Optimal Statistical Hypothesis Testing for Social Choice

LIRONG XIA, RENSSELAER POLYTECHNIC INSTITUTE, XIAL@CS.RPI.EDU

We address the following question: "*What are the optimal statistical hypothesis testing 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* for testing whether a given alternative is the winner under Mallows' model and Condorcet's model, respectively. We propose natural procedures to choose a winner by combining the results of multiple tests.

1 INTRODUCTION

Suppose a group of friends want to choose a restaurant for dinner. Each person uses a full ranking over the restaurants to represent his or her preferences. Assuming that their tastes are similar and 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 give it a high rank. How can they decide the winner?

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, which is the MLE of the model, converges to the ground truth as the number of voters goes to infinity.

However, the Jury Theorem does not identify the *optimal* rule to make a decision w.r.t. the unknown ground truth, especially for three or more alternatives. From a statistical point of view, while it is straightforward to think that an optimal rule should have the best chance to reveal the ground truth, defining the measure for such chance is highly nontrivial and controversial. If we use likelihood of a parameter as the measure, then we may pursue the *maximum likelihood estimation (MLE)* 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 another 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, 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.

While many generic hypothesis testing methods, such as the *generalized likelihood ratio tests* [Hoeffding, 1965, Zeitouni et al., 1992], can be applied, they are often not optimal. It is still unclear what are the optimal statistical hypothesis testing methods for social choice.

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

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 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-α 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 α.

	Non-winner ($H_0 = \{a \text{ wins}\}$)	Winner ($H_1 = \{a \text{ wins}\}$)
Mallows	Y&N (Thm. 3, 4)	Y&N (Thm. 5,6,7)
Condorcet	Y&N (Thm. 10, 11)	Y (Thm. 12)

Table 1. Existence of UMP tests for winner determination.

In Table 1, "Y" in Condorcet-Winner means that for any $0 < \alpha < 1$, there exists a level-α UMP winner test for $H_0 = \bar{H}_1$ vs. H_1, where H_1 consists of rankings where a given alternative a is ranked at the top. "Y&N" means that for some α, no level-α UMP test exists when the other hypothesis consists of all other parameters; but a UMP test exists for all levels for some natural cases of the other hypothesis. For example, "Y&N" in Mallows-Non-winner in Table 1 means that for some α, no level-α UMP test exists for H_0 vs. $H_1 = \bar{H}_0$, 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-α UMP test exists for all $0 < \alpha < 1$. In fact, Theorem 4 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 4) and under Condorcet's model (Theorem 11). Technically, to obtain the characterizations, we leverage the theory of *uniformly least favorable distributions* to finite models and randomized tests (Theorem 1, 2, 8, 9). These theorems generalize the key theorems by Reinhardt [1961] that only hold for continuous parameter space, and they might be of independent interest.

How to apply our results? The most positive result is the existence of UMP winner tests in the Condorcet-Winner column. These tests 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". We also propose two procedures in Section 6 to choose a winner as the social choice by combining results of multiple tests. Interestingly, one procedure corresponds to the Borda rule. This offers a new statistical justification of Borda.

Related Work and Discussions. Most previous work in statistical approaches in social choice focused on extending the Condorcet Jury Theorem and proving asymptotic results. See the survey by Gerlinga et al. [2005] and a more recent one by Nitzan and Paroush [2017]. In terms of decision rules, previous work focused on using commonly-studied voting rules designed for elections [Caragiannis et al., 2016, Conitzer and Sandholm, 2005], maximum likelihood estimators [Conitzer and Sandholm, 2005, Xia and Conitzer, 2011], or Bayesian estimators [Azari Soufiani et al., 2014, Elkind

[1] This is a non-winner test because H_0 is often used as the devil's advocate, and the goal is often to reject H_0.

and Shah, 2014, Pivato, 2013, Procaccia et al., 2012, Xia, 2016, Young, 1988]. We are not aware of a previous work that studies UMP tests under popular models for rank data.

In this paper, we focus on two commonly studied statistical models for rank data, i.e. Mallows' model and Condorcet's model. The goal of this paper is to setup a first framework for using hypothesis testing for social choice. How to extend the results to other models and how to compute the UMP tests are promising directions for future research.

Compared to previous MLE and Bayesian approaches to social choice, optimal rules characterized in this paper are more robust because it offers the best worst-case guarantee for an adversarial 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 are different. And we note that the goal of this paper is to understand optimal rules for robustness decision-making in the principled and widely-applied framework of hypothesis testing.

Marden [1995] studied various properties of distance-based models, especially Mallows' model, such as the probability of certain pairwise comparisons and the expected rank of a given alternative. However, as our analyses are largely based on first-order stochastic dominance (Theorem 1), Marden's results do not help and we need to develop new ones (Lemma 3 and 4). These results provide new insights into Mallows' model, which might be of independent interest.

2 PRELIMINARIES

Let $\mathcal{A} = \{a_1, \dots, a_m\}$ denote a set of m *alternatives* and let $\mathcal{L}(\mathcal{A})$ denote the set of all linear orders over $\mathcal{A}$. Let n denote the number of agents. Each agent's preferences are represented by a linear order in $\mathcal{L}(\mathcal{A})$. We often use $V = [a \succ b \succ \cdots]$ to denote a ranking, and write $a \succ_V b$ if a is preferred to b in V. The collection P_n of all agents' votes is called an *(n-)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* of P, denoted by $\text{WMG}(P)$, is a directed weighted graph where the weight $w_P(a \succ b)$ on any edge $a \to b$ is $w_P(a \succ b) = P(a \succ b) - P(b \succ a)$. Clearly $w_P(a \succ b) = -w_P(b \succ a)$.

Statistical Models for Rank Data. A statistical model $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$ has three parts: the *sample space* $\mathcal{S}$, which is composed of all possible data; the *parameter space* Θ; and the probability distributions $\vec{\pi} = \{\pi_\theta : \theta \in \Theta\}$. If both $\mathcal{S}$ and Θ contain finite elements, then we call $\mathcal{M}$ a *finite model*.

For any pair of linear orders V, W in $\mathcal{L}(\mathcal{A})$, let $\text{KT}(V, W)$ denote the *Kendall-tau distance*, which is the total number of pairwise disagreements between V and W.

Example 1. For $m = 3$ and $\mathcal{A} = \{1, 2, 3\}$, the Kendall-Tau distances among the six rankings are visualized in Figure 1 (a), where 123 represents $1 \succ 2 \succ 3$, and the Kendall-Tau distance between V and W is the closest distance on the circle. For example, $\text{KT}(123, 312) = 2$ and $\text{KT}(213, 231) = 3$.

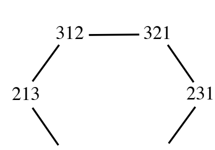

Fig. 1. Kentall-Dau distance for three alternatives.

For $m \geq 4$, the structure of Kendall-Tau distance is much more complicated and cannot be visualized as a circle. ■

DEFINITION 1 (MALLOWS' MODEL [MALLOWS, 1957]). *Given* $0 < \varphi < 1$, Mallows' model *is denoted by* $\mathcal{M}^{Ma} = (\mathcal{L}(\mathcal{A})^n, \mathcal{L}(\mathcal{A}), \vec{\pi})$*, where n linear orders are i.i.d. generated, the parameter space is* $\mathcal{L}(\mathcal{A})$ *and for any* $V, W \in \mathcal{L}(\mathcal{A})$, $\pi_W(V) = \frac{1}{Z}\varphi^{KT(V,W)}$*, where Z is the normalization factor.*

Let $\mathcal{B}(\mathcal{A})$ denote the set of all irreflexive, antisymmetric, and total *binary relations* over $\mathcal{A}$. We have $\mathcal{L}(\mathcal{A}) \subseteq \mathcal{B}(\mathcal{A})$ and the Kendall-tau distance can be easily extended to $\mathcal{B}(\mathcal{A})$ by counting the number of pairwise disagreements.

DEFINITION 2 (CONDORCET'S MODEL [CONDORCET, 1785, YOUNG, 1988]). *Given* $0 < \varphi < 1$, Condorcet's model *is denoted by* $\mathcal{M}^{Co} = (\mathcal{L}(\mathcal{A})^n, \mathcal{B}(\mathcal{A}), \vec{\pi})$*, where the parameter space is* $\mathcal{B}(\mathcal{A})$ *and for any* $W \in \mathcal{B}(\mathcal{A})$ *and* $V \in \mathcal{L}(\mathcal{A})$, $\pi_W(V) = \frac{1}{Z}\varphi^{KT(V,W)}$*, where Z is the normalization factor.*

Condorcet's model differs from Mallows' model mainly in allowing ties in the ground truth.

Statistical Hypothesis Testing: The Neyman-Pearson Framework. Given a statistical model $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$, 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 $P \in \mathcal{S}$. To simplify notation, we let 0 denote retain and let 1 denote reject. A test is characterized by a (randomized) *critical function* $f : \mathcal{S} \to [0, 1]$ such that for any $P \in \mathcal{S}$, 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 α, then f is called a *level-*α 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 π_{h_1}. We would like a test f to have low size and high power, but often tradeoffs must be made.

Example 2. Let $\mathcal{M}^{\text{Ma}}$ denote a Mallows' model with $m = 3$ and $n = 1$. Let $\mathcal{A} = \{1, 2, 3\}$, $h_1 = 123$, and let H_0 denote the other rankings. Let f be a test where $f(123) = 1$, $f(213) = f(132) = 0.5$, and f outputs 0 for all other rankings. We have $\text{Size}(f) = \text{Size}(f, 213) = \text{Size}(f, 132) = (0.5+\varphi+0.5\varphi^2)/Z$, where Z is the normalization factor. $\text{Power}(f, 123) = (1 + \varphi)/Z$.

There is no deterministic test f' whose size is $\varphi^3/2$, because as soon as $f'(V) = 1$ for any ranking V, its size becomes at least φ^3. However, the size of the randomized test f^* that outputs $\varphi^2/2$ at 123 and outputs 0 at other rankings is $\varphi^3/2$. ■

There is a natural analogy between hypothesis testing and the following production problem.

The production problem. *Let* H_0 *denote a set of workers, let* $\mathcal{S}$ *denote the types of* **divisible** *items to be produced, and suppose that there is a single buyer* h_1*. For any* $P^i \in \mathcal{S}$*, any* $h_0 \in H_0$*, producing one unit of item* P^i *costs* $\pi_{h_0}(P^i)$ *hours of worker* h_0*. Suppose that workers have different skills and cannot be substituted, and their labor is the only cost for producing the items. The buyer only wants no more than one unit of each item, and the unit price for item* P^i *is* $\pi_{h_1}(P^i)$*. Each test* f *corresponds to a production plan, where* $f(P^i)$ *is the amount of item* P^i *that will be produced. Size*(f, h_0) *is the total amount of time spent by worker* h_0 *in production, and Size*(f) *is the maximum number of hours of any single worker. Power*(f) *is the total revenue.* ■

Given a statistical model $\mathcal{M}$, H_0, a single parameter $h_1 \notin H_0$, and $0 < \alpha < 1$, a *level-α most powerful test f_α* is a test with the highest power among all tests whose size is no more than α. Most powerful tests correspond to the production plan with the highest revenue given that no one works for more than α hours.

For finite H_0, a most powerful test always exists, and can be computed by the following linear program, where for each sample $P^i \in \mathcal{S}$ there is a variable x_i that represents $f_\alpha(P^i)$.

$$
\begin{aligned}
&\max \sum_{P_i \in \mathcal{S}} \pi_{h_1}(P^i) \cdot x_i && \text{(maximize the power)}\\
s.t.\quad &\forall h_0 \in H_0, \sum_{P^i \in \mathcal{S}} \pi_{h_0}(P^i) \cdot x_i \leq \alpha && \text{(size constraints)}\\
&\forall i, 0 \leq x_i \leq 1
\end{aligned}
$$

Most powerful tests 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-α test f_α that is most powerful for all $h_1 \in H_1$, then f is called a level-α *uniformly most powerful (UMP) test* for H_0 vs. H_1. UMP is a strong notion of optimality and a UMP test may not exists.

The notion of UMP test can be explained in the context of the production problem as follows. Now there are multiple potential buyers, one for each $h_1 \in H_1$, such that buyer h_1 is willing to pay $\pi_{h_1}(P^i)$ for each of P^i. However, only one buyer will show up and the production plan must be made before any buyer shows up. A UMP test corresponds to a robust production plan, such that no matter which buyer shows up, its revenue is always the highest among all production plans where no one works for more than α hours.

For simple H_0, the fundamental lemma of Neyman and Pearson characterizes the most powerful tests as *likelihood ratio tests*.

DEFINITION 3 (LIKELIHOOD RATIO TEST). *Given a model $\mathcal{M}$ and $0 < \alpha < 1$. For any $h_0, h_1 \in \Theta$ with $h_0 \neq h_1$ and any $P \in \mathcal{S}$, we let $Ratio_{h_0,h_1}(P) = \frac{\pi_{h_1}(P)}{\pi_{h_0}(P)}$ denote the* likelihood ratio *of P and let $LR_{\alpha,h_0,h_1}(P)$ defined below denote the* level-α likelihood ratio test*, both w.r.t. the simple vs. simple test (h_0 vs. h_1). For any $P \in \mathcal{S}$,*

$$
LR_{\alpha,h_0,h_1}(P) = \begin{cases} 1 & \text{if } Ratio_{h_0,h_1}(P) > k_\alpha \\ 0 & \text{if } Ratio_{h_0,h_1}(P) < k_\alpha \\ \gamma_\alpha & \text{if } Ratio_{h_0,h_1}(P) = k_\alpha \end{cases},
$$

where $k_\alpha \geq 0$ and γ_α are chosen such that $Size(LR_{\alpha,h_0,h_1}) = \alpha$.

Sometimes h_0 and h_1 in the subscript of Ratio_{h_0,h_1} and $\text{LR}_{\alpha,h_0,h_1}$ are omitted without causing confusion. We note that different k_α and γ_α may correspond to the same likelihood ratio test.

LEMMA 1 (THE NEYMAN-PEARSON LEMMA (SEE E.G. [LEHMANN AND ROMANO, 2008])). *For any simple vs. simple test (h_0 vs. h_1) and any $0 < \alpha < 1$, the likelihood ratio test LR_{α,h_0,h_1} is a level-α most powerful test. Moreover, any most powerful test must agree with LR_{α,h_0,h_1} except on $P \in \mathcal{S}$ with $Ratio_{h_0,h_1}(P) = k_\alpha$.*

Example 3. Given a Mallows' model. Let $H_0 = \{h_0\}$ and $H_1 = \{h_1\}$. For any n-profile P_n, we have $\text{Ratio}(P_n) = \frac{\varphi^{\text{KT}(P_n,h_1)}}{\varphi^{\text{KT}(P_n,h_0)}} = \varphi^{\text{KT}(P_n,h_1)-\text{KT}(P_n,h_0)}$. Therefore, it follows from the Neyman-Pearson lemma that for any $0 < \alpha < 1$, there exist K_α and Γ_α such that the following test f_α is a level-α most powerful test: for any n-profile P_n,

$$f_\alpha(P_n) = \begin{cases} 1 & \text{if } \mathrm{KT}(P_n, h_0) - \mathrm{KT}(P_n, h_1) > K_\alpha \\ 0 & \text{if } \mathrm{KT}(P_n, h_0) - \mathrm{KT}(P_n, h_1) < K_\alpha \\ \Gamma_\alpha & \text{if } \mathrm{KT}(P_n, h_0) - \mathrm{KT}(P_n, h_1) = K_\alpha \end{cases}.$$

■

The idea behind most powerful tests for composite H_0 is to use a distribution Λ over H_0 to compress H_0 into a "combined" parameter, denoted by h_0^Λ. Then, the likelihood ratio test for h_0^Λ vs. h_1 is a most powerful test for H_0 vs. h_1 when two additional conditions are satisfied.

DEFINITION 4. *For any $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$, any $H_0 \subseteq \Theta$, and any $h_1 \in (\Theta - H_0)$. Let Λ denote a distribution over H_0 and let h_0^Λ denote a new parameter whose distribution over $\mathcal{S}$ is the probabilistic mixture of $\{\pi_{h_0} : h_0 \in \Theta\}$ according to Λ. For any $0 < \alpha < 1$ and any $P \in \mathcal{S}$, let $Ratio_{\Lambda, h_1}(P) = \frac{\pi_{h_1}(P)}{\sum_{h_0 \in H_0} \Lambda(h_0)\pi_{h_0}(P)}$ and let $LR_{\alpha,\Lambda,h_1}(P)$ denote the likelihood ratio test for h_0^Λ vs. h_1.*

Sometimes Λ and h_1 in the subscripts of $\mathrm{Ratio}_{\Lambda,h_1}$ and $\mathrm{LR}_{\alpha,\Lambda,h_1}$ are omitted without causing confusion. The next lemma is a generalization of the Neyman-Pearson lemma that provides necessary and sufficient conditions for $\mathrm{LR}_{\alpha,\Lambda,h_1}(P)$ to be most powerful for H_0 vs. h_1. Let Support(Λ) denote the *support* of Λ, namely the elements in H_0 that are assigned non-zero probability by Λ.

LEMMA 2 (THEOREM 3.8.1 AND COROLLARY 3.8.1 IN [LEHMANN AND ROMANO, 2008]). *For composite vs. simple test (H_0 vs. h_1) and any distribution Λ over H_0, the likelihood ratio test LR_{α,Λ,h_1} is a level-α most powerful test if and only if the following two conditions are satisfied.*
(i) For any $h_0^* \in Support(\Lambda)$, we have $Size(LR_{\alpha,\Lambda,h_1}, h_0^*) = \alpha$.
(ii) For any $h_0 \in H_0$, we have $Size(LR_{\alpha,\Lambda,h_1}, h_0) \leq \alpha$.
Moreover, if there is no $P \in \mathcal{S}$ with $Ratio_{\Lambda,h_1}(P) = k_\alpha$, then LR_{α,Λ,h_1} is the unique level-α most powerful test.

The distribution Λ in Lemma 2 is called a *least favorable distribution*. If $|\mathrm{Support}(\Lambda)| = 1$, then Λ is called a *deterministic least favorable distribution*. If Λ 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 THEOREMS ON LEAST FAVORABLE DISTRIBUTIONS

We first present two general theorems on least favorable distributions that will be frequently used in this paper. For any model $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$, any composite vs. simple test (H_0 vs. h_1), any distribution Λ over H_0, and any $h_0 \in H_0$, we define a random variable $X_{h_0}^\Lambda : \mathcal{S} \to \mathbb{R}$ such that for any $P \in \mathcal{S}$, $\Pr(P) = \pi_{h_0}(P)$ and $X_{h_0}^\Lambda(P) = \log \mathrm{Ratio}_{\Lambda,h_1}(P)$. A random variable X *weakly first-order stochastically dominates* (*weakly dominates* for short) another random variable Y, if for any $p \in \mathbb{R}$, $\Pr(X \geq p) \geq \Pr(Y \geq p)$.

THEOREM 1. *Λ is a uniformly least favorable distribution for H_0 vs. h_1 if and only if for any $h_0^* \in Support(\Lambda)$ and any $h_0 \in H_0$, $X_{h_0^*}^\Lambda$ weakly dominates $X_{h_0}^\Lambda$.*

All missing proofs can be found in the appendix.

Example 4. Let $\mathcal{M}$ denote a Mallows' model with $m = 3$ and $n = 1$. Let $\mathcal{A} = \{1, 2, 3\}$, $h_1 = [1 \succ 2 \succ 3]$ and let Λ denote the uniform distribution over $\{[2 \succ 1 \succ 3], [1 \succ 3 \succ 2]\}$. We will apply Theorem 1 to prove that Λ is a uniformly least favorable distribution for $H_0 = (\mathcal{L}(\mathcal{A}) - \{[1 \succ 2 \succ 3]\})$ vs. $[1 \succ 2 \succ 3]$. The likelihood ratios of all rankings are summarized in Table 2 in the increasing order.

V	$3 \succ 2 \succ 1$	others	$1 \succ 2 \succ 3$
$\text{Ratio}_{\Lambda, 1\succ 2\succ 3}(V)$:	φ	$\frac{2\varphi}{1+\varphi^2}$	$\frac{1}{\varphi}$

Table 2. Likelihood ratios.

For any $h_1 \in H_0$, $X^{\Lambda}_{h_0}$ takes three values: $\log \frac{1}{\varphi}$, $\log \frac{2\varphi}{1+\varphi^2}$, and $\log \varphi$. The probabilities for the five random variables taking these three values are summarized in Table 3.

	$\log \varphi$	$\log \frac{2\varphi}{1+\varphi^2}$	$\log \frac{1}{\varphi}$
$X^{\Lambda}_{1\succ 3\succ 2}$ and $X^{\Lambda}_{2\succ 1\succ 3}$	$\frac{\varphi^2}{Z}$	$\frac{1+\varphi+\varphi^2+\varphi^3}{Z}$	$\frac{\varphi}{Z}$
$X^{\Lambda}_{2\succ 3\succ 1}$ and $X^{\Lambda}_{3\succ 1\succ 2}$	$\frac{\varphi}{Z}$	$\frac{1+\varphi+\varphi^2+\varphi^3}{Z}$	$\frac{\varphi^2}{Z}$
$X^{\Lambda}_{3\succ 2\succ 1}$	$\frac{1}{Z}$	$\frac{2(\varphi+\varphi^2)}{Z}$	$\frac{\varphi^3}{Z}$

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 $X^{\Lambda}_{1\succ 3\succ 2}$ and $X^{\Lambda}_{2\succ 1\succ 3}$ weakly dominate other random variables. By Theorem 1, Λ is a uniformly least favorable distribution. ■

Our second theorem states that if we can find a deterministic uniformly least favorable distribution, then it is also uniformly least favorable for the same statistical model with multiple i.i.d. samples.

THEOREM 2. *Suppose Λ is a deterministic uniformly least favorable distribution for composite vs. simple test (H_0 vs. h_1) under $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$. Then for any $n \in \mathbb{N}$, Λ is also a uniformly least favorable distribution for testing H_0 vs. h_1 under $\mathcal{M} = (\mathcal{S}^n, \Theta, \vec{\pi})$ with n i.i.d. samples.*

PROOF. Let $\text{Support}(\Lambda) = \{h_0^*\}$. 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 $X_{n,h_0}(P_n) = \log \text{Ratio}(P_n, h_0^*, h_1)$. It follows that $X_{n,h_0} = \underbrace{X_{h_0} + X_{h_0} + \cdots + X_{h_0}}_{n}$. By Theorem 1, for any $h_0 \in H_0$, $X_{h_0^*}$ 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^*} weakly dominates X_{n,h_0}. The theorem follows after applying Theorem 1. □

4 UMP TESTS FOR MALLOWS' MODEL

Given an alternative a, for Mallows' model we define $L_{a\succ\text{others}} = \{V \in \mathcal{L}(\mathcal{A}) : \forall b \in \mathcal{A}, a \succ_V b\}$. That is, a is ranked at the top of the ground truth parameter. We will focus on two classes of tests, one for testing whether a is not a winner and the other for testing whether a is a winner. Because H_0 is often chosen as a devil's advocate and the goal of testing is often to reject H_0, when setting $H_0 = L_{a\succ\text{others}}$ under Mallows' model, we are hoping to reject H_0, which means that a is not the winner. This idea leads to the following formal definition.

DEFINITION 5. *Given an alternative a, in a* non-winner *test of Mallows' model, we let $H_0 = L_{a\succ others}$; in a* winner *test, we let $H_1 = L_{a\succ others}$.*

We note that the decision-maker still need to specify both H_0 and H_1 in a winner or non-winner test and a level α. Then, the decision-maker performs a level-α test f on the data. Often this is done by computing a *test statistic* and computes its *p-value*.

For any profile P, any $B \subset \mathcal{A}$, and any $a \in (\mathcal{A} - B)$, 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 $\text{WMG}(P)$.

THEOREM 3. *Let $\mathcal{M}^{Ma}$ denote a Mallows' model with any $m \geq 2$, any n, and any φ. For any α, any $a \in \mathcal{A}$, and any $h_1 = [B \succ a \succ \text{others}]$ for non-empty B, let h_0^* denote the ranking that is obtained from h_1 by raising a to the top position. Then the deterministic distribution Λ at $\{h_0^*\}$ is a uniformly least favorable distribution for $H_0 = L_{a \succ others}$ vs. h_1, and $f_{\alpha,a,B}$ defined below is a level-α UMP test. For any n-profile P_n,*

$$f_{\alpha,a,B}(P_n) = \begin{cases} 1 & \text{if } w_{P_n}(B \succ a) > K_\alpha \\ 0 & \text{if } w_{P_n}(B \succ a) < K_\alpha \\ \Gamma_\alpha & \text{if } w_{P_n}(B \succ a) = K_\alpha \end{cases},$$

where K_α and Γ_α are chosen such that the size of $f_{\alpha,a,B}$ is α. Moreover, any most powerful test must agree with $f_{\alpha,a,B}(P_n)$ except on P_n with $w_{P_n}(B \succ a) = K_\alpha$.

Proof sketch: We have $h_0^* = [a \succ B \succ \text{other}]$. In this proof we let LR_α denote $\text{LR}_{\alpha,h_0^*,h_1}$ and let Ratio denote $\text{Ratio}_{h_0^*,h_1}$. We first prove the theorem for $n = 1$, and then extend it to any n by Theorem 2. By Theorem 1, it suffices to prove that for any $h_0 \in H_0$, $X^\Lambda_{h_0^*}$ weakly dominates $X^\Lambda_{h_0}$.

For any ranking V and any pair of alternatives a, b, we let $I(a \succ_V b) = 1$ if $a \succ_V b$, otherwise $I(a \succ_V b) = 0$. For any $P \in \mathcal{L}(\mathcal{A})$, we have

$$\begin{aligned}
&\log \text{Ratio}(P) = \log \varphi(\text{KT}(P,h_1) - \text{KT}(P,h_0)) \\
=& -\log \varphi \sum_{c \succ_P d} (I(d \succ_{h_1} c) - I(d \succ_{h_0} c)) \\
=& -\log \varphi \Big(\sum_{a \succ_P b : b \in B} (I(b \succ_{h_1} a) - I(b \succ_{h_0} a)) \\
&+ \sum_{b \succ_P a : b \in B} (I(a \succ_{h_1} b) - I(a \succ_{h_0} b)) \Big) \\
=& -\log \varphi \cdot (2 w_P(B \succ a) - |B|)
\end{aligned}$$

Then we prove that for any $K \in \mathbb{Z}$, $\pi_{h_0}(\{P : w_P(B \succ a) \geq K\}) \leq \pi_{h_0^*}(\{P : w_P(B \succ a) \geq K\})$.

Consequently, following Theorem 1, we prove that this means that Λ is a uniformly least favorable distribution for $n = 1$. Because Λ is deterministic, by Theorem 2, Λ 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_α is most powerful. It is not hard to verify that $\text{LR}_\alpha = f_{\alpha,a,B}$. Moreover, because Λ 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^* vs. h_1). By the Neyman-Pearson lemma (Lemma 1), f must agree with LR_α except on P_n such that $\text{Ratio}(P_n) = k_\alpha$, which corresponds to P_n with $w_{P_n}(B \succ a) = K_\alpha$. □

The next theorem characterizes all UMP non-winner tests ($H_0 = L_{a \succ \text{others}}$) under Mallows' model. For any $B \subset \mathcal{A}$ and $a \in (\mathcal{A} - B)$, we let $L_{B \succ a} \subseteq \mathcal{L}(\mathcal{A})$ 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 4. *Let $\mathcal{M}^{Ma}$ denote a Mallows' model with any $m \geq 2$, any $n \geq 2$, and any φ. There exists a UMP test for $H_0 = L_{a \succ others}$ vs. H_1 for each $0 < \alpha < 1$ if and only if there exists $B \subseteq \mathcal{A}$ such that $H_1 \subseteq L_{B \succ a}$. Moreover, if $H_1 \subseteq L_{B \succ a}$ then $f_{\alpha,a,B}$ defined in Theorem 3 is UMP.*

PROOF. The "if" part. We note that $f_{\alpha,a,B}$ does not depend on the orderings among alternatives in B in h_1. It follows that for all $h_1 \in H_1$, $f_{\alpha,a,B}$ is a level-α most powerful test for H_0 vs. $\{h_1\}$, which means that $f_{\alpha,a,B}$ is a UMP test.

The "only if" part. Suppose there exist B, B' such that $B \neq B'$ and there exists two rankings $h_1^1 = [B \succ a \succ \text{others}]$ and $h_1^2 = [B' \succ a \succ \text{others}]$ in H_1. W.l.o.g. suppose $B' - B \neq \emptyset$. Let $K_\alpha = n|B| - 0.5$, $\Gamma_\alpha = 0$, and let $f_{\alpha,a,B}$ denote the most powerful test for H_0 vs. h_1^1 guaranteed by Theorem 3. Because K_α is not an integer, there does not exist P_n such that $w_{P_n}(B \succ a) = K_\alpha$. This means that $f_{\alpha,a,B}$ is the unique most powerful level-α test for H_0 vs. h_1^1. We observe that for any P_n, $f_{\alpha,a,B}(P_n)$ is either 0 or 1, and $f_{\alpha,a,B}(P_n) = 1$ if and only if a is ranked above B in all n rankings in P_n. It follows that $f_{\alpha,a,B}$ must be the unique level-α UMP test for H_0 vs. H_1.

By Theorem 3, any most powerful level-α test, in particular $f_{\alpha,a,B}$, must agree with $f_{\alpha,a,B'}$ except for the threshold cases $w_{P_n}(B' \succ a) = K'_\alpha$ for some K'_α. Choose arbitrary $b' \in B' - B$ and $b \in B$. Let P_n^* be composed of n copies of $[B \succ a \succ \text{others}]$ and let P'_n be composed of $n-1$ copies of $[b' \succ B \succ a \succ \text{others}]$ and one copy of $[b' \succ (B - \{b\}) \succ a \succ \text{others}]$. Because $w_{P_n^*}(B \succ a) = n|B| > K_\alpha$, we have $f_{\alpha,a,B}(P_n^*) = 1$. This means that the threshold K'_α for $g_{\alpha,a,B'}$ 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 $f_{\alpha,a,B}(P'_n) = 1$. However, $w_{P'_n}(B \succ a) = n|B| - 1 < n|B|$, which is a contradiction because for any profile P_n, $f_{\alpha,a,B}(P_n) = 1$ if and only if $B \succ a$ in all n rankings in P_n. □

We now consider UMP winner tests ($H_1 = L_{a \succ \text{others}}$). The following corollary follows after Theorem 4.

Corrollary 1. Under Mallows' model, for any $m \geq 2$, any $n \geq 2$, and any $a, b \in \mathcal{A}$, there exists $\alpha > 0$ such that no level-α UMP test exists for $H_0 = L_{b \succ \text{others}}$ vs. $H_1 = L_{a \succ \text{others}}$.

THEOREM 5. *Let $\mathcal{M}^{Ma}$ denote a Mallows' model with any $m \geq 2$, any n and any φ. For any any α, $f_{\alpha,a}$ defined below is a UMP test for $H_0 = L_{others \succ a}$ vs. $H_1 = L_{a \succ others}$. For any n-profile P_n,*

$$f_{\alpha,a}(P_n) = \begin{cases} 1 & \text{if } w_{P_n}(a \succ others) > K_\alpha \\ 0 & \text{if } w_{P_n}(a \succ others) < K_\alpha \\ \Gamma_\alpha & \text{if } w_{P_n}(a \succ others) = K_\alpha \end{cases},$$

where K_α and Γ_α are chosen such that the size of $f_{\alpha,a}$ is α.

PROOF. We note that $f_{\alpha,a}$ is insensitive to permutations over $\mathcal{A} - \{a\}$. Therefore, to prove that $f_{\alpha,a}$ is a UMP test, it suffices to prove that for some $h_1 \in H_1$, $f_{\alpha,a}$ is most powerful for H_0 vs. h_1. Choose an arbitrary $h_1 \in H_1$. Let $h_0^* \in H_0$ be the ranking that is obtained from h_1 by moving a to the bottom position without changing the relative positions of the other alternatives. Similar to the proof of Theorem 3, it is not hard to check that $f_{\alpha,a}$ is equivalent to the likelihood ratio test $\text{LR}_{\alpha,h_0^*,h_1}$. Also because $f_{\alpha,a}$ is invariant to permutations over $\mathcal{A} - \{a\}$, for any $h'_0 \in H_0$ and any permutation M over $\mathcal{A} - \{a\}$, we have $\text{Size}(f_{\alpha,a}, h'_0) = \text{Size}(f_{\alpha,a}, M(h'_0))$. In particular, let M denote the permutation such that $M(h'_0) = h_0^*$. We have $\text{Size}(f_{\alpha,a}, h'_0) = \text{Size}(f_{\alpha,a}, h_0^*)$. It follows from Lemma 2 that $f_{\alpha,a}$ is most powerful. This proves the theorem. □

The next two theorems identify conditions on φ in Mallows' model for the UMP winner test $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a \succ \text{others}}$, when $n = 1$. The proofs are quite involved and can be found in the appendix.

THEOREM 6. *Let $\mathcal{M}^{Ma}$ 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-α UMP test exists for $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a \succ others}$.*

THEOREM 7. *Let $\mathcal{M}^{Ma}$ 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 α, a UMP test exists for $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a \succ others}$.*

5 UMP TESTS FOR CONDORCET'S MODEL

For Condorcet's model, given an alternative a, we define $R_{a>\text{others}} = \{V \in \mathcal{B}(\mathcal{A}) : \forall b \in \mathcal{A}, a >_V b\}$, and we will focus on tests where either $H_0 = R_{a>\text{others}}$ (non-winner tests) or $H_1 = R_{a>\text{others}}$ (winner tests).

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. Given two models $\mathcal{M}_X = (\mathcal{S}_X, \Theta_X, \vec{\pi}_X)$ and $\mathcal{M}_Y = (\mathcal{S}_Y, \Theta_Y, \vec{\pi}_Y)$, we let $\mathcal{M}_X \otimes \mathcal{M}_Y = (\mathcal{S}_X \times \mathcal{S}_Y, \Theta_X \times \Theta_Y, \vec{\pi}_X \times \vec{\pi}_Y)$, where for any $(\pi_{\theta_X}, \pi_{\theta_Y}) \in \vec{\pi}_X \times \vec{\pi}_Y$ and any $P_X \in \mathcal{S}_X$ and $P_Y \in \mathcal{S}_Y$, we have $(\pi_{\theta_X}, \pi_{\theta_Y})(P_X, P_Y) = \pi_{\theta_X}(P_X) \cdot \pi_{\theta_Y}(P_Y)$.

Example 5. Given a Condorcet's model $\mathcal{M}^{\text{Co}}$ with $m = 3$. Let $\mathcal{A} = \{1, 2, 3\}$. For any pair of alternatives $\{a, b\}$, we let $\mathcal{M}_{\{a,b\}} = (\{0, 1\}^n, \{0, 1\}, \vec{\pi})$ denote the restriction of $\mathcal{M}^{\text{Co}}$ on the pairwise comparison between a and b. We have $\mathcal{M}^{\text{Co}} = \mathcal{M}_{\{1,2\}} \otimes \mathcal{M}_{\{2,3\}} \otimes \mathcal{M}_{\{1,3\}}$. ■

THEOREM 8. *For any pair of models $\mathcal{M}_X$ and $\mathcal{M}_Y$, suppose Λ_X is a least favorable distribution for composite vs. simple test ($H_{0,X}$ vs. x_1) under $\mathcal{M}_X$. Given $y_1 \in \Theta_Y$, let Λ^* 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 Λ^* is a least favorable distribution for $H_{0,X} \times \Theta_Y$ vs. (x_1, y_1) under $\mathcal{M}_X \otimes \mathcal{M}_Y$.*

Example 6. Continuing the setting of Example 5, we let $\mathcal{M}_X = \mathcal{M}_{\{1,2\}}$, $H_{0,X} = \{0\}$, $x_1 = 1$, Λ_X is the deterministic distribution $\{0\}$, and let $\mathcal{M}_Y = \mathcal{M}_{\{2,3\}\times\{1,3\}}$ and $y_1 = (1, 1)$. Λ_X is a least favorable distribution according to the Neyman-Pearson lemma. Then, Λ^* is the deterministic distribution $\{(0, 1, 1)\}$ and it is a least favorable distribution for $\{0\} \times \{0, 1\}^2$ vs. $(1, 1, 1)$ under the Condorcet's model with $m = 3$. ■

The next theorem focuses on the setting where we combine $t \in \mathbb{N}$ independent and identical statistical model $\mathcal{M}_X$. Given $\mathcal{M}_X = (\mathcal{S}, \Theta, \vec{\pi})$, a distribution Λ over Θ, any $\theta^* \in \Theta$, and any $t \in \mathbb{N}$, we let $(\mathcal{M}_X)^t = \underbrace{\mathcal{M}_X \otimes \cdots \otimes \mathcal{M}_X}_{t}$ and define the extension of Λ to Θ^t w.r.t. θ^*, denoted by $\text{Ext}(\Lambda, \theta^*, t)$, as follows. Let $\vec{\theta}^* = (\theta^*, \ldots, \theta^*) \in \Theta^t$. For any $j \in t$ and any $\theta \in \Theta$, we have $\text{Ext}(\Lambda, \theta^*, t)(\theta, [\vec{\theta}^*]_{-j}) = \frac{1}{t}\Lambda(\theta)$. That is, $\text{Ext}(\Lambda, \theta^*, t)$ generates a vector $\vec{\theta} \in \Theta^t$ in the following two steps: first, a number $j \leq t$ is chosen uniformly at random. Then, we fix the components of $\vec{\theta}$ to be θ^*, except for the j-th component, which is generated from Θ according to Λ. For any $H_0 \subseteq \Theta$ and any $h_1 \in (\Theta - H_0)$, we let $\vec{h}_1 = (\underbrace{h_1, \ldots, h_1}_{t})$ and let $\text{Ext}(H_0, h_1, t) = (\{H_0 \cup \{h_1\}\}^t - \{\vec{h}_1\})$.

Example 7. Continuing the setting of Example 5, we let $\mathcal{M}_X = \mathcal{M}_{\{1,2\}}$, let Λ denote the deterministic distribution $\{0\}$, let $H_0 = \{0\}$ and $h_1 = 1$. Then $\text{Ext}(\Lambda, 1, 3)$ is the uniform distribution over $\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$, $\vec{h}_1 = (1, 1, 1)$, and $\text{Ext}(H_0, 1, 3) = (\{0, 1\}^3 - \{\vec{1}\})$. ■

THEOREM 9. *For any model $\mathcal{M}_X$ and any $t \in \mathbb{N}$, suppose Λ is a uniformly least favorable distribution for composite vs. simple test (H_0 vs. h_1) under $\mathcal{M}_X$. Then $\text{Ext}(\Lambda, h_1, t)$ is a uniformly least favorable distribution for $\text{Ext}(H_0, h_1, t)$ vs. $\vec{h}_1$ in $(\mathcal{M}_X)^t$.*

Example 8. Continuing the setting of Example 7, it follows from Theorem 9 that the uniform distribution over $\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$ is a uniformly least favorable distribution for testing $\text{Ext}(H_0, 1, 3) = (\{0, 1\}^3 - \{\vec{1}\})$ vs. $\vec{h}_1 = (1, 1, 1)$ under $(\mathcal{M}_X)^3$, which is a Condorcet's model. ■

THEOREM 10. *Let $\mathcal{M}^{Co}$ denote a Condorcet's model with any $m \geq 2$, any n, and any φ. For any $a \in \mathcal{A}$ and any $h_1 \in (\mathcal{B}(\mathcal{A}) - R_{a>others})$, Let h_0^* denote the binary relation obtained from h_1 by letting $a \succ b$ for all $b \in \mathcal{A}$. Then the deterministic distribution $\{h_0^*\}$ is a uniformly least favorable* *distribution for $H_0 = R_{a \succ others}$ vs. h_1 and $g_{\alpha,a,B}$ defined below is a most powerful test. For any n-profile P_n, $g_{\alpha,a,B}(P_n) = \begin{cases} 1 & \text{if } w_{P_n}(B \succ a) > K_\alpha \\ 0 & \text{if } w_{P_n}(B \succ a) < K_\alpha \\ \Gamma_\alpha & \text{if } w_{P_n}(B \succ a) = K_\alpha \end{cases}$, where $B \subset \mathcal{A}$ is the set of alternatives that are preferred to a in h_1. Moreover, when there is no profile P with $w_{P_n}(B \succ a) = K_\alpha$, then $g_{\alpha,a,B}$ is the unique level-α most powerful test.*

PROOF. Let $X = \{\{a, b\} : b \neq a\}$ denote the pairwise comparisons between alternatives in $\mathcal{A}$ that involve a and let Y denote other pairwise comparisons. Let $\mathcal{M}_X = (\mathcal{S}_X, \Theta_X, \vec{\pi}_X)$ denote Condorcet's model $\mathcal{M}^{\text{Co}}$ restricted to X. That is, $\mathcal{S}_X = \{0,1\}^{(m-1)n}, \Theta_X = \{0,1\}^m$ and for any $\theta \in \Theta_X$ and any $P_n \in \mathcal{S}_X$, $\pi_\theta(P_n) \propto \varphi^{\text{KT}(\theta, P_n)}$. Similarly, let $\mathcal{M}_Y$ denote Condorcet's model restricted to Y. It follows that $\mathcal{M}^{\text{Co}} = \mathcal{M}_X \otimes \mathcal{M}_Y$.

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 $b \in \mathcal{A}$. By Neyman-Pearson lemma (Lemma 1), the deterministic distribution $\Lambda_X = \{x_0\}$ is a uniformly least favorable distribution for x_0 vs. x_1. Therefore, by Theorem 8, the deterministic distribution $\Lambda = \{(x_0, y_1)\}$ is uniformly least favorable for $\{x_0\} \times \Theta_Y$ vs. (x_1, y_1). We note that $(x_0, y_1) = h_0^*$ and $(x_1, y_1) = h_1$. It is not hard to verify that $g_{\alpha,a,B}$ is equivalent to the likelihood ratio test $\text{LR}_{\alpha,\lambda,h_1}$, which is most powerful. □

Subsequently, we have the following characterization of UMP non-winner tests under Condorcet's model ($H_0 = R_{a \succ \text{others}}$). For any $B \subset \mathcal{A}$, we let $R_{B \succ a} \subseteq \mathcal{B}(\mathcal{A})$ denote the set of all binary relations where the set of alternatives that are preferred to a is B.

THEOREM 11. *Let $\mathcal{M}^{Co}$ denote a Condorcet's model with any $m \geq 2$ and $n \geq 2$. There exists a UMP test for $H_0 = R_{a \succ others}$ vs. H_1 for every $0 < \alpha < 1$ if and only if there exists $B \subseteq \mathcal{A}$ such that $H_1 \subseteq R_{B \succ a}$. Moreover, if $H_1 \subseteq R_{B \succ a}$ then $g_{\alpha,a,B}$ defined in Theorem 10 is a UMP test.*

The following two results characterize UMP winner tests under Condorcet's model ($H_1 = R_{a \succ \text{others}}$).

Corrollary 2. Let $\mathcal{M}^{\text{Co}}$ denote a Condorcet's model with any $m \geq 2$, any $n \geq 2$, and any φ. For any $a \neq b$, there exists α such that no UMP test exists for $H_0 = R_{b \succ \text{others}}$ vs. $H_1 = R_{a \succ \text{others}}$.

THEOREM 12. *Let $\mathcal{M}^{Co}$ denote a Condorcet's model with any $m \geq 2$, any $n \geq 2$, and any φ. For any α, $g_{\alpha,a}$ defined below is a level-α UMP test for $H_0 = (\mathcal{B}(\mathcal{A}) - H_1)$ vs. $H_1 = R_{a \succ others}$. For any P_n,*

$$g_{\alpha,a}(P_n) = \begin{cases} 1 & \text{if } Ratio(P_n) > k_\alpha \\ 0 & \text{if } Ratio(P_n) < k_\alpha \\ \gamma_\alpha & \text{if } Ratio(P_n) = k_\alpha \end{cases},$$

where $Ratio(P_n) = \dfrac{m-1}{\sum_{b \neq a} \varphi^{w_{P_n}(a \succ b)}}$, and k_α and γ_α are chosen such that the level of $g_{\alpha,a}$ is α.

PROOF. Let $\mathcal{M}_1$ denote the Condorcet's model with a single sample (ranking). Let $X_1, \ldots, X_{m-1}$ denote the $m-1$ pairwise comparisons between a and other alternatives. Similarly to the proof of Theorem 10, we let $\mathcal{M}_{X_1}, \ldots, \mathcal{M}_{X_{m-1}}$ denote the restriction of $\mathcal{M}_1$ on the $m-1$ pairwise comparisons, and let $\mathcal{M}_Y$ denote the restriction of $\mathcal{M}^{\text{Co}}$ on other pairwise comparisons. In fact, $\mathcal{M}_{X_1}, \ldots \mathcal{M}_{X_{m-1}}$ are the same model. It follows that $\mathcal{M}^{\text{Co}} = \mathcal{M}_{X_1} \otimes \mathcal{M}_{X_2} \otimes \cdots \mathcal{M}_{X_{m-1}} \otimes \mathcal{M}_Y$.

In $\mathcal{M}_{X_1}$, let 1 denote 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 Theorem 2 that Λ is still a uniformly least favorable distribution for $\mathcal{M}_{X_{1,n}}$. By Theorem 9,

$\text{Ext}(\Lambda, h_1, m-1)$ is a uniformly least favorable distribution for $\text{Ext}(H_0, h_1, m-1) = (\{0,1\}^{m-1} - \{\vec{1}\})$ vs. $h_1 = \vec{1}$ 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 Λ_{y_1} denote the distribution that is obtained from $\text{Ext}(\Lambda, h_1, m-1)$ by appending y_1 to each parameter. By Theorem 2, Λ_{y_1} is a uniformly least favorable distribution for $\text{Ext}(H_0, h_1, m-1) \times \Theta_{Y,n}$ vs. $(\vec{1}, y_1)$ 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 $\text{Ext}(H_0, h_1, m-1) \times \Theta_{Y,n} = (\{0,1\}^{m-1} - \{\vec{1}\}) \times \Theta_Y = (\mathcal{B}(\mathcal{A}) - R_{a \succ \text{others}})$. This means that the likelihood ratio test $\text{LR}_{\alpha, \Lambda_{y_1}, (\vec{1}, y_1)}$ is a most powerful level-α test for $(\mathcal{B}(\mathcal{A}) - R_{a \succ \text{others}})$ vs. $(\vec{1}, y_1)$. We note that for all y_1, $\text{LR}_{\alpha, \Lambda_{y_1}, (\vec{1}, y_1)}$ is the same test, which means that it is also UMP. It can be verified that $g_{\alpha, a} = \text{LR}_{\alpha, \Lambda_{y_1}, (\vec{1}, y_1)}$. □

6 PROCEDURES FOR SOCIAL 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 now propose two natural procedures to choose the winner by combining multiple winner tests ($H_1 = L_{a \succ \text{others}}$ for Mallows' model and $H_1 = R_{a \succ \text{others}}$ for Condorcet' model) and non-winner tests ($H_0 = L_{a \succ \text{others}}$ for Mallows' model and $H_0 = R_{a \succ \text{others}}$ for Condorcet' model), respectively.

• **The procedure based on winner tests.** We first choose any winner test, such as a UMP test characterized in Theorem 5, then find the alternative a with the minimum α such that H_0 is rejected in the winner test, by conducting binary search on α.[2] This corresponds at a high level to choosing the alternative that is most likely to be the winner according to the tests.

• **The procedure based on non-winner tests.** Similarly, we use binary search on α to find the alternative a with the maximum α 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, the procedure based on winner tests corresponds to the Borda voting rule when we use the UMP winner test for $H_0 = L_{\text{others} \succ a}$ vs. $H_1 = L_{a \succ \text{others}}$ under Mallows' model (Theorem 5). This provides a new theoretical justification for the Borda rule; or vice versa, Borda provides a justification of the proposed procedure.

7 SUMMARY AND FUTURE WORK

We characterized UMP tests for deciding whether a given alternative is the winner under Mallows' model and Condorcet's model, and proposed a procedure to choose a single winner by combining the results of multiple tests. There are many open questions and directions for future research. 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?

REFERENCES

Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. 2014. Statistical Decision Theory Approaches to Social Choice. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. Montreal, Quebec, Canada.

George E. P. Box and Norman R. Draper. 2007. *Empirical Model-Building and Response Surfaces* (2nd edition ed.). John Wiley & Sons.

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

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

[2]Co-winners may exist if they all reject H_0 for the same α's.

Vincent Conitzer and Tuomas Sandholm. 2005. Common Voting Rules as Maximum Likelihood Estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*. Edinburgh, UK, 145–152.

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

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

Bradley Efron. 2005. Bayesians, Frequentists, and Scientists. *J. Amer. Statist. Assoc.* 100, 1–5 (2005).

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

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

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

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

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

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

Andrew Mao, Ariel D. Procaccia, and Yiling Chen. 2013. Better Human Computation Through Principled Voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. Bellevue, WA, USA.

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

Shmuel Nitzan and Jacob Paroush. 2017. Collective Decision Making and Jury Theorems. In *The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts*, Francesco Parisi (Ed.). Oxford University Press.

Jacob Paroush. 1998. Stay away from fair coins: A Condorcet jury theorem. *Social Choice and Welfare* 15, 1 (1998), 15–20.

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

Ariel D. Procaccia, Sashank J. Reddi, and Nisarg Shah. 2012. A Maximum Likelihood Approach For Selecting Sets of Alternatives.. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*.

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

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

Lirong Xia and Vincent Conitzer. 2011. A Maximum Likelihood Approach towards Aggregating Partial Orders. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*. Barcelona, Catalonia, Spain, 446–451.

H. Peyton Young. 1988. Condorcet's Theory of Voting. *American Political Science Review* 82 (1988), 1231–1244.

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

8 APPENDIX: OPTIMAL STATISTICAL HYPOTHESIS TESTING FOR SOCIAL CHOICE

Theorem 1. Λ is a uniformly least favorable distribution for H_0 vs. h_1 if and only if for any $h_0^* \in$ Support(Λ) and any $h_0 \in H_0$, $X^{\Lambda}_{h_0^*}$ weakly dominates $X^{\Lambda}_{h_0}$.

PROOF. To simplify notation we let LR_α and Ratio to denote $\text{LR}_{\alpha,\Lambda,h_1}$ and $\text{Ratio}_{\Lambda,h_1}$, respectively. For any $0 < \alpha < 1$ and any $h_0 \in H_0$, we have

$$\begin{aligned}\text{Size}(\text{LR}_\alpha, h_0) = &\sum\nolimits_{P\in\mathcal{S}:\text{Ratio}(P)>k_\alpha} \pi_{h_0}(P)\\ &+ \gamma_\alpha \sum\nolimits_{P\in\mathcal{S}:\text{Ratio}(P)=k_\alpha} \pi_{h_0}(P)\\ =&\Pr(X^{\Lambda}_{h_0} > \log k_\alpha) + \gamma_\alpha \Pr(X^{\Lambda}_{h_0} = \log k_\alpha)\\ =&(1-\gamma_\alpha)\Pr(X^{\Lambda}_{h_0} > \log k_\alpha) + \gamma_\alpha \Pr(X^{\Lambda}_{h_0} \geq \log k_\alpha)\\ =&(1-\gamma_\alpha)\lim_{x\to\log k_\alpha^-}\Pr(X^{\Lambda}_{h_0} \geq x) + \gamma_\alpha \Pr(X^{\Lambda}_{h_0} \geq k_\alpha)\end{aligned} \tag{1}$$

The "if" direction: for any $h_0 \in H_0$ and any $h_0^* \in$ Support(Λ), 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 $\text{Size}(\text{LR}_\alpha, h_1), h_0^*) \geq \text{Size}(\text{LR}_\alpha, h_0)$. By Lemma 2, LR_α is a level-α most powerful test. Therefore Λ 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 $h_0^* \in$ Support(Λ) 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 Λ is uniformly least favorable, the size of LR_α must be α, where $k_\alpha = 2^x$ and $\gamma_\alpha = 1$. By Lemma 2, $\Pr(X^{\Lambda}_{h_0} \geq x) = \text{Size}(\text{LR}_\alpha, h_0) \geq \text{Size}(\text{LR}_\alpha, h_0^*) = \Pr(X^{\Lambda}_{h_0^*} \geq x)$, which is a contradiction. □

Theorem 3. Let $\mathcal{M}^{\text{Ma}}$ denote a Mallows' model with any $m \geq 2$, any n, and any φ. For any α, any $a \in \mathcal{A}$, and any $h_1 = [B \succ a \succ \text{others}]$ for non-empty B, let h_0^* denote the ranking that is obtained from h_1 by raising a to the top position. Then the deterministic distribution Λ at $\{h_0^*\}$ is a uniformly least favorable distribution for $H_0 = L_{a\succ\text{others}}$ vs. h_1, and $f_{\alpha,a,B}$ defined below is a level-α UMP test. For any n-profile P_n, $f_{\alpha,a,B}(P_n) = \begin{cases} 1 & \text{if } w_{P_n}(B \succ a) > K_\alpha \\ 0 & \text{if } w_{P_n}(B \succ a) < K_\alpha \\ \Gamma_\alpha & \text{if } w_{P_n}(B \succ a) = K_\alpha \end{cases}$, where K_α and Γ_α are chosen such that the size of $f_{\alpha,a,B}$ is α. Moreover, any most powerful test must agree with $f_{\alpha,a,B}(P_n)$ except on P_n with $w_{P_n}(B \succ a) = K_\alpha$.

PROOF. We have $h_0^* = [a \succ B \succ \text{other}]$. In this proof we let LR_α denote $\text{LR}_{\alpha,h_0^*,h_1}$ and let Ratio denote $\text{Ratio}_{h_0^*,h_1}$. We first prove the theorem for $n = 1$ then extend it to arbitrary n by Theorem 2. By Theorem 1, it suffices to prove that for any $h_0 \in H_0$, $X^{\Lambda}_{h_0^*}$ weakly dominates $X^{\Lambda}_{h_0}$.

For any ranking V and any pair of alternatives a, b, we let $I(a \succ_V b) = 1$ if $a \succ_V b$, otherwise $I(a \succ_V b) = 0$. For any $P \in \mathcal{L}(\mathcal{A})$, we have

$$
\begin{aligned}
&\log \text{Ratio}(P) = \log \varphi(\text{KT}(P, h_1) - \text{KT}(P, h_0)) \\
=& -\log \varphi \sum_{c \succ_P d} (I(d \succ_{h_1} c) - I(d \succ_{h_0} c)) \\
=& -\log \varphi (\sum_{a \succ_P b : b \in B} (I(b \succ_{h_1} a) - I(b \succ_{h_0} a)) \\
&+ \sum_{b \succ_P a : b \in B} (I(a \succ_{h_1} b) - I(a \succ_{h_0} b))) \\
=& -\log \varphi \cdot (2 w_P(B \succ a) - |B|)
\end{aligned}
$$

Therefore, to prove that $X^{\Lambda}_{h_0^*}$ weakly dominates $X^{\Lambda}_{h_0}$, 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 $\mathcal{A}$ such that $M(h_0) = h_0^*$. Because $h_0 \in H_0 = L_{a \succ \text{others}}$, we have $M(a) = a$. Let $B' = M(B)$. Because Kendall-Tau distance is invariant to permutations, for any $P \in \mathcal{L}(\mathcal{A})$ we have $\pi_{h_0}(P) = \pi_{M(h_0)}(M(P))$ and

$$
\begin{aligned}
&\pi_{h_0}(\{P : w_P(B \succ a) \geq K\}) \\
=& \pi_{M(h_0)}(\{M(P) : w_{M(P)}(M(B) \succ M(a)) \geq K\}) \\
=& \pi_{h_0^*}(\{M(P) : w_{M(P)}(B' \succ M(a)) \geq K\}) \\
=& \pi_{h_0^*}(\{P : w_P(B' \succ a) \geq K\})
\end{aligned}
$$

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 $W \in \mathcal{L}(\mathcal{A})$ and $C', C \subseteq \mathcal{A}$ 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 3. *Under a Mallows' model, for any φ, any $K \in \mathbb{N}$, any $a \in \mathcal{A}$, any $W \in \mathcal{L}(\mathcal{A})$, and any $C', C \subseteq \mathcal{A}$ such that C dominates C' w.r.t. W, we have*

$$\pi_W(\{P : w_P(C' \succ a) \geq K\}) \leq \pi_W(\{P : w_P(C \succ a) \geq K\})$$

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 $c \notin C'$. Let $c' \in C'$ such that $c' \notin C$. Because C dominates C' in W, we have $c \succ_W c'$.

Let $\mathcal{P} = \{P \in \mathcal{L}(\mathcal{A}) : w_P(C \succ a) \geq K\}$ and $\mathcal{P}' = \{P \in \mathcal{L}(\mathcal{A}) : w_P(C' \succ a) \geq K\}$. We define the following permutation $\mathcal{M}$ over $\mathcal{L}(\mathcal{A})$. For any $P \in \mathcal{L}(\mathcal{A})$, 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 $P \in \mathcal{P} - \mathcal{P}'$, we must have $c \succ_P a \succ_P c'$ and $(C - C') \succ_P a$. Therefore, $\mathcal{M}(\mathcal{P} - \mathcal{P}') = \mathcal{P}' - \mathcal{P}$.

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

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

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 $\pi_W(\{P : w_P(C \succ a) \geq K\}) \geq \pi_W(\{P : w_P(C_1 \succ a) \geq K\}) \geq \cdots \geq \pi_W(\{P : w_P(C' \succ a) \geq K\})$ □

It follows from Lemma 3 that $X^{\Lambda}_{h_0^*}$ weakly dominates $X^{\Lambda}_{h_0}$, which means that Λ is a uniformly least favorable distribution for $n = 1$ by Theorem 1. We note that Λ is deterministic. Therefore, by Theorem 2, Λ 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_α is most powerful. It is not hard to verify that $\text{LR}_\alpha = f_{\alpha,a,B}$. Moreover, because Λ 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^* vs. h_1). By the Neyman-Pearson lemma (Lemma 1), f must agree with LR_α except on P_n such that $\text{Ratio}(P_n) = k_\alpha$, which corresponds to P_n with $w_{P_n}(B \succ a) = K_\alpha$. □

Remarks. We note that $f_{\alpha,a,B}$ does not depend on the ordering among alternatives in B in h_1. $f_{\alpha,a,B}$ is computed in the following way for any given profile P_n: we first build the weighted majority graph, then calculate the total weight $w_{P_n}(B \succ a)$ of all edges from B to a. If the total weight is more than a threshold K_α, then H_0 is rejected; if the total weight is less than K_α, then H_0 is retained; otherwise H_0 is rejected with probability Γ_α. This procedure is intuitive because a larger $w_{P_n}(B \succ a)$ corresponds to more evidence from the data that a should be ranked below B, which means that it is less likely that the ground truth is in H_0, where a is ranked above B. $w_{P_n}(B \succ a)$ is called the *test statistic* and is easy to compute. The threshold K_α and the value Γ_α might be hard to compute. In practice such K_α and Γ_α are pre-computed as a look-up table, and once $w_{P_n}(B \succ a)$ is computed, we can immediately obtain its p-value, which is the smallest α such that $K_\alpha = w_{P_n}(B \succ a)$.

Theorem 6. Let $\mathcal{M}^{\text{Ma}}$ 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-α UMP test exists for $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a\succ\text{others}}$.

Proof. By Lemma 6, if a UMP test exists then $\bar{f}_{\alpha,a}$ is also a UMP test. Therefore, it suffices to prove that $\bar{f}_{\alpha,a}$ is not a level-α 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 $\bar{f}_{\alpha,a}$.

Let $V_1, V_2, \ldots, V_m, V_2' \in \mathcal{L}(\mathcal{A})$ 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 $V_3' = [a_3 \succ a_1 \succ a_4 \succ a_2 \succ \text{others}]$.

We consider the following critical function f. For any $V \in \mathcal{L}_{a\succ\text{others}}$, we let $f(V) = 1$. For any V_j with $j \neq 3$, let $f(V_j) = 1$. We then let $f(V_3) = f(V_3') = \frac{1+\varphi^m}{1+\varphi}$. Let α denote the size of f at V_2. That is, $\alpha = \text{Size}(f, V_2)$. Let $T = \pi_{V_2}(\mathcal{L}_{a\succ\text{others}})$. It follows that

$$
\begin{aligned}
&\alpha - T\\
\propto&\varphi^0 + \frac{1+\varphi^m}{1+\varphi}(\varphi^{\text{KT}(V_2,V_3)} + \varphi^{\text{KT}(V_2,V_3')}) + \sum_{j=5}^{m}\varphi^{\text{KT}(V_2,V_j)}\\
=&1 + \frac{1+\varphi^m}{1+\varphi}(\varphi^3 + \varphi^4) + \varphi^4 + \sum_{j=5}^{m}\varphi^{\text{KT}(V_2,V_j)}\\
>&1 + \varphi^3 + \varphi^4 + \varphi^5
\end{aligned}
$$

For any $j, j^* \geq 2$ such that $j \neq j^*$, it is not hard to verify that $\text{KT}(V_j, V_{j^*}) = j + j^* - 2$. Moreover, $\text{KT}(V_3, V_3') = 1$, $\text{KT}(V_2, V_3') = 4$, $\text{KT}(V_4, V_3') = 4$, and for any $j \geq 5$, we have $\text{KT}(V_3', V_j) = j + 2$. Therefore, we have the following calculations of $\text{Size}(f, V_3)$, $\text{Size}(f, V_3')$, and $\text{Size}(f, V_4)$ (see

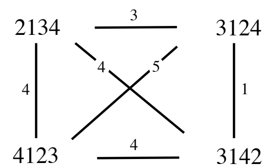

Fig. 2. Kentall-Tau distance for some rankings over four alternatives.

Figure 2 for distances between V_2, V_3, V_3', V_4). We note that $T = \pi_{V_2}(\mathcal{L}_{a>\text{others}}) = \pi_{V_3}(\mathcal{L}_{a>\text{others}}) = \pi_{V_3'}(\mathcal{L}_{a>\text{others}}) = \pi_{V_4}(\mathcal{L}_{a>\text{others}})$ due to symmetry.

$$\text{Size}(f, V_3) - T \propto \varphi^3 + \tfrac{1+\varphi^m}{1+\varphi}(1+\varphi) + \varphi^5 + \textstyle\sum_{j=5} \varphi^{\text{KT}(V_3,V_j)} \leq 1 + \varphi^3 + (m-3)\varphi^5$$
$$\text{Size}(f, V_3') - T \propto \varphi^4 + \tfrac{1+\varphi^m}{1+\varphi}(1+\varphi) + \varphi^4 + \textstyle\sum_{j=5} \varphi^{\text{KT}(V_3',V_j)} \leq 1 + 2\varphi^4 + (m-4)\varphi^6$$
$$\text{Size}(f, V_4) - T \propto \varphi^4 + \tfrac{1+\varphi^m}{1+\varphi}(\varphi^4+\varphi^5) + 1 + \textstyle\sum_{j=5} \varphi^{\text{KT}(V_4,V_j)} \leq 1 + 2\varphi^4 + (m-4)\varphi^7$$

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 $Z_m = \prod_{l=1}^m \frac{1-\varphi^m}{1-\varphi}$ denote the Mallows normalization factor for m alternatives. We have

$$\begin{aligned}
&\text{Ratio}_{V_2,V_1}(T_{m-2}) = \frac{\pi_{V_1}(T_{m-2})}{\pi_{V_2}(T_{m-2})}\\
=&\frac{\varphi Z_{m-1}}{Z_{m-2} + \varphi^2(Z_{m-1} - Z_{m-2})}\\
=&\frac{\varphi\frac{Z_{m-1}}{Z_{m-2}}}{1 + \varphi^2(\frac{Z_{m-1}}{Z_{m-2}} - 1)} = \frac{\varphi + \varphi^2 + \cdots + \varphi^{m-1}}{1 + \varphi^3 + \varphi^4 + \cdots + \varphi^m} < \frac{1}{\varphi}
\end{aligned}$$

$$\begin{aligned}
&\text{Ratio}_{V_2,V_1}(P) = \frac{\varphi + \varphi^2 + \cdots + \varphi^{m-1} + \varphi^{m+2}}{1 + \varphi^3 + \varphi^4 + \cdots + \varphi^m + \varphi^{m+3}}\\
>&\frac{\varphi + \varphi^2 + \cdots + \varphi^{m-1}}{1 + \varphi^3 + \varphi^4 + \cdots + \varphi^m}\\
=&\text{Ratio}_{V_2,V_1}(T_{m-2})
\end{aligned}$$

We note that $\text{Size}(\bar{f}_{\alpha,a}, V_2) = \alpha$. This means that $\text{Power}(\bar{f}_{\alpha,a}, V_1) = \pi_{V_1}(T_{m-1}) + \alpha\text{Ratio}_{T_2,T_1}(T_{m-2}) < \pi_{V_1}(T_{m-1}) + \alpha\text{Ratio}_{T_2,T_1}(P) = \text{Power}(f, V_1)$. This means that $\bar{f}_{\alpha,a}$ is a not a level-α UMP. The theorem follows after Lemma 6. □

Theorem 7. Let $\mathcal{M}^{\text{Ma}}$ 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 α, $\bar{f}_{\alpha,a}$ is a UMP test for $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a>\text{others}}$.

PROOF. We first verify that when $K_\alpha = m - 1$, $\bar{f}_{\alpha,a}$ is a UMP test. For any $h_1 \in H_1$, let $h_0^* \in H_0$ 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 $V \in \mathcal{L}(\mathcal{A})$, $\text{Ratio}_{h_0^*,h_1}(V) \leq 1/\varphi$, and for all $V \in H_1$ we have $\text{Ratio}_{h_0^*,h_1}(V) = 1/\varphi$. This means that for any level-α test for H_0 vs. h_1, the power cannot be more than α/φ. We note that $\bar{f}_{\alpha,a}$ is a level-α test whose power is exactly α/φ. This means that for all $h_1 \in H_1$, $\bar{f}_{\alpha,a}$ is a most powerful test for H_0 vs. h_1. Therefore, when $K_\alpha = m - 1$, $\bar{f}_{\alpha,a}$ is a UMP test.

For any α such that $K_\alpha \leq m-2$, we will prove that for any $h_1 \in H_1$, $\bar{f}_{\alpha,a}$ is a most powerful level-α test for H_0 vs. h_1. This is done in the following steps. Step 1. Find a least favorable distribution $\Lambda_\alpha^{h_1}$

whose support is the set of all rankings where a is ranked at the second position. Step 2. Verify that $\bar{f}_{\alpha,a}$ is the likelihood ratio test w.r.t. $\Lambda_\alpha^{h_1}$, and step 3. verify that the two conditions in Lemma 2 holds for $\Lambda_\alpha^{h_1}$.

Step 1. The main challenge is that in general there does not exist a uniformly least favorable distribution. For different α we define different $\Lambda_\alpha^{h_1}$ as follows. For any α, we let s_α denote the smallest Borda score of the ranking V such that $\bar{f}_{\alpha,a}(V) > 0$. We have that $s_\alpha \leq m-2$. Let the support of $\Lambda_\alpha^{h_1}$ 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 $\Lambda_\alpha^{h_1}$. For any $h_0^* \in T_{m-2}$ there is a variable $x[h_0, s_\alpha]$.

$$\forall V \in T_{s_\alpha}, \sum_{h_0^* \in T_{m-2}} \text{Ratio}_{h_0^*,h_1}^{-1}(V) \cdot x[h_0^*, s_\alpha] = m \qquad (\text{LP}_{s_\alpha}^{h_1})$$

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

Step 2. To simplify notation we let $\text{LR}_\alpha = \text{LR}_{\alpha,\Lambda_\alpha^{h_1},h_1}$ denote the likelihood ratio test and let $\text{Ratio} = \text{Ratio}_{\Lambda_\alpha^{h_1},h_1}$ denote the likelihood ratio function w.r.t. distribution $\Lambda_\alpha^{h_1}$ for H_0 vs. h_1. To prove $\text{LR}_\alpha = \bar{f}_{\alpha,a}$, we first prove that for any $V \in \mathcal{L}(\mathcal{A})$ where a is not ranked at the bottom position, $\text{Ratio}(V) > \text{Ratio}(\text{Down}_a^1(V))$, where we recall that $\text{Down}_a^1(V)$ is the ranking obtained from V by moving a down for one position.

$$\begin{aligned}
&\frac{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \pi_{h_0^*}(\text{Down}_a^1(V))}{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \pi_{h_0^*}(V)} \\
=&\frac{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \varphi^{\text{KT}(h_0^*,\text{Down}_a^1(V))}}{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \varphi^{\text{KT}(h_0^*,V)}} \\
>&\frac{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \varphi^{\text{KT}(h_0^*,V)} \cdot \varphi^{\text{KT}(V,\text{Down}_a^1(V))}}{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \varphi^{\text{KT}(h_0^*,V)}} \\
=&\varphi = \frac{\pi_{h_1}(\text{Down}_a^1(V))}{\pi_{h_1}(V)}
\end{aligned}$$

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 $\text{KT}(h_0^*, V) + \text{KT}(V, \text{Down}_a^1(V)) > \text{KT}(h_0^*, \text{Down}_a^1(V))$, and (3) for all $h_0^* \in T_{m-2}$, $\Lambda_\alpha^{h_1}(h_0^*) > 0$.

It follows from the strict inequality that

$$\begin{aligned}
\text{Ratio}(V) =&\frac{\pi_{h_1}(V)}{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \pi_{h_0^*}(V)} \\
>&\frac{\pi_{h_1}(\text{Down}_a^1(V))}{\sum_{h_0^* \in T_{m-2}} \Lambda_\alpha^{h_1}(h_0^*) \cdot \pi_{h_0^*}(\text{Down}_a^1(V))} \\
=&\text{Ratio}(\text{Down}_a^1(V))
\end{aligned}$$

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

Step 3. Due to the symmetry $f_{\alpha,a}$ among alternatives in $\mathcal{A} - \{a\}$, for any $i \leq m - 2$ and any $h_0, h_0' \in T_i$, we have $\text{Size}(\bar{f}_{\alpha,a}, h_0) = \text{Size}(\bar{f}_{\alpha,a}, h_0')$. Therefore, condition (i) in Lemma 2 is satisfied. Choose arbitrary $h_0^{m-2} \in T_{m-2}$. For any $i \leq m - 3$, let $h_0^i \in T_i$ denote the ranking obtained from h_0^{i+1} 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

$$\pi_{h_0^{m-2}}(\{V : \text{Borda}_a(V) \geq K\}) \geq \pi_{h_0^i}(\{V : \text{Borda}_a(V) \geq K\}) \tag{2}$$

We will prove a slightly stronger lemma.

LEMMA 4. *Under Mallows' model, for any m, any φ, any $W \in \mathcal{L}(\mathcal{A})$, any $b, c \in \mathcal{A}$ such that $b \succ_W c$, and any K, we have $\pi_W(\{V : \text{Borda}_b(V) \geq K\}) \geq \pi_W(\{V : \text{Borda}_c(V) \geq K\})$.*

PROOF. The proof is similar to the proof of Lemma 3. It suffices to prove the lemma for the case where b and c are adjacent in W. Let $\mathcal{P} = \{V \in \mathcal{L}(\mathcal{A}) : \text{Borda}_b(V) \geq K\}$ 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 $\text{KT}(M(V), W) = \text{KT}(V, W) + 1$, which means that $\pi_W(V) = \pi(M(V))/\varphi$. Therefore, we have

$$\begin{aligned}
&\pi_W(\{V : \text{Borda}_b(V) \geq K\}) - \pi_W(\{V : \text{Borda}_c(V) \geq K\}) \\
=&\pi_W(\mathcal{P}) - \pi_W(\mathcal{P}') = \pi_W(\mathcal{P} - \mathcal{P}') - \pi_W(\mathcal{P}' - \mathcal{P}) \\
=&\pi_W(\mathcal{P} - \mathcal{P}') - \pi_W(M(\mathcal{P} - \mathcal{P}')) \\
=&(\frac{1}{\varphi} - 1)\pi_W(\mathcal{P} - \mathcal{P}') \geq 0
\end{aligned}$$

This proves the lemma. □

Let W be an arbitrary ranking and let M_i denote a permutation such that $M_i(h_0^i) = W$. We have $\pi_{h_0^i}(\{V : \text{Borda}_a(V) \geq K\}) = \pi_{M_i(h_0^i)}(\{V : \text{Borda}_{M_i(a)}(V) \geq K\})$. 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 4. This means that condition (ii) in Lemma 2 is also satisfied. Therefore, by Lemma 2, $\bar{f}_{\alpha,a}$ is a level-α most powerful test for H_0 vs. h_1. Since $\bar{f}_{\alpha,a}$ does not depend on h_1, it is a level-α UMP test for H_0 vs. H_1. □

Theorem 8. For any $\mathcal{M}_X$ and $\mathcal{M}_Y$, suppose Λ_X is a least favorable distribution for composite vs. simple test ($H_{0,X}$ vs. x_1) under $\mathcal{M}_X$. Given $y_1 \in \Theta_Y$, let Λ^* 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 Λ^* is a least favorable distribution for $H_{0,X} \times \Theta_Y$ vs. (x_1, y_1) under $\mathcal{M}_X \otimes \mathcal{M}_Y$.

PROOF. Let $x_0^1, \ldots, x_0^K \in \Theta_X$ denote the support of Λ_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 Ratio stands for $\text{Ratio}_{\Lambda^*,(x_1,y_1)}$ and LR_α stands for $\text{LR}_{\alpha,\Lambda^*,(x_1,y_1)}$.

$$\begin{aligned}\text{Ratio}(P_X, P_Y) =& \frac{\pi_{x_1,y_1}(P)}{\sum_{k=1}^K \Lambda^*(x_0^k, y_1)\pi_{(x_0^k,y_1)}(P)}\\ =& \frac{\pi_{x_1}(P_X)\cdot\pi_{y_1}(P_Y)}{\sum_{k=1}^K \Lambda^*(x_0^k, y_1)\pi_{x_0^k}(P_X)\cdot\pi_{y_1}(P_Y)}\\ =& \frac{\pi_{x_1}(P_X)}{\sum_{k=1}^K \Lambda(x_0^k)\pi_{x_0}^k(P_X)} = \text{Ratio}_{\Lambda,x_1}(P_X)\end{aligned}$$

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 $\text{Ratio}_{\Lambda,x}(P_X) \geq \text{Ratio}_{\Lambda,x}(P'_X)$. This means that for any (P_X, P_Y), $\text{LR}_\alpha(P_X, P_Y) = \text{LR}_{\alpha,\Lambda,x_1}(P_X)$. Therefore, for any $x_0 \in H_{0,X}$, we have

$$\begin{aligned}&\text{Size}(\text{LR}_\alpha, (x_0, y_1))\\ =& \sum_{(P_X,P_Y)\in\mathcal{S}_X\times\mathcal{S}_Y} \pi_{x_0}(P_X)\pi_{y_1}(P_Y)\text{LR}_\alpha(P_X, P_Y)\\ =& \sum_{(P_X,P_Y)\in\mathcal{S}_X\times\mathcal{S}_Y} \pi_{x_0}(P_X)\pi_{y_1}(P_Y)\text{LR}_{\alpha,\Lambda,x_1}(P_X)\\ =& \sum_{P_X\in\mathcal{S}_X} \pi_{x_0}(P_X)\text{LR}_{\alpha,\Lambda,x_1}(P_X)\\ =&\text{Size}(\text{LR}_{\alpha,\Lambda,x_1}, x_0)\end{aligned}$$

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

Theorem 9. For any model $\mathcal{M}_X$ and any $t \in \mathbb{N}$, suppose Λ is a uniformly least favorable distribution for composite vs. simple test (H_0 vs. h_1) under $\mathcal{M}_X$. Then $\text{Ext}(\Lambda, h_1, t)$ is a uniformly least favorable distribution for $\text{Ext}(H_0, h_1, t)$ vs. $\vec{h}_1$ in $(\mathcal{M}_X)^t$.

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 $\Lambda^* = \text{Ext}(\Lambda, h_1, t)$, let $H_0^* = \text{Ext}(H_0, h_1, t)$, let $\text{LR}_\alpha = \text{LR}_{\alpha,\Lambda^*,\vec{h}_1}$, $\text{Ratio} = \text{Ratio}_{\Lambda^*,\vec{h}_1}$.

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

PROOF. we have $\text{Ratio}^{-1}(\vec{x}) = \frac{1}{t}\cdot\frac{\sum_{j=1}^t\sum_{h_0\in H_0}\Lambda(h_0)\cdot\pi_{(h_0,[\vec{h}_1]_{-j})}(\vec{x})}{\pi_{\vec{h}_1}(\vec{x})}$
$= \frac{1}{t}\cdot\frac{\sum_{j=1}^t\sum_{h_0\in H_0}\Lambda(h_0)\cdot\pi_{h_0}(x_j)\cdot\pi_{[\vec{h}_1]_{-j}}(x_j)}{\pi_{h_1}(x_j)\cdot\pi_{[\vec{h}_1]_{-j}}(x_j)}$
$= \frac{1}{t}\sum_{j=1}^t \text{Ratio}_{\Lambda,h_1}^{-1}(x_j)$ □

The next lemma proves the following: For any $\vec{z} \in H_0^*$ and any $j \leq t$, suppose the j-th component is not in $\text{Support}(\Lambda) \cup \{h_1\}$. If we fix all components except j-th in $\vec{z}$ and change the j-th component to $h_0^* \in \text{Support}(\Lambda)$, then the size of LR_α will increase. If we further change the j-th component to h_1, then the size of LR_α will further increase.

LEMMA 5. *For any* $0 \leq \alpha \leq 1$, *any* $j \leq t$, *any* $\vec{z}_{-j} \in \Theta^{t-1}$, *any* $h_0 \in H_0$, *and any* $h_0^* \in Support(\Lambda)$, *we have* $Size(LR_\alpha, (h_0, \vec{z}_{-j})) \leq Size(LR_\alpha, (h_0^*, \vec{z}_{-j})) \leq Size(LR_\alpha, (h_1, \vec{z}_{-j}))$.

PROOF. For any $\vec{z}_{-j} \in \Theta^{n-1}$, we have

$$\begin{aligned}&\text{Size}(\text{LR}_\alpha, (h_0, \vec{z}_{-j})) = \pi_{(h_0, \vec{z}_{-j})}(\{\vec{x} \in \mathcal{S}^t : \text{Ratio}(\vec{x}) > k_\alpha^*\})\\ &+ \gamma_\alpha^* \pi_{(h_0, \vec{z}_{-j})}(\{\vec{x} \in \mathcal{S}^t : \text{Ratio}(\vec{x}) = k_\alpha^*\})\end{aligned}$$

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

$$\begin{aligned}&\pi_{(h_0, \vec{z}_{-j})}(\{\vec{x} \in \mathcal{S}^t : \text{Ratio}(\vec{x}) > k_\alpha^*\})\\ =&\pi_{(h_0, \vec{z}_{-j})}(\{\vec{x} \in \mathcal{S}^t : \text{Sum}(\vec{x}) < t/k_\alpha^*\})\\ =&\pi_{(h_0, \vec{z}_{-j})}(\{\vec{x} \in \mathcal{S}^t : \text{Sum}(\vec{x}_{-j}) + \text{Ratio}_{\Lambda,h_1}^{-1}(x_j) < t/k_\alpha^*\})\\ =&\int_0^{t/k_\alpha^*} \sum\nolimits_{\vec{x}_{-j} \in \mathcal{S}^{t-1}:\text{Sum}(\vec{x}_{-j})=p}\\ &\sum\nolimits_{x_j:\text{Ratio}_{\Lambda,h_1}^{-1}(x_j) < t/k_\alpha^* - p} \pi_{(h_0, \vec{z}_{-j})}(\vec{x})dp\\ =&\int_0^{t/k_\alpha^*} \pi_{\vec{z}_{-j}}(\{\vec{x}_{-j} \in \mathcal{S}^{t-1} : \text{Sum}(\vec{x}_{-j}) = p\})\\ &\cdot \pi_{h_0}(\{x_j : \text{Ratio}_{\Lambda,h_1}^{-1}(x_j) < t/k_\alpha^* - p\})dp\\ =&\int_0^{t/k_\alpha^*} Q(\vec{z}_{-j}, p) \cdot \pi_{h_0}(\{x_j : \text{Ratio}_{\Lambda,h_1}^{-1}(x_j) < t/k_\alpha^* - p\})dp\end{aligned}$$

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

$$\text{Size}(\text{LR}_\alpha, (h_0, \vec{z}_{-j})) = \int_0^{t/k_\alpha^*} Q(\vec{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_0)dp \tag{3}$$

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

$$\text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_0) \leq \alpha' \leq \text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_0^*)$$

Then by Equation (3), for any $h_0 \in (H_0 - \text{Support}(\Lambda))$ and any $h_0^* \in \text{Support}(\Lambda)$, we have

$$\begin{aligned}&\text{Size}(\text{LR}_\alpha, (h_0, \vec{z}_{-j}))\\ =&\int_0^{t/k_\alpha^*} Q(\vec{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_0)dp\\ \leq&\int_0^{t/k_\alpha^*} Q(\vec{z}_{-j}, p) \cdot \text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_0^*)dp\\ =&\text{Size}(\text{LR}_\alpha, (h_0^*, \vec{z}_{-j}))\end{aligned}$$

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 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 Λ is a level-η least favorable distribution. Then we have $\text{Size}(\text{LR}_\eta, h_1) \geq \eta = \text{Size}(\text{LR}_\eta, h_0^\Lambda)$. [3]

PROOF. For the sake of contradiction suppose this is not true, that is, for any $h_0^* \in \text{Support}(\Lambda)$ 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

$$\begin{aligned}
&\text{Size}(\text{LR}_\eta, h_1)\\
=&\sum_{P\in\mathcal{S}:\text{Ratio}(P)>k_\eta} \pi_{h_1}(P) + \gamma_\eta \sum_{P\in\mathcal{S}:\text{Ratio}(P)=k_\eta} \pi_{h_1}(P)\\
\geq&\sum_{P\in\mathcal{S}:\text{Ratio}(P)>k_\eta} \pi_\Lambda(P)\cdot k_\eta + \gamma_\eta \sum_{P\in\mathcal{S}:\text{Ratio}(P)=k_\eta} \pi_\Lambda(P)\cdot k_\eta\\
>&\sum_{P\in\mathcal{S}:\text{Ratio}(P)>k_\eta} \pi_\Lambda(P) + \gamma_\eta \sum_{P\in\mathcal{S}:\text{Ratio}(P)=k_\eta} \pi_\Lambda(P) = \eta,
\end{aligned}$$

which is a contradiction. Therefore, we have

$$\begin{aligned}
&1\\
=&\text{Size}(\text{LR}_\eta, h_1) + \textstyle\sum_{P\in\mathcal{S}:\text{Ratio}(P)<k_\eta} \pi_{h_1}(P)\\
&+ (1-\gamma_\eta)\textstyle\sum_{P\in\mathcal{S}:\text{Ratio}(P)=k_\eta} \pi_{h_1}(P)\\
<&\eta + \textstyle\sum_{P\in\mathcal{S}:\text{Ratio}(P)<k_\eta} \pi_\Lambda(P)\cdot k_\eta\\
&+ (1-\gamma_\eta)\textstyle\sum_{P\in\mathcal{S}:\text{Ratio}(P)=k_\eta} \pi_\Lambda(P)\cdot k_\eta\\
\leq&\eta + k_\eta(1-\text{Size}(\text{LR}_\eta, h_0^\Lambda)) \leq 1,
\end{aligned}$$

which is a contradiction. □

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

$$\begin{aligned}
&\text{Size}(\text{LR}_\alpha, (h_0^*, \vec{z}_{-j}))\\
=&\int_0^{t/k_\alpha^*} Q(\vec{z}_{-j}, p)\cdot \text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_0^*)dp\\
\leq&\int_0^{t/k_\alpha^*} Q(\vec{z}_{-j}, p)\cdot \text{Size}(\text{LR}_{\alpha',\Lambda,h_1}, h_1)dp\\
=&\text{Size}(\text{LR}_\alpha, (h_1, \vec{z}_{-j}))
\end{aligned}$$

This finishes the proof of Lemma 5. □

It follows from Lemma 5 that for any $j \leq t$ and any $h_0^* \in \text{Support}(\Lambda)$, we have that $\text{Size}(\text{LR}_\alpha, (h_0^*, [\vec{h}_1]_{-j}))$ is the same. Due to symmetry, for any $\vec{h}_0^* \in H_0^*$, $\text{Size}(\text{LR}_\alpha, h_0^*)$ is the same and is therefore equivalent to α. This verifies condition (i) in Lemma 2.

Condition (ii) in Lemma 2 is verified by recursively applying Lemma 5. Given any $\vec{h}_0 \in H_0^* - \text{Support}(\Lambda^*)$, there must exist $j \leq t$ such that $[\vec{h}_0]_j \neq h_1$. We then change $[\vec{h}_0]_j$ to an arbitrary $h_0^* \in \text{Support}(\Lambda)$, then change the other components of $\vec{h}_0$ to h_1 one by one. Each time we make

[3]We recall that h_0^Λ is the combined H_0 by Λ.

the change the size of LR_α does not decrease according to Lemma 5. At the end of the process we obtain $(h_0^*, [\vec{h}_1]_j) \in \mathrm{Support}(\Lambda^*)$, at which the size of LR_α is α. The theorem follows after applying Lemma 2. □

We now define a test $\bar{f}_{\alpha,a}$ for $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a>\text{others}}$ and prove that if a UMP test exists, then $\bar{f}_{\alpha,a}$ must also be a UMP test. For any $V \in \mathcal{L}(\mathcal{A})$ and any alternative $a \in \mathcal{A}$, we let $\mathrm{Borda}_a(V)$ denote the Borda score of a in V. That is, $\mathrm{Borda}_a(V)$ is the number of alternatives that are ranked below a in V. For any $V \in \mathcal{L}(\mathcal{A})$, we let $\bar{f}_{\alpha,a}(V) = \begin{cases} 1 & \text{if } \mathrm{Borda}_a(V) > K_\alpha \\ 0 & \text{if } \mathrm{Borda}_a(V) < K_\alpha \\ \Gamma_\alpha & \text{if } \mathrm{Borda}_a(V) = K_\alpha \end{cases}$, where K_α and Γ_α are chosen so that the size of $\bar{f}_{\alpha,a}$ is α. In other words, $\bar{f}_{\alpha,a}$ calculates the Borda score of a in the input profile, and if it is larger than a threshold K_α then H_0 is rejected. It is not hard to see that $\bar{f}_{\alpha,a}$ equals to $f_{\alpha',a}$ with a possibly different level α' (defined in Theorem 5).

LEMMA 6. *If there exists a level-α UMP test for $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ vs. $H_1 = L_{a>others}$, then $\bar{f}_{\alpha,a}$ is also a level-α UMP test.*

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

It follows that for any permutation M over $\mathcal{A} - \{a\}$, $M(f_\alpha)$ is also UMP. Therefore, $\bar{f}_\alpha = \frac{1}{(m-1)!}\sum_M M(f_\alpha)$ is also UMP. We note that for any V, V' where a has the same Borda score, there exists a permutation M over $\mathcal{A} - \{a\}$ 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 $\mathrm{Borda}_a(V) > \mathrm{Borda}_a(V')$, 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 $s_1 = \mathrm{Borda}_a(V) > \mathrm{Borda}_a(V') = s_2$, $\bar{f}_\alpha(V') > 0$, 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, $T_s = \{V \in \mathcal{L}(\mathcal{A}) : \mathrm{Borda}_a(V) = s\}$. 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 $\mathrm{Ratio}_{h_0,h_1}(T_{s_1}) > \mathrm{Ratio}_{h_0,h_1}(T_{s_2})$.

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 θ by raising a to the top position. For any $V_{s_1} \in T_{s_1}$, we let $\mathrm{Down}_a^{s_1-s_2}(V_{s_1}) \in T_{s_2}$ 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

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

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, $\text{KT}(h_0, \text{Down}_a^{s_1-s_2}(V)) < \text{KT}(h_0, V) + \text{KT}(V, \text{Down}_a^{s_1-s_2}(V))$. 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 $\text{Ratio}_{h_0,h_1}(T_{s_1}) = \frac{\pi_{h_1}(T_{s_1})}{\pi_{h_0}(T_{s_1})} > \frac{\pi_{h_1}(T_{s_2})}{\pi_{h_0}(T_{s_2})} = \text{Ratio}_{h_0,h_1}(T_{s_2})$.

Therefore, we can find sufficiently small $\epsilon, \delta > 0$, and replace ϵT_{s_2} by δT_{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 theorem. □