Abstract

We address the following question in this paper: “What are the most robust statistical methods for social choice?” By leveraging the theory of uniformly least favorable distributions in the Neyman-Pearson framework to finite models and randomized tests, we characterize uniformly most powerful (UMP) tests, which is a well-accepted statistical optimality w.r.t. robustness, for testing whether a given alternative is the winner under Mallows’ model and under Condorcet’s model, respectively.

1 INTRODUCTION

Suppose a group of seven friends want to choose restaurant a, b, or c for dinner. Each person uses a ranking over the restaurants to represent his or her preferences. Three people rank $a \succ b \succ c$, three people rank $b \succ c \succ a$, and one person ranks $a \succ c \succ b$. Suppose their preferences are correlated and are based on their perception of the quality of the restaurants—the higher the quality of a restaurant, the more likely a person will rank it high. Which restaurant should they choose?

Similar problems exist in a wide range of group decision-making scenarios such as political elections (Condorcet, 1785), meta-search engines (Dwork et al., 2001), recommender systems (Ghosh et al., 1999), and crowdsourcing (Mao et al., 2013). Such problems at the intersection of statistics and social choice can be dated back to Condorcet’s Jury Theorem in the 18th century (Condorcet, 1785). The Jury Theorem states that when there are two alternatives, assuming that the votes are generated i.i.d. from a simple statistical model, then the outcome of majority voting converges to the ground truth as the number of voters goes to infinity.

However, the Jury Theorem does not identify the optimal decision-making rule, especially when there are three alternatives or more. From a statistical point of view, defining the optimality measure is highly nontrivial and controversial. If we use likelihood of a parameter as the measure, then we may pursue the likelihoodist approach. If we view the ground truth parameter as a random variable, and use expected loss w.r.t. the posterior distribution over the parameters as the measure, then we may pursue the Bayesian approach. If we believe that the ground truth is deterministic and unknown, and want to measure the performance of a given rule, then we may pursue the frequentist approach. At a high level, the frequentist approach tries to measure and design the most robust rule, as Efron (2005) noted: “a frequentist is a Bayesian trying to do well, or at least not too badly, against any possible prior distribution”.

Most previous work in the literature of statistical approaches to social choice pursued either an MLE approach or an Bayesian approach. We are not aware of the application of a widely-applied modern frequentists’ decision-making technique—optimal statistical hypothesis testing—to social choice. In the celebrated Neyman-Pearson framework of statistical hypothesis testing (see, e.g. the book by Lehmann and Romano (2008)), a statistical model is given and the decision-maker first chooses two non-overlapping subsets of ground truth parameters $H_0, H_1$, where $H_0$ is called the null hypothesis and $H_1$ is called the alternative hypothesis. Then the decision-maker designs a test for $H_0$ vs. $H_1$, in the form of a critical function $f$, to make a binary decision in $\{0,1\}$ for each observed data. Here 1 means that $H_0$ should be
rejected and 0 means that there is a lack of evidence to reject \( H_0 \). We note that the role of \( H_0 \) and \( H_1 \) are not the same, namely a \( A \) vs. \( B \) test is different from a \( B \) vs. \( A \) test.

While many generic hypothesis testing methods can be applied, such as the generalized likelihood ratio tests (Hoeffding, 1965; Zeitouni et al., 1992), how to make an optimal social choice w.r.t. frequentists’ measure is still an open question.

**Our Contributions.** We answer the question of optimal hypothesis testing for social choice by characterizing uniformly most powerful (UMP) tests for various combinations of \( H_0 \) and \( H_1 \) for winner determination under two popular models for rank data: Mallows’ model and Condorcet’s model. UMP is a strong notion of optimality for hypothesis testing. A test \( f \) is evaluated by two criteria: its size (or level of significance), which is its worst-case probability to wrongly reject \( H_0 \), and its power, which is its probability to correctly reject \( H_0 \). The power of a test is evaluated at each \( h_1 \in H_1 \). A level-\( \alpha \) test \( f \) is a UMP test, if it has the highest power at every \( h_1 \in H_1 \) among all tests whose sizes are no more than \( \alpha \).

We focus on two types of tests for a given alternative \( a \): the non-winner tests, where \( H_0 \) represents \( a \) being the winner\(^2\); and the winner tests, where \( H_1 \) represents \( a \) being the winner. Our main results are summarized in Table 1.

|          | Non-winner \((H_0 = \{a \text{ wins}\})\) | Winner \((H_1 = \{a \text{ wins}\})\) |
|----------|------------------------------------------|-------------------------------------|
| Mallows  | Y\&N (Thm. 1, 2)                        | Y\&N (Thm. 3, 4, 5)                |
| Condorcet| Y\&N (Thm. 6, 7)                        | Y (Thm. 8)                         |

Table 1: UMP tests for Mallows’ model and Condorcet’s model. “Y” in Condorcet-Winner means that for any \( 0 < \alpha < 1 \), there exists a level-\( \alpha \) UMP winner test for \( \mathcal{P}_1 \) vs. \( H_1 \). “Y\&N” means that for some \( \alpha \), no level-\( \alpha \) UMP test exists for \( H_0 = \mathcal{P}_1 \); but a UMP test exists for all levels for some natural special cases.

For example, “Y\&N” in Mallows-Non-winner in Table 1 means that for some \( \alpha \), no level-\( \alpha \) UMP test exists for \( H_0 \) vs. \( \mathcal{P}_0 \), where \( H_0 \) consists of rankings where a given alternative \( a \) is ranked at the top. On the other hand, for some \( H_1 \), a level-\( \alpha \) UMP test exists for all \( 0 < \alpha < 1 \). In fact, Theorem 2 characterizes all such \( H_1 \)’s.

In particular, we obtained a complete characterizations of \( H_1 \) for which UMP non-winner tests (that is, when \( H_0 \) models “\( a \) wins”) exist, under Mallows’ model (Theorem 2) and under Condorcet’s model (Theorem 7). Technically, to obtain the characterizations, we leverage the theory of uniformly least favorable distributions to finite models and randomized tests (Lemma 3, 4, 6, 7). These lemmas generalize the key theorems by Reinhardt-Reinhardt (1961) that only hold for continuous parameter space, and they might be of independent interest.

**Significance of results.** Our results provide the first theoretical characterization of robust social choice w.r.t. frequentists’ measure. Practically, the UMP winner tests in the Condorcet-Winner column can be used for testing whether a given alternative \( a \) is a winner by appropriately setting \( H_0 \) while fixing \( H_1 \) to represent “\( a \) wins”.

**Proof techniques.** This paper focuses on composite vs. composite tests, where both \( H_0 \) and \( H_1 \) contain more than one element. Many results in this paper are proved by applying Lemma 2 (Theorem 3.8.1 and Corollary 3.8.1 in (Lehmann and Romano, 2008)), which offers necessary and sufficient conditions for composite vs. simple tests. However, applying Lemma 2 is more challenging than it appears—the key is to come up with a uniformly least favorable distribution that satisfies the conditions in Lemma 2 for all elements in \( H_1 \), and such distribution is not guaranteed to exist. As we show later in the paper, such distributions indeed exist for non-winner tests for Mallows’ model and Condorcet’s model respectively, and it is non-trivial to verify that they satisfy conditions in Lemma 2. In fact, to this end, we proved new properties (Lemma 5 and Lemma 8 in the appendix) about Mallows’ model and new general theorems (Lemma 6 and 7) that can be applied to Condorcet’s model, which might be of independent interest.

**Related Work and Discussions.** Marden (1995) applied the Neyman-Pearson Lemma (Lemma 1) for simple vs. simple tests under Mallows, as illustrated in Example 2. Most previous work in statistical approaches in social choice focused on extending the Condorcet Jury Theorem and proving asymptotic results (Gerlinga et al., 2005; Nitzan and Paroush, 2017). Previous work focused on using commonly-studied voting rules designed for elections (Conitzer and Sandholm, 2005; Caragiannis et al., 2016), maximum likelihood estimators (Conitzer and Sandholm, 2005; Xia and Conitzer, 2011), or Bayesian estimators (Young, 1988; Procaccia et al., 2012; Pivato, 2013; Elkind and Shah, 2014; Azari Soufiani et al., 2014; Xia, 2016). We are not aware of a previous work on UMP tests for deciding whether a given alternative wins or not in social choice context.

Compared to previous MLE and Bayesian approaches to social choice, optimal rules characterized in this pa-
per are more robust because it offers the best worst-case guarantee against an adversary who controls the ground truth parameter. As in the general Bayesian vs. Frequentists debate, this does not mean that one approach is better than another, because the measures of performance are different.

2 PRELIMINARIES

Let \( A = \{a_1,\ldots,a_m\} \) denote a set of \( m \geq 2 \) alternatives and let \( \mathcal{L}(A) \) denote the set of all linear orders over \( A \). Let \( n \) denote the number of agents. Each agent’s preferences are represented by a linear order in \( \mathcal{L}(A) \). We often use \( V = [a \succ_V b \succ \cdots] \) to denote a ranking, and write \( a \succ_V b \) if \( a \) is preferred to \( b \) in \( V \). Let \( P_n \) denote the collection of \( n \) agents’ votes, called an \((n,p)\)-profile. For any profile \( P \) and any pair of alternatives \( a, b \), we let \( P(a \succ b) \) denote the number of votes in \( P \) where \( a \) is preferred to \( b \).

The weighted majority graph (WMG) of \( P \), denoted by \( WMG(P) \), is a directed weighted graph where the weight \( w_P(a \succ b) \) on any edge \( a \rightarrow b \) is \( w_P(a \succ b) = P(a \succ b) - P(b \succ a) \). By definition \( w_P(a \succ b) = -w_P(b \succ a) \). For example, the WMG of the profile \( P_2 \) of seven linear orders mentioned in the beginning of Introduction has weights \( w_{P_2}(a \succ b) = w_{P_2}(a \succ c) = 1 \) and \( w_{P_2}(b \succ c) = 5 \).

Statistical Models for Rank Data. A statistical model \( \mathcal{M} = (S, \Theta, \pi) \) has three parts: the sample space \( S \), which is composed of all possible lists; the parameter space \( \Theta \); and the probability distributions \( \pi = \{\pi_\theta : \theta \in \Theta\} \). If both \( S \) and \( \Theta \) contain finitely many elements, then we call \( \mathcal{M} \) a finite model. For any pair of linear orders \( V, W \) in \( \mathcal{L}(A) \), let \( KT(V,W) \) denote the Kendall-tau distance, which is the total number of pairwise disagreements between \( V \) and \( W \). Formally,

\[
KT(V,W) = \# \left\{ \{a,b\} \subseteq A : [a \succ_V b \text{ and } b \succ_W a] \right\}
\]

Definition 1 (Mallows’ model with fixed dispersion (Mallows, 1957)) Given the dispersion \( 0 < \varphi < 1 \), Mallows’ model is denoted by \( \mathcal{M}^{Ma} = (\mathcal{L}(A)^n, \mathcal{L}(A), \pi) \), where \( n \) linear orders are i.i.d. generated, the parameter space is \( \mathcal{L}(A) \), and for any \( V,W \in \mathcal{L}(A) \),

\[
\pi_{V,W}(V) = \frac{1}{Z} \varphi^{KT(V,W)},
\]

where \( Z \) is the normalization factor.

Condorcet’s model differs from Mallows’ model by allowing ties in the ground truth and in data. Formally, let \( B(A) \) denote the set of all irreflexive, antisymmetric, and total binary relations over \( A \). We have \( \mathcal{L}(A) \subseteq B(A) \) and the Kendall-tau distance is extended to \( B(A) \) by counting the number of pairwise disagreements.

Definition 2 (Condorcet’s model for binary relations with fixed dispersion) Given the dispersion \( 0 < \varphi < 1 \), Condorcet’s model is denoted by \( \mathcal{M}^{Co} = (B(A)^n, B(A), \pi) \), where the parameter space is \( B(A) \) and for any \( W \in B(A) \) and \( V \in B(A) \),

\[
\pi_{V,W}(V) = \frac{1}{Z} \varphi^{KT(V,W)},
\]

where \( Z \) is the normalization factor.

In classical Condorcet’s model (Condorcet, 1785; Young, 1988), the sample space consists of linear orders and the parameter space consists of binary relations. The model in Definition 2 is a variant of Condorcet’s model, where the sample space consists of binary relations. In other words, each agent is allowed to use a binary relation to represent his or her preferences—transitivity is not required as in classical Condorcet’s model or Mallows’ model.

Statistical Hypothesis Testing: The Neyman-Pearson Framework. Given a statistical model \( \mathcal{M} = (S, \Theta, \pi) \), the decision-maker first chooses two non-overlapping subsets of parameters \( H_0, H_1 \subseteq \Theta \), where \( H_0 \) is called the null hypothesis and \( H_1 \) is called the alternative hypothesis. The goal of hypothesis testing is to decide whether the ground truth parameter is in \( H_0 \) (retaining the null hypothesis) or in \( H_1 \) (rejecting the null hypothesis), based on the observed data \( P \in S \).

To simplify notation, we let 0 denote retain and let 1 denote reject. A test is characterized by a (randomized) critical function \( f : S \rightarrow [0,1] \) such that for any \( P \in S \), with probability \( f(P) \) the outcome of testing is 1 (reject). When \( H_0 \) (or \( H_1 \)) contains a single parameter, it is called a simple hypothesis; otherwise it is called a composite hypothesis.

A test \( f \) is often evaluated by its size and power. The size of \( f \) is the maximum probability for \( f \) to wrongly outputs 1 when the ground truth is in \( H_0 \) (such cases are called Type I errors or false positives), where the max is taken over all parameters in \( H_0 \). More precisely, for any \( h_0 \in H_0 \), we let \( \text{Size}(f,h_0) = E_{P\sim \pi_{h_0}} f(P) \), and \( \text{Size}(f) = \sup_{h_0 \in H_0} \text{Size}(f,h_0) \). If the size of \( f \) is \( \alpha \), then \( f \) is called a level-\( \alpha \) test. For any \( h_1 \in H_1 \), the power of \( f \) at \( h_1 \) is the probability that \( f \) correctly outputs 1 when the ground truth is \( h_1 \). More precisely, we let \( \text{Power}(f,h_1) = E_{P\sim \pi_{h_1}} f(P) \), where the expectation is take over randomly generated profiles from \( \pi_{h_1} \). We would like a test \( f \) to have low size and high power, but often tradeoffs must be made.

Example 1 Let \( \mathcal{M}^{Ma} \) denote a Mallows’ model with \( m = 3 \) and \( n = 1 \). Let \( A = \{1,2,3\} \), \( h_1 = [1 \succ 2 \succ 3] \), and let \( H_0 \) denote the other rankings. Let \( f \) be a test
where \( f(1 \triangleright 2 \triangleright 3) = 1, f(2 \triangleright 1 \triangleright 3) = f(1 \triangleright 3 \triangleright 2) = 0.5, \) and \( f \) outputs \( 0 \) for all other rankings. We have \( \text{Size}(f) = \text{Size}(f, 2 \triangleright 1 \triangleright 3) = \text{Size}(f, 1 \triangleright 3 \triangleright 2) = (0.5 + \varphi + 0.5\varphi^2)/Z, \) where \( Z \) is the normalization factor. Power\((f, 1 \triangleright 2 \triangleright 3) = (1 + \varphi)/Z.\)

Given a statistical model \( M, H_0, \) a parameter \( h_1 \not\in H_0, \) and \( 0 < \alpha < 1, \) a level-\( \alpha \) most powerful test \( f_n \) is a test with the highest power among all tests whose size is no more than \( \alpha. \) For finite \( H_0, \) a most powerful test always exists and may not be unique. For composite \( H_1, \) it is possible that for different \( h_1 \in H_1, \) the most powerful tests are different. If there exists a level-\( \alpha \) test \( f_n \) that is most powerful for all \( h_1 \in H_1, \) then \( f_n \) is called a level-

---

For composite \( H_0 \) vs simple \( H_1 = \{ h_1 \}, \) a generalization of the Neyman-Pearson lemma exists. The idea is to use a distribution \( \Lambda \) over \( H_0 \) to compress \( H_0 \) into a “combined” parameter, defined as follows.

**Definition 4** For any \( M = (S, \Theta, \pi), \) any \( H_0 \subseteq \Theta, \) and any \( h_1 \in (\Theta \setminus H_0). \) Let \( \Lambda \) denote a distribution over \( H_0 \) whose support set is denoted by \( \text{Spt}(\Lambda), \) and let \( h_0^\Lambda \) denote a new parameter whose distribution over \( S \) is the probabilistic mixture of \( \{ \pi_{h_0} : h_0 \in \Theta \} \) according to \( \Lambda. \)

For any \( 0 < \alpha < 1 \) and any \( P \in S, \)

- let \( \text{Ratio}_{\Lambda, h_0}(P) = \frac{\pi_{h_0}(P)}{\sum_{h_0 \in H_0} \Lambda(h_0)\pi_{h_0}(P)}, \) and
- let \( \text{LR}_{\Lambda, h_0}(P) \) denote the likelihood ratio test for \( h_0^\Lambda \) vs. \( h_1 \) as in Definition 3.

The following lemma states that \( \text{LR}_{\Lambda, h_0}(P) \) is a most powerful test for \( H_0 \) vs. \( h_1 \) iff two conditions are satisfied.

**Lemma 2 (Theorem 3.8.1 and Corollary 3.8. by Lehmann and Romano (2008))** For composite vs. simple test \( (H_0 \text{ vs. } h_1) \) and any distribution \( \Lambda \) over \( H_0, \) the likelihood ratio test \( \text{LR}_{\Lambda, h_0}(P) \) is a level-\( \alpha \) most powerful test if and only if the following two conditions are satisfied.

(i) For any \( h_0^\Lambda \in \text{Spt}(\Lambda), \) \( \text{Size}(\text{LR}_{\Lambda, h_0}(P)) = k_\alpha. \)

(ii) For any \( h_0 \in H_0, \) \( \text{Size}(\text{LR}_{\Lambda, h_0}(P)) \leq k_\alpha. \)

Moreover, if there is no \( P \in S \) with \( \text{Ratio}_{\Lambda, h_0}(P) = k_\alpha, \) then \( \text{LR}_{\Lambda, h_0}(P) \) is the unique level-\( \alpha \) most powerful test.

The distribution \( \Lambda \) in Lemma 2 is called a least favorable distribution. If \( |\text{Spt}(\Lambda)| = 1, \) then \( \Lambda \) is called a deterministic least favorable distribution. If \( \Lambda \) is a least favorable distribution for all levels of significance \( 0 < \alpha < 1, \) then it is called a uniformly least favorable distribution (Reinhardt, 1961).

### 3 TEST SETUP AND BASIC LEMMAS

We first introduce two types of hypothesis tests for choices. Given an alternative \( a, \) for Mallows’ model we define \( L_{a>\text{others}} = \{ V \in \mathcal{L}(A) : \forall b \in A, a \succeq_V b \}; \) similarly, for Condorcet’s model we define \( R_{a>\text{others}} = \{ V \in \mathcal{B}(A) : \forall b \in A, a \succ_V b \} \). \( L_{a>\text{others}} \) and \( R_{a>\text{others}} \) naturally correspond to \( a \) being ranked at the top in the the ground truth in Mallows’ model and in Condorcet’s model, respectively.

**Definition 5** (Non-)Winner Tests Given an alternative \( a, \) in a non-winner test for Mallows’ model, we let \( H_0 = L_{a>\text{others}}; \) and in a winner test for Mallows’ model, we let \( H_1 = L_{a>\text{others}}. \)
Given an alternative \(a\), in a non-winner test for Condorcet’s model, we let \(H_0 = R_a\)-others; and in a winner test for Condorcet’s model, we let \(H_1 = R_a\)-others.

The rationale behind the naming of “non-winner” and “winner” is the following. Because \(H_0\) is often chosen as the devil’s advocate and the goal of testing is often to reject \(H_0\), when setting \(H_0 = L\text{ others}\) under Mallows’ model, we are hoping to reject \(H_0\), which means that \(\gamma\) is not the winner. We note that the decision-maker still needs to specify \(H_1\) in a non-winner test and specify \(H_0\) in a winner test. Various natural choices of \(H_1\) or \(H_0\) will be explored in Section 4 and Section 5.

We now present two general lemmas on least favorable distributions that will be frequently used in this paper. For any model \(M = (\Theta, \Theta, \pi)\), any composite vs. simple test (\(H_0\) vs. \(H_1\)), any distribution \(H_{\pi}\) over \(H_0\), and any \(\pi_{h_0}\) in \(H_0\), we define a random variable \(X^\Lambda_{h_0} : \Theta \rightarrow \mathbb{R}\) such that for any \(P \in \Theta\), \(Pr(P) = \gamma_{h_0}(P)\) and \(X^\Lambda_{h_0}(P) = \log \text{Ratio}_{\Lambda,h_1}(P)\). A random variable \(X\) weakly first-order stochastically dominates (weakly dominates for short) another random variable \(Y\), if for all \(p \in \Theta\), \(Pr(X \geq p) \geq Pr(Y \geq p)\).

**Lemma 3** \(\Lambda\) is a uniformly least favorable distribution for \(H_0\) vs. \(H_1\) if and only if for any \(h_0^* \in \text{Spt}(\Lambda)\) and any \(\pi_{h_0} \in H_0\), \(X^\Lambda_{h_0}\) weakly dominates \(X^\Lambda_{h_0^*}\).

**Proof:** To simplify notation we let \(LR_{\pi}\) and Ratio to denote \(LR_{\alpha,\Lambda,h_1}\) and \(\text{Ratio}_{\Lambda,h_1}\), respectively. For any \(0 < \alpha < 1\) and any \(\pi_{h_0} \in H_0\), we have

\[
\begin{align*}
\text{Size}(LR_{\alpha}, h_0) &= \sum_{P \in \Theta: \text{Ratio}(P) > k_{\alpha}} \pi_{h_0}(P) \\
&+ \gamma_{\alpha} \sum_{P \in \Theta: \text{Ratio}(P) = k_{\alpha}} \pi_{h_0}(P) \\
&= Pr(X^\Lambda_{h_0} > \log k_{\alpha}) + \gamma_{\alpha} Pr(X^\Lambda_{h_0} = \log k_{\alpha}) \\
&= (1 - \gamma_{\alpha}) Pr(X^\Lambda_{h_0} > \log k_{\alpha}) + \gamma_{\alpha} Pr(X^\Lambda_{h_0} \geq \log k_{\alpha}) \\
&= (1 - \gamma_{\alpha}) \lim_{x \to \log k_{\alpha}} Pr(X^\Lambda_{h_0} \geq x) + \gamma_{\alpha} Pr(X^\Lambda_{h_0} \geq k_{\alpha})
\end{align*}
\]

The “if” direction: for any \(h_0 \in H_0\) and any \(h_0^* \in \text{Spt}(\Lambda)\), because \(X^\Lambda_{h_0}\) weakly dominates \(X^\Lambda_{h_0^*}\), we have that for any \(x \in \mathbb{R}\), \(Pr(X^\Lambda_{h_0} \geq x) \geq Pr(X^\Lambda_{h_0^*} \geq x)\). It follows from (1) that \(\text{Size}(LR_{\alpha}, h_1, h_0^*) \geq \text{Size}(LR_{\alpha}, h_0)\). By Lemma 2, \(LR_{\alpha}\) is a level-\(\alpha\) most powerful test. Therefore \(\Lambda\) is a uniformly least favorable distribution.

The “only if” direction: suppose for the sake of contradiction that this is not true. Let \(h_0 \in H_0\) and \(h_0^* \in \text{Spt}(\Lambda)\) be such that \(X^\Lambda_{h_0}\) does not weakly dominate \(X^\Lambda_{h_0^*}\). It follows that there exists \(x \in \mathbb{R}\) such that \(Pr(X^\Lambda_{h_0^*} \geq x) < Pr(X^\Lambda_{h_0} \geq x)\). Let \(\alpha = Pr(X^\Lambda_{h_0} \geq x)\).

Because \(\Lambda\) is uniformly least favorable, the size of \(LR_{\alpha}\) must be \(\alpha\), where \(k_{\alpha} = 2^\alpha\) and \(\gamma_{\alpha} = 1\). By Lemma 2, \(Pr(X^\Lambda_{h_0} \geq x) = \text{Size}(LR_{\alpha}, h_0) \geq \text{Size}(LR_{\alpha}, h_0^*) = Pr(X^\Lambda_{h_0^*} \geq x)\), which is a contradiction.

**Example 3** Let \(M\) denote a Mallows’ model with \(m = 3\) and \(n = 1\). Let \(A = \{1, 2, 3\}\), \(h_1 = [1 \succ 2 \succ 3]\) and let \(\Lambda\) denote the uniform distribution over \([2 \succ 1 \succ 3], [1 \succ 3 \succ 2]\). We will apply Lemma 3 to prove that \(\Lambda\) is a uniformly least favorable distribution for testing \(H_0 = (L(A) - \{[1 \succ 2 \succ 3]\})\) vs. \([1 \succ 2 \succ 3]\). The likelihood ratios of all rankings are summarized in Table 2 in the increasing order.

![Table 2: Likelihood ratios.](image)

Table 3: \(X^\Lambda_{h_0}\) for all \(h_0 \in H_0\), where \(Z\) is the normalization factor.

Because \(0 < \varphi < 1\), it is not hard to verify that \(X^\Lambda_{1\succ3\succ2}\) and \(X^\Lambda_{2\succ1\succ3}\) weakly dominate other random variables. By Lemma 3, \(\Lambda\) is a uniformly least favorable distribution.

Our second lemma states that if we can find a deterministic uniformly least favorable distribution for \(n = 1\), then it is also uniformly least favorable for the same statistical model with \(n \geq 2\) i.i.d. samples.

**Lemma 4** Suppose \(\Lambda\) is a deterministic uniformly least favorable distribution for composite vs. simple test (\(H_0\) vs. \(h_1\)) under \(M = (\Theta, \Theta, \pi)\). Then for any \(n \in \mathbb{N}\), \(\Lambda\) is also a uniformly least favorable distribution for testing \(H_0\) vs. \(h_1\) under \(M = (\Theta^n, \Theta, \pi)\) with \(n\) i.i.d. samples.

All missing proofs can be found in the appendix.

### 4 UMP Tests for Mallows

In this section, we present results on UMP non-winner and winner tests for Mallows’ model.
Non-Winner Tests for Mallows. The first theorem (Theorem 1) of this subsection is a warmup, whose main goal is to define a test \( f_{\alpha,B} \) that is UMP for any simple \( H_1 \) that consists in a linear order where \( a \) is not ranked at the top. The main theorem of this section is Theorem 1, which characterizes all UMP non-winner tests for arbitrary choices \( H_1 \).

For any profile \( P \), any \( B \subseteq A \), and any \( a \in (A - B) \), we let \( w_P(B \succ a) = \sum_{b \in B} w_P(b \succ a) \), that is, the total weights on edges from \( B \) to \( a \) in WMG(\( P \)).

**Theorem 1 (A Most powerful non-winner test for Mallows)** Given a Mallows’ model \( M^\alpha \), for any alternative \( a \), any ranking \( h_1 \) where \( a \) is not ranked at the top, any \( \alpha, \beta < 1 \), and any \( n \), the following test is a level-\( \alpha \) UMP for testing \( \Lambda_\alpha \)-others vs. \( h_1 \). For any \( n \)-profile \( P_n \),

\[
f_{\alpha,a,B}(P_n) = \begin{cases} 1 & \text{if } w_P(a \succ B) > K_{\alpha} \\ 0 & \text{if } w_P(a \succ B) < K_{\alpha} \\ 1 - \alpha & \text{if } w_P(a \succ B) = K_{\alpha} \end{cases}
\]

where \( B \) is the set of alternatives ranked above \( a \) in \( h_1 \), and \( K_{\alpha} \) and \( \alpha \) are chosen s.t. the size of \( f_{\alpha,a,B} \) is \( \alpha \).

**Proof:** The proof proceeds by identifying a uniformly least favorable distribution for \( H_0 = \Lambda_\alpha \)-others vs. \( h_1 \). In fact, let \( B \) denote the set of alternatives ranked above \( a \) in \( h_1 \). Let \( h_0 \) denote the ranking that is obtained from \( h_1 \) by raising \( a \) to the top position. We will prove that the deterministic distribution \( \Lambda \) at \( \{h_0\} \) is a uniformly least favorable distribution.

Let \( LR_{\alpha} \) denote \( LR_{\alpha,h_0,h_1} \), and let \( Ratio \) denote \( Ratio_{\alpha,h_0,h_1} \). Recall that both are defined in Definition 3. We first prove the theorem for \( n = 1 \). By Lemma 3, it suffices to prove that for any \( h_0 \in H_0 \), \( X^{\alpha}_{h_0} \) weakly dominates \( X^{\alpha}_{h_1} \). For any ranking \( V \) and any pair of alternatives \( b, c \), we let \( I(b \succ V, c) = 1 \) if \( b \succ V, c \) otherwise \( I(b \succ V, c) = 0 \). For any single-vote profile \( P = \{V\} \), we have:

\[
\log \text{Ratio}(P) = (KT(V, h_1) - KT(V, h_0)) \log \varphi = \log \varphi \sum_{c \succ V} (I(d \succ h_1, c) - I(d \succ h_0, c))
\]

\[
= \log \varphi \left( \sum_{b \in B \cup a} (I(b \succ h_1, a) - I(b \succ h_0, a)) + \sum_{b \in B \cup a} (I(a \succ h_1, b) - I(a \succ h_0, b)) \right)
\]

\[
= \log \varphi \cdot |B| - 2w_P(B \succ a)
\]

Therefore, to prove that \( X^{\alpha}_{h_0} \) weakly dominates \( X^{\alpha}_{h_1} \), it suffices to prove for any \( K \in \mathbb{Z}\),

\[
\pi_{h_0}(\{P : w_P(B \succ a) \geq K\}) \leq \pi_{h_0}(\{P : w_P(B \succ a) \geq K\})
\]

Let \( M \) denote the permutation over \( A \) such that \( M(h_0) = h_0 \). Because \( h_0 \in H_0 = \Lambda_\alpha \)-others, we have \( M(a) = a \). Let \( B' = M(B) \). Because Kendall-Tau distance is invariant to permutations, for any \( P \in \mathcal{L}(A) \) we have \( \pi_{h_0}(P) = \pi_{M(h_0)}(M(P)) \) and

\[
\pi_{h_0}(\{P : w_P(B \succ a) \geq K\}) = \pi_{M(h_0)}(\{M(P) : w_{M(P)}(M(B) \succ M(a)) \geq K\})
\]

Therefore, it suffices to prove that \( \pi_{h_0}(\{P : w_P(B \succ a) \geq K\}) \leq \pi_{h_0}(\{P : w_P(B \succ a) \geq K\}) \). We will prove a stronger lemma. Given any \( W \in \mathcal{L}(A) \) and \( C', C \subseteq A \) with \( C \neq C' \) and \( |C| = |C'| \), we say that \( C \) dominates \( C' \) w.r.t. \( W \) if there exists a one-one mapping \( F \) : \( C - C' \to C' - C \) such that for all \( c \in C \) we have \( c \succ_W F(c) \). In words, \( C' \) can be obtained from \( C \) by lowering some alternatives according to \( W \).

**Lemma 5** Under a Mallows’ model, for any \( \varphi \), any \( K \in \mathbb{N} \), any \( a \in A \), any \( W \in \mathcal{L}(A) \), and any \( C', C \subseteq A \) such that \( C \) dominates \( C' \) w.r.t. \( W \), we have \( \pi_{W}(\{P : w_{W}(C' \succ a) \geq K\}) \leq \pi_{W}(\{P : w_{W}(C \succ a) \geq K\}) \).

It follows from Lemma 5 that \( X^{\alpha}_{h_0} \) weakly dominates \( X^{\alpha}_{h_1} \), which means that \( \Lambda \) is a uniformly least favorable distribution for \( n = 1 \) by Lemma 3. We note that \( \Lambda \) is deterministic. Therefore, by Lemma 4, \( \Lambda \) is also a uniformly least favorable distribution for Mallows’ model with any \( n \in \mathbb{N} \), which means that the corresponding likelihood ratio test \( LR_{\alpha} \) is most powerful. It is not hard to verify that \( LR_{\alpha} = f_{\alpha,a,B} \). Moreover, because \( \Lambda \) is deterministic, any most powerful test \( f \) for \( H_0 \) vs. \( h_1 \) must also be most powerful for the simple vs. simple test \( (h_0^* \text{ vs. } h_1) \). By the Neyman-Pearson lemma (Lemma 1), \( f \) must agree with \( LR_{\alpha} \) except on \( P_n \) such that \( Ratio(P_n) = \kappa_{\alpha} \), which corresponds to \( P_n \) with \( w_{P_n}(B \succ a) = K_{\alpha} \).

Theorem 1 can be extended to the following characterization of all UMP non-winner tests for Mallows’ model. For any \( B \subseteq A \) and any \( a \in (A - B) \), we let \( L_{B,a} \subseteq \mathcal{L}(A) \) denote the set of all rankings where the set of alternatives ranked above \( a \) is exactly \( B \). For example, when \( m = 4 \), \( L_{[c] \succ a} = \{[c \succ a \succ b \succ d], [c \succ a \succ d \succ b] \} \).

**Theorem 2 (Characterization of UMP non-winner tests for Mallows)** Given a Mallows’ model \( M^\alpha \) with \( m \geq 2 \) and \( n \geq 2 \), there exists a UMP test for \( H_0 \) = \( \Lambda_\alpha \)-others vs. \( H_1 \) for all \( 0 < \alpha < 1 \) if and only if there exists \( B \subseteq A \) such that \( H_1 \subseteq L_{B,a} \).

Moreover, when \( H_1 \subseteq L_{B,a} \), we have that \( f_{\alpha,a,B} \) as defined in Theorem 1 is a UMP test.
Example 4 Let $P_7$ denote the profile mentioned in the beginning of Introduction. Suppose we want to test whether there is enough evidence to claim that $α$ cannot be the winner. We can apply a non-winner test on $α$ by letting $H_0 = L_α$ and $H_1 = L_o$. By Theorem 2, $f_{α,B}$ is a UMP test, where $B = \{b,c\}$. The test can be done by computing the test statistic $T = w_{P_7}(B > α) = -2$, and then checking if $T$ is in the critical region $(\alpha_∗, \infty)$ for some pre-computed $α_*$. If $T \in (\alpha_∗, \infty)$, then $H_0$ is rejected, which means that $α$ should be chosen as the winner. If $T = \alpha_*$, then $H_0$ is rejected with a pre-computed probability $Γ_α$. Otherwise $H_0$ cannot be rejected, meaning that there is no enough evidence to claim that $α$ cannot be the winner.

Winner Tests for Mallows. We now consider UMP winner tests under Mallows’ model ($H_1 = L_o$) for two natural choices of $H_0$: $H_0 = L_o$ in Theorem 3, which means that $α$ is ranked in the bottom in the ground truth, and $H_0 = (L(A) - H_1)$ in Theorem 4 and 5, which means that $α$ is not ranked at the top in the ground truth.

Theorem 3 (A UMP winner test under Mallows)
Given a Mallows’ model $M_{α}$, for any alternative $α$, any $0 < \alpha < 1$, and any $n$, the following test is a level-$\alpha$ UMP for testing $H_0 = L_o$ vs. $H_1 = L_α$. For any $n$-profile $P_n$,

$$f_{α,a}(P_n) = \begin{cases} 1 & \text{if } w_{P_n}(a > others) > K_α \\ 0 & \text{if } w_{P_n}(a > others) < K_α \\ \Gamma_α & \text{if } w_{P_n}(a > others) = K_α \end{cases},$$

where $K_α$ and $\Gamma_α$ are chosen s.t. the size of $f_{α,a}$ is $α$.

Proof: For any $h_1 ∈ H_1$, we will prove that $f_{α,a}$ is a most powerful level-$α$ test. Let $h_0^* ∈ H_0$ denote the ranking that is obtained from $h_1$ by moving $a$ to the bottom position without changing the relative positions of the other alternatives. Like the proof of Theorem 1, it is not hard to check that $f_{α,a}$ is equivalent to the likelihood ratio test $LR_{α,h_0^*,h_1}$.

Because $f_{α,a}$ is invariant to permutations over $A \setminus \{α\}$, for any $h_0^* ∈ H_0$ and any permutation $M$ over $A \setminus \{α\}$, we have $Size(f_{α,a}, h_0^*) = Size(f_{α,a}, M(h_0^*))$. In particular, let $M$ denote the permutation such that $M(h_0^*) = h_0^*$. We have $Size(f_{α,a}, h_0^*) = Size(f_{α,a}, h_0^*)$. It follows from Lemma 2 that $f_{α,a}$ is most powerful, by letting $Λ$ to be the deterministic distribution on $\{h_0^*\}$.

Example 5 Let us continue with the setting in Example 4. Suppose we want to test whether there is enough evidence to claim that $α$ is the winner. We can apply a winner test on $α$ by letting $H_0 = L_o$ and $H_1 = L_o$. By Theorem 3, $f_{α,a}$ is a UMP test. The test can be done by computing the test statistic $T = w_{P_7}(α > others) = 2$, and then checking if $T$ is in the critical region $(\alpha_∗, \infty)$ for some pre-computed $α_*$. If $T \in (\alpha_∗, \infty)$, then $H_0$ is rejected, which means that $α$ should be chosen as the winner. If $T = \alpha_*$, then $H_0$ is rejected with a pre-computed probability $Γ_α$. Otherwise $H_0$ cannot be rejected, meaning that there is no enough evidence to claim that $α$ is the winner.

The following two theorems identify conditions on $ϕ$ in Mallows’ model for the UMP winner test $H_0 = (L(A) - H_1)$ vs. $H_1 = L_o$ when $n = 1$.

Theorem 4 Let $M_{α}$ denote a Mallows’ model with $n = 1$, any $m ≥ 4$, and any $ϕ < 1/m$. There exists $0 < \alpha < 1$ such that no level-$α$ UMP test exists for $H_0 = (L(A) - H_1)$ vs. $H_1 = L_o$.

Theorem 5 Let $M_{α}$ denote a Mallows’ model with $n = 1$ and any $m ≥ 4$. There exists $ϵ > 0$ such that for any $ϕ > 1 - ϵ$ and any $α$, a UMP test exists for $H_0 = (L(A) - H_1)$ vs. $H_1 = L_o$.

5 UMP TESTS FOR CONDORCET

We first prove two general theorems on UMP tests for statistical models that combine multiple independent models, and then apply them to characterize UMP tests under Condorcet’s model.

Definition 6 (Combining two models) Given two models $M_X = (S_X, Θ_X, π_X)$ and $M_Y = (S_Y, Θ_Y, π_Y)$, we let $M_X ⊗ M_Y = (S_X × S_Y, Θ_X × Θ_Y, π_X × π_Y)$, where for any $π_{θ_X, π_{θ_Y}} ∈ π_X × π_Y$ and any $P_X ∈ S_X$ and $P_Y ∈ S_Y$, we let $π_{θ_X, π_{θ_Y}}(P_X, P_Y) = π_{θ_X}(P_X) · π_{θ_Y}(P_Y)$.

Example 6 Given a Condorcet’s model $M_{α}$ with $m = 3$. Let $A = \{1, 2, 3\}$. For any pair of alternatives $\{a,b\}$, we let $M_{α} = (\{0,1\}^n, \{0,1\}, π)$ denote the restriction of $M_{α}$ on the pairwise comparison between $a$ and $b$. We have $M_{α} = M_{(1,2)} ⊗ M_{(2,3)} ⊗ M_{(1,3)}$.

Given two models $M_X$ and $M_Y$, the next theorem provides a way to leverage a least favorable distribution for a composite vs. simple test under $M_X$ to a least favorable distribution for a composite vs. simple test under the combined model $M_X ⊗ M_Y$.

Lemma 6 For any pair of models $M_X$ and $M_Y$, suppose $Λ_X$ is a least favorable distribution for composite vs. simple test ($H_0,X ⊗ M_X$) under $M_X$. For any $y_1 ∈ Θ_Y$, let $Λ^*$ be the distribution over $H_{0,X} × Θ_Y$ where for all $x ∈ H_{0,X}$, $Λ^*(x, y_1) = Λ_X(x)$. Then,
First, a number characterize UMP tests for Condorcet’s model by applying Ext. It follows from Lemma 6 that \( \Lambda^* \) is a least favorable distribution for \( (0) \times \{0,1,1\}^2 \) as follows. Let \( \tilde{\theta} \) be a vector \( \in \Theta^t \). Then, Ext is uniformly least favorable for \( \Lambda^* \). Theorem 7 (Characterization of UMP non-winner tests for Condorcet) Given a Condorcet’s model \( \mathcal{M}_X \) with \( m \geq 2 \), we will prove that the deterministic distribution over \( \{0,1\}^m \) is a uniformly least favorable distribution for \( \Lambda^* \). Let \( X = \{\{a,b\} : b \neq a\} \) denote the pairwise comparisons between alternatives in \( A \) that involve \( a \) and let \( Y \) denote the set of all other pairwise comparisons. Let \( \mathcal{M}_X = (S_X, \Theta_X, \pi_X) \) denote Condorcet’s model \( \mathcal{M}_X \) restricted to \( X \). That is, \( S_X = \{0,1\}^m \) and for any \( \theta \in \Theta_X \), any \( \pi \in \pi_X \), and any \( P_n \in S_X \), \( \pi_\theta(P_n) \propto \varphi^{Kt(\theta, P_n)} \). Similarly, let \( \mathcal{M}_Y \) denote Condorcet’s model restricted to \( Y \). It follows that \( \mathcal{M}_Y \) is a least favorable distribution for \( \Lambda^* \). Let \( h_1 = (x_1, y_1) \), where \( x_1 \in \Theta_X \) and \( y_1 \in \Theta_Y \). Let \( x_0 \in \Theta_X \) denote the vector that represents \( a \succ b \) for all \( b \in A \). By Neyman-Pearson lemma (Lemma 1), the deterministic distribution \( \Lambda_X = \{x_0\} \) is uniformly least favorable distribution for \( x_0 \). Therefore, by Lemma 6, the deterministic distribution \( \Lambda = \{x_0, y_1\} \) is uniformly least favorable distribution for \( \{x_0, y_1\} \) vs. \( \{x_1, y_1\} \). We note that the test \( \pi_\theta(h_0) < \pi_\theta(h_1) \). It is not hard to verify that \( g_{a,a,B}(P_n) \) is equivalent to the likelihood ratio test \( LR_{a,A,h_1} \), which is most powerful. The theorem follows after Lemma 2. Subsequently, we have the following characterization of UMP non-winner tests under Condorcet’s model \( (H_0 = R_{a\succ others}) \). For any \( B \subseteq A \), let \( R_{B \succ a} \subseteq B(A) \) denote the set of all binary relations where the set of alternatives that are preferred to \( a \) is \( B \). Theorem 7 (Characterization of UMP non-winner tests for Condorcet) Let \( \mathcal{M}_X \) denote a Condorcet’s model with any \( m \geq 2 \) and \( n \geq 2 \). There exists an UMP test for \( H_0 = R_{a\succ others} \) with \( m \geq 2 \) and \( n \geq 2 \). This test is uniformly least favorable distribution for \( \Lambda^* \).
Moreover, when \( H_1 \subseteq R_{B^\Rightarrow a} \), \( g_{a,a,B} \) defined in Theorem 6 is a UMP test.

The proof is similar to the proof of Theorem 2 and is thus omitted.

**Winner Tests for Condorcet.** Finally, we turn to UMP winner tests for Condorcet’s model \((H_1 = R_{a^\Rightarrow \text{others}})\).

**Theorem 8 (A UMP winner test for Condorcet)** Let \( \mathcal{M}^C_0 \) denote a Condorcet’s model with any \( m \geq 2 \), any \( n \geq 2 \), and any \( \varphi \). For any \( \alpha \), \( g_{a,a} \) defined below is a level-\( \alpha \) UMP test for \( H_0 = (B(A) \setminus H_1) \) vs. \( H_1 = R_{a^\Rightarrow \text{others}} \). For any \( P_n \),

\[
g_{a,a}(P_n) = \begin{cases} 1 & \text{if Ratio}(P_n) > K_\alpha \\ 0 & \text{if Ratio}(P_n) < K_\alpha \\ \Gamma_\alpha & \text{if Ratio}(P_n) = K_\alpha \end{cases},
\]

where \( \text{Ratio}(P_n) = \frac{m-1}{\sum_{b \neq a} w_{ub}(a \succ b)} \) and \( K_\alpha \) and \( \Gamma_\alpha \) are chosen such that the level of \( g_{a,a} \) is \( \alpha \).

**Proof:** Let \( \mathcal{M}_1 \) denote Condorcet’s model with a single sample. Let \( X_1, \ldots, X_{m-1} \) denote the \( m-1 \) pairwise comparisons between \( a \) and other alternatives. Similarly to the proof of Theorem 6, we let \( \mathcal{M}_{X_1}, \ldots, \mathcal{M}_{X_{m-1}} \) denote the restriction of \( \mathcal{M}_3 \) on the \( m-1 \) pairwise comparisons, and let \( \mathcal{M}_Y \) denote the restriction of \( \mathcal{M}^C_0 \) on other pairwise comparisons. In fact, \( \mathcal{M}_{X_1}, \ldots, \mathcal{M}_{X_{m-1}} \) are the same model. It follows that \( \mathcal{M}^C = \mathcal{M}_{X_1} \otimes \mathcal{M}_{X_2} \otimes \cdots \otimes \mathcal{M}_{X_{m-1}} \otimes \mathcal{M}_Y \).

In \( \mathcal{M}_{X_1} \), let 1 represent that \( a \) is more preferred in the pairwise comparison. Due to the Neyman-Pearson lemma (Lemma 1), the deterministic distribution \( \Lambda = \{0\} \) is a uniformly least favorable distribution for \( H_0 = \{0\} \) vs. \( h_1 = 1 \). For any \( n \in \mathbb{N} \), let \( \mathcal{M}_{X_{1,n}} \) denote \( \mathcal{M}_{X_1} \) with \( n \) i.i.d. samples. It follows from Lemma 4 that \( \Lambda \) is still a uniformly least favorable distribution for \( \mathcal{M}_{X_{1,n}} \). By Lemma 7, Ext(\( \Lambda, h_1, m-1 \)) is a uniformly least favorable distribution for Ext(\( H_0, h_1, m-1 \)) = (\{0,1\}^{m-1} \setminus \{\overline{1}\}) vs. \( h_1 = \overline{1} \) under \( \mathcal{M}_{X_{1,n}} \otimes \cdots \otimes \mathcal{M}_{X_{m-1,n}} \).

Let \( \mathcal{M}_{Y,n} \) denote the model obtained from \( \mathcal{M}_Y \) by using \( n \) i.i.d. samples. For any \( y_1 \in \Theta_{Y,n} \), let \( \Lambda_{y_1} \) denote the distribution that is obtained from Ext(\( \Lambda, h_1, m-1 \)) by appending \( y_1 \) to each parameter. By Lemma 4, \( \Lambda_{y_1} \) is a uniformly least favorable distribution for Ext(\( H_0, h_1, m-1 \)) × \( \Theta_{Y,n} \) vs. (\( \overline{1}, y_1 \)) under \( \mathcal{M}_{X_{1,n}} \otimes \cdots \otimes \mathcal{M}_{X_{m-1,n}} \otimes \mathcal{M}_{Y,n} \), which is the Condorcet’s model with \( n \) i.i.d. samples. We note that Ext(\( H_0, h_1, m-1 \)) × \( \Theta_{Y,n} \) = (\{0,1\}^{m-1} \setminus \{\overline{1}\}) × \Theta_{Y} = (B(A) \setminus R_{a^\Rightarrow \text{others}}). \)

This means that the likelihood ratio test \( LR_{\alpha,\Lambda_{y_1},(\overline{1}, y_1)} \) is a most powerful level-\( \alpha \) test for (\( B(A) \setminus R_{a^\Rightarrow \text{others}} \)) vs. (\( \overline{1}, y_1 \)). We note that for all \( y_1 \), \( LR_{\alpha,\Lambda_{y_1},(\overline{1}, y_1)} \) is the same test, which means that it is also UMP. The theorem is proved after noticing that \( g_{a,a} = LR_{\alpha,\Lambda_{y_1},(\overline{1}, y_1)} \).

6 DISCUSSION: BEYOND BINARY CHOICE

All UMP tests we have characterized so far are optimal in making binary decisions, such as whether a given alternative \( a \) is the winner. We propose two natural procedures to choose the winner by combining multiple winner tests \((H_1 = L_{a^\Rightarrow \text{others}} \text{ for Mallows’ model})\) and \((H_1 = R_{a^\Rightarrow \text{others}} \text{ for Condorcet’ model})\) and non-winner tests \((H_0 = L_{a^\Rightarrow \text{others}} \text{ for Mallows’ model})\) and \((H_0 = R_{a^\Rightarrow \text{others}} \text{ for Condorcet’ model})\).

**Procedure based on combining winner tests.** We first choose any winner test, such as a UMP test characterized in Theorem 3, then find the alternative \( a \) with the minimum \( \alpha \) such that \( H_0 \) is rejected in the winner test, by conducting binary search on \( \alpha \).3 This corresponds at a high level to choosing the alternative that is most likely to be the winner according to the tests.

**Procedure based on combining non-winner tests.** Similarly, we use binary search on \( \alpha \) to find the alternative \( a \) with the maximum \( \alpha \) such that \( H_0 \) is rejected in the non-winner test. This corresponds to choosing the alternative that is mostly unlikely to be a non-winner according to the tests.

Interestingly, both procedures correspond to the Borda voting rule when the proposed UMP tests for Mallows’ model are used: in the UMP winner test we let \( H_0 = L_{a^\Rightarrow \text{others}} \) vs. \( H_1 = L_{a^\Rightarrow \text{others}} \) as in Example 5, and in the UMP non-winner test we let \( H_0 = L_{a^\Rightarrow \text{others}} \) vs. \( H_1 = L_{a^\Rightarrow \text{others}} \) as in Example 4. This provides a new theoretical justification for the Borda rule; or vice versa, Borda provides a justification of the proposed procedure.

7 FUTURE WORK

An immediate open question is how to use hypothesis testing for choosing a winner beyond testing whether a given alternative is a winner or not, following the initial thoughts discussed in Section 6. Also, can we characterize UMP tests for other goals of social choice, such as pairwise comparisons? Do UMP tests exist for other statistical models, such as random utility models? How can we efficiently compute the results of the proposed tests?

---

3 Co-winners exist if they all reject \( H_0 \) for the same \( \alpha \).
References

Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Statistical decision theory approaches to social choice. In *Proceedings of Advances in Neural Information Processing Systems*, 2014.

Ioannis Caragiannis, Ariel D. Procaccia, and Nisarg Shah. When Do Noisy Votes Reveal the Truth? *ACM Transactions on Economics and Computation*, 4(3): Article No. 15, 2016.

Marquis de Condorcet. *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L’Imprimerie Royale, 1785.

Vincent Conitzer and Tuomas Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence*, pages 145–152, 2005.

Griselda Deelstra and Guillaume Plantin. *Risk Theory and Reinsurance*. Springer, 2014.

Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th World Wide Web Conference*, pages 613–622, 2001.

Edith Elkind and Nisarg Shah. Electing the Most Probable Without Eliminating the Irrational: Voting Over Intransitive Domains. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, pages 182–191, 2014.

Kerstin Gerlinga, Hans Peter Grünera, Alexandra Kielc, and Elisabeth Schulte. Information acquisition and decision making in committees: A survey. *European Journal of Political Economy*, 21(3):563—597, 2005.

Sumit Ghosh, Manisha Mundhe, Karina Hernandez, and Sandip Sen. Voting for movies: the anatomy of a recommender system. In *Proceedings of the third annual conference on Autonomous Agents*, pages 434–435, 1999.

Wassily Hoeffding. Asymptotically Optimal Tests for Multinomial Distributions. *The Annals of Mathematical Statistics*, 36(2):369–401, 1965.

Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, third edition edition, 2008.

Colin L. Mallows. Non-null ranking model. *Biometrika*, 44(1/2):114–130, 1957.

Andrew Mao, Ariel D. Procaccia, and Yiling Chen. Better human computation through principled voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, Bellevue, WA, USA, 2013.

John I. Marden. *Analyzing and modeling rank data*. Chapman & Hall, 1995.

Shmuel Nitzan and Jacob Paroush. Collective decision making and jury theorems. In Francesco Parisi, editor, *The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts*. Oxford University Press, 2017.

Marcus Pivato. Voting rules as statistical estimators. *Social Choice and Welfare*, 40(2):581–630, 2013.

Ariel D. Procaccia, Sashank J. Reddi, and Nisarg Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 2012.

H. E. Reinhardt. The Use of Least Favorable Distributions in Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 32(4):1034–1041, 1961.

Lirong Xia. Bayesian estimators as voting rules. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, pages 785–794, 2016.

Lirong Xia and Vincent Conitzer. A maximum likelihood approach towards aggregating partial orders. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, pages 446–451, Barcelona, Catalonia, Spain, 2011.

H. Peyton Young. Condorcet’s theory of voting. *American Political Science Review*, 82:1231–1244, 1988.

Ofer Zeitouni, Jacob Ziv, and Neri Merhav. When is the generalized likelihood ratio test optimal? *IEEE Transactions on Information Theory*, 38(5):1597–1602, 1992.
8 Appendix: Proofs

Lemma 4. Suppose $\Lambda$ is a deterministic uniformly least favorable distribution for composite vs. simple test ($H_0$ vs. $h_1$) under $\mathcal{M} = (\mathcal{S}, \Theta, \pi)$. Then for any $n \in \mathbb{N}$, $\Lambda$ is also a uniformly least favorable distribution for testing $H_0$ vs. $h_1$ under $\mathcal{M} = (\mathcal{S}^n, \Theta, \pi)$ with $n$ i.i.d. samples.

Proof: Let $\text{Spt}(\Lambda) = \{h_0^n\}$. For any $n \in \mathbb{N}$ and any $h_0 \in H_0$, we define a random variable $X_{n,h_0} : \mathcal{S}^n \to \mathbb{R}$, where for any $P_n \in \mathcal{S}^n$, $\Pr(P_n) = \pi_{h_0}(P_n) = \prod_{V \in P_n} \pi_{h_0}(V)$, and $X_{n,h_0}(P_n) = \log \text{Ratio}_{h_0^n,h_1}$. It follows that

$$X_{n,h_0} = X_{h_0} + X_{h_0} + \cdots + X_{h_0}$$

By Lemma 3, for any $h_0 \in H_0$, $X_{h_0^n}$ weakly dominates $X_{h_0}$. Because first-order stochastic dominance is preserved under convolution (Deelstra and Plantin, 2014), we have that $X_{n,h_0^n}$ weakly dominates $X_{n,h_0}$. The lemma follows after applying Lemma 3.

Remarks. Lemma 4 is an extension of Theorem 2.3 by Reinhardt (1961) to finite models. Reinhardt’s theorem requires that for any constant $t$, with measure 0 we have $\pi_{h_0^n}(P) = t\pi_{h_1}(P)$. This is an important assumption in Reinhardt’s proof because it assumes away cases with $\text{Ratio}(P) = k_\alpha$ so that the most powerful test is deterministic. Unfortunately, this assumption does not hold for finite models and we must deal with randomized tests.

Lemma 5 Under a Mallows’ model, for any $\varphi$, any $K \in \mathbb{N}$, any $a \in A$, any $W \in \mathcal{L}(A)$, and any $C', C \subseteq A$ such that $C$ dominates $C'$ w.r.t. $W$, we have $\pi_W(\{P : w_P(C' > a) \geq K\}) \leq \pi_W(\{P : w_P(C > a) \geq K\})$.

Proof: We first prove the lemma for a special case where $C$ and $C'$ differ in only one alternative, that is, $|C - C'| = 1$. Let $c \in C$ such that $c \not\in C'$. Let $c' \in C'$ such that $c' \not\in C$. Because $C$ dominates $C'$ in $W$, we have $c \succ_W c'$.

Let $P = \{P \in \mathcal{L}(A) : w_P(C > a) \geq K\}$ and $P' = \{P \in \mathcal{L}(A) : w_P(C' > a) \geq K\}$. We define the following permutation $\mathcal{M}$ over $\mathcal{L}(A)$. For any $P \in \mathcal{L}(A)$, if $c \succ_P a \succ_P c'$ then $\mathcal{M}(P)$ is the ranking that is obtained from $P$ by switching $c$ and $c'$; otherwise $\mathcal{M}(P) = P$. Because $|C - C'| = 1$, it follows that for any $P \in P - P'$, we must have $c \succ_P a \succ_P c'$ and $(C - C') \succ_P a$. Therefore, $\mathcal{M}(P - P') = P' - P$.

We now prove that $\pi_W(P - P') \geq \pi_W(P' - P)$. For any $P \in P - P'$, we have $c \succ_P a \succ_P c'$, which means $\pi_W(P) \geq \pi_W(\mathcal{M}(P))/\varphi$ because $c \succ_W c'$. Therefore, $\pi_W(P - P') \geq \pi_W(P' - P)$ because $\mathcal{M}(P - P') = P' - P$.

We have $\pi_W(P) = \pi_W(P \cap P') + \pi_W(P - P') \geq \pi_W(P \cap P') + \pi_W(P' - P) = \pi_W(P')$.

Therefore, the lemma holds for the case where $|C - C'| = 1$. For general $C$ and $C'$, because $C$ dominates $C'$, there exists a sequence of sets $C = C_0, C_1, \ldots, C_l = C'$ such that for all $0 \leq i \leq l - 1$, (i) $C_i$ dominates $C_{i+1}$; (ii) $|C_i - C_{i+1}| = 1$. It follows that $\pi_W(\{P : w_P(C > a) \geq K\}) \geq \pi_W(\{P : w_P(C_1 > a) \geq K\}) \geq \cdots \geq \pi_W(\{P : w_P(C' > a) \geq K\})$.

Theorem 2 (Characterization of all UMP non-winner tests under Mallows). Given a Mallows’ model $\mathcal{M}^{\alpha}$ with $m \geq 2$ and $n \geq 2$, there exists a UMP test for $H_0 = L_{a>\text{others}}$ vs. $H_1$ for all $0 < \alpha < 1$ if and only if there exists $B \subseteq A$ such that $H_1 \subseteq L_{B>\alpha}$.

Moreover, when $H_1 \subseteq L_{B>\alpha}$, $f_{a,a,B}$ defined in Theorem 1 is a UMP test.

Proof: The “if” part. We note that $f_{a,a,B}$ does not depend on the orderings among alternatives in $B$ in $h_1$. It follows that for all $h_1 \in H_1$, $f_{a,a,B}$ is a level-$\alpha$ most powerful test for $H_0$ vs. $\{h_1\}$, which means that $f_{a,a,B}$ is a UMP test. The “only if” part. Suppose there exist $B, B'$ such that $B \neq B'$ and there exist two rankings $h_1^B = |B > a > \text{others}|$ and $h_1^B = |B' > a > \text{others}|$ in $H_1$. W.l.o.g. suppose $B'-B \neq \emptyset$. Let $\alpha$ denote the number such that $K_\alpha = n|B| - 0.5$, $\Gamma_\alpha = 0$, and let $f_{a,a,B}$ denote the most powerful test for $H_0$ vs. $h_1^B$ guaranteed by Theorem 1. Because $K_\alpha$ is not an integer, there does not exist $P_n$ such that $w_{P_n}(B > a) = K_\alpha$. This means that $f_{a,a,B}$ is the unique most powerful level-$\alpha$ test for $H_0$ vs. $h_1^B$. We observe that for any $P_n$, $f_{a,a,B}(P_n)$ is either 0 or 1, and $f_{a,a,B}(P_n) = 1$ if and only if $a$ is ranked below $B$ in all $n$ rankings in $P_n$. It follows that $f_{a,a,B}$ must be the unique level-$\alpha$ UMP test for $H_0$ vs. $H_1$. 

\[\therefore\]
By Theorem 1, any most powerful level-\(\alpha\) test, in particular \(f_{a,a,B}\), must agree with \(f_{a,a,B'}\) except for the threshold cases \(w_{P_n}(B' \succ a) = K'_a\) for some \(K'_a\). Choose arbitrary \(b' \in B' - B\) and let \(P'_n\) be composed of \(n\) copies of \([B' \succ a \succ \text{others}]\) and one copy of \([b' \succ (B - \{b\}) \succ a \succ \text{others}]\). Because \(w_{P_n}(B \succ a) = n|B| > K_a\), we have \(f_{a,a,B}(P'_n) = 1\). This means that the threshold \(K'_a\) for \(f_{a,a,B}\) is no more than \(w_{P_n}(B' \succ a) = n|B \cap B'|\). Because \(n \geq 2\), we have \(w_{P_n}(B' \succ a) \geq n(|B \cap B'| + 1) - 1 > n|B \cap B'| = w_{P_n}(B' \succ a)\), which means that \(f_{a,a,B}(P'_n) = 1\). However, \(w_{P_n}(B \succ a) = n|B| - 1 < n|B|\), which is a contradiction because for any profile \(P_n\), \(f_{a,a,B}(P_n) = 1\) if and only if \(B \succ a\) in all \(n\) rankings in \(P_n\).

\[\Box\]

**Theorem 4.** Let \(M^{\text{Ma}}\) denote a Mallows’ model with \(n = 1\), any \(m \geq 4\), and any \(\varphi < 1/m\). There exists \(0 < \alpha < 1\) such that no level-\(\alpha\) UMP test exists for \(H_0 = (L(A) - H_1)\) vs. \(H_1 = L_{a>\text{others}}\).

**Proof:** By Lemma 10, if a UMP test exists then \(f_{a,a}\) is also a UMP test. Therefore, it suffices to prove that \(f_{a,a}\) is not a level-\(\alpha\) UMP test. To this end, we explicitly construct a test \(f\) and prove that the rankings assigned value 1 are more cost-effective than that under \(f_{a,a}\).

Let \(V_1, V_2, \ldots, V_m, V'_2 \in L(A)\) denote \(m + 1\) rankings defined as follows. For any \(j \leq m\), let \(V_j = [a_j \succ \text{others}]\), where alternatives in “others” are ranked w.r.t. the increasing order of their subscripts. In other words, \(V_j\) is obtained from \(V_1\) by raising alternative \(a_j\) to the top position. We let \(V'_2 = [a_3 \succ a_1 \succ a_4 \succ a_2 \succ \text{others}]\).

We consider the following critical function \(f\). For any \(V \in L_{a>\text{others}}\), we let \(f(V) = 1\). For any \(V_j\) with \(j \not= 3\), let \(f(V_j) = 1\). We then let \(f(V_3) = f(V'_3) = \frac{1 + \varphi^m}{1 + \varphi}\). Let \(\alpha\) denote the size of \(f\) at \(V_2\). That is, \(\alpha = \text{Size}(f, V_2)\). Let \(T = \pi V_2(L_{a>\text{others}})\). It follows that

\[\alpha - T = \alpha \varphi^0 + \frac{1 + \varphi^m}{1 + \varphi} (\varphi^{KT(V_2,V_3)} + \varphi^{KT(V_2,V'_3)}) + \sum_{j=5}^m \varphi^{KT(V_2,V_j)} \]

\[= 1 + \frac{1 + \varphi^m}{1 + \varphi} (\varphi^3 + \varphi^4) + \varphi^4 + \sum_{j=5}^m \varphi^{KT(V_2,V_j)} \]

\[> 1 + \varphi^3 + \varphi^4 + \varphi^5\]

\[2134\]

\[4\]

\[3\]

\[3124\]

\[4123\]

\[4\]

\[3142\]

\[1\]

\[\text{Figure 1: Kentall-Tau distance for some rankings over four alternatives.}\]

For any \(j, j^* \geq 2\) such that \(j \not= j^*\), it is not hard to verify that \(KT(V_j, V_{j^*}) = j + j^* - 2\). Moreover, \(KT(V_3, V'_3) = 1\), \(KT(V_2, V'_2) = 4\), \(KT(V_4, V'_3) = 4\), and for any \(j \geq 5\), we have \(KT(V'_3, V_j) = j + 2\). Therefore, we have the following calculations of \(\text{Size}(f, V_3)\), \(\text{Size}(f, V'_3)\), and \(\text{Size}(f, V_4)\) (see Figure 1 for distances between \(V_2, V_3, V'_3, V_4\)). We note that \(T = \pi V_2(L_{a>\text{others}}) = \pi V_3(L_{a>\text{others}}) = \pi V'_3(L_{a>\text{others}}) = \pi V_4(L_{a>\text{others}})\) due to symmetry.

\[\text{Size}(f, V_3) - T \propto \varphi^3 + \frac{1 + \varphi^m}{1 + \varphi} (1 + \varphi) + \varphi^5 + \sum_{j=5}^m \varphi^{KT(V_3,V_j)} \leq 1 + \varphi^3 + (m - 3)\varphi^5\]

\[\text{Size}(f, V'_3) - T \propto \varphi^4 + \frac{1 + \varphi^m}{1 + \varphi} (1 + \varphi) + \varphi^4 + \sum_{j=5}^m \varphi^{KT(V'_3,V_j)} \leq 1 + 2\varphi^4 + (m - 4)\varphi^6\]

\[\text{Size}(f, V_4) - T \propto \varphi^4 + \frac{1 + \varphi^m}{1 + \varphi} (4 + \varphi^5) + 1 + \sum_{j=5}^m \varphi^{KT(V_4,V_j)} \leq 1 + 2\varphi^4 + (m - 4)\varphi^7\]

For any other \(h_0' \in H_0\), we have \(\text{Size}(f, h_0') - T \leq m\varphi\). Because \(\varphi < 1/m\), we have \(\text{Size}(f) = \alpha\). Let \(P\) denote a profile that is composed of \(\{V_2, V_4, \ldots, V_m\} \cup \frac{1 + \varphi^m}{1 + \varphi} \{V_3, V'_3\}\). We next prove that \(\text{Ratio}_{V_2,V_1}(P) > \text{Ratio}_{V_2,V_1}(T_{m-2})\).
Let $Z_m = \prod_{l=1}^{m} \frac{1}{1 - \varphi^l}$ denote the Mallows normalization factor for $m$ alternatives. We have

$$\text{Ratio}_{\alpha, f}(V_1, V) = \frac{\pi_{V_1}(T_{m-1})}{\pi_{V_2}(T_{m-2})}$$

$$= \frac{\varphi Z_{m-1}}{Z_{m-2} + \varphi^2(Z_{m-1} - Z_{m-2})}$$

$$= \frac{\varphi Z_{m-1}}{1 + \varphi^2(Z_{m-1} - Z_{m-2})} = \frac{\varphi + \varphi^2 + \cdots + \varphi^{m-1}}{1 + \varphi^3 + \varphi^4 + \cdots + \varphi^m} < \frac{1}{\varphi}$$

$$\text{Ratio}_{\alpha, f}(V_1, V) = \frac{\varphi + \varphi^2 + \cdots + \varphi^{m-1} + \varphi^m + \varphi^{m+2}}{1 + \varphi^3 + \varphi^4 + \cdots + \varphi^{m+3}}$$

$$> \frac{\varphi + \varphi^2 + \cdots + \varphi^{m-1}}{1 + \varphi^3 + \varphi^4 + \cdots + \varphi^{m}} = \text{Ratio}_{\alpha, f}(V_1, T_{m-2})$$

We note that $\text{Size}(\hat{f}_{\alpha, f}, V_2) = \alpha$. This means that $\text{Power}(\hat{f}_{\alpha, f}, V_1) = \pi_{V_1}(T_{m-1}) + \alpha \text{Ratio}_{\alpha, f}(V_1, T_{m-2}) < \pi_{V_1}(T_{m-1}) + \alpha \text{Ratio}_{\alpha, f}(V_1, P) = \text{Power}(f, V_1)$. This means that $\hat{f}_{\alpha, f}$ is not a level-$\alpha$ UMP. The theorem follows after Lemma 10.

**Theorem 5.** Let $\mathcal{M}^{M_i}$ denote a Mallows’ model with $n = 1$ and any $m \geq 4$. There exists $\epsilon > 0$ such that for any $\varphi > 1 - \epsilon$ and any $\alpha$, $\hat{f}_{\alpha, f}$ is a UMP test for $H_0 = (\mathcal{L}(A) - H_1)$ vs. $H_1 = L_{a \triangleright \text{others}}$.

**Proof:** We first verify that when $K_\alpha = m - 1$, $\hat{f}_{\alpha, f}$ is a UMP test. For any $h_1 \in H_1$, let $h_0^\alpha \in H_0$ denote the ranking that is obtained from $h_1$ by moving $a$ down for one position. It is not hard to check that for any $V \in \mathcal{L}(A)$, $\text{Ratio}_{h_0^\alpha, h_1}(V) \leq 1/\varphi$, and for all $V \in H_1$ we have $\text{Ratio}_{h_0^\alpha, h_1}(V) = 1/\varphi$. This means that for any level-$\alpha$ test for $H_0$ vs. $h_1$, the power cannot be more than $\alpha/\varphi$. We note that $\hat{f}_{\alpha, f}$ is a level-$\alpha$ test whose power is exactly $\alpha/\varphi$. This means that for all $h_1 \in H_1$, $\hat{f}_{\alpha, f}$ is a most powerful test for $H_0$ vs. $h_1$. Therefore, when $K_\alpha = m - 1$, $\hat{f}_{\alpha, f}$ is a UMP test.

For any $\alpha$ such that $K_\alpha \leq m - 2$, we will prove that for any $h_1 \in H_1$, $\hat{f}_{\alpha, f}$ is a most powerful level-$\alpha$ test for $H_0$ vs. $h_1$. This is done in the following steps. Step 1. Find a least favorable distribution $\Lambda_{h_0^\alpha}^{h_1}$ whose support is the set of all rankings where $a$ is ranked at the second position. Step 2. Verify that $\hat{f}_{\alpha, f}$ is the likelihood ratio test w.r.t. $\Lambda_{h_0^\alpha}^{h_1}$, and step 3. verify that the two conditions in Lemma 2 holds for $\Lambda_{h_0^\alpha}^{h_1}$.

**Step 1.** The main challenge is that in general there does not exist a uniformly least favorable distribution. For different $\alpha$ we define different $\Lambda_{h_0^\alpha}^{h_1}$ as follows. For any $\alpha$, we let $s_\alpha$ denote the smallest Borda score of the ranking $V$ such that $\hat{f}_{\alpha, f}(V) > 0$. We have that $s_\alpha \leq m - 2$. Let the support of $\Lambda_{h_0^\alpha}^{h_1}$ be $T_{m-2}$, which is the set of rankings where $a$ is ranked at the second position. We will solve the following system of linear equations to determine $\Lambda_{h_0^\alpha}^{h_1}$. For any $h_0^\alpha \in T_{m-2}$ there is a variable $x[h_0^\alpha, s_\alpha]$.

$$\forall V \in T_{s_\alpha}, \sum_{h_0^\alpha \in T_{m-2}} \text{Ratio}_{h_0^\alpha, h_1}(V) \cdot x[h_0^\alpha, s_\alpha] = m \quad (\text{LP}_{h_0^\alpha})$$

We note that as $\varphi \to 1$, $\text{Ratio}_{h_0^\alpha, h_1}(V) = \frac{\pi_{h_1}(V)\pi_{h_0^\alpha}(V)}{\varphi^{KT(h_0^\alpha, h_1) - KT(h_1, V)}} \to 1$. Because there are $m$ variables and $m$ equations, as $\varphi \to 1$ the solution to $\text{LP}_{h_0^\alpha}$ converges to $\hat{1}$. Therefore, there exists $\epsilon > 0$ such that for all $\varphi > 1 - \epsilon$, the linear systems $\{\text{LP}_{h_0^\alpha} : s \leq m - 1, h_1 \in H_1\}$ all have strictly positive solutions. Let $\{x[h_0^\alpha, s_\alpha]|V \in T_{s_\alpha}\}$ denote a solution to $\text{LP}_{h_0^\alpha}$. For any $h_0^\alpha \in T_{m-2}$, we let $\Lambda_{h_0^\alpha}^{h_1}(h_0^\alpha) = x[h_0^\alpha, s_\alpha] / \sum_{h_0^\alpha \in T_{m-2}} x[h_0^\alpha, s_\alpha]$.

**Step 2.** To simplify notation we let $L_{\alpha} = L_{\alpha, \Lambda_{h_0^\alpha}^{h_1}}$ denote the likelihood ratio test and let $\text{Ratio} = \text{Ratio}_{\Lambda_{h_0^\alpha}^{h_1}, h_1}$ denote the likelihood ratio function w.r.t. distribution $\Lambda_{h_0^\alpha}^{h_1}$ for $H_0$ vs. $h_1$. To prove $L_{\alpha} = \hat{f}_{\alpha, f}$, we first prove that for any $V \in \mathcal{L}(A)$ where $a$ is not ranked at the bottom position, $\text{Ratio}(V) > \text{Ratio}(\text{Down}_a(V))$, where we recall that
It follows from the strict inequality that
\[ KT(V) = \frac{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \pi_{h_0^*} (\text{Down}_a(V))}{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \pi_{h_0^*} (V)} = \frac{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \varphi^{KT(h_0^* \text{Down}_a(V))}}{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \varphi^{KT(h_0^* V)}} > \frac{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \varphi^{KT(h_0^* V)} \cdot \varphi^{KT(V \text{Down}_a(V))}}{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \varphi^{KT(h_0^* V)}} = \varphi = \pi_{h_1}(\text{Down}_a(V)) / \pi_{h_1}(V)
\]

The strict inequality holds because of (1) triangle inequality for Kentall-Tau distance, and (2) for any ranking $V$ where the top-ranked alternative in $h_0^*$ is ranked right below $a$, we have $KT(h_0^*,V) + KT(V,\text{Down}_a(V)) > KT(h_0^*, \text{Down}_a(V))$, and (3) for all $h_0^* \in T_{m-2}$, $\Lambda_{h_0^*} (h_0^*) > 0$.

It follows from the strict inequality that
\[ \text{Ratio}(V) = \frac{\pi_{h_1}(V)}{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \pi_{h_0^*} (V)} > \frac{\pi_{h_1}(\text{Down}_a(V))}{\sum_{h_0^* \in T_{m-2}} \Lambda_{h_0^*} (h_0^*) \cdot \pi_{h_0^*} (\text{Down}_a(V))} = \text{Ratio}(\text{Down}_a(V))
\]

Moreover, for any $V, V' \in T_{s_a}$ we have $\text{Ratio}(V) = \text{Ratio}(V')$ by verifying $\text{LP}_{s_a}^{h_1}$. Therefore, for any $V \in T_i$ with $i < s_a$, we can move up the position of $a$ one by one until we reach the $(m-s_a)$-th position. Let $V^* \in T_{s_a}$ denote this ranking. It follows that $\text{Ratio}(V) < \text{Ratio}(V^*)$. Similarly for any $V' \in T_i$ with $i > s_a$ we have $\text{Ratio}(V') > \text{Ratio}(V^*)$ for any $V^* \in T_{s_a}$. This means that for any $V$ where $a$ is ranked above the $(m-s_a)$-th position, we have $\text{LR}_{s_a}(V) = 1$; for any $V$ where $a$ is ranked below the $(m-s_a)$-th position, we have $\text{LR}_{s_a}(V) = 0$; for any $V$ where $a$ is ranked at the $(m-s_a)$-th position, we have that $\text{LR}_{s_a}(V)$ is the same and is between 0 and 1. It follows that $\text{LR}_{s_a} = \tilde{f}_{a,a}$.

**Step 3.** Due to the symmetry of $f_{a,a}$ among alternatives in $\mathcal{A} - \{a\}$, for any $i \leq m - 2$ and any $h_0, h_0' \in T_i$, we have $\text{Size}(f_{a,a}, h_0) = \text{Size}(f_{a,a}, h_0')$. Therefore, condition (i) in Lemma 2 is satisfied. Choose arbitrary $h_0 \in T_i$ denote the ranking obtained from $h_0$ by moving $a$ down for one position. To verify condition (ii) in Lemma 2, it suffices to prove that for any $i \leq m - 3$ and any $K \in \mathbb{N}$, we have
\[ \pi_{h_0}^{m-2}(\{V : \text{Borda}_a(V) \geq K\}) \geq \pi_{h_0}^{m-2}(\{V : \text{Borda}_a(V) \geq K\}) \]

We will prove a slightly stronger lemma.

**Lemma 8** Under Mallows’ model, for any $m$, any $\varphi$, any $W \in \mathcal{L}(\mathcal{A})$, any $b, c \in \mathcal{A}$ such that $b \succ_W c$, and any $K$, we have $\pi_W(\{V : \text{Borda}_a(V) \geq K\}) \geq \pi_W(\{V : \text{Borda}_c(V) \geq K\})$.

**Proof:** The proof is similar to the proof of Lemma 5. It suffices to prove the lemma for the case where $b$ and $c$ are adjacent in $W$. Let $\mathcal{P} = \{V \in \mathcal{L}(\mathcal{A}) : \text{Borda}_a(V) \geq K\}$ and $\mathcal{P}' = \{V \in \mathcal{L}(\mathcal{A}) : \text{Borda}_c(V) \geq K\}$. It follows that $\mathcal{P} \cap \mathcal{P}'$ is the set of rankings where both $b$ and $c$ are ranked within top $m - K$ positions; $\mathcal{P} - \mathcal{P}'$ is the set of rankings where $b$ is ranked within top $m - K$ positions but $c$ is not; and $\mathcal{P}' - \mathcal{P}$ is the set of rankings where $c$ is ranked within top $m - K$ positions but $b$ is not. We let $\mathcal{M}$ be a permutation that switches $b$ and $c$. It is not hard to check that $\mathcal{M}$ is a bijection between $(\mathcal{P} - \mathcal{P}')$ and $(\mathcal{P}' - \mathcal{P})$, and because $b$ and $c$ are adjacent in $W$, for any $V \in \mathcal{P}$, we have $\text{KT}(\mathcal{M}(V), W) = \text{KT}(V, W) + 1$, which means that $\pi_W(V) = \pi(\mathcal{M}(V))/\varphi$. Therefore, we have
\[ \pi_W(\{ V : \text{Borda}_0(V) \geq K \}) - \pi_W(\{ V : \text{Borda}_1(V) \geq K \}) \\
= \pi_W(P) - \pi_W(P') = \pi_W(P - P') = (1 - \varphi) - \pi_W(M(P - P')) \]

This proves the lemma.

Let \( W \) be an arbitrary ranking and let \( M_i \) denote a permutation such that \( M_i(h_0^*) = W \). We have \( \pi_{M_i}(\{ V : \text{Borda}((W_{M_i})) \geq K \}) \). We note that \( M_i(a) \) is the alternative that is ranked at the \((m - i)\)-th position in \( W \). Inequality (2) follows after applying Lemma 8. This means that condition (ii) in Lemma 2 is also satisfied. Therefore, by Lemma 2, \( \tilde{f}_{\alpha,a} \) is a level-\( \alpha \) most powerful test for \( H_0 \) vs. \( h_1 \). Since \( \tilde{f}_{\alpha,a} \) does not depend on \( h_1 \), it is a level-\( \alpha \) UMP test for \( H_0 \) vs. \( H_1 \).

**Lemma 6.** For any \( M_X \) and \( M_Y \), suppose \( \Lambda_X \) is a least favorable distribution for composite vs. simple test \((H_{0,X} \text{ vs. } x_1)\) under \( M_X \). Given \( y_1 \in \Theta_Y \), let \( \Lambda^* \) be the distribution over \( H_{0,X} \times \Theta_Y \) where for all \( x \in H_{0,X}, \Lambda^*(x, y_1) = \Lambda_X(x) \). Then \( \Lambda^* \) is a strict level-\( \alpha \) most powerful distribution for \( H_{0,X} \times \Theta_Y \) vs. \((x_1, y_1)\) under \( M_X \otimes M_Y \).

**Proof:** Let \( x_0^1, \ldots, x_0^K \in \Theta_X \) denote the support of \( \Lambda_X \). The theorem is proved by applying Lemma 2. For any \( 0 < \alpha < 1 \) and any \( P = (P_X, P_Y) \in \mathcal{S}_X \times \mathcal{S}_Y \), we have the following calculation. In this proof \( \text{Ratio} \) stands for \( \text{Ratio}_{\Lambda^*,(x_1,y_1)} \) and \( \text{LR}_\alpha \) stands for \( \text{LR}_{\alpha,\Lambda^*,(x_1,y_1)} \).

\[
\text{Ratio}(P_X, P_Y) = \sum_{k=1}^{K} \Lambda^*(x_0^k, y_1) \pi_{x_0^k, y_1}(P) / \sum_{k=1}^{K} \Lambda^*(x_0^k, y_1) \pi_{x_0^k}(P_X) \cdot \pi_{y_1}(P_Y) \\
= \sum_{k=1}^{K} \Lambda(x_0^k) \pi_{x_0^k}(P_X) / \sum_{k=1}^{K} \Lambda(x_0^k) \pi_{x_0^k}(P_X) = \text{Ratio}_{\Lambda^*,(x_1,y_1)}(P_X)
\]

It follows that for any pair of samples \((P_X, P_Y), (P'_X, P'_Y) \in \mathcal{S}_X \times \mathcal{S}_Y \), \( \text{Ratio}(P_X, P_Y) \geq \text{Ratio}(P'_X, P'_Y) \) if and only if \( \text{Ratio}_{\Lambda^*,(x_1,y_1)}(P_X) \geq \text{Ratio}_{\Lambda^*,(x_1,y_1)}(P'_X) \). This means that for any \((P_X, P_Y), \text{LR}_\alpha(P_X, P_Y) = \text{LR}_{\alpha,\Lambda^*,(x_1,y_1)}(P_X) \). Therefore, for any \( x_0 \in H_{0,X} \), we have

\[
\text{Size}(\text{LR}_\alpha, (x_0, y_1)) = \sum_{(P_X, P_Y) \in \mathcal{S}_X \times \mathcal{S}_Y} \pi_{x_0}(P_X) \pi_{y_1}(P_Y) \text{LR}_\alpha(P_X, P_Y) \\
= \sum_{(P_X, P_Y) \in \mathcal{S}_X \times \mathcal{S}_Y} \pi_{x_0}(P_X) \pi_{y_1}(P_Y) \text{LR}_{\alpha,\Lambda^*,(x_1,y_1)}(P_X) \\
= \sum_{P_X \in \mathcal{S}_X} \pi_{x_0}(P_X) \text{LR}_{\alpha,\Lambda^*,(x_1,y_1)}(P_X) \\
= \text{Size}(\text{LR}_{\alpha,\Lambda^*,(x_1,y_1)}).
\]

Therefore, by Lemma 2, for any \((x_0^1, y_1) \in \text{Spt}(\Lambda^*) \), we have \( \text{Size}(\text{LR}_\alpha, (x_0^1, y_1)) = \text{Size}(\text{LR}_{\alpha,\Lambda^*,(x_1,y_1), x_0^1} = \alpha \) because \( x_0^1 \in \text{Spt}(\Lambda) \); for any \((x_0, y) \in H_{0,X} \times \Theta_Y \), we have \( \text{Size}(\text{LR}_\alpha, (x_0, y)) = \text{Size}(\text{LR}_{\alpha,\Lambda^*,(x_1,y_1), x_0} \leq \alpha \). This means that the two conditions in Lemma 2 are satisfied, which proves the theorem.

**Lemma 7.** For any model \( M_X \) and any \( t \in \mathbb{N} \), suppose \( \Lambda \) is a uniformly least favorable distribution for composite vs. simple test \((H_0 \text{ vs. } h_1)\) under \( M_X \). Then \( \text{Ext}(\Lambda, h_1, t) \) is a uniformly least favorable distribution for \( \text{Ext}(H_0, h_1, t) \) vs. \( h_1 \) in \( (M_X)^t \).
Proof: Again the proof is done by applying Lemma 2. We first prove a claim that characterizes samples whose likelihood ratio is no more than a given threshold. To this end, it is convenient to use the inverse of the likelihood ratio. To simplify notation, in this proof we let $\Lambda^* = \text{Ext}(\Lambda, h_1, t)$, \(l_0^* = \text{Ext}(H_0, h_1, t)\), let $LR_\alpha = LR_{\alpha, \Lambda^*, h_1}$, $\text{Ratio} = \text{Ratio}_{\Lambda^*, h_1}$.

Claim 1 For any $k_\alpha$ and any $\vec{x} \in S^t$, $\sum_{j=1}^t \text{Ratio}^{-1}_{\Lambda^*, h_1} (x_j) = t \cdot \text{Ratio}^{-1}(\vec{x})$.

Proof: we have $\text{Ratio}^{-1}(\vec{x}) = 1 / t \cdot \sum_{j=1}^t \sum_{h_0} \Lambda(h_0) \cdot \pi_{(h_0, [\vec{x}]_{-j})} (\vec{x})$

$\text{Ratio}^{-1}_{\Lambda^*, h_1} (x_j) = 1 / t \cdot \sum_{h_0} \Lambda(h_0) \cdot \pi_{(h_0, [\vec{x}]_{-j})} (\vec{x})$

The next lemma proves the following: For any $\vec{x} \in H_0^*$ and any $j \leq t$, suppose the $j$-th component is not in $\text{Spt}(\Lambda)$, if we change the $j$-th component to $h_0^* \in \text{Spt}(\Lambda)$, then the size of $LR_\alpha$ will increase. If we further change the $j$-th component to $h_1$, then the size of $LR_\alpha$ will further increase.

Lemma 9 For any $0 \leq \alpha \leq 1$, any $j \leq t$, any $\vec{x}_{-j} \in \Theta^{t-1}$, any $h_0 \in \Lambda_0$, and any $h_0^* \in \text{Spt}(\Lambda)$, we have $\text{Size}(LR_\alpha, (h_0, \vec{x}_{-j})) \leq \text{Size}(LR_\alpha, (h_0^*, \vec{x}_{-j})) \leq \text{Size}(LR_\alpha, (h_1, \vec{x}_{-j}))$.

Proof: For any $\vec{x}_{-j} \in \Theta^{t-1}$, we have

\[ \text{Size}(LR_\alpha, (h_0, \vec{x}_{-j})) = \pi(h_0, \vec{x}_{-j}) \{ \vec{x} \in S^t : \text{Ratio}(\vec{x}) > k_\alpha^* \} \]

$+ \gamma_\alpha \pi(h_0, \vec{x}_{-j}) \{ \vec{x} \in S^t : \text{Ratio}(\vec{x}) = k_\alpha^* \}$

For any $\vec{x}$, we let $\text{Sum}(\vec{x}) = \sum_{l=1}^t \text{Ratio}^{-1}_{\Lambda^*, h_1} (x_l)$ and for any $j \leq t$, we let $\text{Sum}(\vec{x}_{-j}) = \sum_{l \neq j} \text{Ratio}^{-1}_{\Lambda^*, h_1}(x_l)$. By Claim 1, we have

\[ \pi(h_0, \vec{x}_{-j}) \{ \vec{x} \in S^t : \text{Ratio}(\vec{x}) > k_\alpha^* \} \]

$= \pi(h_0, \vec{x}_{-j}) \{ \vec{x} \in S^t : \text{Sum}(\vec{x}) < t/k_\alpha^* \}$

$= \pi(h_0, \vec{x}_{-j}) \{ \vec{x} \in S^t : \text{Sum}(\vec{x}_{-j}) + \text{Ratio}^{-1}_{\Lambda^*, h_1} (x_j) < t/k_\alpha^* \}$

$= \int_0^{t/k_\alpha^*} \sum_{\vec{x}_{-j} \in S^{t-1} : \text{Sum}(\vec{x}_{-j}) = p} \pi_{\Lambda^*, h_1} (x_j) \text{dp}

\[ \sum_{\vec{x}_{-j} \in S^{t-1} : \text{Sum}(\vec{x}_{-j}) = p} \text{dp} \]

$= \int_0^{t/k_\alpha^*} \pi_{\Lambda^*, h_1} (x_j) \text{dp} \]

$\cdot \pi_{h_0} \{ \vec{x} : \text{Ratio}_{\Lambda^*, h_1} (x_j) < t/k_\alpha^* - p \}

= \int_0^{t/k_\alpha^*} \pi_{h_0} \{ \vec{x} : \text{Ratio}_{\Lambda^*, h_1} (x_j) < t/k_\alpha^* - p \}

= \int_0^{t/k_\alpha^*} \pi_{h_0} \{ \vec{x} : \text{Ratio}_{\Lambda^*, h_1} (x_j) < t/k_\alpha^* - p \}

= \int_0^{t/k_\alpha^*} Q(\vec{x}_{-j}, p) \cdot \pi_{h_0} \{ \vec{x} : \text{Ratio}_{\Lambda^*, h_1} (x_j) < t/k_\alpha^* - p \}

= \int_0^{t/k_\alpha^*} Q(\vec{x}_{-j}, p) \cdot \text{Size}(LR_\alpha, (h_1, \vec{x}_{-j}))

where $Q(\vec{x}_{-j}, p) = \pi_{\Lambda^*, h_1} (x_j) \{ \vec{x} \in S^{t-1} : \text{Sum}(\vec{x}_{-j}) = p \}$. Given $p$ and $\gamma_\alpha$, let $\alpha'$ denote the size of the likelihood ratio test $LR_{\alpha', \Lambda^*, h_1}$, where the threshold $k_{\alpha'}$ is $1/(t/k_\alpha^* - p)$ and $\gamma_{\alpha'} = \gamma_\alpha$. We have

\[ \text{Size}(LR_\alpha, (h_0, \vec{x}_{-j})) = \int_0^{t/k_\alpha^*} Q(\vec{x}_{-j}, p) \cdot \text{Size}(LR_{\alpha', \Lambda^*, h_1}, h_0) \text{dp} \]

We note that in Equation (3), $\alpha'$ is a function of $t$, $p$, $k_\alpha^*$, and $\gamma_\alpha$. Because $\Lambda$ is a uniformly least favorable distribution, it follows from Lemma 2 that for any $h_0^* \in \text{Spt}(\Lambda)$ and any $h_0 \in (H_0 - \text{Spt}(\Lambda))$, we have

\[ \text{Size}(LR_{\alpha', \Lambda^*, h_1}, h_0) \leq \alpha' \leq \text{Size}(LR_{\alpha', \Lambda^*, h_1}, h_0^*) \]
Then by Equation (3), for any $h_0 \in (H_0 - \text{Spt}(\Lambda))$ and any $h_0^* \in \text{Spt}(\Lambda)$, we have

\[
\text{Size}(\text{LR}_\alpha, (h_0, \bar{z}_{-j})) = \int_0^{t/k_\alpha^*} Q(\bar{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha^*, \Lambda, h_1}, h_0^*)dp
\]

\[
\leq \int_0^{t/k_\alpha^*} Q(\bar{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha^*, \Lambda, h_1}, h_0^*)dp
\]

\[
= \text{Size}(\text{LR}_\alpha, (h_0^*, \bar{z}_{-j}))
\]

To prove the last inequality in the lemma, we prove a claim that holds for any least favorable distribution and the corresponding likelihood ratio test. The $\text{Size}(\cdot)$ function in the claim is extended to $h_1 \in H_1$ in the natural way.

**Claim 2** For any model, any composite vs. simple test ($H_0$ vs. $h_1$), suppose $\Lambda$ is a level-$\eta$ least favorable distribution. Then we have $\text{Size}(\text{LR}_\eta, h_1) \geq \eta = \text{Size}(\text{LR}_{\eta^*, h_0^*})$.

**Proof:** For the sake of contradiction suppose this is not true, that is, for any $h_0^* \in \text{Spt}(\Lambda)$ we have $\text{Size}(\text{LR}_\eta, h_1) < \eta = \text{Size}(\text{LR}_\eta, h_0^*)$. It follows that $k_\eta \leq 1$, otherwise we have

\[
\text{Size}(\text{LR}_\eta, h_1) = \sum_{P \in S: \text{Ratio}(P) > k_\eta} \pi_{h_1}(P) + \gamma_\eta \sum_{P \in S: \text{Ratio}(P) = k_\eta} \pi_{h_1}(P)
\]

\[
\geq \sum_{P \in S: \text{Ratio}(P) > k_\eta} \pi_\Lambda(P) \cdot k_\eta + \gamma_\eta \sum_{P \in S: \text{Ratio}(P) = k_\eta} \pi_\Lambda(P) \cdot k_\eta
\]

\[
> \sum_{P \in S: \text{Ratio}(P) > k_\eta} \pi_\Lambda(P) + \gamma_\eta \sum_{P \in S: \text{Ratio}(P) = k_\eta} \pi_\Lambda(P) = \eta,
\]

which is a contradiction. Therefore, we have

\[
1 = \text{Size}(\text{LR}_\eta, h_1) + \sum_{P \in S: \text{Ratio}(P) < k_\eta} \pi_{h_1}(P) + (1 - \gamma_\eta) \sum_{P \in S: \text{Ratio}(P) = k_\eta} \pi_{h_1}(P)
\]

\[
< \eta + \sum_{P \in S: \text{Ratio}(P) < k_\eta} \pi_\Lambda(P) \cdot k_\eta + (1 - \gamma_\eta) \sum_{P \in S: \text{Ratio}(P) = k_\eta} \pi_\Lambda(P) \cdot k_\eta
\]

\[
\leq \eta + k_\eta(1 - \text{Size}(\text{LR}_\eta, h_0^*)) \leq 1,
\]

which is a contradiction. \qed

Applying Claim 2 to $\text{LR}_{\alpha^*, \Lambda, h_1}$, we have

\[
\text{Size}(\text{LR}_\alpha, (h_0^*, \bar{z}_{-j})) = \int_0^{t/k_\alpha^*} Q(\bar{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha^*, \Lambda, h_1}, h_0^*)dp
\]

\[
\leq \int_0^{t/k_\alpha^*} Q(\bar{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha^*, \Lambda, h_1}, h_1)dp
\]

\[
= \text{Size}(\text{LR}_\alpha, (h_1, \bar{z}_{-j}))
\]

---

4We recall that $h_0^\Lambda$ is the combined $H_0$ by $\Lambda$. 

At the end of the process we obtain $j$ such that Borda $\bar{M}$.

We now define a test $\bar{\text{V}}$. Condition (ii) in Lemma 2 is verified by recursively applying Lemma 9. Given any $(i)$ in Lemma 2.

This finishes the proof of Lemma 9. □

Condition (ii) in Lemma 2 is verified by recursively applying Lemma 9. Given any $\bar{h}_0 \in H_0^\ast - \text{Spt}(\Lambda^*)$, there must exist $j \leq t$ such that $[\bar{h}_0]_j \neq h_1$. We then change $[\bar{h}_0]_j$ to an arbitrary $h_0^* \in \text{Spt}(\Lambda)$, then change the other components of $\bar{h}_0$ to $h_1$ one by one. Each time we make the change the size of LR does not decrease according to Lemma 9. At the end of the process we obtain $(h_0^*, [\bar{h}_1]_j) \in \text{Spt}(\Lambda^*)$, at which the size of LR is $\alpha$. The theorem follows after applying Lemma 2.

We now define a test $\bar{f}_{\alpha,a}$ for $H_0 = (\mathcal{L}(A) - H_1) \text{ vs. } H_1 = L_{a>\text{others}}$ and prove that if a UMP test exists, then $\bar{f}_{\alpha,a}$ also be a UMP test. For any $V \in \mathcal{L}(A)$ and any alternative $a \in A$, we let Borda$_a(V)$ denote the Borda score of $a$ in $V$. That is, Borda$_a(V)$ is the number of alternatives that are ranked below $a$ in $V$. For any $V \in \mathcal{L}(A)$, we let

$$\bar{f}_{\alpha,a}(V) = \begin{cases} 1 & \text{if } \text{Borda}_a(V) > K_\alpha \\ 0 & \text{if } \text{Borda}_a(V) < K_\alpha, \text{ where } K_\alpha \text{ and } \Gamma_\alpha \text{ are chosen so that the size of } \bar{f}_{\alpha,a} \text{ is } \alpha. \end{cases}$$

$\bar{f}_{\alpha,a}$ calculates the Borda score of $a$ in the input profile, and if it is larger than a threshold $K_\alpha$, then $H_0$ is rejected. It is not hard to see that $f_{\alpha,a}$ equals to $\bar{f}_{\alpha,a}$ with a possibly different level $\alpha'$ (defined in Theorem 3).

\textbf{Lemma 10} If there exists a level-$\alpha$ UMP test for $H_0 = (\mathcal{L}(A) - H_1) \text{ vs. } H_1 = L_{a>\text{others}}$, then $\bar{f}_{\alpha,a}$ is also a level-$\alpha$ UMP test.

\textbf{Proof:} Let $f_\alpha$ denote a level-$\alpha$ UMP test. For any permutation $M$ over $A - \{a\}$, we let $M(f_\alpha)$ denote the test such that for any $V \in \mathcal{L}(A)$, $M(f_\alpha)(V) = f_\alpha(M(V))$. Because the Kendall-Tau distance is invariant to permutations, we have that for any $h_0 \in H_0$, Size($f_\alpha, h_0$) = Size($M(f_\alpha), M(h_0)$), and for any $h_1 \in H_1$, Power($f_\alpha, h_1$) = Power($M(f_\alpha), M(h_1)$). Therefore $\text{Size}(M(f_\alpha)) = \alpha$. Also because the multi-set of $\{\text{Power}(f_\alpha, h_1) : h_1 \in H_1\}$ is the same as the multi-set of $\{\text{Power}(M(f_\alpha), h_1) : h_1 \in H_1\}$, for all $h_1 \in H_1$, we must have $\text{Power}(f_\alpha, h_1) = \text{Power}(M(f_\alpha), h_1)$, otherwise there exists $h_1 \in H_1$ such that $\text{Power}(f_\alpha, h_1) < \text{Power}(M(f_\alpha), h_1)$, which contradicts the assumption that $f_\alpha$ is UMP.

It follows that for any permutation $M$ over $A - \{a\}$, $M(f_\alpha)$ is also UMP. Therefore, $\bar{f}_\alpha = \frac{1}{(m-1)!} \sum M(f_\alpha)$ is also UMP. We note that for any $V, V'$ where $a$ has the same Borda score, there exists a permutation $M$ over $A - \{a\}$ so that $M(V) = V'$. This means that $\bar{f}_\alpha(V) = \bar{f}_\alpha(V')$.

We now prove that $\bar{f}_\alpha$ must be $\bar{f}_{\alpha,a}$ as in the statement of the Lemma. More precisely, we will prove that for any $V, V'$ such that Borda$_a(V) > \text{Borda}_a(V')$, if $\bar{f}_\alpha(V') > 0$ then $\bar{f}_\alpha(V) = 1$. Suppose for the sake of contradiction that this is not true, and there exist $V, V'$ such that $s_1 = \text{Borda}_a(V) > \text{Borda}_a(V') = s_2$, $\bar{f}_\alpha(V') > 0$, and $\bar{f}_\alpha(V) < 1$. For any $s \leq m - 1$, we let $T_s$ denote the set of rankings where the Borda score of $a$ is $s$. That is, $T_s = \{V \in \mathcal{L}(A) : \text{Borda}_a(V) = s\}$. We will prove that for any $s_1 > s_2$, $T_{s_1}$ as a whole is more “cost effective” than $T_{s_2}$ as a whole for any $h_0 \in H_0$ against any $h_1 \in H_1$. More precisely, we will prove that $\text{Ratio}_{h_0,h_1}(T_{s_1}) > \text{Ratio}_{h_0,h_1}(T_{s_2})$.

For any $s \leq m - 2$ and any $h_0 \in T_s$, let $h_1$ denote the ranking in $T_{m-1} = H_1$ that is obtained from $\theta$ by raising $a$ to the top position. For any $V_{s_1} \in T_{s_1}$, we let Down$_{a}^{s_1-s_2}(V_{s_1}) \in T_{s_2}$ denote the ranking that is obtained from $V_{s_1}$ by
moving $a$ down for $s_1 - s_2$ positions, that is, from the $(m - s_1)$-th position to the $(m - s_2)$-th position. We have

$$\frac{\pi_{h_0}(T_{s_2})}{\pi_{h_0}(T_{s_1})} = \frac{\sum_{V \in T_{s_2}} \pi_{h_0}(V)}{\sum_{V \in T_{s_1}} \pi_{h_0}(V)} = \frac{\sum_{V \in T_{s_1}} \pi_{h_0}(\text{Down}_{s_1}^{s_1-s_2}(V))}{\sum_{V \in T_{s_1}} \pi_{h_0}(V)} = \frac{\sum_{V \in T_{s_1}} \varphi^{\text{KT}(h_0,\text{Down}_{s_1}^{s_1-s_2}(V))}}{\sum_{V \in T_{s_1}} \varphi^{\text{KT}(h_0,V)}} > \frac{\sum_{V \in T_{s_1}} \varphi^{\text{KT}(h_0,V)} \cdot \varphi^{\text{KT}(V,\text{Down}_{s_1}^{s_1-s_2}(V))}}{\sum_{V \in T_{s_1}} \varphi^{\text{KT}(h_0,V)}} = \varphi^{s_1-s_2} = \frac{\pi_{h_1}(T_{s_2})}{\pi_{h_1}(T_{s_1})}$$

The inequality is due to triangle inequality for Kendall-Tau distance. It is strict because for any $V \in T_{s_1}$ where the top-ranked alternative in $h_0$ is ranked between the $(m - s_1)$-th and $(m - s_2)$-th position, $\text{KT}(h_0,\text{Down}_{s_1}^{s_1-s_2}(V)) < \text{KT}(h_0,V) + \text{KT}(V,\text{Down}_{s_1}^{s_1-s_2}(V))$. Therefore, $\frac{\pi_{h_0}(T_{s_2})}{\pi_{h_0}(T_{s_1})} > \frac{\pi_{h_1}(T_{s_2})}{\pi_{h_1}(T_{s_1})}$, which means that $\text{Ratio}_{h_0,h_1}(T_{s_1}) = \frac{\pi_{h_1}(T_{s_1})}{\pi_{h_0}(T_{s_1})} \pi_{h_0}(T_{s_1}) = \pi_{h_1}(T_{s_1})$.

Therefore, we can find sufficiently small $\epsilon, \delta > 0$, and replace $\epsilon T_{s_2}$ by $\delta T_{s_1}$, without changing the size. This will increase the power of $\bar{f}_{\alpha}$ because $T_{s_1}$ is strictly more cost effective than $T_{s_2}$, which contradicts the assumption that $\bar{f}_{\alpha}$ is a UMP test. Therefore, $\bar{f}_{\alpha} = \bar{f}_{\alpha,a}$, which proves the lemma. $\square$