \tag{2.6}
\]

The following proposition guarantees Condition A.

**Proposition 1.** If \( \lim_{t \to \infty} D_t/t \geq \tau, \tau \in (0, 1] \) and \( M_t = o(t/\log t) \), then \( pr(\text{Condition A holds}) \to 1 \) as \( t \to \infty \).
Now we illustrate the calculation of the sparsity level $D_t/t$. Suppose the $t+1$ subjects are divided into $G$ groups of equal size. There exists direct comparisons for each intra-group pair. The scheme for inter-group pairs is depicted by a graph with $G$ groups as vertices. If two groups are directly connected by a solid line, there exists direct comparisons for any inter-group pair within these two groups. If two groups are connected by a dash line, there exists direct comparisons only for a fixed proportion $0 < q < 1$ of all inter-group pairs within these two groups. If two groups are not directly connected, there exists no direct comparison for any inter-group pair within these two groups. Let $N_s$ and $N_d$ be the numbers of solid and dash lines respectively. In Figure 1, we give three examples. In the left panel, $G = 6$, $N_s = 9$, $N_d = 0$, and $\lim_{t \to \infty} D_t/t = 1/6$. In the middle panel, $G = 8$, $N_s = 8$, $N_d = 0$, and $\lim_{t \to \infty} D_t/t = 1/4$. In the right panel, $G = 8$, $N_s = 8$, $N_d = 8$, and $\lim_{t \to \infty} D_t/t = q/4$. The NFL example we mentioned before is a special case of the right panel example with $t + 1 = 32$ and $q = 1/4$.

First, we establish the consistency of the MLEs uniformly for the merit parameters of all subjects.

**Theorem 1** (uniform consistency). If

$$M_t = o\left(\frac{t}{\log t}\right), \quad \text{and} \quad \lim_{t \to \infty} \frac{D_t}{t} \geq \tau, \, \tau \in (0, 1],$$

(2.7)
then
\[
\max_{i=0,\ldots,t} |\Delta u_i| \leq \max_{i,j=0,\ldots,t} |\Delta u_i - \Delta u_j| = O_p(\delta t) = o_p(1). \tag{2.8}
\]

**Remark 1.** The condition on \(M_t\) is the same as that in Simons and Yao (1999). However, the measure \(\delta t\) on the accuracy of MLES involves \(D_t/t\) and is larger than that in Simons and Yao (1999) due to weaker conditions on \(D_t/t\).

Let \(V_t = (v_{ij})_{i,j=1,\ldots,t}\) denote the covariance matrix of \(a_1, \ldots, a_t\), where
\[
v_{ii} = \sum_{k=0}^{t} \frac{n_{ik} u_i u_k}{(u_i + u_k)^2}, \quad v_{ij} = -\frac{n_{ij} u_i u_j}{(u_i + u_j)^2}, \quad i, j = 0, \ldots, t; j \neq i. \tag{2.9}
\]
Note that \(V_t\) is also the Fisher information matrix for \(\log u_1, \ldots, \log u_t\). Simons and Yao (1999) introduced a \(t \times t\) matrix \(S_t = (s_{ij})\) as an approximation to \(V_t^{-1}\), where
\[
s_{ij} = \frac{\delta_{ij}}{v_{ii}} + \frac{1}{v_{00}}, \quad i, j = 1, \ldots, t, \tag{2.10}
\]
and \(\delta_{ij}\) is the Kronecker delta.

**Theorem 2** (asymptotic normality). If
\[
M_t = o\left(\frac{t^{1/10}}{(\log t)^{1/5}}\right), \quad \text{and} \quad \lim_{t \to \infty} \frac{D_t}{t} \geq \tau, \tau \in (0, 1], \tag{2.11}
\]
then for each fixed \(r \geq 1\), as \(t \to \infty\), the vector \((\Delta u_1, \ldots, \Delta u_r)\) is asymptotically normally distributed with mean 0 and covariance matrix given by the upper left \(r \times r\) block of \(S_t\) as defined at (2.10).

Alternatively, we can define a coarser measure of the sparsity level. Let
\[
A_i = \#\{j : n_{ij} = 0, j \neq i\}, \quad B_t = \max_{i=0,\ldots,t} A_i, \tag{2.12}
\]
where \(A_i\) is the number of subjects that do not have direct comparisons with subject \(i\), and \(B_t\) is the largest \(A_i, i = 0, \ldots, t\). Since
\[
\frac{C_{ij}}{t} \geq 1 - \frac{\#\{k : n_{ik} = 0, n_{jk} > 0\}}{t} - \frac{\#\{k : n_{ik} > 0, n_{jk} = 0\}}{t} \geq 1 - 2\frac{B_t}{t},
\]
we have \(\lim_{t \to \infty} D_t/t \geq 1 - 2\rho > 0\) if \(\lim_{t \to \infty} B_t/t = \rho < 1/2\).

**Corollary 1** (uniform consistency). If
\[
M_t = o\left(\sqrt{\frac{t}{\log t}}\right), \quad \text{and} \quad \frac{B_t}{t} \leq \rho < \frac{1}{2}, \tag{2.13}
\]
then
\[
\max_{i=0,\ldots,t} |\Delta u_i| \leq \max_{i,j=0,\ldots,t} |\Delta u_i - \Delta u_j| = O_p(\delta t) = o_p(1). \tag{2.14}
\]
Corollary 2 (asymptotic normality). If
\[
M_t = o\left(\frac{t^{1/10}}{(\log t)^{1/5}}\right), \quad \text{and} \quad \frac{B_t}{t} \leq \rho < \frac{1}{2},
\]
then for each fixed \( r \geq 1 \), as \( t \to \infty \), the vector \((\Delta u_1, \ldots, \Delta u_r)\) is asymptotically normally distributed with mean 0 and covariance matrix given by the upper left \( r \times r \) block of \( S_t \) as defined at (2.10).

3. Numerical Studies

We conducted simulation studies to evaluate the performance of the maximum likelihood estimates with sparse paired comparisons. By Theorem 2, the asymptotic variances of \( \log(\hat{u}_i/u_i) \) and \( \log(\hat{u}_j/u_i) \) are \( 1/v_{jj} + 1/v_{00} \) and \( 1/v_{ii} + 1/v_{jj} \), which can be estimated by replacing \( u_i \) with \( \hat{u}_i \) in (2.9). Hence 95% confidence interval of \( \log u_i \) and \( \log(u_j/u_i) \) can be constructed accordingly. As a practical matter, the estimates are only available when Condition A does not fail. We report the probabilities that Condition A and coverage both occur, as well as the probabilities that Condition A fails. The average coverage probabilities (ACP) for \( \log u_i \), \( i = 1, \ldots, t \) is calculated to gauge the overall performance of the maximum likelihood estimation. We also list the coverage probabilities for certain pairs. The average length of confidence interval, conditional on that Condition A does not fail, is also reported.

We consider the situation where all subjects are divided into \( G = 8 \) groups of equal size and let \( N = 1 \) (i.e., there is at most one comparison in each pair). There exists direct comparisons for each intra-group pair. For Group 1, the merit parameters of subject \( k \) are \((M_t - 1)k/(G - 1) + 1, k = 1, \ldots, G - 1 \) and the merit parameters of other groups are the same as those of Group 1.

In simulation study 1, the inter-group pair comparison scheme was specified by the middle panel graph in Figure 1 and the sparsity level \( D_t/t = 1/4 \) and \( A_0/t = \cdots = A_t/t = 1/2 \). Thus the sparsity condition was satisfied.

Let \( x = (x_1, \ldots, x_7) \). In simulation study 2, for any inter-group pair \((i, j)\) with subject \( i \) in group \( k \) and subject \( j \) in group \( \ell, 1 \leq k < \ell \leq 8 \), the \( n_{ij} \) were Bernoulli \( Ber(1, x_k) \). Thus, \( EA_i = [(8 - k)(1 - x_k) + \sum_{\ell=1}^{k-1}(1 - x_k)]t/8 \) for each subject \( i \) in group \( k \). We chose \( x = (2/3, 2/3, 2/3, 1/2, 1/2, 1/3, 1/3) \) and \((1/2, 1/2, 1/2, 1/3, 1/3, 1/6, 1/6)\), respectively. By calculation, \( D_t = 0 \) with positive probability and groups vary in \( A_t/t \). We have \( ED_t/t = 0.111 \) and 0.017, respectively.

We chose \( M_t = 1, t^{1/4}, t^{1/2}, t^{3/4}, \) or \( t \). For \( 0 \leq i < j \leq t \), the \( a_{ij} \) were the binomial \( Bin(n_{ij}, p_{ij}) \) and \( a_{ji} = n_{ij} - a_{ij} \). When \( n_{ij} = 0 \), \( a_{ij} \) was set to zero. The results based on 10,000 replications are summarized in Table 1.
Table 1. Coverage probabilities, probabilities that condition A fails (in parentheses), and lengths of confidence intervals (in brackets).

<table>
<thead>
<tr>
<th>t (i,j)/ACP</th>
<th>( M_t = t^{1/4} )</th>
<th>( M_t = t^{3/4} )</th>
<th>( M_t = t^{7/4} )</th>
<th>( M_t = t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>(0,7)</td>
<td>0.947 (0.002)</td>
<td>0.943 (0.011)</td>
<td>0.884 (0.075)</td>
</tr>
<tr>
<td></td>
<td>(30,31)</td>
<td>0.948 (0.002)</td>
<td>0.945 (0.011)</td>
<td>0.894 (0.073)</td>
</tr>
<tr>
<td></td>
<td>(15,16)</td>
<td>0.948 (0.002)</td>
<td>0.940 (0.011)</td>
<td>0.880 (0.073)</td>
</tr>
<tr>
<td></td>
<td>ACP</td>
<td>0.947 (0.002)</td>
<td>0.937 (0.011)</td>
<td>0.878 (0.073)</td>
</tr>
</tbody>
</table>

Simulation study 1: \( D_{t,t} = 3/8 \)

| 79          | (0,7)           | 0.948 (0.012)   | 0.947 (0.011)   | 0.946 (0.002)   | 0.918 (0.040)   | 0.675 (0.299)   |
|             | (30,31)         | 0.954 (0.012)   | 0.951 (0.011)   | 0.950 (0.002)   | 0.914 (0.040)   | 0.659 (0.299)   |
|             | (15,16)         | 0.953 (0.012)   | 0.952 (0.011)   | 0.944 (0.002)   | 0.915 (0.040)   | 0.671 (0.299)   |
|             | ACP             | 0.950 (0.012)   | 0.948 (0.011)   | 0.947 (0.002)   | 0.918 (0.040)   | 0.674 (0.299)   |

Simulation study 2: \( D_{t,t} = 0 \) with positive probability

From Table 1, we see that in simulation study 1, when \( M_t = 1 \) or \( t^{1/4} \), the simulated coverage probabilities are close to the nominal level, but when \( M_t = t^{3/4} \) or \( t \), most results look very bad, especially when \( t = 31 \). The results become worse as \( M_t \) increases. The length of confidence interval also increases as \( M_t \) becomes larger. These observations indicate that the accuracy of the estimation deteriorates as \( M_t \) increases.

In simulation study 2, the sparsity condition fails with positive probability. From Table 1, we observe that the simulated coverage probabilities are quite different from the nominal level in all situations because with positive probability, the undirected graph \((n_{ij})_{(t+1)} \times (t+1)\) is not strongly connected and condition A fails. This is more pronounced when \( M_t \) is larger or when \( D_t = 0 \) with larger probability. On the other hand, the length of confidence interval increases when \( D_t = 0 \) with larger probability which shows that the accuracy of the estimation deteriorates when comparisons are more sparse.
Next, consider the NFL example which corresponds to the right panel in Figure 1 where $t + 1 = 32$ and the sparsity condition is satisfied. The league consists of thirty-two teams that are divided evenly into two conferences and each conference has four divisions that have four teams each. In the regular season, each team plays six games with three intra-division teams and ten games with ten inter-division teams. So $B_t = t - 13 = 31 - 13 = 18$. Accordingly, we have $D_t/t = 2/32 \geq 1/16 > 0$ and $B_t/t = 18/31 = 0.58$. Thus the sparsity condition is satisfied in the theorems but not in the corollaries and this further illustrates that the sparsity condition $D_t/t$ is more useful than $B_t/t$. We used the 2009 NFL regular season data as an example, which can be downloaded from http://en.wikipedia.org/wiki/2009_NFL_season. The fitted merits for the remaining 31 teams are given in Table 2, where we used “Miami Dolphins” as the baseline (with $u_0 = 1$).

It is interesting to compare the ordering of six playoff seeds of the two conferences by the NFL rule with the ordering by their merits obtained from fitting the Bradley-Terry model. The NFL rule can be briefly summarized as follows: the teams in each division with the best regular season won-lost percentage record (PCT) are seeded one through four based on their PCT; another two teams from each conference are seeded five and six based on their PCT. From Table 3, we can see that the ordering of merits of one conference is consistent with that of NFL except that the order of the fourth and fifth seeds is switched. For the other conference, Table 4, we can see that only the top team is seeded according to its merit.

### Table 2. Merits of NFL 2009

<table>
<thead>
<tr>
<th>Division</th>
<th>Team</th>
<th>Merit</th>
<th>Division</th>
<th>Team</th>
<th>Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFC East</td>
<td>New England Patriots</td>
<td>1.980</td>
<td>AFC South</td>
<td>Indianapolis Colts</td>
<td>6.395</td>
</tr>
<tr>
<td></td>
<td>New York Jets</td>
<td>1.481</td>
<td></td>
<td>Houston Texans</td>
<td>1.223</td>
</tr>
<tr>
<td></td>
<td>Miami Dolphins</td>
<td>1.000</td>
<td></td>
<td>Tennessee Titans</td>
<td>1.050</td>
</tr>
<tr>
<td></td>
<td>Buffalo Bills</td>
<td>0.622</td>
<td></td>
<td>Jacksonville Jaguars</td>
<td>0.637</td>
</tr>
<tr>
<td>AFC North</td>
<td>Cincinnati Bengals</td>
<td>1.460</td>
<td>AFC West</td>
<td>San Diego Chargers</td>
<td>4.122</td>
</tr>
<tr>
<td></td>
<td>Baltimore Ravens</td>
<td>1.251</td>
<td></td>
<td>Denver Broncos</td>
<td>1.335</td>
</tr>
<tr>
<td></td>
<td>Pittsburgh Steelers</td>
<td>1.085</td>
<td></td>
<td>Oakland Raiders</td>
<td>0.462</td>
</tr>
<tr>
<td></td>
<td>Cleveland Browns</td>
<td>0.351</td>
<td></td>
<td>Kansas City Chiefs</td>
<td>0.278</td>
</tr>
<tr>
<td>NFC East</td>
<td>Dallas Cowboys</td>
<td>2.142</td>
<td>NFC North</td>
<td>Minnesota Vikings</td>
<td>1.989</td>
</tr>
<tr>
<td></td>
<td>Philadelphia Eagles</td>
<td>2.097</td>
<td></td>
<td>Green Bay Packers</td>
<td>1.414</td>
</tr>
<tr>
<td></td>
<td>New York Giants</td>
<td>1.035</td>
<td></td>
<td>Chicago Bears</td>
<td>0.493</td>
</tr>
<tr>
<td></td>
<td>Washington Redskins</td>
<td>0.194</td>
<td></td>
<td>Detroit Lions</td>
<td>0.063</td>
</tr>
<tr>
<td>NFC South</td>
<td>New Orleans Saints</td>
<td>3.909</td>
<td>NFC West</td>
<td>Arizona Cardinals</td>
<td>1.056</td>
</tr>
<tr>
<td></td>
<td>Atlanta Falcons</td>
<td>1.315</td>
<td></td>
<td>San Francisco 49ers</td>
<td>0.619</td>
</tr>
<tr>
<td></td>
<td>Carolina Panthers</td>
<td>1.143</td>
<td></td>
<td>Seattle Seahawks</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>Tampa Bay Buccaneers</td>
<td>0.227</td>
<td></td>
<td>St. Louis Rams</td>
<td>0.028</td>
</tr>
</tbody>
</table>
Table 3. The six playoff seeds of the AFC.

<table>
<thead>
<tr>
<th>Seed</th>
<th>Team</th>
<th>Won-Lost Percentage</th>
<th>Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Indianapolis Colts</td>
<td>0.875</td>
<td>6.395</td>
</tr>
<tr>
<td>2</td>
<td>San Diego Chargers</td>
<td>0.812</td>
<td>4.122</td>
</tr>
<tr>
<td>3</td>
<td>New England Patriots</td>
<td>0.625</td>
<td>1.980</td>
</tr>
<tr>
<td>4</td>
<td>Cincinnati Bengals</td>
<td>0.625</td>
<td>1.460</td>
</tr>
<tr>
<td>5</td>
<td>New York Jets</td>
<td>0.563</td>
<td>1.481</td>
</tr>
<tr>
<td>6</td>
<td>Baltimore Ravens</td>
<td>0.375</td>
<td>1.251</td>
</tr>
</tbody>
</table>

Table 4. The six playoff seeds of the NFC.

<table>
<thead>
<tr>
<th>Seed</th>
<th>Team</th>
<th>Won-Lost Percentage</th>
<th>Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New Orleans Saints</td>
<td>0.812</td>
<td>3.909</td>
</tr>
<tr>
<td>2</td>
<td>Minnesota Vikings</td>
<td>0.750</td>
<td>1.989</td>
</tr>
<tr>
<td>3</td>
<td>Dallas Cowboys</td>
<td>0.688</td>
<td>2.142</td>
</tr>
<tr>
<td>4</td>
<td>Arizona Cardinals</td>
<td>0.625</td>
<td>1.056</td>
</tr>
<tr>
<td>5</td>
<td>Green Bay Packers</td>
<td>0.688</td>
<td>1.414</td>
</tr>
<tr>
<td>6</td>
<td>Philadelphia Eagles</td>
<td>0.688</td>
<td>2.097</td>
</tr>
</tbody>
</table>

4. Discussion

We study asymptotics of the maximum likelihood estimates in the Bradley-Terry model when the number of subjects goes into infinity and paired comparisons are sparse. With a condition on $M_t$ similar to that in Simons and Yao (1999), we show that the MLEs of merits still have enjoyable properties when the sparse level is controlled by imposing $\lim_{t \to \infty} D_t / t \geq \tau, \tau \in (0, 1]$. Theoretical results show that asymptotics break down when $M_t$ is too large or $D_t$ is too small. This is also verified in our simulation studies that suggest that a proper control of the magnitude of $M_t$ and $D_t$ is crucial in ensuring good estimation accuracy. As noted by a referee, the condition $D_t / t \geq \tau, \tau \in (0, 1]$ is a sufficient but not necessary condition to guarantee Condition A. The sparsity condition $D_t$ only considers the paths between two subjects with length 2 or 3. More generally, it is interesting to control the sparsity by defining $C_{ij}^m = \# \{ (l_1, \cdots, l_m) : n_{i,l_1} > 0, n_{i,l_2} > 0, \cdots, n_{i,l_m,j} > 0 \}$, i.e., the total number of paths between $i$ and $j$ with length $m + 2$, and we will investigate it in future work.

Theorem 1 requires that $M_t = o((t / \log t)^{1/2})$, and the condition on $M_t$ in Theorem 2 is even more stringent. However, simulation studies suggest that the estimation accuracy is still quite good when $M_t = t^{1/2}$. It will be interesting to see if the conditions on $M_t$ in Theorems 1 and 2 can be relaxed.
Acknowledgement

We are very grateful to the Editor, an associate editor, and two referees for helpful comments that have greatly improved the manuscript. Yan and Yang were supported in part by the grants 10801123 and 10971210 from the National Science Foundation of China, respectively, and Xu was supported in part by the grant from the National University of Singapore (R-155-000-112-112).

Appendix

A.1. Proof of Propositions and Theorems

Proof of Proposition 1. Denote the probability that Condition A fails by $P_t$. Note that $M_t \geq 1$ and

$$ \max_{i,j=0,\ldots ,t} p_{ij} = \max_{i,j=0,\ldots ,t} \frac{1}{1 + u_j/u_i} \leq \frac{1}{1 + 1/M_t} \leq \left( \frac{1}{2} \right)^{1/M_t}.$$

Let $\Omega = \{0, 1, \ldots , t\}$ be the set consisting of all $t + 1$ subjects and consider a particular partition of $\Omega$ into two nonempty subset $\Omega_r$ and $\Omega_r^c$, where the subscript $r$ denotes the number of subjects in $\Omega_r$ and $1 \leq r \leq t$. The probability that all the intergroup comparisons between $\Omega_r$ and $\Omega_r^c$ are won by a subject in $\Omega_r$ is bounded above by $\left( \frac{1}{2} \right)^{N_{\Omega_r}/M_t}$, where $N_{\Omega_r} = \sum_{i \in \Omega_r, j \in \Omega_r^c} n_{ij}$. It follows that

$$ P_t \leq \sum_{r=1}^{t} \sum_{\Omega_r \subset \{0, 1, \ldots , t\}} \left( \frac{1}{2} \right)^{N_{\Omega_r}/M_t} \leq 2 \sum_{r=1}^{[t/2]+1} \sum_{\Omega_r \subset \{0, 1, \ldots , t\}} \left( \frac{1}{2} \right)^{N_{\Omega_r}/M_t}.$$

The last inequality holds since the summands $\sum_{\Omega_r \subset \{0, 1, \ldots , t\}} \left( \frac{1}{2} \right)^{N_{\Omega_r}/M_t}$ are symmetric about $(t + 1)/2$. Note that $C_{ij} \geq \tau t$ for all $i \neq j$ and each pair $(i, j)$ with $n_{ij} > 0$ for $i \in \Omega_r, j \in \Omega_r^c, r \leq t/2$, is repeatedly counted at most $t - r$ times among all $C_{ij}, i \in \Omega_r, j \in \Omega_r^c$ since the total number of subjects is $t + 1$. Consequently,

$$ N_{\Omega_r} \geq \sum_{i \in \Omega_r, j \in \Omega_r^c} \frac{C_{ij}}{(t - r)} \geq \tau rt.$$

We have

$$ P_t/2 \leq \sum_{r=1}^{[t/2]+1} \binom{t + 1}{r} \left( \frac{1}{2} \right)^{\tau rt/M_t} \leq (1 + \left( \frac{1}{2} \right)^{\tau M_t})^{t+1} - 1.$$

When $M_t = o(t/\log t)$ and $\tau \in (0, 1]$, this term goes to zero as $t \to \infty$. 


Proof of Theorem 1. Note that \( a_i = \sum_{j=0}^{t} a_{ij} \) is a sum of \( n_i \leq Nt \) independent Bernoulli random variables, and by Hoeffding’s Inequality (Hoeffding (1963)),

\[
P(|a_i - E(a_i)| \geq \sqrt{Nt \log t}) \leq 2 \exp\left(-\frac{2Nt \log t}{n_i} \right) \leq 2 \exp\left(-\frac{2Nt \log t}{Nt} \right) = \frac{2}{t^2}.
\]

It follows that

\[
P\left( \max_{i=0, \ldots, t} |a_i - E(a_i)| \geq \sqrt{Nt \log t} \right) \leq (t+1) \frac{2}{t^2} \to 0, \text{ as } t \to \infty
\]

and, with probability tending to 1,

\[
\max_{i=0, \ldots, t} |a_i - E(a_i)| = \max_{i=0, \ldots, t} \left| \sum_{j=0}^{t} n_{ij} \left( \frac{\hat{u}_i}{\hat{u}_i + \hat{u}_j} - \frac{u_i}{u_i + u_j} \right) \right| \leq \sqrt{Nt \log t}. \quad (A.1)
\]

Let \( d, b \in \{0, \ldots, t\} \) be such that

\[
\hat{\alpha}_t = \max_{j=0, \ldots, t} \frac{\hat{u}_j}{u_j} = \frac{\hat{u}_d}{u_d} \geq \frac{\hat{u}_0}{u_0} = 1,
\]

and

\[
\hat{\beta}_t = \min_{j=0, \ldots, t} \frac{\hat{u}_j}{u_j} = \frac{\hat{u}_b}{u_b} \leq \frac{\hat{u}_0}{u_0} = 1.
\]

Observe that, for \( j = 0, \ldots, t \)

\[
\frac{\hat{u}_d}{\hat{u}_d + \hat{u}_j} - \frac{u_d}{u_d + u_j} = \frac{\hat{u}_d/u_d - \hat{u}_j/u_j}{(\hat{u}_d/u_d + (u_j/u_d)(\hat{u}_j/u_j))(u_d/u_j + 1)} \geq \frac{\hat{\alpha}_t - \hat{u}_j/u_j}{\hat{\alpha}_t(1 + u_j/u_d)(u_d/u_j + 1)} \geq \frac{\hat{\alpha}_t - \hat{u}_j/u_j}{\hat{\alpha}_t(1 + M_t)^2}.
\]

We have

\[
\sum_{j=0}^{t} n_{dj} \left( \frac{\hat{u}_d}{\hat{u}_d + \hat{u}_j} - \frac{u_d}{u_d + u_j} \right) \geq \frac{M_t}{(1 + M_t)^2} \sum_{j=0}^{t} n_{dj} \left( \frac{\hat{\alpha}_t - \hat{u}_j/u_j}{\hat{\alpha}_t} \right). \quad (A.2)
\]

Likewise,

\[
\sum_{j=0}^{t} n_{bj} \left( \frac{\hat{u}_b}{\hat{u}_b + \hat{u}_j} - \frac{u_b}{u_b + u_j} \right) \geq \frac{M_t}{(1 + M_t)^2} \sum_{j=0}^{t} n_{bj} \left( \frac{\hat{u}_j/u_j - \hat{\beta}_t}{\hat{\alpha}_t} \right). \quad (A.3)
\]

Note that,

\[
C_{db} = \# \{ j : n_{dj} > 0, n_{bj} > 0 \} \geq D_t \geq \tau t. \quad (A.4)
\]
By (A.1), (A.2), (A.3), and (A.4),

\[ 2\sqrt{N t \log t} \geq 2 \max_{i=0, \ldots, t} \left| \sum_{j=0}^{t} n_{ij} \left( \frac{\hat{\alpha}_t - \hat{\beta}_t}{\alpha_t} + \frac{\hat{\beta}_t}{\alpha_t} \right) \right| \]

\[ \geq M_t \frac{(1 + M_t)^2}{(1 + M_t)^2} \sum_{j=0}^{t} n_{ij} \left[ \frac{(\hat{\alpha}_t - \hat{\beta}_t)}{\alpha_t} + \frac{\hat{\beta}_t}{\alpha_t} \right] \]

\[ \geq M_t \frac{(1 + M_t)^2}{(1 + M_t)^2} \sum_{j: n_{ij} > 0, n_{ij} > 0} \frac{(\hat{\alpha}_t - \hat{\beta}_t)}{\alpha_t} \]

\[ = \frac{(D_t/t) M_t}{(1 + M_t)^2} \frac{(\hat{\alpha}_t - \hat{\beta}_t)}{\alpha_t} \]

We have

\[ \frac{(\hat{\alpha}_t - \hat{\beta}_t)}{\alpha_t} \leq \frac{2(1 + M_t)^2}{(D_t/t) M_t} \sqrt{N t \log (t + 1)} \leq \frac{8 M_t}{D_t/t} \sqrt{N \log (t + 1) t} = \delta_t. \]

Thus, when (2.13) holds, \( \max_{i=0, \ldots, t} |\Delta u_i| \leq \max_{i=0, \ldots, t} |\Delta u_i - \Delta u_j| = \hat{\alpha}_t - \hat{\beta}_t \leq (\hat{\alpha}_t - \hat{\beta}_t)/\hat{\beta}_t \leq \delta_t/(1 - \delta_t) = o_p(1). \)

Similar to Lemmas 4 and 6 in Simons and Yao (1999), we have the following.

**Lemma A.1.** An upper bound of \( W_t = V_t^{-1} - S_t \) is given by

\[ \| W_t \| \leq \frac{8 N^2 M_t^3}{t^2(D_t/t)^3}, \]

where \( \| A \| = \max_{i,j} |a_{ij}| \) for the matrix \( A = (a_{ij}) \).

**Lemma A.2.** If \( R_t \) is the covariance matrix of \( W_t a \), we have

\[ \| R_t \| \leq \frac{8 N^2 M_t^3(1 + D_t/t)}{t^2(D_t/t)^2}. \]

Write \( a = (a_1, \ldots, a_t)^T \). Since \( a_i \) is a sum of independent bounded random variables, if \( v_{ii} \) diverges, \( a_i - E(a_i) \) is asymptotically normal with variance \( v_{ii} \) (Loève (1977, p.289)) and the following proposition easily follows.

**Proposition 2.** If \( M_t = o(t) \) as \( t \to \infty \) and \( D_t/t \geq \tau > 0 \), then, as \( t \to \infty \), the components of \( a_1 - E(a_1), \ldots, a_r - E(a_r) \) are asymptotically independent and normally distributed with variances \( v_{11}, \ldots, v_{rr} \), respectively, for each fixed integer
r ≥ 1. Moreover, the first r rows of $S_t(a - E(a))$ are asymptotically normal with covariance matrix given by the upper left $r \times r$ block of $S_t$ defined at (2.11), for fixed $r ≥ 1$.

**Proof of Theorem 2.** Let $E_t$ be the event that Condition A holds and $F_t$ be the event that

$$\max_{i=0,\ldots,t} |\Delta u_i| ≤ \delta_t/(1 - \delta_t).$$

By Proposition 1 and Theorem 1, $P(E_t \cap F_t) → 1$ as $t → ∞$ if (2.13) holds. We proceed conditional on the event $E_t \cap F_t$. Let

$$\xi_{ij} = \frac{n_{ij}u_iu_j(\Delta u_i - \Delta u_j)}{(u_i + u_j)^2}, \quad \xi_i = \sum_{j=0, j \neq i}^t \xi_{ij}, \quad i = 1, \ldots, t,$$

$$\xi = (\xi_1, \ldots, \xi_t)^T, \quad \eta = (\eta_1, \ldots, \eta_t)^T = a - E(a) - \xi, \quad \eta_0 = \sum_{j=1}^t \eta_j.$$

Following the proof of Lemma 7 in Simons and Yao (1999), we have

$$|\eta_i| ≤ \frac{\delta_i^2}{(1 - \delta_t)(1 - 2\delta_t)}, \quad i = 0, \ldots, t, \quad (A.5)$$

$$|S_t\eta_i| ≤ \frac{1}{v_{ii}^2} |\eta_i| + \frac{1}{v_{00}^2} |\eta_0| ≤ \frac{2\delta_i^2}{(1 - \delta_t)(1 - 2\delta_t)} = O\left(\frac{M_t^2 \log t}{t(D_t/t)^2}\right). \quad (A.6)$$

Since

$$\frac{D_tM_t}{(M_t + 1)^2} ≤ v_{ii} ≤ \frac{Nt}{4}, \quad i = 0, \ldots, t, \quad (A.7)$$

by (A.5), (A.7), and Lemma A.1,

$$|(W_t\eta)_i| ≤ \frac{8N^2M_t^3}{t^2(D_t/t)^3} \times \sum_{i=1}^t |\eta_i| ≤ \frac{2N^2t^2M_t^3\delta_i^2}{(1 - \delta_t)(1 - 2\delta_t)t^2(D_t/t)^2} = O\left(\frac{M_t^5 \log t}{t(D_t/t)^3}\right). \quad (A.8)$$

By direct calculation, $\xi = V_t\Delta u$, where $\Delta u = (\Delta u_1, \ldots, \Delta u_t)^T$. Thus,

$$\Delta u = V_t^{-1}\xi = V_t^{-1}(a - E(a)) - V_t^{-1}\eta = S_t(a - E(a)) + W_t(a - E(a)) - V_t^{-1}\eta. \quad (A.8)$$
When $M_t = o(t^{1/10}/(\log t)^{1/5})$ and $D_t/t \geq \tau > 0$, $|\mathbf{V}_t^{-1}\eta_i| = o(t^{-1/2})$ and, by Lemma A.2, $|\mathbf{W}_t(a - E\mathbf{a})_i| = o_p(t^{-1/2})$. We have

$$\langle \Delta \mathbf{u} \rangle_i = \langle \mathbf{V}_t^{-1}\xi_i \rangle_i = (S_i(a - E\mathbf{a}))_i + o_p(t^{-1/2}).$$

By (A.7), $v_{ii}v_{00}/(v_{ii} + v_{00}) \leq (v_{00} + v_{ii})/2 \leq Nt/4$ for all $i$ and Theorem 2 follows directly from Proposition 2.

References


Department of Statistics and Finance, University of Science and Technology of China, Anhui, 230026, China.

E-mail: ting0yan@gmail.com; ynyang@ustc.edu.cn

Department of Statistics and Applied Probability, Risk Management Institute, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore.

E-mail: staxj@nus.edu.sg

(Received December 2010; accepted July 2011)