Optimal Ranking in Crowd-Sourcing Problems

Emmanuel Pilliat
Université de Montpellier, INRAE

Observation Model

We observe the correctness of answer of n experts to d questions. If $i \in [n]$ and $k \in [d]$, expert i answers correctly to question k with **unknown probability** $M_{ik} \in [0,1]$:

$$Y_{ik} = \mathrm{Bern}(M_{ik}) \in \mathbb{R}^{n \times d} \ .$$

There exists an **unknown permutation** π^* such that M satisfies either the isotonicity or bi-isotonicity constraints after sorting its rows with π^*.

Aim: Find an estimator of π^*

$$\underbrace{\begin{pmatrix} 1&0&1&0&0&0&1&0&1&1 \\ 0&0&1&1&1&1&0&1&1&1 \\ 0&0&0&0&1&0&1&1&0&1 \\ 0&0&1&1&1&1&1&1&1&1 \end{pmatrix}}_{\text{4 experts} \times \text{10 questions}} \quad \text{Matrix } Y$$

Parametric VS Non-Parametric Models

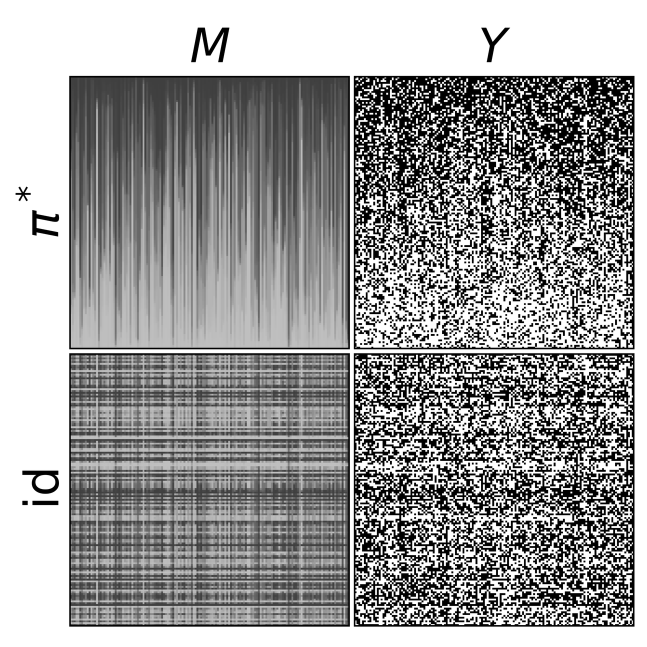
Parametric Models:

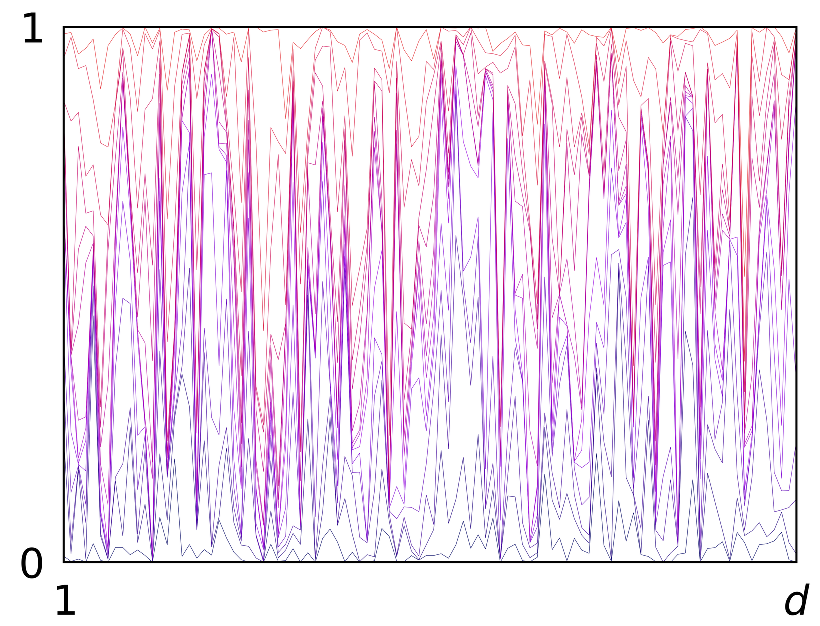
- Questions Equaly Difficult $\rightsquigarrow M_{ik} = a_i \quad \approx$ [Dawid and Skene, 1979]
- Ability/Difficulty $\rightsquigarrow M_{ik} = \phi(\alpha_i - \beta_k) \quad \approx$ [Bradeley and Terry, 1952]

Non-Parametric Models:

- **Isotonicity**: M has Increasing Columns for an **unknown permutation** $\pi^* \ \approx$ [3]
- **Bi-Isotonicity**: Isotonicity **and** M has Increasing Rows $\approx$ [4]

Isotonicity Constraints

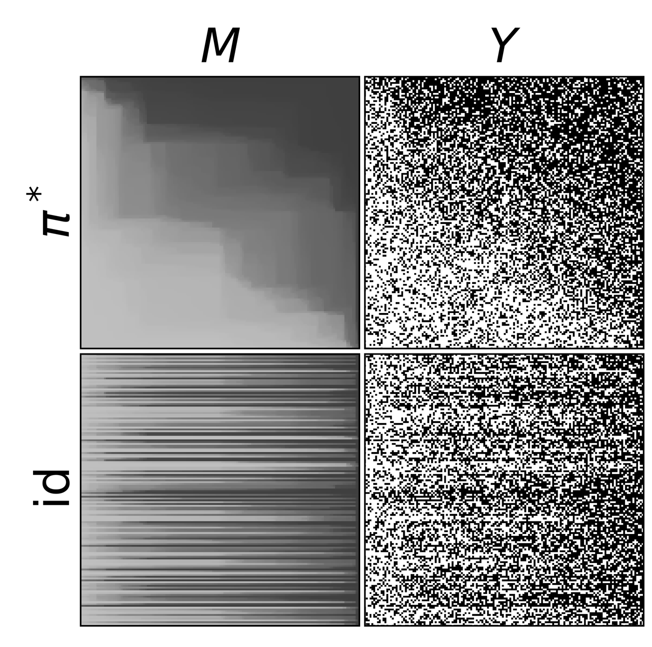


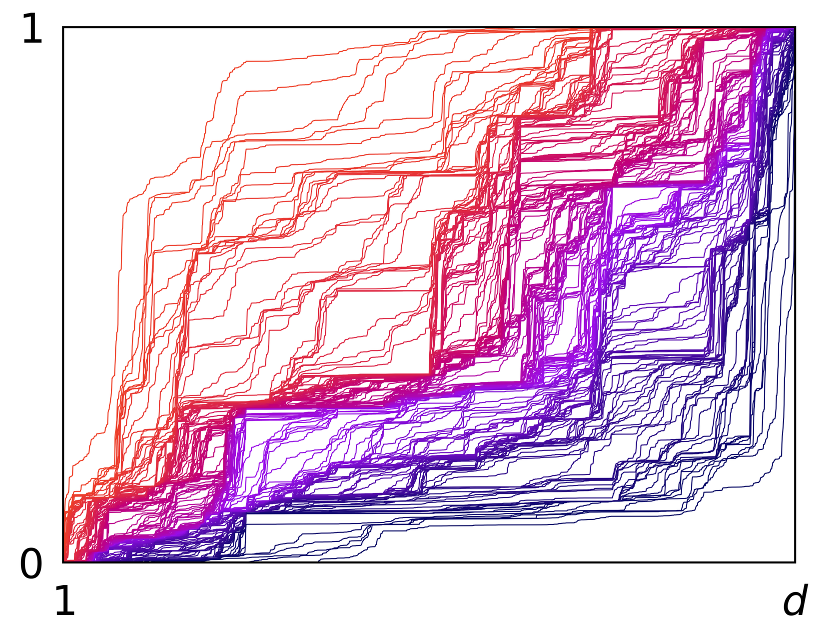


Matrix M_{π^*} plotted by lines

- Increasing columns for an unknown permutation π^*: $M_{\pi^*(i),k} \leq M_{\pi^*(i+1),k}$
- No constraint on the rows

Bi-Isotonicity Constraints





Matrix M_{π^*} plotted by lines

- Increasing columns for an unknown permutation π^*: $M_{\pi^*(i),k} \leq M_{\pi^*(i+1),k}$
- Increasing rows: $M_{ik} \leq M_{i,k+1}$

MiniMax Permutation Risk

In both models, we introduce the following **permutation loss** for any estimator $\hat{\pi}$:

$$l(\hat{\pi}, \pi^*, M) = \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2 \ ,$$

and the associated minimax permutation risk:

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

Carpentier, Pilliat, Verzelen

Assume we are in the isotonic or bi-isotonic model and that we have a polylogarithmic number of samples. There exists an estimator $\hat{\pi}$ computable in **polynomial time** achieving the minimax permutation risk up to polylogarithms:

$$\sup_{\pi^*, M} \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2 \lesssim \mathcal{R}^*_{\mathrm{perm}} \ .$$

Moreover, we give the minimax risks for permutation and estimation, in both models, for any n, d:

[1] **Isotonic Model**:

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

[2] **Bi-Isotonic Model**:

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

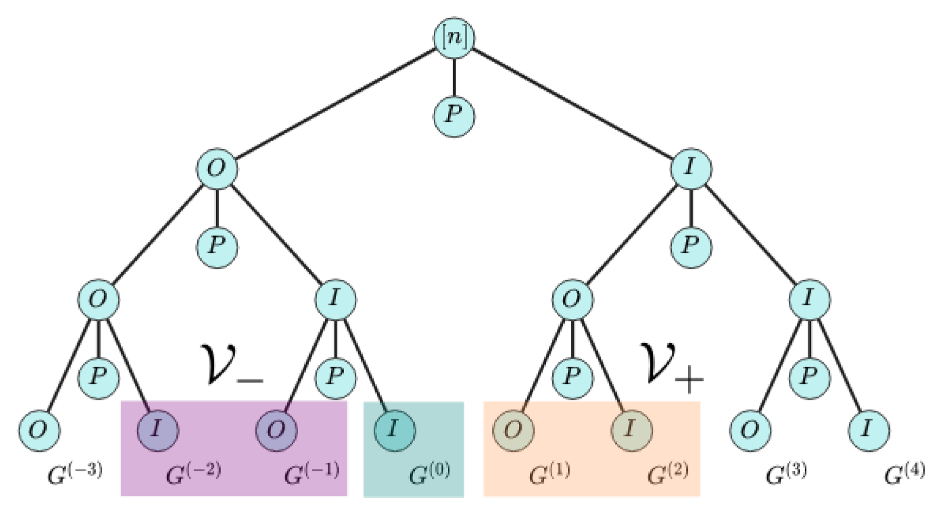
MiniMax Estimation Risk

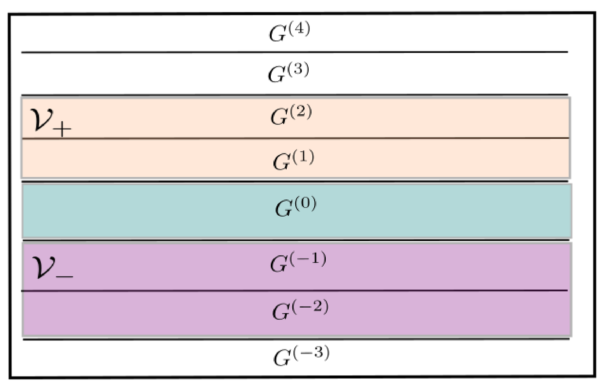
We can also introduce the minimax estimation risk for any estimator $\hat{M}$ of M:

$$\mathcal{R}^*_{\mathrm{est}} = \inf_{\hat{M}} \sup_{\pi^*, M} \mathbb{E}\|\hat{M} - M\|_F^2 \ .$$

A minimax estimator $\hat{\pi}$ can be combined with an isotonic or bi-isotonic regression to obtain an estimator $\hat{M}$ achieving $\mathcal{R}^*_{\mathrm{est}}$

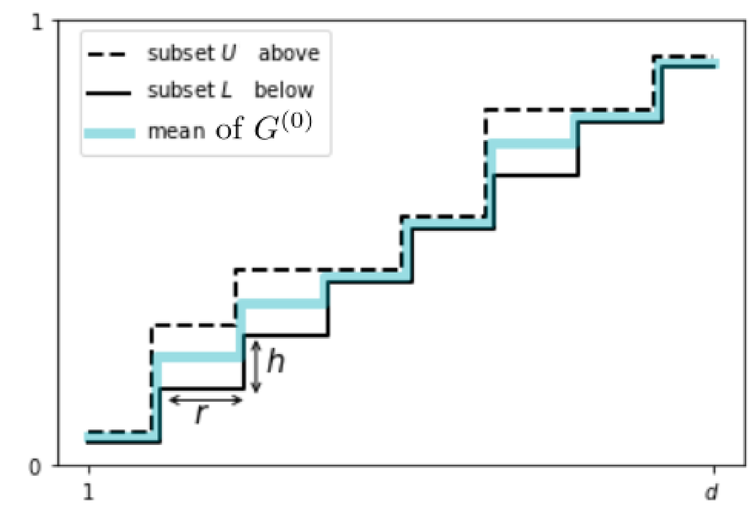
General Idea: Non-Oblivious Hierarchical Clustering





Iteratively trisect any set G of rows of M in (O, P, I) such that with high probability, all the experts in O are below all the experts in I. On the above pictures, a useful information to trisect $G^{(0)}$ is that it is sandwiched between some sets of rows that have already been classified as above or below $G^{(0)}$.

Worst-Case Scenario in the Bi-Isotonic Model



$$\frac{\sqrt{r}h}{2} \begin{pmatrix} 0&-1&-1&0&0&-1&0&0 \\ 0&1&1&0&0&1&0&0 \\ 0&1&1&0&0&1&0&0 \\ 0&-1&-1&0&0&-1&0&0 \\ 0&-1&-1&0&0&-1&0&0 \\ 0&1&1&0&0&1&0&0 \end{pmatrix}$$

A worst case scenario for a given set $G^{(0)}$ is when it contains two types of rows, that are either in a set L or in a set U. The Rank 1 Matrix on the right corresponds to the left picture after a local aggregation of columns. The positive (resp. negative) lines correspond to rows in U (resp. in L), and the zero columns correspond to areas where all the rows of M are equal. This worst case leads to the idea of averaging over local areas around detectable **variations** of the rows and of using **PCA** to compute clusters with the first left singular vector.

[1] E. Pilliat, C. Carpentier, N. Verzelen. Work in Progress, 2023+
[2] E. Pilliat, C. Carpentier, N. Verzelen. Optimal Permutation Estimation in Crowd-Sourcing Problems [arxiv:2211.04092], 2022
[3] N. Flammarion, C. Mao, and P. Rigollet. Optimal rates of statistical seriation. Bernoulli, 25(1):623–653, 2019.
[4] C. Mao, A. Pananjady, and M. J. Wainwright. Towards optimal estimation of bivariate isotonic matrices with unknown permutations. The Annals of Statistics, 48(6):3183–3205, 2020.