Optimal Statistical Hypothesis Testing for Social Choice

Lirong Xia *
Rensselaer Polytechnic Institute, RPI, Troy NY 12180, USA
xial@cs.rpi.edu

Abstract

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

1 INTRODUCTION

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

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

However, the Jury Theorem does not identify the *optimal* decision-making rule, especially when there are three alternatives or more. From a statistical point of view, defining the optimality measure is highly nontrivial and controversial. If we use likelihood of a parameter as the measure, then we may pursue the *likelihoodist* approach. If we view the ground truth parameter as a random variable, and use expected loss w.r.t. the posterior distribution over the parameters as the measure, then we may pursue the *Bayesian* approach. If we believe that the ground truth is deterministic and unknown, and want to measure the performance of a given rule, then we may pursue the *frequentist* approach.[1] At a high level, the frequentist approach tries to measure and design the most *robust* rule, as Efron [2005] noted: "*a frequentist is a Bayesian trying to do well, or at least not too badly, against any possible prior distribution*".

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

*This work is supported by NSF #1453542 and #1716333, and ONR #N00014-17-1-2621. We thank all reviewers for helpful suggestions.

[1]The three approaches differ in philosophy of probability and measure of rules. The same rule, for example the MLE (MAP with uniform prior for Bayesians), might be used in all three approaches due to its optimality w.r.t. the three measures under certain conditions.

Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), PMLR volume 124, 2020.

rejected and 0 means that there is a lack of evidence to reject H_0. We note that the role of H_0 and H_1 are not the same, namely a A vs. B test is different from a B vs. A test.

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

Our Contributions. We answer the question of optimal hypothesis testing for social choice by characterizing *uniformly most powerful (UMP)* tests for various combinations of H_0 and H_1 for winner determination under two popular models for rank data: Mallows' model and Condorcet's model. UMP is a strong notion of optimality for hypothesis testing. A test f is evaluated by two criteria: its *size* (or *level of significance*), which is its worst-case probability to wrongly reject H_0, and its *power*, which is its probability to correctly reject H_0. The power of a test is evaluated at each $h_1 \in H_1$. A level-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.

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

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

Table 1: UMP tests for Mallows' model and Condorcet's model. "Y" in Condorcet-Winner means that for any $0 < \alpha < 1$, there exists a level-α UMP winner test for $\overline{H}_1$ vs. H_1. "Y&N" means that for some α, no level-α UMP test exists for $H_0 = \overline{H}_1$; but a UMP test exists for all levels for some natural special cases.

For example, "Y&N" in Mallows-Non-winner in Table 1 means that for some α, no level-α UMP test exists for H_0 vs. $\overline{H}_0$, 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 2 characterizes all such H_1's.

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

[2]This setting is called a "non-winner test" because when H_0 is rejected, a should not be selected as the winner.

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

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

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

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

2 PRELIMINARIES

Let $\mathcal{A} = \{a_1, \dots, a_m\}$ denote a set of $m \geq 2$ *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. Let P_n denote the collection of n agents' votes, 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 (WMG)* 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 $w_P(a \succ b) = P(a \succ b) - P(b \succ a)$. By definition $w_P(a \succ b) = -w_P(b \succ a)$. For example, the WMG of the profile P_7 of seven linear orders mentioned in the beginning of Introduction has weights $w_{P_7}(a \succ b) = w_{P_7}(a \succ c) = 1$ and $w_{P_7}(b \succ c) = 5$.

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 finitely many 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. Formally,

$$\text{KT}(V, W) = \# \left\{ \begin{array}{r} \{a, b\} \subseteq \mathcal{A} : [a \succ_V b \text{ and } b \succ_W a] \\ \text{or } [b \succ_V a \text{ and } a \succ_W b] \end{array} \right\}$$

Definition 1 (Mallows' model with fixed dispersion [Mallows, 1957]) *Given the dispersion* $0 < \varphi < 1$, Mallows' model *is denoted by* $\mathcal{M}^{Ma} = (\mathcal{L}(\mathcal{A})^n, \mathcal{L}(\mathcal{A}), \vec{\pi})$*, 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.*

Condorcet's model differs from Mallows' model by allowing ties in the ground truth and in data. Formally, 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 is extended to $\mathcal{B}(\mathcal{A})$ by counting the number of pairwise disagreements.

Definition 2 (Condorcet's model for binary relations with fixed dispersion) *Given the dispersion* $0 < \varphi < 1$, Condorcet's model *is denoted by* $\mathcal{M}^{Co} = (\mathcal{B}(\mathcal{A})^n, \mathcal{B}(\mathcal{A}), \vec{\pi})$*, where the parameter space is* $\mathcal{B}(\mathcal{A})$ *and for any* $W \in \mathcal{B}(\mathcal{A})$ *and* $V \in \mathcal{B}(\mathcal{A})$*,* $\pi_W(V) = \frac{1}{Z}\varphi^{KT(V,W)}$*, where* Z *is the normalization factor.*

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

Statistical Hypothesis Testing: The Neyman-Pearson Framework. Given a statistical model $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$, 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 1 *Let* $\mathcal{M}^{Ma}$ *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* H_0 *denote the other rankings. Let* f *be a test* *where* $f(1 \succ 2 \succ 3) = 1, f(2 \succ 1 \succ 3) = f(1 \succ 3 \succ 2) = 0.5$*, and* f *outputs* 0 *for all other rankings. We have* $\textit{Size}(f) = \textit{Size}(f, 2 \succ 1 \succ 3) = \textit{Size}(f, 1 \succ 3 \succ 2) = (0.5+\varphi+0.5\varphi^2)/Z$*, where* Z *is the normalization factor.* $\textit{Power}(f, 1 \succ 2 \succ 3) = (1+\varphi)/Z$*.*

Given a statistical model $\mathcal{M}$, H_0, a 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. For finite H_0, a most powerful test always exists and may not be unique. For composite H_1, it is possible that for different $h_1 \in H_1$, the most powerful tests are different. If there exists a level-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.

For simple H_0 vs. simple H_1, that is, $|H_0| = |H_1| = 1$, the fundamental lemma of Neyman and Pearson characterizes the most powerful tests as *likelihood ratio tests*, defined as follows.

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* $\textit{Ratio}_{h_0,h_1}(P) = \frac{\pi_{h_1}(P)}{\pi_{h_0}(P)}$ *denote the* likelihood ratio *of* P *and let*

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

denote the level-α likelihood ratio test*, where* $k_\alpha \geq 0$ *and* γ_α *are chosen such that* $\textit{Size}(\textit{LR}_{\alpha,h_0,h_1}) = \alpha$*.*

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* $\textit{LR}_{\alpha,h_0,h_1}$ *is a level-*α *most powerful test. Moreover, any most powerful test must agree with* $\textit{LR}_{\alpha,h_0,h_1}$ *except on* $P \in \mathcal{S}$ *with* $\textit{Ratio}_{h_0,h_1}(P) = k_\alpha$*.*

Example 2 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,

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

■

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

Definition 4 *For any* $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$*, any* $H_0 \subseteq \Theta$*, and any* $h_1 \in (\Theta \setminus H_0)$*. Let* Λ *denote a distribution over* H_0 *whose support set is denoted by* $\textit{Spt}(\Lambda)$*, 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* $\textit{Ratio}_{\Lambda,h_1}(P) = \frac{\pi_{h_1}(P)}{\sum_{h_0 \in H_0} \Lambda(h_0)\pi_{h_0}(P)}$*, and*
- *let* $\textit{LR}_{\alpha,\Lambda,h_1}(P)$ *denote the likelihood ratio test for* h_0^Λ *vs.* h_1 *as in Definition 3.*

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

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

(i) For any $h_0^* \in \textit{Spt}(\Lambda)$*,* $\textit{Size}(\textit{LR}_{\alpha,\Lambda,h_1}, h_0^*) = \alpha$*.*

(ii) For any $h_0 \in H_0$*,* $\textit{Size}(\textit{LR}_{\alpha,\Lambda,h_1}, h_0) \leq \alpha$*.*

Moreover, if there is no $P \in \mathcal{S}$ *with* $\textit{Ratio}_{\Lambda,h_1}(P) = k_\alpha$*, then* $\textit{LR}_{\alpha,\Lambda,h_1}$ *is the unique level-*α *most powerful test.*

The distribution Λ in Lemma 2 is called a *least favorable distribution*. If $|\text{Spt}(\Lambda)| = 1$, 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 TEST SETUP AND BASIC LEMMAS

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

Definition 5 ((Non-)Winner Tests) *Given an alternative* a*, in a* non-winner *test for Mallows' model, we let* $H_0 = L_{a \succ \textit{others}}$*; and in a* winner *test for Mallows' model, we let* $H_1 = L_{a \succ \textit{others}}$*.*

Given an alternative a, in a non-winner *test for Condorcet's model, we let $H_0 = R_{a \succ others}$; and in a* winner *test for Condorcet's model, we let $H_1 = R_{a \succ others}$.*

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

We now present two general lemmas on least favorable distributions that will be frequently used in this paper. For any model $\mathcal{M} = (\mathcal{S}, \Theta, \vec{\pi})$, 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 $X^{\Lambda}_{h_0} : \mathcal{S} \to \mathbb{R}$ such that for any $P \in \mathcal{S}$, $\Pr(P) = \pi_{h_0}(P)$ and $X^{\Lambda}_{h_0}(P) = \log \text{Ratio}_{\Lambda,h_1}(P)$. A random variable X *weakly first-order stochastically dominates* (*weakly dominates* for short) another random variable Y, if for all $p \in \mathbb{R}$, $\Pr(X \geq p) \geq \Pr(Y \geq p)$.

Lemma 3 *Λ is a uniformly least favorable distribution for H_0 vs. h_1 if and only if for any $h^*_0 \in Spt(\Lambda)$ 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

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

The "if" direction: for any $h_0 \in H_0$ and any $h^*_0 \in \text{Spt}(\Lambda)$, because $X^{\Lambda}_{h^*_0}$ weakly dominates $X^{\Lambda}_{h_0}$, we have that for any $x \in \mathbb{R}$, $\Pr(X^{\Lambda}_{h^*_0} \geq x) \geq \Pr(X^{\Lambda}_{h_0} \geq x)$. It follows from (1) that $\text{Size}(\text{LR}_\alpha, h_1), h^*_0) \geq \text{Size}(\text{LR}_\alpha, h_0)$. 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 $h^*_0 \in \text{Spt}(\Lambda)$ be such that $X^{\Lambda}_{h^*_0}$ does not weakly dominate $X^{\Lambda}_{h_0}$. It follows that there exists $x \in \mathbb{R}$ such that $\Pr(X^{\Lambda}_{h^*_0} \geq x) < \Pr(X^{\Lambda}_{h_0} \geq x)$. Let $\alpha = \Pr(X^{\Lambda}_{h^*_0} \geq x)$. Because 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. □

Example 3 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 Lemma 3 to prove that Λ is a uniformly least favorable distribution for testing $H_0 = (\mathcal{L}(\mathcal{A}) - \{[1 \succ 2 \succ 3]\})$ 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	$1 \succ 2 \succ 3$
$\text{Ratio}_{\Lambda,1\succ2\succ3}(V)$:	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.

	$\log \varphi$	$\log \frac{2\varphi}{1+\varphi^2}$	$\log \frac{1}{\varphi}$
$X^{\Lambda}_{1\succ3\succ2}$ and $X^{\Lambda}_{2\succ1\succ3}$	$\frac{\varphi^2}{Z}$	$\frac{1+\varphi+\varphi^2+\varphi^3}{Z}$	$\frac{\varphi}{Z}$
$X^{\Lambda}_{2\succ3\succ1}$ and $X^{\Lambda}_{3\succ1\succ2}$	$\frac{\varphi}{Z}$	$\frac{1+\varphi+\varphi^2+\varphi^3}{Z}$	$\frac{\varphi^2}{Z}$
$X^{\Lambda}_{3\succ2\succ1}$	$\frac{1}{Z}$	$\frac{2(\varphi+\varphi^2)}{Z}$	$\frac{\varphi^3}{Z}$

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 $X^{\Lambda}_{1\succ3\succ2}$ and $X^{\Lambda}_{2\succ1\succ3}$ weakly dominate other random variables. By Lemma 3, Λ is a uniformly least favorable distribution. ■

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

Lemma 4 *Suppose 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.*

All missing proofs can be found in the appendix.

4 UMP TESTS FOR MALLOWS

In this section, we present results on UMP non-winner and winner tests for Mallows' model.

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

For any profile P, any $B \subset \mathcal{A}$, 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 WMG(P).

Theorem 1 (A most powerful non-winner test for Mallows) *Given a Mallows' model $\mathcal{M}^{Ma}$, for any alternative a, any ranking h_1 where a is not ranked at the top, any $0 < \alpha < 1$, and any n, the following test is a level-α UMP for testing $L_{a\succ others}$ vs. h_1. 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 B is the set of alternatives ranked above a in h_1, and K_α and Γ_α are chosen s.t. the size of $f_{\alpha,a,B}$ is α.

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

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

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

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

It follows from Lemma 5 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 Lemma 3. We note that Λ is deterministic. Therefore, by Lemma 4, 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 $w_{P_n}(B \succ a) = K_\alpha$. □

Theorem 1 can be extended to the following characterization of all UMP non-winner tests ($H_0 = L_{a\succ\text{others}}$) for Mallows' model. For any $B \subset \mathcal{A}$ and $a \in (\mathcal{A} \backslash B)$, 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 2 (Characterization of UMP non-winner tests for Mallows) *Given a Mallows' model $\mathcal{M}^{Ma}$ with $m \geq 2$ and $n \geq 2$, there exists a UMP test for $H_0 = L_{a\succ others}$ vs. H_1 for all $0 < \alpha < 1$ if and only if there exists $B \subseteq \mathcal{A}$ such that $H_1 \subseteq L_{B\succ a}$.*

Moreover, when $H_1 \subseteq L_{B\succ a}$, we have that $f_{\alpha,a,B}$ as defined in Theorem 1 is a UMP test.

Example 4 Let P_7 denote the profile mentioned in the beginning of Introduction. Suppose we want to test whether there is enough evidence to claim that a cannot be the winner. We can apply a non-winner test on a by letting $H_0 = L_{a\succ \text{others}}$ and $H_1 = L_{\text{others}\succ a}$. By Theorem 2, $f_{\alpha,a,B}$ is a UMP test, where $B = \{b, c\}$. The test can be done by computing the test statistic $\mathcal{T} = w_{P_7}(B \succ a) = -2$, and then checking if $\mathcal{T}$ is in the critical region (K_α, ∞) for some pre-computed K_α. If $\mathcal{T} \in (K_\alpha, \infty)$, then H_0 is rejected, which means that a should not be chosen as the winner. If $\mathcal{T} = K_\alpha$, then H_0 is rejected with probability Γ_α. Otherwise H_0 cannot be rejected, meaning that there is not enough evidence to claim that a cannot be the winner. ■

Winner Tests for Mallows. We now consider UMP winner tests under Mallows' model ($H_1 = L_{a\succ \text{others}}$) for two natural choices of H_0: $H_0 = L_{\text{others}\succ a}$ in Theorem 3, which means that a is ranked in the bottom in the ground truth, and $H_0 = (\mathcal{L}(\mathcal{A}) - H_1)$ in Theorem 4 and 5, which means that a is not ranked at the top in the ground truth.

Theorem 3 (A UMP winner test under Mallows) *Given a Mallows' model $\mathcal{M}^{Ma}$, for any alternative a, any $0 < \alpha < 1$, and any n, the following test is a level-α UMP for testing $H_0 = L_{others\succ a}$ vs. $H_1 = L_{a\succ others}$. For any n-profile P_n,*

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

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

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

Because $f_{\alpha,a}$ is invariant to permutations over $\mathcal{A} \setminus \{a\}$, for any $h_0' \in H_0$ and any permutation M over $\mathcal{A} \setminus \{a\}$, 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, by letting Λ to be the deterministic distribution on $\{h_0^*\}$. □

Example 5 Let us continue with the setting in Example 4. Suppose we want to test whether there is enough evidence to claim that a is the winner. We can apply a winner test on a by letting $H_0 = L_{\text{others}\succ a}$ and $H_1 = L_{a\succ \text{others}}$, i.e. switching the roles of H_0 and H_1 in Example 4. By Theorem 3, $f_{\alpha,a}$ is a UMP test. The test can be done by computing the test statistic $\mathcal{T} = w_{P_7}(a \succ \text{others}) = 2$, and then checking if $\mathcal{T}$ is in the critical region (K_α^*, ∞) for some pre-computed K_α^*. If $\mathcal{T} \in (K_\alpha^*, \infty)$, then H_0 is rejected, which means that a should be chosen as the winner. If $\mathcal{T} = K_\alpha^*$, then H_0 is rejected with a pre-computed probability Γ_α^*. Otherwise H_0 cannot be rejected, meaning that there is no enough evidence to claim that a is the winner. ■

The following two theorems identify conditions on φ in Mallows' model for the UMP winner test $H_0 = (\mathcal{L}(\mathcal{A}) \setminus H_1)$ vs. $H_1 = L_{a\succ \text{others}}$ when $n = 1$.

Theorem 4 *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 5 *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

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

Definition 6 (Combining two models) *Given two models $\mathcal{M}_X = (\mathcal{S}_X, \Theta_X, \vec{\pi}_X)$ 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 let $(\pi_{\theta_X}, \pi_{\theta_Y})(P_X, P_Y) = \pi_{\theta_X}(P_X) \cdot \pi_{\theta_Y}(P_Y)$.*

Example 6 *Given a Condorcet's model $\mathcal{M}^{Co}$ 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 $\mathcal{M}^{Co}$ on the pairwise comparison between a and b. We have $\mathcal{M}^{Co} = \mathcal{M}_{\{1,2\}} \otimes \mathcal{M}_{\{2,3\}} \otimes \mathcal{M}_{\{1,3\}}$.* ■

Given two models $\mathcal{M}_X$ and $\mathcal{M}_Y$, the next theorem provides a way to leverage a least favorable distribution for a composite vs. simple test under $\mathcal{M}_X$ to a least favorable distribution for a composite vs. simple test under the combined model $\mathcal{M}_X \otimes \mathcal{M}_Y$.

Lemma 6 *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 $\mathcal{M}_X$. For any $y_1 \in \Theta_Y$, 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 7 Continuing Example 6, we let $\mathcal{M}_X = \mathcal{M}_{\{1,2\}}$, $H_{0,X} = \{0\}$, $x_1 = 1$, let Λ_X be the deterministic distribution over $\{0\}$, 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 (Lemma 1). Let Λ^* denote the deterministic distribution over $\{(0,1,1)\}$. It follows from Lemma 6 that Λ^* is a least favorable distribution for $(\{0\} \times \{0,1\}^2)$ vs. $(1,1,1)$ under Condorcet's model. ■

The next theorem focuses on the setting where we combine $t \in \mathbb{N}$ identical statistical models $\mathcal{M}_X$. 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 $h_1 \in (\Theta \setminus H_0)$, we let $\vec{h}_1 = \underbrace{(h_1, \ldots, h_1)}_{t}$ and let $\text{Ext}(H_0, h_1, t) = (\{H_0 \cup \{h_1\}\}^t \setminus \{\vec{h}_1\})$.

Example 8 In the setting of Example 6, we let $\mathcal{M}_X = \mathcal{M}_{\{1,2\}}$, let Λ denote the deterministic distribution over $\{0\}$, 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 $\text{Ext}(H_0, 1, 3) = (\{0,1\}^3 \setminus \{(1,1,1)\})$. ■

Lemma 7 *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 Ext*(Λ, h_1, t) *is a uniformly least favorable distribution for Ext*(H_0, h_1, t) *vs.* $\vec{h}_1$ *in* $(\mathcal{M}_X)^t$.

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

Non-Winner Tests for Condorcet. We are now ready to characterize UMP tests for Condorcet's model by applying Lemma 6 and 7. Theorem 6 and Theorem 7 of this section are counterparts of Theorem 1 and Theorem 2 (both are for Mallows' model), respectively, though the proof techniques are quite different.

Theorem 6 (A most powerful non-winner test for Condorcet) *Given a Condorcet's model* $\mathcal{M}^{Co}$ *with* $m \geq 2$, *for any* $a \in \mathcal{A}$, *any* $h_1 \in (\mathcal{B}(\mathcal{A}) \setminus R_{a\succ \text{others}})$, *any* n, *and any* φ, *the following test is most powerful for testing* $R_{a\succ \text{others}}$ *vs.* h_1. *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 *is the set of alternatives that are preferred to* a *in* h_1.

Proof: Let h_0^* denote the binary relation obtained from h_1 by enforcing $a \succ b$ for all $b \in \mathcal{A}$. We will prove that the deterministic distribution over $\{h_0^*\}$ is a uniformly least favorable distribution for $R_{a\succ\text{others}}$ vs. h_1.

Let $X = \{\{a,b\} : b \neq a\}$ denote the pairwise comparisons between alternatives in $\mathcal{A}$ that involve a and let Y denote the set of all 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 Lemma 6, 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. The theorem follows after Lemma 2. □

Subsequently, we have the following characterization of UMP non-winner tests under Condorcet's model ($H_0 = R_{a\succ\text{others}}$). 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 7 (Characterization of UMP non-winner tests for Condorcet) *Let* $\mathcal{M}^{Co}$ *denote a Condorcet's model with any* $m \geq 2$ *and* $n \geq 2$. *There exists a UMP test for* $H_0 = R_{a\succ \text{others}}$ *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, when $H_1 \subseteq R_{B\succ a}$, $g_{\alpha,a,B}$ defined in Theorem 6 is a UMP test.

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

Winner Tests for Condorcet. Finally, we turn to UMP winner tests for Condorcet's model ($H_1 = R_{a\succ\text{others}}$).

Theorem 8 (A UMP winner test for Condorcet) *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 $H_0 = (\mathcal{B}(\mathcal{A}) \setminus H_1)$ vs. $H_1 = R_{a\succ\text{others}}$. 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 Condorcet's model with a single sample. Let $X_1, \ldots, X_{m-1}$ denote the $m-1$ pairwise comparisons between a and other alternatives. Similarly to the proof of Theorem 6, we let $\mathcal{M}_{X_1}, \ldots, \mathcal{M}_{X_{m-1}}$ denote the restriction of $\mathcal{M}_1$ 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 $\mathcal{M}^{\text{Co}} = \mathcal{M}_{X_1} \otimes \mathcal{M}_{X_2} \otimes \cdots \otimes \mathcal{M}_{X_{m-1}} \otimes \mathcal{M}_Y$.

In $\mathcal{M}_{X_1}$, let 1 represent that a is more preferred in the pairwise comparison. Due to the Neyman-Pearson lemma (Lemma 1), the deterministic distribution $\Lambda = \{0\}$ is a uniformly least favorable distribution for $H_0 = \{0\}$ vs. $h_1 = 1$. For any $n \in \mathbb{N}$, let $\mathcal{M}_{X_{1,n}}$ denote $\mathcal{M}_{X_1}$ with n i.i.d. samples. It follows from Lemma 4 that Λ is still a uniformly least favorable distribution for $\mathcal{M}_{X_{1,n}}$. By Lemma 7, $\text{Ext}(\Lambda, h_1, m-1)$ is a uniformly least favorable distribution for $\text{Ext}(H_0, h_1, m-1) = (\{0,1\}^{m-1} \setminus \{\vec{1}\})$ 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 Lemma 4, 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 $\text{Ext}(H_0, h_1, m-1) \times \Theta_{Y,n} = (\{0,1\}^{m-1} \setminus \{\vec{1}\}) \times \Theta_Y = (\mathcal{B}(\mathcal{A}) \setminus R_{a\succ\text{others}})$. 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 $(\mathcal{B}(\mathcal{A}) \setminus R_{a\succ\text{others}})$ 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. The theorem is proved after noticing that $g_{\alpha,a} = \text{LR}_{\alpha,\Lambda_{y_1},(\vec{1},y_1)}$. □

6 DISCUSSION: BEYOND BINARY CHOICE

All UMP tests we have characterized so far are optimal in making binary decisions, such as whether a given alternative a is the winner. We propose two natural procedures to choose the winner by combining multiple winner tests ($H_1 = L_{a\succ\text{others}}$ 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).

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

Procedure based on combining non-winner tests. Similarly, we use binary search on 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, both procedures correspond to the Borda voting rule when the proposed UMP tests for Mallows' model are used: in the UMP winner test we let $H_0 = L_{\text{others}\succ a}$ vs. $H_1 = L_{a\succ\text{others}}$ as in Example 5, and in the UMP non-winner test we let $H_0 = L_{a\succ\text{others}}$ vs. $H_1 = L_{\text{others}\succ a}$ as in Example 4. This provides a new theoretical justification for the Borda rule; or vice versa, Borda provides a justification of the proposed procedure.

7 FUTURE WORK

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

[3]Co-winners exist if they all reject H_0 for the same α.

References

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

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

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

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

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

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

Bradley Efron. Bayesians, Frequentists, and Scientists. *Journal of the American Statistical Association*, 100 (1–5), 2005.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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