The Annals of Applied Probability
2020, Vol. 30, No. 5, 2491–2515
https://doi.org/10.1214/20-AAP1564

ASYMPTOTIC THEORY OF SPARSE BRADLEY–TERRY MODEL

BY RUIJIAN HAN*, ROUGANG YE†, CHUNXI TAN‡ AND KANI CHEN§

Department of Mathematics, Hong Kong University of Science and Technology, *rhanab@connect.ust.hk; †rye@connect.ust.hk; ‡ctanac@connect.ust.hk; §makchen@ust.hk*

The Bradley–Terry model is a fundamental model in the analysis of network data involving paired comparison. Assuming every pair of subjects in the network have an equal number of comparisons, Simons and Yao (*Ann. Statist.* **27** (1999) 1041–1060) established an asymptotic theory for statistical estimation in the Bradley–Terry model. In practice, when the size of the network becomes large, the paired comparisons are generally sparse. The sparsity can be characterized by the probability p_n that a pair of subjects have at least one comparison, which tends to zero as the size of the network n goes to infinity. In this paper, the asymptotic properties of the maximum likelihood estimate of the Bradley–Terry model are shown under minimal conditions of the sparsity. Specifically, the uniform consistency is proved when p_n is as small as the order of $(\log n)^3/n$, which is near the theoretical lower bound $\log n/n$ by the theory of the Erdős–Rényi graph. Asymptotic normality and inference are also provided. Evidence in support of the theory is presented in simulation results, along with an application to the analysis of the ATP data.

1. Introduction. Is Roger Federer a better tennis player than John McEnroe? Is research article A more influential than research article B, among a collection of all research articles in a scientific field? Is webpage A more important than webpage B, among the existing millions of webpages? Is person A more popular than person B in a large social network, such as Twitter users? These questions may be answered by analysis of paired comparison data in a network. The paired comparison may be in terms of head-to-head match outcomes, citation of a research article by another, a webpage containing a link of another webpage, a user retweeting the tweet of another user, etcetera. When the size of the network, such as a total number of webpages, becomes large, paired comparisons are generally sparse. The sparsity may be described in different ways, such as the total number of observed comparisons divided by the total number of subjects in the network. Throughout this paper, the size of the network in study is denoted as n, and we assume any pair has a comparison with probability p_n. The degree of sparsity is then characterized by the size of p_n, the smaller the sparser.

For paired comparison, the Bradley–Terry model (Bradley and Terry (1952)) is one of the most commonly used models. Consider n subjects in a network. Subject i has merit u_i for $i = 0, \dots, n-1$, where $u_i \in \mathbf{R}$ and $u_i > 0$. The Bradley–Terry model assumes the probability that subject i defeats j as

$$p_{ij} = \frac{u_i}{u_i + u_j}, \quad i, j = 0, \dots, n-1; i \neq j. \tag{1.1}$$

The generalizations of the Bradley–Terry model can be seen in, for example, Luce (1959), Rao and Kupper (1967) and Agresti (1990), among many others. For estimation of the merits based on a set of paired comparison data, the maximum likelihood estimation (MLE) is a common choice. It is much desired to justify the asymptotic properties of the MLE, particularly when the comparisons are sparse. A distinct feature of this problem is that the number

Received June 2019; revised October 2019.
MSC2020 subject classifications. Primary 60F05; secondary 62E20, 62F12, 62J15.
Key words and phrases. Bradley–Terry model, sparsity, uniform consistency, asymptotic normality, maximum likelihood estimation.

of parameters, which is the same as the number of subjects, tends to infinity. Moreover, in the case of sparsity, the number of comparisons per pair is 0 with probability tending to 1.

Suppose any pair has a fixed positive number of comparisons, Simons and Yao (1999) proved the uniform consistency and asymptotic normality of the MLE for the Bradley–Terry model. Under this assumption, there would be at least $n(n-1)/2$ comparisons in total. Further extension was reported in Yan, Yang and Xu (2012) with relaxed conditions but still requiring the number of comparisons at the order of n^2. Both papers considered nonsparse cases, where p_n has a positive lower bound and, as a result, does not go to 0. In this paper, we show the asymptotic properties of the MLE for sparse comparisons. In particular, when the maximum ratio of merits are bounded, the uniform consistency holds as long as p_n is greater than $(\log n)^3/n$, and the asymptotic normality is true when p_n is greater than the order of $(\log n)^{1/5}/n^{1/10}$.

The order of $(\log n)^3/n$ required for the uniform consistency is close to the necessary theoretical lower bound, $\log n/n$, below which a unique MLE does not exist. The network we consider can be regarded as the Erdős–Rényi graph $G(n, p_n)$ (Erdős and Rényi (1959)), where each node stands for a subject and each edge stands for the comparison between the two corresponding nodes. Erdős and Rényi (1960) showed that the Erdős–Rényi graph will be disconnected with positive probability if $p_n < \epsilon \log n/n$, for any $\epsilon < 1$. As to be seen, a disconnected graph will fail the condition of the existence and uniqueness of the MLE of Bradley–Terry model, implying that not all subjects are comparable. In the sense of sparsity, the theory established in this paper is nearly optimal.

Although we assume that each pair of subjects in the network has a comparison with the same probability p_n, one can follow our proof to extend it to the case with different comparison probabilities at the order of p_n. The main contribution of this article is to show the asymptotic theorem of MLE when $p_n \to 0$ and how small p_n can be to obtain it.

We note that Negahban, Oh and Shah (2012) and Maystre and Grossglauser (2015) showed the consistency of the MLE under ℓ_2 norm for sparse comparison data. The ℓ_2 norm therein is normalized by $\sqrt{n}$. Since the network size goes to infinity, consistency under the ℓ_2 norm does not ensure the consistency of the merits of any fixed number of subjects. In other words, with their normalized ℓ_2 consistency, one cannot tell for sure that the estimation of merits ratio of any given pair is accurate. In this sense, the uniform consistency is a much desired result.

This paper is organized as follows. In Section 2, we show the large sample properties of the MLE. Simulation results and analysis of the ATP data are given in Section 3. Section 4 contains some concluding remarks. All proofs are relegated to the appendices.

2. Main results. Consider any two subjects i and j with $0 \le i, j \le n-1$. Let t_{ij} denote the number of comparisons between subjects i and j and a_{ij} denote the number of times that i defeats j. Set $a_{ii} = t_{ii} = 0$ for simplicity of notation. Then $t_{ji} = t_{ij} = a_{ij} + a_{ji}$ for all i, j. For slightly more generality, throughout the sequel, we assume t_{ij} follows binomial distribution, $\sim \mathrm{Bin}(T, p_n)$ where T is a positive integer not depending on n. Given t_{ij}, the Bradley–Terry model implies $a_{ij} \sim \mathrm{Bin}(t_{ij}, p_{ij})$, where $p_{ij} = u_i/(u_i + u_j)$. Without loss of generality, one can take T as 1 for ease of understanding.

Based on the observations of paired comparisons $\{a_{ij}, t_{ij} : 0 \le i, j \le n-1\}$, the likelihood function is

$$L(\mathbf{u}) \propto \prod_{\substack{i,j=0 \\ i \ne j}}^{n-1} p_{ij}^{a_{ij}} = \frac{\prod_{i=0}^{n-1} u_i^{a_i}}{\prod_{0 \le i < j \le n-1} (u_i + u_j)^{t_{ij}}}, \tag{2.1}$$

where $a_i = \sum_{j=0}^{n-1} a_{ij}$ is the total number of comparisons that subject i wins and $\boldsymbol{u} = (u_0, u_1, \dots, u_{n-1})$ is the merits vector. By the method of maximum likelihood estimation, the likelihood equations are

$$a_i = \sum_{j=0}^{n-1} \frac{t_{ij}\hat{u}_i}{\hat{u}_i + \hat{u}_j}, \quad i = 0, \dots, n-1, \tag{2.2}$$

where $\hat{\boldsymbol{u}} = (\hat{u}_0, \hat{u}_1, \dots, \hat{u}_{n-1})$ is the MLE of the merits vector u. Since the Bradley–Terry model is invariant under scaling of parameters, we assume that $u_0 = 1$, $\hat{u}_0 = 1$ for the purpose of identifiability.

As noted by Zermelo (1929) and Ford (1957), a necessary and sufficient condition for existence and uniqueness of the MLE is as follows,

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

When Condition A is not satisfied, there exists a nonempty set of subjects, say $\mathcal{A}$, such that the MLEs of the merits of members in $\mathcal{A}$ would be infinitely larger than those of the members not in $\mathcal{A}$. As a result, the MLE cannot be consistent. Under some sparsity conditions given there, Lemma 1 shows Condition A holds with probability approaching 1 as $n \to \infty$. Some more notations are introduced here:

$$\begin{aligned} &M_n = \max_{0 \le i,j \le n-1} \frac{u_i}{u_j}, \qquad \Delta_n = \sqrt{\frac{(\log n)^3}{[\log(np_n)]^2 np_n}} \quad \text{and} \\ &\Delta u_i = \frac{\hat{u}_i - u_i}{u_i}, \quad i = 0, \dots, n-1, \end{aligned} \tag{2.3}$$

where M_n is the largest ratio of u_i and u_j for all i, j, and will be called the largest ratio of merits.

LEMMA 1. *If*

$$\frac{M_n \log n}{np_n} \to 0 \quad \text{as } n \to \infty, \tag{2.4}$$

then P(Condition A is satisfied) $\to 1$ as $n \to \infty$.

REMARK 1. The largest ratio of merits, M_n controls the spread of the merits in the network, while p_n controls the possibility of comparisons. A large M_n and a small p_n both increase the likelihood of the existence of a group of subjects such that any member of this group always wins in a comparison with any member not in this group, and result in Condition A being violated.

Under condition (2.4), p_n can be close to the order of $\log n/n$. Given $T = 1$, $(t_{ij})_{n \times n}$ can be regarded as the adjacency matrix of the Erdős-Rényi graph (Erdős and Rényi (1959)), denoted as $G(n, p_n)$, under our assumption. Erdős and Rényi (1960) showed that if $p_n < \epsilon \log n/n$, for any positive $\epsilon < 1$, $G(n, p_n)$ is disconnected, disagreeing with Condition A, with probability tending to 1. Therefore, in order to satisfy Condition A, it is necessary to require $p_n \ge \log n/n$. According to (2.4), when we fix M_n as a constant, p_n nearly meets the lower bound $\log n/n$.

Condition (2.4) of Lemma 1 ensures the existence and uniqueness of the MLE $(\hat{u}_0, \hat{u}_1, \dots, \hat{u}_{n-1})$. The theorems in this paper all assume conditions that imply (2.4).

2.1. *Uniform consistency.* We first define two notations O_p and o_p to stand for the order of the sequence of random variables. Given a sequence of random variables $\{X_n\}$ and a corresponding sequence of constants $\{a_n\}$. We say that $X_n = O_p(a_n)$ if for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$ such that $P(|X_n/a_n| > M) < \epsilon$ for any $n > N$. Generally speaking, $X_n = O_p(a_n)$ denotes X_n/a_n is stochastically bounded. Another notation $X_n = o_p(a_n)$ means that X_n/a_n converges to zero in probability.

Based on these two notations, we have the following theorem.

THEOREM 2.1. *If*

$$M_n^2 \Delta_n \to 0 \quad \textit{as } n \to \infty, \tag{2.5}$$

then

$$\max_{i=0,\dots,n-1} |\Delta u_i| = O_p\bigl(M_n^2 \Delta_n\bigr) = o_p(1). \tag{2.6}$$

REMARK 2. The condition imposed on the largest ratio of merits M_n and sparsity p_n in (2.5) ensures the uniform consistency of the MLE of the Bradley–Terry model when the comparisons may be sparse and the network is large. For a large value of M_n, the teams with relative poor merits has very little chance to defeat those with relative large merits, thereby making estimation difficult. Meanwhile, for a small p_n, teams have few opportunities to compete with others, thus making a poor estimation.

To prove (2.6), we let

$$i_0 = \arg\max_i \frac{\hat{u}_i}{u_i}, \qquad i_1 = \arg\min_i \frac{\hat{u}_i}{u_i}.$$

Since $\hat{u}_0/u_0 = 1$, it suffices to show that the ratio of subject i_0, $\hat{u}_{i_0}/u_{i_0}$, and the ratio of i_1, $\hat{u}_{i_1}/u_{i_1}$ are very close.

Review that the main idea of the previous work (Simons and Yao (1999), Yan, Yang and Xu (2012)) contains two parts. The first part is that the number of the common neighbors between any two subjects is at least cn for some constant c through their dense assumption. The second part is that for subject $i = i_0$ or i_1, some subjects j with $t_{ij} \neq 0$ have the ratio close to the ratio of i. Then, it can be shown there exists at least one subject, say s (one would suffice), who is a neighbor of i_0 with the ratio $\hat{u}_s/u_s$ close to $\hat{u}_{i_0}/u_{i_0}$ and is also a neighbor of i_1 with the ratio close to $\hat{u}_{i_1}/u_{i_1}$. Such a common neighbor serves as middleman between subjects i_0 and i_1. As a result, $\hat{u}_{i_0}/u_{i_0}$ is close to $\hat{u}_{i_1}/u_{i_1}$. Thus, the uniform consistency holds.

However, in the sparse case, the number of common neighbors of any two subjects tends to 0 as n increases to infinity. If we follow the previous proof directly, no common neighbors of subjects i_0 and i_1 may be found, let alone a common neighbor with desired closeness of its ratio to $\hat{u}_{i_0}/u_{i_0}$ and $\hat{u}_{i_1}/u_{i_1}$. Due to the absence of such a middleman, this approach cannot be applied to the sparse comparison.

It appears to be an obvious extension that one might try to find a chain of subjects, say $l_1, \dots, l_k$, serving as middlemen to bridge the subjects i_0 and i_1. Namely l_{i+1} is a neighbor of l_i and they are close in terms of the ratios. An immediate difficulty, along with other technicalities, arises from the evaluation of this closeness since the previous proof only works for subjects i_0 and i_1. Then, the second extension of our proof is to show for any subject $i = 0, \dots, n-1$, some subjects j with $t_{ij} \neq 0$ have the ratios close to the ratio of i, which is summarized in Lemma 3.

With these two extensions, we prove the uniform consistency by showing the existence of a nonempty intersection between two carefully designed sets. One is the set of subjects having the ratio close to the maximum ratio $\hat{u}_{i_0}/u_{i_0}$. The other is the set of subjects having the ratio close to the minimum ratio $\hat{u}_{i_1}/u_{i_1}$. For the first set, we start from $A = \{i_0\}$ and then constantly expand the size of A through the neighbors of the subjects in A until $|A| > n/2$. Similarly, we can obtain the size of the second set is also larger than $n/2$. Hence there must exist a subject, a middleman, with its ratio close to both $\hat{u}_{i_0}/u_{i_0}$ and $\hat{u}_{i_1}/u_{i_1}$. More details can be found in Remark 5.

The consistency presented in Theorem 2.1 also applies to the special case of nonsparse comparisons, where p_n has a lower bound away from 0, as considered in the previous work. Moreover, the special case of M_n being a constant sheds light on the sparsity required for the uniform consistency. These are summarized in the following corollaries.

COROLLARY 1. *If $M_n = C$ for some constant $C \geq 1$ and there exists an n_0 such that*

$$p_n \geq \frac{(\log n)^3}{n} \tag{2.7}$$

for all $n > n_0$, then

$$\max_{i=0,\ldots,n-1} |\Delta u_i| = O_p\left(\frac{1}{\log(np_n)}\right) = o_p(1). \tag{2.8}$$

COROLLARY 2. *If $p_n = c$ for some constant $c \leq 1$ and*

$$M_n^2 \Delta_n \to 0 \quad \text{as } n \to \infty, \tag{2.9}$$

then

$$\max_{i=0,\ldots,n-1} |\Delta u_i| = O_p\left(M_n^2 \sqrt{\frac{\log n}{n}}\right) = o_p(1). \tag{2.10}$$

REMARK 3. Corollary 1 shows the uniform consistency of the MLE holds when $M_n = C$ and $p_n \geq (\log n)^3/n$, close to the theoretical lower bound $\log n/n$.

2.2. *Asymptotic normality.* Recall that $a_i = \sum_{j=0, j\neq i}^{n-1} a_{ij}$, where a_{ij} is the number of times that subject i prevails over j. Let $\boldsymbol{V}_{n-1} = (v_{ij})_{i,j=1,\ldots,n-1}$ denote the covariance matrix of $a_1, \ldots, a_{n-1}$, where

$$v_{ii} = \sum_{k=0}^{n-1} \frac{t_{ik} u_i u_k}{(u_i + u_k)^2}, \qquad v_{ij} = -\frac{t_{ij} u_i u_j}{(u_i + u_j)^2}, \quad i, j = 1, \ldots, n-1; i \neq j. \tag{2.11}$$

Let $v_{00} = \sum_{i,j=1}^{n-1} v_{ij} = \sum_{k=1}^{n-1} [(t_{0k} u_k)/(1 + u_k)^2]$. Here $\boldsymbol{V}_{n-1}$ is the Fisher information matrix for the parameterization $(\log u_1, \ldots, \log u_{n-1})$. Simons and Yao (1999) used a symmetric matrix $\boldsymbol{S}_{n-1} = (s_{ij})_{(n-1)\times(n-1)}$ to approximate $\boldsymbol{V}_{n-1}^{-1}$, where

$$s_{ij} = \frac{\delta_{ij}}{v_{ii}} + \frac{1}{v_{00}}, \quad i, j = 1, \ldots, n-1, \tag{2.12}$$

and δ_{ij} is the Kronecker delta. With sparse comparisons, we shall re-evaluate the accuracy of this approximation in Lemma 7. The following theorem shows the asymptotic normality of the MLE.

THEOREM 2.2. *If*

$$\frac{M_n}{p_n} \frac{(\log n)^{1/5}}{n^{1/10}} \to 0 \quad \text{as } n \to \infty, \tag{2.13}$$

then for each fixed $r \geq 1$, as $n \to \infty$, 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 $\boldsymbol{S}_{n-1}$ defined in (2.12).

As expected, condition (2.13) involves M_n and p_n. The following two corollaries each deal with the special cases of M_n and p_n being constants.

COROLLARY 3. *If $p_n = c$ for some constant $c \le 1$ and*

$$M_n \frac{(\log n)^{1/5}}{n^{1/10}} \to 0 \quad \text{as } n \to \infty, \tag{2.14}$$

then for each fixed $r \ge 1$, as $n \to \infty$, 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 $\mathbf{S}_{n-1}$ defined in (2.12).

COROLLARY 4. *If $M_n = C$ for some constant $C \ge 1$ and*

$$p_n \frac{n^{1/10}}{(\log n)^{1/5}} \to \infty \quad \text{as } n \to \infty, \tag{2.15}$$

then for each fixed $r \ge 1$, as $n \to \infty$, 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 $\mathbf{S}_{n-1}$ defined in (2.12).

REMARK 4. Corollary 3 is the theorem about asymptotic normality presented in Simons and Yao (1999) and Yan, Yang and Xu (2012), and Corollary 4 gives the sparsity condition to ensure asymptotic normality when the largest ratio of merits is bounded above.

3. Numerical studies.

3.1. *Simulation.* Simulations are carried out to evaluate the finite sample performance of the MLE of the Bradley–Terry model. We assume $T = 1$ in all simulations, which means that any pair has one comparison with probability p_n and no comparison with probability $1 - p_n$.

The result of the first simulation study is shown in Table 1. In order to study the uniform consistent tendency of MLE, we present the mean and median of $\max_{i=0,\ldots,n-1} |\Delta u_i|$ based on 1000 repetitions. In this simulation, the size of network n is taken to be 1000, 2000, 5000, 10,000, 15,000, the sparse probability p_n is chosen as $\log n/n$, $(\log n)^3/n$, $10/\sqrt{n}$ and M_n equals to 1, which implies merits of all subjects are identical. When $p_n = \log n/n$, all the repetitions do not produce the MLE since Condition A is not satisfied. In the case of $p_n = (\log n)^3/n$, both mean and median of $\max_{i=0,\ldots,n-1} |\Delta u_i|$ become closer to 0 with increasing n. For comparison, we also consider p_n as large as $10/\sqrt{n}$. We multiply $1/\sqrt{n}$ by 10 to ensure there are more paired comparisons than $p_n = (\log n)^3/n$ for values of n in the simulation. The result shown in Table 1 indeed indicates the consistency of the MLE under

TABLE 1

The mean and median (in parentheses) of $\max_{i=0,\ldots,n-1} |\Delta u_i|$. In the third column, "–" means all repetitions fail Condition A. The three numbers in the parentheses in the first column represent respectively the average numbers of comparisons one subject has for $p_n = \log n/n$, $p_n = (\log n)^3/n$ and $p_n = 10/\sqrt{n}$

n	M_n	$p_n = \log n/n$	$p_n = (\log n)^3/n$	$p_n = 10/\sqrt{n}$
1000 (7/330/316)	1	–	0.4784 (0.4365)	0.4862 (0.4427)
2000 (8/439/447)	1	–	0.4291 (0.4012)	0.4242 (0.3929)
5000 (9/618/707)	1	–	0.3765 (0.3593)	0.3511 (0.3327)
10,000 (9/781/1000)	1	–	0.3423 (0.3223)	0.2988 (0.2867)
15,000 (10/889/1225)	1	–	0.3226 (0.3085)	0.2727 (0.2595)

TABLE 2
Coverage probabilities and the probabilities that Condition A fails (in parentheses). In the first column, the numbers in the parentheses represent the average numbers of comparisons one subject has

n	(i, j)	$M_n = 1$	$M_n = \sqrt{n}$	$M_n = n$
		$p_n = 1/\sqrt{n}$		
100(10)	(0, 1)	0.277 (0.703)	0.041 (0.958)	0.001 (0.999)
	(0, 99)	0.275 (0.703)	0.041 (0.958)	0.001 (0.853)
	(50, 51)	0.278 (0.703)	0.039 (0.958)	0.001 (0.853)
200(14)	(0, 1)	0.697 (0.260)	0.124 (0.871)	0.001 (0.999)
	(0, 199)	0.696 (0.260)	0.123 (0.871)	0.001 (0.999)
	(100, 101)	0.692 (0.260)	0.120 (0.871)	0.001 (0.999)
500(23)	(0, 1)	0.932 (0.010)	0.416 (0.562)	0.002 (0.998)
	(0, 499)	0.931 (0.010)	0.415 (0.562)	0.002 (0.998)
	(250, 251)	0.930 (0.010)	0.410 (0.562)	0.002 (0.998)
		$p_n = \sqrt{\log n/n}$		
100(21)	(0, 1)	0.938 (0.003)	0.888 (0.070)	0.501 (0.483)
	(0, 99)	0.940 (0.003)	0.886 (0.070)	0.491 (0.483)
	(50, 51)	0.943 (0.003)	0.877 (0.070)	0.483 (0.483)
200(33)	(0, 1)	0.944 (0)	0.941 (0.011)	0.710 (0.262)
	(0, 199)	0.947 (0)	0.939 (0.011)	0.696 (0.262)
	(100, 101)	0.942 (0)	0.931 (0.011)	0.693 (0.262)
500(56)	(0, 1)	0.949 (0)	0.949 (0)	0.897 (0.062)
	(0, 499)	0.945 (0)	0.948 (0)	0.886 (0.062)
	(250, 251)	0.947 (0)	0.944 (0)	0.890 (0.062)
		$p_n = 1$		
100(99)	(0, 1)	0.951 (0)	0.954 (0)	0.954 (0)
	(0, 99)	0.952 (0)	0.954 (0)	0.950 (0)
	(50, 51)	0.941 (0)	0.948 (0)	0.948 (0)
200(199)	(0, 1)	0.954 (0)	0.954 (0)	0.942 (0)
	(0, 199)	0.953 (0)	0.957 (0)	0.951 (0)
	(100, 101)	0.950 (0)	0.946 (0)	0.951 (0)
500(499)	(0, 1)	0.951 (0)	0.953 (0)	0.949 (0)
	(0, 499)	0.953 (0)	0.959 (0)	0.950 (0)
	(250, 251)	0.955 (0)	0.948 (0)	0.950 (0)

the sparsity condition given in Theorem 2.1. This further supports that our sparsity condition nearly meets the lower bound of p_n to ensure the existence of a unique MLE.

The second simulation is done to measure the coverage probabilities of MLE and the result based on 5000 repetitions is given in Table 2. By applying Theorem 2.2, we construct the approximate $1-\alpha$ confidence interval for $\log(u_i/u_j)$ as

$$\log(\hat{u}_i/\hat{u}_j) \pm z_{1-\alpha/2}\sqrt{1/\hat{v}_{ii} + 1/\hat{v}_{jj}},$$

where $z_{1-\alpha/2}$ refers to the quantile of the standard normal distribution at level $1-\alpha/2$. The asymptotic variances are based on (2.11), and $\hat{v}_{ii}$ and $\hat{v}_{jj}$ are computed using $\hat{u}$, the MLE of merits u. We present the coverage probabilities of 95% confidence intervals of some pairs of merits (the first two merits, the middle two merits, the first and the last merits when they are sorted in ascending order) when Condition A is met. The frequencies that Condition A fails are also reported. In this simulation, we choose the size of network $n = 100, 200, 500$, the sparse probability $p_n = 1/\sqrt{n}$, $\sqrt{\log n/n}$, 1 and the largest ratio of merits $M_n = 1, \sqrt{n}$.

TABLE 3
Coverage probabilities and the probabilities that Condition A fails (in parentheses) when $M_n = 1$ and $n = 1000, 2000$. In the first column, the two numbers in the parentheses represent the numbers of comparisons one subject has for $p_n = \sqrt{\log n/n}$ and $p_n = \log n/\sqrt{n}$

n	(i, j)	$p_n = \sqrt{\log n/n}$	$p_n = \log n/\sqrt{n}$
1000 (83/218)	(0, 1)	0.942 (0)	0.938 (0)
	(0, 999)	0.947 (0)	0.943 (0)
	(500, 501)	0.943 (0)	0.946 (0)
2000 (123/340)	(0, 1)	0.945 (0)	0.954 (0)
	(0, 1999)	0.948 (0)	0.950 (0)
	(1000, 1001)	0.940 (0)	0.956 (0)

The average numbers of comparisons one subject has under different sparse probabilities are shown in the parentheses following the numbers of subjects n in Table 2. For example, there are only around 20 and 50 comparisons for each of 500 subjects in two cases with $p_n = 1/\sqrt{n}$ and $\sqrt{\log n/n}$ respectively.

From Table 2, we see coverage probabilities become closer to the nominal level as p_n increases or M_n decreases. With n increasing, the coverage probabilities approach the nominal level and the probabilities that Condition A fails decrease. These phenomena agree with the theoretical asymptotic properties given in Theorem 2.2. Condition A fails mostly when $p_n = 1/\sqrt{n}$ and $M_n = n$, due to extremely sparse comparisons and the large range of merits.

Furthermore, Table 3 reports coverage probabilities for larger network, with $n = 1000$ and 2000. In this simulation, we let $p_n = \sqrt{\log n/n}$, $\log n/\sqrt{n}$ and $M_n = 1$, which means all subjects have equal merits. One can conclude that the coverage probabilities are close to the nominal level from Table 3, showing evidence in support of the theory.

3.2. *The ATP dataset.* We present results of the Bradley–Terry model applied to the 2017 ATP data and then compare its ranking with the official ATP ranking. The ATP matches of one year include four Grand Slams, the ATP World Tour Masters 1000, the ATP World Tour 500 series and other tennis series of the year. There are 203 players in total after removing those who never win or lose for Condition A to be satisfied. Besides, we exclude walkovers and only consider finished games. The results are reported in Table 4. The estimated merits of

TABLE 4
Results of the analysis of the ATP 2017 data

Rank	Player	Games	Winning rate	Merit	ATP Ranking
1	Roger Federer	55	0.909	7.505	2
2	Rafael Nadal	76	0.855	4.085	1
3	Novak Djokovic	36	0.806	2.029	12
4	Juan Martin del Potro	53	0.698	1.440	11
5	Alexander Zverev	73	0.712	1.321	4
6	Grigor Dimitrov	65	0.708	1.303	3
7	Nick Kyrgios	38	0.684	1.287	21
8	Milos Raonic	39	0.718	1.136	24
9	Stan Wawrinka	36	0.694	1.043	9
10	Kei Nishikori	42	0.714	1.000	22

TABLE 5
Results of the analysis of the ATP 1968–2016 *data*

Rank	Player	Games	Winning rate	Merit
1	Novak Djokovic	880	0.831	2.788
2	Rafael Nadal	968	0.818	2.286
3	Roger Federer	1287	0.814	2.128
4	Andy Murray	782	0.776	1.744
5	Ivan Lendl	1274	0.820	1.247
6	Pete Sampras	957	0.770	1.128
7	Andy Roddick	784	0.741	1.111
8	John McEnroe	1035	0.813	1.104
9	Juan Martin del Potro	481	0.709	1.066
10	Andre Agassi	1101	0.756	1.000

Bradley–Terry model are given in the fifth column, and, based on which, the ranks are given in the first column. The number of games played in 2017, winning rates and ATP rankings are also presented. The 10th player, Kei Nishikori, is taken as the baseline ($u_0 = \hat{u}_0 = 1$). We note that there is a difference between ranking by the estimated merits and the ATP ranking. For example, the 7th player in our ranking list, Nick Kyrgios, ranked 21st in the ATP ranking. Yet he defeated Novak Djokovic twice and Rafael Nadal once in 2017. These will be counted heavily in the Bradley–Terry model and less so in the points calculation which the ATP ranking is based on. Notice that Roger Federer and Rafael Nadal have reversed order in the two ranking systems. In fact, Rafael Nadal had 6 winners and 4 runners-up and more ATP points in 2017 than Roger Federer, who had 7 winners and 1 runner-up. On the other hand, Federer defeated Nadal four times in 2017 and had an outstanding winning rate. As a result, the estimated merit of Federer is higher than that of Nadal.

Moreover, we applied the Bradley–Terry model to the ATP matches from 1968 to 2016. There are 2877 players in total after data cleaning. All players are ranked by their estimated merits, and top 10 of them are presented in Table 5. The 10th player, Andre Agassi, is taken as the baseline. The Big Four, Novak Djokovic, Rafael Nadal, Roger Federer and Andy Murray, rank top four in the ranking list. They are considered dominant in terms of ranking and the tournament victories from 2004 onwards. With this dataset and the application of the Bradley–Terry model, the estimated chance that Federer defeats McEnroe in a hypothetical match is $2.128/(2.128+1.104) = 0.6584$, even though the two had very similar winning rates in reality. We remark that the correctness of this answer is limited by the assumption that the players' merits are fixed and may be viewed as averaged over time. Further analysis using more sophisticated dynamic models, such as the whole history rating (Coulom (2008)), may be more appropriate. The static model in this paper serves as a basis for further extensions.

4. Discussion. This paper provides an asymptotic theory of the MLE of the Bradley–Terry model when comparisons between any pair of subjects are sparse. The uniform consistency and asymptotic normality of the MLE are shown under, respectively, conditions (2.5) and (2.13). Two quantities, the largest ratio of merits, M_n, and the probability of paired comparison, p_n, contribute to the accuracy of the MLE. When M_n are bounded, we prove the uniform consistency holds under nearly minimal condition of sparsity. The results of this paper may have broad applications and can be further generalized to other models in the sparse case, such as Plackett–Luce model (Luce (1959)) which fits multiple comparisons, the Rao–Kupper model (Rao and Kupper (1967)) which allow paired comparisons with ties.

APPENDIX A: PROOF OF LEMMA 1

PROOF. Let E_n denote the event that Condition A holds. We will show that under condition (2.4), $P(E_n^c) \to 0$ as $n \to \infty$, that is, the probability that the subjects in the second group never defeat the subjects in the first group, for all partitions of subjects into two nonempty groups, tends to 0. The proof consists of two steps. Step 1 is about estimating the number of comparisons between the first and second groups. Step 2 is about computing the probability of E_n^c and its convergence to 0.

Step 1. Let $\Omega = \{0, 1, \ldots, n-1\}$ be the set of all n subjects, and let Ω_r denote any subset of Ω with r subjects. The number of comparisons between Ω_r and Ω_r^c, denoted by N_{Ω_r}, can be expressed as

$$N_{\Omega_r} = \sum_{i \in \Omega_r, j \in \Omega_r^c} t_{ij}.$$

Recall that t_{ij} is the number comparisons between subjects i and j, which follows binomial distribution. N_{Ω_r} can be viewed as the sum of $Tr(n-r)$ independent identically distributed Bernoulli random variables with common probability ratio p_n.

We first estimate N_{Ω_r} for a fixed r. Condition (2.4) implies $\log n/(np_n) \to 0$ as $n \to \infty$, since $M_n \geq 1$. It follows from the Chernoff bound (Chernoff (1952)) that, for a fixed $r \in \{1, \ldots, \lfloor n/2 \rfloor\}$ and $n > 32 \log n/(Tp_n)$,

$$\begin{aligned}
P\Big(\min_{\Omega_r} N_{\Omega_r} \leq \frac{T}{2} r(n-r)p_n\Big) &\leq \sum_{\Omega_r} P\Big(N_{\Omega_r} \leq \frac{T}{2} r(n-r)p_n\Big) \\
&\leq \binom{n}{r} \sup_{\Omega_r} P\Big(N_{\Omega_r} \leq \frac{T}{2} r(n-r)p_n\Big) \\
&\leq n^r \sup_{\Omega_r} P\Big(N_{\Omega_r} \leq \frac{T}{2} r(n-r)p_n\Big) \\
&\leq \exp\Big\{-\frac{T}{8} r(n-r)p_n + r \log n\Big\} \\
&\leq \exp\Big\{-\frac{T}{16} r(n-r)p_n\Big\} \\
&\leq \exp\Big\{-\frac{T}{16} (n-1)p_n\Big\}.
\end{aligned}$$

The next-to-last inequality holds due to $n > 32 \log n/(Tp_n)$. The definition of N_{Ω_r}, implies the symmetry: $\min_{\Omega_r} N_{\Omega_r} = \min_{\Omega_{n-r}} N_{\Omega_{n-r}}$. Therefore, for any fixed $r \in \{1, \ldots, n-1\}$ and $n > 32 \log n/(Tp_n)$,

$$P\Big(\Big(\min_{|\Omega_r|=r} N_{\Omega_r}\Big) \leq \frac{T}{2} r(n-r)p_n\Big) \leq \exp\Big\{-\frac{T}{16}(n-1)p_n\Big\}.$$

Next, we estimate the lower bound N_{Ω_r} for all $r \in \{1, \ldots, n-1\}$. Let F_n be the event that $N_{\Omega_r} > Tr(n-r)p_n/2$ holds for all $r = 1, \ldots, n-1$ and all partitions of Ω into Ω_r and Ω_r^c.

Then, and $n > 32\log n/(Tp_n)$,

$$
\begin{aligned}
P(F_n) &\geq 1 - \sum_{r=1}^{n-1} P\Big(\Big(\min_{|\Omega_r|=r} N_{\Omega_r}\Big) \leq \frac{T}{2} r(n-r)p_n\Big) \\
&\geq 1 - \sum_{r=1}^{n-1} \exp\Big\{-\frac{T}{16}(n-1)p_n\Big\} \\
&\geq 1 - \exp\Big\{-\frac{T}{16}(n-1)p_n + \log(n-1)\Big\}.
\end{aligned}
$$

Therefore, $P(F_n) \to 1$ as $n \to \infty$, since Condition (2.4) ensures $n > 32\log n/(Tp_n)$ for all large n.

Step 2. Since $M_n = \max_{0\leq i,j\leq n-1} u_i/u_j \geq 1$,

$$
\max_{0\leq i,j\leq n-1} p_{ij} = \max_{0\leq i,j\leq n-1} \frac{1}{1+u_j/u_i} \leq \frac{1}{1+1/M_n} \leq \Big(\frac{1}{2}\Big)^{1/M_n}. \tag{A.1}
$$

Let $G_n^{(r)}$ denote the event that Condition A fails with the first group containing r subjects. Then, by the definition of F_n,

$$
\begin{aligned}
P\big(G_n^{(r)}|F_n\big) &\leq \sum_{\Omega_r}\Big(\max_{0\leq i,j\leq n-1} p_{ij}\Big)^{\frac{Tr(n-r)p_n}{2}} \\
&\leq \sum_{\Omega_r}\Big(\frac{1}{2}\Big)^{\frac{Tr(n-r)p_n}{2M_n}} = \binom{n}{r}\Big(\frac{1}{2}\Big)^{\frac{Tr(n-r)p_n}{2M_n}}.
\end{aligned}
$$

Recall that E_n^c is the event that Condition A fails. Write

$$
\begin{aligned}
P\big(E_n^c|F_n\big) &= P\Big(\bigcup_{r=1}^{n-1} G_n^{(r)}\Big| F_n\Big) = \sum_{r=1}^{n-1} P\big(G_n^{(r)}|F_n\big) \\
&\leq \sum_{r=1}^{n-1}\binom{n}{r}\Big(\frac{1}{2}\Big)^{\frac{Tr(n-r)p_n}{2M_n}} \leq 2\sum_{r=1}^{\lfloor n/2\rfloor}\binom{n}{r}\Big(\frac{1}{2}\Big)^{\frac{Tr(n-r)p_n}{2M_n}} \\
&\leq 2\sum_{r=1}^{\lfloor n/2\rfloor}\binom{n}{r}\Big(\frac{1}{2}\Big)^{\frac{Trnp_n}{4M_n}} \leq 2\Big[\Big(1+\Big(\frac{1}{2}\Big)^{\frac{Tnp_n}{4M_n}}\Big)^n - 1\Big],
\end{aligned}
$$

which tends to 0 as $n \to \infty$, ensured by Condition (2.4). With the law of total probability,

$$
P\big(E_n^c\big) = P\big(E_n^c|F_n\big)P(F_n) + P\big(E_n^c|F_n^c\big)P\big(F_n^c\big),
$$

where F_n^c is the complementary event of F_n. Since $P(E_n^c|F_n) \to 0$ and $P(F_n) \to 1$ as $n \to \infty$, it follows that $P(E_n^c) \to 0$ and $P(E_n) \to 1$ as $n \to \infty$. The proof is complete. □

APPENDIX B: PROOF OF UNIFORM CONSISTENCY

In this appendix, we show the proof of uniform consistency described in Theorem 2.1. Some notations need to be introduced first.

We first make a transformation on $\hat{u}_i$ for $i = 0, \dots, n-1$. Let $\tilde{u}_j = \hat{u}_j / (\max_i \hat{u}_i / u_i)$. As a result, $\max_{0 \le i \le n-1} \tilde{u}_i / u_i = 1$ and we set $\tilde{u}_{i_0} / u_{i_0} = 1$. In addition, let

$$\text{(B.1)} \qquad \begin{aligned} &K = \left\lfloor \frac{2\log n}{\log(np_n)} \right\rfloor, \qquad \phi_n = \frac{\log(np_n)}{\log n}, \qquad q_n = \frac{\phi_n M_n}{(1+M_n)^2 T + \phi_n M_n}, \\ &C_i = \{j : t_{ij} > 0\}, \qquad d_{ij} = I(t_{ij} > 0), \qquad d_i = \sum_{j=0}^{n-1} d_{ij} \quad \text{and} \quad t_i = \sum_{j=0}^{n-1} t_{ij}, \end{aligned}$$

where $\lfloor \cdot \rfloor$ is the floor function and $I(\cdot)$ is the indicator function. We also need a sequence of increasing number $\{D_k\}_{k=1}^K$ to present the level of the closeness in terms of the ratios,

$$\begin{aligned} D_k &= \beta(1+\phi_n)^k M_n \Delta_n \quad \text{for } k = 0, \dots, K-1, \\ D_K &= 40\beta T (1+\phi_n)^K M_n^2 \Delta_n, \end{aligned}$$

and a sequence of nested or increasing sets $\{A_k\}_{k=1}^K$ to collect the subjects which have ratios close to $\max_i \hat{u}_i / u_i$ in the level of D_k,

$$\text{(B.2)} \qquad \begin{aligned} A_k &= \left\{ j : \frac{\tilde{u}_j}{u_j} \ge 1 - \beta(1+\phi_n)^k M_n \Delta_n \right\} \quad \text{for } k = 0, \dots, K-1, \\ A_K &= \left\{ j : \frac{\tilde{u}_j}{u_j} \ge 1 - 40\beta T (1+\phi_n)^K M_n^2 \Delta_n \right\}, \end{aligned}$$

where the constant $\beta = 20T$.

To prove Theorem 2.1, we need four additional lemmas, whose proofs are given after the proof of Theorem 2.1. For ease of illustration, we use the sentence that "for all large n, the condition S_n holds" to stand for that there exists n_0 such that S_n holds for all $n > n_0$.

LEMMA 2. *Assume condition* (2.5) *holds. For all large* n,

$$\text{(B.3)} \qquad P\left(\max_{0 \le i \le n-1} |Z_i^+ - Z_i^-| < 2\sqrt{\frac{\log n}{np_n}} \right) \ge 1 - 3n^{-3},$$

where

$$\text{(B.4)} \qquad \begin{aligned} Z_i^+ &= \frac{1}{t_i} \sum_{j \in C_i^+} t_{ij} \left(\frac{t_{ij}\tilde{u}_i}{\tilde{u}_i + \tilde{u}_j} - \frac{t_{ij} u_i}{u_i + u_j} \right) = \frac{1}{t_i} \sum_{j \in C_i^+} t_{ij} \frac{\tilde{u}_i u_j - \tilde{u}_j u_i}{(\tilde{u}_i + \tilde{u}_j)(u_i + u_j)}, \\ Z_i^- &= \frac{1}{t_i} \sum_{j \in C_i^-} t_{ij} \left(\frac{t_{ij}\tilde{u}_i}{\tilde{u}_i + \tilde{u}_j} - \frac{t_{ij} u_i}{u_i + u_j} \right) = -\frac{1}{t_i} \sum_{j \in C_i^-} t_{ij} \frac{\tilde{u}_i u_j - \tilde{u}_j u_i}{(\tilde{u}_i + \tilde{u}_j)(u_i + u_j)}, \\ C_i^+ &= \left\{ j : \frac{\tilde{u}_i}{u_i} > \frac{\tilde{u}_j}{u_j}, j \in C_i \right\} \quad \textit{and} \quad C_i^- = \left\{ j : \frac{\tilde{u}_i}{u_i} \le \frac{\tilde{u}_j}{u_j}, j \in C_i \right\}. \end{aligned}$$

LEMMA 3. *Assume condition* (2.5) *holds. For any* $i \in A_k$ *where* $k < K - 1$, *let*

$$\text{(B.5)} \qquad C_i^* = \left\{ j : j \in C_i, \frac{\tilde{u}_j}{u_j} \ge 1 - \beta(1+\phi_n)^{k+1} M_n \Delta_n \right\}.$$

Then for all large n,

$$P(|C_i^*| \ge q_n d_i) \ge 1 - 3n^{-3},$$

where q_n *and* d_i *are defined in* (B.1). *For any* $i \in A_{K-1}$, *let*

$$C_i^* = \left\{ j : j \in C_i, \frac{\tilde{u}_j}{u_j} \ge 1 - 40\beta T (1+\phi_n)^K M_n^2 \Delta_n \right\}. \tag{B.6}$$

Then for all large n,

$$P\left(|C_i^*| \ge \frac{19}{20} d_i \right) \ge 1 - 3n^{-3}.$$

LEMMA 4. *Assume condition* (2.5) *holds. For a set* $A \subset \Omega$, *let* $s = |A|$ *denote the size of* A. *Define a set* $B = \{j : \text{there exists } i \in A \text{such that} t_{ij} > 0\}$. *If* $sT < p_n^{-1}$, *then for all large* n,

$$P\left(|B| > \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)(sTnp_n - s^2T^2np_n^2)\right) \ge 1 - 2n^{-3sT}.$$

If $sT = p_n^{-1}$, *then for all large* n,

$$P\left(|B| > \frac{3}{5}\left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)n \right) \ge 1 - 2n^{-3sT}.$$

LEMMA 5. *Assume condition* (2.5) *holds. Recall* A_k *defined in* (B.2), *for all large* n,

$$P\left(|A_k| \ge (np_n)^{\frac{k}{2}}\right) \ge 1 - 6kn^{-2} \quad \text{for } k = 0, \dots, K-2, \tag{B.7}$$

$$P\left(|A_{K-1}| \ge \frac{1}{p_n} \right) \ge 1 - 6(K-1)n^{-2}, \tag{B.8}$$

and

$$P\left(|A_K| \ge \frac{21n}{40} \right) \ge 1 - 6Kn^{-2}. \tag{B.9}$$

REMARK 5. We give some insights to the proof of Lemma 5, which used the facts proved in Lemmas 2 to 4. In particular, Lemma 5 is proved by mathematical induction. For illustration, we assume that $u_i = 1$ for all i. A_k is exactly the set that contains subjects with estimators close to the maximum estimator $\max_{0 \le i \le n-1} \tilde{u}_i$ in the level of D_k. We aim to show there are more than $n/2$ subjects whose estimators are close to the maximum estimator in the level of D_K. In the first step, we begin with $A_0 = \{i_0\}$. With the use of Lemma 2, we know $Z_i^- = 0$ so that $Z_i^+ = O_p(\sqrt{\log n/(np_n)})$. It means there are some $\tilde{u}_j$ very close to $\tilde{u}_{i_0}$, where $j \in C_{i_0}$ (j is the neighbor of i_0). Moreover, Lemma 3 states the proportion of such kind of j in C_{i_0} and Lemma 4 gives the size of C_{i_0}. Eventually, we put i_0 and such kind of j together to generate the set A_1 whose size is obtained from Lemma 5.

Next, by Lemma 2, given $i \in A_1$, the relevant quantities Z_i^+ and Z_i^- associated with subjects in C_i^+ and C_i^- are balanced. If C_i^- has a large size, then C_i^*, the neighbors of i with estimators so large to be in A_2 will automatically be large; if C_i^- does not have a large size, the balance of Z_i^+ and Z_i^- dictates that those in C_i^+ would still have a large size of subset in A_2. The detailed arguments are given in Cases 1 and 2 in the proof of Lemma 3. Then, we find the subjects in A_2 through the neighbors of the subjects in A_1. We repeat the process until the size of A_k is larger than $n/2$.

B.1. Proof of Theorem 2.1.

PROOF. The proof mainly contains two parts. One is to show that the number of subjects whose ratios of estimated merits and real merits $\hat{u}_j/u_j$ are close to the largest ratio $\max_i \hat{u}_i/u_i$ is larger than $n/2$. The other is to show that the number of subjects whose ratios are close to the smallest ratio is also larger than $n/2$.

Step 1. Observe that (B.9), with proof given in that of Lemma 5, implies, for all large n,

$$P\left(\left|\left\{j : \frac{\tilde{u}_j}{u_j} \geq 1 - 40\beta T(1+\phi_n)^K M_n^2 \Delta_n\right\}\right| \geq \frac{21n}{40}\right) \geq 1 - 6Kn^{-2}. \tag{B.10}$$

Under condition (2.5), it follows that

$$K \leq 2\log n \quad \text{and} \quad (1+\phi_n)^K \leq e^2.$$

Let $\lambda = 40\beta T e^2$, from (B.10), we obtain for all large n,

$$P\left(\left|\left\{j : \frac{\tilde{u}_j}{u_j} \geq 1 - \lambda M_n^2 \Delta_n\right\}\right| \geq \frac{21n}{40}\right) \geq 1 - 12n^{-2}\log n. \tag{B.11}$$

Notice that $\tilde{u}_j = \hat{u}_j/(\max_i \hat{u}_i/u_i)$. Thus, (B.11) can be written as

$$P\left(\left|\left\{j : \frac{\hat{u}_j/u_j}{\max_{0\leq i\leq n-1}(\hat{u}_i/u_i)} \geq 1 - \lambda M_n^2 \Delta_n\right\}\right| \geq \frac{21n}{40}\right) \geq 1 - 12n^{-2}\log n.$$

It means that with probability approaching 1 as $n \to \infty$,

$$\left|\left\{j : \frac{\hat{u}_j/u_j}{\max_{0\leq i\leq n-1}(\hat{u}_i/u_i)} \geq 1 - \lambda M_n^2 \Delta_n\right\}\right| \geq \frac{21n}{40} > \frac{n}{2}. \tag{B.12}$$

Step 2. Next, we will show that the number of subjects whose ratios are close to the smallest ratio is also larger than $n/2$.

Let $\bar{u}_j = \hat{u}_j/(\min_i \hat{u}_i/u_i)$. Similar to (B.2), we define

$$\bar{A}_k = \left\{j : \frac{\bar{u}_j}{u_j} \leq 1 + \beta(1+\phi_n)^k M_n \Delta_n\right\} \quad \text{for } k = 0, \dots, K-1,$$

$$\bar{A}_K = \left\{j : \frac{\bar{u}_j}{u_j} \leq 1 + 40\beta T(1+\phi_n)^K M_n^2 \Delta_n\right\}.$$

Compared with (B.9), we can obtain

$$P\left(|\bar{A}_K| \geq \frac{21n}{40}\right) \geq 1 - 6Kn^{-2}$$

with the similar proof of Lemma 2 to 5. Similar to *Step 1*, that with probability approaching 1 as $n \to \infty$,

$$\left|\left\{j : \frac{\hat{u}_j/u_j}{\min_{0\leq i\leq n-1}(\hat{u}_i/u_i)} \leq 1 + \lambda M_n^2 \Delta_n\right\}\right| \geq \frac{21n}{40} > \frac{n}{2}. \tag{B.13}$$

Combining $\hat{u}_0 = u_0 = 1$, (B.12) and (B.13), it can be shown that with probability approaching 1 as $n \to \infty$,

$$\frac{1 - \lambda M_n^2 \Delta_n}{1 + \lambda M_n^2 \Delta_n} \leq \min_{0\leq i\leq n-1} \frac{\hat{u}_i}{u_i} \leq \max_{0\leq i\leq n-1} \frac{\hat{u}_i}{u_i} \leq \frac{1 + \lambda M_n^2 \Delta_n}{1 - \lambda M_n^2 \Delta_n}.$$

Consequently,

$$\max_{0\leq i\leq n-1}\left|\frac{\hat{u}_i}{u_i} - 1\right| \leq \frac{2\lambda M_n^2 \Delta_n}{1 - \lambda M_n^2 \Delta_n}.$$

Since $M_n^2\Delta_n \to 0$ as $n\to\infty$, with probability approaching 1,

$$\max_{0\le i\le n-1}\left|\frac{\hat{u}_i}{u_i}-1\right|\to 0 \quad \text{as } n\to\infty.$$

Except for the deferred proof of Lemma 2 to 5, the proof of Theorem 2.1 is complete. □

B.2. Proof of Lemma 2.

PROOF. The proof of Lemma 2 contains three steps. In the first step, we find the upper bound of $|a_i - E(a_i|t_{ij}, 0\le j\le n-1)|$ for fixed i. In the second step, we find the upper bound of $|Z_i^+ - Z_i^-|$ for fixed i through the first step. In the third step, we find the uniform upper bound of $|Z_i^+ - Z_i^-|$ for $i=0,\dots,n-1$.

Step 1. Recall that $t_i=\sum_{0\le j\le n-1} t_{ij}$ and a_i is the total number of wins of subject i in t_i comparisons. Since the outcome of each comparison is independent of other comparisons, a_i is the sum of m_i independent Bernoulli random variables given $t_{ij}=m_{ij}$, for $j=0,\dots,n-1$, where $m_i=\sum_{j=0}^{n-1} m_{ij}$. With the use of Hoeffding's inequality (Hoeffding (1963)), we have

$$\begin{aligned}
&P\big(\big|a_i - E(a_i|t_{ij},0\le j\le n-1)\big|\ge\sqrt{2Tt_i\log n}\,|t_{ij}=m_{ij},0\le j\le n-1\big)\\
&\quad = P\big(\big|a_i - E(a_i|t_{ij},0\le j\le n-1)\big|\ge\sqrt{2Tm_i\log n}\,|t_{ij}=m_{ij},0\le j\le n-1\big)\\
&\quad \le \exp\{-(4Tm_i\log n)/m_i\} = 2n^{-4T}\le 2n^{-4},
\end{aligned}$$

where $E(a_i|t_{ij},0\le j\le n-1)$ is the conditional expectation given t_{ij} for $0\le j\le n-1$. Note that the upper bound of the above probability does not depend on m_{ij}. With the law of total probability, for fixed i,

$$\begin{aligned}
&P\big(\big|a_i - E(a_i|t_{ij},0\le j\le n-1)\big|\ge\sqrt{2Tt_i\log n}\big)\\
&\quad = \sum_{m_{i0}=0}^{T}\cdots\sum_{m_{i,n-1}=0}^{T} P(t_{ij}=m_{ij},0\le j\le n-1)\\
&\qquad \times P\big(\big|a_i - E(a_i|t_{ij},0\le j\le n-1)\big|\ge\sqrt{2Tt_i\log n}\,|t_{ij}=m_{ij},0\le j\le n-1\big)\\
&\quad \le 2n^{-4}\sum_{m_{i0}=0}^{T}\cdots\sum_{m_{i,n-1}=0}^{T} P(t_{ij}=m_{ij},0\le j\le n-1)\\
&\quad = 2n^{-4}.
\end{aligned}$$

Thus with probability at least $1-2n^{-4}$, for any fixed i,

$$\big|a_i - E(a_i|t_{ij},0\le j\le n-1)\big| < \sqrt{2Tt_i\log n}. \tag{B.14}$$

Step 2. Recall that the maximum likelihood estimator $\hat{u}_i$ satisfies

$$a_i=\sum_{j=0}^{n-1} a_{ij}=\sum_{j=0}^{n-1}\frac{t_{ij}\hat{u}_i}{\hat{u}_i+\hat{u}_j}.$$

Since $\tilde{u}_j=\hat{u}_j/(\max_i \hat{u}_i/u_i)$, we can rewrite the above equation as,

$$a_i=\sum_{j=0}^{n-1} a_{ij}=\sum_{j=0}^{n-1}\frac{t_{ij}\tilde{u}_i}{\tilde{u}_i+\tilde{u}_j}.$$

Then,

$$a_i - E(a_i | t_{ij}, 0 \le j \le n-1) = \sum_{j=0}^{n-1} \left(\frac{t_{ij}\tilde{u}_i}{\tilde{u}_i + \tilde{u}_j} - \frac{t_{ij}u_i}{u_i + u_j} \right) = t_i(Z_i^+ - Z_i^-).$$

Based on (B.14), we can obtain with probability at least $1 - 2n^{-4}$,

$$|Z_i^+ - Z_i^-| < \sqrt{\frac{2T \log n}{t_i}}.$$

Step 3. We first find the uniform lower bound of t_i for $i = 0, \dots, n-1$. Notice that t_i is the sum of n independent and identically distributed (i.i.d.) binomial random variables, $\mathrm{Bin}(T, p_n)$. It can be also regarded as the sum of Tn i.i.d. Bernoulli random variables. With the use of Chernoff bound (Chernoff (1952)), we have

$$P\left(\min_{0 \le i \le n-1} t_i < \frac{T}{2} np_n \right) \le \sum_{i=0}^{n-1} P\left(t_i < \frac{T}{2} np_n \right) \le n \exp\left\{ -\frac{T}{12} np_n \right\}.$$

Thus, with probability at least $1 - n\exp\{-(Tnp_n)/12\}$,

$$\min_{0 \le i \le n-1} t_i \ge \frac{T}{2} np_n \tag{B.15}$$

which means that $t_i = O_p(np_n)$. According to the result of *Step 2*,

$$P\left(\max_{0 \le i \le n-1} |Z_i^+ - Z_i^-| \ge \sqrt{2T \log n / \min_{0 \le i \le n-1} t_i} \right)$$
$$\le \sum_{i=0}^{n-1} P\left(|Z_i^+ - Z_i^-| \ge \sqrt{2T \log n / t_i} \right) \le n \times 2n^{-4} = 2n^{-3}.$$

By (B.15), with probability $1 - 2n^{-3} - n\exp\{-(Tnp_n)/12\}$,

$$\max_{0 \le i \le n-1} |Z_i^+ - Z_i^-| < 2\sqrt{\frac{\log n}{np_n}}.$$

Meanwhile, $M_n^2 \Delta_n \to 0$ as $n \to \infty$ implies $(\log n)/(np_n) \to 0$ as $n \to \infty$. Therefore,

$$n \exp\left\{ -\frac{Tnp_n}{12} \right\} < n^{-3},$$

for all large n. As a result, with probability at least $1 - 3n^{-3}$,

$$\max_{0 \le i \le n-1} |Z_i^+ - Z_i^-| < 2\sqrt{\frac{\log n}{np_n}}.$$

We complete the proof. □

B.3. Proof of Lemma 3.

PROOF. We first consider the case when $k < K - 1$. Recall the definition of A_k is given in (B.2). For any $i \in A_k$, we aim to show that for all large n, with probability at least $1 - 3n^{-3}$,

$$|C_i^*| \ge q_n d_i,$$

where C_i^* is defined in (B.5). Observe that $C_i^- \subset C_i^*$ for any $i \in A_k$.

Case 1. If $|C_i^-| \geq q_n d_i$, then we have

$$|C_i^*| \geq |C_i^-| \geq q_n d_i.$$

Case 2. If $|C_i^-| < q_n d_i$, we set $|C_i^-| = \alpha d_i$, where $\alpha < q_n$. We need to show that for all large n, with probability at least $1 - 3n^{-3}$,

$$\left|\left\{ j : j \in C_i^+, \frac{\tilde{u}_j}{u_j} > 1 - \beta(1+\phi_n)^{k+1} M_n \Delta_n \right\}\right| \geq (q_n - \alpha) d_i.$$

We use Lemma 2 to prove the above inequality and our proof contains three steps. Recall that Z_i^+ and Z_i^- defined in (B.4). The first step is to find the lower bound of Z_i^+ and the upper bound of Z_i^-. The second step is to show that the number of subjects who have comparisons with i and close ratios to $\hat{u}_i/u_i$ is larger than $(q_n - \alpha)d_i$, that is,

$$\left|\left\{ j : \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i} > 1 - \beta\phi_n(1+\phi_n)^k M_n \Delta_n,\ j \in C_i^+ \right\}\right| \geq (q_n - \alpha) d_i.$$

The last step is to prove the subject j included in the above set belongs to C_i^*.

Step 1. For Z_i^-, we have

$$\begin{aligned} Z_i^- &= -\frac{1}{t_i} \sum_{j \in C_i^-} \frac{\tilde{u}_i u_j - \tilde{u}_j u_i}{(\tilde{u}_i + \tilde{u}_j)(u_i + u_j)} \cdot t_{ij} \\ &= \frac{1}{t_i} \sum_{j \in C_i^-} \frac{\tilde{u}_j/u_j - \tilde{u}_i/u_i}{(\tilde{u}_i/u_i + \tilde{u}_j/u_j \times u_j/u_i)(1 + u_i/u_j)} \cdot t_{ij}. \end{aligned}$$

Since $M_n^2 \Delta_n \to 0$ as $n \to \infty$,

$$\beta(1+\phi_n)^k M_n \Delta_n \leq \beta e^2 M_n \Delta_n \leq \frac{1}{2}$$

for all large n. It can be given that for $i \in A_k$, $j \in C_i^-$ and for all large n,

$$\frac{1}{2} \leq 1 - \beta(1+\phi_n)^k M_n \Delta_n \leq \tilde{u}_i/u_i \leq \tilde{u}_j/u_j \leq 1.$$

Therefore, for all large n,

$$Z_i^- \leq \frac{1}{t_i} \sum_{j \in C_i^-} 2t_{ij} \frac{1 - (1 - \beta(1+\phi_n)^k M_n \Delta_n)}{(1 + u_i/u_j)(1 + u_j/u_i)} \leq \frac{\alpha\beta}{2}(1+\phi_n)^k M_n \Delta_n.$$

The last inequality is from

$$4 \leq \left(1 + \frac{u_i}{u_j}\right)\left(1 + \frac{u_j}{u_i}\right) \leq \frac{(1+M_n)^2}{M_n}.$$

Similarly, for Z_i^+, we obtain

$$\begin{aligned} Z_i^+ &= \frac{1}{t_i} \sum_{j \in C_i^+} t_{ij} \frac{\tilde{u}_i u_j - \tilde{u}_j u_i}{(\tilde{u}_i + \tilde{u}_j)(u_i + u_j)} \\ &= \frac{1}{t_i} \sum_{j \in C_i^+} t_{ij} \frac{\tilde{u}_i/u_i - \tilde{u}_j/u_j}{(\tilde{u}_i/u_i + \tilde{u}_j/u_j \times u_j/u_i)(1 + u_i/u_j)} \\ &\geq \frac{M_n}{(1+M_n)^2 T d_i} \sum_{j \in C_i^+} \left(1 - \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i}\right). \end{aligned}$$

From above, we have bounds for Z_i^- and Z_i^+ respectively, for all large n,

$$Z_i^- \le \frac{\alpha\beta}{2}(1+\phi_n)^k M_n \Delta_n, \qquad Z_i^+ \ge \frac{M_n}{(1+M_n)^2 T d_i} \sum_{j\in C_i^+} \left(1 - \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i}\right). \tag{B.16}$$

Step 2. According to Lemma 2, we have for all large n, with probability at least $1-3n^{-3}$,

$$Z_i^+ - \frac{\alpha\beta}{2}(1+\phi_n)^k M_n \Delta_n \le Z_i^+ - Z_i^- \le 2\sqrt{\frac{\log n}{np_n}}$$

Since $\Delta_n\phi_n = \sqrt{\log n/(np_n)}$, for all large n, with probability at least $1-3n^{-3}$,

$$\begin{aligned} Z_i^+ &\le \frac{\alpha\beta}{2}(1+\phi_n)^k M_n \Delta_n + 2\sqrt{\frac{\log n}{np_n}} \\ &= \left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n. \end{aligned}$$

Then, from (B.16), it follows that for all large n, with probability at least $1-3n^{-3}$,

$$\frac{M_n}{(1+M_n)^2 T d_i} \sum_{j\in C_i^+} \left(1 - \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i}\right) \le \left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n.$$

Notice that $|C_i^+| = |C_i| - |C_i^-| = (1-\alpha)d_i$, we can rewrite the above inequality as for all large n, with probability at least $1-3n^{-3}$,

$$\frac{1}{|C_i^+|} \sum_{j\in C_i^+} \left(1 - \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i}\right) \le \frac{T(1+M_n)^2}{M_n(1-\alpha)}\left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n.$$

Set the xth percentile of $\{(\tilde{u}_j/u_j)/(\tilde{u}_i/u_i) : j \in C_i^+\}$ to be b_x, where $x = (1-q_n)/(1-\alpha)$. As

$$\frac{1}{|C_i^+|} \sum_{j\in C_i^+} \left(1 - \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i}\right) = 1 - \frac{1}{|C_i^+|} \sum_{j\in C_i^+} \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i},$$

it follows that for all large n, with probability at least $1-3n^{-3}$,

$$\begin{aligned} \frac{1}{|C_i^+|} \sum_{j\in C_i^+} \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i} &\ge 1 - \frac{T(1+M_n)^2}{M_n(1-\alpha)}\left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n, \\ 1 - x + b_x x &\ge 1 - \frac{T(1+M_n)^2}{M_n(1-\alpha)}\left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n, \\ b_x &\ge 1 - \frac{T(1+M_n)^2}{M_n(1-\alpha)x}\left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n. \end{aligned}$$

Since $x = (1-q_n)/(1-\alpha)$, $0 \le \alpha < q_n$, where q_n is defined in (B.1), we have for all large n, with probability at least $1-3n^{-3}$,

$$\begin{aligned} b_x &\ge 1 - \frac{T(1+M_n)^2}{M_n(1-q_n)}\left[\frac{\alpha\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n \\ &> 1 - \frac{T(1+M_n)^2}{M_n(1-q_n)}\left[\frac{q_n\beta}{2}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n \end{aligned}$$

$$\geq 1 - \frac{(1+M_n)^2T + \phi_n M_n}{M_n}\left[\frac{\phi_n M_n \beta}{2(1+M_n)^2T + 2\phi_n M_n}(1+\phi_n)^k M_n + 2\phi_n\right]\Delta_n$$
$$\geq 1 - \left[\frac{\beta}{2}\phi_n(1+\phi_n)^k M_n + \frac{(1+M_n)^2T + \phi_n M_n}{M_n} \times 2\phi_n\right]\Delta_n$$
$$\geq 1 - \left[\frac{\beta}{2}\phi_n(1+\phi_n)^k M_n + 8TM_n\phi_n + 2\phi_n^2\right]\Delta_n.$$

Given $\beta = 20T$, we can rewrite above inequality as

$$b_x > 1 - \beta\phi_n(1+\phi_n)^k M_n \Delta_n,$$

which means

$$\left|\left\{j : \frac{\tilde{u}_j/u_j}{\tilde{u}_i/u_i} > 1 - \beta\phi_n(1+\phi_n)^k M_n\Delta_n,\ j \in C_i^+\right\}\right| \geq (1-x)(1-\alpha)d_i = (q_n - \alpha)d_i.$$

Step 3. Since $i \in A_k = \{j : (\tilde{u}_j/u_j) \geq 1 - \beta(1+\phi_n)^k M_n\Delta_n\}$, for any $j \in \{j : (\tilde{u}_j/u_j)/(\tilde{u}_i/u_i) > 1 - \beta\phi_n(1+\phi_n)^k M_n\Delta_n,\ j \in C_i^+\}$, it follows that for all large n, with probability at least $1 - 3n^{-3}$,

$$\frac{\tilde{u}_j}{u_j} > \big(1 - \beta\phi_n(1+\phi_n)^k M_n\Delta_n\big)\big(1 - \beta(1+\phi_n)^k M_n\Delta_n\big)$$
$$\geq 1 - \beta\phi_n(1+\phi_n)^k M_n\Delta_n - \beta(1+\phi_n)^k M_n\Delta_n$$
$$= 1 - \beta(1+\phi_n)^{k+1} M_n\Delta_n,$$

which implies

$$\left|\left\{j : \frac{\tilde{u}_j}{u_j} > 1 - \beta(1+\phi_n)^{k+1} M_n\Delta_n,\ j \in C_i^+\right\}\right| \geq (q_n - \alpha)d_i.$$

Therefore, for all large n, with probability at least $1 - 3n^{-3}$,

$$|C_i^*| \geq |C_i^-| + \left|\left\{j : \frac{\tilde{u}_j}{u_j} > 1 - \beta(1+\phi_n)^{k+1} M_n\Delta_n,\ j \in C_i^+\right\}\right|$$
$$\geq \alpha d_i + (q_n - \alpha)d_i$$
$$= q_n d_i.$$

For $k = K - 1$, we can obtain the result with the same proof as above except replacing q_n with $19/20$. Hence, for all large n, with probability at least $1 - 3n^{-3}$,

$$|C_i^*| = \left|\left\{j : j \in C_i, \frac{\tilde{u}_j}{u_j} \geq 1 - 40\beta T(1+\phi_n)^K M_n^2\Delta_n\right\}\right| \geq \frac{19}{20}d_i.$$

The proof is complete. □

B.4. Proof of Lemma 4.

PROOF. We present the proof with two steps. The first step is to find the lower bound of $|B|$ when A is a deterministic set. The second step is to extend the result of the first step to the case when A is a random set.

Step 1. Let A be a nonrandom set with size $s \leq (Tp_n)^{-1}$. For any $j \in \Omega$,

$$P(t_{ij} = 0 \text{ for all } i \in A) = (1 - p_n)^{sT}.$$

Set $y = 1 - (1 - p_n)^{sT}$ and $\eta_j = I(j \in B)$. So $\eta_j = 1$ if j has a comparison with someone in A, otherwise $\eta_j = 0$. We know that $P(\eta_j = 1) = y$.

Since $M_n^2 \Delta_n \to 0$ as $n \to \infty$, $4\sqrt{\log n} < \sqrt{np_n}$ for all large n. Thus, with Chernoff bound (Chernoff (1952)), for all large n,

$$\begin{aligned} P\left(|B| \le \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)ny\right) &\le 2\exp\left\{-\frac{16ny\log n}{2np_n}\right\} \\ &= 2\exp\left\{-\frac{8(1-\exp(sT\log(1-p_n)))\log n}{p_n}\right\} \\ &\le 2\exp\left\{-\frac{8(1-\exp(-sTp_n))\log n}{p_n}\right\} \\ &\le 2\exp\{-4sT\log n\}. \end{aligned}$$

Here, the second and third inequalities are based on $\log(1-x) \le -x$ and $x \le 2(1-\exp(-x))$ respectively when $0 < x < 1$.

Step 2. For any set $A \subset \Omega$ with size s and all large n, it follows that

$$\begin{aligned} &P\left(\left(\min_{|A|=s}\left|\{j : \text{there exists } i \in A \text{ such that } t_{ij} > 0\}\right|\right) \le \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)ny\right) \\ &\le \sum_{|A|=s} P\left(\left|\{j : \text{there exists } i \in A \text{ such that } t_{ij} > 0\}\right| \le \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)ny\right) \\ &\le 2\binom{n}{s}\exp\{-4sT\log n\} \\ &\le 2n^s\exp\{-4sT\log n\} \\ &\le 2n^{-3sT}. \end{aligned}$$

In summary,

$$\begin{aligned} &P\left(\left(\min_{|A|=s}\left|\{j : \text{there exists } i \in A \text{such that} t_{ij} > 0\}\right|\right) \le \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)ny\right) \\ &\quad \le 2n^{-3sT} \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

Since for any set A with the size s,

$$|B| \ge \min_{|A| = s}\left|\{j : \text{there exists } i \in A \text{ such that } t_{ij} > 0\}\right|.$$

Thus, for all large n, with probability at least $1 - 2n^{-3sT}$,

$$|B| > \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)ny.$$

Recall that $y = 1 - (1 - p_n)^{sT}$, for $sT < p_n^{-1}$,

$$ny = n\big(1 - (1 - p_n)^{sT}\big) \ge sTnp_n - s^2T^2np_n^2,$$

while for $sT = p_n^{-1}$,

$$ny = n\big(1 - (1 - p_n)^{sT}\big) \ge n\big(1 - e^{-1}\big) > \frac{3}{5}n.$$

The proof is complete. $\square$

B.5. Proof of Lemma 5.

PROOF. Our aim is to show that there exists a uniform constant C such that when $n > C$, (B.7), (B.8) and (B.9) hold. C is defined as the maximum of n which does not satisfy any following inequalities,

$$n^{-3} > n\exp\left\{-\frac{Tnp_n}{12}\right\}, \qquad \sqrt{np_n} > 4\sqrt{\log n},$$
$$\beta e^2 M_n \Delta_n \leq \frac{1}{2} \quad \text{and} \quad q_n\sqrt{np_n} - 8\sqrt{\log n} > 2, \tag{B.17}$$

where q_n is defined in (B.1) and β is defined in (B.9). Since $M_n^2\Delta_n \to 0$ as $n \to \infty$, it ensures the existence of C. Then, for any $n > C$, n satisfies all inequalities in (B.17).

We show Lemma 5 by mathematical induction.

(1) For $k = 0$, it is obvious that

$$|A_0| \geq \big|\{i_0\}\big| = (np_n)^0 = 1.$$

Therefore, we obtain

$$P\big(|A_0| \geq (np_n)^0\big) \geq 1.$$

(2) For $k < K - 2$, let $\mathcal{A}_k$ denote the event that $|A_k| \geq (np_n)^{k/2}$ happens. Assume that when $n > C$,

$$P(\mathcal{A}_k) \geq 1 - 6kn^{-2}.$$

Then we proceed under the condition that event $\mathcal{A}_k$ happens. Without loss of generality, let $|A_k| = (np_n)^{k/2}/T$, otherwise we consider any of its subsets with size $(np_n)^{k/2}/T$.

For any $i \in A_k$, we have $C_i^* \subset A_{k+1}$. Hence $\bigcup_{i\in A_k} C_i^* \subset A_{k+1}$. Consequently,

$$|A_{k+1}| \geq \left|\bigcup_{i\in A_k} C_i^*\right| \geq \left|\bigcup_{i\in A_k} C_i\right| - \sum_{i\in A_k}|C_i \setminus C_i^*|. \tag{B.18}$$

Next we estimate $\sum_{i\in A_k}|C_i \setminus C_i^*|$ and $|\bigcup_{i\in A_k} C_i|$ by Lemma 3 and Lemma 4 respectively. Note that $\bigcup_{i\in A_k} C_i = \{j : j$ has a comparsion with anyone in $A_k\}$ and

$$|A_k| = \frac{(np_n)^{\frac{k}{2}}}{T} \leq \frac{(np_n)^{\frac{K-3}{2}}}{T} \leq \frac{(np_n^3)^{-\frac{1}{2}}}{T}. \tag{B.19}$$

Based on Lemma 4, we know that when $n > C$, with probability at least $1 - 2n^{-3T|A_k|}$,

$$\left|\bigcup_{i\in A_k} C_i\right| \geq \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)\left[(np_n)^{\frac{k}{2}+1} - (np_n)^k p_n^2 n\right]. \tag{B.20}$$

Meanwhile, from Lemma 3, we know for $i \in A_k$, when $n > C$, with probability at least $1 - 3n^{-3}$,

$$|C_i \setminus C_i^*| \leq (1 - q_n)d_i,$$

where q_n and d_i are defined in (B.1). As a result, when $n > C$, with probability at least $1 - 3|A_k|n^{-3}$,

$$\sum_{i\in A_k}|C_i \setminus C_i^*| \leq (1 - q_n)\sum_{i\in A_k} d_i. \tag{B.21}$$

Based on the Chernoff bound (Chernoff (1952)), the range of d_i is given as

$$P\left(d_i \geq \left(1+\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)Tnp_n\right) \leq P\left(d_i \geq \left(1+\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)E(d_i)\right)$$
$$\leq \exp\left(-\frac{16}{3}\log n\right),$$

where the first inequality is from $E(d_{ij}) = 1-(1-p_n)^T \leq Tp_n$. Consequently,

$$P\left(\left(\max_{0\leq i\leq n-1} d_i\right) \geq \left(1+\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)Tnp_n\right) \leq n\exp\left(-\frac{16}{3}\log n\right) \leq n^{-4},$$

which implies

$$\left(\max_{0\leq i\leq n-1} d_i\right) < \left(1+\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)Tnp_n,$$

with probability at least $1-n^{-4}$. So we can rewrite (B.21) as when $n > C$, with probability at least $1-3|A_k|n^{-3}-n^{-4}$,

$$\sum_{i\in A_k}|C_i \setminus C_i^*| \leq (1-q_n)\left(1+\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)(np_n)^{\frac{k}{2}+1}. \tag{B.22}$$

Based on (B.18), (B.20) and (B.22), when $n > C$, with probability at least $1-2n^{-3T|A_k|}-3|A_k|n^{-3}-n^{-4}$,

$$|A_{k+1}| \geq \left(1-\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)\left[(np_n)^{\frac{k}{2}+1}-(np_n)^k p_n^2 n\right]-(1-q_n)\left(1+\frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)(np_n)^{\frac{k}{2}+1}$$
$$\geq (np_n)^{\frac{k}{2}+1}\left[q_n-\frac{8\sqrt{\log n}}{\sqrt{np_n}}-(np_n)^{\frac{k}{2}}p_n\right].$$

Due to $1\leq |A_k|\leq n$, $1-2n^{-3T|A_k|}-3|A_k|n^{-3}-n^{-4}\geq 1-6n^{-2}$. Thus, when $n > C$, with probability at least $1-6n^{-2}$,

$$|A_{k+1}| \geq (np_n)^{\frac{k}{2}+1}\left[q_n-\frac{8\sqrt{\log n}}{\sqrt{np_n}}-(np_n)^{\frac{k}{2}}p_n\right]. \tag{B.23}$$

According to (B.19), $(np_n)^{k/2}\leq (np_n^3)^{-1/2}$. Meanwhile, when $n > C$, from (B.17), we have

$$\sqrt{np_n}\left[q_n-\frac{8\sqrt{\log n}}{\sqrt{np_n}}-(np_n)^{\frac{k}{2}}p_n\right] = q_n\sqrt{np_n}-8\sqrt{\log n}-(np_n)^{\frac{k}{2}}p_n\sqrt{np_n}$$
$$\geq 2-(np_n)^{\frac{k}{2}}p_n\sqrt{np_n}$$
$$\geq 2-(np_n^3)^{-\frac{1}{2}}\times p_n\sqrt{np_n} = 1.$$

Hence, we can rewrite (B.23) as when $n > C$, with probability at least $1-6n^{-2}$,

$$|A_{k+1}| \geq (np_n)^{\frac{k+1}{2}}.$$

That is, when $n > C$,

$$P\left(|A_{k+1}| \geq (np_n)^{\frac{k+1}{2}} \mid \mathcal{A}_k\right) \geq 1-6n^{-2}.$$

Given $P(\mathcal{A}_k)\geq 1-6kn^{-2}$, we obtain when $n > C$,

$$P\left(|A_{k+1}| \geq (np_n)^{\frac{k+1}{2}}\right) \geq 1-6(k+1)n^{-2}.$$

(3) For $k = K - 2$, we assume that when $n > C$,

$$(|A_{K-2}| \geq (np_n)^{\frac{K-2}{2}}) \geq 1 - 6(K-2)n^{-2}.$$

Notice that $(np_n)^{(K-2)/2} \geq (np_n^3)^{-1/2}$. We choose a subset of A_{K-2} with size $(np_n^3)^{-1/2}/T$. Proceed similarly as (2), so when $n > C$, with probability at least $1 - 6(K-1)n^{-2}$,

$$\begin{aligned}
|A_{K-1}| &\geq \left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)\left[\frac{n}{\sqrt{np_n}} - \frac{1}{p_n}\right] - (1-q_n)\left(1 + \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)\frac{n}{\sqrt{np_n}} \\
&\geq \frac{n}{\sqrt{np_n}}\left(q_n - \frac{8\sqrt{\log n}}{\sqrt{np_n}} - \frac{1}{\sqrt{np_n}}\right) \\
&\geq \frac{1}{p_n}.
\end{aligned}$$

(4) For $k = K - 1$, we assume that when $n > C$,

$$P\left(|A_{K-1}| \geq \frac{1}{p_n}\right) \geq 1 - 6(K-1)n^{-2}.$$

We choose a subset of A_{K-1} with size $(Tp_n)^{-1}$ and can complete the proof similar to (2) by replacing q_n with 19/20 and using the second case of Lemma 3 and Lemma 4. Hence, when $n > C$, with probability at least $1 - 6Kn^{-2}$,

$$\begin{aligned}
|A_K| &\geq \frac{3}{5}\left(1 - \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)n - \frac{1}{20}\left(1 + \frac{4\sqrt{\log n}}{\sqrt{np_n}}\right)n \\
&= \frac{11}{20}n - \frac{11\sqrt{\log n}}{5\sqrt{np_n}} \\
&\geq \frac{21}{40}n.
\end{aligned}$$

The proof is complete. □

APPENDIX C: PROOF OF ASYMPTOTIC NORMALITY

Now we will sketch the proof of Theorem 2.2. Similar to Simons and Yao (1999), we need the following lemmas.

LEMMA 6. *If*

$$\delta_n = 32M_n\sqrt{\frac{\log n}{(n-1)p_n^3}} \to 0 \quad \text{as } n \to \infty, \tag{C.1}$$

then $\max_{i=0,\dots,n-1} |\Delta u_i| = O_p(\delta_n) \to 0$ *as* $n \to \infty$.

LEMMA 7. *If*

$$\boldsymbol{W}_{n-1} := \boldsymbol{V}_{n-1}^{-1} - \boldsymbol{S}_{n-1},$$

then with probability approaching 1 *as* $n \to \infty$,

$$\|\boldsymbol{W}_{n-1}\| \leq \frac{256T^2M_n^3}{(n-1)^2p_n^3},$$

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

Lemma 7 evaluates the quality of the approximation $\boldsymbol{S}_{n-1}$, for $\boldsymbol{V}_{n-1}^{-1}$. This idea was first proposed by Simons and Yao (1998). We are able to establish analogous results with the sparser probability.

Let $\boldsymbol{a} = (a_1, \ldots, a_{n-1})^\top$, where a_i is defined in the (2.2) for $i = 1, \ldots, n-1$.

LEMMA 8. *If $\boldsymbol{R}_{n-1}$ denotes the covariance matrix of $\boldsymbol{W}_{n-1}\boldsymbol{a}$, then with probability approaching* 1 *as $n \to \infty$,*

$$\|\boldsymbol{R}_{n-1}\| \le \frac{256T^2M_n^3}{(n-1)^2p_n^3} + \frac{48TM_n^2}{(n-1)^2p_n^2}.$$

As 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), page 289) and the following lemma is derived.

LEMMA 9. *If $M_n = o(n)$ as $n \to \infty$, then, as $n \to \infty$, 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 $r \ge 1$. Moreover, the first r rows of $\boldsymbol{S}_{n-1}(\boldsymbol{a} - E(\boldsymbol{a}))$ are asymptotically normal with covariance matrix given by the upper left $r \times r$ block of $\boldsymbol{S}_{n-1}$, for fixed $r \ge 1$.*

PROOF OF THEOREM 2.2. Recall that E_n is the event that Condition A holds and let G_n be the event that

$$\max_{0\le i\le n-1} |\Delta u_i| \le 32TM_n\sqrt{\frac{\log n}{(n-1)p_n^3}}.$$

It follows from Lemma 1 and Lemma 6 that $P(E_n \cap G_n) \to 1$ as $n \to \infty$. We proceed under the condition that event $E_n \cap G_n$ happens. Let

$$\xi_{ij} = \frac{t_{ij}u_iu_j(\Delta u_i - \Delta u_j)}{(u_i + u_j)^2}, \qquad \xi_i = \sum_{j=0, j\ne i}^{n-1} \xi_{ij},$$

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{n-1})^\top, \qquad \boldsymbol{\eta} = (\eta_1, \ldots, \eta_{n-1})^\top = \boldsymbol{a} - E(\boldsymbol{a}) - \boldsymbol{\xi}, \qquad \eta_0 = \sum_{j=1}^{n-1} \eta_j.$$

It follows that with probability approaching 1 as $n \to \infty$,

$$\begin{aligned} |\eta_i| &\le 2v_{ii} \max_{0\le j\le n-1} |\Delta u_j|^2 \le v_{ii}\frac{2^{11}T^2M_n^2\log n}{(n-1)p_n^3}, \quad i = 1, \ldots, n-1, \\ |(\boldsymbol{S}_{n-1}\boldsymbol{\eta})_i| &\le \frac{1}{v_{ii}}|\eta_i| + \frac{1}{v_{00}}|\eta_0| \le \frac{2^{12}T^2M_n^2\log n}{(n-1)p_n^3} = O_p\left(\frac{M_n^2\log n}{np_n^3}\right), \end{aligned} \tag{C.2}$$

where v_{ij} is defined in (2.11). With the use of Chernoff bound (Chernoff (1952)), it is easy to show that, with probability approaching 1 as $n \to \infty$,

$$\frac{Tnp_n}{8M_n} \le \frac{M_n}{(M_n+1)^2} \min_{0\le i\le n-1} t_i \le v_{ii} \le \frac{1}{4} \max_{0\le i\le n-1} t_i \le \frac{3Tnp_n}{8}. \tag{C.3}$$

According to Lemma 7 and (C.3),

$$|(\boldsymbol{W}_{n-1}\boldsymbol{\eta})_i| \le \frac{256T^2M_n^3}{(n-1)^2p_n^3} \times \sum_{i=1}^{n-1} |\eta_i| = O_p\left(\frac{M_n^5\log n}{np_n^5}\right). \tag{C.4}$$

By (C.2) and (C.4),

$$\left|(\boldsymbol{V}_{n-1}^{-1}\boldsymbol{\eta})_i\right| \le \left|(\boldsymbol{W}_{n-1}\boldsymbol{\eta})_i\right| + \left|(\boldsymbol{S}_{n-1}\boldsymbol{\eta})_i\right| = O_p\left(\frac{M_n^5 \log n}{np_n^5}\right) + O_p\left(\frac{M_n^2 \log n}{np_n^3}\right).$$

Since $\boldsymbol{\xi} = \boldsymbol{V}_{n-1}\boldsymbol{\Delta u}$, where $\boldsymbol{\Delta u} = (\Delta u_1, \dots, \Delta u_{n-1})^\top$, it can be obtained that

$$\begin{aligned}\boldsymbol{\Delta u} &= \boldsymbol{V}_{n-1}^{-1}\boldsymbol{\xi}\\ &= \boldsymbol{V}_{n-1}^{-1}\big(\boldsymbol{a} - E(\boldsymbol{a})\big) - \boldsymbol{V}_{n-1}^{-1}\boldsymbol{\eta}\\ &= \boldsymbol{S}_{n-1}\big(\boldsymbol{a} - E(\boldsymbol{a})\big) + \boldsymbol{W}_{n-1}\big(\boldsymbol{a} - E(\boldsymbol{a})\big) - \boldsymbol{V}_{n-1}^{-1}\boldsymbol{\eta}.\end{aligned} \tag{C.5}$$

When (2.13) holds, $|(\boldsymbol{V}_{n-1}^{-1}\boldsymbol{\eta})_i| = o_p(n^{-1/2})$, and by Lemma 8, $|(\boldsymbol{W}_{n-1}(\boldsymbol{a} - E(\boldsymbol{a})))_i| = o_p(n^{-1/2})$. So (C.5) is equivalent to

$$\Delta u_i = (\boldsymbol{V}_{n-1}^{-1}\boldsymbol{\xi})_i = \big(\boldsymbol{S}_{n-1}(\boldsymbol{a} - E(\boldsymbol{a}))\big)_i + o_p(n^{-1/2}).$$

Following Lemma 9, the proof is complete. □

Acknowledgments. The authors are thankful to the Editor, the Associate Editor and referees for their constructive comments. The research of Kani Chen is supported by Hong Kong Research Grant Council grants 16300714 and 16309816.

REFERENCES

AGRESTI, A. (1990). *Categorical Data Analysis*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1044993

BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925 https://doi.org/10.2307/2334029

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507. MR0057518 https://doi.org/10.1214/aoms/1177729330

COULOM, R. (2008). Whole-history rating: A Bayesian rating system for players of time-varying strength. In *International Conference on Computers and Games* 113–124. Springer, Berlin.

ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

FORD, L. R. JR. (1957). Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly* **64** 28–33. MR0097876 https://doi.org/10.2307/2308513

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

LOÈVE, M. (1977). *Probability Theory. I*, 4th ed. *Graduate Texts in Mathematics* **45**. Springer, New York. MR0651017

LUCE, R. D. (1959). *Individual Choice Behavior*: *A Theoretical Analysis*. Wiley, New York. MR0108411

MAYSTRE, L. and GROSSGLAUSER, M. (2015). Fast and accurate inference of Plackett–Luce models. In *Advances in Neural Information Processing Systems* 172–180.

NEGAHBAN, S., OH, S. and SHAH, D. (2012). Iterative ranking from pair-wise comparisons. In *Advances in Neural Information Processing Systems* 2474–2482.

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. MR0217963

SIMONS, G. and YAO, Y.-C. (1998). Approximating the inverse of a symmetric positive definite matrix. *Linear Algebra Appl.* **281** 97–103. MR1645343 https://doi.org/10.1016/S0024-3795(98)10038-1

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. MR1724040 https://doi.org/10.1214/aos/1018031267

YAN, T., YANG, Y. and XU, J. (2012). Sparse paired comparisons in the Bradley–Terry model. *Statist. Sinica* **22** 1305–1318. MR2987494 https://doi.org/10.5705/ss.2010.299

ZERMELO, E. (1929). Die Berechnung der Turnier–Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Math. Z.* **29** 436–460. MR1545015 https://doi.org/10.1007/BF01180541