View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by LSE Research Online

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000
© 0000 INFORMS

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

Asymptotically Optimal Sequential Design for Rank Aggregation

Xi Chen
Stern School of Business, New York University, xc13@stern.nyu.edu

Yunxiao Chen
Department of Statistics, London School of Economics and Political Science, y.chen186@lse.ac.uk

Xiaoou Li
School of Statistics, University of Minnesota, lixx1766@umn.edu

A sequential design problem for rank aggregation is commonly encountered in psychology, politics, marketing, sports, etc. In this problem, a decision-maker is responsible for ranking K items by sequentially collecting noisy pairwise comparisons from judges. The decision-maker needs to choose a pair of items for comparison in each step, decide when to stop data collection, and make a final decision after stopping, based on a sequential flow of information. Due to the complex ranking structure, existing sequential analysis methods are not suitable.

In this paper, we formulate the problem under a Bayesian decision framework and propose sequential procedures that are asymptotically optimal. These procedures achieve asymptotic optimality by seeking a balance between exploration (i.e., finding the most indistinguishable pair of items) and exploitation (i.e., comparing the most indistinguishable pair based on the current information). New analytical tools are developed for proving the asymptotic results, combining advanced change of measure techniques for handling the level crossing of likelihood ratios and classic large deviation results for martingales, which are of separate theoretical interest in solving complex sequential design problems. A mirror-descent algorithm is developed for the computation of the proposed sequential procedures.

Key words: Active sequential tests; asymptotically optimal policy; sequential analysis; rank aggregation

MSC2000 subject classification: Primary: 62L10; secondary: 62L15

OR/MS subject classification: Primary: Probability/Stochastic model applications; secondary: Statistics/Sampling:

History:

1. Introduction This paper considers a sequential design problem for rank aggregation. In this problem, a decision maker is responsible for ranking K items by adaptively collecting noisy outcome of pairwise comparison from judges. Sequential rank aggregation has a wide range of applications, including social choice [49], sports [23], search rankings [48], etc. Pairwise comparison is the most popular approach for rank aggregation, as sufficient evidence from cognitive psychology suggests that people make more accurate judgement when making pairwise comparisons (i.e., given a pair of items and asked to indicate which item is preferred to the other) as compared to multi-wise comparison [10] and some applications such as chess gaming have a natural form of pairwise comparison.

In a rank aggregation problem, more comparisons usually lead to a more accurate global ranking. However, each comparison comes with some cost, e.g., in crowdsourcing applications, a requester has to pay crowd workers a fixed amount of monetary reward for each labeled pair. Therefore, to design a cost-efficient ranking procedure, a decision maker faces the following three key challenges:

1. How to adaptively decide the next pair of objects for comparison based on the collected information? The adaptive selection of pairs is important for saving the cost. For example, if we are confident that object 1 is ranked higher than 2 and the object 2 is preferred over 3, there is no need to compare objects 1 and 3.
2. When to stop asking for more comparisons?
3. When stopping the comparison process, how to aggregate the pairwise comparisons to infer the global ranking?

Due to wide applications of rank aggregation, there are several recent machine learning works devoted to the development of ranking algorithms with rigorous theoretical guarantees. For example, Hajek et al. [25], Negahban et al. [46], Shah et al. [51] proposed algorithms and established the estimation error rates under Bradley-Terry-Luce (BTL) model [11, 41], Thurstone model [55], and a more general strong stochastic transitivity model [4, 44]. However, these works mainly focus on a static setting with either given pairs or randomly drawn pairs. In contrast, under a sequential setting, we are interested in designing an adaptive pair selection rule. Moreover, for recent active ranking works (e.g., [26]), optimal stopping is usually not considered. For example, the common studied PAC (Probably Approximately Correct) sample complexity bound from the machine learning literature usually involves some large universal constants and cannot be directly used for an accurate stopping rule. Determining a right stopping time is critical for balancing accuracy and cost in many applications (e.g., ranking via crowdsourcing). Therefore, to address the challenge of optimal stopping, we adopt the sequential analysis framework from statistics that directly optimizes over the random stopping time. On the other hand, due to the complex structure of ranking aggregation, this problem cannot be formulated and solved by existing sequential adaptive design methods [20, 45].

Under a wide class of parametric comparison models (e.g., Bradley-Terry-Luce (BTL) model [11, 41]), we develop new sequential analysis methods to conduct sequential experiments for pairwise comparisons and to balance the ranking accuracy and cost. We first formulate the problem under a general *Bayesian decision framework*. In particular, each item k is represented by a parameter θ_k, which determines its underlying true rank among K items. For example, the parameter θ_k can be viewed as the quality score for item k, and item i has a higher rank than item j if and only if $\theta_i > \theta_j$. The pairwise comparison of items i and j follows a probabilistic comparison model (e.g., [11, 41, 55]) parameterized by θ_i and θ_j. Under the Bayesian framework, the parameter vector for all product $\boldsymbol{\theta}$ is drawn from some prior distribution. A sequential procedure chooses a pair (i,j) for the next comparison in each stage and decides the stopping time T. Upon stopping, the final decision is to choose the global rank $R := (R_1, \ldots, R_K)$ from the set of all permutations of $\{1,2,...,K\}$. To measure the accuracy of a rank R, we adopt the widely used Kendall's tau distance [32], which measures the number of inconsistent pairs between the decision R and the underlying true rank induced by the scores $(\theta_1,...,\theta_K)$. Then, the loss function of this sequential design problem is defined by combining the cost of data collection and the Kendall's tau distance:

$$\sum_{i<j} \left\{ I(\theta_i > \theta_j) I(R_i > R_j) + I(\theta_i < \theta_j) I(R_i < R_j) \right\} + cT, \tag{1}$$

where the constant $c > 0$ indicates the relative cost of each comparison and $I(\cdot)$ denotes an indicator function. The goal is to optimize the expected loss in (1) over pair-selection rule, stopping rule T, and final decision R (see Section 2 for more details). To justify the performance of the proposed policies, we adopt the notion of "asymptotic optimality" from [20] (see Eq. (7) below), that is widely used in sequential analysis [35, 50, 52, 54]. While finding an exact optimal policy is computationally intractable, we prove that the proposed policies are asymptotically optimal.

It is also worthwhile noting that although according to the final decision, our problem seems to be a multi-hypothesis sequential testing problem with adaptive experiment selection considered in [45], there exist fundamental differences. First, Naghshvar and Javidi [45] only consider simple hypotheses, while the ranking problem, when viewed as a multi-hypothesis testing problem, consists of composite hypotheses. Second, typically $0-1$ loss is considered for measuring the decision accuracy in multi-hypothesis testing, while our problem has a more complex loss function based on the Kendall's tau distance that is tailored to rank aggregation. Our problem is also a substantial generalization of classical sequential test of two composite hypotheses [33, 34, 50]. In particular, when the number of items is two $(K = 2)$, our problem degenerates to testing two composite hypotheses without adaptive experiment selection.

1.1. Main contribution

We summarize the main methodological and theoretical contributions of the paper as follows.

- Under a Bayesian decision framework and under a large class of parametric pairwise comparison models, we derive an asymptotic lower bound (Theorem 1) for the Bayes risk of all possible sequential ranking policies. Note that the Bayes risk of the sequential rank aggregation problem, which combines the expected Kendall's tau distance and the expected sample size, is more complex than that of the traditional sequential hypothesis testing problems (e.g., [20, 33, 45]).
- We propose two sequential ranking policies. In particular, we provide two choices of stopping rule and a class of randomized pair selection rules. We quantify the expected Kendall's tau and the sample size of the proposed methods (Theorems 2 and 3) and show that the Bayes risks match the asymptotic lower bound, which further implies that the proposed methods are asymptotically optimal (Corollary 1). Our randomized pair selection rule utilizes an epsilon-greedy strategy to balance the exploration (i.e., randomly selecting pairs to gain information about the underlying parameters $\{\theta_k\}_{k=1}^K$) and exploitation (i.e., choosing the best pair for comparison based on the current information). The exploration is critical for learning the rank, while the exploitation is critical for saving the sample size for comparison.

—For the exploration, we quantify the impact of the exploration rate on the estimation of model parameters and provide an exponential probability bound as an auxiliary result (Lemma 1).

—For the exploitation, we consider a randomized adaptive selection rule (see Section 3). Specifically, in each step, the probability of selecting each pair is obtained by solving a saddle point optimization problem. We further develop a mirror descent algorithm for solving the optimization (see Section 3.4).

- Technically, we develop new analytical tools for quantifying the level crossing probability of a random function (e.g. likelihood function, martingale, or sub-martingale) double-indexed by model parameters and the sample size. As such a probability tends to zero, the problem falls into the rare-event analysis domain, where an exact exponential decay rate is challenging to obtain. Traditional methods, such as the ones adopted in [20, 45], are based on exponential change-of-measure of the log-likelihood ratio statistics, and are not directly applicable to the ranking problem considered here. The method we use in the proof combines a mixture-type of change-of-measure method recently proposed in [1, 37, 39] and large deviation results for martingales.

1.2. Related works

Sequential hypothesis testing, initiated by the seminal works of [58] and [59], is an important area of statistics for processing data taken in a sequential experiment, where the total number of observations is not fixed in advance. A sequential test is characterized by two components: (1) a stopping rule that decides when to stop the data collection process, and (2) a decision rule on choosing the hypothesis upon stopping. A large body of literature on sequential tests with two hypotheses has been developed, a partial list of which includes [27, 34, 50]. Sequential testing with more than two hypotheses and sequential multiple testing have been extensively studied in recent decades (see, e.g., [21, 22, 43, 53, 62]). For a comprehensive review on sequential analysis, we refer the readers to the surveys and books [29, 35, 52, 54] and references therein. In addition to optimizing over the stopping rule and final decision, [20] first introduces the adaptive design into the sequential testing framework, followed by a large body of literature; see, e.g. [2, 33, 45, 47, 57]. Sequential analysis finds many applications in different disciplines, including clinical trials, educational testing, and industrial quality control (see, e.g., [5, 6, 7, 36, 60, 63]).

Rank aggregation has been an active research problem in recent years (see, e.g., [16, 18, 19, 24, 25, 31, 46, 51] and references therein) that finds many applications to social choice, tournament play, search rankings, advertisement placement, etc. With the advent of crowdsourcing services, one can easily ask crowd workers to conduct comparisons among a few objects in an online fashion at a low cost [15, 17]. Therefore, active noisy sorting and ranking problems have received a lot of attentions in recent years. For example, Braverman et al. [12], Braverman and Mossel [13], Mao et al. [42] studied the active sorting problem where each query of (i,j) reveals the true ranking between i and j with a fixed probability $1/2+\gamma$ for some $\gamma > 0$, regardless of the distance between i and j. In contrast, our model associates each item i with a preference score (a.k.a. utility) θ_i. The comparison result between i and j would be based on the values of θ_i and θ_j according to some probabilistic model (e.g., see Eq. (2)). Jamieson and Nowak [30] studied the ranking problem with feature information for each item. Heckel et al. [26] investigated the active top-K ranking under a general class of nonparametric models and also established a lower bound on the number of comparisons for parametric models. However, as we mentioned, although rank aggregation has been extensively studied in the machine learning community, it has not been investigated under the sequential analysis framework, which incorporates the random stopping rule as a decision variable. The techniques developed in this work will enable a sequential rank procedure with optimal stopping and adaptive design.

1.3. Paper Organization

The rest of the paper is organized as follows. In Section 2, we introduce the setup of the problem. Section 3 presents the proposed policies and the theoretical results, and provides further discussions on the proof sketch and model misspecification. The concluding remarks are provided in Section 5. Technical proofs for the Theorems are provided in the Section 6. Proofs for all the lemmas are provided in the supplementary material.

2. Problem Setup We first introduce the comparison model and formulate the sequential ranking problem. Consider the task of inferring a global ranking over K items. Let $\mathcal{A} = \{(i,j) : i,j \in \{1,...,K\}, i<j\}$ be the set of pairs for comparison. At each time n $(n = 1,2,\ldots)$, a pair $a_n := (a_{n,1}, a_{n,2}) \in \mathcal{A}$ is selected for comparison. For example, $a_2 = (1,2)$ means that items 1 and 2 are compared at time two. The comparison outcome is denoted by a random variable $X_n \in \{0,1\}$, where $X_n = 1$ means item $a_{n,1}$ is preferred to item $a_{n,2}$ and $X_n = 0$ otherwise. The comparison outcome X_n is assumed to follow a ranking model, such as the widely used Bradley-Terry-Luce (BTL) model [11, 41] and Thurstone model [55]. Such a ranking model assumes that each item is associated with an unknown latent score $\theta_i \in \mathbb{R}$, for $i = 1,\ldots,K$, where the global rank of the K items is given by the rank of θ_1, ..., θ_K. The distribution of X_n is determined by θ_i and θ_j, when comparing pair (i,j). For example, given pair $a_n := (a_{n,1}, a_{n,2})$, the BTL model assumes that,

$$\begin{aligned}\mathbb{P}(X_n = 1) &= \frac{\exp(\theta_{a_{n,1}})}{\exp(\theta_{a_{n,1}}) + \exp(\theta_{a_{n,2}})};\\ \mathbb{P}(X_n = 0) &= \frac{\exp(\theta_{a_{n,2}})}{\exp(\theta_{a_{n,1}}) + \exp(\theta_{a_{n,2}})}.\end{aligned} \tag{2}$$

Under this model, $\theta_{a_{n,1}} > \theta_{a_{n,2}}$ means that item $a_{n,1}$ is preferred to item $a_{n,2}$, reflected by $\mathbb{P}(X_n = 1) > 0.5$. A common feature for many comparison models is that the distribution of the comparison between items i and j only depends on the pairwise differences $\theta_i - \theta_j$. Consequently, such models are not identifiable up to a location shift. To overcome this issue, we fix $\theta_1 = 0$ and treat $\boldsymbol{\theta} = (\theta_2, ..., \theta_K)$ as the unknown model parameters. The result of this paper applies to a wide class of comparison models and thus we denote the probability mass function of the comparison outcome x given pair a as $f^a_{\boldsymbol{\theta}}(x)$. We point out that while we focus on the case where the distribution of the pairwsise comparison only depends on $\theta_{a_n,1} - \theta_{a_n,2}$, our methods and results can be extended to more general cases without this requirement.

We now describe components in a sequential design for rank aggregation: an adaptive selection rule A, a stopping time T, and a decision rule R on the global rank. For the adaptive selection rule A, we consider the class of randomized adaptive selection rules, which contains deterministic selection rules as special cases. In particular, let $A = \{\boldsymbol{\lambda}_n : n = 1,2,...\}$, where $\boldsymbol{\lambda}_n = (\lambda_n^{i,j})_{(i,j)\in\mathcal{A}} \in \Delta$ denotes the probability of selecting the pair (i,j). Here, $\Delta = \{(\lambda^{i,j}) : \sum_{(i,j)\in\mathcal{A}} \lambda^{i,j} = 1,\ \lambda^{i,j} \geq 0\}$ is a probability simplex over $K(K-1)/2$ pairs. At each time n, a pair a_n is selected according to the categorical distribution $\boldsymbol{\lambda}_n$, where $\boldsymbol{\lambda}_n$ adapts to the filtration sigma-algebra generated by the selected pairs and the observed outcomes, that is, $\mathcal{F}_n = \sigma(X_1, ..., X_{n-1}, a_1, ..., a_{n-1})$. The adaptive comparison process will stop at time T, a stopping time with respect to the filtration $\{\mathcal{F}_n\}_{n\geq 0}$. It is worthwhile to note that the random stopping time T is also the number of samples being collected. Upon stopping, one needs to make a decision $R := (R_1, \ldots, R_K)$, the global ranking of the

K items. For example, when $K=3$, $R=(3,1,2)$ means that one decides $\theta_2>\theta_3>\theta_1$. We further denote P_K the set of permutations over $\{1,\dots K\}$ and thus $R\in P_K$. The adaptive selection rule $A=\{\boldsymbol{\lambda}_n : n=1,2,...\}$, the stopping time T, and the decision R together form a *sequential ranking policy*, denoted by $\pi=(A,T,R)$.

The performance of a sequential ranking policy is measured via its ranking accuracy and the expected stopping time. Specifically, we measure the ranking accuracy by Kendall's tau distance [32], which is one of the most widely used measures for ranking consistency. More precisely, for each $R=(R_1,...,R_K)\in P_K$, we convert it to the binary decisions over pairs $\{R_{i,j}\in\{0,1\}: i,j\in\{1,...,K\}, i<j\}$, where $R_{i,j}=I(R_i<R_j)$, and $R_{i,j}=1$ means that item i is preferred to item j. For example, if $R=(3,1,2)$, we have $R_{1,2}=0$ and $R_{2,3}=1$. The Kendall's tau distance between R and the true ranking induced by $(\theta_1,\dots,\theta_K)$ is defined by

$$L_K(\{R_{i,j}\})=\sum_{i<j}\{I(\theta_i>\theta_j)(1-R_{i,j})+I(\theta_i<\theta_j)R_{i,j}\}. \tag{3}$$

On the other hand, the loss function associated with the random sample size T is defined as,

$$L_c(T)=c\times T, \tag{4}$$

where the constant $c>0$ indicates the *relative* cost of conducting one more pairwise comparison. The choice of c depends on the nature of the ranking problem. Generally, if obtaining each sample is expensive comparing to the cost due to the inaccuracy of the ranking, then a large c will be chosen and vise versa. Note that c is not a tuning parameter to optimize over.

We define the risk associated with a sequential ranking policy under the Bayesian decision framework, in which the model parameter $\boldsymbol{\theta}$ is assumed to be random and follows a prior distribution. To avoid confusion, we write Θ when $\boldsymbol{\theta}$ is viewed as random, and denote by $\rho(\boldsymbol{\theta})$ the prior density function of $\Theta=(\Theta_2,...,\Theta_K)$. Recall that we have fixed $\Theta_1=0$ to ensure identifiability. The Bayes risk combines the risks associated with Kendall's tau distance of the decision and the sampling cost,

$$\begin{aligned}V_c(\rho,\pi)&=\mathbb{E}^\pi\left(L_K(\{R_{i,j}\})+L_c(T)\right)\\&=\mathbb{E}^\pi\Big\{\sum_{i<j}I(\Theta_i>\Theta_j)(1-R_{i,j})+I(\Theta_i<\Theta_j)R_{i,j}\Big\}+c\mathbb{E}^\pi T,\end{aligned} \tag{5}$$

where the expectation $\mathbb{E}^\pi$ is taken under the policy π, with respect to the randomness of the selected pairs, the observed comparison results, the stopping time, as well as the prior distribution ρ. Of particular interest is the minimum risk under the optimal sequential ranking policy given the prior distribution of Θ and sampling cost c

$$V_c^*(\rho)=\inf_\pi V_c(\rho,\pi). \tag{6}$$

For any given cost c, obtaining an analytical form of an optimal policy that achieves $V^*(\rho, c)$ is typically infeasible. Following the literature of sequential analysis, a policy is usually evaluated by the notion of *asymptotic optimality* [20]. In particular, a policy π is said to be asymptotically optimal if

$$\lim_{c\to 0}\frac{V_c(\rho,\pi)}{V_c^*(\rho)}=1, \tag{7}$$

i.e. when the relative sampling cost converges to 0. It is worthwhile noting that in the construction of our policy, we certainly allow the cost c to be non-zero. The notion of asymptotic optimality in (7) has been widely adopted in the sequential analysis literature as an optimality criterion (see, e.g., [20, 33, 45, 50]). The limiting process $c \to 0$ should be interpreted as the *sample size n goes to infinity*, which is a very common limiting process in statistical asymptotic theory. In asymptotic theory, letting n grow to infinity is only for the theoretical study of the properties of an estimator, while in practice no dataset has an infinite number of observations.

3. Sequential Policies and Asymptotic Optimality

In Section 3.1, we propose two sequential ranking policies π_1 and π_2. The asymptotic optimality of the two policies is presented in Section 3.2. Then we provide the proof sketch in Section 3.3, the optimization algorithm for efficient computation in Section 3.4, and the discussions on model misspecification in Section 3.5.

3.1. Two Sequential Policies

We first introduce some notations. Let W be the support of the prior probability density function ρ, i.e., $W=\overline{\{\boldsymbol{\theta}:\rho(\boldsymbol{\theta})>0\}}$, where $\bar{E}$ denotes the closure of a set E. We further define the set $W_{i,j}=\{\boldsymbol{\theta}:\theta_i\geq\theta_j\}\cap W$ for all $i,j\in\{1,...,K\}$. It is worthwhile to note that $W_{i,j}$ and $W_{j,i}$ are different sets and their union is the set W. Given a sequence of selected pairs $a_1,...,a_n$ and observed comparisons $X_1,...,X_n$, the log-likelihood function is defined as,

$$l_n(\boldsymbol{\theta})=\sum_{i=1}^{n}\log f_{\boldsymbol{\theta}}^{a_i}(X_i),$$

and the corresponding maximum likelihood estimator $\widehat{\boldsymbol{\theta}}^{(n)}=(\widehat{\theta}_2^{(n)},...,\widehat{\theta}_K^{(n)})$ is

$$\widehat{\boldsymbol{\theta}}^{(n)}=\arg\sup_{\boldsymbol{\theta}\in W} l_n(\boldsymbol{\theta}). \tag{8}$$

In what follows, we present our proposed sequential policies in terms of the proposed stopping time T, selection rule A, and ranking decision R.

3.1.1. Stopping Times

We then introduce two stopping times based on the generalized likelihood ratio statistic,

$$T_1=\inf\Big\{n>1:\sum_{(i,j)\in\mathcal{A}}\exp\{-|\sup_{\boldsymbol{\theta}\in W_{i,j}} l_n(\boldsymbol{\theta})-\sup_{\boldsymbol{\theta}\in W_{j,i}} l_n(\boldsymbol{\theta})|\}\leq e^{-h(c)}\Big\}, \tag{9}$$

and

$$T_2 = \inf\Big\{n > 1 : \min_{(i,j)\in\mathcal{A}} |\sup_{\boldsymbol{\theta}\in W_{i,j}} l_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in W_{j,i}} l_n(\boldsymbol{\theta})| \ge h(c)\Big\}, \tag{10}$$

where $h(c) = |\log c|(1 + |\log c|^{-\alpha})$ for some constant $\alpha \in (0,1)$ and c is the relative cost introduced in (4). We note that T_2 is obtained by replacing the summation in T_1 by maximization and taking log and minus on both sides. Intuitively, the stopping rule T_2 stops when the likelihood can tell whether $\theta_i \ge \theta_j$ or vice versa for each pair (i,j).

3.1.2. Ranking Decision Upon stopping, the decision about the global rank is made according to the rank of MLE at the stopping time T ($T = T_1$ or T_2). That is,

$$R = r(\widehat{\boldsymbol{\theta}}^{(T)}), \tag{11}$$

where the function $r(\boldsymbol{\theta}) : \mathbb{R}^{K-1} \to P_K$ gives the rank of $(0, \theta_2, ..., \theta_K)$. More precisely, $r(\boldsymbol{\theta}) = (r_1, \dots, r_K) \in P_K$, satisfying $\theta_{r_1} \ge \theta_{r_2} \ge \dots \ge \theta_{r_K}$, where $\theta_1 = 0$.

3.1.3. Randomized Selection Rule We proceed to the randomized selection rule A, which is obtained by solving an optimization program. For a given $\boldsymbol{\theta}$, we define function $D(\boldsymbol{\theta})$,

$$D(\boldsymbol{\theta}) = \max_{\boldsymbol{\lambda}\in\Delta} \min_{\widetilde{\boldsymbol{\theta}}\in W: r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})} \sum_{(i,j)} \lambda^{i,j} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}), \tag{12}$$

where $D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}})$ is the Kullback-Leibler (KL) divergence from $f^{i,j}_{\widetilde{\boldsymbol{\theta}}}(\cdot)$ to $f^{i,j}_{\boldsymbol{\theta}}(\cdot)$, i.e.

$$D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}) := \sum_{x\in\{0,1\}} f^{i,j}_{\boldsymbol{\theta}}(x) \log \frac{f^{i,j}_{\boldsymbol{\theta}}(x)}{f^{i,j}_{\widetilde{\boldsymbol{\theta}}}(x)}$$

and $f^{i,j}_{\boldsymbol{\theta}}(x)$ denotes the probability mass function when the pair (i,j) is selected. We further define

$$\boldsymbol{\lambda}^*(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\lambda}\in\Delta} \min_{\widetilde{\boldsymbol{\theta}}\in W: r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})} \sum_{(i,j)} \lambda^{i,j} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}), \tag{13}$$

and

$$\widehat{\boldsymbol{\lambda}}_n = (\widehat{\lambda}^{i,j}_n) = \boldsymbol{\lambda}^*(\widehat{\boldsymbol{\theta}}^{(n-1)}). \tag{14}$$

That is, $\boldsymbol{\lambda}^*(\boldsymbol{\theta})$ is the solution to the optimization problem (12), and $\widehat{\boldsymbol{\lambda}}_n$ is the solution to the optimization problem given the MLE based on the previous $n-1$ samples. The objective function in (12) is a weighted KL divergence for all pairs with the weights $\lambda^{i,j}$. The inner minimization problem is taken over all the parameter vector $\widetilde{\boldsymbol{\theta}} \in W$, for which the induced rank $r(\widetilde{\boldsymbol{\theta}})$ is different from that of $\boldsymbol{\theta}$. At each time n, given the MLE $\widehat{\boldsymbol{\theta}}^{(n-1)}$, we compute $\widehat{\boldsymbol{\lambda}}_n$, which is the maximizer of $\boldsymbol{\lambda} \in \Delta$ in $D(\widehat{\boldsymbol{\theta}}^{(n-1)})$. We elaborate on the intuition behind the optimization in (12). First, for each $\boldsymbol{\theta}$, $\sum_{(i,j)} \lambda^{i,j} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}})$ gives the drift of the log-likelihood ratio statistics between $f_{\boldsymbol{\theta}}$ and $f_{\widetilde{\boldsymbol{\theta}}}$

under the model $f_{\boldsymbol{\theta}}$ and a randomized sampling scheme specified by $\boldsymbol{\lambda}$, which is also the mutual information between $f_{\boldsymbol{\theta}}$ and $f_{\widetilde{\boldsymbol{\theta}}}$ when the pair is selected according to $\boldsymbol{\lambda}$. Minimizing the inner part with respect to $\widetilde{\boldsymbol{\theta}}$ over the set $\{\widetilde{\boldsymbol{\theta}} \in W : r(\widetilde{\boldsymbol{\theta}}) \neq r(\boldsymbol{\theta})\}$ provides a measure on the distinguishability of the rank of $\boldsymbol{\theta}$ under the sampling scheme $\boldsymbol{\lambda}$. Second, if the true model parameter is $\boldsymbol{\theta}$, we would like to choose a sampling scheme $\boldsymbol{\lambda}$ such that it provides the highest distinguishability obtained by the first step. Thus, we perform maximization in the outer part of (12). Finally, as the true model parameter $\boldsymbol{\theta}$ is unknown, we will replace $\boldsymbol{\theta}$ by the MLE based on the current information. In Section 3.4, we provide a mirror descent algorithm for solving (12).

Unfortunately, directly using $\widehat{\boldsymbol{\lambda}}_n$ in the selection rule A as the choice probability does not guarantee the asymptotic optimality. This is because $\widehat{\boldsymbol{\lambda}}_n$ does not guarantee sufficient exploration of all item pairs, which may lead to the imbalance between the exploration and exploitation for the sequential procedure. To fix this issue, we combine $\widehat{\boldsymbol{\lambda}}_n$ with an ϵ-greedy approach which is widely used in balancing exploration and exploitation in multi-armed bandit and decision-making problems (see, e.g., [61]). Specifically, an exploration probability $p \in (0,1)$ is chosen, which is typically small and may be chosen depending on the value of the relative sampling cost c. At each time n, with probability p, we select the next pair uniformly from $\mathcal{A}$. With probability $1-p$, the next pair is selected according to the categorical distribution specified by $\widehat{\boldsymbol{\lambda}}_n$. In other words, for each pair (i,j), the choice probability of the selection rule at time n is given by

$$\lambda_n^{i,j} = p\frac{2}{K(K-1)} + (1-p)\widehat{\lambda}_n^{i,j}.$$

Remark 1. We clarify that the proposed 'ϵ-greedy' algorithm is one of asymptotically optimal exploration methods, and there may be other exploration methods with similar theoretical properties. For example, the ϵ-greedy algorithm with the exploration probability decaying at a rate $n^{-\beta}$ when the sample size is n may be asymptotically optimal for a range of $\beta > 0$. Theoretical properties of these additional exploration methods is an interesting problem and worth further investigation.

We call the above selection rule A_p, where the subscript emphasizes its dependence on the exploration rate p. The two proposed sequential ranking policies are defined by $\pi_1 := (A_p, T_1, R)$ and $\pi_2 := (A_p, T_2, R)$. The proposed sequential ranking policies are summarized in Algorithm 1, where the prior information of Θ is only utilized through its support W in Steps 1 and 2. Algorithm 1 is an iterative algorithm, which runs in T_1 (or T_2) iterations, where T_1 (or T_2) is a data-dependent stopping time. The major computational complexity for each iteration arises from solving two optimization problems in Step 1 and 2. The Step 1 is a standard maximum likelihood estimation,

```
Algorithm 1: Sequential Ranking Policy
Input: The probability mass (density) function $f^a_{\boldsymbol{\theta}}(x)$ for any pair $a \in \mathcal{A}$, the probability
  $p \in (0,1)$ in $\epsilon$-greedy, and the support $W$ of $\rho(\boldsymbol{\theta})$.
Initialization: Uniformly sample a pair $a_1$ at random and observe the comparison outcome
  $X_1$.
Iterate For $n = 2, 3, \ldots$ until the stopping time $T$ in (9) (or (10)) is reached.
  1. Compute the MLE based on the previous $n-1$ comparisons:
       $$\widehat{\boldsymbol{\theta}}^{(n-1)} = \arg\sup_{\boldsymbol{\theta}\in W} l_{n-1}(\boldsymbol{\theta}).$$
  2. Compute
       $$\widehat{\boldsymbol{\lambda}}_n = \arg\max_{\boldsymbol{\lambda}\in\Delta} \min_{\widetilde{\boldsymbol{\theta}}\in W : r(\widetilde{\boldsymbol{\theta}})\neq r(\widehat{\boldsymbol{\theta}}^{(n-1)})} \sum_{(i,j)\in\mathcal{A}} \lambda^{i,j} D^{i,j}(\widehat{\boldsymbol{\theta}}^{(n-1)} \| \widetilde{\boldsymbol{\theta}}). \qquad (15)$$
  3. Flip a coin with head probability $p$.
  • If the outcome is head, select the pair $a_n$ uniformly at random over all pairs from $\mathcal{A}$.
  • Otherwise, select the pair $a_n$ according to the categorical distribution specified by $\widehat{\boldsymbol{\lambda}}_n$.
  4. Observe the comparison result $X_n$ and update the likelihood function $l_n(\boldsymbol{\theta})$.
Output: The rank $R = r(\widehat{\boldsymbol{\theta}}^{(T)})$, i.e., the global rank induced by $\widehat{\boldsymbol{\theta}}^{(T)}$.
```

which depends on the structure of the loss function l and the constraint W. The computation for solving (13) will be discussed in Section 3.4. The proofs of the theoretical results are provided in Section 6.

3.2. Asymptotic Optimality

This section contains the main results of the paper, including (1) a lower bound on the risk of a general sequential ranking procedure, and (2) theoretical analysis of the proposed procedures, which leads to their asymptotic optimality. The asymptotic optimality of the proposed method is established through the following theorems, which will be introduced later in this section. Theorem 1 provides an asymptotic lower bound for the Bayes risk of an arbitrary sequential ranking policy. Theorems 2 and 3 provide asymptotic upper bounds for the proposed procedures in terms of their expected Kendall's tau and expected stopping time, respectively. These upper bounds together lead to an asymptotic upper bound for the Bayes risk of the proposed procedures that matches the lower bound in Theorem 1. As the asymptotic lower and upper bounds match, we conclude that the proposed method is asymptotically optimal in Corollary 1. As a by-product, an exponential deviation bound for the MLE over a time window is also obtained in Lemma 1. The assumptions for our results are described and discussed.

Notations Throughout the rest of the paper, we write $a_c = O(b_c)$ for two sequences a_c and b_c if $|a_c|/|b_c|$ is bounded, uniformly in $\boldsymbol{\theta}$, as $c \to 0$. Similarly, we write $a_c = \Omega(b_c)$ if $a_c > 0$, $b_c > 0$ and $b_c = O(a_c)$. We will also write $a_c = o(b_c)$ if $a_c/b_c \to 0$ uniformly in $\boldsymbol{\theta}$. The norm $\|\cdot\|$ indicates the ℓ_2 vector norm. Throughout the paper, we use the uppercase Greek letter Θ to indicate the *random* score parameter and the lowercase Greek letter $\boldsymbol{\theta}$ to denote a *deterministic* vector.

Main results We first describe the assumptions. For technical needs, we make some regularity conditions on the prior distribution $\rho(\boldsymbol{\theta})$. Recall that we have fixed $\theta_1 = 0$ and let $\boldsymbol{\theta} = (\theta_2, ..., \theta_K) \in \mathbb{R}^{K-1}$ be the unknown model parameters.

Assumption 1. *The support $W := \overline{\{\boldsymbol{\theta} \in \mathbb{R}^{K-1} : \rho(\boldsymbol{\theta}) > 0\}}$ is a compact set in $\mathbb{R}^{K-1}$, where $\bar{E}$ denotes the closure of a set E. In addition, for any permutation $\sigma \in P_K$, $(\{\boldsymbol{\theta} \in \mathbb{R}^{K-1} : r(\boldsymbol{\theta}) = \sigma\} \cap W)^\circ \neq \emptyset$, where E° denotes the interior of a set E.*

Assumption 2. *There exists a constant $\delta_b > 0$ such that for all $s > 0$ and $\boldsymbol{\theta} \in W$, $m(B(\boldsymbol{\theta}, s) \cap W) \geq \min\{\delta_b s^{K-1}, 1\}$, where $B(\boldsymbol{\theta}, s)$ denotes the open ball centered at $\boldsymbol{\theta}$ with radius s and $m(\cdot)$ denotes the Lebesgue measure.*

Assumption 3. *The function $\log f^a_{\boldsymbol{\theta}}(x)$ is continuously differentiable in $\boldsymbol{\theta}$ for all x uniformly. That is,*

$$\sup_{\boldsymbol{\theta} \in W, a \in \mathcal{A}, x} \|\nabla_{\boldsymbol{\theta}} \log f^a_{\boldsymbol{\theta}}(x)\| < \infty.$$

In addition, $\inf_{\boldsymbol{\theta} \in W, a \in \mathcal{A}, x} f^a_{\boldsymbol{\theta}}(x) > 0$.

Assumption 4. $\inf_{\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in W : r(\widetilde{\boldsymbol{\theta}}) \neq r(\boldsymbol{\theta})} \max_{(i,j)} D^{i,j}(\boldsymbol{\theta} \| \widetilde{\boldsymbol{\theta}}) > 0$.

Assumption 5. $\inf_{\boldsymbol{\theta} \in W^\circ} \rho(\boldsymbol{\theta}) > 0$ *and* $\sup_{\boldsymbol{\theta} \in W} \rho(\boldsymbol{\theta}) < \infty$.

We provide some remarks on the above regularity assumptions. Assumption 1 requires the prior distribution for Θ to have a bounded support, which has a non-empty interior for each rank. Assumption 2 avoids the support W being singular. Assumption 3 requires the smoothness of the likelihood function. It also requires the comparison probability is bounded away from 0 and 1. Assumption 4 requires that there is no tie in the support of the prior distribution. This is a standard assumption in sequential analysis, which corresponds to the classic "indifference zone" assumption in sequential hypothesis testing [33, 40, 50]. In particular, the "indifference zone" condition assumes that the null and alternative hypotheses are separated in the sense that the Kullback-Leibler divergence between the two hypotheses is positive, and if the true model parameter is in between the two hypotheses, then it is considered to be indifferent for selecting the null and alternative hypothesis. For example, for any $\delta > 0, \kappa > 0$, the set

$$W = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq \kappa \text{ and } \forall i \neq j \text{ such that } |\theta_i - \theta_j| \geq \delta\} \tag{16}$$

satisfies Assumptions 1, 2 and 4. Assumption 5 requires the prior distribution to have a positive density function (bounded from zero) over the support. For instance, for the set W described in (16), the uniform prior over W satisfies the Assumption 5. In addition, with such a uniform prior over W, the BTL model defined in (2) satisfies Assumptions 3 and 4. It is worthwhile to note that these technical assumptions are mainly for the theoretical development, while the proposed adaptive ranking policies are applicable in practice regardless of the conditions on W.

Recall the definition of $D(\boldsymbol{\theta})$ in (12). We further define

$$t_c(\boldsymbol{\theta}) = \frac{|\log c|}{D(\boldsymbol{\theta})}. \tag{17}$$

Note that under the Assumption 4, $t_c(\boldsymbol{\theta})$ is always finite. Intuitively, for small c, $|\log c|/\{\min_{\widetilde{\boldsymbol{\theta}}\in W: r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})} \sum_{(i,j)} \lambda^{i,j} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}})\}$ is approximately the smallest expected sample for the simple against simple hypothesis testing problem $H_0 : X_n \sim f_{\boldsymbol{\theta}}^{a_n}$ against $H_1 : X_n \sim f_{\widetilde{\boldsymbol{\theta}}}^{a_n}$ for some $r(\widetilde{\boldsymbol{\theta}}) \neq r(\boldsymbol{\theta})$, where a_n is sampled from $\boldsymbol{\lambda}$. Note that $t_c(\boldsymbol{\theta}) = |\log c|/D(\boldsymbol{\theta}) = \inf_{\boldsymbol{\lambda}\in\triangle}[|\log c|/\{\min_{\widetilde{\boldsymbol{\theta}}\in W: r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})} \sum_{(i,j)} \lambda^{i,j} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}})\}]$. Thus, $t_c(\boldsymbol{\theta})$ is approximately the smallest expected sample size for distinguishing the global rank of $\boldsymbol{\theta}$ from other ranks with an adaptive selection step. We formalize the above heuristic arguments in the following Theorem 1–Theorem 3.

We first present a lower bound on the minimal Bayes risk $V_c^*(\rho)$ defined in (6).

THEOREM 1. *Under Assumptions 1-5, we have*

$$\liminf_{c\to 0} \frac{V_c^*(\rho)}{c\mathbb{E}t_c(\Theta)} \geq 1,$$

where $\mathbb{E}t_c(\Theta) = \int_W t_c(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}$.

Recall the definition in (7) that a policy π is said to be asymptotically optimal if $V_c(\pi,\rho) = (1+o(1))V_c^*(\rho)$ as $c \to 0$. Thus, to show a policy π is indeed asymptotically optimal, we only need to show that $V_c(\pi,\rho) = (1+o(1))c\mathbb{E}t_c(\Theta)$ as $c \to 0$, according to Theorem 1. We proceed to show that the proposed sequential ranking method is asymptotically optimal. In Section 3.1, we propose two policies $\pi_1 = (A_p, T_1, R)$, $\pi_2 = (A_p, T_2, R)$. Their risks consist of two parts, the expected Kendall's tau and the expected sample size.

ASSUMPTION 6. *For each $\boldsymbol{\theta},\boldsymbol{\theta}' \in W$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, there exists $a \in \mathcal{A}$ that can distinguish $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. That is, $\sum_{a\in\mathcal{A}} D^a(\boldsymbol{\theta}\|\boldsymbol{\theta}') > 0$ for $\boldsymbol{\theta},\boldsymbol{\theta}' \in W$. In addition, there is a constant $\delta > 0$ such that $\sum_{a\in\mathcal{A}} D^a(\boldsymbol{\theta}\|\boldsymbol{\theta}') \geq \delta\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2$.*

Assumption 6 requires the identifiability of the model, which is critical for the consistency of the MLE. For the BTL model described in (2), Assumption 6 is satisfied after fixing $\theta_1 = 0$. In what follows, Theorems 2 and 3 provide asymptotic upper bounds for the expected Kendall's tau and expected stopping time of the proposed method, respectively.

THEOREM 2. *Under Assumptions 1- 6, we consider a policy $\pi_l = (A, T_l, R)$ $(l = 1, 2)$, where we choose $p \propto |\log c|^{-\frac{1}{2}+\delta_0}$ for some δ_0 satisfying $0 < \delta_0 < \frac{1}{2}$ in Algorithm 1 and $R = \{R_{i,j}\}$. Then,*

$$\mathbb{E}L_K(\{R_{i,j}\}) = O(c) \text{ for } l = 1, 2.$$

THEOREM 3. *Under Assumptions 1- 6, we consider a policy $\pi_l = (A, T_l, R)$ $(l = 1, 2)$, where we choose $p \propto |\log c|^{-\frac{1}{2}+\delta_0}$ for some δ_0 satisfying $0 < \delta_0 < \frac{1}{2}$ in Algorithm 1 and $R = \{R_{i,j}\}$. Then,*

$$\limsup_{c\to 0} \frac{\mathbb{E}T_l}{\mathbb{E}t_c(\Theta)} \le 1 \text{ for } l = 1, 2.$$

Combining this with the asymptotic lower bound on the minimal Bayes risk in Theorem 1, and noticing that $\lim_{c\to 0} \mathbb{E}t_c(\Theta) = \infty$, we arrive at the asymptotic optimality of the proposed policies.

COROLLARY 1. *Under Assumption 1-6, if we choose $p \propto |\log c|^{-\frac{1}{2}+\delta_0}$ for some δ_0 satisfying $0 < \delta_0 < \frac{1}{2}$, then $\pi_l = (A_p, T_l, R)$, $l = 1, 2$, are asymptotically optimal policies.*

Consistency of MLE An auxiliary result obtained in deriving the upper bound for the expected sample size is the following exponential bound for the MLE over a time window.

Lemma 1 *Let $m \ge n$ and let $\varepsilon_{\lambda,m,n}$ be a sequence of real numbers such that $\min_{n\le t\le m,(i,j)} \lambda_t^{i,j} \ge \varepsilon_{\lambda,m,n}$. In addition, let $\delta_{m,n}$ be a sequence of positive numbers such that $n\varepsilon_{\lambda,m,n}\delta^2_{m,n} \to \infty$ as $n \to \infty$. Then,*

$$\mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{n\le t\le m} \|\widehat{\boldsymbol{\theta}}^{(t)} - \boldsymbol{\theta}\| \ge \delta_{m,n}\Big) \le e^{-\Omega(n\epsilon^2_{\lambda,m,n}\delta^4_{m,n})} \times O(m^K),$$

where we denote $\mathbb{P}_{\boldsymbol{\theta}}(\cdot)$ the conditional probability $\mathbb{P}(\cdot|\Theta = \boldsymbol{\theta})$ and $\widehat{\boldsymbol{\theta}}^{(t)}$ is the MLE defined in (8). Moreover, this upper bound is uniform for $\boldsymbol{\theta} \in W$.

The proof is provided in the supplementary material. From the above lemma, we can derive exponential upper bounds concerning the uniform consistency of $\widehat{\boldsymbol{\theta}}^{(t)}$. In particular, if we let $\delta_{m,n}$ be a fixed positive constant and $\varepsilon^2_{\lambda,m,n} \gg m^{-1}\log m$ as $m \to \infty$, then we can show $\sup_{t\ge n} \|\widehat{\boldsymbol{\theta}}^{(t)} - \boldsymbol{\theta}\| \to 0$ in probability as $n \to \infty$ with additional steps.

3.3. Proof strategy We briefly explain the proof strategy for each of the main theorems. Theorem 1 provides a lower bound on $V(\rho, \pi)$ for an arbitrary policy $\pi = (A, T, R)$ by discussing two cases: $\mathbb{E}L_K(R) \ge c|\log c|^2$ and $\mathbb{E}L_K(R) < c|\log c|^2$. For the first case, Theorem 1 is easily justified. The main technicalities are in the second case, where the main step is to develop an upper bound for the probability $\mathbb{P}\big(T \le (1-\delta)\mathbb{E}t_c(\Theta)\big)$ for any constant $\delta > 0$. Heuristically, we argue that whenever $\mathbb{E}L_K(R)$ is small, it implies that the likelihood ratios between the conditional probability measures of data given that Θ has different ranking patterns will be relatively large, which cannot be achieved with a relatively small sample size T. The rigorous proof for this heuristic statement is done through a change-of-measure argument and a large deviation bound for the likelihood ratio.

The proof of Theorem 2 is based on the analysis of the expected Kendall's tau under the stopping times T_1 and T_2. The analysis under T_2 is based on the equation

$$\mathbb{E}L_K(R)=\sum_{i,j}\int_{\boldsymbol{\theta}\notin W_{j,i}}\mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_2}(\widetilde{\boldsymbol{\theta}})-\sup_{\boldsymbol{\theta}'\in W_{i,j}} l_{T_2}(\boldsymbol{\theta}')>h(c)\Big)\rho(\boldsymbol{\theta})d\boldsymbol{\theta},$$

followed by developing an upper bound for the probability $\mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_2}(\widetilde{\boldsymbol{\theta}})-\sup_{\boldsymbol{\theta}'\in W_{i,j}} l_{T_2}(\boldsymbol{\theta}')>h(c)\Big)$, where $h(c)=|\log c|(1+|\log c|^{-\alpha})$ is slightly larger than $|\log c|$. Intuitively, thanks to the ε-greedy algorithm and the stopping time, a sufficient amount of information has been collected upon stopping so that the error probability is well-controlled. The analysis under T_1 is similar and we omit the details here.

To prove Theorem 3, we first note that $p\propto|\log c|^{-\frac{1}{2}+\delta_0}$ for some positive δ_0 in the ε-greedy algorithm. Thus, we could apply Lemma 1 and show that the MLE $\widehat{\boldsymbol{\theta}}^{(t)}$ is consistent with an exponential error bound. Roughly, this justifies that $\widehat{\boldsymbol{\lambda}}_n$ defined in (14) is close to $\boldsymbol{\lambda}^*(\boldsymbol{\theta})$ given $\Theta=\boldsymbol{\theta}$. Thus, the expected sample size $\mathbb{E}(T_i|\Theta=\boldsymbol{\theta})$ approximates the one given by the selection rule $\boldsymbol{\lambda}^*(\boldsymbol{\theta})$ that can be further approximated by $h(c)/D(\boldsymbol{\theta})=(1+o(1))t_c(\boldsymbol{\theta})$, where we recall that $t_c(\boldsymbol{\theta})=|\log c|/D(\boldsymbol{\theta})$ is defined in (17). We can then justify Theorem 3 by taking the expectation with respect to the prior distribution of Θ on both sides.

3.4. Optimization in Algorithm 1

In this section, we show that the key optimization problem in (13) can be solved efficiently using the mirror descent algorithm (see, e.g., [8]).

Let us first consider the inner optimization problem

$$\widetilde{\boldsymbol{\theta}}^0(\boldsymbol{\lambda})\in\underset{\widetilde{\boldsymbol{\theta}}\in W:r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})}{\arg\max}\ -\sum_{(i,j)}\lambda^{i,j}D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}),\tag{19}$$

in step 1 of Algorithm 2. We clarify that in this optimization, $\boldsymbol{\theta}$ is fixed, $\widetilde{\boldsymbol{\theta}}$ is the decision variable we would like to optimize with, and the resulting $\widetilde{\boldsymbol{\theta}}^0(\boldsymbol{\lambda})$ depends on $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. For almost all the popular comparison models, the objective function $-\sum_{(i,j)}\lambda^{i,j}D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}})$ is smooth in $\widetilde{\boldsymbol{\theta}}$. Moreover, the objective function is also concave in $\widetilde{\boldsymbol{\theta}}$ for comparison models in an exponential family form (e.g., the BTL model in (2)). When the support $\{\widetilde{\boldsymbol{\theta}}\in W:r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})\}$ can be written as the union of a finite number of convex sets (see Eq. (20) below), (19) can be obtained by solving finite maximization problems, each with a smooth concave objective function constrained in a convex set. Therefore, from now on, we assume that the inner optimization problem can be solved.

We then discuss the outer optimization problem

$$\min_{\boldsymbol{\lambda}\in\triangle} h(\boldsymbol{\lambda}),\ \ h(\boldsymbol{\lambda})=\max_{\widetilde{\boldsymbol{\theta}}\in W:r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})}\phi(\boldsymbol{\lambda},\widetilde{\boldsymbol{\theta}}),\ \ \phi(\boldsymbol{\lambda},\widetilde{\boldsymbol{\theta}})=-\sum_{(i,j)}\lambda^{i,j}D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}).$$

```
Algorithm 2: Mirror Descent Algorithm for Solving Eq. (13)
  Input: The MLE estimator $\boldsymbol{\theta}$ and total number of iterations $m$.
  Initialization: A starting point $\boldsymbol{\lambda}^0 \in \Delta$ and a constant $c_0 > 0$.
  Iterate For $t = 1, 2, \ldots, m$:
    1. Compute the maximizer:
         $\widetilde{\boldsymbol{\theta}}^0(\boldsymbol{\lambda}^{t-1}) \in \operatorname*{arg\,max}_{\widetilde{\boldsymbol{\theta}} \in W : r(\widetilde{\boldsymbol{\theta}}) \neq r(\boldsymbol{\theta})} -\sum_{(i,j)} \lambda^{i,j,t-1} D^{i,j}(\boldsymbol{\theta} \| \widetilde{\boldsymbol{\theta}})$
    2. Compute the sub-gradient $\mathbf{g}(\boldsymbol{\lambda}^{t-1})$ where $\mathbf{g}(\boldsymbol{\lambda}^{t-1})_{i,j} = -D^{i,j}(\boldsymbol{\theta} \| \widetilde{\boldsymbol{\theta}}^0(\boldsymbol{\lambda}^{t-1}))$
    3. Update for $\boldsymbol{\lambda}^t$:
         $\boldsymbol{\lambda}^t = \operatorname*{arg\,min}_{\boldsymbol{\lambda} \in \Delta} \left\{ \eta_t \langle \mathbf{g}(\boldsymbol{\lambda}^{t-1}), \boldsymbol{\lambda} \rangle + D(\boldsymbol{\lambda} \| \boldsymbol{\lambda}^{t-1}) \right\},$   (18)
  where $\eta_t = \frac{c_0}{\sqrt{t}}$ and $D(\boldsymbol{\lambda} \| \boldsymbol{\lambda}^{t-1})$ is the KL divergence between $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^{t-1}$, i.e.,
  $D(\boldsymbol{\lambda} \| \boldsymbol{\lambda}^{t-1}) = \sum_{i,j} \lambda_{i,j} \log \frac{\lambda^{i,j}}{\lambda^{i,j,t-1}}$
  Output: The solution $\widehat{\boldsymbol{\lambda}} = \frac{1}{m} \sum_{t=1}^m \boldsymbol{\lambda}^t$.
```

When $\phi(\boldsymbol{\lambda}, \widetilde{\boldsymbol{\theta}})$ is a continuous and bounded function and the set W is compact, further noting that $\phi(\boldsymbol{\lambda}, \widetilde{\boldsymbol{\theta}})$ is convex in $\boldsymbol{\lambda}$ for every $\widetilde{\boldsymbol{\theta}}$, $h(\boldsymbol{\lambda})$ is a convex function in $\boldsymbol{\lambda}$, by the Danskin's Theorem (see Proposition B.25 in [9]). Moreover, for a given $\boldsymbol{\lambda}$, let $\widetilde{\boldsymbol{\theta}}^0(\boldsymbol{\lambda}) \in \arg\max_{\widetilde{\boldsymbol{\theta}} \in W : r(\widetilde{\boldsymbol{\theta}}) \neq r(\boldsymbol{\theta})} \phi(\boldsymbol{\lambda}, \widetilde{\boldsymbol{\theta}})$ be one of the maximizers. Then, by Danskin's theorem, $\mathbf{g}(\boldsymbol{\lambda})$ with $\mathbf{g}(\boldsymbol{\lambda})_{i,j} = -D^{i,j}(\boldsymbol{\theta} \| \widetilde{\boldsymbol{\theta}}^0(\boldsymbol{\lambda}))$ is a sub-gradient of $h(\boldsymbol{\lambda})$, as used in step 2 of Algorithm 2.

Finally, (18) in step 3 of the algorithm has a closed-form solution, obtained by by writing down the KKT condition. That is,

$$\lambda^{i,j,t} = \frac{1}{C} \lambda^{i,j,t-1} \exp\left(-\eta_t \mathbf{g}(\boldsymbol{\lambda}^{t-1})_{i,j}\right),$$

where $\lambda^{i,j,t}$ is the (i,j)-th component of $\boldsymbol{\lambda}^t$ and the normalization constant $C = \sum_{i,j} \lambda^{i,j,t-1} \exp\left(-\eta_t \mathbf{g}(\boldsymbol{\lambda}^{t-1})_{i,j}\right)$.

From [8] or Theorem 4.2 from [14], we have the following convergence rate for Algorithm 2.

Proposition 1 ([8]). *Assuming the inner optimization in (19) can be solved exactly, the mirror descent algorithm in Algorithm 2 is guaranteed to converge to the optimal solution at the rate of* $O\left(\sqrt{1/t}\right)$. *That is when* $t = O(1/\epsilon^2)$, *we have* $h(\widehat{\boldsymbol{\lambda}}) - \min_{\boldsymbol{\lambda} \in \Delta} h(\boldsymbol{\lambda}) \leq \epsilon$.

We clarify that for W defined in the example (16), it is a union over exponentially many convex sets. Thus, the proposed method requires exponential computational time for such a W. On the other hand, it is possible to have a fully polynomial-computational-time algorithm if a mis-specified $\widetilde{W}$ is adopted (see (20) in the next section).

3.5. Model misspeficiation In practice, the support W of the prior distribution $\rho(\cdot)$ maybe unknown. In this case, we may choose

$$\widetilde{W} = \cup_{(i,j)} \widetilde{W}_{i,j} \text{ and } \widetilde{W}_{i,j} = \{\boldsymbol{\theta} : \theta_i \geq \theta_j\} \cap \{\boldsymbol{\theta} : |\theta_i| \leq M, 2 \leq i \leq K\} \tag{20}$$

in the sequential ranking policy for some reasonable positive constant M. With this mis-specified support of $\rho(\cdot)$, the resulting policy may not achieve the asymptotic lower bound of the Bayes risk presented in Theorem 1, due to the incomplete information. On the other hand, the Bayes risk of the resulting ranking procedure can still achieve the same order of the minimal Bayes risk as $c \to 0$. That is, $\limsup_{c\to 0} V_c(\rho,\pi)/V_c^*(\rho)$ is finite but greater than 1. The following assumption is made to guarantee that the function $f_{\boldsymbol{\theta}}^a(x)$ has similar regularity on $\widetilde{W}$ as on W. This assumption is mild. For example, it is satisfied for W, $\widetilde{W}$, and $f_{\boldsymbol{\theta}}^a(x)$ described in (16), (20), and (2), respectively.

Assumption 7. $\sup_{\boldsymbol{\theta}\in\widetilde{W}, a\in\mathcal{A}, x} \|\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}^a(x)\| < \infty$, $\inf_{\boldsymbol{\theta}\in\widetilde{W}, a\in\mathcal{A}, x} f_{\boldsymbol{\theta}}^a(x) > 0$, *and* $\inf_{\boldsymbol{\theta}\in W, \widetilde{\boldsymbol{\theta}}\in\widetilde{W}: r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})} \max_{(i,j)} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}) > 0$. *In addition, there is a constant* $\delta > 0$ *such that* $\sum_{a\in\mathcal{A}} D^a(\boldsymbol{\theta}\|\boldsymbol{\theta}') \geq \delta\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2$ *for all* $\boldsymbol{\theta} \in W$ *and* $\boldsymbol{\theta}' \in \widetilde{W}$.

Theorem 4. *If we replace* W *by* $\widetilde{W}$ *and replace* $W_{i,j}$ *by* $\widetilde{W}_{i,j}$ *(defined in* (20)*) in* (9), (10), (13) *as well as in Algorithm* 1*, and adopt the resulting policy* $\pi_l = (A, T_l, R)$ $(l = 1, 2)$ *with* $p \propto |\log c|^{-\frac{1}{2}+\delta_0}$ *for some* δ_0 *satisfying* $0 < \delta_0 < \frac{1}{2}$*, then under Assumptions* 1*,* 2*,* 5*, and* 7*,*

$$\limsup_{c\to 0} \frac{V_c(\rho,\pi)}{V_c^*(\rho)} \leq \frac{\mathbb{E}\{1/\widetilde{D}(\Theta)\}}{\mathbb{E}\{1/D(\Theta)\}},$$

where $D(\boldsymbol{\theta})$ *is defined in* (12) *and* $\widetilde{D}(\boldsymbol{\theta})$ *is define as*

$$\widetilde{D}(\boldsymbol{\theta}) = \max_{\boldsymbol{\lambda}\in\Delta} \inf_{\widetilde{\boldsymbol{\theta}}\in\widetilde{W}: r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})} \sum_{(i,j)} \lambda^{i,j} D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}}).$$

To obtain Theorem 4, we perform similar analysis as those for Theorem 2 and Theorem 3. Although $\widetilde{W}$ violates the separation property required by Assumption 4, similar proof strategy still applies under Assumption 7. Roughly, this is because the expected sample size $\mathbb{E}(T_l|\Theta = \boldsymbol{\theta})$ is now approximated by $|\log c|/\widetilde{D}(\boldsymbol{\theta})$ and $\widetilde{D}(\boldsymbol{\theta}) > 0$ for $\boldsymbol{\theta} \in W$. Note that to have $\widetilde{D}(\boldsymbol{\theta}) > 0$, we only need the support W to have the separation property and $\widetilde{W}$ can contain ties among the parameters.

4. Numerical Examples

4.1. Behavior of $D(\Theta)$. Our main results suggest that the oracle risk $V_c^*(\rho) \approx c|\log c|\mathbb{E}\{1/D(\Theta)\}$ when cost c is close to zero under the assumptions required by Theorems 2 and 3. The quantity $1/D(\boldsymbol{\theta})$ can be naturally viewed as a measure of difficulty for the rank aggregation

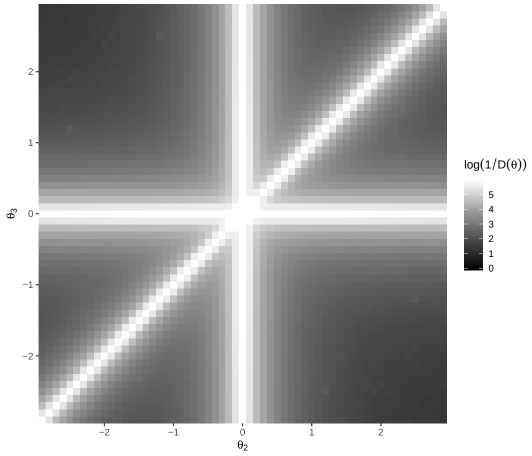

FIGURE 1. A level plot for the value of $\log(1/D(\boldsymbol{\theta}))$ as a function of θ_2 (x-axis) and θ_3 (y-axis), where $K=3$ and $W=\{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|\le 3,\theta_1=0,\text{and }\forall i\ne j\text{ such that }|\theta_i-\theta_j|\ge 0.1\}$.

task when the true parameter vector is $\boldsymbol{\theta}$. In what follows, we numerically investigate the behavior of $1/D(\boldsymbol{\theta})$.

We first show the value of $1/D(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$, when the number of items $K=3$. The support W of the prior distribution is chosen according to (16) that satisfies $W=\{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|\le 3,\theta_1=0,\text{and }\forall i\ne j\text{ such that }|\theta_i-\theta_j|\ge 0.1\}$. Figure 1 provides a level plot for the value of $\log(1/D(\boldsymbol{\theta}))$ as a function of θ_2 and θ_3. As we can see, the value of $1/D(\boldsymbol{\theta})$ becomes larger when the values of θ_1, θ_2, and θ_3 are closer to each other and becomes smaller when they are more distinct.

We further show how the value of $\mathbb{E}(1/D(\Theta))$ depends on the number of items K. For each choice of K, the support W is chosen as (16) with $\theta_1=0$, $\kappa=3$, and $\delta=0.1$. Figure 2 shows that the value of $\mathbb{E}(1/D(\Theta))$ is an increasing function of K, where $\mathbb{E}(1/D(\Theta))$ is approximated by 2000 Monte Carlo simulations. As we can see from Figure 2, $\mathbb{E}(1/D(\Theta))$ increases with K, suggesting that the rank aggregation task becomes more difficult on average, when the number of items becomes larger.

4.2. Effectiveness of adaptive selection. We now show the power of the proposed adaptive selection rule by comparing it with a random selection rule that randomly picks a pair of items in each iteration. For each selection rule, we stop data collection once a fixed number of samples are collected, where sample sizes 20, 40, and 60 are considered. In the adaptive selection method, we set $p=0.2$ for the ϵ-greedy strategy. The adaptive selection is implemented using Algorithm 2 with the number of iterations $m=200$, $\lambda^{i,j,0}=2/(K(K-1))$, and $c_0=1$. Note that the random selection method is essentially an off-line approach. The comparison is conducted under a model with $K=3$, $W=\{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|\le 3,\theta_1=0,\text{ and }\forall i\ne j\text{ such that }|\theta_i-\theta_j|\ge 0.1\}$, and the prior distribution ρ being a uniform distribution on W. For each selection rule and each sample size, 1000 independent

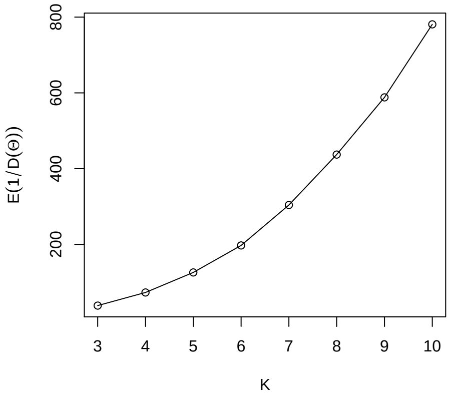

Figure 2. The value of $\mathbb{E}(1/D(\Theta))$ as a function of K, where $K = 3, 4, ..., 10$. For each choice of K, the support $W = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \le 3, \theta_1 = 0, \text{and } \forall i \ne j \text{ such that } |\theta_i - \theta_j| \ge 0.1\}$ and Θ follows a uniform distribution on W. Each $\mathbb{E}(1/D(\Theta))$ is computed by 2000 Monte Carlo simulations.

	Kendall's tau			0-1 Loss		
Sample size	20	40	60	20	40	60
Adaptive selection	0.217	0.115	0.075	0.195	0.113	0.074
Random selection	0.226	0.137	0.114	0.210	0.137	0.111

Table 1. Comparison between adaptive selection and random selection rules, under a fixed-length stopping criterion. Each cell gives the averaged Kendall's tau distance/0-1 loss for global ranking based on 1000 independent simulations.

simulations are conducted. Two performance metrics are considered, including the Kendall's tau distance (3) and the 0-1 loss for the recovery of global ranking that indicates whether or not the global ranking of $\boldsymbol{\theta}$ is completely recovered.

The results are given in Table 1 on the averaged Kendall's tau distance and the averaged 0-1 loss for global ranking. As we can see, for each sample size, both the average Kendall's tau distance and the average 0-1 loss for global ranking are smaller when applying the adaptive selection rule. The advantage of adaptive selection over random selection becomes more substantial as the sample size increases.

Under the current simulation setting, collecting one additional sample takes about 6 seconds[1], which is mainly due to solving optimization problem (15) in Algorithm 1. Note that the complexity for solving (15) depends on the number of disconnected regions that the support W has, which grows exponentially with K. Therefore, for large values of K, it is suggested to simplify the computation by using the misspecified support $\widetilde{W}$ in (20) which can be written as the union of $O(K^2)$ half-planes.

[1] The computation time is evaluated based on our implementation of the proposed method in R version 3.6.1 on a standard desktop PC with Intel(R) Core(TM) i5-5300 @2.3GHZ.

4.3. Effectiveness of adaptive stopping. We further assess the effectiveness of the two stopping rules. The same model as above is used, i.e., $K = 3$, $W = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq 3, \theta_1 = 0,$ and $\forall i \neq j$ such that $|\theta_i - \theta_j| \geq 0.1\}$, and the prior distribution ρ being a uniform distribution on W. For the proposed adaptive stopping rules, we set $h(c) = |\log c|(1 + |\log c|^{-0.5})$, where $\log c = -0.25, -0.5, -0.75, -1, -1.25,$ and -1.5 are considered. The proposed adaptive selection rule is used, with $p = 0.2 \times |\log c|^{-\frac{1}{4}}$. For each stopping rule and each value of c, 1000 independent simulations are conducted, for which the averaged sample size, the Kendall's tau distance, and the Bayes risk (5) are recorded, as shown in Tables 2 and 3.

We then compare these adaptive stopping rules with the fixed-length stopping rule. More precisely, for each value of c and each adaptive stopping rule, we consider a policy with the same adaptive selection rule and the sample size fixed to be the corresponding averaged sample size. The averaged Kendall's tau distance is also obtained based on 1000 independent simulations and is reported in Tables 2 and 3.

Comparing each adaptive stopping rule with the corresponding fixed-length stopping rule, we see that the adaptive stopping rule gives substantially smaller averaged Kendall's tau distances for all choices of c. It suggests that the adaptive stopping rules lead to more accurate ranking aggregation results than the non-adaptive stopping rule.

Comparing the results in Tables 2 and 3, it seems that stopping rule T_1 has slightly better performance than T_2 in terms of Kendall's tau distance when the value of c is large. For example, the averaged Kendall's tau distance for T_1 is 0.107 when the averaged sample size is 31, while that for T_2 is 0.112 when the averaged sample size is 35. Similarly, T_1 achieves an averaged Kendall's tau distance 0.057 when the averaged sample size is 46, while T_2 achieves the same value with an averaged sample size 50. However, as c decays (e.g., when $\log(c) = -1.25, -1.5$), the two procedures have similar performance, in terms of the averaged sample size and Kendall's tau distance. Regarding Bayes risks, we see that for each value of c, the Bayes risks of T_2 tends to be smaller than those of T_1. This is because, sampling cost is the dominant term in the Bayes risk. As T_2 tends to stop slightly earlier than T_1, its Bayes risks tend to be smaller. The difference in the corresponding Bayes risks becomes smaller when c decays. When $\log(c) = -1.25, -1.5$, the Bayes risks of the two methods are quite close to each other. It is worth pointing out that the difference in the finite sample performance when c is relatively large may depend on the choice of $h(c)$, and the two stopping times are asymptotically equivalent when c goes to zero.

5. Concluding Remarks In this paper, we consider the sequential design of rank aggregation with adaptive pairwise comparison. This problem is not only of practical importance due to its wide applications in fields such as psychology, politics, marketing, and sports, but also of theoretical

	Kendall's tau						Bayes risk						
Sample size	31	46	60	76	91	110	log(c)	-0.25	-0.5	-0.75	-1	-1.25	-1.5
T_1	0.107	0.057	0.039	0.019	0.017	0.014	T_1	23.9	27.8	28.5	28.3	26.2	24.5
Fixed length	0.207	0.133	0.100	0.069	0.059	0.052							

Table 2. Comparison between the proposed stopping rule T_1 and a fixed-length stopping rule, with the same adaptive selection rule. For both methods, the averaged Kendall's tau distances are given, each of which is computed based on 1000 independent simulations. For stopping rule T_1, the Bayes risks are also given as a linear combination of Kendall's tau distance and sampling cost.

	Kendall's tau							Bayes risk					
Sample size	19	35	50	64	88	105	log(c)	-0.25	-0.5	-0.75	-1	-1.25	-1.5
T_2	0.190	0.112	0.057	0.029	0.020	0.014	T_2	15.3	21.3	23.8	23.7	25.3	23.5
Fixed length	0.207	0.133	0.100	0.069	0.059	0.052							

Table 3. Comparison between the proposed stopping rule T_2 and a fixed-length stopping rule, with the same adaptive selection rule. For both methods, the averaged Kendall's tau distances are given, each of which is computed based on 1000 independent simulations. For stopping rule T_2, the Bayes risks are also given as a linear combination of Kendall's tau distance and sampling cost.

significance in sequential analysis. Due to the more complex structure of the ranking problem than hypothesis testing problems, no existing sequential analysis framework is suitable. We formulate the problem under a Bayesian decision framework and develop asymptotically optimal policies. Comparing to the existing Bayesian sequential hypothesis testing problems, the problem solved in this paper is technically more challenging due to the more structured risk function. Novel technical tools are developed to solve this problem, which are of separate theoretical interest in solving complex sequential design problems.

The current work may be extended in several directions. First, an even larger class of comparison models may be considered. The models considered in the current paper all assume the judges being homogeneous, i.e., the comparison outcome does not depend on who the judge is. It is of interest to consider the heterogeneity of the judges by incorporating judge-specific random effects into the comparison models and develop corresponding sequential designs. Second, different risk structures will be incorporated into the sequential ranking designs to account for practical needs in different applications. For example, we will consider other metrics for assessing the ranking accuracy (e.g. based on the accuracy of identifying the set of top K items) and non-uniform costs for different judges.

The results for pairwise comparison problem can be extended to the case for multiple choices by extending the BTL model in (2) to the multinomial logit model [56]. More specifically, given an L-tuple at time n: $a_n = (a_{n,1}, \dots, a_{n,L})$, the annotator chooses $X_n \in \{1, \dots, L\}$ following the distribution $\mathbb{P}(X_n = k) = \frac{\exp(\theta_{a_{n,k}})}{\sum_{k'=1}^{L} \exp(\theta_{a_{n,k'}})}$. However, additional challenges arise from solving the corresponding optimization problem in (13) that incurs higher complexity due to exploring more combinations of choices. For example, if there are K items and L choices presented to the annotator each time, we need to solve an optimization problem involving $\binom{K}{L}$ combinations. It is worth further investigation on how to reduce the computational burden while keeping a certain optimality.

6. Proof of Theorems

In this section, we present the proofs of Theorem 1-3. The proof for lemmas are delayed in the supplementary material. Throughout the proof, we will use the constants $\delta_\rho = \inf_{\boldsymbol{\theta}\in W}\rho(\boldsymbol{\theta}) > 0$ and $\sup_{\boldsymbol{\theta}\in W, x, a\in\mathcal{A}} |\nabla f^a_{\boldsymbol{\theta}}(x)| \le \kappa_0$. According to Assumptions 5 and 3, these two constants are finite.

6.1. Proof for Theorem 1

Let $\varepsilon = c|\log c|^2$. For an arbitrary policy $\pi = (A, T, R)$ and a prior probability density function ρ, there are two possibilities: either $\mathbb{E}L_K(R) \ge \varepsilon$ or $\mathbb{E}L_K(R) < \varepsilon$. For the first case, we can see $V(\rho,\pi) \ge \varepsilon \ge (1+o(1))c\mathbb{E}t_c(\Theta)$. For the second case, we have

$$V(\pi,\rho) = \mathbb{E}L_K(R) + c\mathbb{E}T \ge c\mathbb{E}T.$$

Therefore, to prove the theorem it is sufficient to show that

$$\liminf_{c\to 0} \frac{c\mathbb{E}T}{c\mathbb{E}t_c(\Theta)} \ge 1$$

or, equivalently, for each $\delta > 0$ there exists a positive constant $c_0 > 0$ such that for $c < c_0$,

$$\mathbb{E}T \ge (1-\delta)\mathbb{E}t_c(\Theta).$$

Let $t_{c,\delta}(\boldsymbol{\theta}) = (1 - 2\delta/3)t_c(\boldsymbol{\theta})$ for each $\delta > 0$. Then we arrive at a lower bound

$$\begin{aligned}\mathbb{E}T &\ge \mathbb{E}[TI(T > t_{c,\delta}(\Theta))]\\ &\ge \int \rho(\boldsymbol{\theta}) t_{c,\delta}(\boldsymbol{\theta})\mathbb{P}_{\boldsymbol{\theta}}(T > t_{c,\delta}(\boldsymbol{\theta}))d\boldsymbol{\theta}\\ &= \mathbb{E}t_{c,\delta}(\Theta) - \int \rho(\boldsymbol{\theta}) t_{c,\delta}(\boldsymbol{\theta})\mathbb{P}_{\boldsymbol{\theta}}(T \le t_{c,\delta}(\boldsymbol{\theta}))d\boldsymbol{\theta}\\ &\ge \mathbb{E}t_{c,\delta}(\Theta) - t_{\max,\delta}\mathbb{P}(T \le t_{c,\delta}(\Theta)),\end{aligned}$$

where we define $t_{\max,\delta} = \max_{\boldsymbol{\theta}\in W} t_{c,\delta}(\boldsymbol{\theta})$ and recall that $\mathbb{P}_{\boldsymbol{\theta}}$ represents for the conditional probability $\mathbb{P}(\cdot|\Theta = \boldsymbol{\theta})$. According to Assumption 4 we have $t_{\max,\delta} = O(|\log c|) = O(\mathbb{E}t_c(\Theta))$. Therefore, it is sufficient to show

$$\mathbb{P}(T \le t_{c,\delta}(\Theta)) = o(1).$$

We proceed to an upper bound for $\mathbb{P}(T \le t_{c,\delta}(\Theta))$. We abuse the notation a little and write $U_r = \{\boldsymbol{\theta} : r(\boldsymbol{\theta}) = r\}$, the set of parameters that gives the rank r. Then, we have

$$\begin{aligned}\mathbb{P}(T \le t_{c,\delta}(\Theta)) &= \sum_{r\in P_K} \mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r)\\ &= O(1) \times \max_{r\in P_K} \mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r).\end{aligned} \tag{21}$$

We proceed to an upper bound for $\mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r)$ for each $r \in P_K$. Define an event

$$B_r = \Big\{ \frac{\mathbb{P}(\Theta \in U_r | \mathcal{F}_T)}{\max_{(i,j): W_{i,j} \cap U_r = \emptyset} \mathbb{P}(\Theta \in W_{i,j} | \mathcal{F}_T)} > \frac{c^{\frac{\delta}{10}}}{\varepsilon} \Big\}, \tag{22}$$

where $\mathcal{F}_n = \sigma(X_1, ..., X_n, a_1, ..., a_n)$ denotes the σ-algebra generated by $X_1, ..., X_n$ and $a_1, .., a_n$. We split the probability

$$\begin{aligned} &\mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r) \\ =&\mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r, B_r) + \mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r, B_r^c), \end{aligned}$$

which can be bounded from above by

$$\mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r) \le \mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r, B_r) + \mathbb{P}(\Theta \in U_r, B_r^c). \tag{23}$$

We establish upper bounds for the two terms on the right-hand side of the above inequality separately. The next lemma, whose proof is presented in the supplementary material, provides an upper bound for the second term.

Lemma 2 *For all $r \in P_K$, if $\mathbb{E} L_K(R) \le \varepsilon$ then*

$$\mathbb{P}(\Theta \in U_r, B_r^c) \le (1 + \frac{c^{\frac{\delta}{10}}}{\varepsilon})\varepsilon.$$

We proceed to the first term $\mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r, B_r)$ on the right-hand side of (23). Then,

$$\mathbb{P}(T \le t_{c,\delta}(\Theta), \Theta \in U_r, B_r) = \int_{U_r} \mathbb{P}_{\boldsymbol{\theta}}(T \le t_{c,\delta}(\boldsymbol{\theta}), B_r) \rho(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{24}$$

Recall the definition of the event B_r in (22), we have

$$B_r \cap \{T \le t_{c,\delta}(\boldsymbol{\theta})\} \subset \Big\{ \max_{1 \le t \le t_{c,\delta}(\boldsymbol{\theta})} \frac{\mathbb{P}(\Theta \in U_r | \mathcal{F}_t)}{\max_{W_{i,j} \cap U_r = \emptyset} \mathbb{P}(\Theta \in W_{i,j} | \mathcal{F}_t)} > \frac{c^{\frac{\delta}{10}}}{\varepsilon} \Big\}.$$

Consequently,

$$\mathbb{P}_{\boldsymbol{\theta}} \left(T \le t_{c,\delta}(\boldsymbol{\theta}), B_r\right) \le \mathbb{P}_{\boldsymbol{\theta}} \Big(\max_{1 \le t \le t_{c,\delta}(\boldsymbol{\theta})} \frac{\mathbb{P}(\Theta \in U_r | \mathcal{F}_t)}{\max_{W_{i,j} \cap U_r = \emptyset} \mathbb{P}(\Theta \in W_{i,j} | \mathcal{F}_t)} > \frac{c^{\frac{\delta}{10}}}{\varepsilon} \Big). \tag{25}$$

We proceed to an upper bound for the above display. For each $\boldsymbol{\theta}$, we define a random sequence $\{\boldsymbol{\theta}_t^* : 1 \le t \le t_{c,\delta}(\boldsymbol{\theta})\}$ as follows.

$$\boldsymbol{\theta}_t^* = \mathop{\arg\min}_{\widetilde{\boldsymbol{\theta}} \in W : r(\widetilde{\boldsymbol{\theta}}) \ne r(\boldsymbol{\theta})} \sum_{n=1}^{t} \sum_{i,j} \lambda_n^{i,j} D^{i,j}(\boldsymbol{\theta} \| \widetilde{\boldsymbol{\theta}}).$$

Intuitively, $\boldsymbol{\theta}_t^*$ is the score parameter that is most difficult to distinguish from $\boldsymbol{\theta}$ at time t among those that have different rank with $\boldsymbol{\theta}$, given that item selection rules $\lambda_1, ..., \lambda_n$ have been adopted. We further choose the index process (i_t^*, j_t^*) be such that $\boldsymbol{\theta}_t^* \in W_{i_t^*, j_t^*}$ but $\boldsymbol{\theta} \notin W_{i_t^*, j_t^*}$. If there are multiple (i,j)'s satisfying this, then we choose (i_t^*, j_t^*) arbitrarily from them. From the definition, we know $\boldsymbol{\theta}_t^*$ and (i_t^*, j_t^*) are adapted to $\sigma(\lambda_1,...,\lambda_t)$, and thus are adapted to $\mathcal{F}_{t-1}$. We use the next lemma to transform the probability in (25) to a probability based on a martingale parameterized by $\boldsymbol{\theta}$.

Lemma 3 *For each $\boldsymbol{\theta}' \in U_r$, define a martingale with respect to the filtration $\{\mathcal{F}_n : n \geq 1\}$ and probability measure $\mathbb{P}_{\boldsymbol{\theta}}$ as follows,*

$$M_t(\boldsymbol{\theta}') = l_t^{\vec{a}}(\boldsymbol{\theta}') - l_t^{\vec{a}}(\boldsymbol{\theta}_t^*) - \sum_{n=1}^{t}\sum_{(i,j)} \lambda_n^{i,j} D^{i,j}(\boldsymbol{\theta}\|\boldsymbol{\theta}_t^*) + \sum_{n=1}^{t}\sum_{(i,j)} \lambda_n^{i,j} D^{i,j}(\boldsymbol{\theta}\|\boldsymbol{\theta}'),$$

where $l_t^{\vec{a}}(\boldsymbol{\theta}) = \log \prod_{i=1}^t f_{\boldsymbol{\theta}}^{a_i}(X_i)$. Then there exists a positive constant $c_0 > 0$ such that for $0 < c < c_0$,

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\Big(\max_{1\leq t\leq t_{c,\delta}(\boldsymbol{\theta})} \frac{\mathbb{P}(\Theta\in U_r|\mathcal{F}_t)}{\max_{W_{i,j}\cap U_r=\emptyset}\mathbb{P}(\Theta\in W_{i,j}|\mathcal{F}_t)} > \frac{c^{\frac{\delta}{10}}}{\varepsilon}\Big)\\
\leq &\mathbb{P}_{\boldsymbol{\theta}}\Big(\max_{1\leq t\leq t_{c,\delta}(\boldsymbol{\theta}),\boldsymbol{\theta}'\in U_r} M_t(\boldsymbol{\theta}') \geq \frac{\delta}{2}|\log c|\Big). \qquad (26)
\end{aligned}$$

According to the above lemma, to find an upper bound for (25), it is sufficient to find an upper bound for the right-hand side of (26), which is the probability that a stochastic process indexed by $\boldsymbol{\theta}'$ and t goes above a certain level. In this paper, we will use the following two lemmas repeatedly to handle this type of level crossing probabilities. The first one is the Azuma-Hoeffding inequality proved by [3] and [28].

Lemma 4 (Azuma-Hoeffding inequality) *Let M_n be a martingale with respect to the filtration $\{\mathcal{F}_n : n = 1, 2, ..\}$. Let $X_n = M_n - M_{n-1}$. Assume that X_n is bounded and $X_n \in [a_n, b_n]$ where a_n and b_n are deterministic constants. Then, for each $t > 0$ we have*

$$\mathbb{P}(\max_{1\leq m\leq n} M_m \geq t) \leq \exp\Big(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big).$$

The next lemma is the key lemma that allows us to derive level crossing probability by aggregating marginal tail bounds of a random field. Its proof is given in the supplementary material.

Lemma 5 *Let $\{\zeta(\boldsymbol{\theta}) : \boldsymbol{\theta} \in W\}$ be a random field over a compact set $U \subset \mathbb{R}^K$ that satisfies Assumption 2. Let $\beta(\boldsymbol{\theta}, b)$ be defined as follows*

$$\beta(\boldsymbol{\theta}, b) = \mathbb{P}(\zeta(\boldsymbol{\theta}) \geq b),$$

where $\mathbb{P}$ is a probability measure and we assume that $\zeta(\cdot)$ has continuous sample path almost surely under $\mathbb{P}$. Assume that $\zeta(\cdot)$ has a Lipschitz continuous sample path in the sense that there exists a constant κ_L such that for all $\theta, \theta' \in W$

$$|\zeta(\boldsymbol{\theta}) - \zeta(\boldsymbol{\theta}')| \leq \kappa_L \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \text{ almost surely under } \mathbb{P}.$$

Then, we have that for all positive γ

$$\mathbb{P}\Big(\max_{\theta\in W}\zeta(\boldsymbol{\theta})\geq b\Big)\leq\int_W\beta(\boldsymbol{\theta},b-\gamma)d\boldsymbol{\theta}\times\frac{\kappa_L^{K-1}}{\gamma^{K-1}\delta_b},$$

where δ_b is the constant defined in Assumption 2.

Set $n:=t_{c,\delta}(\boldsymbol{\theta})$, $t:=\frac{\delta}{2}|\log c|-1$, $M_n:=M_n(\boldsymbol{\theta}')$, and $a_n=-b_n:=2\max_{x,a\in\mathcal{A},\theta\in W}|\log f^a_{\theta,x}(x)|$ in Lemma 4, we have for each $\boldsymbol{\theta}'$

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\max_{1\leq n\leq t_{c,\delta}(\boldsymbol{\theta})}M_n(\boldsymbol{\theta}')\geq\frac{\delta}{2}|\log c|-1\right)\leq\exp\Big(-\frac{2(\frac{\delta}{2}|\log c|-1)^2}{t_{c,\delta}(\boldsymbol{\theta})a_1^2}\Big).$$

According to Assumption 1 and 3, we have $a_1<\infty$, and consequently,

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\max_{1\leq n\leq t_{c,\delta}(\boldsymbol{\theta})}M_n(\boldsymbol{\theta}')\geq\frac{\delta}{2}|\log c|-1\right)\leq\exp\Big(-\Omega(\delta^2|\log c|)\Big).\tag{27}$$

Note that for $\boldsymbol{\theta}',\widetilde{\boldsymbol{\theta}}\in U_r$,

$$\begin{aligned}&\max_{1\leq n\leq t_{c,\delta}(\boldsymbol{\theta})}M_n(\boldsymbol{\theta}')-\max_{1\leq n\leq t_{c,\delta}(\boldsymbol{\theta})}M_n(\widetilde{\boldsymbol{\theta}})\\ \leq&\max_{1\leq n\leq t_{c,\delta}(\boldsymbol{\theta})}|M_n(\boldsymbol{\theta}')-M_n(\widetilde{\boldsymbol{\theta}})|\\ \leq& t_{c,\delta}(\boldsymbol{\theta})\kappa_0\|\boldsymbol{\theta}'-\widetilde{\boldsymbol{\theta}}\|,\end{aligned}$$

where $\kappa_0=4\sup_{a\in\mathcal{A},\boldsymbol{\theta}'\in W,x}|\nabla\log f^a_{\boldsymbol{\theta}}(x)|<\infty$ denotes the Lipschitz constant of $M_1(\boldsymbol{\theta}')$. Therefore, $M_n(\boldsymbol{\theta}')$ is a Lipschitz continuous random field in $\boldsymbol{\theta}'$. The above display and (27), together with Lemma 5, give

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\max_{1\leq n\leq t_{c,\delta}(\boldsymbol{\theta}),\boldsymbol{\theta}'\in U_r}M_n(\boldsymbol{\theta}')\geq\frac{\delta}{2}|\log c|\right)\\ \leq&\exp\Big(-\Omega(\delta^2|\log c|)\Big)m(U_r)\frac{t_{c,\delta}(\boldsymbol{\theta})^{K-1}\kappa_0^{K-1}}{\delta_b}\\ =&\exp\Big(-\Omega(\delta^2|\log c|)\Big)\times O(|\log c|^{K-1}),\end{aligned}$$

where we recall that $m(\cdot)$ denotes the Lebesgue measure. The above inequality and (24), (25),(26) give

$$\mathbb{P}(T\leq t_{c,\delta}(\Theta),\Theta\in U_r,B_r)\leq\exp\Big(-\Omega(\delta^2|\log c|)\Big)\times O(|\log c|^{K-1}).$$

Combine this with Lemma 2 and (23) we have

$$\mathbb{P}(T\leq t_{c,\delta}(\Theta),\Theta\in U_r)\leq(1+\frac{c^{\frac{\delta}{10}}}{\varepsilon})\varepsilon+\exp\Big(-\Omega(\delta^2|\log c|)\Big)\times O(|\log c|^{K-1}).$$

Combine the above display with (21), we have

$$\mathbb{P}(T\leq t_{c,\delta}(\Theta))\leq O(1)\times\Big\{(1+\frac{c^{\frac{\delta}{10}}}{\varepsilon})\varepsilon+\exp\Big(-\Omega(\delta^2|\log c|)\Big)\times O(|\log c|^{K-1})\Big\}.$$

Therefore, $\mathbb{P}(T\leq t_{c,\delta}(\Theta))=o(1)$ as $c\to 0$. This completes the proof.

6.2. Proof of Theorem 2 We start with the stopping time T_2. With the decision rule D defined in (11), the expected Kendall's tau at the stopping time T_2 is

$$\begin{aligned}\mathbb{E}L_K(R) =&\mathbb{E}\sum_{(i,j)} I(\Theta_i < \Theta_j) R_{i,j}\\ =&\int_W \sum_{(i,j):\boldsymbol{\theta}\notin W_{j,i}} \mathbb{P}_{\boldsymbol{\theta}}(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) > \sup_{\boldsymbol{\theta}'\in W_{i,j}} l_{T_2}(\boldsymbol{\theta}'))\rho(\boldsymbol{\theta})d\boldsymbol{\theta}\\ =&\int_W \sum_{\boldsymbol{\theta}\notin W_{j,i}} \mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}'\in W_{i,j}} l_{T_2}(\boldsymbol{\theta}') > h(c)\Big)\rho(\boldsymbol{\theta})d\boldsymbol{\theta},\end{aligned} \tag{28}$$

where we write $l_t(\boldsymbol{\theta}) = \sum_{n=1}^t \log f_{\boldsymbol{\theta}}^{a_n}(X_n)$ as the log-likelihood function. (28) is bounded from above by

$$\begin{aligned}\mathbb{E}L_K(R) \le& \sup_{\boldsymbol{\theta}\in W} \rho(\boldsymbol{\theta}) \times m(W) \times \frac{K(K-1)}{2}\\ &\times \sup_{\boldsymbol{\theta}\in W} \max_{(i,j):\boldsymbol{\theta}\notin W_{j,i}} \mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) - l_{T_2}(\boldsymbol{\theta}) > h(c)\Big).\end{aligned} \tag{29}$$

To obtain the above inequality, we used the fact that $\sup_{\boldsymbol{\theta}'\in W_{i,j}} l_{T_2}(\boldsymbol{\theta}') \ge l_{T_2}(\boldsymbol{\theta})$ for (i,j) such that $\boldsymbol{\theta}\notin W_{j,i}$ and $\sup_{\boldsymbol{\theta}\in W}\rho(\boldsymbol{\theta}) < \infty$ according to Assumption 5. We split the probability

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) - l_{T_2}(\boldsymbol{\theta}) > h(c)\right)\\ \le&\mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) - l_{T_2}(\boldsymbol{\theta}) > h(c) \text{ and } T_2 \le \tau\right) + \mathbb{P}_{\boldsymbol{\theta}}(T_2\ge\tau).\end{aligned} \tag{30}$$

We clarify that $\boldsymbol{\theta}'$, $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ are deterministic vectors here. The second term on the right-hand side of the above display is controlled by the next lemma.

Lemma 6 *If* $\tau = \Omega(|\log c|^3)$ *then,*

$$\mathbb{P}_{\boldsymbol{\theta}}(T_i \ge \tau) \le c^2 \qquad (i = 1,2).$$

We proceed to an upper bound of the first term on the right-hand side of (30). Define a stopping time $T_2\wedge\tau = \min(T_2,\tau)$, then we have

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) - l_{T_2}(\boldsymbol{\theta}) > h(c) \text{ and } T_2 \le \tau\right)\\ \le&\mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} l_{T_2\wedge\tau}(\widetilde{\boldsymbol{\theta}}) - l_{T_2\wedge\tau}(\boldsymbol{\theta}) > h(c)\right).\end{aligned}$$

Now we consider the random field $\eta(\widetilde{\boldsymbol{\theta}}) = l_{T_2\wedge\tau}(\widetilde{\boldsymbol{\theta}}) - l_{T_2\wedge\tau}(\boldsymbol{\theta})$ for $\widetilde{\boldsymbol{\theta}}\in W_{ji}$. We proceed to an upper bound for $\mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} \eta(\widetilde{\boldsymbol{\theta}}) > h(c)\right)$ through Lemma 5. We first note that $\eta(\widetilde{\boldsymbol{\theta}})$ is a Lipschitz continuous function,

$$|\eta(\widetilde{\boldsymbol{\theta}}) - \eta(\widetilde{\boldsymbol{\theta}}')| \le |l_{T\wedge\tau}(\widetilde{\boldsymbol{\theta}}) - l_{T\wedge\tau}(\widetilde{\boldsymbol{\theta}}')| \le \tau\kappa_0\|\widetilde{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}'\|. \tag{31}$$

We further obtain the marginal tail probability of $\eta(\widetilde{\boldsymbol{\theta}})$ through the next lemma.

Lemma 7 *For all $\widetilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$, and all constant $A > 0$, we have*

$$\mathbb{P}_{\boldsymbol{\theta}}\left(l_{T\wedge\tau}(\widetilde{\boldsymbol{\theta}}) - l_{T\wedge\tau}(\boldsymbol{\theta}) \geq A\right) \leq e^{-A}$$

We take $A = h(c) - 1$ in the above lemma and obtain

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\eta(\widetilde{\boldsymbol{\theta}}) \geq h(c) - 1\right) \leq e^{-h(c)+1}$$

Combining the above display with (31) and Lemma 5, we arrive at

$$\mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} \eta(\widetilde{\boldsymbol{\theta}}) > h(c)\right) \leq O(\tau^{K-1}e^{-h(c)}). \tag{32}$$

We combine (32),(29) and Lemma 6 and arrive at

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} l_{T_2}(\widetilde{\boldsymbol{\theta}}) - l_{T_2}(\boldsymbol{\theta}) > h(c)\Big)\\
\leq& O(c^2) + O(e^{-|\log c| - |\log c|^{1-\alpha} + (K-1)\log\tau})\\
=& O(c^2) + O(ce^{-|\log c|^{1-\alpha} + 3(K-1)\log|\log c|})\\
=& o(c).
\end{aligned}$$

This completes our analysis for T_2. We proceed to the analysis of the policy π_1 and the stopping time T_1. According to the definition of T_1 in (10), we can see that upon stopping,

$$\begin{aligned}
&\max_{(i,j):1\leq i<j\leq K} \exp\left[\min\left\{\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}} l_{T_1}(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_{T_1}(\boldsymbol{\theta}), \sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_1}(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_{T_1}(\boldsymbol{\theta})\right\}\right.\\
\leq& \sum_{(i,j):1\leq i<j\leq K} \exp\left[\min\left\{\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}} l_{T_1}(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_{T_1}(\boldsymbol{\theta}), \sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_{T_1}(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_{T_1}(\boldsymbol{\theta})\right\}\right]\\
\leq& e^{-h(c)}.
\end{aligned}$$

Taking logarithm and rearranging terms in the above display, we have

$$\min_{1\leq i<j\leq K}\left[\sup_{\boldsymbol{\theta}\in W} l_n(\boldsymbol{\theta}) - \min\left\{\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}} l_n(\widetilde{\boldsymbol{\theta}}), \sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_n(\widetilde{\boldsymbol{\theta}})\right\}\right] \geq h(c). \tag{33}$$

With (33) we can follow similar derivations as those for (29) and arrive at

$$\begin{aligned}
&\mathbb{E}L_K(\bar{D}_{T_1})\\
\leq& \sup_{\boldsymbol{\theta}'\in W} \rho(\boldsymbol{\theta})m(W)\\
&\times \frac{K(K-1)}{2} \sup_{\boldsymbol{\theta}\in W} \max_{(i,j):\boldsymbol{\theta}\notin W_{j,i}} \mathbb{P}_{\boldsymbol{\theta}}\Big(\sup_{\widetilde{\boldsymbol{\theta}}\in W_{ji}} l_{T_1}(\widetilde{\boldsymbol{\theta}}) - l_{T_1}(\boldsymbol{\theta}) > h(c)\Big).
\end{aligned}$$

The rest of the proof is similar as that for the stopping time T_2. We omit the details.

6.3. Proof of Theorem 3 Let δ be an arbitrary positive number, we can find an upper bound for the expectation of a stopping time T as follows.

$$\begin{aligned}
&\mathbb{E}T\\
=&\sum_{m=0}^{\infty}\mathbb{E}\Big[TI\big(m(1+\delta)t_c(\Theta)\le T<(m+1)(1+\delta)t_c(\Theta)\big)\Big]\\
\le&(1+\delta)\mathbb{E}t_c(\Theta)+\sum_{m=1}^{\infty}\mathbb{E}\Big[TI\big(m(1+\delta)t_c(\Theta)\le T<(m+1)(1+\delta)t_c(\Theta)\big)\Big]\\
\le&(1+\delta)\mathbb{E}t_c(\Theta)\\
&+(1+\delta)\max_{\boldsymbol{\theta}\in W}t_c(\boldsymbol{\theta})\sum_{m=1}^{\infty}(m+1)\mathbb{P}\left(m(1+\delta)t_c(\Theta)\le T<(m+1)(1+\delta)t_c(\Theta)\right)\\
\le&(1+\delta)\mathbb{E}t_c(\Theta)\\
&+(1+\delta)\max_{\theta\in W}t_c(\boldsymbol{\theta})\sum_{m=1}^{\infty}(m+1)\max_{\boldsymbol{\theta}\in W}\mathbb{P}_{\boldsymbol{\theta}}\left(m(1+\delta)t_c(\boldsymbol{\theta})\le T<(m+1)(1+\delta)t_c(\boldsymbol{\theta})\right)
\end{aligned}\tag{34}$$

We proceed to an upper bound for the probability in the above sum for $T=T_i$ $(i=1,2)$. We start with $T=T_2$. We split the probability for $m\ge 1$,

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(m(1+\delta)t_c(\boldsymbol{\theta})\le T_2<(m+1)(1+\delta)t_c(\boldsymbol{\theta})\right)\\
\le&\mathbb{P}_{\boldsymbol{\theta}}\Big(m(1+\delta)t_c(\boldsymbol{\theta})\le T_2<(m+1)(1+\delta)t_c(\boldsymbol{\theta}),\\
&\qquad\max_{m(1+\delta)\delta_2t_c(\boldsymbol{\theta})\le t\le m(1+\delta)t_c(\boldsymbol{\theta})}\|\widehat{\boldsymbol{\theta}}^{(t)}-\boldsymbol{\theta}\|\le|\log c|^{-\delta_1}\Big)\\
&+\mathbb{P}_{\boldsymbol{\theta}}\Big(\max_{m(1+\delta)\delta_2t_c(\boldsymbol{\theta})\le t\le m(1+\delta)t_c(\boldsymbol{\theta})}\|\widehat{\boldsymbol{\theta}}^{(t)}-\boldsymbol{\theta}\|\ge|\log c|^{-\delta_1}\Big),
\end{aligned}\tag{35}$$

where we choose $\delta_1=\frac{\delta_0}{8}$ and $\delta_2=|\log c|^{-\delta_0/2}$, and δ_0 is defined in the selection rule where we recall that $p\propto|\log c|^{-\frac{1}{2}+\delta_0}$. The second term on the above display is bounded from above according to Lemma 1, where we set $n:=m(1+\delta)\delta_2t_c(\boldsymbol{\theta})$, $m:=m(1+\delta)t_c(\boldsymbol{\theta})$, $\varepsilon_\lambda=\Omega(|\log c|^{-\frac{1}{2}+\delta_0})$ and $\delta_{m,n}=|\log c|^{-\delta_1}$, and arrive at

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\Big(\max_{m(1+\delta)\delta_2t_c(\boldsymbol{\theta})\le t\le m(1+\delta)t_c(\boldsymbol{\theta})}\|\widehat{\boldsymbol{\theta}}^{(t)}-\boldsymbol{\theta}\|\ge|\log c|^{-\delta_1}\Big)\\
\le&e^{-\Omega(m(1+\delta)\delta_2t_c(\boldsymbol{\theta})|\log c|^{-4\delta_1}|\log c|^{-1+2\delta_0})}\times O(m^{K-1}|\log c|^{K-1})\\
=&e^{-\Omega(m|\log c|^{2\delta_0-4\delta_1}\delta_2)}O(m^{K-1}|\log c|^{K-1})\\
=&e^{-\Omega(m|\log c|^{\delta_0})}O(m^{K-1}|\log c|^{K-1}).
\end{aligned}\tag{36}$$

We proceed to the first term on the right-hand side of (35). For $m\ge 1$, we can see that $T_2>m(1+\delta)t_c(\boldsymbol{\theta})$ implies that there exists (i,j) such that $|\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}}l_n(\widetilde{\boldsymbol{\theta}})-\sup_{\boldsymbol{\theta}'\in W_{j,i}}l_n(\boldsymbol{\theta}')|\le h(c)$ for $n=(1+\delta)mt_c(\boldsymbol{\theta})$. Without loss of generality, we assume that $\boldsymbol{\theta}\in W_{i,j}$, then $T_2>m(1+\delta)t_c(\boldsymbol{\theta})$

further implies $l_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}'\in W_{j,i}} l_n(\boldsymbol{\theta}') \le h(c)$. Therefore, an upper bound for the first term on the right-hand side of (35) is

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\Big(m(1+\delta)t_c(\boldsymbol{\theta}) \le T_2 \le (m+1)(1+\delta)t_c(\boldsymbol{\theta}), \max_{m(1+\delta)\delta_2 t_c(\boldsymbol{\theta})\le t\le m(1+\delta)t_c(\boldsymbol{\theta})} \|\widehat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\| \le |\log c|^{-\delta_1}\Big)\\
\le&\mathbb{P}_{\boldsymbol{\theta}}\Big(l_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}'\in W_{j,i}} l_n(\boldsymbol{\theta}') \le h(c), \max_{m(1+\delta)\delta_2 t_c(\boldsymbol{\theta})\le t\le m(1+\delta)t_c(\boldsymbol{\theta})} \|\widehat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\| \le |\log c|^{-\delta_1}\Big),
\end{aligned} \tag{37}$$

We present an upper bound for the above display in the next lemma.

Lemma 8 *If the strategy $\lambda^*(\widehat{\boldsymbol{\theta}}^{(t)})$ is adopted with probability $1-o(1)$ uniformly for $mt_c(\boldsymbol{\theta})(1+\delta)\delta_2 \le t \le m(1+\delta)t_c(\boldsymbol{\theta})$. Then*

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}\left(l_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}'\in W_{j,i}} l_n(\boldsymbol{\theta}') \le h(c), \max_{m(1+\delta)\delta_2 t_c(\boldsymbol{\theta})\le t\le m(1+\delta)t_c(\boldsymbol{\theta})} \|\widehat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\| \le |\log c|^{-\delta_1}\right)\\
\le& e^{-\Omega(m|\log c|)} \times O(|\log c|^{K-1} m^{K-1}),
\end{aligned}$$

where $n = (1+\delta)mt_c(\boldsymbol{\theta})$.

We combine the above lemma with (36) and (35), we arrive at

$$\mathbb{P}_{\boldsymbol{\theta}}\Big(m(1+\delta)t_c(\boldsymbol{\theta}) \le T_2 < (m+1)(1+\delta)t_c(\boldsymbol{\theta})\Big) \le \big(e^{-\Omega(m|\log c|)} + e^{-\Omega(m|\log c|^{\delta_0})}\big) \times O(m^{K-1}|\log c|^{K-1}).$$

This, together with (34) gives

$$\begin{aligned}
&\mathbb{E}T_2\\
\le&(1+\delta)\mathbb{E}t_c(\Theta)\\
&+O(|\log c|)\times\sum_{m=1}^{\infty}(m+1)\{(e^{-\Omega(m|\log c|)} + e^{-\Omega(m|\log c|^{\delta_0})}) \times O(m^{K-1}|\log c|^{K-1})\}]\\
\le&(1+\delta)\mathbb{E}t_c(\Theta) + o(|\log c|).
\end{aligned}$$

This completes our analysis for T_2. We proceed to the analysis of T_1. We can see that the event $T_1 > n$ implies that

$$\sum_{(i,j)} \exp\left[\min\left\{\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}} l_n(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_n(\boldsymbol{\theta}), \sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_n(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_n(\boldsymbol{\theta})\right\}\right] > e^{-h(c)},$$

which further implies that

$$K(K-1)\max_{(i,j)} \exp\left[\min\left\{\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}} l_n(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_n(\boldsymbol{\theta}), \sup_{\widetilde{\boldsymbol{\theta}}\in W_{j,i}} l_n(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W} l_n(\boldsymbol{\theta})\right\}\right] > e^{-h(c)}.$$

Simplifying the above display, we can see it is equivalent to that there exists (i,j) such that

$$|\sup_{\widetilde{\boldsymbol{\theta}}\in W_{i,j}} l_n(\widetilde{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\theta}\in W_{j,i}} l_n(\boldsymbol{\theta})| \le h(c) + \log K(K-1).$$

The analysis is similar for the stopping time T_1 to that of T_2 by replacing $h(c)$ by $h(c) + \log K(K-1)$ in the derivation following (37). We omit the details.

6.4. Proof of Theorem 4 First, to distinguish between the sequential method with and without model misspecification, we will use the notation '$^-$' over a method (e.g., the sequential ranking rule $\bar{\pi}_l=(\bar{A},\bar{T}_l,\bar{R})$, and the MLE $\bar{\boldsymbol{\theta}}^{(t)}$) to indicate that it is based on the algorithm with the misspecified support $\widetilde{W}$ of the prior distribution $\rho(\cdot)$. The proof of Theorem 4 follows similar arguments as those of Corollary 1. That is, we will show the following modified version of Theorem 2 and Theorem 3, whose proof is provided in the supplement.

Proposition 2. *Following the sequential ranking rules $\bar{\pi}_l=(\bar{A},\bar{T}_l,\bar{R})$ (for $l=1,2$), we have*

$$\mathbb{E}L_K(\{\bar{R}_{i,j}\})=O(c)$$

Proposition 3.

$$\limsup_{c\to 0}\frac{\mathbb{E}\bar{T}_l}{\mathbb{E}\widetilde{t}_c(\Theta)}\leq 1$$

where we define $\widetilde{t}_c(\boldsymbol{\theta})=\frac{|\log(c)|}{\widetilde{D}(\boldsymbol{\theta})}$ *and* $\widetilde{D}(\boldsymbol{\theta})=\max_{\boldsymbol{\lambda}\in\Delta}\min_{\widetilde{\boldsymbol{\theta}}\in\widetilde{W}:r(\widetilde{\boldsymbol{\theta}})\neq r(\boldsymbol{\theta})}\sum_{(i,j)}\lambda^{i,j}D^{i,j}(\boldsymbol{\theta}\|\widetilde{\boldsymbol{\theta}})$.

Combining Theorem 2 and 3 with the above Proposition 2 and 3, we arrive at

$$\limsup_{c\to 0}\frac{V_c(\rho,\pi)}{V_c^*(\rho)}\leq\lim_{c\to 0}\frac{O(c)+c\mathbb{E}\widetilde{t}_c(\Theta)}{O(c)+c\mathbb{E}t_c(\Theta)}=\lim_{c\to 0}\frac{O(c)+c|\log c|\mathbb{E}\{1/\widetilde{D}(\Theta)\}}{O(c)+c|\log c|\mathbb{E}\{1/D(\Theta)\}}=\frac{\mathbb{E}\{1/\widetilde{D}(\Theta)\}}{\mathbb{E}\{1/D(\Theta)\}}.$$

Acknowledgments. The authors would like to thank the anonymous associate editor, and two anonymous referees for many useful suggestions and feedback that greatly improve the paper. The authors would also like to thank Dr. Jingchen Liu and Dr. Zhiliang Ying for the helpful discussions. Xi Chen acknowledges the support from National Science Foundation via IIS-1845444. Yunxiao Chen acknowledges the support from the National Academy of Education/Spencer Postdoctoral Fellowship. Xiaoou Li acknowledges the support from NSF under the grant DMS-1712657.

References

[1] Adler RJ, Blanchet JH, Liu J (2012) Efficient monte carlo for high excursions of Gaussian random fields. *The Annals of Applied Probability* 22(3):1167–1214.

[2] Albert AE (1961) The sequential design of experiments for infinitely many states of nature. *The Annals of Mathematical Statistics* 32(3):774–799.

[3] Azuma K (1967) Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series* 19(3):357–367.

[4] Ballinger TP, Wilcox NT (1997) Decisions, error and heterogeneity. *The Economic Journal* 107(443):1090–1105.

[5] Bartroff J, Finkelman M, Lai TL (2008) Modern sequential analysis and its applications to computerized adaptive testing. *Psychometrika* 73:473–486.

[6] Bartroff J, Lai TL (2008) Efficient adaptive designs with mid-course sample size adjustment in clinical trials. *Statistics in Medicine* 27:1593–1611.

[7] Bartroff J, Lai TL, Shih MC (2013) *Sequential Experimentation in Clinical Trials* (New York: Springer).

[8] Beck A, Teboulle M (2003) Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters* 31(3):167–175.

[9] Bertsekas D (1999) *Nonlinear Programming* (Athena Scientific).

[10] Blumenthal AL (1977) *The Process of Cognition.* (Prentice Hall/Pearson Education).

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

[12] Braverman M, Mao J, Weinberg MS (2016) Parallel algorithms for select and partition with noisy comparisons. *Proceedings of Annual Symposium on the Theory of Computing.*

[13] Braverman M, Mossel E (2009) Sorting from noisy information, arXiv preprint arXiv:0910.1191.

[14] Bubeck S (2015) Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning* 8(3-4):231–357.

[15] Chen X, Bennett PN, Collins-Thompson K, Horvitz E (2013) Pairwise ranking aggregation in a crowdsourced setting. *Proceedings of ACM International Conference on Web Search and Data Mining.*

[16] Chen X, Gopi S, Mao J, Schneider J (2018) Optimal instance adaptive algorithm for the top-k ranking problem. *IEEE Transactions on Information Theory* 64(9):6139–6160.

[17] Chen X, Jiao K, Lin Q (2016) Bayesian decision process for cost-efficient dynamic ranking via crowdsourcing. *Journal of Machine Learning Research* 17(217):1–40.

[18] Chen X, Li Y, Mao J (2018) An instance optimal algorithm for top-K ranking under the multinomial logit model. *ACM-SIAM Symposium on Discrete Algorithms (SODA).*

[19] Chen Y, Suh C (2015) Spectral MLE: Top-k rank aggregation from pairwise comparisons. *Proceedings of International Conference on Machine Learning.*

[20] Chernoff H (1959) Sequential design of experiments. *The Annals of Mathematical Statistics* 30(3):755–770.

[21] Dragalin VP, Tartakovsky AG, Veeravalli VV (2000) Multihypothesis sequential probability ratio tests. ii. accurate asymptotic expansions for the expected sample size. *IEEE Transactions on Information Theory* 46(4):1366–1383.

[22] Draglia V, Tartakovsky AG, Veeravalli VV (1999) Multihypothesis sequential probability ratio tests. i. asymptotic optimality. *IEEE Transactions on Information Theory* 45(7):2448–2461.

[23] Elo AE (1978) *The Rating of Chessplayers, Past, and Present* (Arco Pub.).

[24] Garg N, Johari R (2019) Designing optimal binary rating systems. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics.*

[25] Hajek B, Oh S, Xu J (2014) Minimax-optimal inference from partial rankings. *Proceedings of Advances in Neural Information Processing Systems.*

[26] Heckel R, Shah NB, Ramchandran K, Wainwright MJ (2019) Active ranking from pairwise comparisons and when parametric assumptions do not help. *The Annals of Statistics* 47(6):3099 – 3126.

[27] Hoeffding W (1960) Lower bounds for the expected sample size and the average risk of a sequential procedure. *The Annals of Mathematical Statistics* 31(2):352–368.

[28] Hoeffding W (1963) Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58(301):13–30.

[29] Hsiung AC, Ying ZL, Zhang CH, eds. (2004) *Random Walk, Sequential Analysis and Related Topics: A Festschrift in Honor of Yuan-Shih Chow* (World Scientific).

[30] Jamieson K, Nowak R (2011) Active ranking using pairwise comparisons. *Advances in Neural Information Processing Systems.*

[31] Kallus N, Udell M (2020) Dynamic assortment personalization in high dimensions. *Operation Research* 68(4):1020–1037.

[32] Kendall M, Gibbons JD (1990) *Rank Correlation Methods* (A Charles Griffin Title), 5 edition.

[33] Kiefer J, Sacks J (1963) Asymptotically optimum sequential inference and design. *The Annals of Mathematical Statistics* 34(3):705–750.

[34] Lai TL (1988) Nearly optimal sequential tests of composite hypotheses. *The Annals of Statistics* 16(2):856–886.

[35] Lai TL (2001) Sequential analysis: some classical problems and new challenges. *Statistica Sinica* 11:303–408.

[36] Lai TL, Shih MC (2004) Power, sample size and adaptation considerations in the design of group sequential clinical trials. *Biometrika* 91(3):507–528.

[37] Li X, Liu J (2015) Rare-event simulation and efficient discretization for the supremum of Gaussian random fields. *Advances in Applied Probability* 47(03):787–816.

[38] Li X, Liu J, Ying Z (2014) Generalized sequential probability ratio test for separate families of hypotheses. *Sequential analysis* 33(4):539–563.

[39] Li X, Liu J, Ying Z (2018) Chernoff index for cox test of separate parametric families. *The Annals of Statistics* 46(1):1 – 29.

[40] Lorden G (1976) 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size. *The Annals of Statistics* 4(2):281–291.

[41] Luce RD (1959) *Individual choice behavior: A theoretical analysis* (New York: Wiley).

[42] Mao C, Weed J, Rigollet P (2018) Minimax rates and efficient algorithms for noisy sorting. *Proceedings of the Algorithmic Learning Theory.*

[43] Mei Y (2010) Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* 97(2):419–433.

[44] Morrison HW (1963) Testable conditions for triads of paired comparison choices. *Psychometrika* 28:369–390.

[45] Naghshvar M, Javidi T (2013) Active sequential hypothesis testing. *The Annals of Statistics* 41(6):2703–2738.

[46] Negahban S, Oh S, Sha D (2017) Rank centrality: Ranking from pair-wise comparisons. *Operations Research* 65(1):266–287.

[47] Nitinawarat S, Veeravalli VV (2015) Controlled sensing for sequential multihypothesis testing with controlled markovian observations and non-uniform control cost. *Sequential Analysis* 34(1):1–24.

[48] Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

[49] Saaty TL, Vargas LG (2012) The possibility of group choice: pairwise comparisons and merging functions. *Social Choice and Welfare* 38(3):481–496.

[50] Schwarz G (1962) Asymptotic shapes of Bayes sequential testing regions. *The Annals of Mathematical Statistics* 33(1):224–236.

[51] Shah NB, Balakrishnan S, Guntuboyina A, Wainright MJ (2017) Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Transactions on Information Theory* 63(2):934–959.

[52] Siegmund D (1985) *Sequential Analysis: Tests and Confidence Intervals* (Springer New York).

[53] Song Y, Fellouris G (2017) Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals. *Electronic Journal of Statistics* 11(1):338–363.

[54] Tartakovsky A, Nikiforov I, Basseville M (2014) *Sequential Analysis: Hypothesis Testing and Change-point Detection* (Chapman and Hall/CRC).

[55] Thurstone LL (1927) A law of comparative judgement. *Psychological Review* 34(4):273–286.

[56] Train K (2009) *Discrete choice methods with simulation* (Cambridge University Press).

[57] Tsitovich I (1985) Sequential design of experiments for hypothesis testing. *Theory of Probability & Its Applications* 29(4):814–817.

[58] Wald A (1945) Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics* 16:117–186.

[59] Wald A, Wolfowitz J (1948) Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics* 19:326–339.

[60] Wang S, Lin H, Chang HH, Douglas J (2016) Hybrid computerized adaptive testing: from group sequential design to fully sequential design. *Journal of Educational Measurement* 53(1):45–62.

[61] Watkins CJCH (1989) *Learning from Delayed Rewards*. Ph.D. thesis, Cambridge University.

[62] Xie Y, Siegmund DO (2013) Sequential multi-sensor change-point detection. *The Annals of Statistics* 41(2):670–692.

[63] Ye S, Fellouris G, Culpepper S, Douglas J (2016) Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology* 69(2):139–158.