Active Ranking with Subset-wise Preferences

Aadirupa Saha
Indian Institute of Science, Bangalore
aadirupa@iisc.ac.in

Aditya Gopalan
Indian Institute of Science, Bangalore
aditya@iisc.ac.in

Abstract

We consider the problem of probably approximately correct (PAC) ranking n items by adaptively eliciting subset-wise preference feedback. At each round, the learner chooses a subset of k items and observes stochastic feedback indicating preference information of the winner (most preferred) item of the chosen subset drawn according to a Plackett-Luce (PL) subset choice model unknown a priori. The objective is to identify an ϵ-optimal ranking of the n items with probability at least $1-\delta$. When the feedback in each subset round is a single Plackett-Luce-sampled item, we show (ϵ,δ)-PAC algorithms with a sample complexity of $O\left(\frac{n}{\epsilon^2}\ln\frac{n}{\delta}\right)$ rounds, which we establish as being order-optimal by exhibiting a matching sample complexity lower bound of $\Omega\left(\frac{n}{\epsilon^2}\ln\frac{n}{\delta}\right)$—this shows that there is essentially no improvement possible from the pairwise comparisons setting ($k=2$). When, however, it is possible to elicit top-m ($\leq k$) ranking feedback according to the PL model from each adaptively chosen subset of size k, we show that an (ϵ,δ)-PAC ranking sample complexity of $O\left(\frac{n}{m\epsilon^2}\ln\frac{n}{\delta}\right)$ is achievable with explicit algorithms, which represents an m-wise reduction in sample complexity compared to the pairwise case. This again turns out to be order-wise unimprovable across the class of symmetric ranking algorithms. Our algorithms rely on a novel pivot trick to maintain only n itemwise score estimates, unlike $O(n^2)$ pairwise score estimates that has been used in prior work. We report results of numerical experiments that corroborate our findings.

Proceedings of the 22[nd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

1 Introduction

Ranking or sorting is a classic search problem and basic algorithmic primitive in computer science. Perhaps the simplest and most well-studied ranking problem is using (noisy) pairwise comparisons, which started from the work of Feige et al. [19], and which has recently been studied in machine learning under the rubric of ranking in 'dueling bandits' [9].

However, more general *subset-wise* preference feedback arises naturally in application domains where there is flexibility to learn by eliciting preference information from among a *set* of offerings, rather than by just asking for a pairwise comparison. For instance, web search and recommender systems applications typically involve users expressing preferences by clicking on one result (or a few results) from a presented set. Medical surveys, adaptive tutoring systems and multiplayer sports/games are other domains where subsets of questions, problem set assignments and tournaments, respectively, can be carefully crafted to learn users' relative preferences by subset-wise feedback.

In this paper, we explore *active*, probably approximately correct (PAC) ranking of n items using subset-wise, preference information. We assume that upon choosing a subset of $k \geq 2$ items, the learner receives preference feedback about the subset according to the well-known Plackett-Luce (PL) probability model [27]. The learner faces the goal of returning a near-correct ranking of all items, with respect to a tolerance parameter ϵ on the items' PL weights, with probability at least $1-\delta$ of correctness, after as few subset comparison rounds as possible. In this context, we make the following contributions:

1. We consider active ranking with winner information feedback, where the learner, upon playing a subset $S_t \subseteq [n]$ of exactly $k = |S_t|$ elements at each round t, receives as feedback a single winner sampled from the Plackett-Luce probability distribution on the elements of S_t. We design two (ϵ,δ)-PAC algorithms for this problem (Section 5) with sample complexity $O\left(\frac{n}{\epsilon^2}\ln\frac{n}{\delta}\right)$ rounds, for

learning a near-correct ranking on the items.

2. We show a matching lower bound of $\Omega\left(\frac{n}{\epsilon^2}\ln\frac{n}{\delta}\right)$ rounds on the (ϵ,δ)-PAC sample complexity of ranking with winner information feedback (Section 6), which is also of the same order as that for the dueling bandit ($k = 2$) [38]. This implies that despite the increased flexibility of playing larger sets, with just winner information feedback, one cannot hope for a faster rate of learning than in the case of pairwise comparisons.

3. In the setting where it is possible to obtain 'top-rank' feedback – an ordered list of $m \le k$ items sampled from the Plackett-Luce distribution on the chosen subset – we show that natural generalizations of the winner-feedback algorithms above achieve (ϵ,δ)-PAC sample complexity of $O\left(\frac{n}{m\epsilon^2}\ln\frac{n}{\delta}\right)$ rounds (Section 7), which is a significant improvement over the case of only winner information feedback. We show that this is order-wise tight by exhibiting a matching $\Omega\left(\frac{n}{m\epsilon^2}\ln\frac{n}{\delta}\right)$ lower bound on the sample complexity across (ϵ,δ)-PAC algorithms.

4. We report numerical results to show the performance of the proposed algorithms on synthetic environments (Section 8).

By way of techniques, the PAC algorithms we develop leverage the property of independence of irrelevant attributes (IIA) of the Plackett-Luce model, which allows for $O(n)$ dimensional parameter estimation with tight confidence bounds, even in the face of a combinatorially large number of possible subsets of size k. We also devise a generic 'pivoting' idea in our algorithms to efficiently estimate a global ordering using only local comparisons with a pivot or probe element: split the entire pool into playable subsets all containing one common element, learn local orderings relative to this element and then merge. Here again, the IIA structure of the PL model helps to ensure consistency among preferences aggregated across disparate subsets but with a common reference pivot. Our sample complexity lower bounds are information-theoretic in nature and rely on a generic change-of-measure argument but with carefully crafted confusing instances.

Related Work. Over the years, ranking from pairwise preferences ($k = 2$) has been studied in both the batch or non-adaptive setting [20, 32, 37, 30] and the active or adaptive setting [7, 22, 2]. In particular, prior work has addressed the problem of statistical parameter estimation given preference observations from the Plackett-Luce model in the offline setting [30, 15, 26, 21]. There also have been recent developments on the PAC objective for different pairwise preference models, such as those satisfying stochastic triangle inequalities and strong stochastic transitivity [38], general utility-based preference models [36], the Plackett-Luce model [34] and the Mallows model [11]]. Recent work has studied PAC-learning objectives other than identifying the single (near) best arm, e.g. recovering a few of the top arms [10, 28, 13], or the true ranking of the items [12, 18]. There is also work on the problem of Plackett-Luce parameter estimation in the subset-wise feedback setting [23, 26], but for the batch (offline) setup where the sampling is not adaptive. Recent work by Chen et al. [14] analyzes an active learning problem in the Plackett-Luce model with subset-wise feedback; however, the objective there is to recover the top-ℓ (unordered) items of the model, unlike full-rank recovery considered in this work. Moreover, they give instance-dependent sample complexity bounds, whereas we allow a tolerance (ϵ) in defining good rankings, natural in many settings [34, 38, 11].

2 Preliminaries

Notation. We denote the set $[n] = \{1, 2, ..., n\}$. When there is no confusion about the context, we often represent (an unordered) subset S as a vector, or ordered subset, S of size $|S|$ (according to, say, the order induced by the natural global ordering $[n]$ of all the items). In this case, $S(i)$ denotes the item (member) at the ith position in subset S. $\boldsymbol{\Sigma}_S = \{\boldsymbol{\sigma} \mid \boldsymbol{\sigma}$ is a permutation over items of $S\}$. where for any permutation $\boldsymbol{\sigma} \in \Sigma_S$, $\sigma(i)$ denotes the position of element $i \in S$ in the ranking $\boldsymbol{\sigma}$. $\mathbf{1}(\varphi)$ denote an indicator variable that takes the value 1 if the predicate φ is true, and 0 otherwise. $Pr(A)$ is used to denote the probability of event A, in a probability space that is clear from the context. $Ber(p)$ and $Geo(p)$ respectively denote Bernoulli and Geometric [1] random variable with probability of success at each trial being $p \in [0, 1]$. Moreover, for any $n \in \mathbb{N}$, $Bin(n, p)$ and $NB(n, p)$ respectively denote Binomial and Negative Binomial distribution.

2.1 Discrete Choice Models and Plackett-Luce (PL)

A discrete choice model specifies the relative preferences of two or more discrete alternatives in a given set. A widely studied class of discrete choice models is the class of *Random Utility Models* (RUMs), which assume a ground-truth utility score $\theta_i \in \mathbb{R}$ for each alternative $i \in [n]$, and assign a conditional distribution $\mathcal{D}_i(\cdot|\theta_i)$ for scoring item i. To model a winning alternative given any set $S \subseteq [n]$, one first draws a random utility score $X_i \sim \mathcal{D}_i(\cdot|\theta_i)$ for each alternative in S, and selects an item with the highest random score.

[1] this is the 'number of trials before success' version

One widely used RUM is the *Multinomial-Logit (MNL)* or *Plackett-Luce model (PL)*, where the $\mathcal{D}_i$s are taken to be independent Gumbel distributions with parameters θ'_i [3], i.e., with probability densities $\mathcal{D}_i(x_i|\theta'_i) = e^{-(x_j-\theta'_j)}e^{-e^{-(x_j-\theta'_j)}}$, $\theta'_i \in R$, $\forall i \in [n]$. Moreover assuming $\theta'_i = \ln\theta_i$, $\theta_i > 0\ \forall i \in [n]$, it can be shown in this case the probability that an alternative i emerges as the winner in the set $S \ni i$ becomes:$Pr(i|S) = \frac{\theta_i}{\sum_{j\in S}\theta_j}$.

Other families of discrete choice models can be obtained by imposing different probability distributions over the utility scores X_i, e.g. if $(X_1, \ldots X_n) \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ are jointly normal with mean $\boldsymbol{\theta} = (\theta_1, \ldots \theta_n)$ and covariance $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$, then the corresponding RUM-based choice model reduces to the *Multinomial Probit (MNP)*.

Independence of Irrelevant Alternatives A choice model Pr is said to possess the *Independence of Irrelevant Attributes (IIA)* property if the ratio of probabilities of choosing any two items, say i_1 and i_2 from within any choice set $S \ni i_1, i_2$ is independent of a third alternative j present in S [4]. Specifically, $\frac{Pr(i_1|S_1)}{Pr(i_2|S_1)} = \frac{Pr(i_1|S_2)}{Pr(i_2|S_2)}$ for any two distinct subsets $S_1, S_2 \subseteq [n]$ that contain i_1 and i_2. Plackett-Luce satisfies the IIA property.

3 Problem Setup

We consider the PAC version of the sequential decision-making problem of finding the ranking of n items by making subset-wise comparisons. Formally, the learner is given a finite set $[n]$ of $n > 2$ arms. At each decision round $t = 1, 2, \ldots$, the learner selects a subset $S_t \subseteq [n]$ of k items, and receives (stochastic) feedback about the winner (or most preferred) item of S_t drawn from a Plackett-Luce (PL) model with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, a priori unknown to the learner. The nature of the feedback is described in Section 3.1. We assume henceforth that $\theta_i \in [0, 1]$, $\forall i \in [n]$, and also $1 = \theta_1 > \theta_2 > \ldots > \theta_n$ for ease of exposition[2].

Definition 1 (ϵ*-Best-Item*)**.** *For any $\epsilon \in [0, 1)$, an item i is called ϵ-Best-Item if its PL score parameter θ_i is worse than the* Best-Item $i^* = 1$ *by no more than ϵ, i.e. if $\theta_i \geq \theta_1 - \epsilon$. A* 0*-best item is an item with largest PL parameter, which is also a Condorcet winner [33] in case it is unique.*

Definition 2 (ϵ*-Best-Ranking*)**.** *We define a ranking $\boldsymbol{\sigma} \in \Sigma_{[n]}$ to be an ϵ-Best-Ranking when no pair of items in $[n]$ is misranked by $\boldsymbol{\sigma}$ unless their PL scores are ϵ-close to each other. Formally, $\nexists i, j \in [n]$, such that $\sigma(i) > \sigma(j)$ and $\theta_i \geq \theta_j + \epsilon$. A* 0*-Best-Ranking will be called a Best-Ranking* or optimal ranking *of the PL model. With $1 = \theta_1 > \theta_2 > \ldots > \theta_n$, clearly the unique Best-Ranking is $\boldsymbol{\sigma}^* = (1, 2, \ldots, n)$.*

Definition 3 (ϵ*-Best-Ranking-Multiplicative*)**.** *We define a ranking $\boldsymbol{\sigma} \in \Sigma_{[n]}$ of $\boldsymbol{\sigma}^*$ to be ϵ-Best-Ranking-Multiplicative if $\nexists i, j \in [n]$, such that $\sigma(i) > \sigma(j)$, with $Pr(i|\{i, j\}) \geq \frac{1}{2} + \epsilon$.*

Note: The term 'multiplicative' emphasizes the fact that the condition $Pr(i|\{i, j\}) \geq \frac{1}{2} + \epsilon$ equivalently imposes a multiplicative constraint $\theta_i \geq \theta_j \left(\frac{1/2+\epsilon}{1/2-\epsilon}\right)$ on the PL score parameters.

3.1 Feedback models

By feedback model, we mean the information received (from the 'environment') once the learner plays a subset $S \subseteq [n]$ of k items. We consider the following feedback models in this work:

Winner of the selected subset (WI): The environment returns a single item $I \in S$, drawn independently from the probability distribution $Pr(I = i|S) = \frac{\theta_i}{\sum_{j\in S}\theta_j}\ \ \forall i \in S$.

Full ranking on the selected subset (FR): The environment returns a full ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_S$, drawn from the probability distribution $Pr(\boldsymbol{\sigma}|S) = \prod_{i=1}^{|S|}\frac{\theta_{\sigma^{-1}(i)}}{\sum_{j=i}^{|S|}\theta_{\sigma^{-1}(j)}}$, $\sigma \in \boldsymbol{\Sigma}_S$. This is equivalent to picking item $\boldsymbol{\sigma}^{-1}(1) \in S$ according to winner (WI) feedback from S, then picking $\boldsymbol{\sigma}^{-1}(2)$ according to WI feedback from $S \setminus \{\boldsymbol{\sigma}^{-1}(1)\}$, and so on, until all elements from S are exhausted, or, in other words, successively sampling $|S|$ winners from S according to the PL model, without replacement. But more generally, one can define

Top-m ranking from the selected subset (TR-m or TR): The environment successively samples (without replacement) only the first m items from among S, according to the PL model over S, and returns the ordered list. It follows that **TR** reduces to **FR** when $m = k = |S|$ and to **WI** when $m = 1$.

3.2 Performance Objective: (ϵ, δ)-PAC-Rank$-$ Correctness and Sample Complexity

Consider a problem instance with Plackett-Luce (PL) model parameters $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_n)$ and subsetsize $k \leq n$, with its *Best-Ranking* being $\boldsymbol{\sigma}^* = (1, 2, \ldots n)$, and $\epsilon, \delta \in (0, 1)$ are two given constants. A sequential algorithm that operates on this problem instance, with WI feedback model, is said to be (ϵ, δ)**-PAC-Rank** if (a) it stops and outputs a ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$ after a finite number of decision rounds (subset plays) with probability 1, and (b) the probability that its

[2] We naturally assume that this knowledge ordering of the items is not known to the learning algorithm, and note that extension to the case where several items have the same highest parameter value is easily accomplished.

output $\boldsymbol{\sigma}$ is an ϵ-*Best-Ranking* is at least $1-\delta$, i.e, $Pr(\boldsymbol{\sigma}$ is ϵ-*Best-Ranking*$) \geq 1-\delta$. Furthermore, by *sample complexity* of the algorithm, we mean the expected time (number of decision rounds) taken by the algorithm to stop.

In the context of our above problem objective, it is worth noting the work by [34] addressed a similar problem, except in the dueling bandit setup ($k=2$) with the same objective as above, except with the notion of ϵ-*Best-Ranking-Multiplicative*—we term this new objective as (ϵ,δ)-**PAC-Rank-Multiplicative** as referred later for comparing the results. The two objectives are however equivalent under a mild boundedness assumption as follows:

Lemma 4. *Assume $\theta_i \in [a,b], \forall i \in [n]$, for any $a,b \in (0,1)$. If an algorithm is (ϵ,δ)-***PAC-Rank***, then it is also (ϵ',δ)-***PAC-Rank-Multiplicative** *for any $\epsilon' \leq \frac{\epsilon}{4b}$. On the other hand, if an algorithm is (ϵ,δ)-***PAC-Rank-Multiplicative***, then it is also (ϵ',δ)-***PAC-Rank** *for any $\epsilon' \leq 4a\epsilon(1+\epsilon)$.*

4 Parameter Estimation with PL based preference data

We develop in this section some useful parameter estimation techniques based on adaptively sampled preference data from the PL model, which will form the basis for our PAC algorithms later on, in Section 5.1.

4.1 Estimating Pairwise Preferences via Rank-Breaking.

Rank breaking is a well-understood idea involving the extraction of pairwise comparisons from (partial) ranking data, and then building pairwise estimators on the obtained pairs by treating each comparison independently [26, 23], e.g., a winner a sampled from among a,b,c is rank-broken into the pairwise preferences $a \succ b$, $a \succ c$. We use this idea to devise estimators for the pairwise win probabilities $p_{ij} = P(i|\{i,j\}) = \theta_i/(\theta_i+\theta_j)$ in the active learning setting. The following result, used to design Algorithm 1 later, establishes explicit confidence intervals for pairwise win/loss probability estimates under adaptively sampled PL data.

Lemma 5 (Pairwise win-probability estimates for the PL model)**.** *Consider a Plackett-Luce choice model with parameters $\boldsymbol{\theta} = (\theta_1,\theta_2,\ldots,\theta_n)$, and fix two items $i,j \in [n]$. Let $S_1,\ldots,S_T$ be a sequence of (possibly random) subsets of $[n]$ of size at least 2, where T is a positive integer, and $i_1,\ldots,i_T$ a sequence of random items with each $i_t \in S_t$, $1 \leq t \leq T$, such that for each $1 \leq t \leq T$, (a) S_t depends only on $S_1,\ldots,S_{t-1}$, and (b) i_t is distributed as the Plackett-Luce winner of the subset S_t, given $S_1,i_1,\ldots,S_{t-1},i_{t-1}$ and S_t, and (c) $\forall t : \{i,j\} \subseteq S_t$ with probability 1. Let $n_i(T) = \sum_{t=1}^T \mathbf{1}(i_t = i)$ and $n_{ij}(T) = \sum_{t=1}^T \mathbf{1}(\{i_t \in \{i,j\}\})$. Then, for any positive integer v, and $\eta \in (0,1)$,*

$$Pr\left(\frac{n_i(T)}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \geq \eta,\ n_{ij}(T) \geq v\right) \leq e^{-2v\eta^2},$$

$$Pr\left(\frac{n_i(T)}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \leq -\eta,\ n_{ij}(T) \geq v\right) \leq e^{-2v\eta^2}.$$

4.2 Estimating relative PL scores (θ_i/θ_j) using Renewal Cycles

We detail another method to directly estimate (relative) score parameters of the PL model, using renewal cycles and the IIA property.

Lemma 6. *Consider a Plackett-Luce choice model with parameters $(\theta_1,\theta_2,\ldots,\theta_n)$, $n \geq 2$, and an item $b \in [n]$. Let $i_1,i_2,\ldots$ be a sequence of iid draws from the model. Let $\tau = \min\{t \geq \mathbb{N} \mid i_t = b\}$ be the first time at which b appears, and for each $i \neq b$, let $w_i(\tau) = \sum_{t=1}^{\tau} \mathbf{1}(i_t = i)$ be the number of times $i \neq b$ appears until time τ. Then, $\tau - 1$ and $w_i(\tau)$ are Geometric random variables with parameters $\frac{\theta_b}{\sum_{j\in[n]}\theta_j}$ and $\frac{\theta_b}{\theta_i+\theta_b}$, respectively.*

With this in hand, we now show how fast the empirical mean estimates over several renewal cycles (defined by the appearance of a distinguished item) converge to the true relative scores $\frac{\theta_i}{\theta_b}$, a result to be employed in the design of Algorithm 3 later.

Lemma 7 (Concentration of Geometric Random Variables via the Negative Binomial distribution.)**.** *Suppose $X_1,X_2,\ldots X_d$ are d iid $Geo(\frac{\theta_b}{\theta_b+\theta_i})$ random variables, and $Z = \sum_{i=1}^d X_i$. Then, for any $\eta > 0$, $Pr\left(\left|\frac{Z}{d} - \frac{\theta_i}{\theta_b}\right| \geq \eta\right) < 2\exp\left(-\frac{2d\eta^2}{\left(1+\frac{\theta_i}{\theta_b}\right)^2\left(\eta+1+\frac{\theta_i}{\theta_b}\right)}\right)$.*

5 Algorithms for WI Feedback

This section describes the design of (ϵ,δ)-**PAC-Rank** algorithms with winner information (WI) feedback.

A key idea behind our proposed algorithms is to estimate the relative strength of each item with respect to a fixed item, termed as a *pivot-item* b. This helps to compare every item on common terms (with respect to the pivot item) even if two items are not directly compared with each other. Our first algorithm *Beat-the-Pivot* maintains pairwise score estimates P_{ib} of the items $i \in [n]\backslash\{b\}$ with respect to the pivot element by deriving intuition from Lemma 5. The second algorithm *Score-and-Rank* directly estimates the relative scores $\frac{\theta_i}{\theta_b}$ for each item $i \in [n]\backslash\{b\}$, relying on Lemma 6 (Section 4.2). Once all item scores are estimated with enough confidence, the items are simply sorted with respect to their preference scores to obtain a ranking.

5.1 The *Beat-the-Pivot* algorithm

Algorithm 1 *Beat-the-Pivot*

Input:
Set of item: $[n]$ $(n \geq k)$, and subset size: k
Error bias: $\epsilon > 0$, confidence parameter: $\delta > 0$
Initialize:
$\epsilon_b \leftarrow \min(\frac{\epsilon}{2}, \frac{1}{2})$; $b \leftarrow$ *Find-the-Pivot*$(n, k, \epsilon_b, \frac{\delta}{2})$
Set $S \leftarrow [n] \setminus \{b\}$, and divide S into $G := \lceil \frac{n-1}{k-1} \rceil$ sets $\mathcal{G}_1, \mathcal{G}_2, \cdots \mathcal{G}_G$ such that $\cup_{j=1}^{G} \mathcal{G}_j = S$ and $\mathcal{G}_j \cap \mathcal{G}_{j'} = \emptyset$, $\forall j, j' \in [G]$, $|G_j| = (k-1)$, $\forall j \in [G-1]$
If $|\mathcal{G}_G| < (k-1)$, **then** set $\mathcal{R} \leftarrow \mathcal{G}_G$, and $S \leftarrow S \setminus \mathcal{R}$, $S' \leftarrow$ Randomly sample $(k-1-|\mathcal{G}_G|)$ items from S, and set $\mathcal{G}_G \leftarrow \mathcal{G}_G \cup S'$
Set $\mathcal{G}_j = \mathcal{G}_j \cup \{b\}$, $\forall j \in [G]$
for $g = 1, 2, \ldots, G$ **do**
Set $\epsilon' \leftarrow \frac{\epsilon}{16}$ and $\delta' \leftarrow \frac{\delta}{8n}$
Play the subset $\mathcal{G}_g$ for $t := \frac{2k}{\epsilon'^2} \log \frac{1}{\delta'}$ times
Set $w_i \leftarrow$ Number of times i won in m plays of $\mathcal{G}_g$, and estimate $\hat{p}_{ib} \leftarrow \frac{w_i}{w_i + w_b}$, $\forall i \in \mathcal{G}_g$
end for
Choose $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$, such that $\sigma(b) = 1$ and $\boldsymbol{\sigma}(i) < \boldsymbol{\sigma}(j)$ if $\hat{p}_{ib} > \hat{p}_{jb}$, $\forall i, j \in S \cup \mathcal{R}$
Output: The ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$

Beat-the-Pivot (Algorithm 1) first estimates an *approximate Best-Item* b with high probability $(1 - \delta/2)$. We do this using the subroutine *Find-the-Pivot*(n, k, ϵ, δ) (Algorithm *Find-the-Pivot*) that with probability at least $(1-\delta)$ *Find-the-Pivot* outputs an ϵ-*Best-Item* within a sample complexity of $O(\frac{n}{\epsilon^2} \log \frac{k}{\delta})$.

Once the best item b is estimated, *Beat-the-Pivot* divides the rest of the $n-1$ items into groups of size $k-1$, $\mathcal{G}_1, \mathcal{G}_2, \cdots \mathcal{G}_G$, and appends b to each group. This way elements of every group get to compete with b, which aids estimating the pairwise score compared to the pivot item b, $\hat{p}_{ib}$ owing to the *IIA property* of PL model and Lemma 5 (Section 4.1), sorting which we obtain the final ranking. Theorem 8 shows that *Beat-the-Pivot* enjoys the optimal sample complexity guarantee of $O\Big((\frac{n}{\epsilon^2}) \log \Big(\frac{n}{\delta}\Big)\Big))$.

Theorem 8 (*Beat-the-Pivot*: Correctness and Sample Complexity)**.** *Beat-the-Pivot (Algorithm 1) is* (ϵ, δ)**-PAC-Rank** *with sample complexity* $O\big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\big)$.

5.2 The *Score-and-Rank* algorithm

Score-and-Rank (Algorithm 3) differs from *Beat-the-Pivot* in terms of the score estimate it maintains for each item. Unlike our previous algorithm, instead of maintaining pivot-preference scores $p_{ib} = Pr(i \succ b)$, *Beat-the-Pivot*, aims to directly estimate the PL-score θ_i of each item relative to score of the pivot θ_b. In other words, the algorithm maintains the *relative score* estimates $\frac{\theta_i}{\theta_b}$ for every item $i \in [n] \setminus \{b\}$ borrowing results from Lemma 6 and 7, and finally return the ranking sorting the items with respect to their *relative pivotal-score*. *Score-and-Rank* also runs within an optimal sample complexity of $\big(\frac{n}{\epsilon^2} \ln \frac{n}{\delta}\big)$ as shown in Theorem 9. Pseudocode for the algorithm is detailed in Algorithm 3 in the appendix, due to space constraints.

Theorem 9 (*Score-and-Rank*: Correctness and Sample Complexity)**.** *Score-and-Rank (Algorithm 3) is* (ϵ, δ)**-PAC-Rank** *with sample complexity* $O\big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\big)$.

5.3 The *Find-the-Pivot* subroutine (for algorithms 1 and 3)

In this section, we describe the pivot selection procedure *Find-the-Pivot*(n, k, ϵ, δ). The algorithm serves the purpose of finding an ϵ-*Best-Item* with high probability $(1-\delta)$ that is used as *the pivoting element* b both by Algorithm 1 and and 3 (Section 5.1) and 5.2).

Find-the-Pivot is based on the simple idea of tracing the empirical best item–specifically, it maintains a running winner r_ℓ at every iteration ℓ, making it compete with a set of $k-1$ arbitrarily chosen items. After competing long enough ($t := O\big(\frac{k}{\epsilon^2} \ln \frac{n}{\delta}\big)$ rounds), if the empirical winner c_ℓ turns out to be more than $\frac{\epsilon}{2}$-favorable than the running winner r_ℓ, in term of its pairwise preference score: $\hat{p}_{c_\ell, r_\ell} > \frac{1}{2} + \frac{\epsilon}{2}$, then c_ℓ replaces r_ℓ, or else r_ℓ retains its place and status quo ensues. The formal description of *Find-the-Pivot* is in the appendix.

Lemma 10 (*Find-the-Pivot*: Correctness and Sample Complexity with WI)**.** *Find-the-Pivot (Algorithm 2) achieves the* (ϵ, δ)*-PAC objective with sample complexity* $O(\frac{n}{\epsilon^2} \log \frac{n}{\delta})$.

6 Lower Bound

In this section we show the minimum sample complexity required for any *symmetric algorithm* to be (ϵ, δ)**-PAC-Rank** is at least $\Omega\Big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\Big)$ (Theorem 12). Note this in fact matches the sample complexity bounds of our proposed algorithms (recall Theorem 8 and 9) showing the tightness of both our upper and lower bound guarantees. The key observation lies in noting that results are independent of k, which shows the learning problem with k-subsetwise WI feedback is as hard as that of the dueling bandit setup ($k = 2$)—the flexibility of playing a k sized subset does not help in faster information aggregation. We first define the notion of a *symmetric* or label-invariant algorithm.

Definition 11 (Symmetric Algorithm)**.** *A PAC algorithm $\mathcal{A}$ is said to be* symmetric *if its output is insensitive to the specific labelling of items, i.e., if for any PL model* $(\theta_1, \ldots, \theta_n)$*, bijection* $\phi : [n] \rightarrow$

$[n]$ and ranking $\boldsymbol\sigma : [n] \to [n]$, it holds that $Pr(\mathcal{A}$ outputs $\boldsymbol\sigma \,|\, (\theta_1,\ldots,\theta_n)) = Pr(\mathcal{A}$ outputs $\boldsymbol\sigma \circ \phi^{-1} \,|\, (\theta_{\phi(1)},\ldots,\theta_{\phi(n)}))$, where $Pr(\cdot\,|(\alpha_1,\ldots,\alpha_n))$ denotes the probability distribution on the trajectory of $\mathcal{A}$ induced by the PL model $(\alpha_1,\ldots,\alpha_n)$.

Theorem 12 (Lower bound on Sample Complexity with WI feedback)**.** *Given a fixed $\epsilon \in \Big(0, \frac{1}{\sqrt{8}}\Big]$, $\delta \in [0,1]$, and a symmetric (ϵ,δ)-**PAC-Rank** algorithm $\mathcal{A}$ for WI feedback, there exists a PL instance ν such that the sample complexity of $\mathcal{A}$ on ν is at least $\Omega\bigg(\frac{n}{\epsilon^2}\ln\frac{n}{4\delta}\bigg)$.*

Proof. **(sketch)**. The argument is based on the following change-of-measure argument (Lemma 1) of [25]. (restated in Appendix D.1 as Lemma 25). To employ this result, note that in our case, each bandit instance corresponds to an instance of the problem with arm set containing all the subsets of $[n]$ of size k: $\{S = (S(1),\ldots S(k)) \subseteq [n] \mid S(i) < S(j), \forall i < j\}$. The key part of our proof relies on carefully crafting a true instance, with optimal arm 1, and a family of slightly perturbed alternative instances $\{\boldsymbol\nu^a : a \neq 1\}$, each with optimal arm $a \neq 1$.

Designing the problem instances. We first renumber the n items as $\{0,1,2,\ldots n-1\}$. Now for any integer $q \in [n-1]$, we define $\boldsymbol\nu_{[q]}$ to be the set of problem instances where any instance $\nu_S \in \boldsymbol\nu_{[q]}$ is associated to a set $S \subseteq [n-1]$, such that $|S| = q$, and the PL parameters $\boldsymbol\theta$ associated to instance ν_S are set up as follows: $\theta_0 = \theta\bigg(\frac{1}{4}-\epsilon^2\bigg), \theta_j = \theta\bigg(\frac{1}{2}+\epsilon\bigg)^2 \forall j \in S$, and $\theta_j = \theta\bigg(\frac{1}{2}-\epsilon\bigg)^2 \forall j \in [n-1]\setminus S$, for some $\theta \in \mathbb{R}_+$, $\epsilon > 0$. We will restrict ourselves to the class of instances of the form $\boldsymbol\nu_{[q]}$, $q \in [n-1]$.

Corresponding to each problem $\nu_S \in \boldsymbol\nu_{[q]}$, such that $q \in [n-2]$, consider a *slightly altered* problem instance $\nu_{\tilde S}$ associated with a set $\tilde S \subseteq [n-1]$, such that $\tilde S = S \cup \{i\} \subseteq [n-1]$, where $i \in [n-1]\setminus S$. Following the same construction as above, the PL parameters of the problem instance $\nu_{\tilde S}$ are set up as: $\theta_0 = \theta\bigg(\frac{1}{4}-\epsilon^2\bigg), \theta_j = \theta\bigg(\frac{1}{2}+\epsilon\bigg)^2 \forall j \in \tilde S$, and $\theta_j = \theta\bigg(\frac{1}{2}-\epsilon\bigg)^2 \forall j \in [n-1]\setminus \tilde S$.

Remark 1. *Note that any problem instance $\nu_S \in \boldsymbol\nu_{[q]}$, $q \in [n-1]$ is thus can be uniquely defined by its underlying set $S \in [n-1]$. For simplicity we will also use the notations $S \in \boldsymbol\nu_{[q]}$ to define the problem instance.*

Remark 2. *It is easy to verify that, for any $\theta \ge \frac{1}{1-2\epsilon}$, an ϵ-Best-Ranking (Definition. 2) for problem instance ν_S, $S \subseteq [n-1]$, say $\boldsymbol\sigma_S$, has to satisfy the following: $\sigma_S(i) < \sigma_S(0), \forall i \in S$ and $\sigma_S(0) < \sigma_S(j), \forall j \in [n-1]\setminus S$. Thus for any instance S, the items in S should precede item 0 which itself precedes items in $[n-1]\setminus S$.*

For any ranking $\boldsymbol\sigma \in \Sigma_n$, we denote by $\sigma(1:i)$ the set first i items in the ranking, for any $i \in [n]$.

We now fix any set $S^* \subset [n-1]$, $|S^*| = q = \lfloor \frac{n}{2} \rfloor$. Theorem 12 is now obtained by applying Lemma 25 on pair of instances $(\nu_{S^*}, \nu_{\tilde S^*})$, for all possible choices of $\tilde S = S \cup \{i\}$, $i \in [n-1]\setminus S$, and for the event $\mathcal{E} := \{\boldsymbol\sigma_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\}$. However we apply a tighter upper bounds for the KL-divergence term of in the right hand side of Lemma 25. It is easy to note that as $\mathcal{A}$ is (ϵ,δ)-**PAC-Rank** , obviously $Pr_{S^*}\bigg(\boldsymbol\sigma_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\bigg) > 1-\delta$, and $Pr_{\tilde S^*}\bigg(\boldsymbol\sigma_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\bigg) < Pr_{\tilde S^*}\bigg(\boldsymbol\sigma_{\mathcal{A}}(1:q+1) \neq \tilde S^*\bigg) < \delta$. Further using $kl(Pr_{\nu_{S^*}}(\mathcal{E}), Pr_{\nu_{\tilde S^*}}(\mathcal{E})) \ge kl(1-\delta,\delta) \ge \ln\frac{1}{4\delta}$ (due to Lemma 26) leads to a lower bound guarantee of $\Omega\bigg(\frac{n}{\epsilon^2}\ln\frac{1}{\delta}\bigg)$, but that is loose by an $\Omega\big(\frac{n}{\epsilon^2}\log n\big)$ additive factor. Novelty of our analysis lies in further utilising the *symmetric property* of $\mathcal{A}$ to prove a tighter upper bound od the kl-divergence with the following result:

Lemma 13. *For any symmetric (ϵ,δ)-**PAC-Rank** algorithm $\mathcal{A}$, and any problem instance $\nu_S \in \boldsymbol\nu_{[q]}$ associated to the set $S \subseteq [n-1]$, $q \in [n-1]$, and for any item $i \in S$, $Pr_S\bigg(\boldsymbol\sigma_{\mathcal{A}}(1:q) = S\setminus\{i\} \cup \{0\}\bigg) < \frac{\delta}{q}$, where $Pr_S(\cdot)$ denotes the probability of an event under the underlying problem instance ν_S and the internal randomness of the algorithm $\mathcal{A}$ (if any).*

For our purpose, we use the above result for $S = \tilde S^*$ which leads to the desired tighter upper bound for $kl(Pr_{\nu_{S^*}}(\mathcal{E}), Pr_{\nu_{\tilde S^*}}(\mathcal{E})) \ge kl(1-\delta, \frac{\delta}{q}) \ge \ln\frac{q}{4\delta}$, the last inequality follows due to Lemma 26 (Appendix D.2). The complete proof can be found in Appendix D. □

Remark 3. *Theorem 12 shows, rather surprisingly, that the PAC-ranking with winner feedback information from size-k subsets, does not become easier (in a worst-case sense) with k, implying that there is no reduction in hardness of learning from the pairwise comparisons case (k = 2). While one may expect sample complexity to improve as the number of items being simultaneously tested in each round (k) becomes larger, there is a counteracting effect due to the fact that it is intuitively 'harder' for a high-value item to win in just a single winner draw against a (large) population of $k-1$ other competitors. A useful heuristic here is that the number of bits of information that a single winner draw from a size-k subset provides is $O(\ln k)$, which is not significantly larger than when $k > 2$; thus, an algorithm cannot accumulate significantly more information per round compared to the pairwise case.*

We also have a similar lower bound result for the (ϵ,δ)**-PAC-Rank-Multiplicative** objective of Szörényi et al. [34] (Section 3):

Theorem 14. *Given a fixed $\epsilon \in \big(0, \frac{1}{\sqrt{8}}\big]$, $\delta \in [0,1]$, and a symmetric (ϵ,δ)**-PAC-Rank-Multiplicative** algorithm $\mathcal{A}$ for WI feedback model, there exists a PL instance ν such that the sample complexity of $\mathcal{A}$ on ν is at least $\Omega\bigg(\frac{n}{\epsilon^2}\ln\frac{n}{4\delta}\bigg)$.*

7 Analysis with Top Ranking (TR) feedback

We now proceed to analyze the problem with *Top-m Ranking* (TR) feedback (Section 3.1). We first show that unlike WI feedback, the sample complexity lower bound here scales as $\Omega\bigg(\frac{n}{m\epsilon^2}\ln\frac{n}{\delta}\bigg)$ (Theorem 15), which is a factor m smaller than that in Theorem 12 for the WI feedback model. At a high level, this is because TR reveals preference information for m items per feedback round, as opposed to just a single (noisy) information sample of the winning item (WI). Following this, we also present two algorithms for this setting which are shown to enjoy an exact optimal sample complexity guarantee of $O\bigg(\frac{n}{m\epsilon^2}\ln\frac{n}{\delta}\bigg)$ (Section 7.2).

7.1 Lower Bound for Top-m Ranking (TR) feedback

Theorem 15 (Sample Complexity Lower Bound for TR)**.** *Given $\epsilon \in \Big(0, \frac{1}{8}\Big]$ and $\delta \in (0,1]$, and a symmetric (ϵ,δ)**-PAC-Rank** algorithm $\mathcal{A}$ with top-m ranking (TR) feedback ($2 \le m \le k$), there exists a PL instance ν such that the expected sample complexity of $\mathcal{A}$ on ν is at least $\Omega\bigg(\frac{n}{m\epsilon^2}\ln\frac{n}{4\delta}\bigg)$.*

Remark 4. *The sample complexity lower bound for (ϵ,δ)**-PAC-Rank** with top-m ranking (TR) feedback model is $\frac{1}{m}$-times that of the WI model (Theorem 12). Intuitively, revealing a ranking on m items in a k-set provides about $\ln\Big(\binom{k}{m}m!\Big) = O(m\ln k)$ bits of information per round, which is about m times as large as that of revealing a single winner, yielding an acceleration by a factor of m.*

Corollary 16. *Given $\epsilon \in \Big(0, \frac{1}{\sqrt{8}}\Big]$ and $\delta \in (0,1]$, and a symmetric (ϵ,δ)**-PAC-Rank** algorithm $\mathcal{A}$ with full ranking (FR) feedback ($m = k$), there exists a PL instance ν such that the expected sample complexity of $\mathcal{A}$ on ν is at least $\Omega\bigg(\frac{n}{k\epsilon^2}\ln\frac{1}{4\delta}\bigg)$.*

7.2 Algorithms for Top-m Ranking (TR) feedback model

In this section we present a modification of *Beat-the-Pivot* (Algorithm 1) for (ϵ,δ)-PAC objective with top-m ranking feedback. Algorithm 5 (Appendix E.1) shows that how a simple generalization of *Beat-the-Pivot* can proved to be (ϵ,δ)**-PAC-Rank** with optimal sample complexity guarantee (Theorem 17), using the idea of Rank-Breaking [26] on top-m ranking feedback.

Algorithm 5: Generalizing ***Beat-the-Pivot*** **for top-m ranking (TR) feedback.** The main trick we use in modifying *Beat-the-Pivot* for TR feedback is *Rank Breaking*, which essentially extracts pairwise comparisons from subset-wise feedback as described below:

Rank-Breaking [26]. Given any set S of size k, if $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{S_m}$, ($S_m \subseteq S$, $|S_m| = m$) denotes a possible top-m ranking of S, the *Rank Breaking* subroutine considers each item in S to be beaten by its preceding items in $\boldsymbol{\sigma}$ in a pairwise sense. See Algorithm 4 for detailed description of the Rank-Breaking procedure.

Using *Rank-Break* (Algorithm 4), our modified *Beat-the-Pivot* algorithm now essentially maintains the empirical pivotal preferences $\hat{p}_{ib}$ for each item $i \in [n]\backslash\{b\}$ by applying *Rank Breaking* on the TR feedback $\boldsymbol{\sigma}$ of each subsetwise play. Of course in general, *Rank Breaking* may lead to arbitrarily inconsistent estimates of the underlying model parameters [3]. However, owing to the *IIA property* of Plackett-Luce model, we get clean concentration guarantees on p_{ij} using Lemma 5. This is precisely the idea used for obtaining the $\frac{1}{m}$ factor improvement in the sample complexity guarantees of *Beat-the-Pivot* as analysed in Theorem 8. The formal descriptions of *Beat-the-Pivot* generalized to the setting of TR feedback, is given in Algorithm 5.

Theorem 17 (*Beat-the-Pivot*: Correctness and Sample Complexity with TR)**.** *With top-m ranking (TR) feedback model, Beat-the-Pivot (Algorithm 5) is (ϵ,δ)**-PAC-Rank** with sample complexity $O(\frac{n}{m\epsilon^2}\log\frac{n}{\delta})$.*

Remark 5. *Comparing Theorems 8 and 17 shows that the sample complexity of Beat-the-Pivot with TR feedback (Algorithm 5) is m times smaller than its corresponding counterpart for WI feedback, owing to the additional information gain revealed from preferences of top-m items instead of just 1 (i.e. only the winner).*

8 Experiments

We first present the setup of our empirical evaluations:

Algorithms. We simulate the results on our two proposed algorithms (1). *Beat-the-Pivot* and (2). *Score-and-Rank.* We also compare our ranking performance with the *PLPAC-AMPR* method, the only existing method (to the best of our knowledge) that addresses the online PAC ranking problem, although only in the dueling bandit setup (i.e. $k = 2$).

Ranking Performance Measure. We use the popular pairwise *Kendall's Tau ranking loss* [29] for measuring the accuracy of the estimated ranking $\boldsymbol{\sigma}$ with respect to the *Best-Ranking* $\boldsymbol{\sigma}^*$ with an additive ϵ-relaxation: $d_\epsilon(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) = \frac{1}{\binom{n}{2}} \sum_{i<j}(g_{ij} + g_{ji})$, where each $g_{ij} = \mathbf{1}\left((\theta_i > \theta_j + \epsilon) \wedge (\sigma(i) > \sigma(j))\right)$. All reported performances are averaged across 50 runs.

Environments. We use four PL models: 1. *geo8* (with $n = 8$) 2. *arith10* (with $n = 10$) 3. *har20* (with $n = 20$) and 4. *arith50* (with $n = 50$). Their individual score parameters are as follows: **1. geo8:** $\theta_1 = 1$, and $\frac{\theta_{i+1}}{\theta_i} = 0.875$, $\forall i \in [7]$. **2. arith10:** $\theta_1 = 1$ and $\theta_i - \theta_{i+1} = 0.1$, $\forall i \in [9]$. **3. har20:** $\theta = 1/(i)$, $\forall i \in [20]$. **4. arith50:** $\theta_1 = 1$ and $\theta_i - \theta_{i+1} = 0.02$, $\forall i \in [9]$.

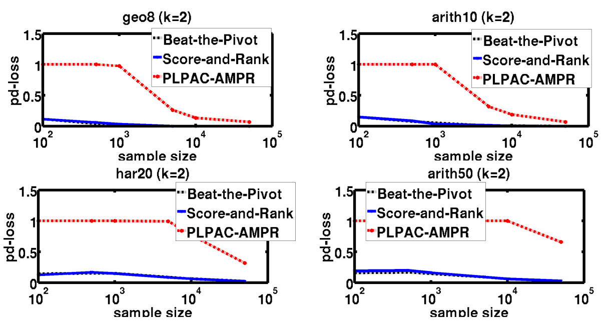

Figure 1: Ranking performance vs. sample size (# rounds) with dueling plays ($k = 2$)

Ranking with Pairwise Preferences ($k = 2$). We first compare the above three algorithms with pairwise preference feedback, i.e. with $k = 2$ and $m = 1$ (WI feedback model). We set $\epsilon = 0.01$ and $\delta = 0.1$. Figure 1 clearly shows superiority of our two proposed algorithms over *PLPAC-AMPR* [34] as they give much higher ranking accuracy given the sample size, rightfully justifying our improved theoretical guarantees as well (Theorem 8 and 9). Note that *geo8* and *arith50* are the easiest and hardest PL model instances, respectively; the latter has the largest n with gaps $\theta_i - \theta_{i+1} = 0.02$. This also reflects in our experimental results as the ranking estimation loss being the highest for *arith50* for all the algorithms, specifically *PLPAC-AMPR* very poorly till 10^4 samples.

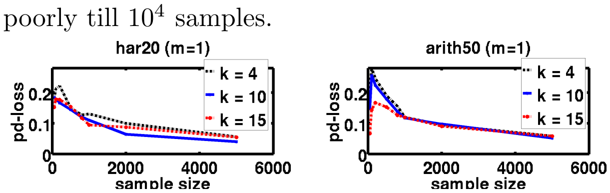

Figure 2: Ranking performance vs. subset size (k) with WI feedback ($m = 1$)

Ranking with Subsetwise-Preferences ($k > 2$) with Winner feedback. We next move to general subsetwise preference feedback ($k \geq 2$) for WI feedback model (i.e. for $m = 1$) [3]. We fix $\epsilon = 0.01$ and $\delta = 0.1$ and report the performance of *Beat-the-Pivot* on the datasets *har20* and *arith50*, varying k over the range 4 - 40. As expected from Theorem 8 and explained in Remark 3, the ranking performance indeed does not seem to be varying with increasing subsetsize k for WI feedback model for both PL models (Figure 2).

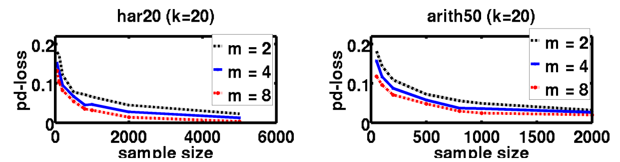

Figure 3: Ranking performance vs. feedback size (m) for fixed subset size (k)

Ranking with Subsetwise-Preferences ($k > 2$) with Top-rank feedback. We finally report the performance of *Beat-the-Pivot* (Algorithm 5) for top-m ranking (TR) feedback model on two PL models: *har20* (for $k = 20$) and *arith50* (for $k = 45$), varying the range of m from 2 to 40 (Figure 3). We set $\epsilon = 0.01$ and $\delta = 0.1$ as before. As expected, in this case indeed larger m improves the ranking accuracy given a fixed sample size which reflects over theoretical guarantee of $\frac{1}{m}$-factor improvement of the sample complexity for TR feedback (Theorem 15 and Remark 5).

9 Conclusion and Future Work

We have considered the PAC version of the problem of adaptively ranking n items from k-subset-wise comparisons, in the Plackett-Luce (PL) preference model with winner information (WI) and top ranking (TR) feedback. With just WI, the required sample complexity lower bound is $\Omega\left(\frac{n}{\epsilon^2} \ln \frac{n}{\delta}\right)$, which is surprisingly independent of the subset size k. We have also designed two algorithms enjoying optimal sample complexity guarantees, and based on a novel *pivoting-trick*. With TR feedback, a $\frac{1}{m}$-times faster learning rate is achievable, and we have given an algorithm with optimal sample complexity guarantees.

In the future, it would be of interest to analyse the problem with other choice models (e.g. multinomial probit, Mallows, nested logit, generalized extreme-value models, etc.), and perhaps to extend this theory to newer formulations such as assortment selection [5, 16], revenue maximization with item prices [35, 1], or even in contextual scenarios [17] where every individual user comes with their own model parameter.

[3] *PLPAC-AMPR* only works for $k = 2$ and is no longer applicable henceforth.

Acknowledgements

The authors are grateful to the anonymous reviewers for valuable feedback. This work is supported by a Qualcomm Innovation Fellowship 2018, and an Indigenous 5G Test Bed project grant from the Dept. of Telecommunications, Government of India.

References

[1] Shipra Agrawal, Vashist Avandhanula, Vineet Goyal, and Assaf Zeevi. A near-optimal exploration-exploitation approach for assortment selection. 2016.

[2] Nir Ailon. An Active Learning Algorithm for Ranking from Pairwise Preferences with an Almost Optimal Query Complexity. *Journal of Machine Learning Research*, 13(Jan):137–164, 2012.

[3] Hossein Azari, David Parkes, and Lirong Xia. Random utility theory for social choice. In *Advances in Neural Information Processing Systems*, pages 126–134, 2012.

[4] Austin R Benson, Ravi Kumar, and Andrew Tomkins. On the relevance of irrelevant alternatives. In *Proceedings of the 25th International Conference on World Wide Web*, pages 963–973. International World Wide Web Conferences Steering Committee, 2016.

[5] Gerardo Berbeglia and Gwenaël Joret. Assortment optimisation under a general discrete choice model: A tight analysis of revenue-ordered assortments. *arXiv preprint arXiv:1606.01371*, 2016.

[6] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[7] Mark Braverman and Elchanan Mossel. Noisy Sorting without Resampling. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 268–276. Society for Industrial and Applied Mathematics, 2008.

[8] Daniel G Brown. How i wasted too long finding a concentration inequality for sums of geometric variables. *Found at https://cs. uwaterloo. ca/˜ browndg/negbin. pdf*, 6, 2011.

[9] Róbert Busa-Fekete and Eyke Hüllermeier. A survey of preference-based online learning with bandit algorithms. In *International Conference on Algorithmic Learning Theory*, pages 18–39. Springer, 2014.

[10] Róbert Busa-Fekete, Balazs Szorenyi, Weiwei Cheng, Paul Weng, and Eyke Hüllermeier. Top-k selection based on adaptive sampling of noisy preferences. In *International Conference on Machine Learning*, pages 1094–1102, 2013.

[11] Róbert Busa-Fekete, Eyke Hüllermeier, and Balázs Szörényi. Preference-based rank elicitation using statistical models: The case of mallows. In *Proceedings of The 31st International Conference on Machine Learning*, volume 32, 2014.

[12] Róbert Busa-Fekete, Balázs Szörényi, and Eyke Hüllermeier. Pac rank elicitation through adaptive sampling of stochastic pairwise preferences. In *AAAI*, pages 1701–1707, 2014.

[13] Xi Chen, Sivakanth Gopi, Jieming Mao, and Jon Schneider. Competitive analysis of the top-k ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1245–1264. SIAM, 2017.

[14] Xi Chen, Yuanzhi Li, and Jieming Mao. A nearly instance optimal algorithm for top-k ranking under the multinomial logit model. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2504–2522. SIAM, 2018.

[15] Yuxin Chen and Changho Suh. Spectral mle: Top-k rank aggregation from pairwise comparisons. In *International Conference on Machine Learning*, pages 371–380, 2015.

[16] Antoine Désir, Vineet Goyal, Srikanth Jagabathula, and Danny Segev. Assortment optimization under the mallows model. In *Advances in Neural Information Processing Systems*, pages 4700–4708, 2016.

[17] Miroslav Dudík, Katja Hofmann, Robert E Schapire, Aleksandrs Slivkins, and Masrour Zoghi. Contextual dueling bandits. In *Conference on Learning Theory*, pages 563–587, 2015.

[18] Moein Falahatgar, Yi Hao, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar. Maxing and ranking with few assumptions. In *Advances in Neural Information Processing Systems*, pages 7063–7073, 2017.

[19] Uriel Feige, Prabhakar Raghavan, David Peleg, and Eli Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001–1018, 1994.

[20] David F Gleich and Lek-heng Lim. Rank Aggregation via Nuclear Norm Minimization. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011.

[21] Bruce Hajek, Sewoong Oh, and Jiaming Xu. Minimax-optimal inference from partial rankings.

In *Advances in Neural Information Processing Systems*, pages 1475–1483, 2014.

[22] Kevin G Jamieson and Robert Nowak. Active Ranking using Pairwise Comparisons. In *Advances in Neural Information Processing Systems*, pages 2240–2248, 2011.

[23] Minje Jang, Sunghyun Kim, Changho Suh, and Sewoong Oh. Optimal sample complexity of m-wise data for top-k ranking. In *Advances in Neural Information Processing Systems*, pages 1685–1695, 2017.

[24] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, volume 12, pages 655–662, 2012.

[25] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

[26] Ashish Khetan and Sewoong Oh. Data-driven rank breaking for efficient rank aggregation. *Journal of Machine Learning Research*, 17(193):1–54, 2016.

[27] John I. Marden. *Analyzing and Modeling Rank Data*. Chapman and Hall/CRC, 1996.

[28] Soheil Mohajer, Changho Suh, and Adel Elmahdy. Active learning for top-k rank aggregation from noisy comparisons. In *International Conference on Machine Learning*, pages 2488–2497, 2017.

[29] Bernard Monjardet. On the Comparison of the Spearman and Kendall Metrics between Linear Orders. *Discrete mathematics*, 1998.

[30] Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative Ranking from Pair-wise Comparisons. In *Advances in Neural Information Processing Systems*, pages 2474–2482, 2012.

[31] Pantelimon G Popescu, Silvestru Dragomir, Emil I Slusanschi, and Octavian N Stanasila. Bounds for Kullback-Leibler divergence. *Electronic Journal of Differential Equations*, 2016, 2016.

[32] Arun Rajkumar and Shivani Agarwal. When Can We Rank Well from Comparisons of $O(n \log n)$ Non-Actively Chosen Pairs? In *Conference on Learning Theory*, pages 1376–1401, 2016.

[33] Siddartha Y Ramamohan, Arun Rajkumar, and Shivani Agarwal. Dueling bandits: Beyond condorcet winners to general tournament solutions. In *Advances in Neural Information Processing Systems*, pages 1253–1261, 2016.

[34] Balázs Szörényi, Róbert Busa-Fekete, Adil Paul, and Eyke Hüllermeier. Online rank elicitation for plackett-luce: A dueling bandits approach. In *Advances in Neural Information Processing Systems*, pages 604–612, 2015.

[35] Kalyan Talluri and Garrett Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50 (1):15–33, 2004.

[36] Tanguy Urvoy, Fabrice Clerot, Raphael Féraud, and Sami Naamane. Generic exploration and k-armed voting bandits. In *International Conference on Machine Learning*, pages 91–99, 2013.

[37] Fabian Wauthier, Michael Jordan, and Nebojsa Jojic. Efficient Ranking from Pairwise Comparisons. In *International Conference on Machine Learning*, pages 109–117, 2013.

[38] Yisong Yue and Thorsten Joachims. Beat the mean bandit. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 241–248, 2011.

Supplementary for Active Ranking with Subset-wise Preferences

A Appendix for Section 2

A.1 Proof of Lemma 4

Lemma 4. *Assume $\theta_i \in [a,b], \forall i \in [n]$, for any $a,b \in (0,1)$. If an algorithm is (ϵ,δ)*-**PAC-Rank**, *then it is also (ϵ',δ)-***PAC-Rank-Multiplicative** *for any $\epsilon' \le \frac{\epsilon}{4b}$. On the other hand, if an algorithm is (ϵ,δ)*-**PAC-Rank-Multiplicative**, *then it is also (ϵ',δ)-***PAC-Rank** *for any $\epsilon' \le 4a\epsilon(1+\epsilon)$.*

Recall that an algorithm is defined to be (ϵ,δ)**-PAC-Rank** (or (ϵ,δ)**-PAC-Rank**2) if it returns an ϵ-*Best-Ranking* (ϵ-*Best-Ranking-Multiplicative*) with probability $(1-\delta)$.

Proof. **Case 1.** Suppose the algorithm is (ϵ,δ)-**PAC-Rank**. So if $\boldsymbol{\sigma}$ is the ranking returned by it, with high probability $(1-\delta)$, $\nexists$ two items $i,j \in [n]$ such that $\sigma(i) > \sigma(j)$ but $\theta_i - \theta_j \ge \epsilon$. But then this implies, $\nexists$ two items $i,j \in [n]$ with $\sigma(i) > \sigma(j)$ such that

$$Pr(i|\{i,j\}) - \frac{1}{2} = \frac{\theta_i - \theta_j}{2(\theta_i - \theta_j)} \ge \frac{\theta_i - \theta_j}{4b} = \frac{\epsilon}{4b} \ge \epsilon',$$

which proves our first claim.

Case 2. Now suppose the algorithm is (ϵ,δ)**-PAC-Rank-Multiplicative**. So if $\boldsymbol{\sigma}$ is the ranking returned by it, with high probability $(1-\delta)$, $\nexists$ two items $i,j \in [n]$ such that $\sigma(i) > \sigma(j)$ but $Pr(i|\{i,j\}) - \frac{1}{2} \ge \epsilon$. But since $Pr(i|\{i,j\}) = \frac{\theta_i}{\theta_i+\theta_j}$, this them equivalently implies, $\nexists$ two items $i,j \in [n]$ with $\sigma(i) > \sigma(j)$ such that

$$\begin{aligned}
&\frac{\theta_i}{\theta_j} \ge \frac{1/2+\epsilon}{1/2-\epsilon}\\
&\implies \theta_i \ge \theta_j \left(\frac{1/2+\epsilon}{1/2-\epsilon}\right) \ge \theta_j\left(\frac{1}{2}+\epsilon\right)^2\\
&\implies \theta_i - \theta_j \ge \theta_j\big(4\epsilon^2+4\epsilon\big) \ge 4a\epsilon(1+\epsilon) \ge \epsilon',
\end{aligned}$$

which proves our second claim and concludes the proof.

□

A.2 Proof of Lemma 5

Lemma 5 (Pairwise win-probability estimates for the PL model)**.** *Consider a Plackett-Luce choice model with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$, and fix two items $i,j \in [n]$. Let $S_1, \ldots, S_T$ be a sequence of (possibly random) subsets of $[n]$ of size at least 2, where T is a positive integer, and $i_1, \ldots, i_T$ a sequence of random items with each $i_t \in S_t$, $1 \le t \le T$, such that for each $1 \le t \le T$, (a) S_t depends only on $S_1, \ldots, S_{t-1}$, and (b) i_t is distributed as the Plackett-Luce winner of the subset S_t, given $S_1, i_1, \ldots, S_{t-1}, i_{t-1}$ and S_t, and (c) $\forall t: \{i,j\} \subseteq S_t$ with probability 1. Let $n_i(T) = \sum_{t=1}^T \mathbf{1}(i_t = i)$ and $n_{ij}(T) = \sum_{t=1}^T \mathbf{1}(\{i_t \in \{i,j\}\})$. Then, for any positive integer v, and $\eta \in (0,1)$,*

$$\begin{aligned}
Pr\left(\frac{n_i(T)}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \ge \eta,\ n_{ij}(T) \ge v\right) \le e^{-2v\eta^2},\\
Pr\left(\frac{n_i(T)}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \le -\eta,\ n_{ij}(T) \ge v\right) \le e^{-2v\eta^2}.
\end{aligned}$$

Proof. We prove the lemma by using a coupling argument. Consider the following 'simulator' or probability space for the Plackett-Luce choice model that specifically depends on the item pair i,j, constructed as follows.

Let $Z_1, Z_2, \ldots$ be a sequence of iid Bernoulli random variables with success parameter $\theta_i/(\theta_i+\theta_j)$. A counter is first initialized to 0. At each time t, given $S_1, i_1, \ldots, S_{t-1}, i_{t-1}$ and S_t, an independent coin is tossed with probability of heads $(\theta_i+\theta_j)/\sum_{k\in S_t}\theta_k$. If the coin lands tails, then i_t is drawn as an independent sample from the Plackett-Luce distribution over $S_t \setminus \{i,j\}$, else, the counter is incremented by 1, and i_t is returned as i if $Z(C)=1$ or j if $Z(C)=0$ where C is the present value of the counter.

It may be checked that the construction above indeed yields the correct joint distribution for the sequence $i_1, S_1, \ldots, i_T, S_T$ as desired, due to the independence of irrelevant alternatives (IIA) property of the Plackett-Luce choice model:

$$Pr(i_t = i | i_t \in \{i,j\}, S_t) = \frac{Pr(i_t = i|S_t)}{Pr(i_t \in \{i,j\}|S_t)} = \frac{\theta_i/\sum_{k\in S_t}\theta_k}{(\theta_i+\theta_j)/\sum_{k\in S_t}\theta_k} = \frac{\theta_i}{\theta_i+\theta_j}.$$

Furthermore, $i_t \in \{i,j\}$ if and only if C is incremented at round t, and $i_t = i$ if and only if C is incremented at round t and $Z(C)=1$. We thus have

$$\begin{aligned}
&Pr\left(\frac{n_i(T)}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \ge \eta,\ n_{ij}(T)\ge v\right) = Pr\left(\frac{\sum_{\ell=1}^{n_{ij}(T)} Z_\ell}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \ge \eta,\ n_{ij}(T)\ge v\right)\\
&= \sum_{m=v}^{T} Pr\left(\frac{\sum_{\ell=1}^{n_{ij}(T)} Z_\ell}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \ge \eta,\ n_{ij}(T) = m\right)\\
&= \sum_{m=v}^{T} Pr\left(\frac{\sum_{\ell=1}^{m} Z_\ell}{m} - \frac{\theta_i}{\theta_i+\theta_j} \ge \eta,\ n_{ij}(T) = m\right)\\
&\overset{(a)}{=} \sum_{m=v}^{T} Pr\left(\frac{\sum_{\ell=1}^{m} Z_\ell}{m} - \frac{\theta_i}{\theta_i+\theta_j} \ge \eta\right) Pr\left(n_{ij}(T) = m\right)\\
&\overset{(b)}{\le} \sum_{m=v}^{T} Pr\left(n_{ij}(T) = m\right)\, e^{-2m\eta^2} \le e^{-2v\eta^2},
\end{aligned}$$

where (a) uses the fact that $S_1, \ldots, S_T, X_1, \ldots, X_T$ are independent of $Z_1, Z_2, \ldots,$, and so $n_{ij}(T) \in \sigma(S_1, \ldots, S_T, X_1, \ldots, X_T)$ is independent of $Z_1, \ldots, Z_m$ for any fixed m, and (b) uses Hoeffding's concentration inequality for the iid sequence Z_i.

Similarly, one can also derive

$$Pr\left(\frac{n_i(T)}{n_{ij}(T)} - \frac{\theta_i}{\theta_i+\theta_j} \le -\eta,\ n_{ij}(T)\ge v\right) \le e^{-2v\eta^2},$$

which concludes the proof. □

B Appendix for Section 4

B.1 Proof of Lemma 6

Lemma 6. *Consider a Plackett-Luce choice model with parameters $(\theta_1, \theta_2, \ldots, \theta_n)$, $n \ge 2$, and an item $b \in [n]$. Let $i_1, i_2, \ldots$ be a sequence of iid draws from the model. Let $\tau = \min\{t \ge \mathbb{N} \mid i_t = b\}$ be the first time at which b appears, and for each $i \ne b$, let $w_i(\tau) = \sum_{t=1}^{\tau} \mathbf{1}(i_t = i)$ be the number of times $i \ne b$ appears until time τ. Then, $\tau - 1$ and $w_i(\tau)$ are Geometric random variables with parameters $\frac{\theta_b}{\sum_{j\in[n]}\theta_j}$ and $\frac{\theta_b}{\theta_i+\theta_b}$, respectively.*

Proof. (1) follows from the simple observation probability of item b winning at any trial t is independent and identically distributed (*iid*) $Ber\left(\frac{\theta_b}{\sum_{j\in S}}\right)$ and T essentially denotes the number of trials till first success (win of b). (Recall from Section 2 that $Geo(p)$ denote the 'number of trials before success' version of the Geometric random variable with probability of success at each trial being $p \in [0,1]$).

We proof (2) by deriving the moment generating function (MGF) of the random variable $w_i(T)$ which gives:

Lemma 18 (MGF of $w_i(T)$)**.** *For any item $i \in S \setminus \{b\}$, the moment generating function of the random variable $w_i(T)$ is given by:* $\mathbf{E}\Big[e^{\lambda w_i(T)}\Big] = \frac{1}{1-\frac{\theta_i}{\theta_b}(e^{\lambda-1})}, \forall \ell \in [w_1],$ *for any* $\lambda \in \big(0, \ln(1+\eta)\big)$, *with* $\eta < \min_{j \in S} \frac{\theta_b}{\theta_j}$.

See Appendix B.2 for the proof. Now firstly recall that the MGF of any random variable $X \sim Geo(p)$ is given by $\mathbf{E}[e^{\lambda X}] = \frac{p}{(1-e^{\lambda}(1-p))}, \forall \lambda \in \Big(0, -\ln(1-p)\Big)$. In the current case $p = \frac{\theta_b}{\theta_b + \theta_i}$. Thus we have $\Bigg(\frac{1}{p} - 1\Bigg) = \frac{\theta_i}{\theta_b}$ or $\frac{1}{1-p} = \frac{\theta_i + \theta_b}{\theta_i}$, and the MGF holds good for any $\lambda \in (0, \ln(1+\eta))$ as long as $\eta < \min_{j \in S} \frac{\theta_b}{\theta_j}$.
The proof now follows from straightforward reduction of Lemma 18. Formally, we have:

$$\begin{aligned}
\mathbf{E}\Big[e^{\lambda X}\Big] &= \frac{p}{(1-e^{\lambda}(1-p))} \quad \text{for any } \lambda \in \big(0, \ln(1+\eta)\big), \text{ where } \eta < \min_{j \in S} \frac{\theta_b}{\theta_j}, \\
&= \frac{1}{\frac{1}{p} - e^{\lambda}\Big(\frac{1-p}{p}\Big)} = \frac{1}{1 + \frac{\theta_i}{\theta_b} - e^{\lambda}\big(\frac{\theta_i}{\theta_b}\big)} \\
&= \frac{1}{1 - \frac{\theta_i}{\theta_b}(e^{\lambda-1})} = \mathbf{E}\Big[e^{\lambda w_i(T)}\Big],
\end{aligned}$$

where the last equality follows from Lemma 18 as two random variables with same MGF must have same distributions. This concludes the proof. □

B.2 Proof of Lemma 18

Lemma 18 (MGF of $w_i(T)$)**.** *For any item $i \in S \setminus \{b\}$, the moment generating function of the random variable $w_i(T)$ is given by:* $\mathbf{E}\Big[e^{\lambda w_i(T)}\Big] = \frac{1}{1-\frac{\theta_i}{\theta_b}(e^{\lambda-1})}, \forall \ell \in [w_1],$ *for any* $\lambda \in \big(0, \ln(1+\eta)\big)$, *with* $\eta < \min_{j \in S} \frac{\theta_b}{\theta_j}$.

Proof. The proof follows from using standard MGF results of Bernoulli and Geometric random variables. We denote $S_{-b} = S \backslash \{b\}$, $\hat{T} = (T-1)$, $p = \frac{\theta_b}{\sum_{j \in S} \theta_j}$, and $p' = \frac{\theta_i}{\sum_{j \in S_{-b}} \theta_j}$. As argued in Lemma 6, we know that $\hat{T} \sim \text{Geo}\Big(p\Big)$. Also given a fixed (non-random) $\hat{T}$, $w_i(T) \sim \text{Bin}\Big(\hat{T}, p'\Big)$. Then using law of iterated expectation:

$$\begin{aligned}
\mathbf{E}\Big[e^{\lambda w_i(T)}\Big] &= \mathbf{E}_{\hat{T}}\Big[\mathbf{E}\big[e^{\lambda w_i(T)} \mid \hat{T}\big]\Big] \\
&= \mathbf{E}_{\hat{T}}\Big[(p'e^{\lambda} + 1 - p')^{\hat{T}}\Big],
\end{aligned}$$

where the last equality follows from the MGF of Binomial random variables. Note that, since $\lambda > 0$, we have $(p'e^{\lambda} + 1 - p') = 1 + p'(e^{\lambda} - 1) > 1$. Let us denote $\lambda' = \ln(1 + p'(e^{\lambda} - 1))$. Clearly $\lambda' > 0$ as both $\lambda, p' > 0$. Then from above equation, one can write:

$$\begin{aligned}
\mathbf{E}\Big[e^{\lambda w_i(T)}\Big] &= \mathbf{E}_{\hat{T}}\Big[e^{\lambda' \hat{T}}\Big], \\
&= \frac{p}{(1-e^{\lambda'}(1-p))} \\
&= \frac{p}{1-(1+p'(e^{\lambda}-1))(1-p)} = \frac{1}{1-\frac{\theta_i}{\theta_b}(e^{\lambda-1})},
\end{aligned}$$

where the second equality follows from the result that MGF of a geometric random variable $X \sim \text{Geo}(p)$ is: $\mathbf{E}[e^{\lambda' X}] = \frac{p}{(1-e^{\lambda'}(1-p))}, \forall \lambda' \in \Big(0, -\ln(1-p)\Big)$.

Thus the only remaining thing to show is λ' indeed satisfies the above range. As argues above, clearly $\lambda' > 0$ as both $\lambda, p' > 0$. To verify the upper bound, note that by choice $\lambda < \ln\Big(1+\frac{\theta_b}{\theta_j}\Big)$, $\forall j \in S$, which implies $e^{\lambda} < (1+\frac{\theta_b}{\theta_i})$, for any $i \in S_{-b}$. This further implies $(e^{\lambda}-1)\frac{\theta_i}{\theta_b} < 1 \implies (1-\frac{\theta_i}{\theta_b}(e^{\lambda}-1)) > 0 \implies (1-p)(1+p'(e^{\lambda}-1)) < 1$ rearranging which leads to the desired bound $\lambda' < -\ln(1-p)$ (recall $\lambda' = \ln(1+p'(e^{\lambda}-1))$, and thus the above MGF holds good. This concludes the proof. $\square$

B.3 Proof of Lemma 7

Lemma 7 (Concentration of Geometric Random Variables via the Negative Binomial distribution.)**.** *Suppose* $X_1, X_2, \ldots X_d$ *are* d *iid* $Geo(\frac{\theta_b}{\theta_b+\theta_i})$ *random variables, and* $Z = \sum_{i=1}^{d} X_i$*. Then, for any* $\eta > 0$*,* $Pr\Big(\Big|\frac{Z}{d} - \frac{\theta_i}{\theta_b}\Big| \ge \eta\Big) < 2\exp\Bigg(-\frac{2d\eta^2}{\big(1+\frac{\theta_i}{\theta_b}\big)^2\big(\eta+1+\frac{\theta_i}{\theta_b}\big)}\Bigg)$.

Proof. The result follows from the concentration of Geometric random variable as shown in [8]. Note that, Z denotes the number of trials needed to get n wins of item b, where the probability of success (i.e. item b winning) at each trial is $\frac{\theta_b}{\theta_b+\theta_i}$. Thus $Z \sim NB(n, \frac{\theta_b}{\theta_b+\theta_i})$. Clearly, by applying union bounding we get:

$$Pr\Big(\Big|\frac{Z}{n} - \frac{\theta_i}{\theta_b}\Big| > \eta\Big) \le Pr\Big(\frac{Z}{n} - \frac{\theta_i}{\theta_b} > \eta\Big) + Pr\Big(\frac{Z}{n} - \frac{\theta_i}{\theta_b} < -\eta\Big).$$

Let us start by analysing the first term $Pr\Big(\frac{Z}{n} - \frac{\theta_i}{\theta_b} > \eta\Big)$.

$$\begin{aligned}
Pr\Big(\frac{Z}{n} - \frac{\theta_i}{\theta_b} > \eta\Big) &= Pr\Big(Z > n\frac{\theta_i}{\theta_b} + n\eta\Big) \\
&\le Pr\Bigg(Bin\Big(n(\frac{\theta_i}{\theta_b}+1) + n\eta, \frac{\theta_b}{\theta_b+\theta_i}\Big) < n\Bigg) \\
&\le Pr\Bigg(Bin\Big(n(\frac{\theta_i}{\theta_b}+1) + n\eta, \frac{\theta_b}{\theta_b+\theta_i}\Big) - \Big[n(\frac{\theta_i}{\theta_b}+1) + n\eta\Big]\frac{1}{1+\frac{\theta_i}{\theta_b}} < -\frac{n\eta}{1+\frac{\theta_i}{\theta_b}}\Bigg) \\
&\le \exp\Big(-\frac{2}{m}\tilde{\eta}^2\Big) = \exp\Bigg(-\frac{2n\eta^2}{\big(1+\frac{\theta_i}{\theta_b}\big)^2\big(\eta+1+\frac{\theta_i}{\theta_b}\big)}\Bigg)
\end{aligned} \tag{1}$$

where the last inequality follows simply apply Hoeffding's inequality with $m = n(\frac{\theta_i}{\theta_b}+1) + n\eta$ and $\tilde{\eta} = \frac{n\eta}{1+\frac{\theta_i}{\theta_b}}$. Using a similar derivation as before, one can also show:

$$Pr\Big(\frac{Z}{n} - \frac{\theta_i}{\theta_b} < -\eta\Big) \le \exp\Bigg(-\frac{2n\eta^2}{\big(1+\frac{\theta_i}{\theta_b}\big)^2\big(\eta+1+\frac{\theta_i}{\theta_b}\big)}\Bigg) \tag{2}$$

The result now follows combining (1) and (2). $\square$

C Appendix for Section 5

C.1 Pseudocode for algorithms

Algorithm 2 *Find-the-Pivot* subroutine

Input:
  Set of items: $[n]$, Subset size: $n \geq k > 1$
  Error bias: $\epsilon > 0$, confidence parameter: $\delta > 0$
Initialize:
  $r_1 \leftarrow$ Any (random) item from $[n]$, $\mathcal{A} \leftarrow$ Randomly select $(k-1)$ items from $[n] \setminus \{r_1\}$
  Set $\mathcal{A} \leftarrow \mathcal{A} \cup \{r_1\}$, and $S \leftarrow [n] \setminus \mathcal{A}$
while $\ell = 1, 2, \ldots$ **do**
  Play the set $\mathcal{A}$ for $t := \frac{2k}{\epsilon^2} \ln \frac{2n}{\delta}$ rounds
  $w_i \leftarrow \#$ (i won in t plays of $\mathcal{A}$), $\forall i \in \mathcal{A}$
  $c_\ell \leftarrow \underset{i \in \mathcal{A}}{\text{argmax}}\, w_i$; $\hat{p}_{ij} \leftarrow \frac{w_i}{w_i + w_j}$, $\forall i, j \in \mathcal{A}, i \neq j$
  if $\hat{p}_{c_\ell, r_\ell} > \frac{1}{2} + \frac{\epsilon}{2}$: $r_{\ell+1} \leftarrow c_\ell$; **else** $r_{\ell+1} \leftarrow r_\ell$
  if $(S == \emptyset)$ **then**
    Break (exit the while loop)
  else if $|S| < k-1$ **then**
    $\mathcal{A} \leftarrow$ Select $(k-1-|S|)$ items from $\mathcal{A} \setminus \{r_\ell\}$ uniformly at random, $\mathcal{A} \leftarrow \mathcal{A} \cup \{r_\ell\} \cup S$; $S \leftarrow \emptyset$
  else
    $\mathcal{A} \leftarrow$ Select $(k-1)$ items from S uniformly at random, $\mathcal{A} \leftarrow \mathcal{A} \cup \{r_\ell\}$; $S \leftarrow S \setminus \mathcal{A}$
  end if
end while
Output: The item r_ℓ

Algorithm 3 *Score-and-Rank*

Input:
  Set of item: $[n]$ $(n \geq k)$, and subset size: k
  Error bias: $\epsilon > 0$, confidence parameter: $\delta > 0$
Initialize:
  $\epsilon_b \leftarrow \min(\frac{\epsilon}{2}, \frac{1}{2})$, $b \leftarrow$ *Find-the-Pivot*$(n, k, \epsilon_b, \frac{\delta}{4})$
  Set $S \leftarrow [n] \setminus \{b\}$, and divide S into $G := \lceil \frac{n-1}{k-1} \rceil$ sets $\mathcal{G}_1, \mathcal{G}_2, \cdots \mathcal{G}_G$ such that $\cup_{j=1}^{G} \mathcal{G}_j = S$ and $\mathcal{G}_j \cap \mathcal{G}_{j'} = \emptyset$, $\forall j, j' \in [G]$, $|\mathcal{G}_j| = (k-1)$, $\forall j \in [G-1]$
  If $|\mathcal{G}_G| < (k-1)$, **then** set $\mathcal{R} \leftarrow \mathcal{G}_G$, and $S \leftarrow S \setminus \mathcal{R}$, $S' \leftarrow$ Randomly sample $(k-1-|\mathcal{G}_G|)$ items from S, and set $\mathcal{G}_G \leftarrow \mathcal{G}_G \cup S'$
  Set $\mathcal{G}_j = \mathcal{G}_j \cup \{b\}$, $\forall j \in [G]$
for $g = 1, 2, \ldots, G$ **do**
  Set $\epsilon' \leftarrow \frac{\epsilon}{24}$ and $\delta' \leftarrow \frac{\delta}{8n}$
  repeat
    Play $\mathcal{G}_g$ and observe the winner.
  until b is chosen for $t = \frac{1}{\epsilon'^2} \ln \frac{1}{\delta'}$ times
  If $w_i \leftarrow$ is the total number of wins of item i in $\mathcal{G}_g$, set $\hat{\theta}_i^b \leftarrow \frac{w_i}{t}$, $\forall i \in \mathcal{G}_g \setminus \{b\}$
end for
Choose $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$, such that $\sigma(b) = 1$ and $\boldsymbol{\sigma}(i) < \boldsymbol{\sigma}(j)$ if $\hat{\theta}_i^b > \hat{\theta}_j^b$, $\forall i, j \in S \cup \mathcal{R}$
Output: The ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$

C.2 Proof of Lemma 10

Lemma 10 (*Find-the-Pivot*: Correctness and Sample Complexity with WI)**.** *Find-the-Pivot (Algorithm 2) achieves the (ϵ, δ)-PAC objective with sample complexity $O(\frac{n}{\epsilon^2} \log \frac{n}{\delta})$.*

Proof. We start by analyzing the required sample complexity first. Note that the 'while loop' of Algorithm 2 always discards away $k-1$ items per iteration. Thus, n being the total number of items the loop can be executed is at most for $\lceil \frac{n}{k-1} \rceil$ many number of iterations. Clearly, the sample complexity of each iteration being $t = \frac{2k}{\epsilon^2}\ln\frac{2\delta}{n}$, the total sample complexity of the algorithm becomes $\big(\lceil \frac{n}{k-1} \rceil\big)\frac{2k}{\epsilon^2}\ln\frac{n}{2\delta} \le \big(\frac{n}{k-1}+1\big)\frac{2k}{\epsilon^2}\ln\frac{n}{2\delta} = \big(n + \frac{n}{k-1} + k\big)\frac{2}{\epsilon^2}\ln\frac{n}{2\delta} = O(\frac{n}{\epsilon^2}\ln\frac{n}{\delta})$.

We now prove the (ϵ,δ)-PAC correctness of the algorithm. As argued before, the 'while loop' of Algorithm 2 can run for maximum $\lceil \frac{n}{k-1} \rceil$ many number of iterations. We denote the iterations by $\ell = 1, 2, \ldots \lceil \frac{n}{k-1} \rceil$, and the corresponding set $\mathcal{A}$ of iteration ℓ by $\mathcal{A}_\ell$.

Note that our idea is to retain the estimated best item in 'running winner' r_ℓ and compare it with the 'empirical best item' c_ℓ of $\mathcal{A}_\ell$ at every iteration ℓ. The crucial observation lies in noting that at any iteration ℓ, r_ℓ gets updated as follows:

Lemma 19. *At any iteration $\ell = 1, 2 \ldots \lfloor \frac{n}{k-1} \rfloor$, with probability at least $(1 - \frac{\delta}{2n})$, Algorithm 2 retains $r_{\ell+1} \leftarrow r_\ell$ if $p_{c_\ell r_\ell} \le \frac{1}{2}$, and sets $r_{\ell+1} \leftarrow c_\ell$ if $p_{c_\ell r_\ell} \ge \frac{1}{2} + \epsilon$.*

Proof. Consider any set $\mathcal{A}_\ell$, by which we mean the state of $\mathcal{A}$ in the algorithm at iteration ℓ. The crucial observation to make is that since c_ℓ is the empirical winner of t rounds, then $w_{c_\ell} \ge \frac{t}{k}$. Thus $w_{c_\ell} + w_{r_\ell} \ge \frac{t}{k}$. Let $n_{ij} := w_i + w_j$ denotes the total number of pairwise comparisons between item i and j in t rounds, for any $i, j \in \mathcal{A}_\ell$. Then clearly, $0 \le n_{ij} \le t$ and $n_{ij} = n_{ji}$. Specifically we have $\hat{p}_{r_\ell c_\ell} = \frac{w_{r_\ell}}{w_{r_\ell} + w_{c_\ell}} = \frac{w_{r_\ell}}{n_{r_\ell c_\ell}}$. We prove the claim by analyzing the following cases:

Case 1. (If $p_{c_\ell r_\ell} \le \frac{1}{2}$, *Find-the-Pivot* retains $r_{\ell+1} \leftarrow r_\ell$): Note that *Find-the-Pivot* replaces $r_{\ell+1}$ by c_ℓ only if $\hat{p}_{c_\ell, r_\ell} > \frac{1}{2} + \frac{\epsilon}{2}$, but this happens with probability:

$$
\begin{aligned}
& Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} > \frac{1}{2} + \frac{\epsilon}{2}\bigg\}\Bigg) \\
& = Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} > \frac{1}{2} + \frac{\epsilon}{2}\bigg\} \cap \bigg\{n_{c_\ell r_\ell} \ge \frac{t}{k}\bigg\}\Bigg) + \cancelto{0}{Pr\bigg\{n_{c_\ell r_\ell} < \frac{t}{k}\bigg\}} Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} > \frac{1}{2} + \frac{\epsilon}{2}\bigg\}\bigg|\bigg\{n_{c_\ell r_\ell} < \frac{t}{k}\bigg\}\Bigg) \\
& \le Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} - p_{c_\ell r_\ell} > \frac{\epsilon}{2}\bigg\} \cap \bigg\{n_{c_\ell r_\ell} \ge \frac{t}{k}\bigg\}\Bigg) \le \exp\Big(-2\frac{t}{k}\Big(\frac{\epsilon}{2}\Big)^2\Big) = \frac{\delta}{2n},
\end{aligned}
$$

where the first inequality follows as $p_{c_\ell r_\ell} \le \frac{1}{2}$, and the second inequality is by applying Lemma 5 with $\eta = \frac{\epsilon}{2}$ and $v = \frac{t}{k}$. We now proceed to the second case:

Case 2. (If $p_{c_\ell r_\ell} \ge \frac{1}{2} + \epsilon$, *Find-the-Pivot* sets $r_{\ell+1} \leftarrow c_\ell$): Recall again that *Find-the-Pivot* retains $r_{\ell+1} \leftarrow r_\ell$ only if $\hat{p}_{c_\ell, r_\ell} \le \frac{1}{2} + \frac{\epsilon}{2}$. This happens with probability:

$$
\begin{aligned}
& Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} \le \frac{1}{2} + \frac{\epsilon}{2}\bigg\}\Bigg) \\
& = Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} \le \frac{1}{2} + \frac{\epsilon}{2}\bigg\} \cap \bigg\{n_{c_\ell r_\ell} \ge \frac{t}{k}\bigg\}\Bigg) + \cancelto{0}{Pr\bigg\{n_{c_\ell r_\ell} < \frac{t}{k}\bigg\}} Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} \le \frac{1}{2} + \frac{\epsilon}{2}\bigg\}\bigg|\bigg\{n_{c_\ell r_\ell} < \frac{t}{k}\bigg\}\Bigg) \\
& = Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} \le \frac{1}{2} + \epsilon - \frac{\epsilon}{2}\bigg\} \cap \bigg\{n_{c_\ell r_\ell} \ge \frac{t}{k}\bigg\}\Bigg) \\
& \le Pr\Bigg(\bigg\{\hat{p}_{c_\ell r_\ell} - p_{c_\ell r_\ell} \le -\frac{\epsilon}{2}\bigg\} \cap \bigg\{n_{c_\ell r_\ell} \ge \frac{t}{k}\bigg\}\Bigg) \le \exp\Big(-2\frac{t}{k}\Big(\frac{\epsilon}{2}\Big)^2\Big) = \frac{\delta}{2n},
\end{aligned}
$$

where the first inequality holds as $p_{c_\ell r_\ell} \ge \frac{1}{2} + \epsilon$, and the second one by applying Lemma 5 with $\eta = \frac{\epsilon}{2}$ and $v = \frac{t}{k}$. Combining the above two cases concludes the proof. $\square$

Given Algorithm 2 satisfies Lemma 19, and taking union bound over $(k-1)$ elements in $\mathcal{A}_\ell \setminus \{r_\ell\}$, we get that with probability at least $(1 - \frac{(k-1)\delta}{2n})$,

$$p_{r_{\ell+1}r_\ell} \geq \frac{1}{2} \text{ and, } p_{r_{\ell+1}c_\ell} \geq \frac{1}{2} - \epsilon. \tag{3}$$

Above suggests that for each iteration ℓ, the estimated 'best' item r_ℓ only gets improved as $p_{r_{\ell+1}r_\ell} \geq \frac{1}{2}$. Let, ℓ_* denotes the specific iteration such that $1 \in \mathcal{A}_\ell$ for the first time, i.e. $\ell_* = \min\{\ell \mid 1 \in \mathcal{A}_\ell\}$. Clearly $\ell_* \leq \lceil \frac{n}{k-1} \rceil$. Now (3) suggests that with probability at least $(1 - \frac{(k-1)\delta}{2n})$, $p_{r_{\ell_*+1}1} \geq \frac{1}{2} - \epsilon$. Moreover (3) also suggests that for all $\ell > \ell_*$, with probability at least $(1 - \frac{(k-1)\delta}{2n})$, $p_{r_{\ell+1}r_\ell} \geq \frac{1}{2}$, which implies for all $\ell > \ell_*$, $p_{r_{\ell+1}1} \geq \frac{1}{2} - \epsilon$ as well – This holds due to the following transitivity property of the Plackett-Luce model: For any three items $i_1, i_2, i_3 \in [n]$, if $p_{i_1 i_2} \geq \frac{1}{2}$ and $p_{i_2 i_3} \geq \frac{1}{2}$, then we have $p_{i_1 i_3} \geq \frac{1}{2}$ as well.

This argument finally leads to $p_{r_* 1} \geq \frac{1}{2} - \epsilon$. Since failure probability at each iteration ℓ is at most $\frac{(k-1)\delta}{2n}$, and Algorithm 2 runs for maximum $\lceil \frac{n}{k-1} \rceil$ many number of iterations, using union bound over ℓ, the total failure probability of the algorithm is at most $\lceil \frac{n}{k-1} \rceil \frac{(k-1)\delta}{2n} \leq (\frac{n}{k-1} + 1)\frac{(k-1)\delta}{2n} = \delta\Big(\frac{n+k-1}{2n}\Big) \leq \delta$ (since $k \leq n$). This concludes the correctness of the algorithm showing that it indeed satisfies the (ϵ, δ)-PAC objective. $\square$

C.3 Proof of Theorem 8

Theorem 8 (*Beat-the-Pivot*: Correctness and Sample Complexity)**.** *Beat-the-Pivot (Algorithm 1) is* (ϵ, δ)-**PAC-Rank** *with sample complexity* $O\big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\big)$.

Proof. We first analyze the sample complexity of *Beat-the-Pivot*. Clearly, there are at most $G = \lceil \frac{n-1}{k-1} \rceil \leq \frac{n-1}{k-1} + 1 \leq \frac{2(n-1)}{k-1}$ groups. Here the last inequality follows since $n \geq k$. Now each group $\mathcal{G}_g$ (set of k items) is played (queried) for at most $t := \frac{2k}{\epsilon'^2} \log \frac{1}{\delta'}$ times, which gives the total sample complexity of the algorithm to be

$$G * t \leq \frac{2(n-1)}{(k-1)} * \frac{2k}{\epsilon'^2} \log \frac{1}{\delta'} = \frac{2048(n-1)}{\epsilon^2} \log \frac{8n}{\delta} = O\Big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\Big),$$

where we used the fact $k \geq 2$ and bound $\frac{k}{k-1} \leq 2$.

Moreover the sample complexity of *Find-the-Pivot* is also $O\Big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\Big)$ as proved in Lemma 10. Combining this with above thus makes the total sample complexity of *Beat-the-Pivot* $O\Big(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\Big)$.

We are now only left to show the correctness of the algorithm, i.e. *Beat-the-Pivot* indeed (ϵ, δ)-**PAC-Rank** in the above sample complexity. We start by proving the following lemma which would be crucial throughout the analysis. Let us first denote $\Delta^b_{ij} = P(i \succ b) - P(j \succ b)$, for any i and $j \in [n]$.

Lemma 20. *If b is the pivot-item returned by Algorithm 2 (Line 5 of Beat-the-Pivot), then for any two items $i, j \in [n]$, such that $\theta_i \geq \theta_j$, $\frac{(\theta_i - \theta_j)}{8} \leq \Delta^b_{ij} \leq 4(\theta_i - \theta_j)$, with probability at least $\big(1 - \frac{\delta}{2}\big)$.*

Proof. First let us assume if $b = 1$. Then $\Delta^b_{ij} = P(i \succ 1) - P(j \succ 1) = \frac{\theta_1(\theta_i - \theta_j)}{(\theta_j + \theta_1)(\theta_i + \theta_1)} \geq \frac{(\theta_i - \theta_j)}{4}$, since $\theta_1 = 1$ and $\theta_i \leq 1,\ \forall i \in [n]$. On the other hand, we also have $\Delta^b_{ij} = P(i \succ 1) - P(j \succ 1) = \frac{\theta_1(\theta_i - \theta_j)}{(\theta_j + \theta_1)(\theta_i + \theta_1)} \leq (\theta_i - \theta_j) = \epsilon$, since $\theta_1 = 1$ and $\theta_i \geq 0, \forall i \in [n]$.

However even if $b \neq 1$, with high probability $(1 - \frac{\delta}{2})$, it is ensured that $\theta_b > \theta_1 - \frac{1}{2}$, as with high probability $(1 - \frac{\delta}{2})$, Algorithm 2 returns an ϵ-*Best-Item* (see Lemma 10, proof in Appendix C.2). Then similarly as before, $\Delta^b_{ij} = P(i \succ 1) - P(j \succ 1) = \frac{\theta_b(\theta_i - \theta_j)}{(\theta_j + \theta_b)(\theta_i + \theta_b)} \geq \frac{(\theta_i - \theta_j)}{8}$, $\theta_b \geq \theta_1 - \frac{1}{2} = \frac{1}{2}$, and $\theta_i \leq 1,\ \forall i \in [n]$. On the other hand, we also have $\Delta^b_{ij} = P(i \succ b) - P(j \succ b) = \frac{\theta_b(\theta_i - \theta_j)}{(\theta_j + \theta_b)(\theta_i + \theta_b)} \leq 4(\theta_i - \theta_j) = \epsilon$, since $\theta_b \in \big(\frac{1}{2}, 1]$, and $\theta_i \geq 0,\ \forall i \in [n]$. This proves our claim. $\square$

Now to ensure the correctness of *Beat-the-Pivot*, recall that all we need to show it returns an ϵ-*Best-Ranking* $\boldsymbol{\sigma} \in \Sigma_{[n]}$. The main idea is to plug in the pivot item b in every group $\mathcal{G}_g$ and estimate the pivot-preference score $p_{ib} = Pr(i \succ b)$ of every item $i \notin \mathcal{G}_g \setminus \{b\}$, i.e. with respect to the pivot item b. We finally output the ranking simply sorting the items w.r.t. p_{ib} – the intuition is if item i beats j in terms of their actual BTL scores (i.e. $\theta_i > \theta_j$), then i beats j in terms of their pivot-preference scores as well (i.e. $p_{ib} > p_{jb}$).

More formally, as $\boldsymbol{\sigma}$ denotes the ranking returned by *Beat-the-Pivot*, the algorithm fails if $\boldsymbol{\sigma}$ is not ϵ-*Best-Ranking*. We denote by $Pr_b(\cdot) = Pr(\cdot | b \text{ is } \epsilon_b\text{-}\textit{Best-Item})$ the probability of an event conditioned on the event that b is indeed an ϵ_b-*Best-Item* (Recall we have set $\epsilon_b = \min(\frac{\epsilon}{2}, \frac{1}{2})$). Formally, we have:

$$
\begin{aligned}
Pr_b(\text{ \textit{Beat-the-Pivot} fails }) &= Pr_b(\exists i, j \in [n] \mid \theta_i > \theta_j + \epsilon \text{ but } \sigma(i) > \sigma(j)) \qquad (4)\\
&= Pr_b(\exists i, j \in [n] \mid \theta_i > \theta_j + \epsilon \text{ but } \hat{p}_{ib} < \hat{p}_{jb})
\end{aligned}
$$

Now, assuming b to be indeed an ϵ_b-*Best-Item*, since $\theta_i > \theta_j \implies \Delta_{ij}^b \geq \frac{\epsilon}{8}$ (from Lemma 20), from Eqn. 4, we further get:

$$
\begin{aligned}
Pr_b(\text{ \textit{Beat-the-Pivot} fails }) &= Pr_b(\exists i, j \in [n] \mid \theta_i > \theta_j + \epsilon \text{ but } \sigma(i) > \sigma(j))\\
&\leq Pr_b\Big(\exists i, j \in [n] \setminus \{b\} \mid p_{ib} > p_{jb} + \frac{\epsilon}{8} \text{ but } \sigma(i) > \sigma(j)\Big)\\
&= Pr_b(\exists i, j \in [n] \setminus \{b\} \mid \Delta_{ij}^b > \frac{\epsilon}{8} \text{ but } \hat{p}_{ib} < \hat{p}_{jb}), \qquad (5)
\end{aligned}
$$

where the inequality follows due to Lemma 20. In the inequality of the above analysis, it is also crucial to note that under the assumption of b to be indeed an ϵ_b-*Best-Item* setting $\sigma(1) = b$ does not incur an error since $\theta_b > \theta_1 - \frac{\epsilon}{2}$. So if we can estimate each p_{ib} within a confidence interval of $\frac{\epsilon}{16}$, that should be enough to ensure correctness of the algorithm. Thus the only thing remaining to show is *Beat-the-Pivot* indeed estimates p_{ib} tightly enough with high confidence – formally, it is enough to show that for any group $g \in [G]$ and any item $i \in \mathcal{G}_g \setminus \{b\}$, $Pr_b\Big(|p_{ib} - \hat{p}_{ib}| > \frac{\epsilon}{16}\Big) \leq \frac{\delta}{4n}$.

We prove this using the following two lemmas. We first show that in any set $\mathcal{G}_g$, if it is played for m times, then with high probability of at least $(1-\delta)$, the pivot item would gets selected at least for $\frac{m}{4k}$ times. Formally:

Lemma 21. *Conditioned on the event that b is indeed an ϵ_b-*Best-Item*, for any group $g \in [G]$ with probability at least $\Big(1 - \dfrac{\delta}{8n}\Big)$, the empirical win count $w_b > (1-\eta)\frac{t}{2k}$, for any $\eta \in \big(\frac{1}{8\sqrt{2}}, 1\big]$.*

Proof. We will assume the event that b to be indeed an ϵ_b-*Best-Item* throughout the proof and use the shorthand notation $Pr_b(\cdot)$ as defined earlier. The proof now follows from an straightforward application of Chernoff-Hoeffding's inequality [6]. Recall that the algorithm plays each set $\mathcal{G}_g$ for $t = \frac{2k}{\epsilon'^2} \ln \frac{1}{\delta'}$ number of times. Now consider a fixed group $g \in [G]$ and let i_τ denotes the winner of the τ-th play of $\mathcal{G}_g$, $\tau \in [t]$. Clearly, for any item $i \in \mathcal{G}_g$, $w_i = \sum_{\tau=1}^t \mathbf{1}(i_\tau == i)$, where $\mathbf{1}(i_\tau == i)$ is a Bernoulli random variable with parameter $\frac{\theta_i}{\sum_{j \in \mathcal{G}_g} \theta_j}$, $\forall \tau \in [t]$, just by the definition of the PL query model with winner information (WI) (Sec. 3.1). Thus the random variable $w_i \sim Bin\Big(t, \frac{\theta_i}{\sum_{j \in \mathcal{G}_g} \theta_j}\Big)$. In particular, for the pivot item, $i = b$, we have $Pr_b(\{i_\tau = b\}) = \frac{\theta_b}{\sum_{j \in \mathcal{G}_g} \theta_j} \geq \dfrac{\frac{1}{2}}{k}$. Hence $\mathbf{E}[w_b] = \sum_{\tau=1}^t \mathbf{E}[\mathbf{1}(i_\tau == b)] \geq \frac{t}{2k}$. Now applying multiplicative Chernoff-Hoeffdings bound for w_b, we get that for any $\eta \in \Big(\frac{1}{8}, 1\Big]$,

$$
\begin{aligned}
Pr_b\Big(w_b \leq (1-\eta)\mathbf{E}[w_b]\Big) &\leq \exp\Big(-\frac{\mathbf{E}[w_b]\eta^2}{2}\Big) \leq \exp\Big(-\frac{t\eta^2}{4k}\Big), \ \Big(\text{since } \mathbf{E}[w_b] \geq \frac{t}{2k}\Big)\\
&\leq \exp\Big(-\frac{\eta^2}{2\epsilon'^2} \ln\Big(\frac{1}{\delta'}\Big)\Big) \leq \exp\Big(-\ln\Big(\frac{1}{\delta'}\Big)\Big) \leq \frac{\delta}{8n},
\end{aligned}
$$

where the second last inequality holds for any $\eta > \frac{1}{8\sqrt{2}}$ as it has to be the case that $\epsilon' < \frac{1}{16}$ since $\epsilon \in (0,1)$. Thus for any $\eta > \frac{1}{8\sqrt{2}}$, $\eta^2 \geq 4\epsilon'^2$. $\square$

In particular, choosing $\eta = \frac{1}{2}$ in Lemma 21, we have with probability at least $\bigg(1 - \dfrac{\delta}{8n}\bigg)$, the empirical win count of the pivot element b is at least $w_b > \frac{t}{4k}$. We next proof under $w_b > \frac{t}{4k}$, the estimate of pivot-preference scores p_{ib} can not be too bad for any item $i \in \mathcal{G}_g$ at any group $g \in [G]$. The formal statement is given in Lemma 22. For the ease of notation we define the event $\mathcal{E}_g := \{\exists i \in \mathcal{G}_g \setminus \{b\} \text{ s.t. } |p_{ib} - \hat p_{ib}| > \frac{\epsilon}{16}\}$.

Lemma 22. *Conditioned on the event that b is indeed an ϵ_b-*Best-Item*, for any group $g \in [G]$, $Pr_b\Big(\mathcal{E}_g\Big) \le \frac{k\delta}{4n}$.*

Proof. We will again assume the event that b to be indeed an ϵ_b-*Best-Item* throughout the proof and use the shorthand notation $Pr_b(\cdot)$ as defined previously. Let us first fix a group $g \in [G]$. We find convenient to define the event $\mathcal{F}_g = \{w_b \ge \frac{t}{4k} \text{ for group } \mathcal{G}_g\}$ and denote by $n_{ib}^g = w_i + w_b$ the total number of times item i and b has won in group $\mathcal{G}_g$. Clearly, $n_{ib}^g \le t$, moreover under $\mathcal{F}_g$, $n_{ib}^g \ge \frac{t}{4k}$, $\forall i \in \mathcal{G}_g$. Then for any item $i \in \mathcal{G}_g \setminus \{b\}$,

$$\begin{aligned} Pr_b\Bigg(\bigg\{|p_{ib} - \hat p_{ib}| > \frac{\epsilon}{16}\bigg\} \cap \mathcal{F}_g\Bigg) &\le Pr_b\Bigg(\bigg\{|p_{ib} - \hat p_{ib}| > \frac{\epsilon}{16}\bigg\} \cap \bigg\{n_{ib}^g \ge \frac{t}{4k}\bigg\}\Bigg) \\ &\le 2\exp\Big(-2\frac{t}{4k}\big(\frac{\epsilon}{16}\big)^2\Big) = \frac{\delta}{4n}, \end{aligned} \tag{6}$$

where the first inequality follows since $\mathcal{F}_g \implies n_{ib}^g \ge \frac{t}{4k}$, the second inequality holds due to Lemma 5 with $\eta = \frac{\epsilon}{16}$, and $v = \frac{t}{4k}$. Its crucial to note that while applying 5, we can so we can drop the notation $Pr_b(\cdot)$ as the event $\{|p_{ib} - \hat p_{ib}| > \frac{\epsilon}{16}\} \cap \bigg\{n_{ib}^g \ge \frac{t}{4k}\bigg\}$ is independent of b to be ϵ_b-*Best-Item* or not.

Then probability that *Beat-the-Pivot* fails to estimate the pivot-preference scores p_{ib} for group $\mathcal{G}_g$,

$$\begin{aligned} Pr_b(\mathcal{E}_g) &= Pr_b(\mathcal{E}_g \cap \mathcal{F}_g) + Pr_b(\mathcal{E}_g \cap \mathcal{F}_g^c) \\ &\le Pr_b(\mathcal{E}_g \cap \mathcal{F}_g) + Pr_b(\mathcal{F}_g^c) \\ &\le Pr_b(\mathcal{E}_g \cap \mathcal{F}_g) + \frac{\delta}{8n} \quad \Big(\text{From Lemma 21}\Big) \\ &\le \sum_{i \in \mathcal{G}_g \setminus \{b\}} Pr_b\Bigg(\bigg\{|p_{ib} - \hat p_{ib}| > \frac{\epsilon}{16}\bigg\} \cap \mathcal{F}_g\Bigg) + \frac{\delta}{8n} \\ &= (k-1)\frac{\delta}{4n} + \frac{\delta}{4n} \le \frac{k\delta}{4n}, \end{aligned}$$

where the last inequality follows by taking union bound. The last equality follows from (6) and hence the proof follows. $\square$

Thus using Lemma 22 and from (5) we get,

$$\begin{aligned} Pr_b(\textit{Beat-the-Pivot} \text{ fails }) &\le Pr_b(\exists i, j \in [n] \setminus \{b\} \mid \Delta_{ij}^b > \frac{\epsilon}{8} \text{ but } \hat p_{ib} < \hat p_{jb}) \\ \le Pr_b(\exists g \in [G] \text{ s.t. } \mathcal{E}_g) &= \Big(\Big\lceil \frac{n-1}{k-1} \Big\rceil\Big)\frac{k\delta}{4n} \le 2\Big(\frac{n-1}{k-1}\Big)\frac{k\delta}{4n} \le \frac{\delta}{2}, \end{aligned} \tag{7}$$

where the last inequality follows taking union bound over all groups $g \in [G]$. Finally analysing all the previous claims together:

$$\begin{aligned} & Pr(\textit{Beat-the-Pivot} \text{ fails }) \\ & \le Pr(\textit{Beat-the-Pivot} \text{ fails } \big| b \text{ is an } \epsilon_b\text{-}\textit{Best-Item}) Pr(b \text{ is an } \epsilon_b\text{-}\textit{Best-Item}) + Pr(b \text{ is not an } \epsilon_b\text{-}\textit{Best-Item}) \\ & \le Pr_b(\textit{Beat-the-Pivot} \text{ fails })\Big(1 - \frac{\delta}{2}\Big) + \frac{\delta}{2} \le \frac{\delta}{2} + \frac{\delta}{2} = \delta, \end{aligned}$$

where the last inequality follows from (7), which concludes the proof. $\square$

C.4 Proof of Theorem 9

Theorem 9 (*Score-and-Rank*: Correctness and Sample Complexity)**.** *Score-and-Rank (Algorithm 3) is* (ϵ, δ)**-PAC-Rank** *with sample complexity* $O\big(\frac{n}{\epsilon^2}\log\frac{n}{\delta}\big)$.

Proof. Before proving the sample complexity, we first show the correctness of the algorithm, i.e. *Beat-the-Pivot* is indeed (ϵ, δ)**-PAC-Rank**. The following lemma would be crucially used throughout the proof analysis. Let us first denote $\theta_i^b = \frac{\theta_i}{\theta_b}$ the score of item $i \in [n]$ with respect to that of item b, we will term it as *pivotal-score* of item i. Also let $\theta_{ij}^b = \theta_i^b - \theta_j^b$, for any i and $j \in [n]$. It is easy to note that since with high probability $(1-\frac{\delta}{4})$, $\frac{1}{2} \le \theta_b \le 1$ (Lemma 10), and hence $\theta_i \le \theta_i^b \le 2\theta_i$. This further leads to the following claim:

Lemma 23. *If b is the pivot-item returned by Algorithm 2 (Line 5 of Score-and-Rank), then for any two items $i, j \in [n]$, such that $\theta_i \ge \theta_j$, $(\theta_i - \theta_j) \le \theta_{ij}^b \le 2(\theta_i - \theta_j)$, with probability at least $\big(1-\frac{\delta}{4}\big)$.*

Proof. First let us assume if $b = 1$, which implies $\theta_i^b = \frac{\theta_i}{\theta_b} = \theta_i$ and the claims holds trivially.

Now let us assume that $b \neq 1$, but by Lemma 10, with high probability, $\theta_b \ge \theta_1 - \frac{1}{2} = \frac{1}{2}$ as *Find-the-Pivot* returns an ϵ_b-*Best-Item* with probability at least $(1-\frac{\delta}{4})$. Also $\theta_b \le 1$ for any $b \neq 1$. The above bounds on θ_b clearly implies $\theta_{ij}^b = \frac{\theta_i - \theta_j}{\theta_b} \in \Big[(\theta_i - \theta_j), 2(\theta_i - \theta_j)\Big]$. □

Now to ensure the correctness of *Beat-the-Pivot*, recall that all we need to show it returns an ϵ-*Best-Ranking* $\boldsymbol{\sigma} \in \Sigma_{[n]}$. Same as Algorithm 1, here also we plug in the pivot item b in every group $\mathcal{G}_g$. But now it estimates the *pivotal score* θ_i^b of every item $i \notin \mathcal{G}_g \setminus \{b\}$ instead of *pivotal preference score* p_{ib} where lies the uniqeness of *Score-and-Rank*. We finally output the ranking simply sorting the items w.r.t. θ_i^b – the intuition is if item i beats j in terms of their actual BTL scores (i.e. $\theta_i > \theta_j$), then i beats j in terms of their *pivotal scores* as well (i.e. $\theta_i^b > \theta_j^b$).

More formally, as $\boldsymbol{\sigma}$ denotes the ranking returned by *Beat-the-Pivot*, the correctness of algorithm *Score-and-Rank* fails if $\boldsymbol{\sigma}$ is not an ϵ-*Best-Ranking*. Formally, we have:

$$\begin{aligned} Pr(\text{ Correctness of } \textit{Beat-the-Pivot} \text{ fails }) &= Pr(\exists i, j \in [n] \mid \theta_i > \theta_j + \epsilon \text{ but } \sigma(i) > \sigma(j)) \\ &= Pr(\exists i, j \in [n] \mid \theta_i > \theta_j + \epsilon \text{ but } \hat\theta_i^b < \hat\theta_j^b) \end{aligned} \tag{8}$$

Now, assuming b to be indeed an ϵ_b-*Best-Item*, since $\theta_i > \theta_j \implies \theta_{ij}^b \ge \epsilon$ (from Lemma 23), from Eqn. 8, we further get:

$$\begin{aligned} Pr(\text{ Correctness of } \textit{Beat-the-Pivot} \text{ fails }) &= Pr(\exists i, j \in [n] \mid \theta_i > \theta_j + \epsilon \text{ but } \sigma(i) > \sigma(j)) \\ &\le Pr\Big(\exists i, j \in [n] \setminus \{b\} \mid \theta_i^b > \theta_j^b + \epsilon \text{ but } \sigma(i) > \sigma(j)\Big) \\ &= Pr(\exists i, j \in [n] \setminus \{b\} \mid \theta_i^b > \theta_j^b + \epsilon \text{ but } \hat\theta_i^b < \hat\theta_j^b), \end{aligned} \tag{9}$$

where the inequality follows due to Lemma 23. In the inequality of the above analysis, it is also crucial to note that under the assumption of b to be indeed an ϵ_b-*Best-Item* setting $\sigma(1) = b$ does not incur an error since $\theta_b > \theta_1 - \frac{\epsilon}{2}$. So if we can estimate each $\hat\theta_i^b$ within a confidence interval of $\frac{\epsilon}{2}$, that should be enough to ensure correctness of the algorithm. Thus the only thing remaining to show is *Score-and-Rank* indeed estimates θ_i^b tightly enough with high confidence – formally, it is enough to show that for any group $g \in [G]$ and any item $i \in \mathcal{G}_g \setminus \{b\}$, $Pr\Big(|\theta_i^b - \hat\theta_i^b| > \frac{\epsilon}{2}\Big) \le \frac{\delta}{4n}$.

For this we will be crucially using the result of Lemma 6, which shows that $\hat\theta_i^b \sim Geo\Big(\dfrac{\theta_b}{\theta_i + \theta_b}\Big)$ for any item $i \in \mathcal{G}_g$ and at any group $g \in [G]$.

Using the above insight, we will show that the estimate of *pivotal scores* $\hat{\theta}_i^b$ can not be too bad for any item $i \in \mathcal{G}_g$ at any group $g \in [G]$. The formal statement is given in Lemma 22. For the ease of notation let us define the event $\mathcal{E}_g := \{\exists i \in \mathcal{G}_g \setminus \{b\} \text{ s.t. } |\theta_i^b - \hat{\theta}_i^b| > \frac{\epsilon}{2}\}$.

Lemma 24. *For any group $g \in [G]$, $Pr\Big(\mathcal{E}_g\Big) \le \frac{(k-1)\delta}{4n}$.*

Proof. We will again assume the event that b to be indeed an ϵ_b-*Best-Item* throughout the proof and use the shorthand notation $Pr_b(\cdot)$ as defined previously. Let us first fix a group $g \in [G]$. Then for any item $i \in \mathcal{G}_g \setminus \{b\}$,

$$Pr\Big(|\hat{\theta}_i^b - \theta_i^b| > \frac{\epsilon}{2}\Big) \le 2\exp\left(-\frac{2t(\epsilon/2)^2}{\Big(1+\frac{\theta_i}{\theta_b}\Big)^2\Big((\epsilon/2)+1+\frac{\theta_i}{\theta_b}\Big)}\right) = \frac{\delta}{4n}, \tag{10}$$

where the inequality follows from Lemma 7. Then summing over, all the items $i \in \mathcal{G}_g \backslash \{b\}$ in group $g \in [G]$,

$$Pr_b(\mathcal{E}_g) \le \sum_{i \in \mathcal{G}_g \backslash \{b\}} Pr\left(|\hat{\theta}_i^b - \theta_i^b| > \frac{\epsilon}{2}\right) = (k-1)\frac{\delta}{4n},$$

where the inequality follows from (10). $\square$

Now applying Lemma 24 over all groups $g \in [G]$, and using (9), the probability that correctness of *Score-and-Rank* fails:

$$\begin{aligned} &Pr(\text{ Correctness of } \textit{Beat-the-Pivot} \text{ fails }) \le Pr(\exists i,j \in [n]\setminus\{b\} \mid \theta_i^b > \theta_j^b + \epsilon \text{ but } \hat{\theta}_i^b < \hat{\theta}_j^b) \\ &\quad \le Pr_b(\exists g \in [G] \text{ s.t. } \mathcal{E}_g) = \Big(\Big\lceil\frac{n-1}{k-1}\Big\rceil\Big)\frac{(k-1)\delta}{4n} \le 2\Big(\frac{n-1}{k-1}\Big)\frac{(k-1)\delta}{4n} \le \frac{\delta}{2}, \end{aligned} \tag{11}$$

where the last inequality follows taking union bound over all groups $g \in [G]$.

Thus we are now only left to prove that the correctness of *Score-and-Rank* indeed holds within the desired sample complexity of $O\big(\frac{n}{\epsilon^2}\log\frac{n}{\delta}\big)$. Towards this, let us first define $t' = \frac{5}{2}tk = \frac{5*567k}{2\epsilon^2}\ln\big(\frac{8n}{\delta}\big)$. Also let $Pr_b(\cdot) = Pr(\cdot \big| b \text{ is } \epsilon_b\text{-}\textit{Best-Item})$ denotes the probability of an event conditioned on the event that b is indeed an ϵ_b-*Best-Item* (Recall we have set $\epsilon_b = \min(\frac{\epsilon}{2}, \frac{1}{2})$). For any group $g \in [G]$, we denote by $\mathcal{T}_g$ the total number of times $\mathcal{G}_g$ was played until t wins of item b were observed. Also recall the probability of item b being the winner at any subsetwise play of $\mathcal{C}_g$ is given by $p := \frac{\theta_b}{\sum_{j \in \mathcal{G}_g}\theta_j} > \frac{1}{2k}$, given b is indeed an ϵ_b-*Best-Item* (from Lemma 10). So we have the $\mathcal{T}_g \sim NB(t, \frac{\theta_b}{\sum_{j\in\mathcal{G}_g}\theta_j})$. Then for any fixed group $g \in [G]$, the probability that $\mathcal{G}_g$ needs to be played (queried) for more than t' times to get at least t wins of item b:

$$\begin{aligned} Pr_b(\mathcal{T}_g > t') &= Pr\Big(Bin(t',p) < t\Big) \\ &\le Pr\Big(Bin(t',p) < \frac{4}{5}pt'\Big) \quad \left[\text{Since } pt' > \frac{5tk}{4k}\right] \\ &= Pr\Big(Bin(t',p) - pt' < -\frac{1}{5}pt'\Big) \\ &\le Pr\Big(Bin(t',p) - pt' < (1-\frac{4}{5})pt'\Big) \\ &\le \exp\Big(-\frac{pt'(4/5)^2}{2}\Big) \quad \Big[\text{By multiplicative Chernoff bound}\Big] \end{aligned}$$

$$\leq \exp\Big(-\frac{2}{5}t\Big) \quad \Bigg[\text{Using } pt' > \frac{5t}{4}\Bigg]$$
$$\leq \frac{\delta}{8n} \quad \Bigg[\text{Recall we have } t = \frac{567}{4\big(\epsilon/2\big)^2}\ln\big(\frac{8n}{\delta}\big)\Bigg]$$

Then taking union bound over all the groups

$$\begin{aligned} Pr(\exists g \in [G] \mid \mathcal{T}_g > t') &\leq Pr_b(\exists g \in [G] \mid \mathcal{T}_g > t')Pr(\theta_b > \theta_1 - \epsilon_b) + Pr(\theta_b < \theta_1 - \epsilon_b) \\ &\leq Pr_b(\exists g \in [G] \mid \mathcal{T}_g > t') + \frac{\delta}{4} \leq \sum_{g \in [G]} Pr_b(\mathcal{T}_g > t') + \frac{\delta}{4} \\ &\leq \Big(\lceil \frac{n-1}{k-1} \rceil\Big)\frac{\delta}{8n} + \frac{\delta}{4} \leq 2\Big(\frac{n-1}{k-1}\Big)\frac{\delta}{8n} + \frac{\delta}{4} \leq \frac{\delta}{2}. \end{aligned} \tag{12}$$

Then with high probability $(1-\frac{\delta}{2})$, $\forall g \in [G]$ we have $\mathcal{T}_g < t' = \frac{5}{2}tk$, which makes the total sample complexity of *Score-and-Rank* to be at most $\Big\lceil \frac{n-1}{k-1} \Big\rceil t' \leq \frac{2n}{k}\frac{5tk}{2} = \frac{2835n}{\epsilon^2}\ln\frac{8n}{\delta} = O\bigg(\frac{n}{\epsilon^2}\ln\frac{n}{\delta}\bigg)$, as total number of groups are $G = \lceil \frac{n-1}{k-1} \rceil$.

Moreover the sample complexity of *Find-the-Pivot* is also $O\Big(\frac{n}{\epsilon^2}\log\frac{n}{\delta}\Big)$ as proved in Lemma 10. Combining this with above the total sample complexity of *Beat-the-Pivot* remains $O\Big(\frac{n}{\epsilon^2}\log\frac{n}{\delta}\Big)$. Finally analysing all the previous claims together:

$$\begin{aligned} &Pr(\textit{Beat-the-Pivot} \text{ fails }) \\ &\leq Pr(\text{ Correctness of } \textit{Beat-the-Pivot} \text{ fails }) + Pr(\text{ Sample complexity of } \textit{Beat-the-Pivot} \text{ fails }) \\ &\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta, \end{aligned}$$

where the last inequality follows from (11) and (12), which concludes the proof. □

D Appendix for Section 6

D.1 Restating Lemma 1 of [25]

Consider a multi-armed bandit (MAB) problem with n arms. At round t, let A_t and Z_t denote the arm played and the observation (reward) received, respectively. Let $\mathcal{F}_t = \sigma(A_1, Z_1, \ldots, A_t, Z_t)$ be the sigma algebra generated by the trajectory of a sequential bandit algorithm upto round t.

Lemma 25 (Lemma 1, [25])**.** *Let ν and ν' be two bandit models (assignments of reward distributions to arms), such that ν_i (resp. ν'_i) is the reward distribution of any arm $i \in \mathcal{A}$ under bandit model ν (resp. ν'), and such that for all such arms i, ν_i and ν'_i are mutually absolutely continuous. Then for any almost-surely finite stopping time τ with respect to $(\mathcal{F}_t)_t$,*

$$\sum_{i=1}^{n} \mathbf{E}_\nu[N_i(\tau)]KL(\nu_i, \nu'_i) \geq \sup_{\mathcal{E} \in \mathcal{F}_\tau} kl(Pr_\nu(\mathcal{E}), Pr_{\nu'}(\mathcal{E})),$$

where $kl(x,y) := x\log(\frac{x}{y}) + (1-x)\log(\frac{1-x}{1-y})$ is the binary relative entropy, $N_i(\tau)$ denotes the number of times arm i is played in τ rounds, and $Pr_\nu(\mathcal{E})$ and $Pr_{\nu'}(\mathcal{E})$ denote the probability of any event $\mathcal{E} \in \mathcal{F}_\tau$ under bandit models ν and ν', respectively.

D.2 Proof of Lemma 13

Lemma 13. *For any symmetric (ϵ,δ)-**PAC-Rank** algorithm $\mathcal{A}$, and any problem instance $\nu_S \in \boldsymbol{\nu}_{[q]}$ associated to the set $S \subseteq [n-1]$, $q \in [n-1]$, and for any item $i \in S$, $Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S\backslash\{i\} \cup \{0\}\Big) < \frac{\delta}{q}$, where $Pr_S(\cdot)$*

denotes the probability of an event under the underlying problem instance ν_S and the internal randomness of the algorithm $\mathcal{A}$ (if any).

Proof. **Base Case.** The claim follows trivially for $q = 1$, just from the definition of (ϵ, δ)**-PAC-Rank** property of the algorithm. Suppose $S = \{i\}$ for some $i \in [n-1]$. Then clearly

$$Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1) = \{0\}\Big) < Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1) \neq \{i\}\Big) < \delta.$$

So for the rest of the proof we focus only on the regime where $2 \le q \le n-1$. Let us first fix an $q' \in [n-2]$ and set $q = q'+1$. Clearly $2 \le q \le n-1$. Consider a problem instance $\nu_S \in \boldsymbol{\nu}_{[q]}$. Recall from Remark 1 that we use the notation $S \in \boldsymbol{\nu}_{[q]}$ to denote a problem instance in $\boldsymbol{\nu}_{[q]}$. Then probability of doing an error over all possible choices of $S \in \boldsymbol{\nu}_{[q]}$:

$$\begin{aligned}\sum_{S \in \boldsymbol{\nu}_{[q]}} Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) \neq S\Big) &\ge \sum_{S \in \boldsymbol{\nu}_{[q'+1]}} \sum_{i \in S} Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S\backslash\{i\} \cup \{0\}\Big)\\ &= \sum_{S' \in \boldsymbol{\nu}_{[q']}} \sum_{i \in [n-1]\backslash S'} Pr_{S' \cup \{i\}}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S' \cup \{0\}\Big), \end{aligned} \tag{13}$$

where the above analysis follows from a similar result proved by [24] to derive sample complexity lower bound for classical multi-armed bandit setting towards recovering top-q items (see Theorem 8, [24]).

Clearly the possible number of instances in $\boldsymbol{\nu}_{[q']}$, i.e. $|\boldsymbol{\nu}_{[q']}| = \binom{n-1}{q'}$, as any set $S \subset [n-1]$ of size q' can be chosen from $[n-1]$ in $\binom{n-1}{q'}$ ways. Similarly, $|\boldsymbol{\nu}_{[q]}| = \binom{n-1}{q} = \binom{n-1}{q'+1}$.

Now from *symmetry* of algorithm $\mathcal{A}$ and by construction of the class of our problem instances $\boldsymbol{\nu}_{[q']}$, for any two instances S_1' and S_2' in $\boldsymbol{\nu}_{[q']}$, and for any choices of $i \in [n-1] \setminus S_1'$ and $j \in [n-1] \setminus S_2'$ we have that:

$$Pr_{S_1' \cup \{i\}}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S_1' \cup \{0\}\Big) = Pr_{S_2' \cup \{j\}}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S_2' \cup \{0\}\Big).$$

Let us denote $Pr_{S_1' \cup \{i\}}\Big(\boldsymbol{\sigma}_{\mathcal{A}} = \boldsymbol{\sigma}_{S_1'}\Big) = p' \in (0,1)$. Then above equivalently implies that for all $S \in \boldsymbol{\nu}_{[q]}$ and any $i \in [n-1] \setminus S$,

$$Pr_S(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S \setminus \{i\} \cup \{0\}) = p'.$$

Then using above in (13) we can further derive,

$$\begin{aligned}\sum_{S \in \boldsymbol{\nu}_{[q]}} Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) \neq S\Big) &\ge \sum_{S' \in \boldsymbol{\nu}_{[q']}} \sum_{i \in [n-1]\backslash S'} Pr_{S' \cup \{i\}}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S' \cup \{0\}\Big)\\ &= \sum_{S' \in \boldsymbol{\nu}_{[q']}} (n-1-q')p'\\ &= \binom{n-1}{q'}(n-1-q')p'\\ &= \binom{n-1}{q'}\frac{n-1-q'}{q'+1}(q+1)p'\\ &= \binom{n-1}{q'+1}(q'+1)p' = \binom{n-1}{q}qp'. \end{aligned}$$

But now observe that $|\boldsymbol{\nu}_{[q]}| = \binom{n-1}{q}$. Thus if $p' \ge \frac{\delta}{q}$, this implies

$$\sum_{S \in \boldsymbol{\nu}_{[q]}} Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) \neq S\Big) \ge \binom{n-1}{q}\delta,$$

which in turn implies that there exist at least one instance $\nu_S \in \boldsymbol{\nu}_{[q+1]}$ such that $Pr_S\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) \neq S\Big) \geq \delta$, which violates the (ϵ,δ)**-PAC-Rank** property of algorithm $\mathcal{A}$. Thus it has to be the case that $p' < \frac{\delta}{q}$. Recall that we have chosen $2 \leq q \leq n-1$, and for any $S \in \boldsymbol{\nu}_{[q]}$, and any $i \in [n-1] \setminus S$ we proved that

$$Pr_S(\boldsymbol{\sigma}_{\mathcal{A}}(1:q) = S \setminus \{i\} \cup \{0\}) = p' < \frac{\delta}{q},$$

which concludes the proof. □

D.3 Proof of Lemma 26

Lemma 26. *For any $\delta \in (0,1)$, and $q \in \mathbb{R}_+$, $kl\bigg(1-\delta, \dfrac{\delta}{q}\bigg) > \ln \dfrac{q}{4\delta}$.*

Proof. The proof simply follows from the definition of KL divergence. Recall that for any $p_1, p_2 \in (0,1)$,

$$kl(p_1,p_2) = p_1 \ln \frac{p_1}{p_2} + (1-p_1)\ln\frac{1-p_1}{1-p_2}.$$

Applying above in our case we get,

$$\begin{aligned}
kl\bigg(1-\delta, \frac{\delta}{q}\bigg) &= (1-\delta)\ln\frac{q(1-\delta)}{\delta} + \delta \ln \frac{q(\delta)}{(q-\delta)}\\
&= (1-\delta)\ln\frac{q(1-\delta)}{\delta} + \delta \ln \frac{q(\delta)}{(q-\delta)}\\
&= \ln q + (1-\delta)\ln\frac{(1-\delta)}{\delta} + \delta\ln\frac{(\delta)}{(q-\delta)}\\
&\geq \ln q + (1-\delta)\ln\frac{(1-\delta)}{\delta} + \delta\ln\frac{(\delta)}{(1-\delta)} \quad [\text{ since } q \geq 1]\\
&= \ln q + (1-2\delta)\ln\frac{(1-\delta)}{\delta}\\
&\geq \ln q + (1-2\delta)\ln\frac{1}{2\delta} \quad \bigg[\text{since the second term is negative for } \delta \geq \frac{1}{2}\bigg]\\
&= \ln q + \ln\frac{1}{2\delta} + 2\delta\ln 2\delta\\
&\geq \ln q + \ln\frac{1}{2\delta} + 2\delta\Big(1-\frac{1}{2\delta}\Big) \quad \Big[\text{since } x\ln x \geq (x-1), \forall x > 0\Big]\\
&= \ln q + \ln\frac{1}{2\delta} - (1-2\delta)\\
&\geq \ln q + \ln\frac{1}{2\delta} + \ln\frac{1}{2} = \ln\frac{q}{4\delta}.
\end{aligned}$$

□

D.4 Proof of Theorem 12

Theorem 12 (Lower bound on Sample Complexity with WI feedback)**.** *Given a fixed $\epsilon \in \big(0, \frac{1}{\sqrt{8}}\big]$, $\delta \in [0,1]$, and a symmetric (ϵ,δ)***-PAC-Rank** *algorithm $\mathcal{A}$ for WI feedback, there exists a PL instance ν such that the sample complexity of $\mathcal{A}$ on ν is at least $\Omega\bigg(\frac{n}{\epsilon^2}\ln\frac{n}{4\delta}\bigg)$.*

Proof. The main idea lies in constructing 'hard enough' problem instances on which no algorithm can be (ϵ,δ)**-PAC-Rank** without observing $\Omega\bigg(\frac{n}{\epsilon^2}\ln\frac{1}{4\delta}\bigg)$ number of samples. We crucially use the results of [25] (Lemma 25) for the purpose.

Towards this we first fix our *true problem instance* (ν in Lemma 25) to be $\nu_{S^*} \in \boldsymbol{\nu}_{[q]}$, for some $q \in [n-2]$ (the actual value of q to be decided later). Note that the arm set $\mathcal{B}$ (of Lemma 25) for our current problem setup is set of all k-sized subsets of $[n-1] \cup \{0\}$, i.e. $\mathcal{B} = \{S \subseteq [n-1] \cup \{0\} \mid |S| = k\}$.

We now fix the altered problem instance (ν' in Lemma 25) to be $\nu_{\tilde{S}^*} \in \boldsymbol{\nu}_{[q+1]}$ such that $\tilde{S}^* = S^* \cup \{a\}$, of some $a \in [n-1] \setminus S^*$. Now if $\boldsymbol{\sigma}_{\mathcal{A}} \in \Sigma_{[n]}$ is the ranking returned by Algorithm $\mathcal{A}$, then clearly owing to the (ϵ, δ)**-PAC-Rank** property of $\mathcal{A}$,

$$Pr_{S^*}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\Big) > Pr_{S^*}\big(\{\boldsymbol{\sigma}\}_{\mathcal{A}} \text{ is an } \epsilon\text{-}\textit{Best-Ranking}\}\big) > (1-\delta), \tag{14}$$

Moreover the (ϵ, δ)**-PAC-Rank** property of $\mathcal{A}$ also implies that:

$$Pr_{\tilde{S}^*}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\Big) \le Pr_{\tilde{S}^*}\Big(\boldsymbol{\sigma}_{\mathcal{A}} \neq \boldsymbol{\sigma}_{\tilde{S}^*}\Big) < \delta.$$

But owing to Lemma 13, we are able to claim a stronger bound (using $S = \tilde{S}^*$):

$$Pr_{\tilde{S}^*}\Big(\boldsymbol{\sigma}_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\Big) < \frac{\delta}{q}. \tag{15}$$

We will crucially use (14) and (15) in the following analysis. But before that note that for problem instance $\nu_{S^*} \in \boldsymbol{\nu}_{[q]}$, the probability distribution associated with a particular arm (set of size k in our case) $B \in \mathcal{B}$ is given by:

$$\nu_{S^*}^B \sim Categorical(p_1, p_2, \ldots, p_k), \text{ where } p_i = Pr(i|B), \quad \forall i \in [k], \forall B \in \mathcal{B},$$

where $Pr(i|S)$ is as defined in Sec. (3.1). Now applying Lemma 25, for some event $\mathcal{E} \in \mathcal{F}_\tau$ we get,

$$\sum_{\{B \in \mathcal{B}: a \in B\}} \mathbf{E}_{\nu_{S^*}^B}[N_B(\tau_{\mathcal{A}})] KL(\nu_{S^*}^B, \nu_{\tilde{S}^*}^B) \ge kl(Pr_{\nu_{S^*}^B}(\mathcal{E}), Pr_{\nu_{\tilde{S}^*}^B}(\mathcal{E})), \tag{16}$$

where $N_B(\tau_{\mathcal{A}})$ denotes the number of times arm (subset of size k) B is played by $\mathcal{A}$ in τ rounds. Above clearly follows due to the fact that for any arm $B \in \mathcal{B}$ such that $a \notin B$, ν_S^B is same as $\nu_{\tilde{S}^*}^B$, and hence $KL(\nu_{S^*}^B, \nu_{\tilde{S}^*}^B) = 0$, $\forall S \in \mathcal{A}$, $a \notin S$. For the notational convenience we will henceforth denote $\mathcal{B}^a = \{B \in \mathcal{B} : a \in S\}$.

Now let us first analyse the right hand side of (16), for any set $B \in \mathcal{B}^a$.

Case 1. Assume $0 \notin B$, and denote by $r = |B \cap S^*|$ the number of "good" arms with PL parameter $\theta\Big(\frac{1}{2} + \epsilon\Big)^2$. Note that for problem instance $\nu_{S^*}^B$,

$$\nu_{S^*}^B(i) = \begin{cases} \frac{\theta(\frac{1}{2}+\epsilon)^2}{r\theta(\frac{1}{2}+\epsilon)^2 + (k-r)\theta(\frac{1}{2}-\epsilon)^2} = \frac{R^2}{rR^2+(k-r)}, \forall i \in [k], \text{ such that } B(i) \in S^*, \\ \frac{\theta(\frac{1}{2}-\epsilon)^2}{r\theta(\frac{1}{2}+\epsilon)^2 + (k-r)\theta(\frac{1}{2}-\epsilon)^2} = \frac{1}{rR^2+(k-r)}, \text{ otherwise.} \end{cases}$$

Similarly, for problem instance $\nu_{\tilde{S}^*}^B$, we have:

$$\nu_{\tilde{S}^*}^B(i) = \begin{cases} \frac{\theta(\frac{1}{2}+\epsilon)^2}{(r+1)\theta(\frac{1}{2}+\epsilon)^2 + (k-r-1)\theta(\frac{1}{2}-\epsilon)^2} = \frac{R^2}{(r+1)R^2+(k-r-1)}, \forall i \in [k], \text{ such that } B(i) \in \tilde{S}^* = S^* \cup \{a\}, \\ \frac{\theta(\frac{1}{2}-\epsilon)^2}{(r+1)\theta(\frac{1}{2}+\epsilon)^2 + (k-r-1)\theta(\frac{1}{2}-\epsilon)^2} = \frac{1}{(r+1)R^2+(k-r-1)}, \text{ otherwise.} \end{cases}$$

Now using the following upper bound on $KL(\mathbf{p}_1, \mathbf{p}_2) \le \sum_{x \in \mathcal{X}} \frac{p_1^2(x)}{p_2(x)} - 1$, $\mathbf{p}_1$ and $\mathbf{p}_2$ be two probability mass functions on the discrete random variable $\mathcal{X}$ [31] we get:

$$
\begin{aligned}
KL(\nu^B_{S^*}, \nu^B_{\tilde S^*}) &\le \frac{(r+1)R^2 + (k-r-1)}{(rR^2+k-r)^2}\Big[rR^2 + \frac{1}{R^2} + (k-r-1)\Big] - 1 \\
&= \Big(R - \frac{1}{R}\Big)^2 \Big[\frac{rR^2 + (k-r-1)}{(rR^2+k-r)^2}\Big] \qquad (17)
\end{aligned}
$$

Case 2. Now assume $0 \in B$, and denote by $r = |B \cap S^* \cup \{0\}|$ the number of "non-bad" arms with PL parameter greater than $\theta\Big(\frac{1}{2} - \epsilon\Big)^2$. Clearly $r \ge 1$ as $0 \in B$. Similar to **Case 1**, for problem instance $\nu^B_{S^*}$,

$$
\nu^B_{S^*}(i) = \begin{cases} \frac{\theta(\frac{1}{2}+\epsilon)^2}{(r-1)\theta(\frac{1}{2}+\epsilon)^2 + (k-r)\theta(\frac{1}{2}-\epsilon)^2 + \theta(\frac{1}{4}-\epsilon^2)} = \frac{R^2}{(r-1)R^2+(k-r)+R}, \forall i \in [k], \text{ such that } B(i) \in S^*, \\ \frac{\theta(\frac{1}{4}-\epsilon^2)}{(r-1)\theta(\frac{1}{2}+\epsilon)^2 + (k-r)\theta(\frac{1}{2}-\epsilon)^2 + \theta(\frac{1}{4}-\epsilon^2)} = \frac{R}{(r-1)R^2+(k-r)+R}, \forall i \in [k], \text{ such that } B(i) = 0, \\ \frac{\theta(\frac{1}{2}-\epsilon)^2}{(r-1)\theta(\frac{1}{2}+\epsilon)^2 + (k-r)\theta(\frac{1}{2}-\epsilon)^2 + \theta(\frac{1}{4}-\epsilon^2)} = \frac{1}{(r-1)R^2+(k-r)+R}, \text{ otherwise.} \end{cases}
$$

Similarly, for problem instance $\nu^B_{\tilde S^*}$, we have:

$$
\nu^B_{\tilde S^*}(i) = \begin{cases} \frac{\theta(\frac{1}{2}+\epsilon)^2}{r\theta(\frac{1}{2}+\epsilon)^2 + (k-r-1)\theta(\frac{1}{2}-\epsilon)^2 + \theta(\frac{1}{4}-\epsilon^2)} = \frac{R^2}{rR^2+(k-r-1)+R}, \forall i \in [k], \text{ such that } B(i) \in \tilde S^* = S^* \cup \{a\}, \\ \frac{\theta(\frac{1}{4}-\epsilon^2)}{r\theta(\frac{1}{2}+\epsilon)^2 + (k-r-1)\theta(\frac{1}{2}-\epsilon)^2 + \theta(\frac{1}{4}-\epsilon^2)} = \frac{R}{rR^2+(k-r-1)+R}, \forall i \in [k], \text{ such that } B(i) = 0, \\ \frac{\theta(\frac{1}{2}-\epsilon)^2}{r\theta(\frac{1}{2}+\epsilon)^2 + (k-r-1)\theta(\frac{1}{2}-\epsilon)^2 + \theta(\frac{1}{4}-\epsilon^2)} = \frac{1}{rR^2+(k-r-1)+R}, \text{ otherwise.} \end{cases}
$$

Same as before, again using the following upper bound on $KL(\mathbf{p}_1, \mathbf{p}_2) \le \sum_{x \in \mathcal{X}} \frac{p_1^2(x)}{p_2(x)} - 1$, $\mathbf{p}_1$ and $\mathbf{p}_2$ be two probability mass functions on the discrete random variable $\mathcal{X}$ [31] we get:

$$
\begin{aligned}
KL(\nu^B_{S^*}, \nu^B_{\tilde S^*}) &\le \frac{rR^2 + R + (k-r-1)}{((r-1)R^2+R+k-r)^2}\Big[rR^2 + \frac{1}{R^2} + (k-r-1)\Big] - 1 \\
&= \Big(R - \frac{1}{R}\Big)^2 \left[\frac{(r-1)R^2 + (k-r) + R - 1}{\big((r-1)R^2 + (k-r) + R\big)^2}\right] \qquad (18)
\end{aligned}
$$

Now, consider $\mathcal{E}_0 \in \mathcal{F}_\tau$ be an event such that $\mathcal{E}_0 := \{\boldsymbol{\sigma}_{\mathcal{A}}(1:q+1) = S^* \cup \{0\}\}$. Note that since algorithm $\mathcal{A}$ is (ϵ,δ)**-PAC-Rank**, clearly $Pr_{\nu^B_{S^*}}(\mathcal{E}_0) > Pr_{\nu^B_{S^*}}\big(\{\boldsymbol{\sigma})_{\mathcal{A}} \text{ is an } \epsilon\textit{-Best-Ranking}\}\big) > (1-\delta)$. On the other hand, Lemma 13 implies that $Pr_{\nu^B_{\tilde S^*}}(\mathcal{E}_0)) < \frac{\delta}{q}$. Then analysing the left hand side of (16) for $\mathcal{E} = \mathcal{E}_0$ along with (14) and (15), we get that

$$
kl(Pr_{\nu^B_{S^*}}(\mathcal{E}_0), Pr_{\nu^B_{\tilde S^*}}(\mathcal{E}_0)) \ge kl\Big(1-\delta, \frac{\delta}{q}\Big) \ge \ln \frac{q}{4\delta} \qquad (19)
$$

where the last inequality follows from Lemma 26.

Now applying (16) for each altered problem instance $\nu^B_{\tilde S^*}$, each corresponding to any one of the $(n-1-q)$ different choices of $a \in [n-1] \setminus S^*$, and summing all the resulting inequalities of the form (16):

$$
\sum_{a \in [n-1]\setminus S^*} \sum_{B \in \mathcal{B}^a} \mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_{\mathcal{A}})] KL(\nu^B_{S^*}, \nu^B_{\tilde S^*}) \ge (n-1-q) \ln \frac{q}{4\delta}. \qquad (20)
$$

A crucial observation here is that in the left hand side of (20) above, any $B \in \mathcal{B}^a$ shows up for exactly $k-r$ may times, where r is as defined in **Case 1** and **Case 2** above. Thus, given a fixed set B, the coefficient of the term $\mathbf{E}_{\nu^B_{S^*}}$ becomes for:

Case 1. From (17), $(k-r)\Big(R-\frac{1}{R}\Big)^2\Big[\frac{rR^2+(k-r-1)}{(rR^2+k-r)^2}\Big] \le \Big(R-\frac{1}{R}\Big)^2$, as $r \ge 0$.

Case 2. From (18), $(k-r)\Big(R-\frac{1}{R}\Big)^2\left[\frac{(r-1)R^2+(k-r)+R-1}{\Big((r-1)R^2+(k-r)+R\Big)^2}\right] \le \Big(R-\frac{1}{R}\Big)^2$, as in this case $r \ge 1$, and note that $R = \frac{\frac{1}{2}+\epsilon}{\frac{1}{2}-\epsilon} > 1$ by definition.

Thus from (20) we further get

$$\sum_{a=2}^{n}\sum_{\{S\in\mathcal{A}|a\in S\}}\mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)]KL(\nu^B_{S^*},\nu^B_{\tilde{S}^*}) \le \sum_{S\in\mathcal{A}}\mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)]\Bigg(R-\frac{1}{R}\Bigg)^2$$
$$\le 256\epsilon^2\sum_{S\in\mathcal{A}}\mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)]\quad\Bigg[\text{since, }\Bigg(R-\frac{1}{R}\Bigg) = \frac{8\epsilon}{(1-4\epsilon^2)} \le 16\epsilon, \forall \epsilon \in \Bigg(0,\frac{1}{\sqrt{8}}\Bigg]\Bigg]. \tag{21}$$

Finally noting that $\tau_A = \sum_{B\in\mathcal{B}}[N_B(\tau_A)]$, and combining (20) and (21), we get

$$256\epsilon^2\mathbf{E}_{\nu^B_{S^*}}[\tau_A] = \sum_{S\in\mathcal{A}}\mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)](256\epsilon^2) \ge (n-1-q)\ln\frac{q}{4\delta}.$$

The proof now follows choosing $q = \lfloor\frac{n}{2}\rfloor$ and the fact that $n \ge 4$, as $(n-1-q) \ge \frac{n}{2}-1 \ge \frac{n}{4}$ for any $n \ge 4$. Also $\ln q \ge \ln(\frac{n-1}{2}) \ge \ln\frac{n}{4}$ for any $n \ge 2$. Thus above construction shows the existence of a problem instance $\boldsymbol{\nu} = \nu^B_{S^*}$, such that $\mathbf{E}_{\nu^B_{S^*}}[\tau_A] = \frac{n}{1024\epsilon^2}\ln\frac{n}{16\delta} = \Omega\Big(\frac{n}{\epsilon^2}\ln\frac{n}{\delta}\Big)$, which concludes the proof. □

D.5 Proof of Theorem 14

Proof. The result can be obtained following an exact same proof as that of Theorem 12 with the observation that for any $\theta > 0$, for any of the problem instances ν_S, $S \subseteq [n-1]$, $\boldsymbol{\sigma}_S$ is the only ϵ-*Best-Ranking* for ν_S as follows from the observation that:

Case 1. For any $i \in [n-1]\backslash S$

$$Pr_S(0|\{i,0\}) = \frac{\theta\Big(\frac{1}{4}-\epsilon^2\Big)}{\theta\Big(\frac{1}{4}-\epsilon^2\Big)+\theta\Big(\frac{1}{2}-\epsilon\Big)^2} = \frac{1}{2}+\epsilon.$$

So i must follow 0, in the any ϵ-*Best-Ranking*.

Case 2. For any $i \in S$

$$Pr_S(i|\{i,0\}) = \frac{\theta\Big(\frac{1}{2}+\epsilon\Big)^2}{\theta\Big(\frac{1}{4}-\epsilon^2\Big)+\theta\Big(\frac{1}{2}+\epsilon\Big)^2} = \frac{1}{2}+\epsilon.$$

So 0 must follow i, in the any ϵ-*Best-Ranking*. So the only choice of ϵ-*Best-Ranking* for problem instance $\boldsymbol{\nu}_S$

□

D.6 Proof of Theorem 15

Theorem 15 (Sample Complexity Lower Bound for TR)**.** *Given* $\epsilon \in \Big(0,\frac{1}{8}\Big]$ *and* $\delta \in (0,1]$, *and a symmetric* (ϵ,δ)**-PAC-Rank** *algorithm* $\mathcal{A}$ *with top-m ranking (TR) feedback* ($2 \le m \le k$), *there exists a PL instance* ν *such that the expected sample complexity of* $\mathcal{A}$ *on* ν *is at least* $\Omega\Big(\frac{n}{m\epsilon^2}\ln\frac{n}{4\delta}\Big)$.

Proof. The proof follows exactly following the same lines of argument as of Theorem 12. The only difference lies in computing the KL-divergence terms in the left hand side of Lemma 25 for TR-m feedback model. We consider the exact same set of problem instances for the purpose as defined in Theorem 12.

The interesting thing however to note is that how top-m ranking feedback affects the KL-divergence analysis in this case. Precisely, using chain rule of the KL-divergence along the two case analyses of Eqn. 18 and 18, it can be shown that with TR feedback

$$KL(\nu^B_{S^*}, \nu^B_{\tilde{S}^*}) \le m\Big(R - \frac{1}{R}\Big)^2$$

for any set $B \in \mathcal{B}$. This shows a multiplicative m-factor blow up in the KL-divergence terms compared to earlier case with WI, owning to top-m ranking feedback–this in fact triggers the $\frac{1}{m}$ reduction in regret learning rate as shown below. Similar to (21) we here get

$$\sum_{a=2}^{n} \sum_{\{S\in\mathcal{A} | a\in S\}} \mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)] KL(\nu^B_{S^*}, \nu^B_{\tilde{S}^*}) \le m \sum_{S\in\mathcal{A}} \mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)] \left(R - \frac{1}{R}\right)^2$$
$$\le 256 m\epsilon^2 \sum_{S\in\mathcal{A}} \mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)] \quad \left[\text{since, } \left(R - \frac{1}{R}\right) = \frac{8\epsilon}{(1-4\epsilon^2)} \le 16\epsilon, \forall \epsilon \in \left(0, \frac{1}{\sqrt{8}}\right]\right]. \tag{22}$$

Now noting that $\tau_A = \sum_{B\in\mathcal{B}}[N_B(\tau_A)]$, and combining (20) and (22), we further get

$$256 m\epsilon^2 \mathbf{E}_{\nu^B_{S^*}}[\tau_A] = \sum_{S\in\mathcal{A}} \mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)](256 m\epsilon^2) \ge (n-1-q)\ln\frac{q}{4\delta}.$$

The proof now similarly follows choosing $q = \lfloor \frac{n}{2} \rfloor$ with $n \ge 4$. The above inequality implies that

$$\sum_{S\in\mathcal{A}} \mathbf{E}_{\nu^B_{S^*}}[N_B(\tau_A)] \ge \frac{1}{256 m\epsilon^2}\left(\frac{n}{4}\right)\ln\frac{n}{4\delta},$$

which certifies existence of a problem instance $\boldsymbol{\nu} = \nu^B_{S^*}$, such that $\mathbf{E}_{\nu^B_{S^*}}[\tau_A] = \frac{n}{1024 m\epsilon^2}\ln\frac{n}{16\delta} = \Omega\Big(\frac{n}{m\epsilon^2}\ln\frac{n}{\delta}\Big)$, and the claim follows. □

E Appendix for Section 7

E.1 Pseudocode for algorithms

Algorithm 4 *Rank-Break* (for updating the pairwise win counts w_{ij} with TR feedback for Algorithm 5)

Input:
    Subset $S \subseteq [n]$, $|S| = k$ $(n \ge k)$
    A top-m ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{S_m}$, $S_m \subseteq [n]$, $|S_m| = m$
    Pairwise (empirical) win-count w_{ij} for each item pair $i, j \in S$
while $\ell = 1, 2, \ldots m$ **do**
    Update $w_{\sigma(\ell)i} \leftarrow w_{\sigma(\ell)i} + 1$, for all $i \in S \setminus \{\sigma(1), \ldots, \sigma(\ell)\}$
end while

Algorithm 5 *Beat-the-Pivot* (for TR feedback)

Input:
  Set of item: $[n]$ $(n \geq k)$, and subset size: k
  Error bias: $\epsilon > 0$, confidence parameter: $\delta > 0$
Initialize:
  $\epsilon_b \leftarrow \min(\frac{\epsilon}{2}, \frac{1}{2})$; $b \leftarrow$ *Find-the-Pivot*$(n, k, \epsilon_b, \frac{\delta}{2})$
  Set $S \leftarrow [n] \setminus \{b\}$, and divide S into $G := \lceil \frac{n-1}{k-1} \rceil$ sets $\mathcal{G}_1, \mathcal{G}_2, \cdots \mathcal{G}_G$ such that $\cup_{j=1}^{G} \mathcal{G}_j = S$ and $\mathcal{G}_j \cap \mathcal{G}_{j'} = \emptyset$, $\forall j, j' \in [G]$, $|G_j| = (k-1)$, $\forall j \in [G-1]$
  If $|\mathcal{G}_G| < (k-1)$, **then** set $\mathcal{R} \leftarrow \mathcal{G}_G$, and $S \leftarrow S \setminus \mathcal{R}$, $S' \leftarrow$ Randomly sample $(k-1-|\mathcal{G}_G|)$ items from S, and set $\mathcal{G}_G \leftarrow \mathcal{G}_G \cup S'$
  Set $\mathcal{G}_j = \mathcal{G}_j \cup \{b\}$, $\forall j \in [G]$
for $g = 1, 2, \ldots, G$ **do**
  Set $\epsilon' \leftarrow \frac{\epsilon}{16}$, $\delta' \leftarrow \frac{\delta}{8n}$ and $t := \frac{2k}{m\epsilon'^2} \log \frac{1}{\delta'}$
  Initialize pairwise (empirical) win-count $w_{ij} \leftarrow 0$, for each item pair $i, j \in \mathcal{G}_g$
  for $\tau = 1, 2, \ldots t$ **do**
    Play the set $\mathcal{G}_g$
    Receive feedback: The top-m ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{\mathcal{G}_{gm}^\tau}$, where $\mathcal{G}_{gm}^\tau \subseteq \mathcal{G}_g$, $|\mathcal{G}_{gm}^\tau| = m$
    Update win-count w_{ij} of each item pair $i, j \in \mathcal{G}_g$ using *Rank-Break*$(\mathcal{G}_g, \boldsymbol{\sigma})$
  end for
  Estimate $\hat{p}_{ib} \leftarrow \frac{w_{ib}}{w_{ib} + w_{bi}}$, $\forall i \in \mathcal{G}_g \setminus \{b\}$
end for
Choose $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$, such that $\sigma(b) = 1$ and $\boldsymbol{\sigma}(i) < \boldsymbol{\sigma}(j)$ if $\hat{p}_{ib} > \hat{p}_{jb}$, $\forall i, j \in S \cup \mathcal{R}$
Output: The ranking $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_{[n]}$

E.2 Proof of Theorem 17

Theorem 17 (*Beat-the-Pivot*: Correctness and Sample Complexity with TR)**.** *With top-m ranking (TR) feedback model, Beat-the-Pivot (Algorithm 5) is* (ϵ, δ)**-PAC-Rank** *with sample complexity* $O(\frac{n}{m\epsilon^2} \log \frac{n}{\delta})$.

Proof. Note that the only difference of Algorithm 5 from that of Algorithm 1 is former plays each group $\mathcal{G}_g$ only for $\frac{1}{m}$ fraction of the later (as t is set to be $t := \frac{2k}{m\epsilon'^2} \log \frac{1}{\delta'}$ for Algorithm 5). The sample complexity bound of Theorem 17 thus holds straightforwardly, same as the derivation shown in the proof of Theorem 8 for proving the sample complexity guarantee of Algorithm 1.

The main novely lies in showing the with TR feedback how does the same guarantee of Theorem 8 still holds. This essentially holds due to the rank breaking updates on each pair w_{ij} as formally justified below.

The proof follows exactly the same analysis till Lemma 21 as the expression of t is not used till that part. The crucial claim now is to prove an equivalent statement of Lemma 21 for the current value of t. We show this as follows:

Consider any particular set $\mathcal{G}_g$ at any iteration $\ell \in \lfloor \frac{n-1}{k-1} \rfloor$ and define $q_i := \sum_{\tau=1}^{t} \mathbf{1}(i \in \mathcal{G}_{gm}^\tau)$ as the number of times any item $i \in \mathcal{G}_g$ appears in the top-m rankings in t rounds of play of the subset $\mathcal{G}_g$. Then conditioned on the event that b is indeed an ϵ_b-*Best-Item*, for any group $g \in [G]$, then with probability at least $\Big(1 - \frac{\delta}{8n}\Big)$, the empirical win count $w_b > (1-\eta)\frac{t}{2k}$, for any $\eta \in \big(\frac{1}{8\sqrt{2}}, 1\big]$. More formally,

Lemma 27. *Conditioned on the event that b is indeed an ϵ_b-*Best-Item*, for any group $g \in [G]$ with probability at least* $\Big(1 - \dfrac{\delta}{8n}\Big)$, $q_b \geq (1-\eta)\frac{mt}{2k}$, *for any* $\eta \in \big(\frac{1}{8\sqrt{2}}, 1\big]$.

Proof. Fix any iteration ℓ and a set $\mathcal{G}_g$, $g \in 1, 2, \ldots G$. Define $i^\tau := \mathbf{1}(i \in \mathcal{G}_{gm}^\tau)$ as the indicator variable if i^{th} element appeared in the top-m ranking at iteration $\tau \in [t]$. Recall the definition of TR feedback model (Sec. 3.1). Using this we get $\mathbf{E}[b^\tau] = Pr(\{b \in \mathcal{G}_{gm}^\tau\}) = Pr\big(\exists j \in [m] \mid \sigma(b) = j\big) = \sum_{j=1}^{m} Pr\big(\sigma(b) = j\Big) = \sum_{j=0}^{m-1} \frac{1}{2(k-j)} \geq \frac{m}{2k}$, since $Pr(\{b|S\}) = \frac{\theta_b}{\sum_{j \in S}\theta_j} \ge \frac{1/2}{|S|}$ for any $S \subseteq [\mathcal{G}_g]$, as b is assumed to be an ϵ_b-*Best-Item* where $\epsilon_b \ge \frac{1}{2} \implies \theta_b \ge 1 - \frac{1}{2} = \frac{1}{2}$. Thus we get $\mathbf{E}[q_b] = \sum_{\tau=1}^{t} \mathbf{E}[b^\tau] \ge \frac{mt}{2k}$.

Now applying Chernoff-Hoeffdings bound for w_b, we get that for any $\eta \in (\frac{3}{32}, 1]$,

$$
\begin{aligned}
Pr_b\Big(q_b \le (1-\eta)\mathbf{E}[q_b]\Big) &\le \exp\Big(-\frac{\mathbf{E}[q_b]\eta^2}{2}\Big) \le \exp\Big(-\frac{mt\eta^2}{4k}\Big), \ \Big(\text{since } \mathbf{E}[q_b] \ge \frac{mt}{2k}\Big)\\
&\le \exp\Big(-\frac{\eta^2}{2\epsilon'^2}\ln\Big(\frac{1}{\delta'}\Big)\Big) \le \exp\Big(-\ln\Big(\frac{1}{\delta'}\Big)\Big) \le \frac{\delta}{8n},
\end{aligned}
$$

where the second last inequality holds for any $\eta > \frac{1}{8\sqrt{2}}$ as it has to be the case that $\epsilon' < \frac{1}{16}$ since $\epsilon \in (0,1)$. Thus for any $\eta > \frac{1}{8\sqrt{2}}$, $\eta^2 \ge 4\epsilon'^2$.

Thus we finally derive that with probability at least $\Big(1 - \frac{\delta_\ell}{8n}\Big)$, one can show that $q_b > (1-\eta)\mathbf{E}[q_b] \ge (1-\eta)\frac{tm}{2k}$, and the proof follows henceforth. $\square$

Above is the crucial most result due to which it is possible to prove Lemma 22 to be true in this case as well, even with an m-factor reduced sample complexity. This follows since:

$$
\begin{aligned}
Pr_b\Bigg(\Big\{|p_{ib} - \hat{p}_{ib}| > \frac{\epsilon}{16}\Big\} \cap \mathcal{F}_g\Bigg) &\le Pr_b\Bigg(\Big\{|p_{ib} - \hat{p}_{ib}| > \frac{\epsilon}{16}\Big\} \cap \Big\{n_{ib}^g \ge \frac{mt}{4k}\Big\}\Bigg)\\
&\le 2\exp\Big(-2\frac{mt}{4k}\big(\frac{\epsilon}{16}\big)^2\Big) = \frac{\delta}{4n},
\end{aligned} \tag{23}
$$

where same as before the first inequality follows as $\mathcal{F}_g \implies n_{ib}^g \ge \frac{t}{4k}$, and the second inequality holds due to Lemma 5 with $\eta = \frac{\epsilon}{16}$, and $v = \frac{mt}{4k}$.

Thus the rest of the proof can be derived following the exact same analysis as of Theorem 8 post Lemma 22, which shows that *Beat-the-Pivot* indeed returns an ϵ-*Best-Ranking* with probability at least $(1-\delta)$ for TR feedback model, and the claim of Theorem 17 holds good. $\square$