Pairwise comparison models for ranking:
Statistical and computational issues

Martin Wainwright

UC Berkeley
Statistics and EECS

Joint work with: Nihar Shah
Sivaraman Balakrishnan
Aditya Guntuboyina
Pairwise ranking data

- collection of \( n \) items (e.g., football teams, chess players, web pages, politicians, cars)
Pairwise ranking data

- collection of $n$ items (e.g., football teams, chess players, web pages, politicians, cars)
- observe $Y_{ij} \in \{0, 1\}$ corresponding to $(i, j)$-comparison
Pairwise ranking data

- collection of $n$ items (e.g., football teams, chess players, web pages, politicians, cars)
- observe $Y_{ij} \in \{0, 1\}$ corresponding to $(i, j)$-comparison
- want to estimate matrix $M^* \in [0, 1]^{n \times n}$ of probabilities

$$M^*_{ij} = \mathbb{P}[i \text{ beats } j]$$
One “possible” underlying ranking:

- collection of $n$ items (e.g., football teams, chess players, web pages, politicians, cars)
- observe $Y_{ij} \in \{0, 1\}$ corresponding to $(i, j)$-comparison
- want to estimate matrix $M^* \in [0, 1]^{n \times n}$ of probabilities
  
  $$M^*_{ij} = P[i \text{ beats } j]$$

- also of interest: underlying ranking of $n$ items
Classical parametric models

- underlying weight vector $\omega^* \in \mathbb{R}^n_+$
  $\omega^*_j = $ quality assignment to item $j$

- generative model

$$\mathbb{P}[i \text{ beats item } j] = \Phi(\omega^*_i - \omega^*_j)$$

where $\Phi : \mathbb{R} \to [0, 1]$ is a CDF.
Classical parametric models

- underlying weight vector $\omega^* \in \mathbb{R}_+^n$
  
  $\omega^*_j = \text{quality assignment to item } j$

- generative model

  $\mathbb{P}[i \text{ beats item } j] = \Phi(\omega^*_i - \omega^*_j)$

  where $\Phi : \mathbb{R} \rightarrow [0, 1]$ is a CDF.

Classical examples:

- Thurstone model (1927): $\Phi$ is standard Gaussian CDF

  $\Phi(t) = \int_{-\infty}^{t} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \, du.$

- Bradley-Terry-Luce model (1952, 1959): $\Phi$ is logistic CDF

  $\Phi(t) = \frac{e^t}{1 + e^t}.$
Classical parametric models

- underlying weight vector $\omega^* \in \mathbb{R}^n_{+}$
  
  $\omega_j^* =$ quality assignment to item $j$

- generative model

  $$\mathbb{P}[i \text{ beats item } j] = \Phi(\omega_i^* - \omega_j^*)$$

  where $\Phi : \mathbb{R} \to [0, 1]$ is a CDF.

Some recent work:

- Hajek, Oh & Xu (2014): study of minimax rates
- N. Shah et al. (2015): minimax analysis of topology dependence
How to relax parametric assumptions?

For each pair \((i, j) \in [n] \times [n]\):

- \(M_{ij}^* = \mathbb{P}[i \succ j]\): probability that item \(i\) preferred to \(j\) in a comparison
- skew symmetry condition: \(M_{ji}^* = 1 - M_{ij}^*\)
How to relax parametric assumptions?

For each pair \((i, j) \in [n] \times [n]\):

- \(M_{ij}^* = \mathbb{P}[i \succ j]\): probability that item \(i\) preferred to \(j\) in a comparison
- skew symmetry condition: \(M_{ji}^* = 1 - M_{ij}^*\)

**Strong stochastic transitivity (SST):** (Davidson & Marschak, 1959, Fishburn, 1973)

- **unknown** permutation \(\pi^*\) defines ordering of items \([n]\)
- if \(\pi^*(i) > \pi^*(j)\), then

\[
\mathbb{P}[i \succ k] \geq \mathbb{P}[j \succ k] \quad \text{for all other } k \in [n] \setminus \{i, j\}.
\]
How to relax parametric assumptions?

For each pair \((i, j) \in [n] \times [n]\):
- \(M^*_{ij} = \mathbb{P}[i \succ j]\): probability that item \(i\) preferred to \(j\) in a comparison
- skew symmetry condition: \(M^*_{ji} = 1 - M^*_{ij}\)

**Ballinger et al., The Economic Journal, 1997**

All of these parametric CDFs are soundly rejected by our data. However, SST usually survives scrutiny.

**Strong stochastic transitivity (SST):** (Davidson & Marschak, 1959, Fishburn, 1973)
- unknown permutation \(\pi^*\) defines ordering of items \([n]\)
- if \(\pi^*(i) > \pi^*(j)\), then

\[
\begin{align*}
\mathbb{P}[i \succ k] & \geq \mathbb{P}[j \succ k] \\
M^*_{ik} & \geq M^*_{jk}
\end{align*}
\]

for all other \(k \in [n]\setminus\{i, j\}\).
Illustration of SST condition

\[ [n] = \{1, 2, 3, 4\} \quad 3 > 1 > 4 > 2 \]

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th>0.7</th>
<th>0.2</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.6</td>
<td>0.2</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

\[ M^* \]
Illustration of SST condition

\[ [n] = \{1, 2, 3, 4\} \]

\[ 3 > 1 > 4 > 2 \]

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th>0.7</th>
<th>0.2</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.6</td>
<td>0.2</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

\( M^* \)

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th>0.8</th>
<th>0.8</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

\( \pi^*(M^*) \)

Bi-isotonic after permutation:

Permuted matrix \( \pi^*(M^*) \) is bi-isotonic:
- rows: entries \{left \rightarrow right\} are non-decreasing
- columns: entries \{up \rightarrow down\} are non-increasing
Poor fit of all (semi)-parametric models

\[ \text{Number of items } n \]

\[ \text{Error } \frac{||\hat{\Lambda} - \Lambda^*||_F^2}{n^2} \]

Thurstone MLE
CRL
Poor fit of all (semi)-parametric models

<table>
<thead>
<tr>
<th>Number of items $n$</th>
<th>Error $|\hat{M} - M^*|^2/F^2/n^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^6$</td>
<td>$2^{-6}$</td>
</tr>
<tr>
<td>$2^7$</td>
<td>$2^{-7}$</td>
</tr>
<tr>
<td>$2^8$</td>
<td>$2^{-8}$</td>
</tr>
<tr>
<td>$2^9$</td>
<td>$2^{-9}$</td>
</tr>
</tbody>
</table>

Lemma

There exist matrices $M^* \in C_{SST}$ such that

$$\inf_{\text{all valid } \Phi} \inf_{M \in \mathcal{C}_{par}(F)} \|M - M^*\|^2_F \geq c_0 n^2.$$
Question:

With what accuracy can we estimate an SST matrix $M^*$ based on noisy comparisons (each entry observed with prob. $p \in (0, 1)$)?
Global minimax risk for SST class

**Question:**
With what accuracy can we estimate an SST matrix $M^*$ based on noisy comparisons (each entry observed with prob. $p \in (0, 1)$)?

**Theorem**

There are universal constants $0 < c_\ell < c_u$ such that

$$
\frac{c_\ell}{pn} \leq \inf_{\tilde{M}} \sup_{M^* \in \mathbb{C}_{SST}} \frac{1}{n^2} \mathbb{E}[\|\tilde{M} - M^*\|_F^2] \leq c_u \frac{\log^2(n)}{pn},
$$

where infimum ranges over all measurable functions $\tilde{M}$ of the data matrix.
Global minimax risk for SST class

**Question:**
With what accuracy can we estimate an SST matrix $M^*$ based on noisy comparisons (each entry observed with prob. $p \in (0, 1)$)?

**Theorem**
There are universal constants $0 < c_\ell < c_u$ such that

$$\frac{c_\ell}{p n} \leq \inf_{\tilde{M}} \sup_{M^* \in \mathbb{C}_{SST}} \frac{1}{n^2} \mathbb{E}[\|\tilde{M} - M^*\|_F^2] \leq c_u \frac{\log^2(n)}{p n},$$

where infimum ranges over all measurable functions $\tilde{M}$ of the data matrix.

- upper bound: achieved by least-squares over SST class

$$\hat{M}_{LS} = \arg \min_{M \in \mathbb{C}_{SST}} \left\{ \sum_{(i,j) \in \Omega} (Y_{ij} - M_{ij})^2 \right\},$$

where $\Omega \subseteq [n] \times [n]$ are observed entries.
Global minimax risk for SST class

**Theorem**

There are universal constants $0 < c_\ell < c_u$ such that

$$
\frac{c_\ell}{p n} \leq \inf_{\tilde{M}} \sup_{M^* \in \mathcal{C}_{SST}} \frac{1}{n^2} \mathbb{E}[\|\tilde{M} - M^*\|_F^2] \leq c_u \frac{\log^2(n)}{p n},
$$

where infimum ranges over all measurable functions $\tilde{M}$ of the data matrix.

- **upper bound:** achieved by least-squares over SST class

$$
\hat{M}_{LS} = \arg \min_{M \in \mathcal{C}_{SST}} \left\{ \sum_{(i,j) \in \Omega} (Y_{ij} - M_{ij})^2 \right\},
$$

where $\Omega \subseteq [n] \times [n]$ are observed entries.

- **relevant complexity parameters:**
  - expected sample size $\approx p n^2$
  - log cardinality of permutations: $\log(n!) \lesssim n \log n$
  - metric entropy of bi-isotonic function classes (Gao & Wellner, 2007)
Beyond (global) minimax....

Geared for Canadian winter
Beyond (global) minimax....

Geared for Canadian winter

Unhappy in Hawaii

Observation:
Worst-case optimal solutions can have undesirable properties.
Indifference sets

A partition \([n] = \bigcup_{i=1}^{S} P_i\) such that

\[ M_{ac}^* = M_{bd}^* \quad \text{whenever } (a, b) \text{ in same partition, and } (c, d) \text{ in same partition.} \]
Illustration of indifference set

\[
M^* = \begin{bmatrix}
0.5 & 0.5 & 0.7 & 0.7 & 0.9 \\
0.5 & 0.5 & 0.7 & 0.7 & 0.9 \\
0.3 & 0.3 & 0.5 & 0.5 & 0.8 \\
0.3 & 0.3 & 0.5 & 0.5 & 0.8 \\
0.1 & 0.1 & 0.2 & 0.2 & 0.5 \\
\end{bmatrix}
\]

Partition: \( P_1 = \{1, 2\} \quad P_2 = \{3, 4\} \quad P_3 = \{5\} \)

Indifference vector: \( k = (2, 2, 1) \)
Adaptivity index

Given indifference vector $\mathbf{k} = (k_1, \ldots, k_s)$, define oracle risk:

$$R_n(\mathbf{k}) := \inf_{\tilde{M}(\mathbf{k})} \sup_{M^* \in \mathcal{C}_{\text{SST}}(\mathbf{k})} \mathbb{E}[\|\tilde{M}(\mathbf{k}) - M^*\|_F^2]$$
Adaptivity index

- given indifference vector $\mathbf{k} = (k_1, \ldots, k_s)$, define oracle risk:

$$R_n(\mathbf{k}) := \inf_{\widehat{M}(\mathbf{k})} \sup_{M^* \in \mathcal{C}_{\text{SST}}(\mathbf{k})} \mathbb{E}[\| \widehat{M}(\mathbf{k}) - M^* \|_F^2]$$

- rescaled risk $R_n(\mathbf{k})/n^2$ varies from $\frac{1}{n^2}$ to $\frac{1}{n}$
Adaptivity index

- given indifference vector \( \mathbf{k} = (k_1, \ldots, k_s) \), define oracle risk:

\[
R_n(\mathbf{k}) := \inf_{\mathbf{M}(\mathbf{k})} \sup_{M^* \in \mathbb{C}_{\text{SST}}(\mathbf{k})} \mathbb{E}[\|\mathbf{M}(\mathbf{k}) - M^*\|_F^2]
\]

- rescaled risk \( R_n(\mathbf{k}) / n^2 \) varies from \( \frac{1}{n^2} \) to \( \frac{1}{n} \)

Measure performance of estimator \( \widehat{\mathbf{M}} \) relative to oracle:

\[
\alpha_n(\widehat{\mathbf{M}}) := \max_{\mathbf{k}} \frac{\sup_{M^* \in \mathbb{C}_{\text{SST}}(\mathbf{k})} \mathbb{E}[\|\widehat{\mathbf{M}} - M^*\|_F^2]}{R_n(s, \mathbf{k})}
\]
Adaptivity of unregularized least squares?

\[ \hat{M}_{LS} \in \arg \min_{M \in \mathcal{C}_{SST}} \| Y - M \|_F^2 \]

- Minimax optimal for estimating SST matrices
- **Without** unknown permutation: known to adapt in shape-constrained estimation (e.g., Guntuboyina et al., 2014)
Adaptivity of unregularized least squares?

\[ \hat{M}_{LS} \in \arg \min_{M \in \mathcal{C}_{SST}} \|Y - M\|_F^2 \]

- Minimax optimal for estimating SST matrices
- **Without** unknown permutation: known to adapt in shape-constrained estimation (e.g., Guntuboyina et al., 2014)

**Theorem**

Least-squares over SST has adaptivity index lower bounded as \( \alpha_n(\hat{M}_{LS}) \geq \frac{n}{4} \).
Adaptivity of unregularized least squares?

\[ \hat{M}_{\text{LS}} \in \arg \min_{M \in \mathcal{C}_{\text{SST}}} \|Y - M\|_F^2 \]

- Minimax optimal for estimating SST matrices
- Without unknown permutation: known to adapt in shape-constrained estimation (e.g., Guntuboyina et al., 2014)

**Theorem**

Least-squares over SST has adaptivity index lower bounded as \( \alpha_n(\hat{M}_{\text{LS}}) \geq \frac{n}{4} \).

- Compare with an estimator \( \hat{M}_{\frac{1}{2}} \) that always outputs an all-half matrix
  \[ \alpha_n(\hat{M}_{\frac{1}{2}}) \leq cn(\log n)^{-2} \]
- Least squares’ adaptivity nearly as bad as estimator that ignores the data!
Adaptivity of unregularized least squares?

$$\hat{M}_{LS} \in \arg \min_{M \in \mathcal{C}_{SST}} \|Y - M\|_F^2$$

- Minimax optimal for estimating SST matrices
- **Without** unknown permutation: known to adapt in shape-constrained estimation (e.g., Guntuboyina et al., 2014)

**Theorem**

Least-squares over SST has adaptivity index lower bounded as $$\alpha_n(\hat{M}_{LS}) \geq \frac{n}{4}$$.

- Compare with an estimator $$\hat{M}_{1/2}$$ that always outputs an all-half matrix
  $$\alpha_n(\hat{M}_{1/2}) \leq cn(\log n)^{-2}$$

- Least squares’ adaptivity nearly as bad as estimator that ignores the data!
- **Why?** Overfitting effect of $$n!$$ permutations.
Fundamental limits of adaptivity?

Consider least-squares over SST combined with a penalty:

$$\hat{M}_{\text{REG}} \in \arg \min_{M \in \mathbb{C}_{\text{SST}}} \left\{ \|Y - M\|_F^2 \underbrace{-k_{\text{max}}(M)(\log n)^3}_{\text{Encourages large indifference}} \right\}$$
Fundamental limits of adaptivity?

Consider least-squares over SST combined with a penalty:

\[
\hat{M}_{REG} \in \arg \min_{M \in \mathcal{C}_{SST}} \left\{ \|Y - M\|_F^2 - k_{\max}(M)(\log n)^3 \right\} \\
\text{Encourages large indifference}
\]

**Theorem**

The penalized LS estimator satisfies the bound

\[
\frac{1}{n^2} \|M^* - \hat{M}_{REG}\|_F^2 \leq c_u \frac{(n - k_{\max}(M^*) + 1)}{n^2} (\log n)^3
\]

with high probability.
Consider least-squares over SST combined with a penalty:

$$\hat{M}_{\text{REG}} \in \arg \min_{M \in \mathcal{C}_{\text{SST}}} \left\{ \| Y - M \|_F^2 - k_{\text{max}}(M)(\log n)^3 \right\}$$

Encourages large indifference

**Theorem**

The penalized LS estimator satisfies the bound

$$\frac{1}{n^2} \| M^* - \hat{M}_{\text{REG}} \|_F^2 \leq c_u \frac{(n - k_{\text{max}}(M^*) + 1)}{n^2} (\log n)^3$$

with high probability. Consequently, it has adaptivity index at most

$$\alpha_n(\hat{M}_{\text{REG}}) \leq c_u (\log n)^3.$$
Step 1: Count and re-order

$$Y = \begin{bmatrix}
1 & 0 & 1 & 1 & \ldots & 1 & 45 \\
1 & 0 & 0 & 1 & \ldots & 0 & 90 \\
0 & 0 & 0 & 1 & \ldots & 1 & 47 \\
1 & 0 & 1 & 0 & \ldots & 0 & 2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
1 & 1 & 1 & 1 & \ldots & 1 & 43
\end{bmatrix}$$

- Count **number of pairwise comparisons won** by each item
Count-Randomize-Least-Squares (CRL) Estimator

Step 1: Count and re-order

\[ Y' = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 & 2 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 43 \\ 1 & 0 & 1 & 1 & \cdots & 1 & 45 \\ 0 & 0 & 0 & 1 & \cdots & 1 & 47 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & \cdots & 0 & 90 \end{bmatrix} \]

- Count **number of pairwise comparisons won** by each item
- Order items based on these counts
Count-Randomize-Least-Squares (CRL) Estimator

Step 2: Randomize

\[
Y' = \begin{bmatrix}
1 & 0 & 1 & 0 & \cdots & 0 & 2 \\
1 & 1 & 1 & 1 & \cdots & 1 & 43 \\
1 & 0 & 1 & 1 & \cdots & 1 & 45 \\
0 & 0 & 0 & 1 & \cdots & 1 & 47 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & 0 & 0 & 1 & \cdots & 0 & 90 \\
\end{bmatrix}
\]

- Identify largest (contiguous) set of items whose counts are separated by at most \(10\sqrt{n}\log n\)
Count-Randomize-Least-Squares (CRL) Estimator

Step 2: Randomize

\[ Y'' = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 1 & \cdots & 0 \end{bmatrix} \]

- Identify largest (contiguous) set of items whose counts are separated by at most \(10\sqrt{n\log n}\)
- Permute all items in this set uniformly at random
Step 3: Least squares

\[ Y'' = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 1 & \cdots & 0 \end{bmatrix} \]

Least-squares fit over subset of SST matrices with this ordering:

\[ \widehat{M}_{\text{CRL}} \in \arg \min_{M \in \mathcal{C}_{\text{SST}}} \| Y'' - M \|_F^2. \]

faithful to this order
Adaptivity of the CRL estimator

\[
\text{Error } \left\| \hat{M} - M^* \right\|_F^2 / n^2
\]

Size of largest clique \( \|k\|_\infty \)

\( n = 64 \)
Guarantees on CRL adaptivity

Theorem

The CRL estimator satisfies the bound

\[
\frac{1}{n^2} \| \hat{M}_{CRL} - M^* \|^2_F \leq c_u \frac{(n - k_{\text{max}}(M^*) + 1)}{n^{3/2}} (\log n)^8
\]

with high probability.
Guarantees on CRL adaptivity

**Theorem**

The CRL estimator satisfies the bound

\[
\frac{1}{n^2} \| \hat{M}_{CRL} - M^* \|_F^2 \leq c_u \left( \frac{n - k_{\text{max}}(M^*) + 1}{n^{3/2}} \right) (\log n)^8
\]

with high probability. Consequently, it has adaptivity index at most

\[
\alpha_n(\hat{M}_{CRL}) \leq c_u \sqrt{n} (\log n)^8.
\]
Guarantees on CRL adaptivity

Theorem

The CRL estimator satisfies the bound

$$\frac{1}{n^2} \| \hat{M}_{CRL} - M^* \|_F^2 \leq c_u \left( \frac{n - k_{\text{max}}(M^*) + 1}{n^{3/2}} \right) (\log n)^8$$

with high probability. Consequently, it has adaptivity index at most

$$\alpha_n(\hat{M}_{CRL}) \leq c_u \sqrt{n} (\log n)^8.$$

differs by $\sqrt{n}$ from best achievable (using a complicated algorithm)
Guarantees on CRL adaptivity

Theorem

The CRL estimator satisfies the bound

\[
\frac{1}{n^2} \| \hat{M}_{CRL} - M^* \|_F^2 \leq c_u \frac{(n - k_{\text{max}}(M^*) + 1)}{n^{3/2}} (\log n)^8
\]

with high probability. Consequently, it has adaptivity index at most

\[
\alpha_n(\hat{M}_{CRL}) \leq c_u \sqrt{n}(\log n)^8.
\]

- differs by \( \sqrt{n} \) from best achievable (using a complicated algorithm)

Question:

Is this an optimal polynomial-time algorithm?
A computational lower bound

Planted clique conjecture: Hard to detect a planted clique of size $o(\sqrt{n})$ in an Erdős-Rényi random graph.
A computational lower bound

Planted clique conjecture: Hard to detect a planted clique of size $o(\sqrt{n})$ in an Erdős-Rényi random graph.
A computational lower bound

Planted clique conjecture: Hard to detect a planted clique of size $o(\sqrt{n})$ in an Erdős-Rényi random graph.

Conjecture used in past work in machine learning:
- Rigollet & Berthet, 2013: sparse PCA
- Sub-matrix detection: Ma & Wu, 2014
A computational lower bound

**Planted clique conjecture:** Hard to detect a planted clique of size $o(\sqrt{n})$ in an Erdős-Rényi random graph.

**Theorem**

Under planted clique hardness, adaptivity index of any polynomial-time computable estimator $\hat{M}$ is lower bounded as

$$\alpha_n(\hat{M}) \geq c_\ell \sqrt{n} (\log n)^{-2},$$
A computational lower bound

Planted clique conjecture: Hard to detect a planted clique of size \(o(\sqrt{n})\) in an Erdős-Rényi random graph.

Theorem

Under planted clique hardness, adaptivity index of any polynomial-time computable estimator \(\hat{M}\) is lower bounded as

\[
\alpha_n(\hat{M}) \geq c_\ell \sqrt{n}(\log n)^{-2},
\]

Consequence: CRL achieves best possible adaptivity index (up to log factors) over all computationally efficient estimators
A computational lower bound

**Planted clique conjecture:** Hard to detect a planted clique of size $o(\sqrt{n})$ in an Erdős-Rényi random graph.

**Theorem**

Under planted clique hardness, adaptivity index of any polynomial-time computable estimator $\widehat{M}$ is lower bounded as

$$\alpha_n(\widehat{M}) \geq c\sqrt{n}(\log n)^{-2},$$

**Consequence:** CRL achieves best possible adaptivity index (up to log factors) over all computationally efficient estimators

- proof via reduction to planted clique
  - form matrix $M^*$ with three-way partition $(n - 2k, k, k)$ with $k = \frac{\sqrt{n}}{\log \log n}$
  - construct a certain permuted ensemble, and embed it into planted clique $(n/2, k)$
Summary

• SST class of models for pairwise comparison
  ▶ more flexible than popular parametric models
  ▶ rates no worse than parametric estimators (even if data comes from parametric models)

• least-squares over SST class
  ▶ minimax optimal but computationally challenging
  ▶ without regularization: no adaptivity!

• Count-Randomize-Least-Squares (CRL)
  ▶ $\sqrt{n}$-suboptimal in minimax rate and adaptivity index
  ▶ but optimal among polynomial-time algorithms!
Summary

- SST class of models for pairwise comparison
  - more flexible than popular parametric models
  - rates no worse than parametric estimators (even if data comes from parametric models)
- least-squares over SST class
  - minimax optimal but computationally challenging
  - without regularization: no adaptivity!
- Count-Randomize-Least-Squares (CRL)
  - $\sqrt{n}$-suboptimal in minimax rate and adaptivity index
  - but optimal among polynomial-time algorithms!


Stochastically transitive models for pairwise comparisons, Shah, Balakrishnan, Guntuboyina, & W. Arxiv, 2015