Ranking the Rows of a Permuted Isotonic Matrix in Noise

Emmanuel Pilliat – ENSAI

Presented at SODA24 [https://arxiv.org/abs/2310.01133],
Joint work with

Alexandra Carpentier – Uni Potsdam, Germany
Nicolas Verzelen – INRAE Montpellier, France

Motivations for Ranking

Crowdsourcing data

Tournaments

Motivations for Ranking

Crowdsourcing data

- Correctness of answer for pairs of expert/questions

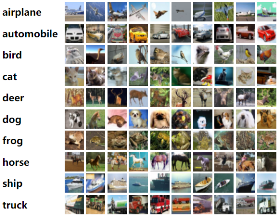
Tournaments

Motivations for Ranking

Crowdsourcing data

- Correctness of answer for pairs of expert/questions

CIFAR10H Dataset



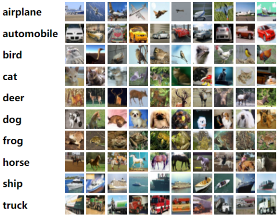
Tournaments

Motivations for Ranking

Crowdsourcing data

- Correctness of answer for pairs of expert/questions

CIFAR10H Dataset



Tournaments

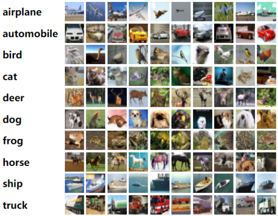
- Pairwise comparison between players

Motivations for Ranking

Crowdsourcing data

- Correctness of answer for pairs of expert/questions

CIFAR10H Dataset



Tournaments

- Pairwise comparison between players

Sports and video games

e.g. [Cattelan et al., 2013]

General Question

Given the correctness of answer from n experts to d questions.

- How accurately can we recover the **ranking of the experts?**

Illustration

Illustration

9 questions

$$
\text{4 experts}\quad \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}
$$

1: Correct answer   0: Wrong answer

Illustration

9 questions

4 experts

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

1: Correct answer   0: Wrong answer

Good Experts   **Bad Experts**

Illustration

9 questions

4 experts

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

1: Correct answer 0: Wrong answer

Good Experts **Bad Experts**

Illustration

9 questions

4 experts
$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

1: Correct answer   0: Wrong answer

Good Experts   **Bad Experts**

Goal: **Ranking of Experts**

1: Correct   0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

1: Correct   0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

Experts $i \in \{1, \ldots, n\}$

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

Experts $i \in \{1, \ldots, n\}$
Questions $k \in \{1, \ldots, d\}$

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

Experts $i \in \{1, \ldots, n\}$
Questions $k \in \{1, \ldots, d\}$
We observe for all i, k:

$$Y_{ik} \sim \text{Bern}\left(M_{ik}\right)$$

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

Experts $i \in \{1, \ldots, n\}$
Questions $k \in \{1, \ldots, d\}$
We observe for all i, k:

$$Y_{ik} \sim \mathrm{Bern}\left(M_{ik}\right)$$

- $M_{ik} = 1/2$: random choice of expert i at question k

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

Experts $i \in \{1, \ldots, n\}$
Questions $k \in \{1, \ldots, d\}$
We observe for all i, k:

$$Y_{ik} \sim \text{Bern}\,(M_{ik})$$

- $M_{ik} = 1/2$: random choice of expert i at question k
- $M_{ik} = 1$: Expert i knows perfectly the answer of question k

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

Experts/Questions Setting in Crowdsourcing

Experts $i \in \{1, \dots, n\}$
Questions $k \in \{1, \dots, d\}$
We observe for all i, k:

$$Y_{ik} \sim \text{Bern}\,(M_{ik})$$

- $M_{ik} = 1/2$: random choice of expert i at question k
- $M_{ik} = 1$: Expert i knows perfectly the answer of question k

1: Correct 0: Wrong

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

expert i is correct at question k

$$\Leftrightarrow Y_{ik} = 1$$

We will often assume that $n = d$

Pairwise Comparison Setting in Tournaments

Pairwise Comparison Setting in Tournaments

Player $i \in \{1, \ldots, n\}$
Player $k \in \{1, \ldots, n\}$
We observe for all i, k:

$$Y_{ik} = 1 - Y_{ki} \sim \mathrm{Bern}\left(M_{ik}\right)$$

1: Wins 0: Loses

$$\begin{pmatrix} \times & 1 & 1 \\ 0 & \times & 0 \\ 0 & 1 & \times \end{pmatrix}$$

3 players

Player i wins against player k

$$\Leftrightarrow Y_{ik} = 1$$

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$, with $M \in [0,1]^{n\times d}$

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- $\mathrm{Poisson}(\lambda)$ observations per entry (i, k)

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(λ) observations per entry (i,k)

- $\lambda \leq 1$: partial observations
- For simplicity, we assume that $\lambda = 1$

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(λ) observations per entry (i, k)

- $\lambda \leq 1$: partial observations
- For simplicity, we assume that $\lambda = 1$

Isotonic Model: [Flammarion et al., 2019]

$\exists$ unknown permutation π^*: l

M_{π^*} is isotonic

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n\times d}$

- Independent observations
- Poisson(λ) observations per entry (i,k)

- $\lambda \leq 1$: partial observations
- For simplicity, we assume that $\lambda = 1$

Isotonic Model: [Flammarion et al., 2019]

$\exists$ unknown permutation π^*: l

M_{π^*} is isotonic

$$M_{\pi^*} = \begin{pmatrix} 0.9 & 0.8 & 0.9 & 1 \\ 0.8 & 0.7 & 0.9 & 0.8 \\ 0.6 & 0.7 & 0.7 & 0.6 \\ 0.5 & 0.7 & 0.5 & 0.6 \end{pmatrix}$$

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(λ) observations per entry (i, k)

- $\lambda \leq 1$: partial observations
- For simplicity, we assume that $\lambda = 1$

Isotonic Model: [Flammarion et al., 2019]

$\exists$ unknown permutation π^*: l

M_{π^*} is isotonic

$$M_{\pi^*} = \begin{pmatrix} 0.9 & 0.8 & 0.9 & 1 \\ 0.8 & 0.7 & 0.9 & 0.8 \\ 0.6 & 0.7 & 0.7 & 0.6 \\ 0.5 & 0.7 & 0.5 & 0.6 \end{pmatrix}$$

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(λ) observations per entry (i,k)

- $\lambda \leq 1$: partial observations
- For simplicity, we assume that $\lambda = 1$

Isotonic Model: [Flammarion et al., 2019]

$\exists$ unknown permutation π^*: l

M_{π^*} is isotonic

$$M = \begin{pmatrix} 0.6 & 0.7 & 0.7 & 0.6 \\ 0.8 & 0.7 & 0.9 & 0.8 \\ 0.5 & 0.7 & 0.5 & 0.6 \\ 0.9 & 0.8 & 0.9 & 1 \end{pmatrix}$$

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(λ) observations per entry (i,k)

- $\lambda \leq 1$: partial observations
- For simplicity, we assume that $\lambda = 1$

Isotonic Model: [Flammarion et al., 2019]

$\exists$ unknown permutation π^*: 1

M_{π^*} is isotonic

$$M = \begin{pmatrix} 0.6 & 0.7 & 0.7 & 0.6 \\ 0.8 & 0.7 & 0.9 & 0.8 \\ 0.5 & 0.7 & 0.5 & 0.6 \\ 0.9 & 0.8 & 0.9 & 1 \end{pmatrix}$$

- Goal: recover π^*, the ranking of the experts/rows

Isotonic model

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape constraints:

- Increasing Columns **for an unknown permutation** π^*

Isotonic model

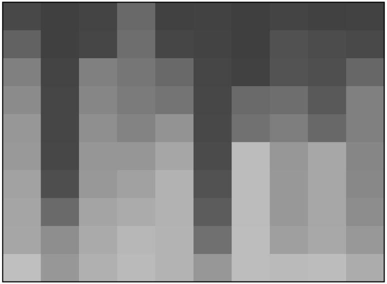
Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i,k)

Shape constraints:

- Increasing Columns **for an unknown permutation** π^*



Matrix M_{π^*}. (isotonic)

Isotonic model

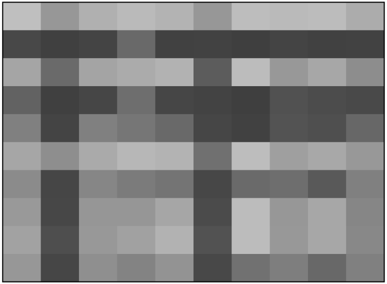
Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape constraints:

- Increasing Columns **for an unknown permutation** π^*



Matrix M (isotonic up to a permutation of rows)

Isotonic model

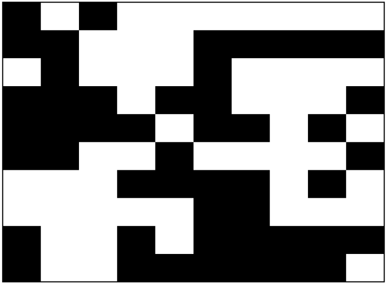
Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape constraints:

- Increasing Columns **for an unknown permutation** π^*



Matrix Y (M in noise)

Isotonic model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

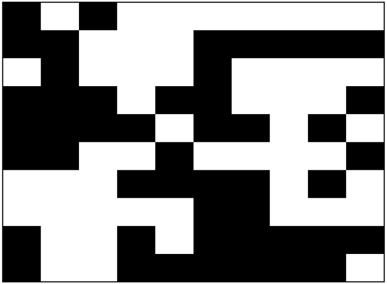
- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape constraints:

- Increasing Columns **for an unknown permutation** π^*

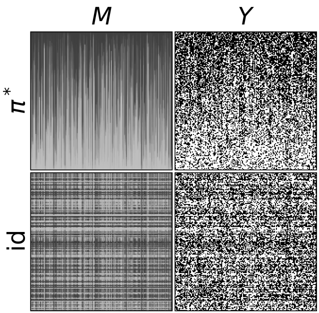
Aim

Estimation of π^*



Matrix Y (M in noise)

Illustration



Other Models

Parametric Models

Non-Parametric Models

Other Models

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$ ~[Bradley and Terry, 1952]
  - a_i: abilities of the experts
  - b_k: difficulties of the questions

Non-Parametric Models

Other Models

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$ ~[Bradley and Terry, 1952]
  - a_i: abilities of the experts
  - b_k: difficulties of the questions

Non-Parametric Models

- Isotonic: M_{π^*} is isotonic
- Bi-Isotonic: $M_{\pi^*\eta^*}$ is bi-isotonic

  ~[Shah et al., 2016],[Mao et al., 2018],[Liu and Moitra, 2020]

Other Models

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$ ∼[Bradley and Terry, 1952]
  - a_i: abilities of the experts
  - b_k: difficulties of the questions

Non-Parametric Models

- Isotonic: M_{π^*} is isotonic
- Bi-Isotonic: $M_{\pi^* \eta^*}$ is bi-isotonic
  ∼[Shah et al., 2016],[Mao et al., 2018],[Liu and Moitra, 2020]
  - The rows are increasing up to an unknown permutation η^*

Other Models

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$ ∼[Bradley and Terry, 1952]
  - a_i: abilities of the experts
  - b_k: difficulties of the questions

Non-Parametric Models

- Isotonic: M_{π^*} is isotonic
- Bi-Isotonic: $M_{\pi^* \eta^*}$ is bi-isotonic
  ∼[Shah et al., 2016],[Mao et al., 2018],[Liu and Moitra, 2020]
  - The rows are increasing up to an unknown permutation η^*
  - η^* represents a ranking of the questions by difficulty

Other Models

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$ ~[Bradley and Terry, 1952]
  - a_i: abilities of the experts
  - b_k: difficulties of the questions

Non-Parametric Models

- Isotonic: M_{π^*} is isotonic
- Bi-Isotonic: $M_{\pi^* \eta^*}$ is bi-isotonic
  ~[Shah et al., 2016],[Mao et al., 2018],[Liu and Moitra, 2020]
  - The rows are increasing up to an unknown permutation η^*
  - η^* represents a ranking of the questions by difficulty

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i,k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation Loss

For an estimator $\hat{\pi}$ of π^*

$$\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

$\|A\|_F^2 = \sum_{i,k} A_{ik}^2$

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation Loss

For an estimator $\hat{\pi}$ of π^*

$$\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

$\|A\|_F^2 = \sum_{i,k} A_{ik}^2$

$$\frac{1}{2} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & h & h & h \end{pmatrix}$$

If the two lines are misclassified:
$\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2 = 6h^2$

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation Loss

For an estimator $\hat{\pi}$ of π^*

$$\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

Reconstruction Loss

For an estimator $\hat{M}$ of M

$$\|\hat{M} - M\|_F^2$$

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation Risk

For an estimator $\hat{\pi}$ of π^*

$$\mathbb{E}\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

Reconstruction Risk

For an estimator $\hat{M}$ of M

$$\mathbb{E}\|\hat{M} - M\|_F^2$$

Isotonic Model

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation Risk

For an estimator $\hat{\pi}$ of π^*

$$\mathbb{E}\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

Reconstruction Risk

For an estimator $\hat{M}$ of M

$$\mathbb{E}\|\hat{M} - M\|_F^2$$

Aim

Estimation of π^*

Isotonic Model

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$

- Independent obs.
- Poisson(1) obs. per entry (i, k)

Shape Constraints:

- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation Risk

For an estimator $\hat{\pi}$ of π^*

$$\mathbb{E}\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

Reconstruction Risk

For an estimator $\hat{M}$ of M

$$\mathbb{E}\|\hat{M} - M\|_F^2$$

Aim

Estimation of π^*

Max Risks

Max Risks

Permuation Estimation:

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d) = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Max Risks

Permuation Estimation:

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d) = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

- **Supremum** over all isotonic matrices $M_{\pi^*} \in [0,1]^{n\times d}$

Max Risks

Permuation Estimation:

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d) = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

- **Supremum** over all isotonic matrices $M_{\pi^*} \in [0,1]^{n \times d}$
- $\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d,\lambda)$ when partial observations

Max Risks

Permuation Estimation:

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d) = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Matrix Reconstruction:

$$\mathcal{R}^{\hat{M}}_{\text{reco}}(n,d) = \sup_{M,\pi^*} \mathbb{E}[\|\hat{M} - M\|_F^2]$$

- **Supremum** over all isotonic matrices $M_{\pi^*} \in [0,1]^{n\times d}$
- $\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d,\lambda)$ when partial observations

Minimax Risks

Permuation Estimation:

Matrix Reconstruction:

- **Supremum** over all isotonic matrices $M_{\pi^*} \in [0,1]^{n\times d}$
- $\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d,\lambda)$ when partial observations

Minimax Risks

Permuation Estimation:

$$\mathcal{R}^*_{\text{perm}}(n,d) =$$

Matrix Reconstruction:

$$\mathcal{R}^*_{\text{reco}}(n,d) =$$

- **Supremum** over all isotonic matrices $M_{\pi^*} \in [0,1]^{n\times d}$
- $\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d,\lambda)$ when partial observations

Minimax Risks

Permuation Estimation:

$$\mathcal{R}^*_{\text{perm}}(n,d) = \inf_{\hat{\pi}} \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Matrix Reconstruction:

$$\mathcal{R}^*_{\text{reco}}(n,d) = \inf_{\hat{M}} \sup_{M,\pi^*} \mathbb{E}[\|\hat{M} - M\|_F^2]$$

- **Supremum** over all isotonic matrices $M_{\pi^*} \in [0,1]^{n\times d}$
- $\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,d,\lambda)$ when partial observations

Main Questions

Permuation Estimation:

$$\mathcal{R}^*_{\text{perm}}(n, d) = \inf_{\hat{\pi}} \sup_{M, \pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Matrix Reconstruction:

$$\mathcal{R}^*_{\text{reco}}(n, d) = \inf_{\hat{M}} \sup_{M, \pi^*} \mathbb{E}[\|\hat{M} - M\|_F^2]$$

Main Questions

Permuation Estimation:

$$\mathcal{R}^*_{\text{perm}}(n,d) = \inf_{\hat{\pi}} \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Matrix Reconstruction:

$$\mathcal{R}^*_{\text{reco}}(n,d) = \inf_{\hat{M}} \sup_{M,\pi^*} \mathbb{E}[\|\hat{M} - M\|_F^2]$$

- Is there a **computational-statistical gap**?

Main Questions

Permuation Estimation:

$$\mathcal{R}^*_{\text{perm}}(n,d) = \inf_{\hat{\pi}} \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Matrix Reconstruction:

$$\mathcal{R}^*_{\text{reco}}(n,d) = \inf_{\hat{M}} \sup_{M,\pi^*} \mathbb{E}[\|\hat{M} - M\|_F^2]$$

- Is there a **computational-statistical gap**?
- Is estimating π^* easier than reconstructing M?

Main Questions

Permuation Estimation:

$$\mathcal{R}^*_{\text{perm}}(n,d) = \inf_{\hat{\pi}} \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Matrix Reconstruction:

$$\mathcal{R}^*_{\text{reco}}(n,d) = \inf_{\hat{M}} \sup_{M,\pi^*} \mathbb{E}[\|\hat{M} - M\|_F^2]$$

- Is there a **computational-statistical gap**?
- Is estimating π^* easier than reconstructing M? ($\mathcal{R}^*_{\text{perm}} \ll \mathcal{R}^*_{\text{reco}}$?)

Short Story

Parametric Models

Non-Parametric Models

Short Story

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$

Non-Parametric Models

- No computational gap for **parametric models**

  e.g. [Chen et al., 2022]

Short Story

Parametric Models

- BTL: $M_{ik} = \phi(a_i - b_k)$

Non-Parametric Models

- Isotonic: M_{π^*} isotonic
- Bi-Isotonic: $M_{\pi^* \eta^*}$ bi-isotonic

- No computational gap for **parametric models**

  e.g. [Chen et al., 2022]

- Mostly unknown for **non-parametric models**: computational gaps were conjectured

  e.g. [Flammarion et al., 2019, Mao et al., 2018]

Contributions

Overview

Contributions

Overview

- There is no significant computational gap

Contributions

Overview

- There is no significant computational gap
- Much easier to estimate π^* than M in many regimes

Contributions

Overview

- There is no significant computational gap
- Much easier to estimate π^* than M in many regimes

Corollary for the bi-isotonic model ($M_{\pi^*\eta^*}$ bi-isotonic):

Contributions

Overview

- There is no significant computational gap
- Much easier to estimate π^* than M in many regimes

Corollary for the bi-isotonic model ($M_{\pi^*\eta^*}$ bi-isotonic):

- Poly. time algo achieves better rates than state of the art [Mao et al., 2018, Liu and Moitra, 2020]

Existing Methods

Existing Methods

- Simple Global Average Comparison

Existing Methods

- Simple Global Average Comparison
- Least-Square

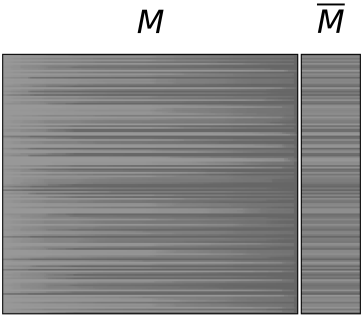
Existing Methods

- Simple Global Average Comparison
- Least-Square
- Poly. Time Methods only in the Bi-isotonic Model

Existing Methods

- Simple Global Average Comparison
- Least-Square
- Poly. Time Methods only in the Bi-isotonic Model

Global Average Comparison

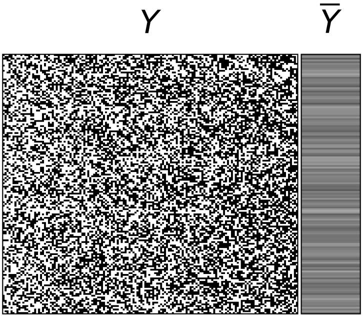


Matrix M

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

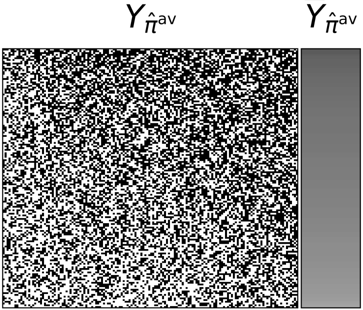


Matrix Y (M in noise).

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$



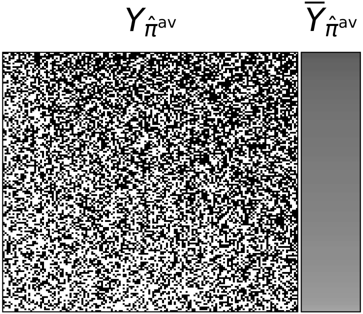
Matrix $Y_{\hat{\pi}^{\mathrm{av}}}$ (M in noise).

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$



Matrix $Y_{\hat{\pi}^{\text{av}}}$ (M in noise).

Global Average Comparison

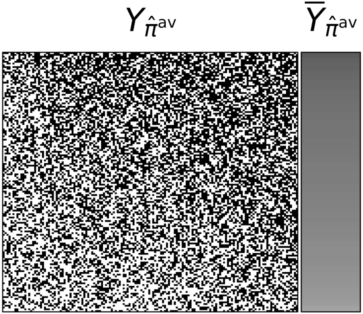
- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}^{\hat\pi}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat\pi} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat\pi^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}^{\hat\pi^{\text{av}}}_{\text{perm}}(n,n) \in [\log^c(n)n^{3/2}, \log^C(n)n^{3/2}]$$



Matrix $Y_{\hat\pi^{\text{av}}}$ (M in noise).

Global Average Comparison

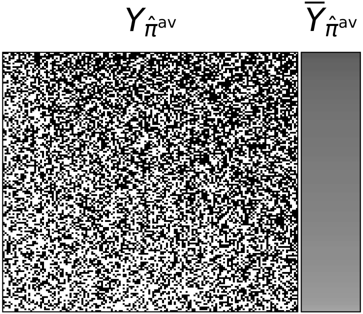
- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}^{\hat{\pi}^{\text{av}}}_{\text{perm}}(n,n) \asymp n^{3/2}$$



Matrix $Y_{\hat{\pi}^{\text{av}}}$ (M in noise).

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}^{\text{av}}}(n,n) \asymp n^{3/2}$$

Lower Bound for Two Rows

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}^{\hat{\pi}^{\text{av}}}_{\text{perm}}(n,n) \asymp n^{3/2}$$

Lower Bound for Two Rows

$$M_{1,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2} \underbrace{\tfrac{1}{2} \dots \tfrac{1}{2}})$$

$$M_{2,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2} \underbrace{1 \dots 1}_{\sim\sqrt{d}})$$

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}^{\hat{\pi}^{\text{av}}}_{\text{perm}}(n,n) \asymp n^{3/2}$$

Lower Bound for Two Rows

$$M_{1,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2}\, \underbrace{\tfrac{1}{2} \dots \tfrac{1}{2}})$$

$$M_{2,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2}\, \underbrace{1 \dots 1}_{\sim\sqrt{d}})$$

- 1 and 2 are likely to be inversed by π^{av}

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}_{\text{perm}}^{\hat\pi} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat\pi} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat\pi^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}_{\text{perm}}^{\hat\pi^{\text{av}}}(n,n) \asymp n^{3/2}$$

Lower Bound for Two Rows

$$M_{1,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2}\ \underbrace{\tfrac{1}{2} \dots \tfrac{1}{2}})$$

$$M_{2,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2}\ \underbrace{1 \dots 1}_{\sim\sqrt{d}})$$

- 1 and 2 are likely to be inversed by π^{av}
- $\|M_{1,\cdot} - M_{2,\cdot}\|_F^2 \asymp \sqrt{d}$

Global Average Comparison

- Rank according to the averages:

$$\overline{Y}_i = \frac{1}{d}\sum_{k=1}^{d} Y_{ik}$$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}^{\text{av}}$ [Shah et al., 2016]

$$\mathcal{R}^{\hat{\pi}^{\text{av}}}_{\text{perm}}(n,n) \asymp n^{3/2}$$

Lower Bound for Two Rows

$$M_{1,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2}\ \underbrace{\tfrac{1}{2} \dots \tfrac{1}{2}})$$

$$M_{2,\cdot} = (\tfrac{1}{2} \dots \tfrac{1}{2}\ \underbrace{1 \dots 1}_{\sim\sqrt{d}})$$

- 1 and 2 are likely to be inversed by π^{av}
- $\|M_{1,\cdot} - M_{2,\cdot}\|_F^2 \asymp \sqrt{d}$
- $\mathcal{R}^{\hat{\pi}^{\text{av}}}_{\text{perm}} \gtrsim n\sqrt{d}$

Existing Methods

- Simple Global Average Comparison
- Least-Square
- Poly. Time Methods only in the Bi-isotonic Model

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

$$\hat{\pi} = \arg\min_{\pi} \; d_{\text{iso}}(Y_{\pi^{-1}}) \; ,$$

d_{iso}: distance to isotonic matrices in Frobenius norm.

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

$$\hat{\pi} = \arg\min_{\pi} \; d_{\text{iso}}(Y_{\pi^{-1}}) \; ,$$

d_{iso}: distance to isotonic matrices in Frobenius norm.

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

$$\hat{\pi} = \arg\min_{\pi} \; d_{\mathrm{iso}}(Y_{\pi^{-1}}) \; ,$$

d_{iso}: distance to isotonic matrices in Frobenius norm.

$$\mathcal{R}^{\hat{\pi}}_{\mathrm{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}$

$$\mathcal{R}^{\hat{\pi}}_{\mathrm{perm}}(n, n) \lesssim n^{4/3}$$

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

$$\hat{\pi} = \arg\min_{\pi} \; d_{\text{iso}}(Y_{\pi^{-1}}) \;\; ,$$

d_{iso}: distance to isotonic matrices in Frobenius norm.

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,n) \lesssim n^{4/3}$$

- Better than global average

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

$$\hat{\pi} = \arg\min_{\pi} \; d_{\text{iso}}(Y_{\pi^{-1}}) \;\;,$$

d_{iso}: distance to isotonic matrices in Frobenius norm.

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,n) \lesssim n^{4/3}$$

- Better than global average
- Not poly. time

Least-Square [Flammarion et al., 2019]

Brute-Force all $n!$ permutations π until $Y_{\pi^{-1}}$ is close to an isotonic matrix:

$$\hat{\pi} = \arg\min_{\pi} \; d_{\text{iso}}(Y_{\pi^{-1}}) \;\; ,$$

d_{iso}: distance to isotonic matrices in Frobenius norm.

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee on $\hat{\pi}$

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n,n) \lesssim n^{4/3}$$

- Better than global average
- Not poly. time
- $n^{4/3}$ still not optimal

Existing Methods

- Simple Global Average Comparison
- Least-Square
- Poly. Time Methods only in the Bi-isotonic Model

Polynomial Time Method in The Bi-Isotonic Model

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection
  [Liu and Moitra, 2020]

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection
  [Liu and Moitra, 2020]
- Hierarchical clustering
  [Liu and Moitra, 2020]

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection [Liu and Moitra, 2020]
- Hierarchical clustering [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection [Liu and Moitra, 2020]
- Hierarchical clustering [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee in bi-isotonic model when $n = d$ [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}}(n, n) \leq n^{7/6+o(1)}$$

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection [Liu and Moitra, 2020]
- Hierarchical clustering [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee in bi-isotonic model when $n = d$ [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}}(n, n) \leq n^{7/6+o(1)}$$

- Not adapted to the isotonic model (Here $M_{\pi^*\eta^*}$ bi-isotonic)

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection [Liu and Moitra, 2020]
- Hierarchical clustering [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee in bi-isotonic model when $n = d$ [Liu and Moitra, 2020]

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}}(n, n) \leq n^{7/6+o(1)}$$

- Not adapted to the isotonic model (Here $M_{\pi^*\eta^*}$ bi-isotonic)
- Not optimal when $n \ll d$

Polynomial Time Method in The Bi-Isotonic Model

- Local averages [Mao et al., 2018]
- Change-Point detection [Liu and Moitra, 2020]
- Hierarchical clustering [Liu and Moitra, 2020]

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

Guarantee in bi-isotonic model when $n = d$ [Liu and Moitra, 2020]

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}}(n, n) \leq n^{7/6+o(1)}$$

- Not adapted to the isotonic model (Here $M_{\pi^*\eta^*}$ bi-isotonic)
- Not optimal when $n \ll d$
- Not optimal when we have partial observations ($\lambda \ll 1$)

Contributions

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

$$\mathcal{R}^{*}_{\text{perm}} = \inf_{\hat{\pi}} \mathcal{R}^{\hat{\pi}}_{\text{perm}}$$

Contributions

$$\mathcal{R}_{\text{perm}}^{\hat{\pi}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

$$\mathcal{R}_{\text{perm}}^{*} = \inf_{\hat{\pi}} \mathcal{R}_{\text{perm}}^{\hat{\pi}}$$

Theorem

In the Isotonic model, for all n, d, λ,

Contributions

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

$$\mathcal{R}^{*}_{\text{perm}} = \inf_{\hat{\pi}} \mathcal{R}^{\hat{\pi}}_{\text{perm}}$$

Theorem

In the Isotonic model, for all n, d, λ,

$\exists$ poly. time estimator $\hat{\pi}$ that achieves, up to a numerical constant C,

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} \leq \left[\log^C\left(\frac{nd}{\lambda}\right)\right] \mathcal{R}^{*}_{\text{perm}}$$

Contributions

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

$$\mathcal{R}^{*}_{\text{perm}} = \inf_{\hat{\pi}} \mathcal{R}^{\hat{\pi}}_{\text{perm}}$$

Theorem

In the Isotonic model, for all n, d, λ,

$\exists$ poly. time estimator $\hat{\pi}$ that achieves, up to a numerical constant C,

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} \leq \left[\log^C\left(\frac{nd}{\lambda}\right)\right] \mathcal{R}^{*}_{\text{perm}}$$

When $\lambda = 1$ and $n \leq d$, up to polylogs:

$$\mathcal{R}^{*}_{\text{perm}} \asymp n^{2/3}\sqrt{d}$$

Contributions

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2]$$

$$\mathcal{R}^{*}_{\text{perm}} = \inf_{\hat{\pi}} \mathcal{R}^{\hat{\pi}}_{\text{perm}}$$

Theorem

In the Isotonic model, for all n, d, λ,

$\exists$ poly. time estimator $\hat{\pi}$ that achieves, up to a numerical constant C,

$$\mathcal{R}^{\hat{\pi}}_{\text{perm}} \leq \left[\log^C\left(\frac{nd}{\lambda}\right)\right] \mathcal{R}^{*}_{\text{perm}}$$

When $\lambda = 1$ and $n \leq d$, up to polylogs:

$$\mathcal{R}^{*}_{\text{perm}} \asymp n^{2/3}\sqrt{d}$$

When $n = d$, we recover the $n^{7/6}$ rate in the easier bi-isotonic model

Weighted Comparison Graph Updates

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n\times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

1. Neighborhood $G^{(0)}$ of rows around i according to $\mathcal{W}$

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

1. Neighborhood $G^{(0)}$ of rows around i according to $\mathcal{W}$
2. Subset of columns $\hat{Q} \subset [d]$

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

1. Neighborhood $G^{(0)}$ of rows around i according to $\mathcal{W}$
2. Subset of columns $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

1. Neighborhood $G^{(0)}$ of rows around i according to $\mathcal{W}$
2. Subset of columns $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

1. Neighborhood $G^{(0)}$ of rows around i according to $\mathcal{W}$
2. Subset of columns $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$, Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Weighted Comparison Graph Updates

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \quad \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

$\mathcal{W}_{ij}$ is large: confident in estimating row i above row j

For all i, compute:

1. Neighborhood $G^{(0)}$ of rows around i according to $\mathcal{W}$
2. Subset of columns $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

$$\text{If } |\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|, \text{ Update } \mathcal{W}_{ij} := \mathcal{U}_{ij}$$

- Perform a polylog number of iterations

$$\mathcal{W} \in \mathbb{R}^{n \times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

$$\mathcal{W} \in \mathbb{R}^{n\times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

Permutation Estimation

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

$$\mathcal{W} \in \mathbb{R}^{n \times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

Permutation Estimation

Thresholded Graphs

$$\mathcal{G}(\gamma) = \{(i,j)\ :\ \mathcal{W}^{\mathrm{f}}_{ij} \geq \gamma\}$$

$$\mathcal{W} \in \mathbb{R}^{n \times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

Permutation Estimation

Thresholded Graphs

$$\mathcal{G}(\gamma) = \{(i,j) \ : \ \mathcal{W}^{\mathrm{f}}_{ij} \geq \gamma\}$$

- $\mathcal{G}(\gamma)$ is $\emptyset$ if γ is large
- $\mathcal{G}(0)$ is the complete graph

$$\mathcal{W} \in \mathbb{R}^{n\times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u\rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

Permutation Estimation

Thresholded Graphs

$$\mathcal{G}(\gamma) = \{(i,j)\ :\ \mathcal{W}^{\mathrm{f}}_{ij} \geq \gamma\}$$

Minimal threshold

$$\hat{\gamma} = \inf_\gamma\{\mathcal{G}(\gamma) \text{ is acyclic}\}$$

- $\mathcal{G}(\gamma)$ is $\emptyset$ if γ is large
- $\mathcal{G}(0)$ is the complete graph

$$\mathcal{W} \in \mathbb{R}^{n \times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

Permutation Estimation

Thresholded Graphs

$$\mathcal{G}(\gamma) = \{(i,j)\ :\ \mathcal{W}^{\mathrm{f}}_{ij} \geq \gamma\}$$

Minimal threshold

$$\hat{\gamma} = \inf_\gamma \{\mathcal{G}(\gamma) \text{ is acyclic}\}$$

Estimator $\hat{\pi}$ is such that

$$(i,j) \in \mathcal{G}(\hat{\gamma}) \implies \hat{\pi}(i) > \hat{\pi}(j)$$

- $\mathcal{G}(\gamma)$ is $\emptyset$ if γ is large
- $\mathcal{G}(0)$ is the complete graph

$$\mathcal{W} \in \mathbb{R}^{n \times n}, \ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

Permutation Estimation

Thresholded Graphs

$$\mathcal{G}(\gamma) = \{(i,j) \ : \ \mathcal{W}^{\mathrm{f}}_{ij} \geq \gamma\}$$

Minimal threshold

$$\hat{\gamma} = \inf_\gamma \{\mathcal{G}(\gamma) \text{ is acyclic}\}$$

Estimator $\hat{\pi}$ is such that

$$(i,j) \in \mathcal{G}(\hat{\gamma}) \implies \hat{\pi}(i) > \hat{\pi}(j)$$

- $\mathcal{G}(\gamma)$ is $\emptyset$ if γ is large
- $\mathcal{G}(0)$ is the complete graph

Choosing $\hat{Q}$

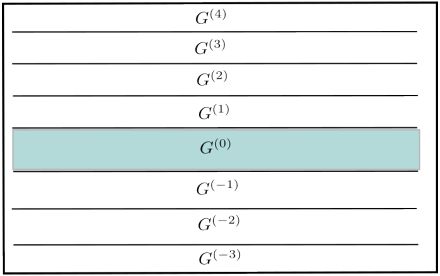
Choosing $\hat{Q}$

For a given row i, consider:

Choosing $\hat{Q}$

For a given row i, consider:

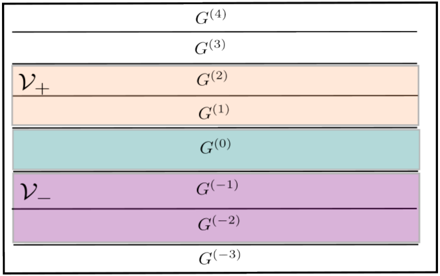
- Neighborhood $G^{(0)}$ of i



Choosing $\hat{Q}$

For a given row i, consider:

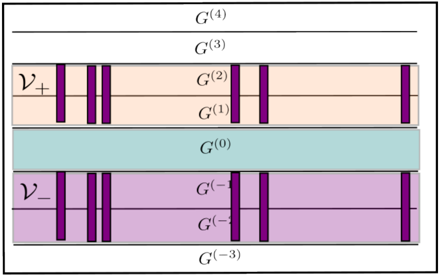
- Neighborhood $G^{(0)}$ of i
- Set $\mathcal{V}^+$ of rows j above $G^{(0)}$
- Set $\mathcal{V}^-$ of rows j below $G^{(0)}$



Choosing $\hat{Q}$

For a given row i, consider:

- Neighborhood $G^{(0)}$ of i
- Set $\mathcal{V}^+$ of rows j above $G^{(0)}$
- Set $\mathcal{V}^-$ of rows j below $G^{(0)}$



Set $\hat{Q}$

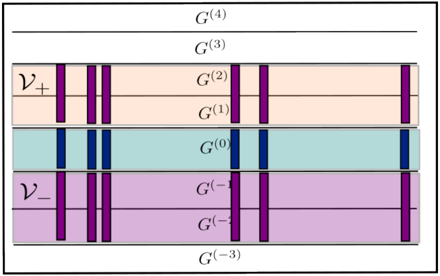
Reduce the dim. of columns to

$$\hat{Q} = \{k \ : \ \frac{1}{|\mathcal{V}^+|} \sum_{i \in \mathcal{V}^+} Y_{ik} - \frac{1}{|\mathcal{V}^-|} \sum_{i \in \mathcal{V}^-} Y_{ik} > h\}$$

Choosing $\hat{Q}$

For a given row i, consider:

- Neighborhood $G^{(0)}$ of i
- Set $\mathcal{V}^+$ of rows j above $G^{(0)}$
- Set $\mathcal{V}^-$ of rows j below $G^{(0)}$



$Y(G^{(0)}, \hat{Q})$

Reduce the dim. of columns to

$$\hat{Q} = \{k \ : \ \frac{1}{|\mathcal{V}^+|} \sum_{i \in \mathcal{V}^+} Y_{ik} - \frac{1}{|\mathcal{V}^-|} \sum_{i \in \mathcal{V}^-} Y_{ik} > h\}$$

$$\mathcal{W} \in \mathbb{R}^{n \times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

$$\mathcal{W} \in \mathbb{R}^{n \times n},\ \mathcal{W}_{ij} = -\mathcal{W}_{ji}$$

For all i, compute:

1. Neighborhood $G^{(0)}$ of i
2. Subset $\hat{Q} \subset [d]$
3. Vector $u \in \mathbb{R}_+^{\hat{Q}}$
4. $\mathcal{U}_{ij} = \langle Y_{i\cdot} - Y_{j\cdot}, u \rangle$

Spectral method on $Y(G^{(0)}, \hat{Q})$

If $|\mathcal{U}_{ij}| \geq |\mathcal{W}_{ij}|$,
Update $\mathcal{W}_{ij} := \mathcal{U}_{ij}$

Final weighted graph $\mathcal{W}^{\mathrm{f}}$

Conclusion

Isotonic Model

Conclusion

Isotonic Model

- Weak non-parametric assumptions

Conclusion

Isotonic Model

- Weak non-parametric assumptions
- Poly. time and near minimax optimal ranking method

Conclusion

Isotonic Model

- Weak non-parametric assumptions
- Poly. time and near minimax optimal ranking method
- Results adapted to rectangular cases and partial observations

Conclusion

Isotonic Model

- Weak non-parametric assumptions
- Poly. time and near minimax optimal ranking method
- Results adapted to rectangular cases and partial observations
- Can be associated with isotonic regression for optimal estimation of the whole matrix M

Conclusion

Isotonic Model

- Weak non-parametric assumptions
- Poly. time and near minimax optimal ranking method
- Results adapted to rectangular cases and partial observations
- Can be associated with isotonic regression for optimal estimation of the whole matrix M
- Improve existing poly. time rates in all the regimes in the easier bi-isotonic model

Crowdsourcing Problems with Unknown Labels

Vector of unknown labels $x^* \in \{-1, 1\}^d$

$$Y_{ik} = \begin{cases} x_k^*, & \text{with probability } M_{ik}, \\ -x_k^*, & \text{with probability } 1 - M_{ik}. \end{cases}$$

Objective: recover the vector of labels x^* and close the computational gap conjectured in [Shah et al., 2020]

Crowdsourcing Problems with Unknown Labels

Vector of unknown labels $x^* \in \{-1, 1\}^d$

$$Y_{ik} = \begin{cases} x_k^*, & \text{with probability } M_{ik}, \\ -x_k^*, & \text{with probability } 1 - M_{ik}. \end{cases}$$

Objective: recover the vector of labels x^* and close the computational gap conjectured in [Shah et al., 2020]

Partial observations:

$$Y_{ik} = \pm x_k^* \text{ with probability } \lambda.$$

Spectral Method

Simple scenario in **isotonic** model:

$$M := M(G^{(0)}, \hat{Q}) = \frac{1}{2} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & h & h & 0 & h & 0 & 0 & h \\ 0 & h & h & 0 & h & 0 & 0 & h \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & h & h & 0 & h & 0 & 0 & h \end{pmatrix}$$

- Rank 1 matrix: spectral gap if h is large.
- Compute:

$$\hat{v} = \sup_{\|v\|=1} \|v^T Y\|_2^2 \ .$$

- In the isotonic model, use that

$$\|M - \overline{M}\|_{\text{op}} \geq \frac{1}{\log^C(nd)} \|M - \overline{M}\|_F \ .$$

- Iterate a spectral method a polylogarithmic number of time.

Parametric Models

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(1) observations per entry (i, k)

Noisy Sorting:
[Braverman and Mossel, 2008]

$$|M_{ik} - \frac{1}{2}| \geq \gamma \ ,$$

For some $\gamma > 0$

- Goal: Estimate π^* such that $M_{\pi^*(i+1), \pi^*(i)} \geq \frac{1}{2} + \gamma$

BTL: [Bradley and Terry, 1952]
Unknown vector $\theta \in \mathbb{R}^n$

$$M_{ik} = \frac{e^{(\theta_i - \theta_k)}}{1 + e^{(\theta_i - \theta_k)}}$$

- θ_i: ability of player i
- Goal: estimate a ranking π^* such that $\theta_{\pi^*(1)} \leq \cdots \leq \theta_{\pi^*(n)}$

Non-Parametric Models

Observation Model

$Y_{ik} = \text{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(1) observations per entry (i, k)

SST: [Shah et al., 2016]

$\exists$ unknown permutation π^*:

$$M_{\pi^*,\pi^*} \text{ is bi-isotonic}$$

$$\text{e.g.: } \begin{pmatrix} 0.5 & 0.6 & 0.8 \\ 0.4 & 0.5 & 0.7 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

- $M_{ij} \geq 1/2$ implies $M_{ik} \geq M_{jk}$
- Goal: estimate π^*

Bi-Isotonic-2D: [Mao et al., 2018]

$\exists$ unknown permutations π^*, η^*:

$$M_{\pi^*,\eta^*} \text{ is bi-isotonic}$$

- η^* represents a ranking of the difficulty of the questions

Non-Parametric Bi-Isotonic-1D and Isotonic Models

Observation Model

$Y_{ik} = \mathrm{Bern}(M_{ik})$, with $M \in [0,1]^{n \times d}$

- Independent observations
- Poisson(1) observations per entry (i, k)

Bi-Isotonic-1D: [Mao et al., 2018]

$\exists$ unknown permutations π^*:

$$M_{\pi^*} \text{ is bi-isotonic}$$

- Corresponds to $\eta^* = \mathrm{id}$ in the bi-isotonic-2D model
- The questions are ordered by difficulty

Isotonic: [Flammarion et al., 2019]

$\exists$ unknown permutation π^*:

$$M_{\pi^*} \text{ is isotonic}$$

- No isotonicity constraint on the rows of M
- More flexible than the bi-isotonic 1D, 2D models and SST

Summary

Parametric Models

- BTL: $M_{ik} = \frac{e^{(\theta_i - \theta_k)}}{1+e^{(\theta_i - \theta_k)}}$
- Noisy Sorting: $|M_{ik} - \frac{1}{2}| \geq \gamma$

Non-Parametric Models

- SST: $M_{\pi^*\pi^*}$ is bi-isotonic
- Bi-Isotonic-2D: $M_{\pi^*\eta^*}$ is bi-isotonic
- Bi-Isotonic-1D: M_{π^*} is bi-isotonic
- Isotonic: M_{π^*} is isotonic

Sorting models by statistical difficulty:

$$\text{Bi-isotonic-1D} \prec \text{Bi-Isotonic-2D} \prec \text{Isotonic}$$

Rates in Isotonic and Bi-Isotonic-1D Models

Isotonic Model:

	$n \lesssim d^{3/2}$	$d^{3/2} \lesssim n$
$\mathcal{R}^*_{\text{perm}}$	$n^{2/3}\sqrt{d}$	n
$\mathcal{R}^*_{\text{est}}$	$n^{1/3}d$	n

Bi-Isotonic-1D Model:

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
$\mathcal{R}^*_{\text{perm}}$	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
$\mathcal{R}^*_{\text{est}}$	$nd^{1/3}$	$\sqrt{nd}$	n

References I

Bradley, R. A. and Terry, M. E. (1952).
Rank analysis of incomplete block designs: I. the method of paired comparisons.
Biometrika, 39(3/4):324–345.

Braverman, M. and Mossel, E. (2008).
Noisy sorting without resampling.
In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 268–276.

Cattelan, M., Varin, C., and Firth, D. (2013).
Dynamic bradley–terry modelling of sports tournaments.
Journal of the Royal Statistical Society Series C: Applied Statistics, 62(1):135–150.

References II

Chen, P., Gao, C., and Zhang, A. Y. (2022).
Partial recovery for top-k ranking: Optimality of mle and suboptimality of the spectral method.
The Annals of Statistics, 50(3):1618–1652.

Flammarion, N., Mao, C., and Rigollet, P. (2019).
Optimal rates of statistical seriation.

Liu, A. and Moitra, A. (2020).
Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation.
In Abernethy, J. and Agarwal, S., editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2780–2829. PMLR.

References III

Mao, C., Pananjady, A., and Wainwright, M. J. (2018).
Towards optimal estimation of bivariate isotonic matrices with unknown permutations.
arXiv preprint arXiv:1806.09544.

Shah, N., Balakrishnan, S., Guntuboyina, A., and Wainwright, M. (2016).
Stochastically transitive models for pairwise comparisons: Statistical and computational issues.
In *International Conference on Machine Learning*, pages 11–20. PMLR.

Shah, N. B., Balakrishnan, S., and Wainwright, M. J. (2020).
A permutation-based model for crowd labeling: Optimal estimation and robustness.
IEEE Transactions on Information Theory, 67(6):4162–4184.