Statistical Properties of Social Choice Mechanisms

Lirong Xia

Abstract

Most previous research on statistical approaches to social choice focused on the computation and characterization of maximum likelihood estimators (MLE) of various parametric ranking models. In this paper, we take a traditional statistical approach by evaluating social choice mechanisms w.r.t. two important statistical criteria: (1) *consistency*, a minimum requirement for reasonable estimators, and (2) *minimaxity*, a well-accepted optimality criterion.

For consistency, we propose a new and general class of social choice mechanisms called *generalized outcome scoring rules (GOSR)* that include many commonly studied social choice mechanisms. Given any ranking model, we fully characterize all GOSRs that are consistent estimators of it, and derive an upper bound on the convergence rate. We also showed that the bound is asymptotically tight for some GOSR and ranking model. This allows us to fully characterize all GOSRs that are consistent w.r.t. *some* model. For minimaxity, we characterize a class of minimax estimators for *neutral* ranking models. As a corollary, the uniformly randomized MLE has the highest probability to correctly reveal the ground truth among all estimators for a number of natural ranking models.

1 Introduction

Social choice theory studies how to combine agents' opinions or preferences to a joint decision. Traditional social choice theory concerns how to reach a consensus among subjective preferences of the agents, and evaluates these mechanisms based on agents' *subjective* satisfaction of the joint decision. Ideally we would like to make a joint decision that every agent is happy about, which is often impossible due to conflicting preferences of the agents. A typical example of this scenario is presidential elections.

In many other scenarios, our goal is to aggregate agents' opinions and preferences to figure out the *ground truth*. For example, online retailers (including Amazon.com) aggregate reviewers' ratings of an item to provide an estimation to the ground truth quality of this item. In such settings, instead of making a joint decision (e.g. an aggregated score) to make agents (e.g. reviewers of an item) happy, we want to make a joint decision that is evaluated w.r.t. the *objective* quality of the decision.

To reveal the ground truth by aggregating agents' preferences, it is natural to take a statistical approach. This can be dated back to the *Condorcet Jury Theorem* in the 18th century [11], which states that fix $p > 1/2$, for two alternatives, suppose agents' votes are generated i.i.d. such that each agent has probability p to be correct, then the majority aggregation of agents' votes converges to the ground truth as the number of agents goes to infinity.

Let us take a closer look at the framework of the Condorcet Jury Theorem. It has an outcome space, which is also the parameter space and consists of all possible outcomes of the voting rule, i.e. the two alternatives. It assumes a probability distribution over agents' votes for each outcome (alternative) assuming that the alternative is the ground truth. Then, the majority rule is evaluated by the probability that it successfully reveals the ground truth for randomly-generated votes, and justified by showing that as the number of agents goes to infinity, such probability goes to 1. In modern statistical terms, Condorcet proposed a statistical parametric model to capture the random generation of votes, and proved that the majority rule, as a statistical estimator, satisfies an important statistical property called *consistency*.

Most previous work on statistical approaches towards social choice focus on the computation and characterization of the *maximum likelihood estimators (MLEs)* of various ranking models. For example, Conitzer and Sandholm [13] investigated commonly studied voting rules that are MLEs of *some* statistical models. Conitzer et al. [14] further examined this relationship for preference functions (social choice mechanisms that output rankings). MLE inference algorithms for extensions of Condorcet's model, the Mallows model, and random utility models have been studied in [27, 24, 20, 1, 3]; Braverman and Mossel [7] proposed an algorithm that computes the MLE of the Mallows model with high probability; Elkind et al. [15] studied MLE voting rules that are distance-rationalizable; Caragiannis et al. [8] justified uniformly randomized MLE (that is, the estimator that uniformly selects an outcome with the maximum likelihood) under the Mallow model by proving that it has the smallest sample complexity among all estimators.

Despite the great interests in studying MLEs in social choice applications, the usage of MLEs in the first place seems to lack a statistical theoretical justification. It is well-known that for continuous parameter space, e.g. the space of all real numbers, MLEs satisfy many desired statistical properties including *asymptotic normality* and *asymptotic efficiency* (i.e. they have the lowest asymptotic variance among all unbiased estimators) [16, Theorem 14.1], which justifies the (asymptotic) optimality of MLEs. However, in many commonly studied ranking models in social choice, the parameter space is finite and thus discrete, which means that MLEs are not automatically justified as "optimal" estimators.

More importantly, while it is an important first step to understand properties of MLEs of various statistical models for social choice, as done in many previous works, it is arguably more important to move on to tackle the social choice problem by designing novel mechanisms that *best* reveal the ground truth. This can be achieved in two interrelated directions: (1) propose new statistical models to better capture agents' preferences [27, 24, 8, 25, 12], and (2) learn "optimal" estimators for existing models and use them as social choice mechanisms [20, 2, 5]. The first direction was mainly pursued in the computational social choice community while the second was mainly pursued in the machine learning community, famously known as "learning to rank" [19].

The topic of this paper is along the second direction. We take a traditional statistical approach towards social choice mechanisms to study mechanisms beyond MLEs, and evaluate them w.r.t. well-established statistical properties. This is non-trivial due to the finite/discrete parameter space in popular ranking models and the structure of data, which are rankings over alternatives. We focus on two well-known statistical properties, i.e. *consistency* and *minimaxity*. Consistency is often regarded as a minimum requirements for reasonable estimators, which states that as the data size goes to infinity, the output of the estimator is the same as the ground truth with probability 1. In fact, consistency is exactly the property proved in the Condorcet Jury theorem for the majority rule. *Minimaxity* is a well-accepted optimality criterion for estimators, which states that the estimator always minimizes the worst-case frequentist loss, where the worst case is taken over all parameters used to generate data (see Definition 5).

Our technical contributions are two-fold. First, we propose a new and general class of social choice mechanisms called *generalized outcome scoring rules (GOSR)*, which are natural extensions of *generalized scoring rules* [26] to an arbitrary outcome spaces, including a winner, multiple winners, and rankings. We think GOSRs are interesting in their own right, but in this paper they mainly serve as a framework to obtain general results on consistency. For any given parametric ranking model, we characterize all GOSRs that are consistent estimators w.r.t. it, and derive bounds on the convergence rate for the output of a GOSR to reveal the ground truth; this allows us to investigate a similar question asked by Conitzer and Sandholm [13] and fully characterize all GOSRs that are consistent w.r.t. *some* parametric ranking models. For minimaxity, we prove that the *uniform Bayesian estimator* is minimax for *neutral* parametric ranking models (with some loss functions). As a corollary, for any fixed number of agents and any neutral parametric ranking model, the uniformly randomized MLE reveals the ground truth with the highest probability among all (deterministic and randomized) estimators, which shows the optimality of the uniformly randomized MLE. The implication of these two theoretical results on the design of new social choice mechanisms is that given a parametric ranking model, we know which GOSRs are reasonable estimators (i.e. they satisfy consistency), and which estimators (not necessarily GOSRs) are optimal (i.e. they satisfy minimaxity).

1.1 Related Work

Minimax estimators for various statistical models with continuous parameter spaces have been characterized by Berger [6]. Statistical inference under models with discrete parameter space has attracted a lot of attention recently, see [10] and references therein. Specifically, Choirat and Seri [10] provided a sufficient condition on discrete-parameter models for MLEs to be minimax. However, their work focused on deterministic estimators, while our work focused on the minimaxity for randomized estimators.

In social choice, the closest previous work to ours is by Caragiannis et al. [8], who focused on the Mallows model and other distance-based models. Even though Caragiannis et al. did not explicitly mention consistency and minimaxity, they essentially proved that many commonly studied voting rules are consistent estimators w.r.t. the Mallows model. They also obtained fine-grained convergence rates, and proved the maximality for the uniformly randomized MLE under the Mallows model. We take a different approach towards consistency and minimaxity, and our results on consistency and minimaxity are for all *neutral* model, including the Mallows model.

More recently, Caragiannis et al. [9] discovered that there is a unique preference function called *modal rank* within a large family of voting rules that is consistent w.r.t. all *d-monotonic* models. In this paper we focused on consistency and minimaxity for any parametric ranking model.

Our work is also related to a recent work by Pivato [23], who studied voting rules that can be viewed as MLEs, maximum *a posteriori* estimators (MAP), and expected utility maximizers, which is closely related to the Bayesian estimators studied in this paper. Our work has a different focus: we are interested in statistical properties of estimators, while Pivato focused on understanding the mathematical structures of these voting rules.

Lastly, for social choice problems where the parameter space of the statistical model is different from the outcome space (e.g. when we want to use Mallows model to model agents' preferences but want to select a single winner based on agents' preferences), we refer to an ongoing work on applying *statistical decision theory* to social choice [4], in which the framework is more general but the results there are significantly different from the results in this paper.

2 Preliminary

Let $\mathcal{C} = \{c_1, \ldots, c_m\}$ denote a set of m *alternatives* and let $\mathcal{L}(\mathcal{C})$ denote the set of all linear orders over $\mathcal{C}$. Each agent uses a linear order in $\mathcal{L}(\mathcal{C})$ to represent her preferences, called her *vote*. The collection of all agents' votes P is called a *profile*. Let $\mathcal{L}(\mathcal{C})^* = \mathcal{L}(\mathcal{C}) \cup \mathcal{L}(\mathcal{C})^2 \cup \cdots$ denote the set of all profiles.

Let $\mathcal{O}$ denote the set of *outcomes*. A (deterministic) *voting rule* r is a mapping that assigns to each profile a single outcome in $\mathcal{O}$. Common choices of $\mathcal{O}$ are: (1) $\mathcal{C}$, where voting rules are often called *resolute voting rules*; (2) $(2^{\mathcal{C}} \setminus \emptyset)$, where voting rules often are called *irresolute voting rules*; and (3) $\mathcal{L}(\mathcal{C})$, where voting rules are often called *preference functions* (a.k.a. *social welfare function*). A randomized voting rule assigns to each profile a probability distribution over $\mathcal{O}$.

Many commonly studied voting rules have resolute, irresolute, and preference function versions. For example, an irresolute *positional scoring rule* is characterized by a scoring vector $\vec{s} = (s_1, \ldots, s_m)$ with $s_1 \geq s_2 \geq \cdots \geq s_m$. For any alternative c and any linear order V, we let $\vec{s}(V, c) = s_j$, where j is the position of c in V. Given a profile P, the positional scoring rule chooses all alternatives c that maximize $\sum_{V \in P} \vec{s}(V, c)$, where P is viewed as a multi-set of votes. The resolute version of a positional scoring rule chooses a single alternative by further applying a

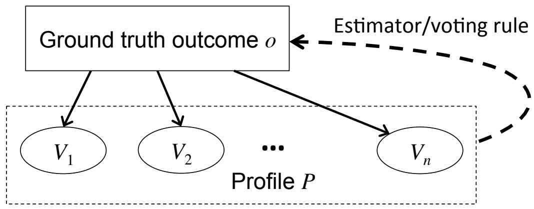

Figure 1: Voting rules as statistical estimators.

tie-breaking mechanism, and the preference function version ranks the alternatives w.r.t. their scores, and sometimes uses a mechanism to break ties.

As another example, the *single transferable vote (STV)* rule is naturally defined as a preference function that outputs a ranking in the following $m-1$ steps: in each step, the alternative ranked in the top positions least often is eliminated from the profile;[1] the outcome ranking is the inverse of the elimination order. The resolute version of STV simply outputs the top-ranked alternative in the winning ranking, and an irresolute version contains all alternatives that can be made the winner for some tie-breaking mechanisms (c.f. the *parallel-universes tiebreaking* [14]).

In the statistical approach towards social choice, we assume that agents' preferences are i.i.d. generated from a statistical model. Given the set of alternatives $\mathcal{C}$ and the number of agents n, a *parametric ranking model* $\mathcal{M}_\mathcal{C} = (\mathcal{O}, \vec{\pi})$ has two parts: a *parameter space* $\mathcal{O}$, and a set of distributions over $\mathcal{L}(\mathcal{C})$, denoted by $\vec{\pi} = \{\pi_o : o \in \mathcal{O}\}$. Throughout the paper, we let P_n denote an i.i.d.-generated profile of n votes from a distribution π_o that will be clear from the context.

In this paper, we focus on cases where the parameter space is exactly the same as the outcome space, and require that $\pi_o(V) > 0$ for all $V \in \mathcal{L}(\mathcal{C})$ and all $o \in \mathcal{O}$. When the parameter space is different from the outcome space, one can use *statistical decision theory* with a decision function to select an outcome as social choice [4]. Given $\mathcal{M}_\mathcal{C} = (\mathcal{O}, \vec{\pi})$, a deterministic *estimator* T is a function that maps each profile to a parameter (outcome) in $\mathcal{O}$; a randomized estimator T maps each profile to a distribution over $\mathcal{O}$. The statistical approach is illustrated in Figure 1.

Given a statistical model $\mathcal{M}_c = (\mathcal{O}, \vec{\pi})$, the *maximum likelihood estimator (MLE)* is a statistical estimator that outputs an outcome o with the maximum likelihood, that is, for any profile P, $MLE(P) \in \arg\max_o \prod_{V\in P} \pi_o(V)$. The *uniformly randomized MLE* chooses all outcomes in $\arg\max_o \prod_{V\in P} \pi_o(V)$ uniformly at random.

We recall two classes of parametric ranking models: the first is a variant of the *Mallows model* [21] and the second was introduced by [13] whose MLEs are positional scoring rules.

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

Definition 1 *Given* $0 < \varphi < 1$*, the* Mallows model with fixed dispersion parameter φ *is denoted by* $\mathcal{M}_\varphi = (\mathcal{O}, \vec{\pi})$*, where* $\mathcal{O} = \mathcal{L}(\mathcal{C})$*, and for any* $V, W \in \mathcal{L}(\mathcal{C})$*,* $\pi_W(V) = \frac{1}{Z}\varphi^{\textit{Kendall}(V,W)}$*, where* Z *is the normalization factor with* $Z = \sum_{V\in\mathcal{L}(\mathcal{C})} \varphi^{\textit{Kendall}(V,W)}$.

Definition 2 ([13]) *For any* $\vec{s} = (s_1, \ldots, s_m)$ *with* $s_1 \geq s_2 \geq \cdots \geq s_m$*, let* $\mathcal{M}_{\vec{s}} = (\mathcal{O}, \vec{\pi})$ *be the parametric ranking model where* $\mathcal{O} = \mathcal{C}$*, and for any* $V \in \mathcal{L}(\mathcal{C})$ *and any* $c \in \mathcal{C}$*,* $\pi_c(V) = \frac{1}{Z} \cdot 2^{\vec{s}(V,c)}$*, where* Z *is the normalization factor with* $Z = \sum_{V\in\mathcal{L}(\mathcal{C})} 2^{\vec{s}(V,c)}$.

Given a parameter space $\mathcal{O}$, a *loss function* $L(o, o')$ takes two elements in $\mathcal{O}$ as inputs, and outputs a real number representing the loss of mispredicting o as o'. A popular loss function is the 0-1 loss function $L_{\text{0-1}}$, where $L_{\text{0-1}}(o, o') = 0$ if and only if $o = o'$; otherwise $L_{\text{0-1}}(o, o') = 1$. The loss function can be naturally generalized to evalute the loss of a distribution over $\mathcal{O}$ w.r.t. a fixed o. There are mainly two ways to evaluate the expected loss of a (deterministic or randomized) estimator.

[1] In case there is a tie, we use some tie-breaking mechanism.

Definition 3 *Given a parametric ranking model $\mathcal{M}_\mathcal{C}$, a loss function L, and n, the* frequentist expected loss $R_F^n(o,T)$ *is a function of $o \in \mathcal{O}$ and an estimator T, such that $R_F^n(o,T) = \sum_P \pi_o(P)L(o,T(P))$, where the summation is taken for all profiles P with n votes.*

Given a prior distribution over $\mathcal{O}$, the Bayesian expected loss $R_B(P,T(P))$ *takes a profile P and $T(P) \in \mathcal{O}$ as inputs, and $R_B(P,T(P)) = \sum_{o\in\mathcal{O}} \Pr(o|P)L(o,T(P))$.*

In words, the frequentist expected loss is calculated for a fixed parameter $o \in \mathcal{O}$, such that we first generate the data (profile) from π_o, and then evaluate the expected loss of output of the estimator T w.r.t. o. The Bayesian expected loss is calculated for a fixed profile P, and is evaluated by the expected loss of the output $T(P)$ w.r.t. a random ground truth o drawn from the posterior distribution.

We now recall two well-known properties for statistical estimators.

Definition 4 ([16]) *An estimator T is* consistent *w.r.t. a parametric ranking model $\mathcal{M}_\mathcal{C}$ if for all $o \in \mathcal{O}$, $\lim_{n\to\infty} \Pr(T(P_n) = o) \to 1$, where votes in P_n are generated i.i.d. from π_o.*

That is, a consistent estimator correctly reveals the ground truth as the number of i.i.d. generated votes goes to infinity.

Definition 5 ([6]) *Given a parametric ranking model $\mathcal{M}_\mathcal{C} = (\mathcal{O}, \vec{\pi})$, a loss function L, and n, an estimator T is* minimax, *if $T \in \arg\min_{T^*} \max_{o\in\mathcal{O}} R_F^n(o,T^*)$.*

That is, a minimax estimator minimizes the maximum frequentist expected loss for $o \in \mathcal{O}$. In the above definition, we distinguish the minimaxity where T^* must be deterministic and the minimaxity where T^* can be any deterministic or randomized estimstor. For the Bayesian expected loss, the optimality is achieved by *Bayesian estimators*. That is, given a statistical model, a loss function, and a prior distribution over the parameters, T is a Bayesian estimator if and only if for all profile P, $R_B(P,T(P)) = \min_{o'\in\mathcal{O}} R_B(P,o')$.

Example 1 *For any parametric ranking model, the MLE is the Bayesian estimator for the* 0*-*1 *loss function and the uniform prior.*

3 Generalized Outcome Scoring Rules

To define generalized outcome scoring rules as a natural extension of *generalized scoring rules* [26], we need some notation. For any $K \in \mathbb{N}$, let $\mathcal{B}_K = \{b_1,\ldots,b_K\}$. A *total preorder* (*preorder* for short) is a reflexive, transitive, and total relation. Let Pre($\mathcal{B}_K$) denote the set of all preorders over $\mathcal{B}_K$. For any $\vec{p} \in \mathbb{R}^K$, we let Order($\vec{p}$) denote the preorder $\trianglerighteq$ over $\mathcal{B}_K$ where $b_{k_1} \trianglerighteq b_{k_2}$ if and only if $p_{k_1} \geq p_{k_2}$. That is, the k_1-th component of $\vec{p}$ is at least as large as the k_2-th component of $\vec{p}$. For any preorder $\trianglerighteq$, if $b \trianglerighteq b'$ and $b' \trianglerighteq b$, then we write $b =_\trianglerighteq b'$. Each preorder $\trianglerighteq$ naturally induces a (partial) strict order $\triangleright$, where $b \triangleright b'$ if and only if $b \trianglerighteq b'$ and $b' \ntrianglerighteq b$.

Definition 6 (Generalized outcome scoring rules) *Given an outcome space $\mathcal{O}$, $K \in \mathbb{N}$, $f : \mathcal{L}(\mathcal{C}) \to \mathbb{R}^K$ and g : Pre($\mathcal{B}_K$) $\to \mathcal{O}$, we define a* generalized outcome scoring rule (GOSR), *denoted by GOS$_{(f,g)}$, to be a mapping such that for any profile P, GOS$_{(f,g)}(P) = g($Order$(f(P)))$, where $f(P) = \sum_{V\in P}^{n} f(V)$.*

In words, a GOSR first uses the f function to transform the input profile P to a vector $f(P) = \sum_{V\in P} f(V)$ in $\mathbb{R}^K$, then use g to select the winner based on the *order* of the components in $f(P)$.

For any $V \in \mathcal{L}(\mathcal{C})$, $f(V)$ is called a *generalized scoring vector*, $f(P)$ is called a *total generalized score vector*. To simplify notation, we let Order$_f(P)$ = Order$(f(P))$. We note that Order$_f(P)$ is a preorder over $\mathcal{B}_K$, which means that it may not be a linear order. For any distribution π over $\mathcal{L}(\mathcal{C})$, we define $f(\pi) = \sum_{V\in\mathcal{L}(\mathcal{C})} \pi(V)f(V)$ and Order$_f(\pi)$ = Order$(f(\pi))$. In this paper, we assume that no component in generalized scoring vectors is *redundant* for better presentation. That is, for all $k_1 \neq k_2 \leq K$, there exists $V \in \mathcal{L}(\mathcal{C})$ such that $[f(V)]_{k_1} \neq [f(V)]_{k_2}$. This is without loss of generality because we can always remove the redundant component without changing the definition of the rule.

The next proposition shows that $\text{GOS}_{(f,g)}$ is a general class of voting rules, whose proof is by construction and resembles the proof for GSRs [26].

Proposition 1 *Generalized scoring rules are GOSRs with $\mathcal{O} = \mathcal{C}$. The irresolute versions and the ranking versions of positional scoring rules, Bucklin, Copeland, maximin, ranked pairs, STV are GOSRs, with $\mathcal{O} = (2^{\mathcal{C}} \setminus \emptyset)$ and $\mathcal{O} = \mathcal{L}(\mathcal{C})$, respectively.*[2]

We now introduce more notation to present the results.

Definition 7 (Extension of a preorder) *We say that $\trianglerighteq' \in Pre(\mathcal{B}_K)$ is an* extension *of $\trianglerighteq \in Pre(\mathcal{B}_K)$, if for all $b, b' \in \mathcal{B}_K$, we have $(b \rhd b') \Rightarrow (b \rhd' b')$. For any $\trianglerighteq, \trianglerighteq' \in Pre(\mathcal{B}_K)$, we let $\trianglerighteq \oplus \trianglerighteq'$ denote the preorder in $Pre(\mathcal{B}_K)$ obtained from $\trianglerighteq$ by using $\trianglerighteq'$ to break ties. That is, b_i is strictly preferred to b_j in $(\trianglerighteq \oplus \trianglerighteq')$ if and only if (1) $b_i \rhd b_j$, or (2) $b_i =_{\trianglerighteq} b_j$ and $b_i \rhd' b_j$.*

Definition 8 (Possible linear orders) *Give a generalized scoring function f, we define the set of* possible linear orders over $\mathcal{B}_K$, *denoted by $PL(f)$, to be the linear orders over $\mathcal{B}_K$ that are the orders of the total score vector of* some *profile. Formally, $PL(f) = \{Order_f(P) : P \in \mathcal{L}(\mathcal{C})^*\} \cap \mathcal{L}(\mathcal{B}_K)$.*

Definition 9 (Neighborhood) *For any $\trianglerighteq \in Order(\mathcal{B}_K)$, we define the* neighborhood *of $\trianglerighteq$ w.r.t. f, denoted by $Nbr_f(\trianglerighteq)$, to be all linear orders over $\mathcal{B}_K$ that can be obtained from $\trianglerighteq$ by using a linear order in $PL(f)$ to break ties. That is, $Nbr_f(\trianglerighteq) = \{\trianglerighteq \oplus \rhd^* : \rhd^* \in PL(f)\}$. Given f, the neighborhood of a distribution π, denoted by $Nbr_f(\pi)$, is the neighborhood of $f(\pi)$ w.r.t. f, that is, $Nbr_f(\pi) = Nbr_f(Order_f(\pi))$.*

We note that the definition of neighborhood does not involve the g function.

The next lemma characterizes the asymptotic behavior of $\text{Order}_f(P_n)$ and will be frequently used in the proof of the next section. It states that for any distribution π over $\mathcal{L}(\mathcal{C})$, if we generate votes in P_n i.i.d. from π, then $\text{Order}_f(P_n)$ asymptotically almost surely (a.a.s.) falls in the neighborhood of π w.r.t. f. As a corollary, $\text{Order}_f(P_n)$ is a linear order over $\mathcal{B}_K$ a.a.s. We recall that it is assumed that the generalized scoring vectors have no redundant components, which means that no components in the total generalized score vector are always equal.

Lemma 1 *Given a generalized scoring function f, for any distribution π that is positive everywhere on $\mathcal{L}(\mathcal{C})$, we have:*

(1) for any $\rhd \in Nbr_f(\pi)$, there exists a constant $\delta_{\rhd} > 0$ so that for sufficiently large n, $\Pr(Order_f(P_n) = \rhd) > \delta_{\rhd}$.

(2) for any $\trianglerighteq \notin Nbr_f(\pi)$, $\lim_{n\to\infty} \Pr(Order_f(P_n) = \trianglerighteq) = 0$.

Proof: We first illustrate the idea behind the proof in a very special case where $\text{Order}_f(\pi)$ is a linear order. In this case $\text{Nbr}_f(\pi) = \{\text{Order}_f(\pi)\}$. Then, the Central Limit Theorem tells us that for each linear order V, the frequency of V in P_n goes to $\pi(V)$ as $n \to \infty$, and the noise is $o(n^{-1})$. That is, with probability that goes to 1, votes in P_n are distributed as $\pi + o(n^{-1})\pi_{noise}$. Then, when n is sufficiently large, the π_{noise} part cannot affect $\text{Order}_f(\pi)$. Hence we have that with probability that goes to 1, $\text{Order}_f(P_n) = \text{Order}_f(\pi) \in \text{Nbr}_f(\pi)$.

The proof for the general case is more involved, because if $\text{Order}_f(P_\pi)$ is not a linear order, then the noise part π_{noise} acts as a tie-breaker and thus cannot be overlooked even for large n. Our main tool is to estimate the distribution of π_{noise} by the Multivariate Lindeberg-Lévy Central Limit Theorem (CLT) [16, Theorem D.18A], which states that for i.i.d. generated vector-valued random variables X_i, if the covariance matrix Σ for the components of X_i is nonsingular, then

[2] The definition of these rules can be found in [22].

$(\sum_{i=1}^{n} X_i - nE(X_i))/\sqrt{n}$ converges in probability to a multivariate normal distribution $\mathcal{N}(0, \Sigma)$. However, this theorem cannot be directly applied to analyze the asymptotic frequencies of the linear orders because the resulting covariance matrix is singular, since for any given n, the number of occurrences of all $m!$ linear orders must sum up to n, which means that they are linearly correlated.

Let $\mathcal{L}(\mathcal{C}) = \{l_1, \ldots, l_{m!}\}$ denote the set of all $m!$ linear orders. To avoid the singularity, our analysis will focus on $l_1, \ldots, l_{m!-1}$. For any $j \leq m! - 1$, let $\vec{v}_j$ denote the vector in $\{0, 1\}^{m!-1}$ where the j-th component is 1 and all other components are zeros. We then define i.i.d. multivariate random variables $X_1, \ldots, X_n$, where each X_i takes $\vec{v}_j$ with probability $\pi(l_j)$, and takes $\vec{0}$ with probability $\pi(l_{m!})$. It is not hard to verify that the mean of X_1 is $E(X_1) = (\pi(l_1), \ldots, \pi(m! - 1))$ and the covariance matrix is the following.

$$\Sigma_\pi = \begin{bmatrix} \pi(l_1) - \pi(l_1)^2 & -\pi(l_1)\pi(l_2) & \cdots -\pi(l_1)\pi(l_{m!-1}) \\ -\pi(l_2)\pi(l_1) & \pi(l_2) - \pi(l_2)^2 & \cdots -\pi(l_2)\pi(l_{m!-1}) \\ \vdots & & \vdots \\ -\pi(l_{m!-1})\pi(l_1) & -\pi(l_{m!-1})\pi(l_2) & \cdots \pi(l_{m!-1}) - \pi(l_{m!-1})^2 \end{bmatrix}$$

Since each diagonal element is strictly larger than the sum of the absolute values of other elements in the same row, Σ_π is non-singular according to the Levy-Desplanques Theorem [18]. Let $Y_n = X_1 + \ldots + X_n$. Each Y_n naturally corresponds to a profile P_n of n votes, where for all $j \leq m! - 1$, $[Y_n]_j$ is the number of occurrences of l_j, and $n - \sum_{j=1}^{m!-1} [Y_n]_j$ is the number of occurrences of $l_{m!}$. By the multivariate Central Limit Theorem, $Y_{noise} = \frac{Y_n - nE(X_1)}{\sqrt{n}}$ converges in distribution to the multivariate normal distribution $\mathcal{N}(0, \Sigma_\pi)$.

Part (1) of the lemma. For any $\rhd \in$ Nbr$_f(\pi)$, there exists a profile P such that (1) Order$_f(P) \in \mathcal{L}(\mathcal{C})$, and (2) Order$_f(\pi) \oplus$ Order$_f(P) = \rhd$. We define $\vec{p} \in \mathbb{R}^{m!}$ such that for all $j \leq m!$, $[\vec{p}]_j = P(l_j)/|P| - 1/m!$, where $P(l_j)$ is the number of occurrences of l_j in P. That is, $\sum_j [\vec{p}]_j = 0$ and for all j, $|[\vec{p}]_j| \leq 1$. Since Order$(\vec{p})$ = Order$_f(P)$ and is a strict order, there exist positive numbers $\delta_1, \ldots, \delta_{m!-1}$ such that for any vector $\vec{q} \in \mathbb{R}^{m!}$ with (1) $\sum_j [\vec{q}]_j = 0$ and (2) for all $j \leq m! - 1$, $|[\vec{p}]_j - [\vec{q}]_j| < \delta_j$, we have Order$_f(\vec{q})$ = Order$_f(\vec{p})$.

Let $S = \prod_{j=1}^{m!-1} [\pi(l_j) - \frac{1}{m!} - \delta_j, \pi(l_j) - \frac{1}{m!} + \delta_j]$ denote a hypercube in $\mathbb{R}^{m!-1}$. When $\vec{x}$ is generated from $\mathbb{N}(0, \Sigma_\pi)$, the probability that $\vec{x} \in S$ is strictly positive because $\mathbb{N}(0, \Sigma_\pi)$ has full support. It is not hard to prove that for any Y_n, if $Y_{noise} \in S$, then for the corresponding profile P_n we have Order$_f(P_n)$ = Order$_f(\pi) \oplus$ Order$_f(\vec{p}) = \rhd$. Hence the probability for Order$_f(P_n) = \rhd$ is at least $\Pr(Y_{noise} \in S)$, which converges to $\Pr(\vec{x} \in S)$ when $\vec{x}$ is generated from $\mathcal{N}(0, \Sigma_\pi)$. This proves part (1).

Part (2) of the lemma. For any $\unrhd \notin$ Nbr$_f(\pi)$, we prove the lemma in the following three cases.

Case 1: $\unrhd$ does not extend Order$_f(\pi)$. Following a similar argument with the case where Order$_f(\pi)$ is a linear order, if b_i is strictly preferred to b_j in Order$_f(\pi)$, then with probability that goes to 1, b_i is strictly preferred to b_j in Order$_f(P_n)$. So the probability of Order$_f(P_n) = \unrhd$ goes to 0.

Case 2: $\unrhd$ is not a linear order. We recall that for any pair of $k_1, k_2 \leq K$ with $k_1 \neq k_2$, there exists a linear order l such that $[f(l)]_{k_1} \neq [f(l)]_{k_2}$. Therefore, following the Berry-Esseen theorem, the probability of seeing a tie between the k_1-th component and k_2-th component of $f(P_n)$ for i.i.d. generated P_n is $O(n^{-0.5})$, which goes to 0 as n goes to infinity.

Case 3: $\unrhd$ is a linear order and extends Order$_f(\pi)$, but there is no profile $\rhd \in$ PL(f) such that $\unrhd =$ Order$_f(\pi) \oplus \rhd$. It follows that $\unrhd \notin$ PL(f), otherwise $\unrhd =$ Order$_f(\pi) \oplus \unrhd$, which is a contradiction. Hence for any profile P_n, $f(P_n) \neq \unrhd$. □

Corollary 1 *For any π over $\mathcal{L}(\mathcal{C})$, if Order$_f(\pi)$ is a linear order $\rhd$, then Nbr$_f(\pi) = \{\rhd\}$ and* $\lim_{n\to\infty} \Pr(\text{Order}_f(P_n) = \rhd) = 1$.

4 GOSRs as Consistent Estimators

Theorem 1 *Given $\mathcal{M}_{\mathcal{C}} = (\mathcal{O}, \vec{\pi})$, f, and g, $GOS_{(f,g)}$ is consistent w.r.t. $\mathcal{M}_{\mathcal{C}}$ if and only if for all $o \in \mathcal{O}$ and all $\rhd \in Nbr_f(\pi_o)$, we have $g(\rhd) = o$.*

Proof: The "if" direction follows after Lemma 1. To prove the "only if" direction, if there exists o and $\rhd \in \text{Nbr}_f(\pi_o)$ with $g(\rhd) \neq o$, then by Lemma 1, as $n \to \infty$, the probability for the order over the components of the total generalized score vector to be $\rhd$ is non-negligible. Hence, with non-negligible probability $\text{GOS}_{(f,g)}$ will not output o when P_n is generated i.i.d. from π_o, which means that $\text{GOS}_{(f,g)}$ is not consistent. □

The condition in Theorem 1 might be hard to check since it might be hard to enumerate elements in $\text{Nbr}_f(\pi_o)$. In fact, it suffices to prove that for all extensions $\rhd$ of $\text{Nbr}_f(\pi_o)$, $g(\rhd) = o$, then the condition in Theorem 1 automatically holds. This leads to the following corollary.

Corollary 2 *Given a parametric ranking model $\mathcal{M}_{\mathcal{C}} = (\mathcal{O}, \vec{\pi})$, if for all $o \in \mathcal{O}$ and all extensions $\rhd \in \mathcal{L}(\mathcal{B}_K)$ of $Order_f(\pi_o)$ we have $g(\rhd) = o$, then $GOS_{(f,g)}$ is a consistent estimator for $\mathcal{M}_{\mathcal{C}}$.*

We now give an example of applying Corollary 2 to STV.

Proposition 2 *STV (preference function) is a consistent estimator for $\mathcal{M}_\varphi$ for all φ. STV (resolute rule) is a consistent estimator for $\mathcal{M}_{\vec{s}}$ for all scoring vector $\vec{s}$.*

Proof: We first present the GOSR formulation of STV (as a resolute rule and as a preference function), which is similar to the GSR representation of STV in [26]. We will use generalized scoring vectors with exponentially many components. For every proper subset S of alternatives, for every alternative c outside of S, there is a component in the vector that contains the number of times that c is ranked first if all alternatives in S are removed. We define $\text{GOS}_{(f,g)}$ as follows.

- $K = \sum_{i=0}^{m-1} \binom{m}{i}(m-i)$; the elements of $\mathcal{B}_K$ are indexed by (S, j), where S is a proper subset of $\mathcal{C}$ and $j \leq m, c_j \notin S$.
- $(f(V))_{(S,j)} = 1$, if after removing S from V, c_j is at the top of the modified V; otherwise, let $(f(V))_{(S,j)} = 0$.
- g mimics the execution of STV to select a winner (for resolute version) or a ranking (for preference function version).

W.l.o.g. suppose the ground truth parameter $o = [c_1 \succ \cdots \succ c_m]$. By Corollary 2, to show that STV (preference function) is consistent for $\mathcal{M}_\varphi$, it suffices to show that for any $2 \leq k \leq m$, after removing $C_k = \{c_{k+1}, \ldots, c_m\}$, c_k has strictly the lowest expected plurality score, where the expectation is taken over a randomly generated ranking from $\mathcal{M}_\varphi$ given o. To this end, for any $i < k$, we consider the following one-one mapping. For any rankings V where c_i is ranked at the top after all alternatives in C_k are removed, we switch the positions of c_i and c_{i+1}. This will give us another ranking V' where c_{i+1} is ranked in the top position if alternatives in C_k are removed. It is easy to check that $\text{Kendall}(V, o) = \text{Kendall}(V', o) - 1$, which means that the expected plurality score of c_i is higher than the expected plurality score of c_{i+1} after C_k is removed. This shows that if C_k is removed, then the expected score of c_k is strictly smaller than all other remaining alternatives. Hence for any $\rhd$ that is an extension of $\text{Order}_f(\pi_o)$, $g(\rhd) = o$. By Corollary 2, STV is consistent for $\mathcal{M}_\varphi$.

To prove that STV (resolute rule) is consistent w.r.t. $\mathcal{M}_{\vec{s}}$, w.l.o.g. suppose $o = c_1$, it suffices to show that for any $C \subseteq \mathcal{C}$, after removing C, c_1 has the strictly largest expected plurality score. This can be proved by a similar argument to the proof for $\mathcal{M}_\varphi$: for any other alternative $c \neq c_1$, for any linear order V where c_1 is ranked in the top after removing C, we can obtain another linear order V' by switching the positions of c_1 and c'. Since the position of c_1 in V is strictly higher than the position of c_1 in V', we have $\pi_o(V) \geq \pi_o(V')$, and the inequality is strict for some V. The proposition follows after a similar argument as for $\mathcal{M}_\varphi$. □

Given a parametric ranking model $\mathcal{M}_\mathcal{C} = (\mathcal{O}, \vec{\pi})$ and a consistent GOS$_{(f,g)}$, we next give an upper bound on the convergence rate of the outcome of GOS$_{(f,g)}$ to the ground truth. We let s_{max} denote the maximum absolute value of the components in all generalized scoring vectors. That is, $s_{max} = \max_{V,j} |[f(V)]_j|$. Let s_{min} denote the minimum non-zero absolute value of the components in all generalized scoring vectors. Let d_{min} denote the smallest non-zero difference between the components in all $f(\pi_o)$. That is, $d_{min} = \min_{i,j\leq K,o}\{|[f(\pi_o)]_i - [f(\pi_o)]_j| : [f(\pi_o)]_i \neq [f(\pi_o)]_j\}$. Let p_{min} denote the minimum probability of any linear order under any parameter, that is, $p_{min} = \min_{V,o} \pi_o(V)$.

Theorem 2 *Suppose GOS$_{(f,g)}$ is a consistent estimator for $\mathcal{M} = (\mathcal{O}, \vec{\pi})$. For any $o \in \mathcal{O}$ and $n \in \mathbb{N}$, we have:*

$$\Pr\left(GOS_{(f,g)}(P_n) \neq o\right) < K \cdot \exp\left(-n \cdot \frac{d_{min}}{8s_{max}^2}\right) + \frac{(K(K-1)s_{max})^3}{(2p_{min})^{1.5}(s_{min})^3\sqrt{n}} = O(n^{-0.5})$$

Proof: Let Strict(π_o) denote the set of strict pairwise comparisons in Order$_f(\pi_o)$, that is, $(b_i, b_j) \in$ Strict(π_o) if and only if b_i is strictly preferred to b_j in Order$_f(\pi_o)$. For any profile P, if Order$_f(P)$ is a linear order that extends Order$_f(\pi_o)$, then Order$_f(P) \in$ Nbr$_f(\pi_o)$. By Theorem 1, GOS$_{(f,g)}(P) = o$. Hence, if GOS$_{(f,g)}(P) \neq o$, then there are only two possibilities: (1) for some $(b_i, b_j) \in$ Strict(π_o), $[f(P)]_j \geq [f(P)]_i$, or (2) there exist $i \neq j$ with $[f(P)]_i = [f(P)]_j$.

For case (1), for any pair of $(b_i, b_j) \in$ Strict(π_o), we let $X_1, \ldots, X_n$ denote i.i.d. variables that represents $[f(l)]_i - [f(l)]_j$ for randomly generated l from π_o. Let $Y_n = (X_1 + \cdots + X_n)/n$. We have $E(X_i) \geq d_{min}$, $Var(X_i) < 2s_{max}$, and each X_i takes a value in $[-2s_{max}, 2s_{max}]$. By Hoeffding's inequality [17], we have: $\Pr(Y_n \leq 0) = \Pr(Y_n - E(X_1) \leq -E(X_1)) \leq \exp\left(-\frac{2n^2E(X_1)}{n(4s_{max})^2}\right) \leq \exp\left(-n \cdot \frac{d_{min}}{8s_{max}^2}\right)$.

For case (2), for any pair of i, j with $[f(\pi_o)]_i = [f(\pi_o)]_j$, we define X_i and Y_n similarly as in case (1). The third moment of X_1 is no more than s_{max}^3 and $Var(X_i) \geq p_{min}s_{min}^2$. By Berry-Esseen theorem, the probability for $Y_n = 0$ is no more than $\frac{(s_{max})^3}{(p_{min})^{1.5}(s_{min})^3\sqrt{n}}$.

Combining the above calculations, for (1) we only need to consider adjacent pairs in Strict(π_o) and for (2) we need to consider all pairs of tied components. Hence the probability that either (1) or (2) holds is at most $K \cdot \exp\left(-n \cdot \frac{d_{min}}{8s_{max}^2}\right) + \frac{(K(K-1)s_{max})^3}{(2p_{min})^{1.5}(s_{min})^3\sqrt{n}}$, which proves the theorem. □

We next show that the $O(n^{-0.5})$ bound proved in Theorem 2 is asymptotically tight.

Theorem 3 *There exists a parametric ranking model $\mathcal{M}_\mathcal{C}$ where $\mathcal{O} = \mathcal{C}$ and a GOSR r such that (1) r is consistent w.r.t. $\mathcal{M}_\mathcal{C}$, and (2) there exists $o \in \mathcal{O}$ such that for all even numbers n, $\Pr(r(P_n) \neq o) = \Omega(n^{-0.5})$, where votes in P_n are generated i.i.d. from π_o.*

Proof: Let there be three alternatives $\{c_1, c_2, c_3\}$ and $\mathcal{O} = \mathcal{C}$, that is, we want to select a single winner. Let the parametric ranking model $\mathcal{M}_\mathcal{C} = (\mathcal{O}, \vec{\pi})$ be the following:

$$\begin{array}{lll} \text{Groundtruth} = \text{c}_1: & \pi_{c_1}(c_1 \succ c_2 \succ c_3) = 0.5, & \pi_{c_1}(c_1 \succ c_3 \succ c_2) = 0.5 \\ \text{Groundtruth} = \text{c}_2: & \pi_{c_2}(c_2 \succ c_1 \succ c_3) = 0.5, & \pi_{c_2}(c_2 \succ c_3 \succ c_1) = 0.5 \\ \text{Groundtruth} = \text{c}_3: & \pi_{c_3}(c_3 \succ c_1 \succ c_2) = 0.5, & \pi_{c_3}(c_3 \succ c_2 \succ c_1) = 0.5 \end{array}$$

Let r be the Borda rule with fixed order tie-breaking $c_1 \succ c_2 \succ c_3$, except in one case: if c_1's total score is strictly the largest, and the total scores of c_2 and c_3 are exactly the same, then the winner is c_2 (instead of c_1 for Borda). It is not hard to verify that r is a GOSR. By Theorem 1, r is consistent w.r.t. $\mathcal{M}_\mathcal{C}$.

Let for any profile P and alternative c, let $s(P, c)$ denote the Borda score of c in P. For any even n, when the ground truth is c_1, the probability that $s(P_n, c_2) = s(P_n, c_3)$ is $\binom{n}{n/2}/2^n$. By Stirling's formula, we have

$$\frac{\binom{n}{n/2}}{2^n} = \frac{n!}{(\frac{n}{2}!)^2 2^n} \approx \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{(\sqrt{\pi n}(\frac{n}{2e})^{n/2})^2 2^n} = \frac{\sqrt{2}}{\sqrt{\pi n}} = \Omega(n^{-0.5})$$

Similar to the proof of Theorem 2, it is not hard to show that the probability for the total score of c_1 to be the highest is $1-\exp(-\Omega(n)) = 1-o(n^{-0.5})$. So the probability for $s(P_n, c_1) > s(P_n, c_2) = s(P_n, c_3)$ is $\Omega(n^{-0.5})$. In all such cases $r(P_n) = c_2 \neq c_1$, which proves the theorem. □

For specific distributions and GOSRs we may improve the bound as follows.

Proposition 3 *Suppose $GOS_{(f,g)}$ is a consistent estimator for $\mathcal{M}_\mathcal{C} = (\mathcal{O}, \vec{\pi})$. For any $o \in \mathcal{O}$ and $n \in \mathbb{N}$, if for all extensions $\trianglerighteq$ of $Order_f(\pi_o)$, $g(\trianglerighteq) = o$, then:*

$$\Pr\left(GOS_{(f,g)}(P_n) \neq o\right) < K \cdot \exp\left(-n \cdot \frac{d_{min}}{8s_{max}^2}\right)$$

Example 2 *The bound in Proposition 3 can be applied to STV (preference function) w.r.t. $\mathcal{M}_\varphi$ for all φ and STV (resolute rule) w.r.t. $\mathcal{M}_{\vec{s}}$ for all $\vec{s}$, following the proof of Proposition 2.*

In the next theorem we fully characterize all GOSRs that are consistent w.r.t. *some* parametric ranking models. The theorem states that a GOSR $\text{GOS}_{(f,g)}$ is a consistent estimator for some parametric ranking model if and only if for any outcome o, there exists a profile P such that (1) $\text{GOS}_{(f,g)}(P) = o$ and (2) $\text{Order}_f(P)$ is a linear order.

Theorem 4 *A GOSR is a consistent estimator for* some *parametric ranking model if and only if for all $o \in \mathcal{O}$, $g^{-1}(o) \cap PL(f) \neq \emptyset$.*

Proof: The "if" direction: For any profile P' with $\text{Order}_f(P') \in g^{-1}(o) \cap \text{PL}(f)$, since $\text{Order}_f(P')$ is strict, there exists a $t \in \mathbb{N}$ so that $\text{Order}_f(tP' \cup \mathcal{L}(\mathcal{C})) = \text{Order}_f(P') =\rhd$, where $tP' \cup \mathcal{L}(\mathcal{C})$ is the profile composed of t copies of P' plus each linear order in $\mathcal{L}(\mathcal{C})$. We note that $P_o = tP' \cup \mathcal{L}(\mathcal{C})$ is a profile that contains all types of linear orders. Then, we define a distribution π_o such that for any linear order V, $\pi_o(V) = \frac{P_o(V)}{|P_o|}$, where $P_o(V)$ is the number of occurrences of V in P_o. Consistency follows after Corollary 1.

The "only if" direction: Let o denote the outcome with $g^{-1}(o) \cap \text{PL}(f) = \emptyset$. We need the following lemma.

Lemma 2 *For any distribution π and generalized scoring function f, $Nbr_f(\pi) \subseteq PL(f)$.*

Proof: For any $\rhd \in \text{Nbr}_f(\pi)$, let $P_\rhd$ be a profile such that $\rhd = \text{Order}_f(\pi) \oplus \text{Order}_f(P_\rhd)$. The existence of $P_\rhd$ is guaranteer by the definition of $\text{Nbr}_f(\pi)$ (Definition 9). For any n, we let $Q_n = Q_n^1 \cup Q_n^2$ be a profile composed of the following two parts.

1. The first part Q_n^1 contains the following votes: for any linear order $V \in \mathcal{L}(\mathcal{C})$, there are $\lfloor \pi(V) \cdot n \rfloor$ copies of V.

2. The first part Q_n^2 contains $\lfloor \sqrt{n} \rfloor$ copies of $P_\rhd$.

By the Central Limit Theorem, for any i, j such that b_i is strictly preferred to b_j in $\text{Order}_f(\pi)$, as n goes to infinity $[f(Q_n^1)]_i - [f(Q_n^1)]_j = \Theta(n)$; for any i, j such that b_i is tied with b_j in $\text{Order}_f(\pi)$, as n goes to infinity $|[f(Q_n^1)]_i - [f(Q_n^1)]_j| = O(1)$. Therefore, Q_n^2 effectively acts as a tie-breaker for $\text{Order}_f(\pi)$, in the same way as $\text{Order}_f(P_\rhd)$ does. This shows that there exists n such that $\text{Order}_f(Q_n) =\rhd$ and proves the lemma. □

By this lemma, because $g^{-1}(o) \cap \text{PL}(f) = \emptyset$, for any distribution π and any $\rhd \in \text{Nbr}_f(\pi)$, $g(\rhd) \neq o$. By Therom 1, $\text{GOS}_{(f,g)}$ is not consistent w.r.t. any model. □

5 Minimax Estimators for Neutral Parametric Ranking Models

In this section we fully characterize minimax estimators for *neutral* parametric ranking models, which will be defined momentarily. We recall that minimaxity is the optimality of an estimator w.r.t. the frequentist risk. We first define neutral parametric models and loss functions. Intuitively, it states that if a permutation over $\mathcal{C}$ is applied to the data as well as the parameter, then the probability of the permuted data under the permuted parameter should stay the same. Moreover, the loss function should also be invariant to any such permutation. However, due to the generality of the parameter space, a permutation over $\mathcal{C}$ may not be directly applied to the parameter space. Hence, for any permutation M over $\mathcal{C}$ we need to define a corresponding permutation over the parameter space $\mathcal{O}$.

Definition 10 *A parametric ranking model $\mathcal{M}$ with a loss function L is* neutral*, if there exists a permutation S_M over $\mathcal{O}$ for every permutation M over $\mathcal{C}$ that satisfies the following conditions.*

1. For any pair of permutations M_1 and M_2, $S_{M_1 \circ M_2} = S_{M_1} \circ S_{M_2}$. For any $o \in \mathcal{O}$, $S_I(o) = o$, where I is the identity permutation.

2. For all $o \in \mathcal{O}$, $V \in \mathcal{L}(\mathcal{C})$, and permutation M, we have $\pi_o(V) = \pi_{S_M(o)}(M(V))$.

3. For all $o, o' \in \mathcal{O}$, $L(o, o') = L(S_M(o), S_M(o'))$.

4. For any pair $o, o' \in \mathcal{O}$, there exists a permutation M such that $S_M(o) = o'$.

Condition 1 requires that there is a group homomorphism between permutations on $\mathcal{O}$ and permutations on $\mathcal{L}(\mathcal{C})$. Condition 2 states that when the ground truth is permuted using S_M, then its corresponding distribution over $\mathcal{L}(\mathcal{C})$ is also permuted using M. Condition 3 states that the loss function is invariant to permutations, and Condition 4 requires that every outcome in $\mathcal{O}$ can be obtained from another outcome in $\mathcal{O}$ by applying some permutation S_M. We note that in our definition, neutrality is a property for a parametric ranking model together with a loss function.

To show the generality of neutral models, we define a natural extension of $\mathcal{M}_\varphi$ using a different distance function from the Kendall-tau distance. A distance function $d : \mathcal{L}(\mathcal{C}) \times \mathcal{L}(\mathcal{C}) \to \mathbb{R}$ is *neutral*, if for any permutation M over $\mathcal{C}$ and any $V, W \in \mathcal{L}(\mathcal{C})$, we have $d(V, W) = d(M(V), M(W))$. For example, Spearman's footrule is neutral.

Definition 11 *Given $0 < \varphi$ and a neutral distance d, let $\mathcal{M}_\varphi^d = (\mathcal{O}, \vec{\pi})$, where $\mathcal{O} = \mathcal{L}(\mathcal{C})$, and for any $V, W \in \mathcal{L}(\mathcal{C})$, $\pi_W(V) = \frac{1}{Z}\varphi^{d(V,W)}$, where Z is the normalization factor with $Z = \sum_{V \in \mathcal{L}(\mathcal{C})} \varphi^{d(V,W)}$.*

When $\varphi < 1$, $\mathcal{M}_\varphi^d$ is a special case of *d-Monotonic Noise Models* [8].

Example 3 *$\mathcal{M}_\varphi$ with $L_{0\text{-}1}$ is neutral, where $S_M = M$. For any neutral distance function d, $\mathcal{M}_\varphi^d$ with $L_{0\text{-}1}$ is neutral, where $S_M = M$. For any scoring vector $\vec{s}$, $\mathcal{M}_{\vec{s}}$ with $L_{0\text{-}1}$ is neutral, where $S_M = M$.*

Proposition 4 *If $\mathcal{M}_\mathcal{C}$ is neutral with any loss function L, then $\mathcal{M}_\mathcal{C}$ is neutral with $L_{0\text{-}1}$.*

Proof: Let S_M denote the permutation over $\mathcal{O}$ for $\mathcal{M}_\mathcal{C}$ and L. It suffices to prove that for any $o_1 \neq o_2$ and any M, $S_M(o_1) \neq S_M(o_2)$. This is true because otherwise $o_1 = S_{M^{-1}} \circ S_M(o_1) = S_{M^{-1}} \circ S_M(o_2) = o_2$, which is a contradiction. □

Given a parametric ranking model, a loss function, and a fixed n, we define the *uniform Bayesian estimator*, denoted by T_U, to be the randomized Bayesian estimator w.r.t. the uniform distribution over $\mathcal{O}$ that outputs uniformly at random a parameter that minimizes the Bayesian expected loss.

Example 4 *For any parametric ranking model with the* 0-1 *loss function $L_{0\text{-}1}$, T_U is the uniformly randomized MLE.*

We now present the main theorem of this section.

Theorem 5 *For any neutral parametric ranking model $\mathcal{M}$ with loss function L and any n, the uniform Bayesian estimator is minimax.*

Proof: We will use the following lemma.

Lemma 3 [6, Section 5.3.2 III] *Given a statistical model with loss function L, let T_{π^*} denote a Bayesian estimator for prior π^*. If $R_F^n(o, T_{\pi^*})$ are equal for all $o \in \mathcal{O}$, then T_{π^*} is minimax.*

By Lemma 3, it suffices to show that for all $o \in \mathcal{O}$, $R_F(o, T_U)$ are equal. For any pair of different $o_1, o_2 \in \mathcal{O}$, let M denote any permutation with $M(o_1) = o_2$.

We first prove that for any profile P of n votes and any $o^* \in \mathcal{O}$, $T_U(P)(o^*) > 0$ if and only if $T_U(M(P))(S_M(o^*)) > 0$. The former holds if and only if o^* minimizes the Bayesian expected loss at P, which is equivalent to that for all o', $\sum_o L(o, o^*) \Pr(o|P) \le \sum_o L(o, o') \Pr(o|P)$. We have the following calculation.

$$
\begin{aligned}
& \sum\nolimits_o L(o, o^*) \Pr(o|P) \le \sum\nolimits_o L(o, o') \Pr(o|P) \\
\Leftrightarrow & \sum\nolimits_o L(o, o^*) \Pr(P|o) \le \sum\nolimits_o L(o, o') \Pr(P|o) && (1) \\
\Leftrightarrow & \sum\nolimits_o L(S_M(o), S_M(o^*)) \Pr(M(P)|S_M(o)) \\
& \quad \le \sum\nolimits_o L(S_M(o), S_M(o')) \Pr(M(P)|S_M(o)) && (2) \\
\Leftrightarrow & \sum\nolimits_o L(o, S_M(o^*)) \Pr(M(P)|o) \\
& \quad \le \sum\nolimits_o L(o, S_M(o')) \Pr(M(P)|o) && (3) \\
\Leftrightarrow & \sum\nolimits_o L(o, S_M(o^*)) \Pr(o|M(P)) \\
& \quad \le \sum\nolimits_o L(o, S_M(o')) \Pr(o|M(P)) && (4)
\end{aligned}
$$

(1) and (4) are due to Bayes' rule and the uniform prior assumption. (2) is because $\mathcal{M}$ and L are neutral. (3) is a change of variables, which is possible because for any $o_1 \neq o_2$ and any M, $S_M(o_1) \neq S_M(o_2)$, otherwise $o_1 = S_{M^{-1}} \circ S_M(o_1) = S_{M^{-1}} \circ S_M(o_2) = o_2$, which is a contradiction.

It follows that $T_U(M(P)) = M(T_U(P))$, which means that $L(o_1, T_U(P)) = L(S_M(o_1), M(T_U(P))) = L(o_2, T_U(M(P)))$

Therefore, we have: $R_F^n(o_1, T_U) = \sum_P L(o_1, T_U(P)) \Pr(P|o_1) = \sum_P L(o_2, T_U(M(P))) \Pr(M(P)|o_2) = R_F^n(o_2, T_U)$

The theorem follows after Lemma 3. □

Combining Theorem 5 and Proposition 4, and recall from Example 4 that the uniformly randomized MLE is the uniform Bayesian estimator for the 0-1 loss function, we have the following corollary.

Corollary 3 *For any neutral parametric ranking model $\mathcal{M}$ with any loss function L, the uniformly randomized MLE is minimax.*

Combining Corollary 3 and Example 3, we obtain the following corollary.

Corollary 4 *For $\mathcal{M}_\varphi$, $\mathcal{M}_\varphi^d$ with neutral distance function, and $\mathcal{M}_{\vec{s}}$, the uniformly randomized MLE is a minimax estimator w.r.t. the 0-1 loss function. That is, for each of these models, for any n, any estimator T^*, and any outcome o, $\Pr_{P_n \sim \pi_o}(T(P_n) = o) \ge \Pr_{P_n \sim \pi_o}(T^*(P_n) = o)$.*

The case for $\mathcal{M}_\varphi$ in Corollary 4 was proved directly by [8], where the minimax property is equivalently defined as having the lowest sample complexity. We note that in Appendix A of [8], an example was shown to illustrate that MLE is not always the minimax rule w.r.t. $L_{0\text{-}1}$. This does not contradict Theorem 5 since the model there is not neutral with any loss function.

6 Future Work

We plan to refine the upper bound described in Theorem 2. The notion of GOSRs and neutral parametric ranking models are of independent interest. We also plan to evaluate common voting rules from a statistical viewpoint for other parametric ranking models, and extend our study to cases where votes are represented by partial orders.

7 Acknowledgments

This work is supported by a startup fund at RPI. We thank Marcus Pivato, Ariel Procaccia, and all anonymous AAAI reviewers and COMSOC reviewers for helpful comments and suggestions.

References

[1] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Random utility theory for social choice. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 126–134, Lake Tahoe, NV, USA, 2012.

[2] Hossein Azari Soufiani, William Chen, David C. Parkes, and Lirong Xia. Generalized method-of-moments for rank aggregation. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2013.

[3] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Preference Elicitation For General Random Utility Models. In *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, Bellevue, Washington, USA, 2013.

[4] Hossein Azari Soufiani, David Parkes, and Lirong Xia. *Statistical Decision Theory Approaches to Social Choice*. Working paper, `cs.rpi.edu//~xial/Files/SDT.pdf`, 2014.

[5] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Computing Parametric Ranking Models via Rank-Breaking. In *Proceedings of the 31st International Conference on Machine Learning*, Beijing. China, 2014.

[6] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. James O. Berger, 2nd edition, 1985.

[7] Mark Braverman and Elchanan Mossel. Noisy Sorting Without Resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 268–276, Philadelphia, PA, USA, 2008.

[8] Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, Philadelphia, PA, 2013.

[9] Ioannis Caragiannis, Ariel D. Procaccia, and Nisarg Shah. Modal Ranking: A Uniquely Robust Voting Rule. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014.

[10] Christine Choirat and Raffaello Seri. Estimation in discrete parameter models. *Statistical Science*, 27(2): 278–293, 2012.

[11] 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.

[12] Vincent Conitzer. The maximum likelihood approach to voting on social networks. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton-13)*, pages 1482–1487, Monticello, IL, USA, 2013.

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

[14] Vincent Conitzer, Matthew Rognlie, and Lirong Xia. Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 109–115, Pasadena, CA, USA, 2009.

[15] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Good rationalizations of voting rules. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 774–779, Atlanta, GA, USA, 2010.

[16] William H. Greene. *Econometric Analysis*. Prentice Hall, 7th edition, 2011.

[17] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13—30, 1963.

[18] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[19] Tie-Yan Liu. *Learning to Rank for Information Retrieval*. Springer, 2011.

[20] Tyler Lu and Craig Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the Twenty-Eighth International Conference on Machine Learning (ICML 2011)*, pages 145–152, Bellevue, WA, USA, 2011.

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

[22] Hannu Nurmi. *Comparing voting systems*. Springer, 1987.

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

[24] 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.

[25] Amirali Salehi-Abari and Craig Boutilier. Empathetic social choice on social networks. In *Proceedings of COMSOC*, 2012.

[26] Lirong Xia and Vincent Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 109–118, Chicago, IL, USA, 2008.

[27] 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.

Lirong Xia
Computer Science Department
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, USA
Email: xial@cs.rpi.edu