# Ranking Items from Discrete Ratings: The Cost of Unknown User Thresholds

Oscar Villemaud  Suryanarayana Sankagiri  Matthias Grossglauser

EPFL, Switzerland

## Motivation

For many tasks, user feedback is collected in the form of discrete ratings in order to produce a ranking. We study the cost of obtaining a ranking from discrete ratings, and show that this cost is dominated by estimating the unknown personal discretization threshold of every new user.

## The Threshold Rating Model

- $n \in \mathbb{N}$ items and $m \in \mathbb{N}$ users
- $(X_i)_{i\in[n]} \in [0,1]$: the item scores, of density $f_X$
- $(Y_u)_{u\in[m]} \in [0,1]$: the user thresholds, of density $f_Y$

### Query Model:

$$\forall u \in \mathbb{N}, \forall i \in [n], q(u,i) \triangleq \mathbf{1}(X_i > Y_u)$$

*i.e* we can only observe the relative ordering of an item-threshold pair. This naturally splits items into "bins" (equivalence classes).

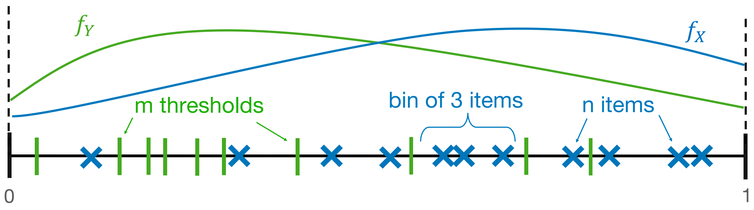



Figure 1. Item scores and user thresholds are sampled *iid*, respectively with densities $f_X$ and $f_Y$.

## Impossibility Result

**Lemma 1:** Let $X_1, \dots, X_n$ be i.i.d. item scores of density $f_X$ on $[0,1]$. Let $M$ be the random number of users needed to obtain a total order. Then we have

$$\mathbb{E}[M] = \infty$$

This cost stems from the difficulty of separating pairs of items that are very close.

$\longrightarrow$ We relax the problem by considering partial rankings.

## Maximum Spearman Footrule

### Definition:

For a given partial ranking, we define the Maximum Spearman Footrule $F$ as the diameter of the set $\mathcal{S}$ of compatible rankings with respect to the Spearman Footrule ranking error:

$$F(\mathcal{S}) \triangleq \max_{\sigma,\sigma' \in \mathcal{S}} \mathtt{SF}(\sigma, \sigma')$$

## Main Results

### Theorem 1

Assume that there exists $r \in \mathbb{R}^+ s.t. m \sim rn$ as $n$ goes to infinity, then

$$\left| \mathbb{E}[F] - n\left(\frac{1}{2} + \frac{1}{r}\mathbb{E}\left[\frac{f_X(Y)^2}{f_Y(Y)^2}\right]\right)\right| \leq \frac{r}{2}n + o(n)$$

### Theorem 2

Assume that there exists $r \in \mathbb{R}^+, \gamma > 1 \;\; s.t. \;\; m \sim rn^\gamma$ as $n$ goes to infinity, then

$$\mathbb{E}[F] \sim \frac{2}{r} n^{2-\gamma} \mathbb{E}\left[\frac{f_X(Y)^2}{f_Y(Y)^2}\right]$$

where $Y$ has density $f_Y$.

### Interpretation:

- Linear number of users $\Rightarrow \mathbb{E}[F]$ is linear.
- Quadratic number of users $\Rightarrow \mathbb{E}[F]$ is constant.
- Mismatch between the densities $\Rightarrow$ higher constants.

$\longrightarrow$ **In order to get constant ranking error, the number of users and queries must be at least quadratic in the number of items.**

## Algorithm for Querying Users

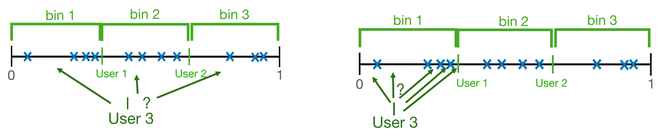



Figure 2. We find the bin containing the threshold, then locate the threshold inside the bin. The first part uses a binary search that returns to adjacent bins.

### Threshold Binary Search

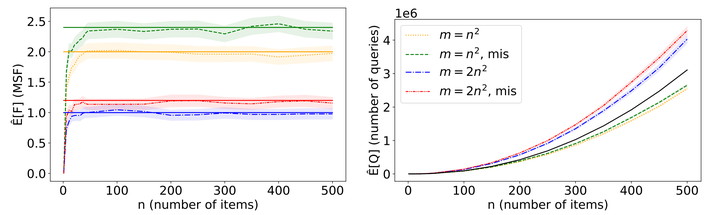
**for** each user $u$ **do**
- Find a pair of bins containing $Y_u$, the threshold of user $u$
- Rate items of the two bins until correct bin is found
- Rate all items of the selected bin to split it in two (if possible)

**return** The sequence of bins created, *i.e.*, partial ranking.

Estimated query complexity of the algorithm:

$$\mathbb{E}[Q] = O(n\log(m) + m\log(n))$$

## Experiments




(a) Expected Error (b) Expected number of queries

Figure 3. Quadratic number of users. The MSF rapidly converges to the values predicted by Theorem 2 (the solid lines of (a)). On (b), we see that the empirical query complexity is of the order $n^2\log(n)$ (the solid black line shows a reference of $y = 2n^2\log(n)$). *mis* (mismatch) means $f_X \neq f_Y$.

### Interpretation:

- Error term of Theorem 2 vanishes after $\simeq$ 50 items.
- Number of queries of `TBS` follows the $m\log(n)$ tendency.
- Mismatch only slightly increases the average number of queries per user.

## Conclusion

- Query complexity of $\Omega(n^2)$, tight up to a logarithmic factor.
- Compared to the $O(n\log(n))$ complexity of ranking from comparisons, the excess sample complexity can be interpreted as the "cost of estimating the latent threshold of every user".
- Ranking recovery decreases when the thresholds and scores have different distributions, via a quadratic divergence term.

**Main Takeway**

In order to keep a constant ranking error when ranking from discrete ratings with unknown user thresholds, we prove a quadratic lower bound on the number of queries with respect to the number of items.

### Limitations:

- Shared utilities across users.
- No noise in the model.

**How easily do these principles extend to more realistic models?**

Scan for paper!

[oscar.villemaud,suryanarayana.sankagiri]@epfl.ch