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

$$p_{ij} = \frac{u_i}{u_i + u_j}, \tag{1.1}$$

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 (1967); 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

$$L(\mathbf{u}) = \prod_{\substack{i,j=0 \\ i \neq j}}^{t} p_{ij}^{a_{ij}} = \frac{\prod_{i=0}^{t} u_i^{a_i}}{\prod_{0 \leq i < j \leq t} (u_i + u_j)^{n_{ij}}}, \tag{2.1}$$

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

$$a_i = \sum_{j=0}^{t} \frac{n_{ij}\hat{u}_i}{\hat{u}_i + \hat{u}_j} \quad i = 1, \ldots, t. \tag{2.2}$$

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: $n_{i,l_1} > 0, n_{l_1,l_2} > 0, \cdots, n_{l_m,j} > 0$.

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

$$n_i = \sum_{j=0}^{t} n_{ij}, \tag{2.3}$$

$$C_{ij} = \#\{k : n_{ik} > 0, n_{jk} > 0\} + I(n_{ij} > 0), \quad D_t = \min_{0 \le i < j \le t} C_{ij}, \tag{2.4}$$

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 $C_{ij}, 0 \le i < j \le t$. 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

$$M_t = \max_{i,j=0,\ldots,t} \frac{u_i}{u_j}, \qquad \delta_t = \frac{8M_t}{(D_t/t)}\sqrt{\frac{N\log(t+1)}{t}}, \tag{2.5}$$

$$\Delta u_i = \frac{\hat{u}_i - u_i}{u_i}. \tag{2.6}$$

The following proposition guarantees Condition A.

Proposition 1. If $\lim_{t\to\infty} D_t/t \ge \tau, \tau \in (0,1]$ and $M_t = o(t/\log t)$, then $pr(\text{Condition A holds}) \to 1$ as $t \to \infty$.

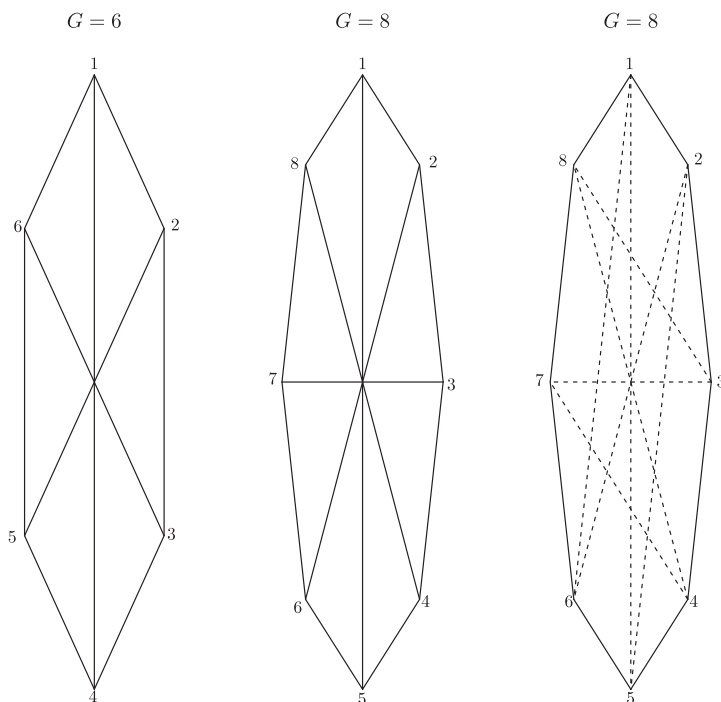$G = 6$ $G = 8$ $G = 8$

Figure 1. The sparse designs.

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*

$$M_t = o(\sqrt{\frac{t}{\log t}}), \quad and \quad \lim_{t\to\infty} \frac{D_t}{t} \geq \tau, \tau \in (0, 1], \tag{2.7}$$

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 δ_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 $\mathbf{V}_t = (v_{ij})_{i,j=1,\ldots,t}$ 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 $\mathbf{S}_t = (s_{ij})$ as an approximation to $\mathbf{V}_t^{-1}$, where

$$s_{ij} = \frac{\delta_{ij}}{v_{ii}} + \frac{1}{v_{00}}, \quad i, j = 1, \ldots, t, \tag{2.10}$$

and δ_{ij} is the Kronecker delta.

Theorem 2 (asymptotic normality). *If*

$$M_t = o\left(\frac{t^{1/10}}{(\log t)^{1/5}}\right), \quad and \quad \lim_{t\to\infty} \frac{D_t}{t} \geq \tau, \tau \in (0, 1], \tag{2.11}$$

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

$$\frac{C_{ij}}{t} \geq 1 - \frac{\#\{k : n_{ik} = 0, n_{jk} > 0\}}{t} - \frac{\#\{k : n_{ik} > 0, n_{jk} = 0\}}{t} \geq 1 - \frac{2B_t}{t},$$

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*

$$M_t = o\left(\sqrt{\frac{t}{\log t}}\right), \quad and \quad \frac{B_t}{t} \leq \rho < \frac{1}{2}, \tag{2.13}$$

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*

$$M_t = o(\frac{t^{1/10}}{(\log t)^{1/5}}), \quad and \quad \frac{B_t}{t} \leq \rho < \frac{1}{2}, \tag{2.15}$$

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 $\log(\hat{u}_j/\hat{u}_i)$ 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 ℓ, $1 \leq k < \ell \leq 8$, the n_{ij} were Bernoulli $Ber(1, x_k)$. Thus, $EA_i = [(8-k)(1-x_k) + \sum_{\ell=1}^{k-1}(1-x_\ell)]t/8$ for each subject i in group k. We chose $\mathbf{x} = (2/3, 2/3, 2/3, 1/2, 1/2, 1/3, 1/3)$ 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).

t	(i,j)/ACP	$M_t = 1$	$M_t = t^{1/4}$	$M_t = t^{1/2}$	$M_t = t^{3/4}$	$M_t = t$
			Simulation study 1: $D_t/t = 3/8$			
31	(0,7)	0.947(0.002)[3.09]	0.943(0.011)[3.23]	0.884(0.073)[3.62]	0.702(0.265)[4.14]	0.485(0.487)[4.69]
	(30,31)	0.948(0.002)[3.09]	0.945(0.011)[3.18]	0.894(0.073)[3.39]	0.700(0.265)[3.61]	0.486(0.487)[3.74]
	(15,16)	0.948(0.002)[3.08]	0.940(0.011)[3.23]	0.880(0.073)[3.61]	0.693(0.265)[4.13]	0.481(0.487)[4.73]
	ACP	0.947(0.002)[3.09]	0.937(0.011)[3.21]	0.878(0.073)[3.62]	0.699(0.265)[4.19]	0.485(0.487)[4.82]
79	(0,7)	0.948(0)[1.82]	0.947(0)[1.92]	0.946(0.002)[2.23]	0.918(0.040)[2.78]	0.675(0.299)[3.62]
	(30,31)	0.954(0)[1.82]	0.951(0)[1.89]	0.950(0.002)[2.01]	0.914(0.040)[2.08]	0.659(0.299)[2.12]
	(15,16)	0.953(0)[1.82]	0.952(0)[1.92]	0.944(0.002)[2.23]	0.915(0.040)[2.78]	0.671(0.299)[3.58]
	ACP	0.950(0)[1.82]	0.948(0)[1.91]	0.947(0.002)[2.22]	0.918(0.040)[2.82]	0.674(0.299)[3.70]
			Simulation study 2: $D_t/t = 0$ with positive probability			
			$(x_1, \cdots, x_7) = (2/3, 2/3, 2/3, 1/2, 1/2, 1/3, 1/3)$			
31	(0,7)	0.918(0.032)[2.57]	0.909(0.044)[2.68]	0.854(0.101)[2.97]	0.731(0.222)[3.38]	0.589(0.362)[3.83]
	(30,31)	0.925(0.032)[2.89]	0.916(0.044)[2.98]	0.864(0.101)[3.15]	0.750(0.222)[3.32]	0.612(0.362)[3.40]
	(15,16)	0.914(0.032)[2.80]	0.902(0.044)[2.92]	0.857(0.101)[3.23]	0.735(0.222)[3.66]	0.591(0.362)[4.10]
	ACP	0.915(0.032)[2.67]	0.906(0.044)[2.78]	0.853(0.101)[3.08]	0.736(0.222)[3.53]	0.596(0.362)[4.05]
			$(x_1, \cdots, x_7) = (1/2, 1/2, 1/2, 1/3, 1/3, 1/6, 1/6)$			
	(0,7)	0.737(0.229)[2.95]	0.708(0.258)[3.08]	0.593(0.376)[3.39]	0.421(0.555)[3.84]	0.271(0.704)[4.34]
	(30,31)	0.742(0.229)[3.41]	0.712(0.258)[3.53]	0.601(0.376)[3.69]	0.427(0.555)[3.84]	0.283(0.704)[3.88]
	(15,16)	0.722(0.229)[3.28]	0.692(0.258)[3.40]	0.589(0.376)[3.73]	0.422(0.555)[4.20]	0.277(0.704)[4.71]
	ACP	0.7250.229)[3.11]	0.698(0.258)[3.23]	0.590(0.376)[3.54]	0.420(0.555)[4.03]	0.275(0.704)[4.60]
			$(x_1, \cdots, x_7) = (2/3, 2/3, 2/3, 1/2, 1/2, 1/3, 1/3)$			
79	(0,7)	0.924(0.017)[1.55]	0.935(0.017)[1.64]	0.928(0.021)[1.87]	0.884(0.064)[2.31]	0.730(0.228)[2.98]
	(30,31)	0.924(0.017)[1.75]	0.930(0.017)[1.80]	0.934(0.021)[1.90]	0.888(0.064)[1.96]	0.734(0.228)[1.97]
	(15,16)	0.932(0.017)[1.67]	0.930(0.017)[1.76]	0.930(0.021)[2.03]	0.901(0.064)[2.50]	0.740(0.228)[3.13]
	ACP	0.934(0.017)[1.61]	0.933(0.017)[1.68]	0.932(0.021)[1.92]	0.891(0.064)[2.39]	0.738(0.228)[3.10]
			$(x_1, \cdots, x_7) = (1/2, 1/2, 1/2, 1/3, 1/3, 1/6, 1/6)$			
	(0,7)	0.785(0.175)[1.76]	0.784(0.178)[1.85]	0.764(0.195)[2.13]	0.649(0.314)[2.67]	0.418(0.559)[3.30]
	(30,31)	0.781(0.175)[2.00]	0.784(0.178)[2.08]	0.769(0.195)[2.20]	0.658(0.314)[2.28]	0.416(0.559)[2.29]
	(15,16)	0.773(0.175)[1.93]	0.770(0.178)[2.04]	0.760(0.195)[2.36]	0.660(0.314)[2.90]	0.422(0.559)[3.63]
	ACP	0.783(0.175)[1.84]	0.777(0.178)[1.92]	0.765(0.195)[2.21]	0.653(0.314)[2.77]	0.421(0.559)[3.45]

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 $M_t = t^{1/2}, t^{3/4}$ 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.

Table 2. Merits of NFL 2009

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

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 3. The six playoff seeds of the AFC.

Seed	team	won-lost percentage	merit
1	Indianapolis Colts	0.875	6.395
2	San Diego Chargers	0.812	4.122
3	New England Patriots	0.625	1.980
4	Cincinnati Bengals	0.625	1.460
5	New York Jets	0.563	1.481
6	Baltimore Ravens	0.375	1.251

Table 4. The six playoff seeds of the NFC.

Seed	team	won-lost percentage	merit
1	New Orleans Saints	0.812	3.909
2	Minnesota Vikings	0.750	1.989
3	Dallas Cowboys	0.688	2.142
4	Arizona Cardinals	0.625	1.056
5	Green Bay Packers	0.688	1.414
6	Philadelphia Eagles	0.688	2.097

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 $C_{ij}^m = \#\{(l_1, \cdots, l_m) : n_{i,l_1} > 0, n_{l_1,l_2} > 0, \cdots, n_{l_m,j} > 0\}$, 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

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

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

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

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

The last inequality holds since the summands $\sum_{\Omega_r \subset \{0,1,\dots,t\}} (1/2)^{N_{\Omega_r}/M_t}$ 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,

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

We have

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

$$\leq (1 + (\frac{1}{2})^{\frac{\tau t}{M_t}})^{t+1} - 1.$$

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)),

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

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 $j = 0, \ldots, t$

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

We have

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

Likewise,

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

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),

$$2\sqrt{Nt\log t} \geq 2\max_{i=0,\dots,t}|\sum_{j=0}^{t} n_{ij}(\frac{\hat{u}_i}{\hat{u}_i+\hat{u}_j} - \frac{u_i}{u_i+u_j})|$$

$$\geq \frac{M_t}{(1+M_t)^2}\sum_{j=0}^{t}[n_{dj}\frac{(\hat{\alpha}_t - \hat{u}_j/u_j)}{\hat{\alpha}_t} + n_{bj}\frac{(\hat{u}_j/u_j - \hat{\beta}_t)}{\hat{\alpha}_t}]$$

$$\geq \frac{M_t}{(1+M_t)^2}\sum_{j:\ n_{dj}>0,n_{bj}>0}[\frac{(\hat{\alpha}_t - \hat{u}_j/u_j)}{\hat{\alpha}_t} + \frac{(\hat{u}_j/u_j - \hat{\beta}_t)}{\hat{\alpha}_t}]$$

$$= \frac{M_t}{(1+M_t)^2}\sum_{j:\ n_{dj}>0,n_{bj}>0}\frac{\hat{\alpha}_t - \hat{\beta}_t}{\hat{\alpha}_t}$$

$$\geq \frac{(D_t/t)tM_t}{(1+M_t)^2}\times\frac{\hat{\alpha}_t - \hat{\beta}_t}{\hat{\alpha}_t}.$$

We have

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

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

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

Lemma A.1. *An upper bound of* $\mathbf{W}_t = \mathbf{V}_t^{-1} - \mathbf{S}_t$ *is given by*

$$\|\mathbf{W}_t\| \leq \frac{8N^2M_t^3}{t^2(D_t/t)^3},$$

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

Lemma A.2. *If* $\mathbf{R}_t$ *is the covariance matrix of* $\mathbf{W}_t\mathbf{a}$, *we have*

$$\|\mathbf{R}_t\| \leq \frac{8N^2M_t^3(1+D_t/t)}{t^2(D_t/t)^2}.$$

Write $\mathbf{a} = (a_1,\dots,a_t)^T$. 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* $a_1 - E(a_1),\dots,a_r - E(a_r)$ *are asymptotically independent and normally distributed with variances* $v_{11},\dots,v_{rr}$, *respectively, for each fixed integer*

$r \geq 1$. *Moreover, the first r rows of $S_t(\mathbf{a} - E(\mathbf{a}))$ are asymptotically normal with covariance matrix given by the upper left $r \times r$ block of S_t defined at (2.10), for fixed $r \geq 1$.*

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

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

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

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

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

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

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

$$|(\mathbf{S}_t \boldsymbol{\eta})_i| \leq \frac{1}{v_{ii}} |\eta_i| + \frac{1}{v_{00}} |\eta_0| \leq \frac{2\delta_t^2}{(1 - \delta_t)(1 - 2\delta_t)} = O\left(\frac{M_t^2 \log t}{t(D_t/t)^2}\right). \tag{A.6}$$

Since

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

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

$$|(\mathbf{W}_t \boldsymbol{\eta})_i| \leq \frac{8N^2 M_t^3}{t^2 (D_t/t)^3} \times \sum_{i=1}^{t} |\eta_i|$$

$$\leq \frac{2N^2 t^2 M_t^3 \delta_t^2}{(1 - \delta_t)(1 - 2\delta_t) t^2 (D_t/t)^2} = O\left(\frac{M_t^5 \log t}{t(D_t/t)^3}\right).$$

By (A.6),

$$|(\mathbf{V}_t^{-1} \boldsymbol{\eta})_i| \leq |(\mathbf{S}_t \boldsymbol{\eta})_i| + |(\mathbf{W}_t \boldsymbol{\eta})_i| = O\left(\frac{M_t^2 \log t}{t(D_t/t)^2}\right) + O\left(\frac{M_t^5 \log t}{t(D_t/t)^3}\right).$$

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

$$\boldsymbol{\Delta u} = \mathbf{V}_t^{-1} \boldsymbol{\xi} = \mathbf{V}_t^{-1} (\mathbf{a} - E\mathbf{a}) - \mathbf{V}_t^{-1} \boldsymbol{\eta} = \mathbf{S}_t (\mathbf{a} - E\mathbf{a}) + \mathbf{W}_t (\mathbf{a} - E\mathbf{a}) - \mathbf{V}_t^{-1} \boldsymbol{\eta}. \tag{A.8}$$

When $M_t = o(t^{1/10}/(\log t)^{1/5})$ and $D_t/t \geq \tau > 0$, $|(\mathbf{V}_t^{-1}\boldsymbol{\eta})_i| = o(t^{-1/2})$ and, by Lemma A.2, $|(\mathbf{W}_t(\mathbf{a} - E\mathbf{a}))_i| = o_p(t^{-1/2})$. We have

$$(\boldsymbol{\Delta}\mathbf{u})_i = (\mathbf{V}_t^{-1}\boldsymbol{\xi})_i = (\mathbf{S}_t(\mathbf{a} - E\mathbf{a}))_i + o_p(t^{-1/2}).$$

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

Bradley, R. A. and Terry, M. E. (1952). The rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika* **39**, 324-345.

Colonius, H. (1980). Representation and uniqueness of the Bradley-Terry-Luce model for pair comparisons. *British J. Math. Statist. Psych.* **33**, 99-103.

David, H. A. (1988). *The Method of Paired Comparisons*. 2nd edition. Griffin, London.

Ford, L. R., Jr. (1957). Solution of the ranking problem from binary comparisons. *Amer. Math. Monthly* **64**, 28-33.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Huang, T. K., Weng, R. C. and Lin, C. J. (2006) Generalized Bradley-terry models and multiclass probability estimates. *J. Machine Learning Research* **7**, 85-115.

Hunter, D. (2004). MM algorithms for generalized Bradley-Terry models. *Ann. Statist.* **32**, 384-406.

Loève, M. (1977). *Probability Theory I*, 4th ed. Springer-Verlag, New York.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, NewYork.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1-32.

Rao, P. V. and Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *J. Amer. Statist. Assoc.* **62**, 194-204. [Corrigendum *J. Amer. Statist. Assoc.* **63**, 1550-1551.]

Sham P. C. and Curtis D. (1995). An extended transmission/disequilibrium test (TDT) for multiallele marker loci. *Ann Hum Genet* **59**, 323-336.

Simons, G. and Yao, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.* **27**, 1041-1060.

Stigler, S. (1994). Citation patterns in the journals of statistics and probability. *Statist. Sci.* **9**, 94-108.

Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Math. Zeit.* **29**, 436-460.

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