# Uncertainty Quantification in The Bradley-Terry-Luce Model

Anderson Ye Zhang

Department of Statistics
University of Pennsylvania


Yandi Shen
UChicago

Chao Gao
UChicago

# Ranking Examples

Sports and Gaming:


Image source: `www.psdcovers.com/wp-content/uploads/2012/07/NFL-vector-logos-1024x772.jpg`

# Ranking Examples

## Recommendation System and Web Search:

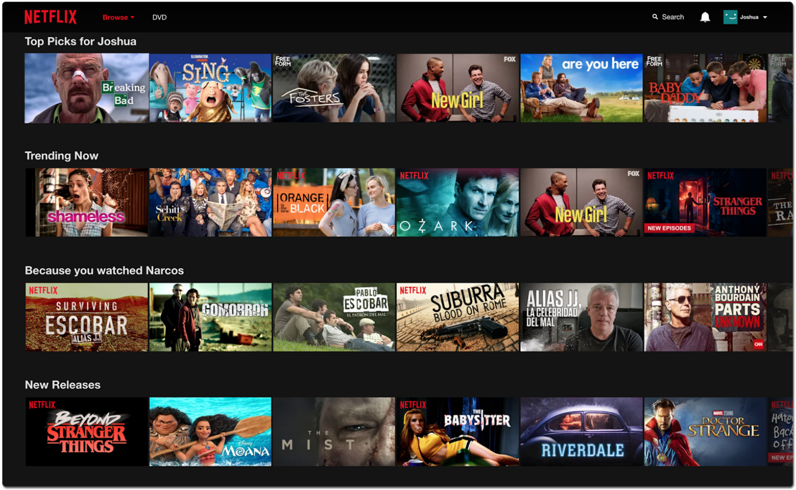



Image source: `https://miro.medium.com/max/2400/1*dMR3xmufnmKiw4crlisQUA.png`

# Ranking Examples

Ranked Choice Voting:

**Instructions to Voters**

To vote, fill in the oval like this ●

To rank your candidate choices, fill in the oval:

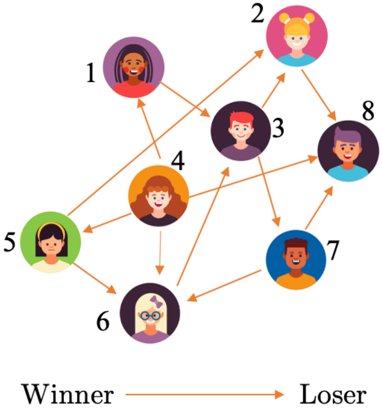
- In the 1st column for your 1st choice candidate.
- In the 2nd column for your 2nd choice candidate, and so on.

Continue until you have ranked as many or as few candidates as you like.

| Governor | 1st Choice | 2nd Choice | 3rd Choice | 4th Choice | 5th Choice | 6th Choice | 7th Choice | 8th Choice |
|---|---|---|---|---|---|---|---|---|
| **Cote, Adam Roland**<br>Sanford | O | O | O | O | O | O | O | O |
| **Dion, Donna J.**<br>Biddeford | O | O | O | O | O | O | O | O |
| **Dion, Mark N.**<br>Portland | O | O | O | O | O | O | O | O |
| **Eves, Mark W.**<br>North Berwick | O | O | O | O | O | O | O | O |
| **Mills, Janet T.**<br>Farmington | O | O | O | O | O | O | O | O |
| **Russell, Diane Marie**<br>Portland | O | O | O | O | O | O | O | O |
| **Sweet, Elizabeth A.**<br>Hallowell | O | O | O | O | O | O | O | O |
| **Write-in** | O | O | O | O | O | O | O | O |

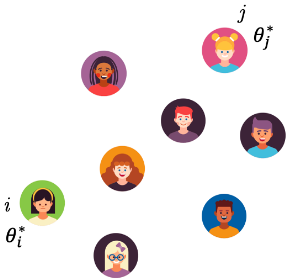
www.mainepublic.org/sites/mpbn/files/styles/x_large/public/201805/DEMRCV2Contests20180420.jpg

# Ranking from Pairwise Comparisons



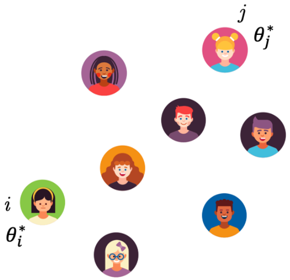

- ✓ Player 4 is the strongest
- ✓ Player 8 is the weakest
- ? Other Players?

# Bradley-Terry-Luce (BTL) Model




- $n$ players
- A skill parameter vector $\theta^* \in \mathbb{R}^n$.
  For player $i$, her skill is $\theta_i^*$

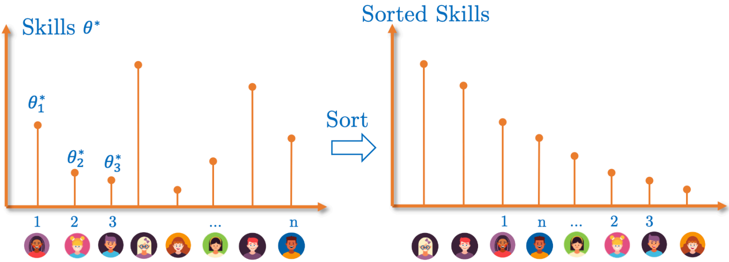
# Bradley-Terry-Luce (BTL) Model




- $n$ players
- A skill parameter vector $\theta^* \in \mathbb{R}^n$. For player $i$, her skill is $\theta_i^*$
- Rank vector $r^*$: a permutation of $1, \ldots, n$ such that $\theta_i^* = \theta^*_{(r_i^*)}$. For player $i$, her rank is $r_i^*$

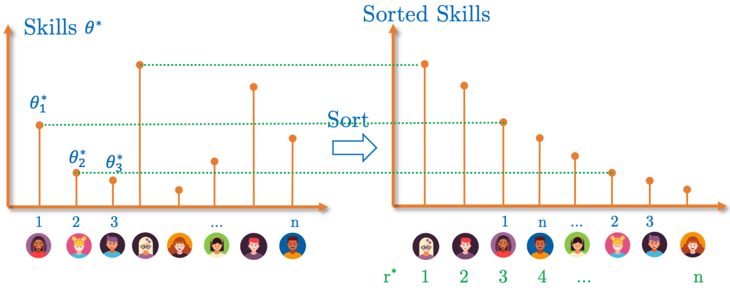
# Bradley-Terry-Luce (BTL) Model


- $n$ players
- A skill parameter vector $\theta^* \in \mathbb{R}^n$. For player $i$, her skill is $\theta_i^*$
- Rank vector $r^*$: a permutation of $1, \ldots, n$ such that $\theta_i^* = \theta^*_{(r_i^*)}$. For player $i$, her rank is $r_i^*$

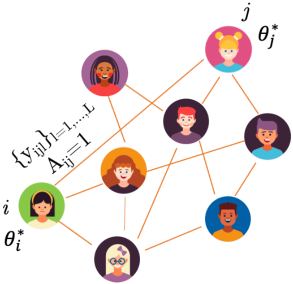


# Bradley-Terry-Luce (BTL) Model


- $n$ players
- A skill parameter vector $\theta^* \in \mathbb{R}^n$. For player $i$, her skill is $\theta_i^*$
- Rank vector $r^*$: a permutation of $1, \dots, n$ such that $\theta_i^* = \theta_{(r_i^*)}^*$. For player $i$, her rank is $r_i^*$



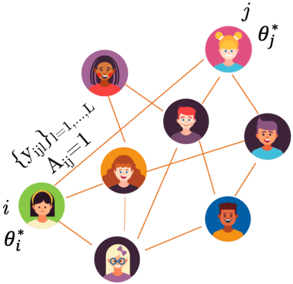
# Bradley-Terry-Luce (BTL) Model




$$\mathbb{P}(i \text{ beats } j) \propto \exp\left(\theta_i^*\right)$$

$$\mathbb{P}(j \text{ beats } i) \propto \exp(\theta_j^*)$$

# Bradley-Terry-Luce (BTL) Model



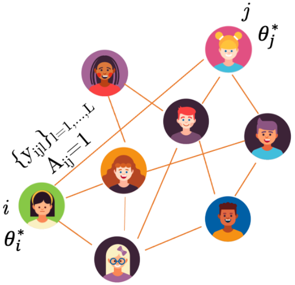

$$\mathbb{P}(i \text{ beats } j) \propto \exp\left(\theta_i^*\right)$$

$$\mathbb{P}(j \text{ beats } i) \propto \exp(\theta_j^*)$$

$$\mathbb{P}(i \text{ beats } j) = \frac{\exp\left(\theta_i^*\right)}{\exp\left(\theta_i^*\right) + \exp(\theta_j^*)} = \psi(\theta_i^* - \theta_j^*)$$

where $\psi(x) = \frac{e^x}{e^x+1}$

# Bradley-Terry-Luce (BTL) Model




$$\mathbb{P}(i \text{ beats } j) \propto \exp(\theta_i^*)$$
$$\mathbb{P}(j \text{ beats } i) \propto \exp(\theta_j^*)$$

$$\mathbb{P}(i \text{ beats } j) = \frac{\exp(\theta_i^*)}{\exp(\theta_i^*) + \exp(\theta_j^*)} = \psi(\theta_i^* - \theta_j^*)$$

where $\psi(x) = \frac{e^x}{e^x+1}$

- Missing Data: comparison graph $A_{ij} \overset{iid}{\sim} \text{Ber}(p)$
- $L$ outcomes for each observed pair $(i,j)$:

$$y_{ijl} | A_{ij} = 1 \overset{ind}{\sim} \text{Ber}\left(\psi(\theta_i^* - \theta_j^*)\right)$$

# An Incomplete List of Prior Art

- Dwork, Kumar, Naor, and Sivakumar (2001)
- Hunter (2004)
- Jiang, Lim, Yao, and Ye (2011)
- Negahban and Wainwright (2012)
- Rajkumar and Agarwal (2014)
- Chen and Suh (2015)
- Jin, Zhang, Balakrishnan, Wainwright, and Jordan (2016)
- Shah, Balakrishnan, Guntuboyina, and Wainwright (2016)
- Shah and Wainwright (2017)
- Agarwal, Agarwal, Assadi, and Khanna (2017)
- Pananjady, Mao, Muthukumar, Wainwright, and Courtade (2017)
- Mao, Pananjady, and Wainwright (2018a)
- Mao, Weed, and Rigollet (2018b)
- Balakrishnan, Wainwright, and Yu (2017)
- Negahban, Oh, and Shah (2017)
- Chen, Fan, Ma, and Wang (2019)
- Shah, Balakrishnan, and Wainwright (2019)
- Heckel, Shah, Ramchandran, and Wainwright (2019)
- …

# Existing Literature

Focuses on the estimation of $\theta^*$:

$$\|\hat{\theta} - \theta^*\|, \quad \|\hat{\theta} - \theta^*\|_\infty$$

It remains unclear

- Uncertainty quantification for $\theta^*$
  - Entrywise distribution of $\hat{\theta}$?
  - Confidence interval and hypothesis testing for $\theta_i^*$?
  - Confidence interval and hypothesis testing for $r_i^*$?

- Recovery of $r^*$?

# Existing Literature

Focuses on the estimation of $\theta^*$:

$$\|\hat{\theta} - \theta^*\|, \quad \|\hat{\theta} - \theta^*\|_\infty$$

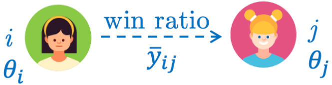
It remains unclear

- Uncertainty quantification for $\theta^*$ Focus of this talk
  - Entrywise distribution of $\hat{\theta}$?
  - Confidence interval and hypothesis testing for $\theta_i^*$?
  - Confidence interval and hypothesis testing for $r_i^*$?
- Recovery of $r^*$?

# Uncertainty Quantification for MLE

# Maximum Likelihood Estimator

Step 1: Compute $\bar{y}_{ij} = \frac{1}{L}\sum_{l=1}^{L} y_{ijl}$




Step 2: Obtain the negative log-likelihood function

$$\ell(\theta) = \sum_{i,j:i<j} A_{ij}\left(\bar{y}_{ij}\log\frac{1}{\psi(\theta_i-\theta_j)} + (1-\bar{y}_{ij})\log\frac{1}{1-\psi(\theta_i-\theta_j)}\right)$$

Step 3: Find the MLE $\hat{\theta}$ by convex optimization

$$\hat{\theta} = \operatorname*{argmin}_{\theta\in\mathbb{R}^n:\mathbb{1}_n^\top\theta=0} \ell(\theta)$$

Identifiability: $\theta$ is identifiable up to a global shift $a \in \mathbb{R}$, i.e., $\ell(\theta) = \ell(\theta + a\mathbb{1}_n)$

# Existing Results

The skill parameter $\theta^*$ is assumed to satisfy

- Dynamic range:

$$\max_{i \in [n]} \theta_i^* - \min_{i \in [n]} \theta_i^* \leq \kappa = \mathcal{O}(1)$$

- Identifiability: $\mathbb{1}_n^\top \theta^* = 0$

**Proposition (CFMW19, CGZ20)**

*Assume $np \gtrsim \log n$, then w.h.p.,*

$$\|\hat{\theta} - \theta^*\|^2 \lesssim \frac{1}{pL} \quad \textit{and} \quad \|\hat{\theta} - \theta^*\|_\infty^2 \lesssim \frac{\log n}{npL}$$

$np \gtrsim \log n$ is necessary as otherwise the comparison graph $A \sim G(n, p)$ is disconnected.

# Existing Results

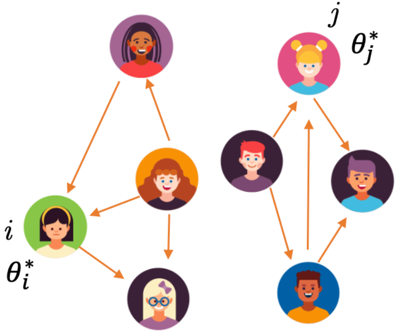
**Proposition (CFMW19, CGZ20)**

*Assume* $np \gtrsim \log n$*, then w.h.p.,*

$$\|\hat{\theta} - \theta^*\|^2 \lesssim \frac{1}{pL} \quad \text{and} \quad \|\hat{\theta} - \theta^*\|_\infty^2 \lesssim \frac{\log n}{npL}$$

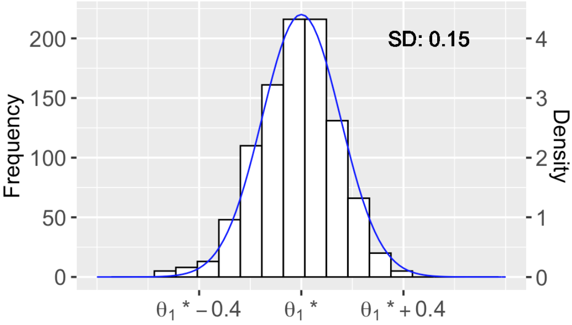
$np \gtrsim \log n$ is necessary as otherwise the comparison graph $A \sim G(n, p)$ is disconnected.

If $np \lesssim \log n$:



# Entrywise Distribution

Player 1  $\hat{\theta}_1 \sim ?$




Histogram of $\hat{\theta}_1$ from 100 independent datasets generated from $\theta^*$

# Existing Results

**Proposition (SY99, HYTC20)**

*Assume* $n^{1/10}p \to \infty$*, then for any fixed* $k \geq 1$*,*

$$(\hat{\theta}_1 - \theta_1^*, \ldots, \hat{\theta}_k - \theta_k^*)^\top \xrightarrow{d} \mathcal{N}_k(0, S)$$

Questions:

- Weaker assumption on $p$?
- Non-asymptotic results?

# Our Result: Non-asymptotic Expansion

**Theorem**

*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any* $i \in [n]$,

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \frac{\sum_{j:j \neq i} A_{ij} \left( \bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*) \right)}{\sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)} + \eta_i.$$

*Here* $\epsilon, \eta \in \mathbb{R}^n$ *such that* $\|\epsilon\|_\infty = \mathfrak{o}(1)$, $\|\eta\|_\infty = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$ *w.h.p..*

# Our Result: Non-asymptotic Expansion

**Theorem**

*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any* $i \in [n]$,

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \frac{\sum_{j:j \neq i} A_{ij} \left( \bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*) \right)}{\sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)} + \eta_i.$$

*Here* $\epsilon, \eta \in \mathbb{R}^n$ *such that* $\|\epsilon\|_\infty = \mathfrak{o}(1), \|\eta\|_\infty = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$ *w.h.p..*

the main error term $\asymp \frac{1}{\sqrt{npL}}$

# Our Result: Non-asymptotic Expansion

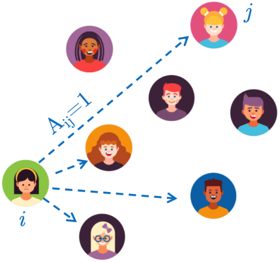
**Theorem**

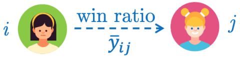
*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any* $i \in [n]$,

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)} + \eta_i.$$

*Here* $\epsilon, \eta \in \mathbb{R}^n$ *such that* $\|\epsilon\|_\infty = \mathfrak{o}(1), \|\eta\|_\infty = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$ *w.h.p..*

the main error term $\asymp \frac{1}{\sqrt{npL}}$




$$\mathbb{E}\bar{y}_{ij} = \psi(\theta_i^* - \theta_j^*)$$

$$\mathrm{Var}(\bar{y}_{ij}) = L^{-1}\psi'(\theta_i^* - \theta_j^*)$$

# Our Result: Non-asymptotic Expansion

**Theorem**

*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any* $i \in [n]$,

$$\hat{\theta}_i - \theta_i^* = (1+\epsilon_i)\frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)} + \eta_i.$$

*Here* $\epsilon, \eta \in \mathbb{R}^n$ *such that* $\|\epsilon\|_\infty = \mathfrak{o}(1), \|\eta\|_\infty = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$ *w.h.p..*

Remarks:

- Uniform, Explicit
- Near optimal assumption on $p$
- No assumption on $L$ (we can let $L = 1$)
- Immediately imply bounds on $\|\hat{\theta} - \theta^*\|^2, \|\hat{\theta} - \theta^*\|_\infty^2$, and the asymptotic normality

# Consequence I: $\ell_2$, $\ell_\infty$ bounds

**Our result**

$$\hat{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)} \asymp \frac{1}{\sqrt{npL}}$$

It explains why

$$\|\hat{\theta} - \theta^*\|^2 = \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i^*)^2 \lesssim \frac{n}{npL} = \frac{1}{pL}$$

$$\|\hat{\theta} - \theta^*\|_\infty^2 = \max_{i\in[n]} |\hat{\theta}_i - \theta_i^*|^2 \lesssim \frac{\log n}{npL}$$

# Consequence II: Asymptotic Normality

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \quad \frac{\sum_{j:j\neq i} A_{ij}\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)} \quad + \eta_i.$$

# Consequence II: Asymptotic Normality

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \quad \mathcal{N}\left(0, \frac{1}{L \sum_{j: j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}\right) \quad + \eta_i.$$

# Consequence II: Asymptotic Normality

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \quad \mathcal{N}\left(0, \frac{1}{L \sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}\right) \quad + \eta_i.$$

$$\sqrt{L \sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)} \, (\hat{\theta}_i - \theta_i^*) \xrightarrow{d} \mathcal{N}(0, 1).$$

# Consequence II: Asymptotic Normality

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \quad \mathcal{N}\left(0, \frac{1}{L \sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}\right) \quad + \eta_i.$$

$$\sqrt{L \sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}\, (\hat{\theta}_i - \theta_i^*) \xrightarrow{d} \mathcal{N}(0, 1).$$

**Corollary**

*Assume $np \gg (\log n)^{\frac{3}{2}}$, then for any fixed $k \geq 1$,*

$$\left(\rho_1(\theta^*)(\hat{\theta}_1 - \theta_1^*), \ldots, \rho_k(\theta^*)(\hat{\theta}_k - \theta_k^*)\right)^\top \xrightarrow{d} \mathcal{N}_k(0, I_k),$$

*where $\rho_i(\theta^*) = \sqrt{L \sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}$.*

# Consequence II: Asymptotic Normality

$$\hat{\theta}_i - \theta_i^* = (1 + \epsilon_i) \quad \mathcal{N}\left(0, \frac{1}{L \sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}\right) \quad + \eta_i.$$

$$\sqrt{L \sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}\,(\hat{\theta}_i - \theta_i^*) \xrightarrow{d} \mathcal{N}(0,1).$$

**Corollary**

*Assume $np \gg (\log n)^{\frac{3}{2}}$, then for any fixed $k \geq 1$,*

$$\left(\rho_1(\hat{\theta})(\hat{\theta}_1 - \theta_1^*), \ldots, \rho_k(\hat{\theta})(\hat{\theta}_k - \theta_k^*)\right)^\top \xrightarrow{d} \mathcal{N}_k(0, I_k),$$

*where $\rho_i(\hat{\theta}) = \sqrt{L \sum_{j:j\neq i} A_{ij}\psi'(\hat{\theta}_i - \hat{\theta}_j)}$.*

# Application I: CI/HT for Skills $\theta^*$


vs.

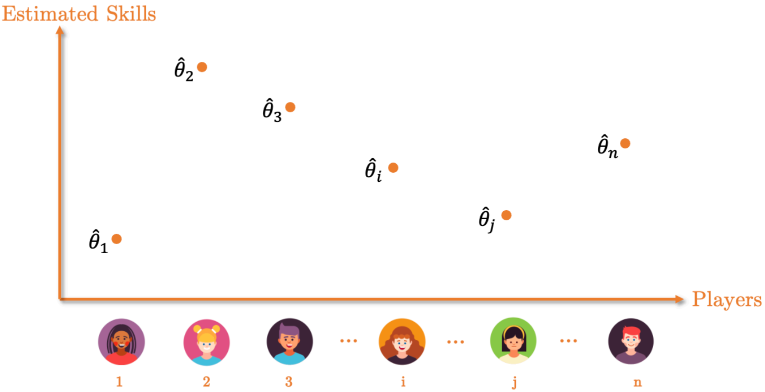

- CI for $\theta_i^* - \theta_j^*$
- HT for $\mathbb{H}_0 : \theta_i^* = \theta_j^*, \quad \mathbb{H}_1 : \theta_i^* \neq \theta_j^*$

$$\begin{pmatrix} \rho_i(\hat{\theta})(\hat{\theta}_i - \theta_i^*) \\ \rho_j(\hat{\theta})(\hat{\theta}_j - \theta_j^*) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(0, I_2)$$

# Application II: CI for Rank $r^*$

Statistical inference for the rank of a player of interest $r_i^*$

- $r_i^*$ is the order of $\theta_i^*$ in $\theta^*$
- Point estimation: $\hat{r}_i$ is the order of $\hat{\theta}_i$ in $\hat{\theta}$



# Application II: CI for Rank $r^*$

What about constructing an $(1 - \alpha)$ CI for $r_i^*$?

# Application II: CI for Rank $r^*$

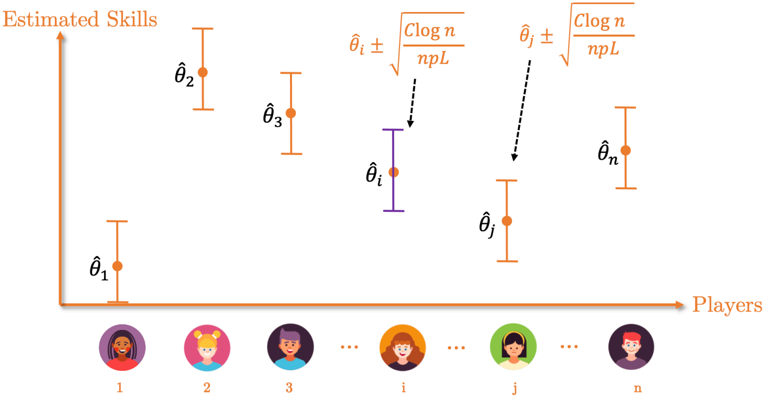
What about constructing an $(1-\alpha)$ CI for $r_i^*$?

- Use the $\ell_\infty$ result from existing literature: $\|\hat{\theta} - \theta^*\|_\infty^2 \le \frac{C \log n}{npL}$

# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

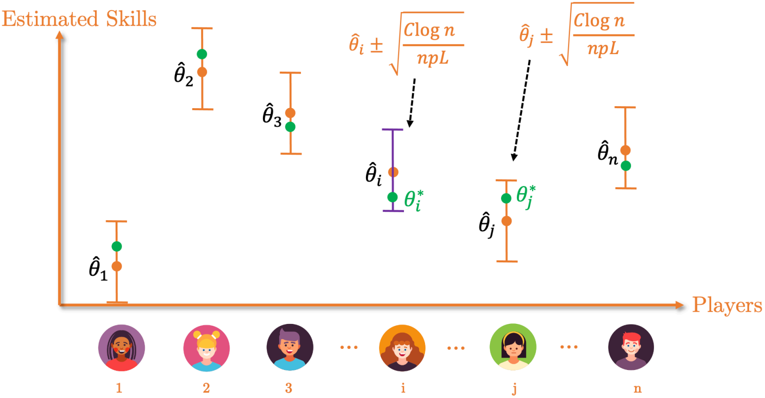
- Use the $\ell_\infty$ result from existing literature: $\|\hat{\theta}-\theta^*\|_\infty^2 \leq \frac{C\log n}{npL}$



# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

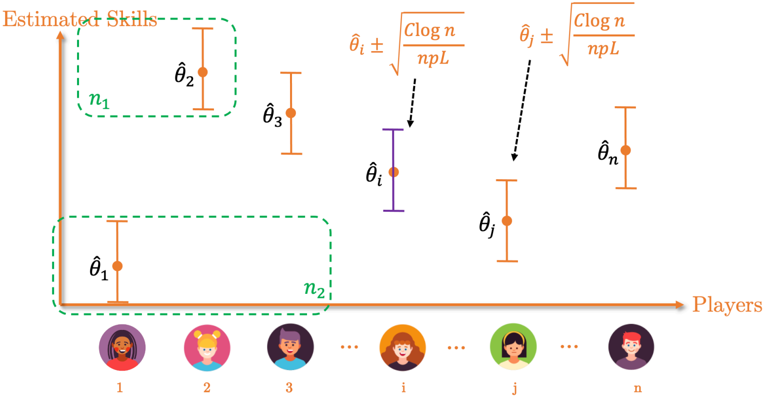
- Use the $\ell_\infty$ result from existing literature: $\|\hat{\theta} - \theta^*\|_\infty^2 \leq \frac{C \log n}{npL}$



# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

- Use the $\ell_\infty$ result from existing literature: $\|\hat{\theta} - \theta^*\|_\infty^2 \le \frac{C \log n}{npL}$

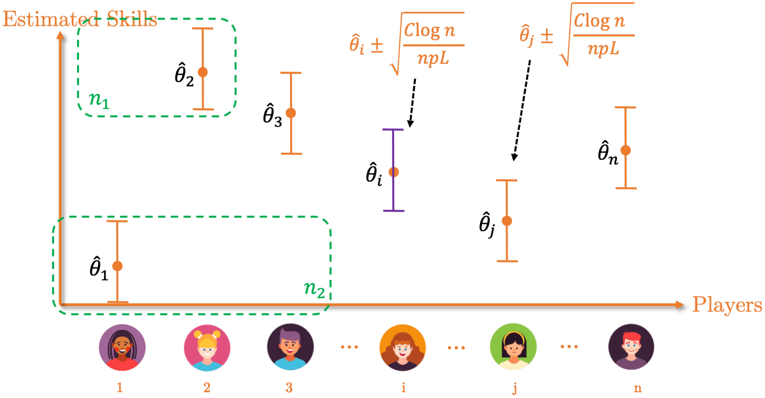



CI: $[n_1 + 1, n - n_2]$

# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

- Use the $\ell_\infty$ result from existing literature: $\|\hat{\theta}-\theta^*\|_\infty^2 \leq \frac{C \log n}{npL}$




CI: $[n_1+1, n-n_2]$ very conservative!!

# Application II: CI for Rank $r^*$

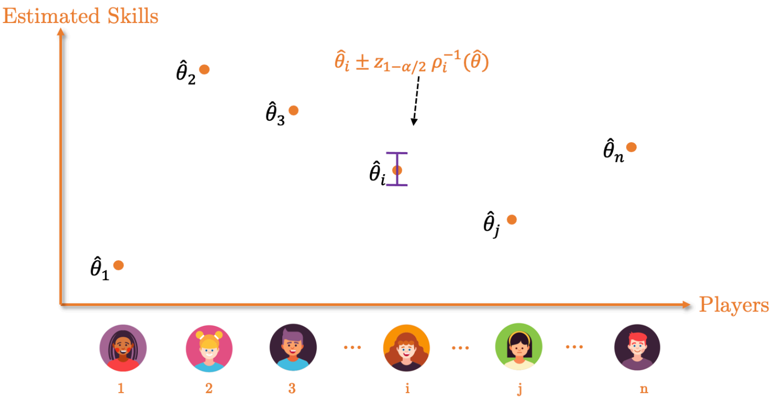
What about constructing an $(1-\alpha)$ CI for $r_i^*$?

- Use our new result

# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

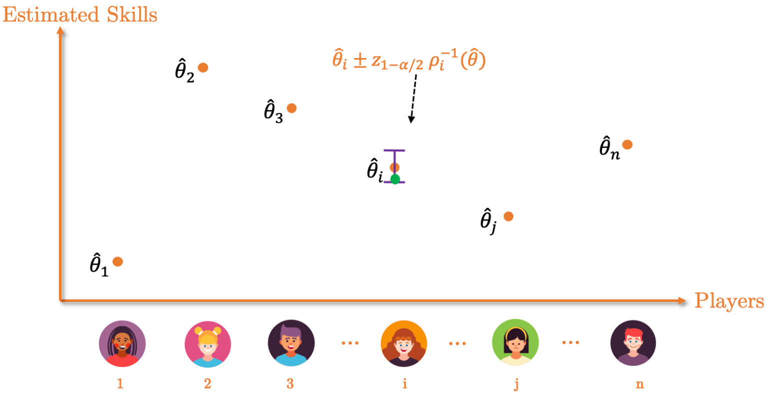
- Use our new result



# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

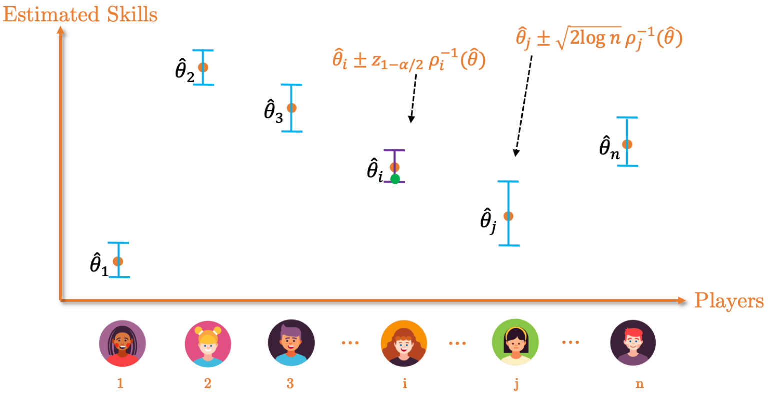
- Use our new result



# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

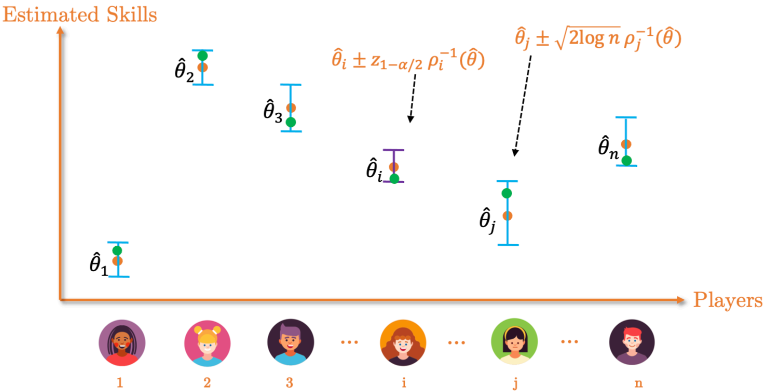
- Use our new result



# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

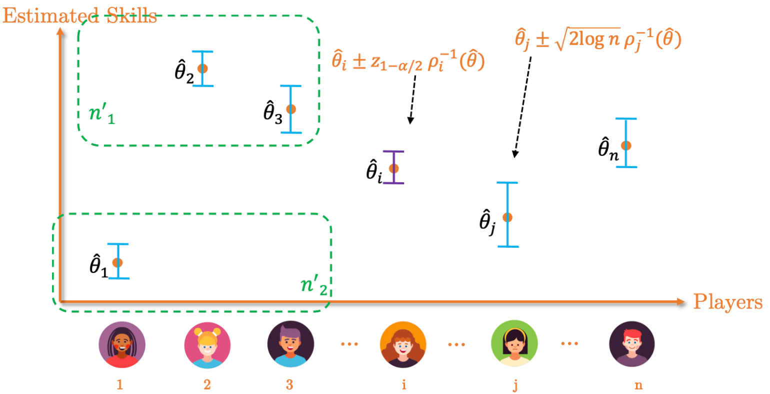
- Use our new result



# Application II: CI for Rank $r^*$

What about constructing an $(1-\alpha)$ CI for $r_i^*$?

- Use our new result




CI: $[n_1' + 1, n - n_2']$ $\mathbb{P}\left(r_i^* \in [n_1'+1, n-n_2']\right) \approx 1-\alpha$

# Intuition

## Our result revisit

$$\hat{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}$$

Intuition for the main error term?

# Intuition

## Our result revisit

$$\hat{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}$$

Intuition for the main error term?

The global likelihood function $\quad \hat{\theta} = \text{argmin}_\theta\, \ell(\theta)$

$$\ell(\theta) = \sum_{i,j:i<j} A_{ij}\left(\bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \theta_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\theta_i - \theta_j)}\right)$$

# Intuition

## Our result revisit

$$\hat{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}$$

Intuition for the main error term?

The global likelihood function $\hat{\theta} = \text{argmin}_\theta\, \ell(\theta)$

$$\ell(\theta) = \sum_{i,j:i<j} A_{ij}\left(\bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \theta_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\theta_i - \theta_j)}\right)$$

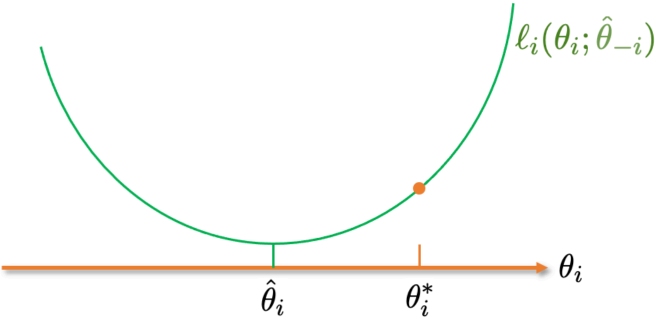
The local likelihood function $\hat{\theta}_i = \text{argmin}_{\theta_i}\, \ell(\theta_i; \hat{\theta}_{-i})$

$$\ell_i(\theta_i; \theta_{-i}) = \sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \theta_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\theta_i - \theta_j)}\right)$$

# Intuition

$\hat{\theta}_i = \text{argmin}_{\theta_i} \ell(\theta_i; \hat{\theta}_{-i})$ where

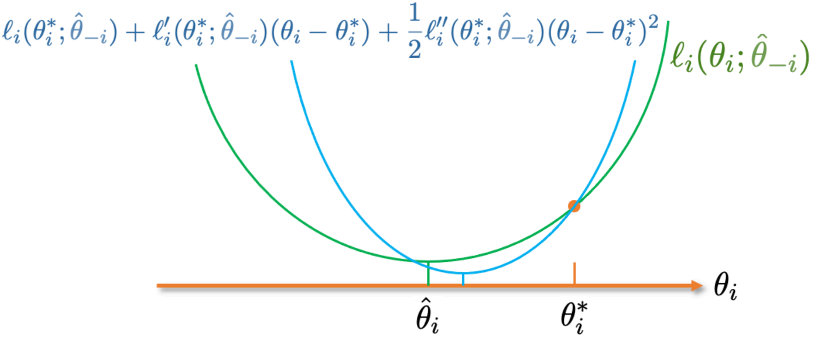
$$\ell_i(\theta_i; \hat{\theta}_{-i}) = \sum_{j:j\neq i} A_{ij} \left( \bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \hat{\theta}_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\theta_i - \hat{\theta}_j)} \right)$$



# Intuition

$\hat{\theta}_i = \text{argmin}_{\theta_i} \, \ell(\theta_i; \hat{\theta}_{-i})$ where

$$\ell_i(\theta_i; \hat{\theta}_{-i}) = \sum_{j:j \neq i} A_{ij} \left( \bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \hat{\theta}_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\theta_i - \hat{\theta}_j)} \right)$$

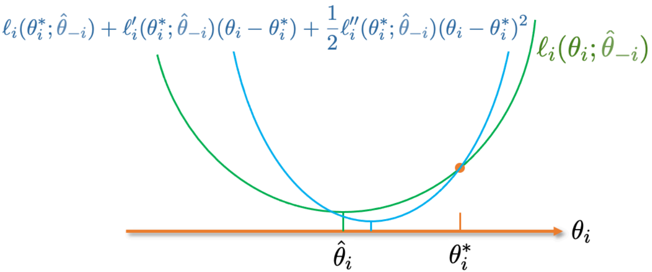


# Intuition

$$\hat{\theta}_i \approx \underset{\theta_i}{\text{argmin}} \left( \ell_i(\theta_i^*; \hat{\theta}_{-i}) + \ell_i'(\theta_i^*; \hat{\theta}_{-i})(\theta_i - \theta_i^*) + \frac{1}{2}\ell_i''(\theta_i^*; \hat{\theta}_{-i})(\theta_i - \theta_i^*)^2 \right)$$

$$= \theta_i^* - \frac{\ell_i'(\theta_i^*; \hat{\theta}_{-i})}{\ell_i''(\theta_i^*; \hat{\theta}_{-i})} = \theta_i^* + \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \hat{\theta}_j)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \hat{\theta}_j)}$$

$$\approx \theta_i^* + \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}$$

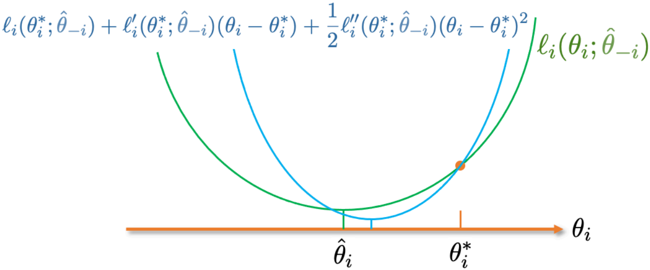


# Intuition

$$\hat{\theta}_i \approx \underset{\theta_i}{\text{argmin}} \left( \ell_i(\theta_i^*; \hat{\theta}_{-i}) + \ell_i'(\theta_i^*; \hat{\theta}_{-i})(\theta_i - \theta_i^*) + \frac{1}{2}\ell_i''(\theta_i^*; \hat{\theta}_{-i})(\theta_i - \theta_i^*)^2 \right)$$

$$= \theta_i^* - \frac{\ell_i'(\theta_i^*; \hat{\theta}_{-i})}{\ell_i''(\theta_i^*; \hat{\theta}_{-i})} = \theta_i^* + \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \hat{\theta}_j)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \hat{\theta}_j)}$$

$$\approx \theta_i^* + \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j\neq i} A_{ij}\psi'(\theta_i^* - \theta_j^*)}$$



# Uncertainty Quantification for Spectral Method

# Spectral Method

(Rank Centrality Algorithm [NW12, NOS17])

Step 1: Construct the Markov transition matrix

$$P_{ij} = \begin{cases} \frac{1}{d} A_{ij} \bar{y}_{ji}, & i \neq j \\ 1 - \frac{1}{d} \sum_{l \neq i} A_{il} \bar{y}_{li}, & i = j \end{cases}$$

Step 2: Find the stationary distribution $\hat{\pi}$

Step 3: Obtain the spectral estimator $\tilde{\theta}$ by

$$\tilde{\theta}_i = \log \hat{\pi}_i - \frac{1}{n} \sum_{j=1}^{n} \log \hat{\pi}_j$$

# Spectral Method

(Rank Centrality Algorithm [NW12, NOS17])

Why spectral method works?

Population version:

$$M_{ij} = \mathbb{E}(P_{ij}|A) = \begin{cases} \frac{1}{d} A_{ij} \psi(\theta_j^* - \theta_i^*), & i \neq j \\ 1 - \frac{1}{d} \sum_{l \neq i} A_{il} \psi(\theta_l^* - \theta_i^*), & i = j \end{cases}$$

$$\pi^* = \left( \frac{\exp\left(\theta_1^*\right)}{\sum_l \exp\left(\theta_l^*\right)}, \ldots, \frac{\exp(\theta_n^*)}{\sum_l \exp\left(\theta_l^*\right)} \right)^\top$$

- Easy to check $\pi^*$ is the stationary distribution of $M$
- $\theta_i^* = \log(\pi_i^*)$ up to a global shift

# Spectral Method

(Rank Centrality Algorithm [NW12, NOS17])

Step 1: Construct the Markov transition matrix

$$P_{ij} = \begin{cases} \frac{1}{d} A_{ij} \bar{y}_{ji}, & i \neq j \\ 1 - \frac{1}{d} \sum_{l \neq i} A_{il} \bar{y}_{li}, & i = j \end{cases}$$

Step 2: Find the stationary distribution $\hat{\pi}$

Step 3: Obtain the spectral estimator $\tilde{\theta}$ by

$$\tilde{\theta}_i = \log \hat{\pi}_i - \frac{1}{n} \sum_{j=1}^{n} \log \hat{\pi}_j$$

Identifiability: $\mathbb{1}_n^\top \tilde{\theta} = 0$

# Existing Results for The Spectral Method

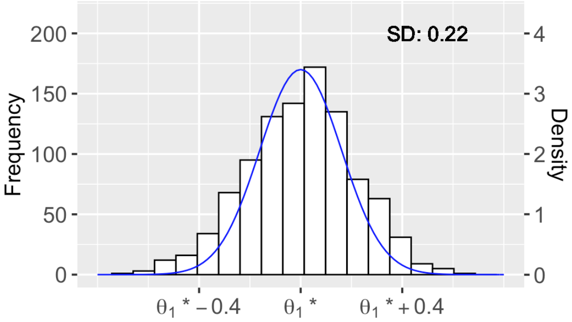
**Proposition (NOS17, CFMW19)**

*Assume* $np \gtrsim \log n$*, then w.h.p.,*

$$\|\tilde{\theta} - \theta^*\|^2 \lesssim \frac{1}{pL} \quad \textit{and} \quad \|\tilde{\theta} - \theta^*\|_\infty^2 \lesssim \frac{\log n}{npL}$$

# Entrywise Distribution of The Spectral Method

Player 1  $\tilde{\theta}_1 \sim ?$




Histogram of $\tilde{\theta}_1$ from 100 independent datasets generated from $\theta^*$

# Our Result for The Spectral Method

**Theorem**

*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any* $i \in [n]$,

$$\tilde{\theta}_i - \theta_i^* = (1 + \tilde{\epsilon}_i) \frac{\sum_{j:j\neq i} A_{ij}(e^{\theta_i^*} + e^{\theta_j^*})\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{e^{\theta_i^*} \cdot \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)} + \tilde{\eta}_i.$$

*Here* $\tilde{\epsilon}, \tilde{\eta} \in \mathbb{R}^n$ *such that* $\|\tilde{\epsilon}\|_\infty = \mathfrak{o}(1), \|\tilde{\eta}\|_\infty = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$ *w.h.p..*

the main error term $\asymp \frac{1}{\sqrt{npL}}$

# Our Result for The Spectral Method

**Theorem**

*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any* $i \in [n]$,

$$\tilde{\theta}_i - \theta_i^* = (1 + \tilde{\epsilon}_i) \frac{\sum_{j:j \neq i} A_{ij}(e^{\theta_i^*} + e^{\theta_j^*})\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{e^{\theta_i^*} \cdot \sum_{j:j \neq i} A_{ij} \psi(\theta_j^* - \theta_i^*)} + \tilde{\eta}_i.$$

*Here* $\tilde{\epsilon}, \tilde{\eta} \in \mathbb{R}^n$ *such that* $\|\tilde{\epsilon}\|_\infty = \mathfrak{o}(1), \|\tilde{\eta}\|_\infty = \mathfrak{o}\left(\frac{1}{\sqrt{npL}}\right)$ *w.h.p..*

the main error term $\asymp \frac{1}{\sqrt{npL}}$

- Recall: the main error term of the MLE

$$\frac{\sum_{j:j \neq i} A_{ij}\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\sum_{j:j \neq i} A_{ij} \psi'(\theta_i^* - \theta_j^*)}$$

# Asymptotic Normality of The Spectral Method

**Corollary**

*Assume* $np \gg (\log n)^{\frac{3}{2}}$*, then for any fixed* $k \geq 1$*,*

$$\left(\tilde{\rho}_1(\theta^*)(\tilde{\theta}_1 - \theta_1^*), \ldots, \tilde{\rho}_k(\theta^*)(\tilde{\theta}_k - \theta_k^*)\right)^\top \xrightarrow{d} \mathcal{N}_k(0, I_k),$$

*where* $\tilde{\rho}_i(\theta^*) = \sqrt{L \cdot \frac{\left(\sum_{j:j\neq i} A_{ij}(e^{\theta_i^*}+e^{\theta_j^*})\psi'(\theta_i^*-\theta_j^*)\right)^2}{\sum_{j:j\neq i} A_{ij}(e^{\theta_i^*}+e^{\theta_j^*})^2\psi'(\theta_i^*-\theta_j^*)}}.$

# Asymptotic Normality of The Spectral Method

**Corollary**

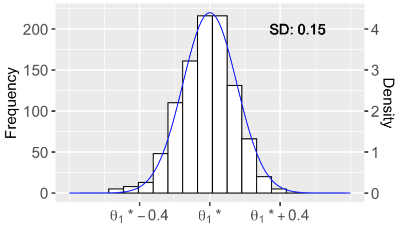
*Assume* $np \gg (\log n)^{\frac{3}{2}}$, *then for any fixed* $k \geq 1$,

$$\Big(\tilde{\rho}_1(\tilde{\theta})(\tilde{\theta}_1 - \theta_1^*), \ldots, \tilde{\rho}_k(\tilde{\theta})(\tilde{\theta}_k - \theta_k^*)\Big)^\top \xrightarrow{d} \mathcal{N}_k(0, I_k),$$

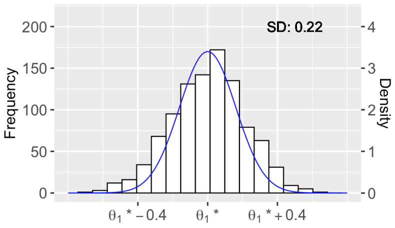
*where* $\tilde{\rho}_i(\tilde{\theta}) = \sqrt{L \cdot \frac{\left(\sum_{j:j\neq i} A_{ij}(e^{\tilde{\theta}_i} + e^{\tilde{\theta}_j})\psi'(\tilde{\theta}_i - \tilde{\theta}_j)\right)^2}{\sum_{j:j\neq i} A_{ij}(e^{\tilde{\theta}_i} + e^{\tilde{\theta}_j})^2\psi'(\tilde{\theta}_i - \tilde{\theta}_j)}}$.

# MLE vs. Spectral Method

# Asymptotic Entrywise Variances




Histogram of $\hat{\theta}_1$ from 100 datasets generated from $\theta^*$



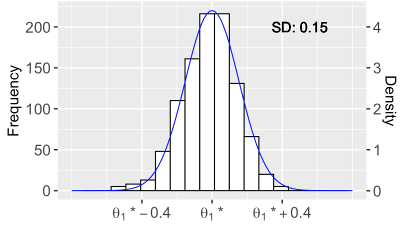

Histogram of $\tilde{\theta}_1$ from 100 datasets generated from $\theta^*$

$$\rho_i(\theta^*)(\hat{\theta}_i - \theta_i^*) \xrightarrow{d} \mathcal{N}(0,1) \qquad \tilde{\rho}_i(\theta^*)(\tilde{\theta}_i - \theta_i^*) \xrightarrow{d} \mathcal{N}(0,1)$$

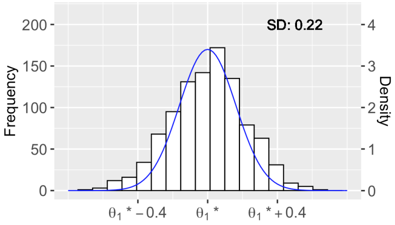
Cauchy-Schwarz yields

$$\tilde{\rho}_i(\theta^*) \leq \rho_i(\theta^*)$$

# Asymptotic Entrywise Variances




Histogram of $\hat{\theta}_1$ from 100 datasets generated from $\theta^*$




Histogram of $\tilde{\theta}_1$ from 100 datasets generated from $\theta^*$

$$\tilde{\rho}_i(\theta^*) \leq \rho_i(\theta^*)$$

**Conclusion 1**

The MLE has a smaller entrywise asymptotic variance than the spectral method.

# Exact Constants in $\ell_2$ Estimation

**Proposition (NOS17, CFMW19, CGZ20)**

*Assume* $np \gtrsim \log n$, *then w.h.p.,*

$$\textit{MLE:} \quad \|\hat{\theta} - \theta^*\|^2 \lesssim \frac{1}{pL}$$

$$\textit{Spectral:} \quad \|\tilde{\theta} - \theta^*\|^2 \lesssim \frac{1}{pL}$$

- $\frac{1}{pL}$ is the optimal rate for the $\ell_2$ estimation
- Both methods are rate-optimal

# Exact Constants in $\ell_2$ Estimation

**Theorem**

*Assume* $np \gg \log n$, *then w.h.p.,*

*MLE:* $$\|\hat{\theta} - \theta^*\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \Big( \sum_{k:k \neq i} \psi'(\theta_i^* - \theta_k^*) \Big)^{-1}$$

*Spectral:*

$$\|\tilde{\theta} - \theta^*\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \frac{\sum_{j:j \neq i} (e^{\theta_i^*} + e^{\theta_j^*})^2 \psi'(\theta_i^* - \theta_j^*)}{\Big( \sum_{j:j \neq i} (e^{\theta_i^*} + e^{\theta_j^*}) \psi'(\theta_i^* - \theta_j^*) \Big)^2}$$

# Exact Constants in $\ell_2$ Estimation

**Theorem**

*Assume* $np \gg \log n$, *then w.h.p.,*

*MLE:* $$\|\hat{\theta} - \theta^*\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \Big( \sum_{k:k \neq i} \psi'(\theta_i^* - \theta_k^*) \Big)^{-1}$$

*Spectral:*

$$\|\tilde{\theta} - \theta^*\|^2 = \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \frac{\sum_{j:j \neq i} (e^{\theta_i^*} + e^{\theta_j^*})^2 \psi'(\theta_i^* - \theta_j^*)}{\Big( \sum_{j:j \neq i} (e^{\theta_i^*} + e^{\theta_j^*}) \psi'(\theta_i^* - \theta_j^*) \Big)^2}$$

- Sharp constants with both upper and lower bounds
- MLE is better with a smaller constant

# Exact Constants in $\ell_2$ Estimation

Q: Is the MLE optimal?

A: Yes, it achieves the exact asymptotic minimax error.

**Theorem**

$$\inf_{\hat{\theta}} \sup_{\theta \in B(\theta^*)} \mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 \geq \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \Big( \sum_{j: j \neq i} \psi'(\theta_i^* - \theta_j^*) \Big)^{-1}$$

# Exact Constants in $\ell_2$ Estimation

Q: Is the MLE optimal?

A: Yes, it achieves the exact asymptotic minimax error.

**Theorem**

$$\inf_{\hat{\theta}} \sup_{\theta \in B(\theta^*)} \mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 \geq \frac{1 + \mathfrak{o}(1)}{pL} \cdot \sum_{i=1}^{n} \Big( \sum_{j: j \neq i} \psi'(\theta_i^* - \theta_j^*) \Big)^{-1}$$

**Conclusion 2**

The MLE is optimal in $\ell_2$ estimation; the spectral method is sub-optimal with a worse constant.

# Intuition for The Spectral Method

## Our result revisit

$$\tilde{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)}$$

Intuition for the main error term?

# Intuition for The Spectral Method

## Our result revisit

$$\tilde{\theta}_i - \theta_i^* \approx \frac{\sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)\big(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\big)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)}$$

Intuition for the main error term?

$$\tilde{\theta}_i = \log \hat{\pi}_i - \frac{1}{n}\sum_{j=1}^{n} \log \hat{\pi}_j$$

$$\theta_i^* = \log \pi_i^* - \frac{1}{n}\sum_{j=1}^{n} \log \pi_j^*$$

$$\tilde{\theta}_i - \theta_i^* = \log\left(1 + \frac{\hat{\pi}_i - \pi^*}{\pi^*}\right) - \frac{1}{n}\sum_{j=1}^{n}\left(1 + \frac{\hat{\pi}_j - \pi_i^*}{\pi_i^*}\right)$$

$$\approx \frac{\hat{\pi}_i - \pi_i^*}{\pi_i^*}$$

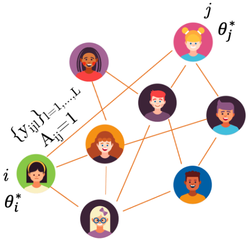
# Intuition for The Spectral Method

$$\hat{\pi}^\top = \hat{\pi}^\top P$$

$$\hat{\pi}_i = \frac{\sum_{j:j\neq i} A_{ij}\bar{y}_{ij}\hat{\pi}_j}{\sum_{j:j\neq i} A_{ij}\bar{y}_{ji}} \approx \frac{\sum_{j:j\neq i} A_{ij}\bar{y}_{ij}\pi_j^*}{\sum_{j:j\neq i} A_{ij}\bar{y}_{ji}}$$

$$\begin{aligned}
\tilde{\theta}_i - \theta_i^* &\approx \frac{\hat{\pi}_i - \pi_i^*}{\pi_i^*} = \frac{\sum_{j:j\neq i} A_{ij}\left(\bar{y}_{ij}\pi_j^* - \bar{y}_{ji}\pi_i^*\right)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\bar{y}_{ji}} \\
&= \frac{\sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\bar{y}_{ji}} \\
&\approx \frac{\sum_{j:j\neq i} A_{ij}(\pi_i^* + \pi_j^*)\left(\bar{y}_{ij} - \psi(\theta_i^* - \theta_j^*)\right)}{\pi_i^* \cdot \sum_{j:j\neq i} A_{ij}\psi(\theta_j^* - \theta_i^*)}
\end{aligned}$$

# Summary: Uncertainty Quantification in BTL Model




- Non-asymptotic expansion for the MLE
  - HT/CI for $\theta_i^*$ and $r_i^*$
- Non-asymptotic expansion for the spectral method
  - MLE vs. spectral method

Chao Gao, Yandi Shen, and Anderson Y Zhang. Uncertainty quantification in the bradley-terry-luce model.
*arXiv preprint arXiv:2110.03874*, 2021

# Thank You