A General Pairwise Comparison Model for Extremely Sparse Networks

Ruijian Han, Yiming Xu, and Kani Chen

CONTACT Ruijian Han ruijianhan@cuhk.edu.hk
Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China.

ABSTRACT Statistical estimation using pairwise comparison data is an effective approach to analyzing large-scale sparse networks. In this article, we propose a general framework to model the mutual interactions in a network, which enjoys ample flexibility in terms of model parameterization. Under this setup, we show that the maximum likelihood estimator for the latent score vector of the subjects is uniformly consistent under a near-minimal condition on network sparsity. This condition is sharp in terms of the leading order asymptotics describing the sparsity. Our analysis uses a novel chaining technique and illustrates an important connection between graph topology and model consistency. Our results guarantee that the maximum likelihood estimator is justified for estimation in large-scale pairwise comparison networks where data are asymptotically deficient. Simulation studies are provided in support of our theoretical findings. Supplementary materials for this article are available online.

1. Introduction

Pairwise comparison data arise frequently in network data analysis. They can be used to assist people in finding vital information underlying many modern interaction systems such as social webs and sports tournaments. For example, match points can be used to evaluate the team strengths in sports competitions. Consider \([n] = \{1, \ldots, n\}\) as \(n\) subjects in a network of interest, and assume that every \(i \in [n]\) is assigned a latent score \(u_i\), with \(u = (u_i)_{i \in [n]}\) denoting the corresponding score vector. Given pairwise comparison data, defined as a set of independent random variables \(\{X_{ij}\}_{1 \leq i < j \leq n}\), one wishes to accrue knowledge on \(u\) via statistical estimation processes. In the case of team ranking, \(u\) is a vector measuring the strength of \(n\) teams and \(X_{ij}\) is the competition outcome between teams \(i\) and \(j\), which is a binary random variable (denoting win and loss) depending on \(u_i\) and \(u_j\). Whereas in other scenarios such as online assessment, \(X_{ij}\) represents the average rating of subject \(i\) against subject \(j\). Under such circumstances, a continuous spectrum of the rating outcome is more appropriate.

Mutual comparison data from different sources may take different forms and can enhance our understanding of the model as long as they are relevant to \(u\), and this is usually manifested in assumptions on the link function between \(u\) and \(\{X_{ij}\}_{1 \leq i < j \leq n}\). Statistical methods can then be applied to estimate \(u\). The framework described here provides a simplified characterization of many parametric pairwise comparison models in the literature.

The study of parametric pairwise comparison models emerged in the early 20th century and gradually gained popularity. Among them, the Bradley–Terry (BT) model, first introduced in Bradley and Terry (1952), attracted much attention. The BT model is a specification of the team ranking example introduced before:

\[
\mathbb{P}\left(\text{team } i \text{ beats team } j \right) = \mathbb{P}\left( X_{ij} = 1 \right) = \Phi(u_i - u_j),
\]

where \(\Phi(\cdot)\) is the logistic link function defined as \(\Phi(x) = (1 + e^{-x})^{-1}\). In practice, not every pair of subjects admits a comparison, and even if it does, not all comparison data can be observed. To accommodate such sampling scenarios, the comparison graph is often assumed as an incomplete general graph; a comparison outcome between two subjects is observed if there is an edge between them. A common approach to modeling the comparison graph structure is through a generalized random graph. In particular, each pair \((i, j)\) is associated with an independent Bernoulli random variable \(n_{ij} \sim \text{Ber}(p_{ij,n})\) to determine the availability of \(X_{ij}\): \(n_{ij} = 1\) if \(X_{ij}\) is observed and \(n_{ij} = 0\) otherwise. Comparison rates \(p_{ij,n}\) are constants measuring the (edge) density of a network. When \(p_{ij,n}\) is independent of \(i\) and \(j\), namely \(p_{ij,n} = p_n\), the comparison graph is homogeneous and is called the Erdős-Rényi graph \(G(n, p_n)\). A detailed discussion of the BT models with random graph structure can be found in David (1988).

Despite its popularity, the BT model fails to capture the truth when \(X_{ij}\) are not binary. The specific choice of \(\Phi(x)\) also limits the use of the BT model in practice. To generalize, numerous variants of the BT model have been invented either by considering Likert-scale responses, or by replacing \(\Phi(x)\) with a different link function (Thurstone 1927; Rao and Kupper 1967; Davidson 1970; Stern 1990). Nevertheless, these are usually case-dependent and a unified treatment is yet to...
be found, which gives the motivation for the current article. Before going further to explain how to generalize the BT model as well as developing a consistent estimation theory under the generalized framework, we recall a few existing results in the BT model.

A natural estimator for $\mathbf{u}$ is the maximum likelihood estimator (MLE), which is denoted by $\hat{\mathbf{u}}$ and will be the main focus in this article. The consistency of the MLE in the BT model has been well studied when the comparison graph is the Erdős–Rényi graph $G(n, p_n)$. For instance, it was established in Simons and Yao (1999) and Yan, Yang, and Xu (2012) that the MLE is consistent in the $L_\infty$ norm, also called the uniformly consistent, if $\liminf_{n \to \infty} p_n > 0$. Consequently, for a network of $n$ subjects, at least $cn^2$ ($c > 0$) samples are needed to ensure the convergence of the MLE based on their results. The quadratic requirement on samples can be restrictive even when $n$ is only moderately large. In fact, many large-scale networks arising from realistic applications are sparse, that is, the degree of most subjects is sublinear in the size of the network.

To address the issue of sparsity, some researchers studied consistency of the MLE under a weaker condition on $p_n$, but in a different metric; see Maystre and Grossglauser (2015) and Negahban, Oh, and Shah (2017). They showed that $\|\hat{\mathbf{u}} - \mathbf{u}\|_2$ cannot grow faster than $\sqrt{n(\log n)^{\kappa}}$ if $p_n \geq n^{-1}(\log n)^{\kappa+1}$ ($\kappa > 0$). A similar result for the $L_2$ norm weighted by the graph Laplacian is obtained in Shah et al. (2016) without requiring homogeneity on $p_{ij,n}$. Although the sparsity condition in these results is optimal, the (weighted) $L_2$ norm only reflects the averaging behavior of the estimator, from which one cannot deduce convergence for each component unless the scores of all subjects are of the same order.

Recently, Han et al. (2020) and Chen et al. (2019) made progress by establishing the uniform consistency of the MLE and the regularized MLE in sparse homogenous BT models, respectively. Their analysis heavily depends on the special parameterization of the BT model as well as the homogeneity assumption on the comparison rates. It is also worth noting that the uniform consistency of the regularized MLE mentioned above does not directly imply the same result for the MLE. Chen, Gao, and Zhang (2020) improved the sparsity condition in Han et al. (2020) and showed that the MLE is superior to the spectral method in Negahban, Oh, and Shah (2017) in terms of the multiplicative constant factors in the sample complexity.

In this article, we develop a uniform consistency theory of the MLE for a general class of comparison models under a near-optimal sparsity condition. Our contribution can be briefly summarized as follows:

- We build a general probabilistic framework for pairwise comparison network data analysis. Our framework enjoys sufficient flexibility in terms of model parameterization, covering a wide variety of existing models in the literature such as the BT model, the Thurstone–Mosteller model (Thurstone 1927; Mosteller 2006), the Davidson model (Davidson 1970) and many others.
- Under this framework, we identify a sufficient condition for the uniform consistency of the MLE for the latent score vector. Our condition can be succinctly characterized using the graph topology and is satisfied in many random graph ensembles with varying parameters. In particular, when the sampling model is the Erdős–Rényi graph $G(n, p_n)$, we show that the comparison rate $p_n$ can be chosen as small as of order $(\log n)^{3+\epsilon}/n$ ($\epsilon > 0$) to ensure the uniform consistency of the MLE (with convergence rate at least of order $(\log n)^{-c/2}$), matching the graph connectivity threshold $\log n/n$ up to logarithmic factors.

The uniform consistency results guarantee the entry-wise convergence of the estimator at a uniform rate. Our approach decouples the randomness in pairwise comparison modeling and graph sampling, demonstrating a deep connection between graph topology and consistency of the MLE. On the practical side, our results justify that the MLE can be used in large-scale complex network estimation even if comparison data are asymptotically deficient, that is, the ratio of the observed comparisons and the theoretical total comparisons goes to zero (at a certain rate) as the number of subjects goes to infinity. In addition, numerical results are provided in strong support of our theory.

The rest of this article is organized as follows. Section 2 introduces a general framework for parametric pairwise comparison models and an inhomogeneous assumption on the comparison graph. Section 3 establishes both the unique existence and uniform consistency of the MLE in the setup introduced in Section 2 under a near-optimal sparsity condition. Section 4 shows that uniform consistency of the MLE in many existing models can be deduced as corollaries from our results in Section 3. Section 5 discusses a few graph topological conditions that will be used to formulate the uniform consistency result. Sections 6 and 7 are devoted to providing numerical experiments to support our theoretical findings and discussing future research directions, respectively. The proofs of our the main results can be found in the supplementary material.

### 1.1. Notation

For $m \in \mathbb{N}$, we write $[1, \ldots, m]$ as $[m]$. For sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$, $a_n \lesssim b_n$ if there exists an absolute constant $C$ such that $a_n \leq Cb_n$ for all $n \in \mathbb{N}$. Particularly, $a_n = \Omega(b_n)$ if $a_n \gtrsim b_n$ and $b_n \lesssim a_n$. For a univariate function $f(x)$, $f'(x)$ and $f''(x)$ denote the first and second derivative of $f(x)$, respectively. For a function with two arguments $f(x; y)$, for $i \in [2]$, $f_i(x; y)$ denotes the derivative of $f(x; y)$ with respect to the $i$th argument. We reserve $\mathbf{u}$ and $\hat{\mathbf{u}}$ for the true and the MLE of the latent score vector in the pairwise comparison model, respectively.

We use $G = (V, E)$ to denote an undirected connected graph where $V$ is the vertices set and $E$ is the edges set. In this work, we allow the existence of multiple edges in $E$. For any $U \subseteq V$, its boundary edges are defined as $\partial U = \{(i, j) \in E : i \in U, j \in U^c\}$. Moreover, for $U_1, U_2 \subseteq V$, we use $E(U_1, U_2)$ to denote the set of cross edges between $U_1$ and $U_2$, that is, $E(U_1, U_2) = \{(i, j) \in E : i \in U_1, j \in U_2\}$. Specifically, $E(U, U^c) = \partial U$.

### 2. Problem Setup

In this section, we introduce a general framework for analyzing pairwise comparison networks.
2.1. Pairwise Comparisons

Let \( u \in \mathbb{R}^n \) be the latent score vector of the subjects of interest. In the general framework, the comparison outcome between subjects \( i, j \in [n] \) is modeled via a random variable \( X_{ij} \), with density (mass) function given by \( f(x; u_i - u_j) \) for some valid function \( f \), which is defined below. Note that the distribution of \( X_{ij} \) depends only on \( u_i - u_j \), which is called the relative score between \( i \) and \( j \).

**Definition 1.** A function \( f : A \times \mathbb{R} \to \mathbb{R}^+ \), where \( A \) is a symmetric subset of \( \mathbb{R} \) denoting the possible comparison outcomes, is said to be **valid** if it satisfies the following assumptions:

**Assumption 1.** For \( y \in \mathbb{R} \), \( \int_A f(x; y) \, dx = 1 \) if \( A \) is continuous, and \( \sum_{x \in A} f(x; y) = 1 \) if \( A \) is discrete.

**Assumption 2.** \( f(x; y) \) is even with respect to \( (x; y) \):

\[
  f(x; y) = f(-x; -y) \quad (x, y) \in A \times \mathbb{R}.
\]

**Assumption 3.** For \( x < 0 \), \( f(x; y) \) is decreasing in \( y \), and \( f(x; y) \to 0 \) as \( y \to \infty \).

**Assumption 4.** \( \sup_{y \in \mathbb{R}} f(x; y) < +\infty \) for every \( x \in A \).

Assumption 1 guarantees that \( \{f(x; y)\}_{y \in \mathbb{R}} \) is a family of probability density (mass) functions indexed by \( y \). Assumption 2 states that \( i \) beats \( j \) by \( x \) is the same as that \( j \) beats \( i \) by \(-x \). Assumption 3 implies that a large relative score makes the comparison outcome more predictable. Assumption 4 is unconditionally true when \( A \) is discrete, and is generic for continuous \( A \) under appropriate regularity conditions on \( f \).

In accordance with the terms in the comparison data analysis, we say subject \( i \) beats \( j \) if we observe \( X_{ij} > 0 \). If we assign “beat” and “not beat” as a binary relation among the subjects, this binary relation is strongly stochastically transitive (Davidson and Marschak 1959; Fishburn 1973), that is, if \( \mathbb{P}(X_{ij} > 0) \geq 1/2 \) and \( \mathbb{P}(X_{jk} > 0) \geq 1/2 \), then

\[
  \mathbb{P}(X_{ik} > 0) \geq \max\{\mathbb{P}(X_{ij} > 0), \mathbb{P}(X_{jk} > 0)\}.
\]

Stochastic transitivity ensures that latent scores directly translate into rankings in practice, which in our case is verified under Assumptions 1–4.

**Proposition 1 (strong stochastic transitivity).** Under Assumptions 1–4, the binary relation “beat” and “not beat” satisfies the strong stochastic transitivity.

**Proof.** According to Definition 1,

\[
  \mathbb{P}(X_{ij} > 0) = f(u_i - u_j), \quad \text{where} \quad f(y) = \int_{(0,\infty)} f(x; y) \, dx.
\]

According to Assumption 2 and 3, \( f(y) \) is an increasing function of \( y \) with \( f(0) \leq 1/2 \). Therefore, \( \mathbb{P}(X_{ij} > 0) \geq 1/2 \) implies \( u_i \geq u_j \) for any \( i, j \in [n] \). The strong stochastic transitivity follows from \( u_i - u_k \geq \max\{u_i - u_j, u_j - u_k\} \).

It is worth mentioning that none of the above assumptions requires comparison data to be discrete or continuous. In particular, various choices of \( A \) in the literature fit here:

- Binary outcome: \( A = \{-1, 1\} \);
- Multiple outcome: \( A = \{-k, \ldots, k\} \) where \( k \in \mathbb{Z}^+ \) or \( A = \mathbb{Z} \);
- Continuous outcome: \( A = [-a, a] \) where \( a \in \mathbb{R}^+ \) or \( A = \mathbb{R} \).

**Remark 1.** Some works use nonsymmetric sets to parameterize the outcomes. For example, in Dittrich et al. (2007), a 5-point Likert scale (\( A = \{5\} \) was employed to represent different levels of preference. This is equivalent to \( A = \{-2, -1, 0, 1, 2\} \) in our case. The symmetry of \( A \) is not special but will make our analysis more convenient and statements more elegant.

**Remark 2.** A similar idea appeared in Shah et al. (2017) and Heckel et al. (2019) when considering binary comparison problems. Specifically, they assume that the distribution of \( X_{ij} \) is determined by some symmetric cumulative distribution function \( \Phi(t) \):

\[
  \mathbb{P}(X_{ij} = 1) = \Phi(u_i - u_j), \quad \mathbb{P}(X_{ij} = -1) = 1 - \Phi(u_i - u_j).
\]

This generalizes the logistic link function in the BT model and can be regarded as a special case under our setup with \( A = \{-1, 1\} \) and \( f(1, y) = \Phi(y) \).

2.2. Comparison Graphs

Random structures on comparison graphs could be added to the framework introduced in Section 2.1 closer to reality. A popular comparison graph structure hypothesizes that the number of comparisons between any pair of items follows Ber\((p_n)\) multiplied by some constant \( T \); see (Chen and Suh 2015; Jang et al. 2016; Negahban, Oh, and Shah 2017; Chen et al. 2019). When \( T = 1 \), their comparison graphs are Erdős-Rényi graphs \( G(n, p_n) \). The homogeneous assumption makes the analysis easier by avoiding pathological configurations. Nevertheless, statistical results obtained in this vein often have subtle dependence on the assumptions of sparse random graphs that shadows its connection to the graph topology.

In our framework, we take a slightly different approach by considering the generalized random graph model \( G(n, p_n, q_n) \) as follows:

**Definition 2.** \( G(n, p_n, q_n) \) is a random graph with vertices set \( V_n = [n] \) where each edge \((i, j) \in V_n \times V_n, i \neq j \) is formed independently with probability \( p_{ij, n} \in [p_n, q_n] \).

According to Definition 2, \( p_n \) and \( q_n \) could be taken as the minimum and maximum comparison rate, respectively, that is, \( p_n := \min_{i,j \in [n]} p_{ij, n} \) and \( q_n := \max_{i,j \in [n]} p_{ij, n} \). It is well known in Erdős and Rényi (1960) that \( G(n, p_n) \) is disconnected with high probability if \( p_n < (1 - \epsilon)n^{-1} \log n \), for any constant \( \epsilon > 0 \). As a result, there exist at least two connected components which cannot be estimated together unless additional constraints are imposed. Therefore, the connectivity threshold \( n^{-1} \log n \) is the best possible lower bound on \( p_n \) that one can hope for. Since \( G(n, p_n, q_n) \) reduces to the Erdős-Rényi graph \( G(n, p_n) \) when \( p_n = q_n \), it shares the same optimal lower bound on \( p_n \).

Given two subjects may have multiple comparisons between them, we assume that the number of comparisons between \( i \) and \( j, n_{ij} \), satisfies \( n_{ij} \sim \text{Bin}(T, p_{ij, n}) \). One can also adopt the setup in Chen et al. (2019) by assuming \( n_{ij} \sim T \times \text{Ber}(p_{ij, n}) \) and the
proof is similar. Since $T$ is a fixed constant, the reader could take $T = 1$ to avoid the additional multiplicative constant for ease of understanding.

**Remark 3.** We adopt $G(n, p_n, q_n)$ as the comparison graph when deriving the entry-wise error in Section 3. Compared to the Erdős-Rényi graph, our choice takes into account the potential heterogeneity of degree sequences. Our analysis in fact applies to a much wider class of comparison graphs beyond random graph models. A more comprehensive discussion on comparison graph structure will be carried out in Section 5. It is worth emphasizing that we require the comparison graph does not depend on the latent score $u$ in the whole article.

### 3. Main Results

We consider estimating $u$ via the MLE under the general framework introduced in Section 2. Let $\{n_{ij}\}_{i,j \in [n], i \neq j}$ be the number of comparisons observed between subjects $i$ and $j$. $\{X_{ij}^{(t)}\}_{t \in [n]}$ denote the outcomes between $i$ and $j$ in $n_{ij}$ comparisons. The conditional log-likelihood function given $\{n_{ij}\}_{i,j \in [n], i \neq j}$ is

$$l(v) = \frac{1}{2} \sum_{i,j \in [n]} \sum_{t \in [n]} \log f(X_{ij}^{(t)}; v_i - v_j),$$

where $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$. Since $l(v) = l(v + z)$ for $z \in \mathbb{R}$, an additional constraint on the parameter space is required to make $u$ identifiable. For this we take $v_1 = 0$; see (Cattelan 2012). The MLE $\hat{u}$ is defined as

$$\bar{u} = \arg\max_{v \in \mathbb{R}^n} l(v).$$

(1)

Note (1) only holds formally unless it admits a unique maximizer. We will show that, for sufficiently large $n$, $l(v)$ attains a unique maximum under an appropriate condition on $G(n, p_n, q_n)$.

#### 3.1. Existence and Uniqueness

**Condition 1.** Denote by $G = (V, E)$ the comparison graph where $V$ is the set of vertices and $E$ is the set of edges. For any nonempty subset $V_1 \subseteq V$, there exist $i \in V_1$, $j \in V \setminus V_1$ and $t \in [n_{ij}]$ such that $X_{ij}^{(t)} > 0$. In other words, for every partition of $V$ into two nonempty sets, a subject in the first set has defeated a subject in the second at least once.

Note that Condition 1 is well known in the BT model (Zermelo 1929; Ford Jr. 1957). We show Condition 1 is also enough to ensure the unique existence of $\hat{u}$ in the generalized model, which is stated in the following lemma.

**Lemma 1.** Under Condition 1, $\bar{u}$ defined in (1) uniquely exists.

Next, we demonstrate the Condition 1 holds almost surely under certain assumption. The dynamic range of $u$ plays a key role in our analysis and is defined by

$$M_n := \max_{i,j} |u_i - u_j|.$$
**Theorem 2 (uniform consistency, bounded case).** Suppose that \( f(x; y) \) is strictly log-concave with respect to \( y \), and
\[
\Delta_n := \omega_n \sqrt{q_n^2 (\log n)^3 / np_n^3} \to 0 \quad \text{as} \quad n \to \infty,
\]
where \( \omega_n \) is defined in (3). If (2) holds true, then there exists an absolute constant \( C > 0 \), such that for sufficiently large \( n \), with probability at least \( 1 - n^{-2} \), \( \hat{u} \) uniquely exists and satisfies
\[
\| \hat{u} - u \|_\infty \leq C \Delta_n.
\]

In other words, \( \hat{u} \) is a uniformly consistent estimator for \( u \).

Although we target at a general graph \( G(n, p_n, q_n) \), it is interesting to obtain a simplified lower bound on \( p_n \) when \( q_n \lesssim p_n \) and \( \max\{M_n, \omega_n\} \lesssim 1 \):

**Corollary 1.** Suppose \( \max\{M_n, \omega_n, q_n/p_n\} \lesssim 1 \). If there exists \( \epsilon > 0 \) such that
\[
p_n \gtrsim \frac{(\log n)^{3+\epsilon}}{n},
\]
then \( \hat{u} \) uniquely exists a.s. for all but finitely many \( n \), and is uniformly consistent for \( u \) with convergence rate at least of order \( (\log n)^{-\epsilon/2} \).

**Remark 5.** When \( A \) is bounded, \( \omega_n < \infty \) can be easily satisfied by imposing some regularity conditions on \( g \). In fact, if both \( g_1(x; y) \) and \( g_2(x; y) \) are continuous functions on \( A \times \mathbb{R} \), or \( g_2(x; y) \) is a continuous function in \( y \) and \( A \) is a finite set, one can deduce that \( C_n^{(2)} < \infty \) and \( C_n^{(3)} > 0 \), which follows from the fact that \( A \times [-M_n - 1, M_n + 1] \) is contained in a compact set in \( \mathbb{R}^2 \) and the strict log-concavity assumption on \( f \).

**Remark 6.** Corollary 1 gives a lower bound on \( p_n \) which only differs from the theoretical possible lower bound \( n^{-1} \log n \) by a logarithmic factor, implying that (5) stated in Corollary 1 is almost optimal. The effective lower bound in our result is \( n^{-1}(\log n)^3 \), and the larger the \( p_n \), the faster the convergence rate.

### 3.3. Uniform Consistency for General \( A \)

We now consider the case when \( A \) is unbounded. Note that Theorem 2 is vacuous unless \( C_n^{(2)} < \infty \) and \( C_n^{(3)} > 0 \), which can be easily verified under proper regularity assumptions when \( A \) is bounded. As we will see next, the same remains true when \( A \) is unbounded except for \( C_n^{(2)} < \infty \), which requires an extra condition to be imposed.

**Lemma 2.** Let \( f(x; y) = h(x) \exp \left\{ yT(x) - a(y) \right\} \) be an exponential family, where \( y \) is the natural parameter and \( T(x) \) is the sufficient statistic. Let \( X_{xy} \) be a random variable whose distribution is \( f(x; y) \). Suppose that \( V(y) := \var[T(X_{xy})] > 0 \) is a continuous function in \( y \). Then \( f(x; y) \) is strictly log-concave, and \( C_n^{(3)} > 0 \).

**Proof.** It is easy to see from direct computation that
\[
g_2(x; y) = -a''(y) = -\var[T(X_{xy})] = -V(y).
\]

The following theorem establishes the uniform consistency of the MLE for general \( A \):

**Theorem 3 (Uniform consistency of the MLE, general case).** Suppose that \( f(x; y) \) is strictly log-concave with respect to \( y \). Suppose that Condition 2 holds and
\[
\Delta_n := \omega_n \sqrt{q_n^2 (\log n)^3 / np_n^3} \to 0 \quad \text{as} \quad n \to \infty,
\]
where \( \omega_n \) is defined in (6). If (2) holds true, then there exists an absolute constant \( C > 0 \), such that for sufficiently large \( n \), with probability at least \( 1 - n^{-2} \), \( \hat{u} \) uniquely exists and satisfies
\[
\| \hat{u} - u \|_\infty \leq C \Delta_n.
\]

In other words, \( \hat{u} \) is a uniformly consistent estimator for \( u \).

**Remark 7.** Although Theorem 3 may seem restrictive, it covers many common unbounded models of interest, that is, the exponential family parameterization mentioned in Lemma 2 with sub-Gaussian property. The example of normal distribution is given in the following section.
4. Examples

In this section, we demonstrate that a number of well-known parametric pairwise comparison models can be covered within our setup. Particularly, the uniform consistency of the MLE can be proved for all these models provided $p_n$ is reasonably sparse. To the best of our knowledge, except for the BT model, the uniform consistency property of all the other models in this section is new.

4.1. Binary Outcomes

We first consider the BT model and the Thurstone–Mosteller model (Thurstone 1927; Mosteller 2006). In both cases, $A = \{-1, 1\}$ and $f(x; y)$ is given by

\[ f(1; y) = \Phi(y) \quad f(-1; y) = 1 - \Phi(y), \]

with

\[ \Phi(y) = \frac{e^y}{1 + e^y}, \quad \text{(Bradley–Terry model)}, \]

\[ \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} dx, \quad \text{(Thurstone–Mosteller model)}. \]

It is easy to check that in both cases $f(x; y)$ is valid and strictly log-concave with respect to $y$. According to Theorem 2, we have the following results:

**Corollary 2 (Bradley–Terry model).** In the Bradley–Terry model, if $q_n, p_n$ and $M_n$ satisfy the following bound:

\[ \Delta_n = e^{M_n} \sqrt{\frac{q_n^2 \log(n)^3}{np_n^3}} \to 0 \quad \text{as} \quad n \to \infty, \]

then for sufficiently large $n$, with probability at least $1 - n^{-2}$, $\hat{u}$ uniquely exists and satisfies $\|\hat{u} - u\|_\infty \lesssim \Delta_n$. In particular, when $M_n = \Omega(1)$ and taking $G(n, p_n, q_n)$ as the Erdős–Rényi graph, $p_n \gtrsim n^{-1}(\log n)^{3+\epsilon}$ for some $\epsilon > 0$ is sufficient for the uniform consistency of $\hat{u}$, namely, $\|\hat{u} - u\|_\infty \lesssim (\log n)^{-\epsilon/2}$.

**Corollary 3 (Thurstone–Mosteller model).** In the Thurstone–Mosteller model, if $q_n, p_n$ and $M_n$ satisfy

\[ \Delta_n = e^{M_n^2/2} \sqrt{\frac{q_n^2 \log(n)^3}{np_n^3}} \to 0 \quad \text{as} \quad n \to \infty, \]

then for sufficiently large $n$, with probability at least $1 - n^{-2}$, $\hat{u}$ uniquely exists and satisfies $\|\hat{u} - u\|_\infty \lesssim \Delta_n$. In particular, when $M_n = \Omega(1)$ and taking $G(n, p_n, q_n)$ as the Erdős–Rényi graph, $p_n \gtrsim n^{-1}(\log n)^{3+\epsilon}$ for some $\epsilon > 0$ is sufficient for the uniform consistency of $\hat{u}$, namely, $\|\hat{u} - u\|_\infty \lesssim (\log n)^{-\epsilon/2}$.

In both models, condition (4) implies condition (2). Corollary 3 gives the first entry-wise error bound for the MLE in the Thurstone–Mosteller model while Corollary 2 provides the entry-wise error bound for the MLE in the BT model under a general comparison graph. It is worth pointing out that the uniform consistency of the MLE in the BT model has been well studied under the structure of the Erdős–Rényi graphs. Although we focus on comparison models with more general structure, our result is effectively the same as Han et al. (2020) and only slightly worse than Chen, Gao, and Zhang (2020) up to logarithmic factors. With some technical refinement of our proof we can exactly recover the result in Han et al. (2020).

**Remark 8.** Both our work and Han et al. (2020) use the idea of chaining to deal with the entry-wise error, which builds an upward nested sequence of vertices at the ends of the estimation error spectrum. However, a direct application of the proof in Han et al. (2020) in our case will lead to an overlapping scenario that is difficult to analyze. Instead, our new proof exploits a symmetry property (Assumption 2) to avoid the technical difficulty.

4.2. Multiple Outcomes

Agresti (1992) provided two generalizations the BT model by taking account of multiple outcomes (ordinal data): one is the cumulative link model and the other is the adjacent categories model. It can be verified that both models satisfy the validity and strict log-concavity assumptions which lead to the uniform consistency of the corresponding MLE under our framework. Specifically, when there are only three outcomes ("win," "tie," "loss," or $A = \{-1, 0, 1\}$), the cumulative link model and the adjacent categories model reduce to the Rao-Kupper model (Rao and Kupper 1967) and the Davidson model (Davidson 1970), respectively. In this section, we prove the uniform consistency of the MLE for these two models.

The link function $f(x; y)$ in the Rao–Kupper model is given by

\[ f(1; y) = \frac{e^\theta}{e^\theta + \theta}; \quad f(0; y) = \frac{\theta e^\theta}{(\theta + \theta)(\theta e^\theta + 1)}; \]

\[ f(-1; y) = \frac{1}{\theta e^\theta + 1}, \]

where $\theta > 1$ is the threshold parameter which is assumed to be fixed. The following corollary is straightforward from Theorem 2.

**Corollary 4 (Rao–Kupper model).** In the Rao–Kupper model, for fixed $\theta > 1$, if $q_n, p_n$ and $M_n$ satisfy

\[ \Delta_n = e^{M_n} \sqrt{\frac{q_n^2 \log(n)^3}{np_n^3}} \to 0 \quad \text{as} \quad n \to \infty, \]

then for sufficiently large $n$, with probability at least $1 - n^{-2}$, $\hat{u}$ uniquely exists and satisfies $\|\hat{u} - u\|_\infty \lesssim \Delta_n$. In particular, when $M_n = \Omega(1)$ and taking $G(n, p_n, q_n)$ as the Erdős–Rényi graph, $p_n \gtrsim n^{-1}(\log n)^{3+\epsilon}$ for some $\epsilon > 0$ is sufficient for the uniform consistency of $\hat{u}$, namely, $\|\hat{u} - u\|_\infty \lesssim (\log n)^{-\epsilon/2}$.

As opposed to the Rao–Kupper model, the Davidson model (Davidson 1970) considers an alternative outcome of being tie:

\[ f(1; y) = \frac{e^\theta}{e^\theta + \theta e^\frac{\theta}{2} + 1}; \quad f(0; y) = \frac{\theta e^\frac{\theta}{2}}{e^\theta + \theta e^\frac{\theta}{2} + 1}; \]

\[ f(-1; y) = \frac{1}{e^\theta + \theta e^\frac{\theta}{2} + 1}, \]

where $\theta > 0$ is assumed to be fixed. Similarly, we have the following corollary.
\textbf{Corollary 5 (Davidson model).} In the Davidson model, for fixed $\theta > 0$, if $q_n, p_n$ and $M_n$ satisfy
\begin{equation*}
\Delta_n = e^{M_n/2} \frac{q_n^3}{np_n^3} \log (n)^3 \to 0 \quad \text{as } n \to \infty,
\end{equation*}
then for sufficiently large $n$, with probability at least $1 - n^{-2}$, $\hat{u}$ uniquely exists and satisfies $\|\hat{u} - u\|_\infty \lesssim \Delta_n$. In particular, when $M_n = \Omega(1)$ and taking $G(n, p_n, q_n)$ as the Erdős–Rényi graph, $p_n \geq n^{-1} (\log n)^{3+\epsilon}$ for some $\epsilon > 0$ is sufficient for the uniform consistency of $\hat{u}$, namely, $\|\hat{u} - u\|_\infty \lesssim (\log n)^{-\epsilon/2}$.

Similar to the examples in the previous section, condition (4) implies condition (2) for both the Rao–Kupper model and the Davidson model. Utilizing a similar argument, the uniform consistency result can be proved for the extensions of the Thurstone–Mosteller model (Maydeu-Olivares 2001, 2002), which we do not state here.

\subsection{4.3. Continuous Outcomes}

We consider the paired cardinal model introduced and studied in Shah et al. (2016). In this case, $A = \mathbb{R}$, and $X_{ij}$ follows a normal distribution with mean $u_i - u_j$ and variance $\sigma^2$, that is, $X_{ij} \sim \mathcal{N}(u_i - u_j, \sigma^2)$ and
\begin{equation*}
f(x; y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-y)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.
\end{equation*}

Cardinal models take a continuous spectrum of measurement values and often contain more information than ordinal models under certain conversion assumptions (Shah et al. 2016). Particularly, the Thurstone–Mosteller model studied in Section 4.1 can be viewed as a thresholded version of the paired cardinal model; the binary outcome between $i$ and $j$ corresponds to the cases where $X_{ij} > 0$ and $X_{ij} \leq 0$.

Validity and log-concavity assumptions on $f$ are easy to verify for the paired cardinal model. To apply Theorem 3, we have

\textbf{Corollary 6 (Paired cardinal model).} In the paired cardinal model, for fixed $\sigma > 0$, if $q_n, p_n$ and $M_n$ satisfy
\begin{equation*}
M_n e^{\frac{3\sigma^2}{2n^2}} \log n \to 0 \quad \text{as } n \to \infty,
\end{equation*}
and
\begin{equation*}
\Delta_n = e^{M_n/2} \frac{q_n^3}{np_n^3} \log (n)^3 \to 0 \quad \text{as } n \to \infty,
\end{equation*}
then for sufficiently large $n$, with probability at least $1 - n^{-2}$, $\hat{u}$ uniquely exists and satisfies $\|\hat{u} - u\|_\infty \lesssim \Delta_n$. In particular, when $M_n = \Omega(1)$ and taking $G(n, p_n, q_n)$ as the Erdős–Rényi graph, $p_n \geq n^{-1} (\log n)^{3+\epsilon}$ for some $\epsilon > 0$ is sufficient for the uniform consistency of $\hat{u}$, namely, $\|\hat{u} - u\|_\infty \lesssim (\log n)^{-\epsilon/2}$.

\section{5. Comparison Graph Structure}

In this section, we summarize two topological conditions of a graph that will be used in the consistency analysis. In particular, we first assume that comparison graphs are deterministic and enjoy certain topological properties; we then verify these properties in several classes of random graph models of practical interest. Our approach decouples the randomness in the comparison model and comparison graph, thus is more transparent in terms of illustrating how graph topology influences the statistical procedures in pairwise data analysis.

Some definitions and notations in the graph theory are introduced here (Chung 1997). Let $G = (V, E)$ be an undirected graph. For $U \subset V$, the connectivity between $U$ and $U^c$ can be measured by the following ratio cut $h_G(U)$:
\begin{equation*}
h_G(U) = \frac{|\partial U|}{\min(|U|, |U^c|)}.
\end{equation*}
A large value of $h_G(U)$ suggests that $U$ and $U^c$ are well-connected. The global connectivity of $G$ can be measured by taking the minimum of $h_G(U)$ over $U$,
\begin{equation*}
h_G = \min_{U \subset V} h_G(U),
\end{equation*}
which is called the modified Cheeger constant or the isoperimetric number, of $G$.

In the analysis of comparison graph models, we are concerned with the asymptotic behavior of the model, which requires us to work with a sequence of graphs. For convenience, we let $\{G_n\}_{n \in \mathbb{N}}$ denote a graph sequence where $G_n = (V_n, E_n)$ with $|V_n| = n$. We now introduce two topological properties that are central to our analysis:

\textbf{Definition 4 (ASC).} A sequence of graphs $\{G_n\}_{n \in \mathbb{N}}$ is said to be asymptotically strongly connected (ASC) with rate $\{\omega_n\}_{n \in \mathbb{N}}$ if
\begin{equation*}
\lim_{n \to \infty} \omega_n \Gamma_{G_n}^{ASC} = 0, \quad \text{where } \Gamma_{G_n}^{ASC} = \sqrt{\frac{\log n}{h_{G_n}}}.
\end{equation*}

\textbf{Definition 5 (RE).} For $G = G(V, E)$, an upward nested sequence of nonempty vertices set $\{A_k\}_{k=1}^K$ (that is $A_k \subseteq A_{k+1} \subseteq V$) is called admissible if $|\mathcal{E}(A_k, A_{k+1} \setminus A_k)| \geq |\partial A_k|/2$ for all $k < K$. Denote the set of admissible sequences of $G$ as $\mathcal{A}(G)$. $\{G_n\}_{n \in \mathbb{N}}$ is called rapidly expanding (RE) with rate $\{\omega_n\}_{n \in \mathbb{N}}$ if
\begin{equation*}
\lim_{n \to \infty} \omega_n \Gamma_{G_n}^{RE} = 0, \quad \text{where } \Gamma_{G_n}^{RE} = \max_{\{A_k\}_{k=1}^K \in \mathcal{A}(G_n)} \sum_{k=1}^{K-1} \sqrt{\frac{\log n}{h_{G_n}(A_k)}}.
\end{equation*}

ASC is a global property that requires “small” subsets of $G_n$ have relatively large edge boundary as $n \to \infty$, and RE is a cumulative version of ASC (defined for all rapidly expanding sequences). It is easy to verify that RE with rate $\{\omega_n\}_{n \in \mathbb{N}}$ implies ASC with the same rate by taking any admissible sequence with $A_1$ that satisfies $h_{G_n}(A_1) = \omega_n$. Note that an admissible sequence $\{A_k\}_{k=1}^K$ is strictly increasing by definition, therefore there exists a natural upper bound for $K$, that is, $K \leq n$.

A sufficient condition for the uniform consistency result can be formulated using RE, with the convergence rate of $\Gamma_{G_n}^{RE}$ (that is, $\omega_n$) appropriately chosen to encode the information of the pairwise comparison parameterization. The uniform consistency result for the MLE can be stated as follows:
**Theorem 4 (uniform consistency).** Suppose \( f(x; y) \) is strictly log-concave with respect to \( y \) and Conditions 1 and 2 hold. If the comparison graph sequence \( \{G_n\}_{n \in \mathbb{N}} \) is RE with rate \( \{\omega_n\}_{n \in \mathbb{N}} \), where \( \omega_n \) is defined in (6), that is,

\[
\Delta_n^{\text{RE}} := \omega_n \Gamma_n^{\text{RE}} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty,
\]

then there exists an absolute constant \( C > 0 \), such that for sufficiently large \( n \), with probability at least \( 1 - n^{-2} \), \( \hat{u} \) uniquely exists and satisfies

\[
\| \hat{u} - u \|_{\infty} \leq C \Delta_n^{\text{RE}}. \tag{7}
\]

In other words, \( \hat{u} \) is a uniformly consistent estimator for \( u \).

**Remark 9.** Theorem 4 requires the comparison graph sequence \( \{G_n\}_{n \in \mathbb{N}} \) is RE. The requirement \( \| E(A_k, A_{k+1} \setminus A_k) \| \geq |\partial A_k|/2 \) in the definition of admissible sequences could be changed to \( \| E(A_k, A_{k+1} \setminus A_k) \| \geq |\partial A_k|/(1 + \epsilon) \) for any absolute constant \( \epsilon > 0 \). In that case, the result in Theorem 4 remains unchanged up to a multiplicative constant in (7).

Theorem 4 provides the uniform consistency of the MLE in the generalized comparison model under a general comparison graph. The convergence rate consists of two parts: \( \omega_n \) and \( \Gamma_n^{\text{RE}} \). \( \omega_n \) is defined as \( \min\{c_n^{(4)}, c_n^{(2)}\} \), which relies only on the comparison model \( f \) and the dynamic range of \( u \). \( \Gamma_n^{\text{RE}} \) is concerned with the topological property of the comparison graph. Compared with Theorems 3, Theorem 4 replaces \( \Delta_n \) with \( \Delta_n^{\text{RE}} \). Consequently, Theorem 3 implies Theorem 4 if the \( \Gamma_n^{\text{RE}} \) in \( G(n, p_n, q_n) \) is bounded by \( \sqrt{q_n^2 (\log n)^3} / np_n^3 \). To further demonstrate the utility of Theorem 4, we also prove that RE is satisfied in the stochastic block model (Holland, Laskey, and Leinhardt 1983) with an additional structure. These results are summarized in the following proposition:

**Proposition 2.** Let \( \{G_n\}_{n \in \mathbb{N}} \) be a (random) graph sequence. For all sufficiently large \( n \), the following events hold with probability at least \( 1 - n^{-2} \):

1. If \( G_n = G(n, p_n) \), then \( \Gamma_n^{\text{RE}} \leq \sqrt{\log n}^3 / np_n^3 \).
2. If \( G_n = G(n, p_n, q_n) \) (see Definition 2), then \( \Gamma_n^{\text{RE}} \leq \sqrt{q_n^2 (\log n)^3} / np_n^3 \).
3. If \( G_n \) is a stochastic block model with finite blocks and its lower bound of edge density is \( p_n \), then \( \Gamma_n^{\text{RE}} \leq \sqrt{\log n}^3 / np_n^3 \).

6. **Numerical Results**

In this section, we first conduct numerical simulations to evaluate the large-sample performance of the MLE in the Davidson model with threshold parameter \( \theta = 1 \) and the paired cardinal model with variance parameter \( \sigma^2 = 1 \). Since extensive numerical results exist for both models using real datasets (Shah et al. 2016; Agresti 2019), our simulations are more focused on the synthetic data, which mainly serve to verify the asymptotic results in Sections 3 and 4. The corresponding results are reported in Section 6.1. Moreover, as our framework provides ample flexibility for model parameterization, it is tempting to test different model parameterizations on a dataset and select the optimal one for use in practice using model selection methods. We empirically investigate this problem on a real dataset in Section 6.2.

**6.1. Asymptotic Performance**

We first test the asymptotic uniform convergence of the MLE when the network is sparse. Note that a large value of \( T \) can inadvertently make the network dense even if \( p_n \) is small. As such, we set \( T = 1 \) in the following simulations. The comparison graph model is set as \( G(n, p_n, q_n) \), with its size chosen in an increasing manner to demonstrate the expected convergence. Specifically, we test on six different values for \( n \): 2000, 4000, 6000, 8000, 10,000, and 12,000. For each \( n \), the latent score vector \( u \) is generated independently with its components from the uniform distribution on \([-0.5, 0.5]\), which guarantees that \( M_n \leq 1 \). The minimum comparison rate \( p_n \) is taken as \( n^{-1}(\log n)^3, \sqrt{n^{-1}(\log n)^3} \), and \( 1/2 \), corresponding to the underlying network being sparse, moderately sparse and dense, respectively. For convenience, we let the maximum comparison rate \( q_n \) be proportional to \( p_n \), that is, \( q_n = 2p_n \). Values of \( p_n \) under different \( n \) are presented in Table 1.

For every fixed \( n, p_n \) and \( u \), the comparison data is generated under the respective model with \( \hat{u} \) computed using a minorization–maximization (MM) algorithm in Hunter (2004). We then calculate the \( \ell_\infty \) error \( \| \hat{u} - u \|_{\infty} \). To check uncertainty, for each \( n \) and \( p_n \), the experiment is repeated 300 times with its quartiles recorded. The results are reported in the first two plots in Figure 1.

In both models and three different sparsity regimes, \( \| \hat{u} - u \|_{\infty} \) decreases to 0 as \( n \) grows to infinity. This numerically verifies the uniform consistency of the MLE as proved in Theorems 2 and 3. Another observation, which is not unexpected, is that the convergence rate of the MLE closely depends on the density parameter \( p_n \): the larger the \( p_n \), the faster the convergence. Particularly, when \( p_n \) is chosen at the critical level obtained in our analysis, \( \| \hat{u} - u \|_{\infty} \) decays rather slowly compared to the denser regimes. Such drawback seems mitigated by increasing the size of the network, suggesting that networks with extremely large size would be more tolerable for a low comparison rate. This demonstrates the potential applicability of our results in studying large complex networks (such as social networks) using under-observed comparison data.

We also investigate how the convergence of the MLE depends on the varying dynamic range \( M_n \). To do so, we fix \( p_n = 0.5 \) and take \( M_n \) as 1, \( \log \log n/2 \) and \( 2 \log \log n \), respectively. The values

**Table 1.** The value of \( p_n \) and \( M_n \) given the different \( n \).

| \( n \) | \( p_n = \sqrt{\log n}^3 / np_n^3 \) | \( M_n = \log \log n/2 \) | \( M_n = 2 \log \log n \) |
|---|---|---|---|
| 2000 | 0.469(937) | 0.220(439) | 1.014 | 4.057 |
| 4000 | 0.378(1511) | 0.143(571) | 1.058 | 4.231 |
| 6000 | 0.331(1988) | 0.110(658) | 1.082 | 4.327 |
| 8000 | 0.301(2410) | 0.091(726) | 1.098 | 4.392 |
| 10,000 | 0.280(2795) | 0.078(781) | 1.110 | 4.441 |
| 12,000 | 0.263(3153) | 0.069(829) | 1.120 | 4.480 |

**Note:** In addition, the average numbers of comparisons one subject has (in parentheses) is in the column of \( p_n \).
of $M_n$ under different $n$ can be found in Table 1. According to our results, small $M_n$’s are better for uniform consistency of the MLE in the Davidson model, which is numerically verified in Figure 1. As a contrast, the paired cardinal model seems not sensitive to the changing magnitude of $M_n$. This may be due to the fact that the convergence rate $\Delta_n$ in the paired cardinal model is independent of $M_n$; see Corollary 6.

### 6.2. ATP Data Analysis

In this section, we model a real pairwise comparison network using three different parameterizations. The dataset under our consideration is the ATP dataset\(^1\) from 2000 to 2018. The ATP match contains four Grand Slams, the ATP World Tour Masters 1000, the ATP World Tour 500 series and several tennis series of the year. For convenience, we focus on the Best of 3 (BO3) matches, which contain two or three sets and have four possible outcomes $A = \{-2, -1, 1, 2\}$. For example, the outcome 2:1 between $i$ and $j$ corresponds to $X_{ij} = 2 - 1 = 1$. As a result, we remove the competitions from Grand slams. In order to let Condition 1 hold, we tease out the players who never win or lose the games. After cleaning, the dataset includes nearly 26,000 competitions and 954 players.

For model parameterization, we consider three relevant models: the general BT model, the cumulative link model (CLM4) and the adjacent categories logit model (ACLM4) with four outcomes (Agresti 1992). Specifically, in the general BT model, if there is a comparison between $i$ and $j$, then

$$\mathbb{P} \text{(the outcome is 2:1)} = 2\Phi^2(u_i - u_j)(1 - \Phi(u_i - u_j)),$$

$$\mathbb{P} \text{(the outcome is 2:0)} = \Phi^2(u_i - u_j)$$

where $\Phi(x)$ is the logistic link function. It can be verified that (8) implies that the results of the three sets are independent, and each of them follows the same BT model. In CLM4,

$$\mathbb{P} \text{(the outcome is 2:1)} = \frac{(\theta - 1)e^{u_i - u_j}}{(\theta + e^{u_i - u_j})(1 + e^{u_i - u_j})}.$$
and developed in Section 2 and satisfy the log-concavity condition.

For example, if the validation data is the comparison between error in LOOCV and the cross-entropy (negative log-likelihood) for model evaluation. In particular, we choose the prediction criterion (BIC) and Leave-one-out cross-validation (LOOCV), Akaike information criterion (AIC), Bayesian information property. In particular, for almost homogeneous random graph consistent, which can be summarized as a graph topological provided a sufficient condition for the MLE to be uniformly log-concave with respect to the parameterization variable, we

\[
P(\text{the outcome is } 2:0) = \frac{\theta e^{\theta j - a}}{\theta + e^{\theta n - a}}, \quad \theta > 1.
\]

In ACLM4,

\[
P(\text{the outcome is } 2:1) = \frac{\theta^2 e^{2(\theta j - a)}}{1 + \theta e^{(\theta j - a)/3} + \theta e^{(\theta n - a)/3} + e^{a - n}},
\]

\[
P(\text{the outcome is } 2:0) = \frac{\theta e^{\theta j - a}}{1 + \theta e^{(\theta j - a)/3} + \theta e^{(\theta n - a)/3} + e^{a - n}}, \quad \theta > 0.
\]

It can be checked that all models are within the framework developed in Section 2 and satisfy the log-concavity condition.

We apply several model selection criteria, including the Akaike information criterion (AIC), Bayesian information criterion (BIC) and Leave-one-out cross-validation (LOOCV), for model evaluation. In particular, we choose the prediction error in LOOCV as the cross entropy (negative log-likelihood). For example, if the validation data is the comparison between i and j with outcome \(a^*\), then the prediction error is given as

\[
\text{error} = - \sum_{a \in A} 1_{\{a = a^*\}} \log f(a; \hat{u}_i - \hat{u}_j).
\]

The result is presented in Table 2. We observe that both ACLM4 and CLM4 yield a better performance than the general BT model in terms of AIC, BIC and LOOCV, for which a possible explanation is that the outcomes of different sets in the same match are not independent. Without assuming independence, both ACLM4 and CLM4 seem to lead to a better fit for the data. On the other hand, LOOCV reflects the overall prediction error. If we consider using random guessing as the benchmark, LOOCV of random guessing is 1.3863 (log 4). Consequently, all the models we have tested in this example (the general BT model, ACLM4 and CLM4) achieve better predictions than random guessing.

7. Discussion

In this article, we introduced a general framework for statistical network analysis using pairwise comparison data. Our framework enjoys abundant parameterization flexibility for practical purposes. Assuming the link function is valid and strictly log-concave with respect to the parameterization variable, we provided a sufficient condition for the MLE to be uniformly consistent, which can be summarized as a graph topological property. In particular, for almost homogeneous random graph models (e.g., Erdős-Rényi graphs), the condition is satisfied when \(p_n = \Omega(n^{-1}(\log n)^{3+\epsilon})\) for any \(\epsilon > 0\), which almost matches the best possible lower bound \(n^{-1}\log n\). We think the potential gap between our result and the lower bound is an artifact of our proof, which arises owing to a cumulative effect of the large deviation bound in the maximal inequality and the number of times that the maximal inequality is applied for chaining. A more elegant design for the proof is needed to avoid applying the maximal inequalities multiple times if one wants to pursue an optimal result. We leave it as one possible direction for future work.

Although the framework considered in this article is rather inclusive, there are also some other possible directions one can go to make it better. For example, one may incorporate a global parameter into the framework to model environmental factors. For instance, the home-field advantage model (Agresti 2019) contains a global parameter measuring the strength of home-field advantage which is not contained in the latent score vector \(u\). The distribution of the outcome will be different depending on which subject is at home.

Second, the assumption on pairwise comparison data can be generalized to multiple comparison data. For example, the Plackett–Luce model (Plackett 1975; Luce 1959) is the multiple-comparison version of the BT model. Compared to pairwise comparison models, multiple comparison models involve data measuring the interaction between more than two items in a single observation, resulting in the comparison graph being a hypergraph. This may cause difficulty in obtaining the asymptotic properties of the MLE. Particularly, the entry-wise error of the MLE in the multiple comparison models is currently elusive to us.

Finally, it is worth pointing out that although our framework allows great flexibility in terms of selecting the link function, it is generally unknown which choice fits the true model best. This poses the natural question of model selection among different valid parameterizations. In Section 6.2, we conducted an empirical study toward this direction using existing model selection techniques. However, it is not clear to us which method has the best capability of deciphering the most relevant model (both in theory and in practice), which we leave as a direction for future investigation.

Supplementary Materials

Detailed proofs of Theorems 1–4, Lemma 1 and Proposition 2.

Acknowledgments

The authors are very grateful to the Editor Prof. McKeague, the Associate Editor, and two anonymous referees for their very helpful comments which significantly improved the presentation of the article. The authors also thank Prof. Tom Alberts for going through an early version of the draft, Prof. Fan R. K. Chung for explaining a proposed graph condition in the manuscript and Prof. Zhigang Bao for helpful discussion.

Funding

Ruijian Han’s research is partially supported by the Hong Kong Research Grants Council (grant no. 14301821) and Direct Grants for Research, The Chinese University of Hong Kong. Kani Chen’s research is partially supported by the Hong Kong Research Grants Council (grant no. 16302881).

ORCID

Ruijian Han http://orcid.org/0000-0002-9225-2218
Yiming Xu http://orcid.org/0000-0001-7223-8147
Kani Chen http://orcid.org/0000-0003-0117-8065
References

Agresti, A. (1992), “Analysis of Ordinal Paired Comparison Data,” *Journal of the Royal Statistical Society*, Series C, 41, 287–297. [2427,2430]

——— (2019), *An Introduction to Categorical Data Analysis*, Wiley Series in Probability and Statistics (3rd ed.), Hoboken, NJ: Wiley. [2429]

Bradley, R. A., and Terry, M. E. (1952), “Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons,” *Biometrika*, 39, 324–345. [2422]

Cattelan, M. (2012), “Models for Paired Comparison Data: A Review with Emphasis on Dependent Data,” *Statistical Science*, 27, 412–433. [2425]

Chen, P., Gao, C., and Zhang, A. Y. (2020), “Partial Recovery for Top-k Ranking: Optimality of MLE and Sub-optimality of Spectral Method,” arXiv preprint arXiv:2006.16485. [2423,2427]

Chen, Y., Fan, J., Ma, C., and Wang, K. (2019), “Spectral Method and Regularized MLE are both Optimal for Top-k Ranking,” *The Annals of Statistics*, 47, 2204–2235. [2423,2424]

Chen, Y., and Suh, C. (2015), “Spectral MLE: Top-k Rank Aggregation from Pairwise Comparisons,” in *International Conference on Machine Learning*, pp. 371–380. [2424]

Chung, F. R. K. (1997), *Spectral Graph Theory* (Volume 92 of CBMS Regional Conference Series in Mathematics). Published for the Conference Board of the Mathematical Sciences, Washington, DC. Providence, RI: American Mathematical Society. [2428]

David, H. A. (1988), *The Method of Paired Comparisons* (Volume 41 of Griffin's Statistical Monographs & Courses, 2nd ed.), London: Charles Griffin & Co., Ltd.; New York: The Clarendon Press, Oxford University Press. [2422]

Davidson, D., and Marschak, J. (1959), “Experimental Tests of a Stochastic Decision Theory,” *Measurement: Definitions and Theories*, 17, 233–269. [2424]

Davidson, R. R. (1970), “On Extending the Bradley–Terry Model to Accommodate Ties in Paired Comparison Experiments,” *Journal of the American Statistical Association*, 65, 317–328. [2422,2423,2427]

Dittrich, R., Francis, B., Hatzinger, R., and Katzenbeisser, W. (2007), “A Paired Comparison Approach for the Analysis of Sets of Likert-scale Responses,” *Statistical Modelling*, 7, 3–28. [2424]

Erdős, P., and Rényi, A. (1960), “On the Evolution of Random Graphs,” *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének közleményei* 5, 17–61. [2424]

Fishburn, P. C. (1973), “Binary Choice Probabilities: On the Varieties of Stochastic Transitivity,” *Journal of Mathematical Psychology*, 10, 327–352. [2424]

Ford Jr., L. R. (1957), “Solution of a Ranking Problem from Binary Comparisons,” *The American Mathematical Monthly*, 64, 28–33. [2425]

Han, R., Ye, R., Tan, C., and Chen, K. (2020), “Asymptotic Theory of Sparse Bradley–Terry Model,” *Annals of Applied Probability*, 30, 2491–2515. [2423,2427]

Heckel, R., Shah, N. B., Ramchandran, K., and Wainwright, M. J. (2019), “Active Ranking from Pairwise Comparisons and When Parametric Assumptions do not Help,” *The Annals of Statistics*, 47, 3099–3126. [2424]

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983), “Stochastic Blockmodels: First Steps,” *Social Networks*, 5, 109–137. [2429]

Hunter, D. R. (2004), “MM Algorithms for Generalized Bradley–Terry Models,” *The Annals of Statistics*, 32, 384–406. [2429]

Jang, M., Kim, S., Suh, C., and Oh, S. (2016), “Top-k Ranking from Pairwise Comparisons: When Spectral Ranking is Optimal,” arXiv preprint arXiv:1603.04153. [2424]

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley; London: Chapman & Hall, Ltd. [2431]

Maydeu-Olivares, A. (2001), “Limited Information Estimation and Testing of Thurstonian Models for Paired Comparison Data Under Multiple Judgment Sampling,” *Psychometrika*, 66, 209–227. [2428]

——— (2002), “Limited Information Estimation and Testing of Thurstonian Models for Preference Data,” *Mathematical Social Sciences*, 43, 467–483. [2428]

Maystre, L., and Grossglauser, M. (2015), “Fast and Accurate Inference of Plackett–Luce Models,” in *Advances in Neural Information Processing Systems*, pp. 172–180. [2423]

Mosteller, F. (2006), “Remarks on the Method of Paired Comparisons I: The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations,” in Selected Papers of Frederick Mosteller, eds. S. E. Fienberg and D. C. Hoaglin, pp. 157–162, New York: Springer. [2423,2427]

Negahban, S., Oh, S., and Shah, D. (2017), “Rank Centrality: Ranking from Pairwise Comparisons,” *Operations Research*, 65, 266–287. [2423,2424]

Plackett, R. L. (1975), “The Analysis of Permutations,” *Journal of the Royal Statistical Society*, Series C, 24, 193–202. [2431]

Rao, P. V., and Kupper, L. L. (1967), “Ties in Paired-Comparison Experiments: A Generalization of the Bradley–Terry Model,” *Journal of the American Statistical Association*, 62, 194–204. [2422,2427]

Shah, N. B., Balakrishnan, S., Bradley, J., Parekh, A., Ramchandran, K., and Wainwright, M. J. (2016), “Estimation from Pairwise Comparisons: Sharp Minimax Bounds with Topology Dependence,” *Journal of Machine Learning Research*, 17: Paper No. 58, 47. [2423,2428,2429]

Shah, N. B., Balakrishnan, S., Guntuboyina, A., and Wainwright, M. J. (2017), “Stochastically Transitive Models for Pairwise Comparisons: Statistical and Computational Issues,” *IEEE Transactions on Information Theory*, 63, 934–959. [2424]

Simons, G., and Yao, Y.-C. (1999), “Asymptotics When the Number of Parameters Tends to Infinity in the Bradley–Terry Model for Paired Comparisons,” *The Annals of Statistics*, 27, 1041–1060. [2423]

Stern, H. (1990), “A Continuum of Paired Comparisons Models,” *Biometrika*, 77, 265–273. [2422]

Thurstone, L. L. (1927), “A Law of Comparative Judgment,” *Psychological Review*, 34, 273–286. [2422,2423,2427]

Yan, T., Yang, Y., and Xu, J. (2012), “Sparse Paired Comparisons in the Bradley–Terry Model,” *Statistica Sinica*, 22, 1305–1318. [2423]

Zermelo, E. (1929), “Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung,” *Mathematische Zeitschrift*, 29, 436–460. [2425]