Recent Advances in Ranking: Adversarial Respondents and Lower Bounds on the Bayes Risk

Vincent Y. F. Tan

Department of Electrical and Computer Engineering,
Department of Mathematics,
National University of Singapore

McMaster University
June 15, 2018
1. Introduction to Statistical Models for Ranking

2. Fundamental Limits of Top-$K$ Ranking with Adversaries

3. Lower Bounds on the Bayes Risk of a Bayesian BTL Model
1. Introduction to Statistical Models for Ranking

2. Fundamental Limits of Top-$K$ Ranking with Adversaries

3. Lower Bounds on the Bayes Risk of a Bayesian BTL Model
A fundamental problem in a wide range of contexts
A fundamental problem in a wide range of contexts

Applications: web search, recommendation systems, social choice, sports competitions, voting, etc.
A fundamental problem in a wide range of contexts

Applications: web search, recommendation systems, social choice, sports competitions, voting, etc.

Efforts in developing various ranking algorithms
- A fundamental problem in a wide range of contexts
- Applications: web search, recommendation systems, social choice, sports competitions, voting, etc.
- Efforts in developing various ranking algorithms
- A variety of statistical models introduced for evaluating ranking algorithms
Example: Web search
Example: Web search

- $n = 10^9$ websites
- $\binom{n}{2} \approx n^2 = 10^{18}$ comparisons
- Do we really need $\Theta(n^2)$ comparisons?
Suppose that
- we want a total ordering
- pairwise comparisons are randomly given (probabilistically).

This indeed requires $\Theta(n^2)$ comparisons
Suppose that

- we want a total ordering
- pairwise comparisons are randomly given (probabilistically).

This indeed requires $\Theta(n^2)$ comparisons

No way to identify the ordering between 1 and 2 without a direct comparison, i.e., comparison must be made w.p. 1
Suppose that
- we want a total ordering
- pairwise comparisons are randomly given (probabilistically).

This indeed requires $\Theta(n^2)$ comparisons.

- No way to identify the ordering between 1 and 2 without a direct comparison, i.e., comparison must be made w.p. 1

- Worse with noisy data
Suppose that
- we want a **total ordering**
- pairwise comparisons are **randomly given** (probabilistically).

This indeed requires $\Theta(n^2)$ comparisons

No way to identify the ordering between 1 and 2 without a direct comparison, i.e., comparison must be made w.p. 1

Worse with noisy data

Adopt a Shannon-theoretic approach in our analyses
Top-$K$ Ranking Usually Suffices

Huge number of movies

Find only top $K = 3$
Top-$K$ Ranking Usually Suffices

Huge number of movies

Find only top $K = 3$

Pairwise sample

$\omega$ score vector

$Y := \{Y_{ij}^{(\ell)}\}$

(top-$K$ ranking $\psi(\cdot)$)

$\hat{S}_K$
Adopt the Bradley-Terry-Luce or BTL model in which there is an underlying unknown score vector

\[ \mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_+, \]

where \( w_i \) is the likeability of movie \( i \).
Adopt the Bradley-Terry-Luce or BTL model in which there is an underlying unknown score vector

\[ \mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_{++}, \]

where \( w_i \) is the likeability of movie \( i \).

Decide which items to compare via a comparison graph.
The outcome of the comparison between item 1 and 2 is

\[ Y_{12} = \mathbb{I}\{\text{item 1} \succ \text{item 2}\} \sim \text{Bern}\left(\frac{w_1}{w_1 + w_2}\right). \]
The outcome of the comparison between item 1 and 2 is
\[ Y_{12} = \mathbb{I}\{\text{item 1} > \text{item 2}\} \sim \text{Bern}\left(\frac{w_1}{w_1 + w_2}\right). \]

E.g., \( w_1 = 0.9 \) and \( w_2 = 0.1 \), then item 1 beats item 2 w.p. 90\%. 

We have \( L \) independent copies \( Y^{(1)}_{ij}, \ldots, Y^{(L)}_{ij} \) for each observed edge \( \{i, j\} \in E \) of the observation graph. Determine fundamental limits on \( L \) (as a function of \( n \) and other parameters) so that recovery of top-\( K \) set is successful.
The outcome of the comparison between item 1 and 2 is

\[ Y_{12} = \mathbb{I}\{\text{item 1} \succ \text{item 2}\} \sim \text{Bern} \left( \frac{w_1}{w_1 + w_2} \right). \]

E.g., \( w_1 = 0.9 \) and \( w_2 = 0.1 \), then item 1 beats item 2 w.p. 90%.

We have \( L \) independent copies

\[ Y_{ij}^{(1)}, \ldots, Y_{ij}^{(L)} \]

for each observed edge \( \{i,j\} \in \mathcal{E} \) of the observation graph.
The outcome of the comparison between item 1 and 2 is

\[ Y_{12} = \mathbb{I}\{\text{item 1} \succ \text{item 2}\} \sim \text{Bern}\left(\frac{w_1}{w_1 + w_2}\right). \]

E.g., \( w_1 = 0.9 \) and \( w_2 = 0.1 \), then item 1 beats item 2 w.p. 90\%.

We have \( L \) independent copies

\[ Y_{ij}^{(1)}, \ldots, Y_{ij}^{(L)} \]

for each observed edge \( \{i,j\} \in \mathcal{E} \) of the observation graph.

Determine fundamental limits on \( L \) (as a function of \( n \) and other parameters) so that recovery of top-\( K \) set is successful.
Outline

1. Introduction to Statistical Models for Ranking

2. Fundamental Limits of Top-$K$ Ranking with Adversaries

3. Lower Bounds on the Bayes Risk of a Bayesian BTL Model
Top-$K$ Ranking with Adversaries

Joint work with

Changho Suh (KAIST)  Renbo Zhao (NUS)
Top-$K$ Ranking with Adversaries

Joint work with

Changho Suh (KAIST)  Renbo Zhao (NUS)

Ranking with Adversaries: Crowdsourced Setting

\[ Y_{ij} \sim \text{Bern}(\eta \cdot w_i + w_j + (1 - \eta) \cdot w_j w_i + w_j) \]

Spammers provide answers in an adversarial manner

\[ Y_{ij} \sim \text{Bern}(w_j w_i + w_j) \]
Y_{ij} \sim \text{Bern}\left(\frac{w_i}{w_i + w_j}\right)
Y_{ij} \sim \text{Bern}\left(\frac{w_i}{w_i + w_j}\right)

Y_{ij} \sim \text{Bern}\left(\frac{1/w_i}{1/w_i + 1/w_j}\right)
Y_{ij} \sim \text{Bern} \left( \frac{w_i}{w_i + w_j} \right)

Y_{ij} \sim \text{Bern} \left( \frac{1/w_i}{1/w_i + 1/w_j} \right)

= \text{Bern} \left( \frac{w_j}{w_i + w_j} \right)
Spammers provide answers in an adversarial manner
Spammers provide answers in an adversarial manner

$$Y_{ij} \sim \text{Bern} \left( \eta \cdot \frac{w_i}{w_i + w_j} + (1 - \eta) \cdot \frac{w_j}{w_i + w_j} \right)$$
Given an observed pair, each sample has different distributions

\[ Y_{ij}^{(l)} \sim \text{Bern} \left( \frac{\eta_l \cdot w_i}{w_i + w_j} + (1 - \eta_l) \cdot \frac{1/w_i}{1/w_i + 1/w_j} \right) \]

where \( \eta_l \) is a quality parameter of measurement \( l \)

---

\(^1\)X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, “Pairwise ranking aggregation in a crowdsourced setting,” in WSDM, 2013
Given an observed pair, each sample has different distributions

\[ Y_{ij}^{(l)} \sim \text{Bern} \left( \eta_l \cdot \frac{w_i}{w_i + w_j} + (1 - \eta_l) \cdot \frac{1/w_i}{1/w_i + 1/w_j} \right) \]

where \( \eta_l \) is a quality parameter of measurement \( l \)

Subsumes as a special case our adversarial BTL model when all quality parameters are the same

---

\(^1\)X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, “Pairwise ranking aggregation in a crowdsourced setting,” in WSDM, 2013
Related Work: Crowdsourced BTL [Chen et al. ’13]¹

Given an observed pair, each sample has different distributions

\[ Y_{ij}^{(l)} \sim \text{Bern} \left( \eta_l \cdot \frac{w_i}{w_i + w_j} + (1 - \eta_l) \cdot \frac{1/w_i}{1/w_i + 1/w_j} \right) \]

where \( \eta_l \) is a quality parameter of measurement \( l \)

Subsumes as a special case our adversarial BTL model when all quality parameters are the same

The authors developed a ranking algorithm but without theoretical guarantees

¹X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, “Pairwise ranking aggregation in a crowdsourced setting,” in WSDM, 2013
Related Work: Crowdsourced BTL [Chen et al. ’13]

- Given an observed pair, each sample has different distributions

\[ Y_{ij}^{(l)} \sim \text{Bern} \left( \eta_l \cdot \frac{w_i}{w_i + w_j} + (1 - \eta_l) \cdot \frac{1/w_i}{1/w_i + 1/w_j} \right) \]

where \( \eta_l \) is a quality parameter of measurement \( l \)

- Subsumes as a special case our adversarial BTL model when all quality parameters are the same

- The authors developed a ranking algorithm but without theoretical guarantees

- More difficult to analyze as there are many more parameters

---

1X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, “Pairwise ranking aggregation in a crowdsourced setting,” in WSDM, 2013
Goal of Adversarial Top-$K$ Ranking

Erdős-Rényi comparison graph

\[ P_e := \Pr[\{\text{top-}K\} \neq \{\text{top-}K\}] \]
Goal of Adversarial Top-$K$ Ranking

Erdös-Rényi comparison graph

\[ P_e := \Pr[\{\text{top-$K$}\} \neq \{\text{top-$K$}\}] \]

- Ranking infeasible: \( P_e \not\to 0 \)
- Ranking feasible: \( P_e \to 0 \)

Sample size

Vincent Y. F. Tan (NUS)
Contribution #1: $1/2 < \eta < 1$ known

\[ \eta = \text{Fraction of non-adversaries}; \ \Delta_K \approx w_K - w_{K+1} \]

\[ \eta = 1 \text{ studied by Chen and Suh (2015)} \]

Contribution #1: $1/2 < \eta < 1$ known

$\eta = \text{Fraction of non-adversaries}; \Delta_K \asymp w_K - w_{K+1}$

Sample complexity

$\approx \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2}$

$\eta = 1$ studied by Chen and Suh (2015)$^2$

---

Contribution #1: \(1/2 < \eta < 1\) known

\(\eta = \text{Fraction of non-adversaries}; \ \Delta_K \asymp w_K - w_{K+1}\)

\[\text{Sample complexity} \approx \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2}\]

\(\eta = 1\) studied by Chen and Suh (2015)\(^2\)

Contribution #1: $1/2 < \eta < 1$ known

$\eta = \text{Fraction of non-adversaries; } \Delta_K \asymp w_K - w_{K+1}$

Sample complexity

$$\sim \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2}$$

$\eta = 1$ studied by Chen and Suh (2015)$^2$

---

Contribution #1: $1/2 < \eta < 1$ known

Experimental Results for $n = 1000$ and $K = 10$

\[ \hat{S} \approx \frac{C}{(2\eta - 1)^2} \]

\[ \hat{S} = \binom{n}{2} p \hat{L}_{\text{ave}} \]

\[ \hat{S}_{\text{norm}} = \frac{\hat{S}}{(n \log n)/\Delta_2^K} \]
Contribution #2: \( \frac{1}{2} < \eta < 1 \) unknown

\[ \eta = \text{Fraction of non-adversaries}; \quad \Delta_K \approx w_K - w_{K+1} \]

sample complexity

0.5

1
Contribution #2: $1/2 < \eta < 1$ unknown

$\eta = \text{Fraction of non-adversaries;} \; \Delta_K \asymp w_K - w_{K+1}$

Sample complexity

\[ \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2} \]

Infeasible
Contribution #2: $1/2 < \eta < 1$ unknown

\[ \eta = \text{Fraction of non-adversaries; } \Delta_K \asymp w_K - w_{K+1} \]

\[
\frac{n \log^2 n}{(2\eta - 1)^4 \Delta^4_K} \\
\frac{n \log n}{(2\eta - 1)^2 \Delta^2_K}
\]

sample complexity

polynomial time algorithm

infeasible

0.5

1

\eta
Contribution #2: $1/2 < \eta < 1$ unknown

$\eta = \text{Fraction of non-adversaries; } \Delta_K \asymp w_K - w_{K+1}$

Sample complexity

$$\frac{n \log^2 n}{(2\eta - 1)^4 \Delta_K^4}$$

Infeasible

Polynomial time algorithm

??
Optimality

\[
\text{ranking infeasible} \quad \iff \quad \text{ranking feasible}
\]

\[
P_e \to 0 \quad \iff \quad P_e \to 0
\]

\[
\sim \frac{n \log n}{(2\eta - 1)^2 \Delta^2_K}
\]
Optimality

Minimax optimality: Construct “worst-case” score vectors

\[ P_e \rightarrow 0 \quad \sim \quad \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2} \]

- Minimax optimality: Construct “worst-case” score vectors
Minimax optimality: Construct “worst-case” score vectors

Translation to $M$-ary hypothesis testing: Construction of multiple hypotheses
Minimax optimality: Construct “worst-case” score vectors

Translation to $M$-ary hypothesis testing: Construction of multiple hypotheses

Information-theoretic ideas applied to statistical learning
Construction of $M := \min\{K, n - K\} + 1 \leq n/2$ hypotheses:

$$\Pr(\sigma([K]) = S) = \frac{1}{M}, \text{ for } S = \{2, \ldots, K\} \cup \{i\}, \text{ } i = 1, K + 1, \ldots, n$$
Construction of $M := \min\{K, n-K\} + 1 \leq n/2$ hypotheses:

$$\Pr(\sigma([K]) = S) = \frac{1}{M}, \text{ for } S = \{2, \ldots, K\} \cup \{i\}, \ i = 1, K+1, \ldots, n$$

Bound mutual info. of permutation and “erased” version of $Y_{ij}^{(l)}$:

$$I(\sigma; Z) \leq \frac{p}{M^2} \sum_{\sigma_1, \sigma_2 \in M} \sum_{l=1}^{L} \left\{ \sum_{i \neq j} D \left( P_{Y_{ij}^{(l)}|\sigma_1} \parallel P_{Y_{ij}^{(l)}|\sigma_2} \right) \right\}$$
Construction of $M := \min\{K, n - K\} + 1 \leq n/2$ hypotheses:

$$\Pr(\sigma([K]) = S) = \frac{1}{M}, \text{ for } S = \{2, \ldots, K\} \cup \{i\}, \quad i = 1, K + 1, \ldots, n$$

Bound mutual info. of permutation and “erased” version of $Y_{ij}^{(l)}$:

$$I(\sigma; \mathbf{Z}) \leq \frac{p}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{l=1}^{L} \left\{ \sum_{i \neq j} D \left( P_{Y_{ij}^{(l)}|\sigma_1} \left\| P_{Y_{ij}^{(l)}|\sigma_2} \right. \right) \right\}$$

Bound the divergence using reverse Pinsker’s inequality. Here is where $\Delta_K$ comes in

$$\sum_{i \neq j} D \left( P_{Y_{ij}^{(l)}|\sigma_1} \left\| P_{Y_{ij}^{(l)}|\sigma_2} \right. \right) \leq n \cdot (2\eta - 1)^2 \cdot \Delta_K^2$$
Optimality: Tools

- Construction of $M := \min\{K, n - K\} + 1 \leq n/2$ hypotheses:

  $$\Pr(\sigma([K]) = S) = \frac{1}{M}, \text{ for } S = \{2, \ldots, K\} \cup \{i\}, \ i = 1, K + 1, \ldots, n$$

- Bound mutual info. of permutation and “erased” version of $Y_{ij}^{(l)}$:

  $$I(\sigma; Z) \leq \frac{p}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{l=1}^{L} \left\{ \sum_{i \neq j} D\left( P_{Y_{ij}^{(l)}|\sigma_1} \parallel P_{Y_{ij}^{(l)}|\sigma_2} \right) \right\}$$

- Bound the divergence using reverse Pinsker’s inequality. Here is where $\Delta_K$ comes in

  $$\sum_{i \neq j} D\left( P_{Y_{ij}^{(l)}|\sigma_1} \parallel P_{Y_{ij}^{(l)}|\sigma_2} \right) \leq n \cdot (2\eta - 1)^2 \cdot \Delta_K^2$$

- Fano’s inequality
Ranking Algorithm for $\eta$ Known: Part I

pairwise samples $\{Y_{ij}\}$ → \text{ESTIMATE} $\mathbf{w}$ → \text{RETURN A RANKING} → \{top-K\}
Scores determine the ranking
Scores determine the ranking

- Adopt a two-step approach
Ranking Algorithm for $\eta$ Known: Part II

Key Message:
Small $\text{MSE} \Rightarrow$ Small $\ell_\infty$ Error of $\hat{w}$ $\Rightarrow$ High Ranking Accuracy

Vincent Y. F. Tan
(NUS)

Spectral MLE

Stage 1
SPECTRAL METHODS
RankCentrality
Negabahn et.al. 12

Stage 2
POINT-WISE MLE

\[ \hat{w} \]

RETURN A RANKING

\{Top-K\}

Pairwise samples \( \{Y_{i,j}\} \)
Key Message:

Small MSE $\rightarrow$ Small $\ell_\infty$ Error of $\hat{w}$ $\rightarrow$ High Ranking Accuracy
How to ensure small MSE for $\eta = 1$?

\[
\frac{1}{L} \sum_{\ell=1}^{L} Y_{ij}^{(\ell)}
\]

Detailed balance equation:

\[
\pi_i \cdot w_j w_i + w_j = \pi_j \cdot w_i w_i + w_j
\]

where $\pi := [\pi_1, \pi_2, \ldots, \pi_n]$ is the stationary distribution of the chain. Stationary distribution converges to $w$ (up to constant scaling), i.e.,

\[
\lim_{L \to \infty} \pi(L) = \alpha w.
\]
How to ensure small MSE for $\eta = 1$?

- Recall $\eta = 1$ (no adversaries)
How to ensure small MSE for \( \eta = 1 \)?

- Recall \( \eta = 1 \) (no adversaries)
- \( L \) independent copies \( Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)} \)
How to ensure small MSE for $\eta = 1$?

- Recall $\eta = 1$ (no adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Convergence to stationary distribution

$$\frac{1}{L} \sum_{\ell=1}^{L} Y_{ij}^{(\ell)} \rightarrow \frac{w_i}{w_i + w_j}$$

where $\pi := [\pi_1, \pi_2, \ldots, \pi_n]$ is the stationary distribution of the chain.

$$\lim_{L \to \infty} \pi(\ell) = \alpha w.$$
How to ensure small MSE for $\eta = 1$?

- Recall $\eta = 1$ (no adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Convergence to stationary distribution

$$\frac{1}{L} \sum_{\ell=1}^{L} Y_{ij}^{(\ell)} \to \frac{w_i}{w_i + w_j}$$

- Detailed balance equation:

$$\pi_i \cdot \frac{w_j}{w_i + w_j} = \pi_j \cdot \frac{w_i}{w_i + w_j}$$

where $\pi := [\pi_1, \pi_2, \ldots, \pi_n]$ is the stat. distn. of the chain.
How to ensure small MSE for $\eta = 1$?

- Recall $\eta = 1$ (no adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Convergence to stationary distribution

$$\frac{1}{L} \sum_{\ell=1}^{L} Y_{ij}^{(\ell)} \rightarrow \frac{w_i}{w_i + w_j}$$

- Detailed balance equation:

$$\pi_i \cdot \frac{w_j}{w_i + w_j} = \pi_j \cdot \frac{w_i}{w_i + w_j}$$

where $\pi := [\pi_1, \pi_2, \ldots, \pi_n]$ is the stat. distn. of the chain.

- Stationary distribution converges to $\mathbf{w}$ (up to constant scaling), i.e.,

$$\lim_{L \to \infty} \pi^{(L)} = \alpha \mathbf{w}.$$
How to ensure small MSE for \( \eta \in (1/2, 1] \)?

\[
Y_{ij}, \quad i, j
\]

\[
\tilde{Y}_{ij}
\]

\[
\sum_{l=1}^{L} Y_{ij} \to \eta w_i w_i + w_j + (1 - \eta) w_j w_i + w_j = (2\eta - 1) w_i w_i + w_j + (1 - \eta)
\]

Redefine Markov chain with transition probabilities \( \{\tilde{Y}_{ij}\} \).

Vincent Y. F. Tan (NUS)
How to ensure small MSE for $\eta \in (1/2, 1]$?

- Arbitrary $\eta \in (1/2, 1]$ (adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Redefine Markov chain
How to ensure small MSE for $\eta \in (1/2, 1]$?

- Arbitrary $\eta \in (1/2, 1]$ (adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Redefine Markov chain

- We instead have the following convergence:

$$
\frac{1}{L} \sum_{l=1}^{L} Y_{ij}^{(l)} \rightarrow \eta \frac{w_i}{w_i + w_j} + (1 - \eta) \frac{w_j}{w_i + w_j} = (2\eta - 1) \frac{w_i}{w_i + w_j} + (1 - \eta)
$$
How to ensure small MSE for $\eta \in (1/2, 1]$?

- Arbitrary $\eta \in (1/2, 1]$ (adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Redefine Markov chain

We instead have the following convergence:

$$\frac{1}{L} \sum_{l=1}^{L} Y_{ij}^{(l)} \rightarrow \eta \frac{w_i}{w_i + w_j} + (1 - \eta) \frac{w_j}{w_i + w_j} = (2\eta - 1) \frac{w_i}{w_i + w_j} + (1 - \eta)$$

Redefine “shifted” samples with range scaled by $2\eta - 1$:

$$\tilde{Y}_{ij} = \frac{1}{2\eta - 1} \left[ \frac{1}{L} \sum_{l=1}^{L} Y_{ij}^{(l)} - (1 - \eta) \right] \rightarrow \frac{w_i}{w_i + w_j}$$
How to ensure small MSE for $\eta \in (1/2, 1]$?

- Arbitrary $\eta \in (1/2, 1]$ (adversaries)
- $L$ independent copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \ldots, Y_{ij}^{(L)}$
- Redefine Markov chain

We instead have the following convergence:

$$\frac{1}{L} \sum_{l=1}^{L} Y_{ij}^{(l)} \to \eta \frac{w_i}{w_i + w_j} + (1 - \eta) \frac{w_j}{w_i + w_j} = (2\eta - 1) \frac{w_i}{w_i + w_j} + (1 - \eta)$$

- Redefine “shifted” samples with range scaled by $2\eta - 1$:

$$\tilde{Y}_{ij} = \frac{1}{2\eta - 1} \left[ \frac{1}{L} \sum_{l=1}^{L} Y_{ij}^{(l)} - (1 - \eta) \right] \to \frac{w_i}{w_i + w_j}$$

- Construct Markov chain with transition probabilities $\{\tilde{Y}_{ij}\}$. \

Vincent Y. F. Tan (NUS)
Use several concentration inequalities (Hoeffding, Bernstein, Tropp, etc.), we can show that if
\[ \text{sample size} = L(n^2 \log n)^2 \]
\[ \geq n \log n (2 \eta - 1)^2 \Delta K \]
\( \Rightarrow \) Feasible Top-K Ranking.
Use several concentration inequalities (Hoeffding, Bernstein, Tropp, etc.), we can show that if

\[
\text{sample size } = L \left( \begin{array}{c} n \\ 2 \end{array} \right) p \geq \frac{n \log n}{(2\eta - 1)^2 \Delta^2_K} \implies \text{Feasible Top-}K \text{ Ranking}
\]
What if $\eta$ is unknown?

- Adversarial BTL model is a mixture model

---


What if $\eta$ is unknown?

- Adversarial BTL model is a mixture model

- Obtaining global optimality guarantees for mixture model problems is difficult in general

---


What if $\eta$ is unknown?

- Adversarial BTL model is a mixture model

- Obtaining global optimality guarantees for mixture model problems is difficult in general

- Recent developments:
  - Tensor methods: Jain and Oh$^3$ and Anandkumar et al.$^4$
  - Key idea: Exact 2nd and 3rd moments yield sufficient statistics

---


What if \( \eta \) is unknown?

- Adversarial BTL model is a mixture model
- Obtaining global optimality guarantees for mixture model problems is difficult in general
- Recent developments:
  - Tensor methods: Jain and Oh\(^3\) and Anandkumar et al.\(^4\)
  - Key idea: Exact 2nd and 3rd moments yield sufficient statistics
- Our setting:
  - Can obtain estimates of 2nd and 3rd moments
  - Can estimate \( \eta \)

---


Turn weights into distribution vectors

$$\pi_0 = \begin{bmatrix} \cdots & \frac{w_i}{w_i + w_j} & \frac{w_j}{w_i + w_j} & \frac{w_i'}{w_i' + w_j'} & \frac{w_j'}{w_i' + w_j'} & \cdots \end{bmatrix}^T$$
Turn weights into distribution vectors

$$\pi_0 = \begin{bmatrix} \cdots & w_i & w_j & w_i' & w_j' & \cdots \end{bmatrix}^T$$

Estimate moments. Ground truth moment matrix and tensor are:

$$M_2 := \eta \pi_0 \otimes \pi_0 + (1 - \eta) \pi_1 \otimes \pi_1,$$

$$M_3 := \eta \pi_0 \otimes \pi_0 \otimes \pi_0 + (1 - \eta) \pi_1 \otimes \pi_1 \otimes \pi_1.$$
1. Turn weights into distribution vectors

$$\pi_0 = \left[ \cdots \frac{w_i}{w_i + w_j} \frac{w_j}{w_i + w_j} \frac{w_i'}{w_i' + w_j'} \frac{w_j'}{w_i' + w_j'} \cdots \right]^T$$

2. Estimate moments. Ground truth moment matrix and tensor are:

$$M_2 := \eta \pi_0 \otimes \pi_0 + (1 - \eta) \pi_1 \otimes \pi_1,$$

$$M_3 := \eta \pi_0 \otimes \pi_0 \otimes \pi_0 + (1 - \eta) \pi_1 \otimes \pi_1 \otimes \pi_1.$$  

3. Solves a Least Squares Problem

$$\hat{G} \in \arg \min_{Z \in \mathbb{R}^{2 \times 2 \times 2}} \left\| \mathcal{P}_{\Omega_3} \left( Z \left[ P_{\hat{M}_2} \right]_3 - \frac{1}{|\mathcal{I}_2|} \sum_{t \in \mathcal{I}_2} \otimes^3 Y(t) \right) \right\|^2_F$$
1. Turn weights into distribution vectors

\[ \pi_0 = \left[ \ldots \begin{array}{cccc} w_i & w_j & w_i' & w_j' \\ w_i + w_j & w_i + w_j & w_i' + w_j' & w_i' + w_j' \end{array} \ldots \right]^T \]

2. Estimate moments. Ground truth moment matrix and tensor are:

\[
M_2 := \eta \pi_0 \otimes \pi_0 + (1 - \eta) \pi_1 \otimes \pi_1,
\]

\[
M_3 := \eta \pi_0 \otimes \pi_0 \otimes \pi_0 + (1 - \eta) \pi_1 \otimes \pi_1 \otimes \pi_1.
\]

3. Solves a Least Squares Problem

\[
\hat{G} \in \arg \min_{Z \in \mathbb{R}^{2 \times 2 \times 2}} \left\| \mathcal{P}_{\Omega_3} \left( Z \left[ P_{\hat{M}_2} \right]_3 - \frac{1}{|\mathcal{I}_2|} \sum_{t \in \mathcal{I}_2} \otimes^3 Y(t) \right) \right\|^2_F
\]

4. Find leading eigenvalue \( \lambda_1(\hat{G}) \) of \( \hat{G} \) which is related to \( \eta \) as follows:

\[
\hat{\eta} = \lambda_1(\hat{G})^{-2}
\]
How does the quality of the estimation of $\eta$ affect overall sample complexity?
How does the quality of the estimation of $\eta$ affect overall sample complexity?
With very careful analysis, we can derive a meta-lemma

\[ |\hat{\eta} - \eta| \leq \epsilon \implies \text{Sample size } = \frac{L(n^2)}{p} \geq \frac{n \log^2 n}{\epsilon^2}. \]
Tradeoff Between $|\hat{\eta} - \eta|$ and Sample Complexity

- With very careful analysis, we can derive a meta-lemma

\[ |\hat{\eta} - \eta| \leq \epsilon \implies \text{Sample size} = L\binom{n}{2}p \geq \frac{n \log^2 n}{\epsilon^2} \]

- This implies that
  - $|\hat{\eta} - \eta|$ ↓ implies that $\|\hat{w} - w\|_\infty$ ↓ but sample size ↑
With very careful analysis, we can derive a meta-lemma

$$\hat{\eta} - \eta \leq \epsilon \implies \text{Sample size} = L\left(\binom{n}{2}\right)p \geq \frac{n \log^2 n}{\epsilon^2}$$

This implies that

- $|\hat{\eta} - \eta| \downarrow$ implies that $\|\hat{w} - w\|_\infty \downarrow$ but sample size $\uparrow$
- $|\hat{\eta} - \eta| \uparrow$ implies that sample size $\downarrow$ but $\|\hat{w} - w\|_\infty \uparrow$
With very careful analysis, we can derive a meta-lemma

\[ |\hat{\eta} - \eta| \leq \epsilon \implies \text{Sample size} = L \left( \frac{n}{2} \right) p \geq \frac{n \log^2 n}{\epsilon^2} \]

This implies that

- \( |\hat{\eta} - \eta| \downarrow \) implies that \( \|\hat{w} - w\|_\infty \downarrow \) but sample size \( \uparrow \)
- \( |\hat{\eta} - \eta| \uparrow \) implies that sample size \( \downarrow \) but \( \|\hat{w} - w\|_\infty \uparrow \)

Find a sweet spot to show that

\[ \text{sample size} \geq \frac{n \log^2 n}{(2\eta - 1)^4 \Delta^4_K}, \implies \text{Feasible Top-}K \text{ ranking} \]
Conclusion for Adversarial Top-$K$ Ranking

- Explored a Top-$K$ ranking problem for an adversarial setting
Conclusion for Adversarial Top-$K$ Ranking

- Explored a Top-$K$ ranking problem for an adversarial setting
- Characterized exact order-wise optimal sample complexity for $\eta$-known case
Conclusion for Adversarial Top-$K$ Ranking

- Explored a Top-$K$ ranking problem for an adversarial setting
- Characterized exact order-wise optimal sample complexity for $\eta$-known case
- Established an upper bound on the sample complexity for the $\eta$-unknown case
Conclusion for Adversarial Top-$K$ Ranking

- Explored a Top-$K$ ranking problem for an adversarial setting
- Characterized exact order-wise optimal sample complexity for $\eta$-known case
- Established an upper bound on the sample complexity for the $\eta$-unknown case
- Developed computationally efficient algorithms for both cases (using state-of-the-art tensor methods for the $\eta$-unknown case)
Conclusion for Adversarial Top-\(K\) Ranking

- Explored a Top-\(K\) ranking problem for an adversarial setting
- Characterized exact order-wise optimal sample complexity for \(\eta\)-known case
- Established an upper bound on the sample complexity for the \(\eta\)-unknown case
- Developed computationally efficient algorithms for both cases (using state-of-the-art tensor methods for the \(\eta\)-unknown case)
- C. Suh, VYFT and R. Zhao “Adversarial Top-\(K\) Ranking”, IEEE Trans. on Inf. Theory, Apr 2017
Outline

1. Introduction to Statistical Models for Ranking

2. Fundamental Limits of Top-$K$ Ranking with Adversaries

3. Lower Bounds on the Bayes Risk of a Bayesian BTL Model
Lower Bounds on the Risk of a Bayesian BTL Model

Joint work with

Mine Alsan
(NUS)

Ranjitha Prasad
(TCS Innovation Labs, Delhi)
Lower Bounds on the Risk of a Bayesian BTL Model

Joint work with

Mine Alsan
(NUS)

Ranjitha Prasad
(TCS Innovation Labs, Delhi)

Summary of contributions

- Study the fundamental performance limits of ranking algorithms in the Bradley-Terry-Luce model within a Bayesian framework:

  1. Derive lower bounds on the Bayes Risk of estimators.
     - A family of information-theoretic lower bounds for norm-based distortion functions \( \| \cdot \|_r \), for any \( r \geq 1 \).
     - The Bayesian Cramér-Rao bound for the MSE, i.e., \( r = 2 \).

  2. Explore optimal comparison graph structures to design experiments minimizing distortion.
Summary of contributions

- Study the fundamental performance limits of ranking algorithms in the Bradley-Terry-Luce model within a Bayesian framework:

  1. Derive lower bounds on the Bayes Risk of estimators.

     - A family of information-theoretic lower bounds for norm-based distortion functions $\|\cdot\|_r$, for any $r \geq 1$.

     - The Bayesian Cramér-Rao bound for the MSE, i.e., $r = 2$. 

Summary of contributions

- Study the fundamental performance limits of ranking algorithms in the Bradley-Terry-Luce model within a Bayesian framework:

  1. Derive lower bounds on the Bayes Risk of estimators.
     - A family of information-theoretic lower bounds for norm-based distortion functions $\| \cdot \|^r$, for any $r \geq 1$.
     - The Bayesian Cramér-Rao bound for the MSE, i.e., $r = 2$.

  2. Explore optimal comparison graph structures to design experiments minimizing distortion.
Recall the BTL Model

BTL model: To each item $i \in [n]$, a skill parameter $w_i \in \mathbb{R}^+$ s.t.

$$P_{ij} := \Pr[item \ i \succ item \ j] = \frac{w_i}{w_i + w_j}.$$
Recall the BTL Model

- **BTL model**: To each item $i \in [n]$, a skill parameter $w_i \in \mathbb{R}^{++}$ s.t.

$$P_{ij} := \Pr[\text{item } i \succ \text{ item } j] = \frac{w_i}{w_i + w_j}.$$ 

⇒ Instead of Top-$K$ ranking, now we want to estimate the vector

$$\mathbf{w} := (w_1, \ldots, w_n) \in \mathbb{R}_+^n$$
Recall the BTL Model

- BTL model: To each item \( i \in [n] \), a skill parameter \( w_i \in \mathbb{R}_{++} \) s.t.
  \[
P_{ij} := \Pr[\text{item } i \succ \text{item } j] = \frac{w_i}{w_i + w_j}.
\]

  ⇒ Instead of Top-\( K \) ranking, now we want to estimate the vector
  \[
  \mathbf{w} := (w_1, \ldots, w_n) \in \mathbb{R}_+^n
  \]

- Given
  \[
  m = \sum_{(i,j):i \neq j} m_{ij} \in \mathbb{N}
  \]
  indep. pairwise comparisons, we count:
Recall the BTL Model

- **BTL model**: To each item \( i \in [n] \), a skill parameter \( w_i \in \mathbb{R}_{++} \) s.t.

\[
P_{ij} := \Pr[\text{item } i > \text{ item } j] = \frac{w_i}{w_i + w_j}.
\]

\( \Rightarrow \) Instead of Top-\( K \) ranking, now we want to estimate the vector

\[
w := (w_1, \ldots, w_n) \in \mathbb{R}^n_{++}
\]

- Given

\[
m = \sum_{(i,j):i \neq j} m_{ij} \in \mathbb{N}
\]

indep. pairwise comparisons, we count:

1. \( m_{ij} \): Num. of pairwise comparisons between items \( i \) & \( j \),
2. \( b_{ij} \): Num. of comparisons in which \( i \) is preferred over \( j \).

\( \Rightarrow \) \( M := \{m_{ij}\} \in \mathbb{N}^{n \times n} \) and \( B := \{b_{ij}\} \in \mathbb{N}^{n \times n} \).
We assume that the matrix $\mathbf{M} = \{m_{ij}\}$ is fixed a priori.
We assume that the matrix $M = \{m_{ij}\}$ is fixed a priori.

The BTL model induces the following distributions:

1. For fixed $m_{ij}$,

   $$p(b_{ij}|w_i, w_j) = \text{Bin}(b_{ij}; m_{ij}, P_{ij}).$$

2. For fixed $M$,

   $$p(B|\lambda) = \prod_{(i,j):i<j} \text{Bin}(b_{ij}; m_{ij}, P_{ij}),$$
Bayesian BTL Model

- Adopt the Bayesian BTL framework by Caron & Doucet\textsuperscript{5}:

\textsuperscript{5}F. Caron and A. Doucet, “Efficient Bayesian Inference for Generalized Bradley-Terry Models”, in JCGS, 2012
Bayesian BTL Model

- Adopt the Bayesian BTL framework by Caron & Doucet\(^5\):

1. **Prior distribution**: They assign

   \[ p(w_i) = \text{Gam}(w_i; \alpha_i, \beta_i) \]

   to each item \( i \in [n] \), where \( \alpha = \{\alpha_i\}_{i=1}^n \), \( \beta := \{\beta_i\}_{i=1}^n \in \mathbb{R}^n_{++} \).

---

Bayesian BTL Model

- Adopt the Bayesian BTL framework by Caron & Doucet\(^5\):

1. **Prior distribution:** They assign
   \[
   p(w_i) = \text{Gam}(w_i; \alpha_i, \beta_i)
   \]
to each item \(i \in [n]\), where \(\alpha = \{\alpha_i\}_{i=1}^n, \beta := \{\beta_i\}_{i=1}^n \in \mathbb{R}^{n+}\).

2. **Latent random variables:** They introduce \(Z := \{Z_{ij}\} \in \mathbb{R}^{n \times n}\)

   \[
   Z_{ij} = Z_{ji} := \sum_{s=1}^{m_{ij}} \min\{Y_{si}, Y_{sj}\},
   \]
   for \(i, j \in [n]\) such that \(i < j\), where
   \[
   Y_i \sim \text{Exp}(w_i) \quad \& \quad Y_j \sim \text{Exp}(w_j) \quad \text{such that} \quad P_{ij} = \Pr[Y_i < Y_j].
   \]

Known as Thurstonian interpretation of the BTL model.

Prior:

\[ p(w) = \prod_{i=1}^{n} p(w_i) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i), \]
Induced Probabilities by Bayesian BTL Model

1. Prior:

\[ p(w) = \prod_{i=1}^{n} p(w_i) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i), \]

2. Prior × Likelihood:

\[ p(w, B) = p(w)p(B|w) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i) \prod_{i<j} \text{Bin} \left( b_{ij}; m_{ij}, \frac{w_i}{w_i + w_j} \right). \]
Induced Probabilities by Bayesian BTL Model

1. Prior:

\[ p(w) = \prod_{i=1}^{n} p(w_i) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i), \]

2. Prior \times Likelihood:

\[ p(w, B) = p(w)p(B|w) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i) \prod_{i<j} \text{Bin} \left( b_{ij}; m_{ij}, \frac{w_i}{w_i + w_j} \right). \]

3. Latent Variable:

\[ p(Z_{ij}|w_i, w_j) = \text{Gam}(Z_{ij}; m_{ij}, w_i + w_j). \]
Induced Probabilities by Bayesian BTL Model

1. Prior:
\[
p(w) = \prod_{i=1}^{n} p(w_i) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i),
\]

2. Prior × Likelihood:
\[
p(w, B) = p(w)p(B|w) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i, \beta_i) \prod_{i<j} \text{Bin} \left( b_{ij}; m_{ij}, \frac{w_i}{w_i + w_j} \right).
\]

3. Latent Variable:
\[
p(Z_{ij}|w_i, w_j) = \text{Gam}(Z_{ij}; m_{ij}, w_i + w_j).
\]

4. Posterior:
\[
p(w|B, Z) = \prod_{i=1}^{n} \text{Gam}(w_i; \alpha_i + b_i, \beta_i + Z_i).
\]
where \( b_i := \sum_{j \neq i} b_{ij} \) and \( Z_i := \sum_{j \neq i} Z_{ij} \).
For any $r \geq 1$, we define the family of Bayes risks for estimating $\mathbf{w}$.
For any $r \geq 1$, we define the family of Bayes risks for estimating $w$ from only $B$ as

$$R_B := \inf_{\varphi} \mathbb{E} \left[ \| w - \varphi(B) \|_r^r \right],$$

where $\varphi(B)$ is an estimator of $w$. 

$R_B \geq R^*_B$. 

Vincent Y. F. Tan (NUS)
For any $r \geq 1$, we define the family of Bayes risks for estimating $w$

1 from only $B$ as

$$R_B := \inf_\varphi \mathbb{E} \left[ \|w - \varphi(B)\|_r \right],$$

where $\varphi(B)$ is an estimator of $w$.

2 from $B$ and the latent variable $Z$ as

$$R_B^* := \inf_{\varphi^*} \mathbb{E} \left[ \|w - \varphi^*(B, Z)\|_r \right],$$

where $\varphi^*(B, Z)$ is an estimator of $w$. 
For any $r \geq 1$, we define the family of Bayes risks for estimating $w$

1. from only $B$ as

$$R_B := \inf_{\varphi} \mathbb{E} \left[ \| w - \varphi(B) \|^r \right],$$

where $\varphi(B)$ is an estimator of $w$.

2. from $B$ and the latent variable $Z$ as

$$R_B^* := \inf_{\varphi^*} \mathbb{E} \left[ \| w - \varphi^*(B, Z) \|^r \right],$$

where $\varphi^*(B, Z)$ is an estimator of $w$.

$$R_B \geq R_B^*.$$
Bayesian Network of All Variables

\[ w_i \sim \text{Gam}(w_i; \alpha_i, \beta_i) \quad \text{Prior on } w_i \]
Bayesian Network of All Variables

\[ P_{ij} = \frac{w_i}{w_i + w_j} \]

BTL model
Bayesian Network of All Variables

\[ Y_{si} \sim \text{Exp}(w_i) \quad \text{Latent “Arrival Times”} \]
Bayesian Network of All Variables

\[ b_{ij} \sim \text{Bin}(b_{ij}; m_{ij}, P_{ij}) \]

Num of times \( i \) beats \( j \) out of \( m_{ij} \) games
\[ Z_{ij} = \sum_{s=1}^{m_{ij}} \min\{Y_{si}, Y_{sj}\} : \text{Latent variables} \]
Bayesian Network of All Variables

\[ \varphi(B) \quad \text{and} \quad \varphi^*(B, Z) : \quad \text{Functions to estimate} \; w \]
For $r = 2$, can compute the Bayesian Cramér-Rao bound on $R_B$. 

---

General Lower Bounds on the Bayes Risk

- For \( r = 2 \), can compute the Bayesian Cramér-Rao bound on \( R_B \).
- We compute a family of information-theoretic lower bounds:

\[ R^*_B \geq n r e \left( V_n \cdot \Gamma \left( 1 + \frac{n r}{n} \right) \right) - \frac{r}{n} \exp \left[ -r n \left( I(w; B, Z) - h(w) \right) \right], \]

where \( V_n \) is the volume of the unit ball in \((\mathbb{R}^n, \| \cdot \|_r)\).

---

For \( r = 2 \), can compute the Bayesian Cramér-Rao bound on \( R_B \).

We compute a family of information-theoretic lower bounds:

1. Theorem 3 of Xu and Raginsky\(^6\) reads: For any \( r \geq 1 \),

\[
R_B^* \geq \frac{n}{re} \left( V_n \cdot \Gamma \left( 1 + \frac{n}{r} \right) \right)^{-r/n} \exp \left[ - \frac{r}{n} \left( I(w; B, Z) - h(w) \right) \right],
\]

where \( V_n \) is the volume of the unit ball in \((\mathbb{R}^n, \| \cdot \|_r)\).

---

For $r = 2$, can compute the Bayesian Cramér-Rao bound on $R_B$.

We compute a family of information-theoretic lower bounds:

Theorem 3 of Xu and Raginsky\(^6\) reads: For any $r \geq 1$,

$$R_B^* \geq \frac{n}{re} \left( V_n \cdot \Gamma \left( 1 + \frac{n}{r} \right) \right)^{-r/n} \exp \left[- \frac{r}{n} \left( I(w; B, Z) - h(w) \right) \right],$$

where $V_n$ is the volume of the unit ball in $(\mathbb{R}^n, \|\cdot\|_r)$.

---

For $r = 2$, can compute the Bayesian Cramér-Rao bound on $R_B$. 

We compute a family of information-theoretic lower bounds:

1. Theorem 3 of Xu and Raginsky\(^6\) reads: For any $r \geq 1$,
   \[
   R_B^* \geq \frac{n}{r e} \left( V_n \cdot \Gamma \left( 1 + \frac{n}{r} \right) \right)^{-r/n} \exp \left[ - \frac{r}{n} \left( I(w; B, Z) - h(w) \right) \right],
   \]
   where $V_n$ is the volume of the unit ball in $(\mathbb{R}^n, \| \cdot \|_r)$.

2. Using Stirling’s approximation, we upper bound
   \[
   I(w; B, Z) - h(w) = \mathbb{E} \left[ \log p(w|B, Z) \right].
   \]

Theorem

For all $i \in [n]$, let

$$m_i := \frac{1}{2} \sum_{j \neq i} m_{ij}.$$  

half the total num. of games $i$ plays
Theorem

For all \( i \in [n] \), let

\[
m_i := \frac{1}{2} \sum_{j \neq i} m_{ij}.
\]

half the total num. of games \( i \) plays

Then, the Bayes risk is asymptotically lower bounded by

\[
R_{B} \gtrsim \frac{n}{re} \left( V_n \cdot \Gamma\left(1 + \frac{n}{r}\right) \right)^{-r/n} \exp \left[ - r E(B, \alpha, \beta) \right],
\]

where
Theorem

For all $i \in [n]$, let

$$m_i := \frac{1}{2} \sum_{j \neq i} m_{ij}. \quad \text{half the total num. of games } i \text{ plays}$$

Then, the Bayes risk is asymptotically lower bounded by

$$R_B \gtrsim \frac{n}{re} \left( V_n \cdot \Gamma \left( 1 + \frac{n}{r} \right) \right)^{-r/n} \exp \left[ -r E(B, \alpha, \beta) \right],$$

where

$$E(B, \alpha, \beta) := \sum_{i=1}^{n} \left( -\frac{1}{2} \log (2\pi) + \log \beta_i - \psi(\alpha_i) + \frac{1}{2} \log (\alpha_i + m_i) \right).$$
Take $\alpha_i = \alpha$ and $\beta_i = \beta$, for each $i \in [n]$. 
Take $\alpha_i = \alpha$ and $\beta_i = \beta$, for each $i \in [n]$.

For the $L^1$ norm ($r = 1$),

$$R_B^* \gtrsim \sqrt{\frac{\pi}{2}} \exp \left[ - (\log \beta - \psi(\alpha) + 1) \right] \frac{n}{\sqrt{\alpha/n + m}},$$
Information-Theoretic Lower Bounds

- Take $\alpha_i = \alpha$ and $\beta_i = \beta$, for each $i \in [n]$.

- For the $L^1$ norm ($r = 1$),

\[ R_B^* \gtrsim \sqrt{\frac{\pi}{2}} \exp \left[ - (\log \beta - \psi(\alpha) + 1) \right] \frac{n}{\sqrt{\alpha/n + m}}, \]

- For the squared $L^2$ norm ($r = 2$),

\[ R_B \gtrsim \exp \left[ - 2(\log \beta - \psi(\alpha)) - 1 \right] \frac{n}{\alpha/n + m}. \]
Performance of Lower Bounds: $L^1$ error

Figure: $L^1$ error of the EM algo. and the information-theoretic lower bound (for $n = 100$, $\alpha = 5$ and $\beta = \alpha n - 1$).
Figure: $L^2$ error of the EM algo., the IT lower bound and the BCRB (for $n = 100$, $\alpha = 5$, and $\beta = \alpha n - 1$).
Given a fixed budget of $m = \sum_{i \neq j} m_{ij}$ games, how to allocate games among $n$ players to minimize the bounds?
Given a fixed budget of \( m = \sum_{i \neq j} m_{ij} \) games, how to allocate games among \( n \) players to minimize the bounds?

**Corollary (Optimal Connected Graphs)**

*Regular Connected Graphs are Optimal!*
**Proof:**

- Minimizing the lower bound is equivalent to maximizing

\[
 f(\{m_i\}_{i \in [n]}):= \sum_{i=1}^{n} \frac{1}{2} \log (\alpha_i + m_i)
\]

subject to \(\sum_{i=1}^{n} m_i = m\) and \(m_i \in \mathbb{N}\).
Application to General Comparison Graphs

Proof:

- Minimizing the lower bound is equivalent to maximizing
  \[ f(\{m_i\}_{i\in[n]}) := \sum_{i=1}^{n} \frac{1}{2} \log (\alpha_i + m_i) \]

  subject to \( \sum_{i=1}^{n} m_i = m \) and \( m_i \in \mathbb{N} \).

- Solution given by water-filling formula:
  \[ m_i = |\mu - \alpha_i|_+, \quad \forall \ i \in [n], \]

  where \( \mu > 0 \) is chosen such that
  \[ \sum_{i=1}^{n} |\mu - \alpha_i|_+ = m. \]
Proof:

- Minimizing the lower bound is equivalent to maximizing

\[ f(\{m_i\}_{i \in [n]}) := \sum_{i=1}^{n} \frac{1}{2} \log (\alpha_i + m_i) \]

subject to \( \sum_{i=1}^{n} m_i = m \) and \( m_i \in \mathbb{N} \).

- Solution given by water-filling formula:

\[ m_i = |\mu - \alpha_i|_+, \quad \forall \ i \in [n], \]

where \( \mu > 0 \) is chosen such that

\[ \sum_{i=1}^{n} |\mu - \alpha_i|_+ = m. \]

- But when \( \alpha_i = \alpha \) for all \( i \), \( m_i \) are all equal.
The Gamma Distribution with Fixed $\beta = 1$

$\alpha_i \uparrow \implies$ Greater belief that $w_i \uparrow$

$\implies$ Games $i$ plays with others $m_i \downarrow$
Corollary (Optimal Tree Graphs)

Best: Minimizes the (lower bound on the) Bayes Risk

Worst: Maximizes the (lower bound on the) Bayes Risk
Corollary (Optimal Tree Graphs)

Chain is best!

Star is worst!
Corollary (Optimal Tree Graphs)

1. Best: Minimizes the (lower bound on the) Bayes Risk
2. Worst: Maximizes the (lower bound on the) Bayes Risk
Proof for Star:

- Maximizing the lower bound on Bayes risk equivalent to

\[
\min_{\mathbf{m}: \sum_i m_i = m} g(\mathbf{m}) := \frac{1}{2} \log \left( \alpha + 2m + \sum_{i' \neq i^*} m_i' \right) + \sum_i \frac{1}{2} \log (\alpha + m_i)
\]

where \( \mathbf{m} = \{m_i\}_{i \in [n]} \) and \( i^* = 1 \) is the central node.
Proof for Star:

- Maximizing the lower bound on Bayes risk equivalent to

$$\min_{m: \sum_i m_i = m} g(m) := \frac{1}{2} \log \left( \frac{\alpha + \sum_{i' \neq i*} m_{i'}}{\sum_i \frac{1}{2} \log (\alpha + m_i)} \right)$$

where $m = \{m_i\}_{i \in [n]}$ and $i^* = 1$ is the central node.

- Shift part of weight of an edge $m_{1j} > 0$, for $j \neq 1$, to create a new edge with weight $m_{ji}$ such that $i \neq 1$. Can show that

$$\frac{\partial g(m_1, \ldots, m_n)}{\partial m_i} > 0$$

implying that $f$ will be increased by the new configuration.
Figure: IT bounds for diff. graph structures (for $n = 100$, $\alpha = 5$, $\beta = \alpha n - 1$).
Effect of Tree Graph Structure on BCRB

Figure: BCRB for diff. graph structures (for $n = 100$, $\alpha = 5$, $\beta = \alpha n - 1$).
Final Remarks

- Also derived lower bounds for the **home-field advantage** scenario:

\[ P_{ij} = \begin{cases} 
Q_{ij} := \frac{\theta \lambda_i}{\theta \lambda_i + \lambda_j}, & \text{if } i \text{ is home}, \\
\bar{Q}_{ij} := \frac{\lambda_i}{\lambda_i + \theta \lambda_j}, & \text{if } j \text{ is home}, 
\end{cases} \]

where \( \theta \in \mathbb{R}_{++} \) models the strength of advantage (\( \theta > 1 \))
Final Remarks

- Also derived lower bounds for the **home-field advantage** scenario:

\[
P_{ij} = \begin{cases} 
Q_{ij} := \frac{\theta \lambda_i}{\theta \lambda_i + \lambda_j}, & \text{if } i \text{ is home}, \\
\overline{Q}_{ij} := \frac{\lambda_i}{\lambda_i + \theta \lambda_j}, & \text{if } j \text{ is home},
\end{cases}
\]

where \( \theta \in \mathbb{R}_{++} \) models the strength of advantage (\( \theta > 1 \))

- Future works: Matching information-theoretic **upper bounds**
Also derived lower bounds for the home-field advantage scenario:

\[
P_{ij} = \begin{cases} 
Q_{ij} := \frac{\theta \lambda_i}{\theta \lambda_i + \lambda_j}, & \text{if } i \text{ is home,} \\
\overline{Q}_{ij} := \frac{\lambda_i}{\lambda_i + \theta \lambda_j}, & \text{if } j \text{ is home,}
\end{cases}
\]

where \( \theta \in \mathbb{R}_{++} \) models the strength of advantage (\( \theta > 1 \))

Future works: Matching information-theoretic upper bounds

Other questions related to comparing graph structure, e.g.,
Also derived lower bounds for the home-field advantage scenario:

\[ P_{ij} = \begin{cases} 
Q_{ij} := \frac{\theta \lambda_i}{\theta \lambda_i + \lambda_j}, & \text{if } i \text{ is home,} \\
\bar{Q}_{ij} := \frac{\lambda_i}{\lambda_i + \theta \lambda_j}, & \text{if } j \text{ is home,}
\end{cases} \]

where \( \theta \in \mathbb{R}_{++} \) models the strength of advantage (\( \theta > 1 \))

Future works: Matching information-theoretic upper bounds

Other questions related to comparing graph structure, e.g.,

"Does the fully-connected graph outperform a simple cycle?"
Final Remarks

- Also derived lower bounds for the **home-field advantage** scenario:

\[
P_{ij} = \begin{cases} 
Q_{ij} := \frac{\theta \lambda_i}{\theta \lambda_i + \lambda_j}, & \text{if } i \text{ is home}, \\
\overline{Q}_{ij} := \frac{\lambda_i}{\lambda_i + \theta \lambda_j}, & \text{if } j \text{ is home},
\end{cases}
\]

where \( \theta \in \mathbb{R}_{++} \) models the strength of advantage (\( \theta > 1 \))

- Future works: Matching information-theoretic **upper bounds**

- Other questions related to comparing **graph structure**, e.g.,

  “Does the fully-connected graph outperform a simple cycle?”