A Statistical Decision-Theoretic Framework for Social Choice

Hossein Azari
SEAS, Harvard University
Cambridge, MA 02138, USA
azari@fas.harvard.edu

David Parkes
SEAS, Harvard University
Cambridge, MA 02138, USA
parkes@eecs.harvard.edu

Lirong Xia
Computer Science Department, RPI
Troy, NY 12110, USA
xial@cs.rpi.edu

Abstract

In this paper, we take a statistical decision-theoretic viewpoint on social choice, putting a focus on the decision to be made on behalf of a system of agents. In our framework, we are given a statistical ranking model, a decision space, and a loss function, and formulate social choice mechanisms as statistical choice functions that minimize expected loss. This suggests a general framework for the design and analysis of new social choice mechanisms. We compare *Bayesian estimators*, which minimize Bayesian expected loss, for two variants of the Mallows model and the Kemeny rule. We consider various normative properties, in addition to computational complexity and asymptotic behavior.

Introduction

Social choice studies the design and evaluation of voting rules (or rank aggregation rules) that satisfy desired properties. There have been two main perspectives: reach a compromise among subjective preferences of agents, or make an objectively correct decision. The former has been extensively studied in classical social choice in the context of political elections, and the latter is less well-developed, even though it can be dated back to the Condorcet Jury Theorem in the 18th century (Condorcet, 1785).

In many multi-agent scenarios the main consideration is to achieve the second objective, and make an objectively correct decision. This is the focus of this paper. Along this line, most work views agents' votes as i.i.d. samples from a statistical model, and computes the *maximum likelihood estimator (MLE)* to estimate the parameters that maximize the likelihood. Popular statistical ranking models include (1) the *Mallows model* (Mallows, 1957), where a parameter (ground truth) includes a "correct" ranking over the alternatives and a value between 0 and 1 to measure dispersion, and (2) *random utility models* (Thurstone, 1927), where each parameter includes specifications of utility distributions for the alternatives.

A limitation of these approaches is that they estimate the parameters of the model, but may not directly inform the right *decision* to make in the multi-agent context. The main approach has been to return the modal rank order implied by the estimated parameters, or the alternative with the highest, predicted marginal probability of being ranked in the top position.

In this paper, we develop a general framework that adopts *statistical decision theory* (Berger, 1985) for the design and analysis of new voting rules. The approach couples a statistical ranking model with an explicit decision space and loss function. Given these, we can adopt *Bayesian estimators* as social choice mechanisms, which make decisions to minimize the expected loss w.r.t. the posterior distribution on the parameters (called the *Bayesian risk*). This provides a principled methodology for the design of new voting rules.

To show the viability of the framework, we focus on selection of a single alternative under a natural extension of the 0-1 loss function for two variants of the Mallows model. In both models the dispersion parameter, denoted φ, is fixed. The difference is that in model 1, $\mathcal{M}^1_\varphi$, the parameter space is composed of all linear orders over alternatives as in the Mallows model, while in model 2, $\mathcal{M}^2_\varphi$, the parameter space is composed of all possibly cyclic rankings over alternatives (irreflexive, antisymmetric, and total binary relations). $\mathcal{M}^2_\varphi$ is a natural model that captures real-world scenarios where the ground truth may contain cycles, or agents' preferences are cyclic, but are required to report a linear order by the protocol. More importantly, as we will show later, $\mathcal{M}^2_\varphi$ is superior from a computational viewpoint.

Through this approach, we obtain two new voting rules. Although our primary concern is that of statistical decision theory, we also evaluate the rules with respect to various normative properties, including *anonymity, neutrality, monotonicity, the majority criterion, the Condorcet criterion* and *consistency*, since this gives additional support for the approach. Both rules satisfy anonymity, neutrality, and monotonicity, but fail the majority criterion, Condorcet criterion,[1] and consistency. Admittedly, the two new rules are not very superior w.r.t. normative properties, but they are not bad either. We also investigate the computational complexity of the two rules. Strikingly, despite the similarity of the two models, the Bayesian estimator for $\mathcal{M}^2_\varphi$ can be computed in polynomial time, while computing the Bayesian estimator for $\mathcal{M}^1_\varphi$ is $\mathsf{P}^{\mathsf{NP}}_{||}$-hard, which means that it is at least NP-hard. Our results are summarized in Table 1.

[1] The new voting rule for $\mathcal{M}^1_\varphi$ fails them for all $\varphi < 1/\sqrt{2}$.

	Anonymity, neutrality	Monotonicity	Majority, Condorcet	Consistency	Complexity	Min. Bayesian risk
Kemeny (for alternatives)	Y	Y	Y	N	NP-hard, $P_{\|\|}^{NP}$-hard	N
Bayesian est. of $\mathcal{M}_\varphi^1$ (uni. prior)	Y	Y	N	N	NP-hard, $P_{\|\|}^{NP}$-hard (Theorem 3)	Y
Bayesian est. of $\mathcal{M}_\varphi^2$ (uni. prior)	Y	Y	N	N	P (Theorem 4)	Y

Table 1: Kemeny vs. Bayesian estimators to choose *alternatives*. All results for the two Bayesian estimators are new.

We also compare the asymptotic outcomes of the two new rules with the Kemeny rule for wining alternatives, which is a natural extension of the MLE of $\mathcal{M}_\varphi^1$ proposed by Fishburn (1977). It turns out that when n votes are generated according to $\mathcal{M}_\varphi^1$, all three rules select the same winner asymptotically almost surely (a.a.s.) as $n \to \infty$. When the votes are generated according to $\mathcal{M}_\varphi^2$, the new rule for $\mathcal{M}_\varphi^1$ still selects the same winner as Kemeny a.a.s.; however, for some parameters, the winner selected by the new rule for $\mathcal{M}_\varphi^1$ and Kemeny is different from the winner selected by the new rule for $\mathcal{M}_\varphi^2$ with non-negligible probability. Our experiments on synthetic datasets confirm these observations.

Related work. In addition to Condorcet's statistical approach to social choice (Condorcet, 1785; Young, 1988), the MLE approach towards social choice has been widely studied (Conitzer and Sandholm, 2005; Conitzer, Rognlie, and Xia, 2009; Xia, Conitzer, and Lang, 2010; Xia and Conitzer, 2011; Azari Soufiani, Parkes, and Xia, 2012; Procaccia, Reddi, and Shah, 2012; Caragiannis, Procaccia, and Shah, 2013). There have also been some proposals to go beyond MLE in social choice, for example,Young (1988) proposed to select a winning alternative that is *"most likely to be the best (i.e., top-ranked in the true ranking)."* We will see that this is a special case of our proposed framework in Example 2, and can be naturally applied to $\mathcal{M}_\varphi^1$ and $\mathcal{M}_\varphi^2$. Pivato (2013) conducted a similar study to Conitzer and Sandholm (2005), examining voting rules that can be interpreted as expect-utility maximizers.

We are not aware of previous work that frames the problem of social choice from the viewpoint of statistical decision theory. Our framework studies the reverse question: how to design new voting rules given particular loss functions and probabilistic models? The statistical decision-theoretic framework is quite general, allowing considerations such as estimators that minimize the maximum expected loss, or the maximum expected regret (Berger, 1985). In a different context, focused on uncertainty about the availability of alternatives, Lu and Boutilier (2010) adopt a decision-theoretic view of the design of an optimal voting rule.

Preliminaries

In social choice, we have a set of m alternatives $\mathcal{C} = \{c_1, \dots, c_m\}$ and a set of n agents. Let $\mathcal{L}(\mathcal{C})$ denote the set of all linear orders over $\mathcal{C}$. For any alternative c, let $\mathcal{L}_c(\mathcal{C})$ denote the set of linear orders over $\mathcal{C}$ where c is ranked in the top. Agent j uses a linear order $V_j \in \mathcal{L}(\mathcal{C})$ to represent her preferences, called her *vote*. The collection of agents votes is called a *profile*, denoted by $P = \{V_1, \dots, V_n\}$. A *(irresolute) voting rule* $r : \mathcal{L}(\mathcal{C})^n \to (2^\mathcal{C} \setminus \emptyset)$ selects a set of winners for every profile profile of n votes.

For any pair of linear orders V, W, let Kendall(V, W) denote the *Kendall-tau distance* between V and W, that is, the number of different pairwise comparisons in V and W.

The *Kemeny rule* (a.k.a. *Kemeny-Young method*) (Kemeny, 1959; Young and Levenglick, 1978) selects all *linear orders* with the minimum Kendall-tau distance from the preference profile P, that is, Kemeny$(P) = \arg\min_W$ Kendall(P, W). The extension of Kemeny to select winning alternatives was due to Fishburn (1977), who defined it to select all alternatives that are ranked in the top position of some winning linear orders by the Kemeny rule. That is, it outputs $\{top(V) : V \in$ Kemeny$(P)\}$, where $top(V)$ is the top-ranked alternative in V.

Voting rules are often evaluated by the following axiomatic normative properties. An irresolute rule r satisfies:

- *anonymity*, if r is insensitive to permutations over agents;
- *neutrality*, if r is insensitive to permutations over alternatives;
- *monotonicity*, if for any P, $c \in (P)$, and any P' that is obtained from P by only raising the positions of c in one or multiple votes, then $c \in r(P')$;
- *Condorcet criterion*, if for any profile P where a Condorcet winner exists, then it must be the unique winner. A Condorcet winner is the alternative that beats every other alternative in pair-wise elections.
- *majority criterion*, if for any profile P where an alternative c is ranked in the top positions for more than half of the votes, then $r(P) = \{c\}$. If r satisfies the majority criterion then it also satisfies Condorcet criterion.
- *consistency*, if for any pair of profiles P_1, P_2 with $r(P_1) \cap r(P_2) \neq \emptyset$, then $r(P_1 \cup P_2) = r(P_1) \cap r(P_2)$.

For any profile P, its *weighted majority graph (WMG)*, denoted by WMG(P), is a weighted directed graph whose vertices are $\mathcal{C}$, and there is an edge between any pair of alternatives (a, b) with weight $w_P(a, b) = \#\{V \in P : a \succ_V b\} - \#\{V \in P : b \succ_V a\}$.

A parametric model $\mathcal{M} = (\Theta, \mathcal{S}, \Pr)$ is composed of three parts: a *parameter space* Θ, a *sample space* $\mathcal{S}$, and a set of probability distributions over $\mathcal{S}$ indexed by elements of Θ: for each $\theta \in \Theta$, the distribution indexed by θ is denoted by $\Pr(\cdot|\theta)$.

Given a parametric model $\mathcal{M}$, a *maximum likelihood estimator (MLE)* is a function $f_{\text{MLE}} : \mathcal{S} \to \Theta$ such that for any data $P \in \mathcal{S}$, $f_{\text{MLE}}(P)$ is the parameter that maximizes the likelihood of the data. That is, $f_{\text{MLE}}(P) \in \arg\max_{\theta\in\Theta} \Pr(P|\theta)$.

In social choice, the data are composed of i.i.d. linear or-

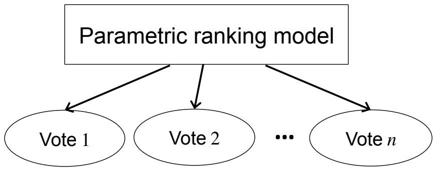

Figure 1: Statistical approach to social choice.

ders. Therefore, in this paper we focus on *parametric ranking models*. Given $\mathcal{C}$, a parametric ranking model $\mathcal{M}_{\mathcal{C}} = (\Theta, \Pr)$ is composed of a parameter space Θ and a distribution $\Pr(\cdot|\theta)$ over $\mathcal{L}(\mathcal{C})$ for each $\theta \in \Theta$, such that for any number of voters n, the sample space is $\mathcal{S}_n = \mathcal{L}(\mathcal{C})^n$, where each vote is generated i.i.d. from $\Pr(\cdot|\theta)$. Hence, for any profile $P \in \mathcal{S}$ and any $\theta \in \Theta$, we have $\Pr(P|\theta) = \prod_{V \in P} \Pr(V|\theta)$. We omit the sample space for parametric ranking models since it is determined by $\mathcal{C}$ and n. The framework of the statistical approach to social choice is illustrated in Figure 1.

We now recall the *Mallows model* (Mallows, 1957).

Definition 1 *In the* Mallows model*, the parameter space Θ is composed of a linear order in $\mathcal{L}(\mathcal{C})$ and a* dispersion parameter φ *with* $0 < \varphi < 1$. *For any profile P and $\theta = (W, \varphi)$, $\Pr(P|\theta) = \prod_{V \in P} \frac{1}{Z} \varphi^{\text{Kendall}(V,W)}$, where Z is the normalization factor with $Z = \sum_{V \in \mathcal{L}(\mathcal{C})} \varphi^{\text{Kendall}(V,W)}$.*

Statistical decision theory (Berger, 1985) studies scenarios where the decision maker must make a *decision* $d \in \mathcal{D}$ based on the data P generated from a parametric model, generally $\mathcal{M} = (\Theta, \mathcal{S}, \Pr)$. The quality of the decision is evaluated by a *loss function* $L : \Theta \times \mathcal{D} \to \mathbb{R}$, which takes the *true* parameter and the decision as inputs.

In this paper, we focus on the *Bayesian principle* of statistical decision theory to design social choice mechanisms as choice functions that minimize the *Bayesian risk* under a prior distribution over Θ. More precisely, the Bayesian risk, $R_B(P, d)$, is the expected loss of the decision d when the parameter is generated according to the posterior distribution. That is, $R_B(P, d) = E_{\theta|P} L(\theta, d)$. Given a parametric model $\mathcal{M}$, a loss function L, and a prior distribution over Θ, a (deterministic) *Bayesian estimator* f_B is a decision rule that makes a deterministic decision in $\mathcal{D}$ to minimizes the Bayesian risk, that is, for any $P \in \mathcal{S}$, $f_B(P) \in \arg\min_d R_B(P, d)$. We focus on deterministic estimators in this work and leave randomized estimators for future research.

Example 1 *When Θ is discrete, an MLE of a parametric model $\mathcal{M}$ is a Bayesian estimator of the statistical decision problem $(\mathcal{M}, \mathcal{D} = \Theta, L_{0\text{-}1})$ under the uniform prior distribution, where $L_{0\text{-}1}$ is the* 0-1 loss function *such that $L_{0\text{-}1}(\theta, d) = 0$ if $\theta = d$, otherwise $L_{0\text{-}1}(\theta, d) = 1$.*

In this sense, all previous MLE approaches in social choice can be viewed as the Bayesian estimators of a statistical decision-theoretic framework for social choice where $\mathcal{D} = \Theta$, a 0-1 loss function, and the uniform prior.

Our Framework

The power of our framework is that we have freedom to choose any parametric ranking model, any decision space, any loss function, and any prior to use the Bayesian estimators social choice mechanisms. Common choices of Θ and $\mathcal{D}$ are $\mathcal{L}(\mathcal{C})$, $\mathcal{C}$, and $(2^{\mathcal{C}} \setminus \emptyset)$, respectively.

Definition 2 *A statistical decision-theoretic framework for social choice is a tuple $\mathcal{F} = (\mathcal{M}_{\mathcal{C}}, \mathcal{D}, L)$, where $\mathcal{C}$ is the set of alternatives, $\mathcal{M}_{\mathcal{C}} = (\Theta, \Pr)$ is a parametric ranking model over $\mathcal{C}$, $\mathcal{D}$ is the decision space, and $L : \Theta \times \mathcal{D} \to \mathbb{R}$ is a loss function.*

Let $\mathcal{B}(\mathcal{C})$ denote the set of all irreflexive, antisymmetric, and total binary relations over $\mathcal{C}$. For any $c \in \mathcal{C}$, $\mathcal{B}_c(\mathcal{C})$ denote the relations in $\mathcal{B}(\mathcal{C})$ where $c \succ a$ for all $a \in \mathcal{C} - \{c\}$. It follows that $\mathcal{L}(\mathcal{C}) \subseteq \mathcal{B}(\mathcal{C})$ and the Kendall-tau distance can be defined similarly between relations in $\mathcal{B}(\mathcal{C})$.

In the rest of the paper, we will focus on the following two variants of the Mallows model.

Definition 3 (Two variants of the Mallows model) *In the first model, $\mathcal{M}^1_\varphi$, the parameter space is $\Theta = \mathcal{L}(\mathcal{C})$ and given any $W \in \Theta$, $\Pr(\cdot|W)$ is the $\Pr(\cdot|(W, \varphi))$ in the Mallows model.*

In the second model, $\mathcal{M}^2_\varphi$, the parameter space is $\Theta = \mathcal{B}(\mathcal{C})$. For any $W \in \Theta$ and any profile P, we have $\Pr(P|W) = \prod_{V \in P} \left(\frac{1}{Z} \varphi^{\text{Kendall}(V,W)}\right)$, where Z is the normalization factor such that $Z = \sum_{V \in \mathcal{B}(\mathcal{C})} \varphi^{\text{Kendall}(V,W)}$.

In words, $\mathcal{M}^1_\varphi$ is the Mallow model with fixed φ, while $\mathcal{M}^2_\varphi$ extends the parameter space in $\mathcal{M}^1_\varphi$ by including "cyclic" orders. Both $\mathcal{M}^1_\varphi$ and $\mathcal{M}^2_\varphi$ degenerate to Condorcet's model for two alternatives (Condorcet, 1785). The Kemeny rule (for linear orders) is the MLE of $\mathcal{M}^1_\varphi$, for any φ.

We now formally define two statistical decision-theoretic framework associated with $\mathcal{M}^1_\varphi$ and $\mathcal{M}^2_\varphi$, which are the focus of the rest of our paper.

Definition 4 *Let $\mathcal{F}^1_\varphi = (\mathcal{M}^1_\varphi, 2^{\mathcal{C}} \setminus \emptyset, L_{0\text{-}1})$ and $\mathcal{F}^2_\varphi = (\mathcal{M}^2_\varphi, 2^{\mathcal{C}} \setminus \emptyset, L_{0\text{-}1})$, where for any $C \subseteq \mathcal{C}$, $L_{0\text{-}1}(\theta, C) = \sum_{c \in C} L_{0\text{-}1}(\theta, c)$. Let f^1_B (respectively, f^2_B) denote the Bayesian estimator of $\mathcal{F}^1_\varphi$ (respectively, $\mathcal{F}^2_\varphi$) under the uniform prior.*

We note that the 0-1 loss function in the above definition takes a parameter and a decision in $2^{\mathcal{C}} \setminus \emptyset$ as inputs, which is different from the usual 0-1 loss function for parameter estimation that takes a pair of parameters as inputs, as the one in Example 1. Hence, f^1_B and f^2_B are *not* the MLEs of their respective models. In this paper we focus on the 0-1 loss function in Definition 4 to illustrate our framework. Certainly our framework is not limited to 0-1 loss functions.

Example 2 *Bayesian estimators f^1_B and f^2_B coincide with Young (1988)'s idea of selecting the alternative that is* "most likely to be the best (i.e., top-ranked in the true ranking)", *under $\mathcal{F}^1_\varphi$ and $\mathcal{F}^2_\varphi$ respectively. This gives a theoretical justification of Young's idea under our framework.*

Normative Properties of Bayesian Estimators

Due to the space constraint, many proofs are omitted.

Theorem 1 *For any φ, f_B^1 satisfies anonymity, neutrality, and monotonicity. It does not satisfy majority or the Condorcet criterion for all $\varphi < \frac{1}{\sqrt{2}}$,*[2] *and it does not satisfy consistency.*

Proof: Anonymity and neutrality are obviously satisfied.
Monotonicity. Suppose $c \in f_B^1(P)$. Monotonicity is proved by showing that for any profile P' obtained from P by raising the position of c in one vote, $c \in f_B^1(P')$. This is a consequence of the following lemma.

Lemma 1 *For any $W \in \mathcal{L}_c(\mathcal{C})$, $Pr(P'|W) = \varphi \cdot Pr(P|W)$. For any $W' \in \mathcal{L}_c(\mathcal{C})$, $Pr(P'|W') \leq \varphi \cdot Pr(P|W')$.*

Majority and the Condorcet criterion. Let $\mathcal{C} = \{c, b, c_3, \ldots, c_m\}$. We construct a profile P^* where c is ranked in the top positions for more than half of the votes, but $c \notin f_B^1(P^*)$.

For any k, let P^* denote a profile composed of k copies of $[c \succ b \succ c_3 \succ \cdots \succ c_m]$, 1 of $[c \succ b \succ c_m \succ \cdots \succ c_3]$ and $k-1$ copies of $[b \succ c_m \succ \cdots \succ c_3 \succ c]$. It is not hard to verify that the WMG of P^* is as in Figure 2.

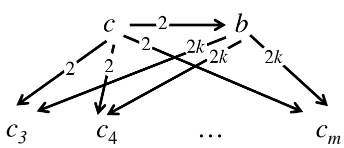

Figure 2: The WMG of the profile P^*, showing positive edges.

Then, we prove that for any $\varphi < \frac{1}{\sqrt{2}}$, we can find m and k so that $\frac{\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P|V)}{\sum_{W \in \mathcal{L}_b(\mathcal{C})} \Pr(P|W)} = \frac{1+\varphi^2+\cdots+\varphi^{2(m-2)}}{1+\varphi^{2k}+\cdots+\varphi^{2k(m-2)}} \cdot \varphi^2 < 1$. It follows that c is the Condorcet winner in P^* but it does not minimize the Bayesian risk under $\mathcal{M}_\varphi^1$, which means that it is not the winner under f_B^1.
Consistency. We construct an example to show that f_B^1 does not satisfy consistency. In our construction m and n are even, and $\mathcal{C} = \{c, b, c_3, c_4\}$. Let P_1 and P_2 denote profiles whose WMGs are as shown in Figure 3, respectively.

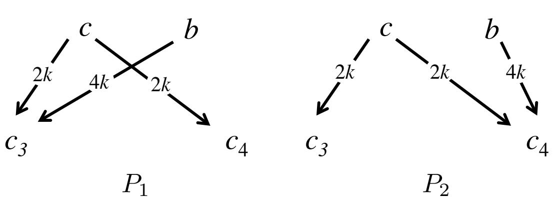

Figure 3: The WMGs of the profile P_1 and P_2.

We provide the following lemma to compare the Bayesian risk of c vs. b.

Lemma 2 *Let $P \in \{P_1, P_2\}$, $\frac{\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P|V)}{\sum_{W \in \mathcal{L}_b(\mathcal{C})} \Pr(P|W)} = \frac{3(1+\varphi^{4k})}{2(1+\varphi^{2k}+\varphi^{4k})}$*

For any $0 < \varphi < 1$, $\frac{3(1+\varphi^{4k})}{2(1+\varphi^{2k}+\varphi^{4k})} > 1$ for all k. It is not hard to verify that $f_B^1(P_1) = f_B^1(P_2) = \{c\}$ and $f_B^1(P_1 \cup P_2) = \{c, b\}$, which means that f_B^1 is not consistent. □

Similarly, we can prove the following theorem for f_B^2.

Theorem 2 *For any φ, f_B^2 satisfies anonymity, neutrality, and monotonicity. It does not satisfy majority, the Condorcet criterion, or consistency.*

By Theorem 1 and 2, the two new voting rules do not satisfy as many desired normative properties as the Kemeny rule (for alternatives). On the other hand, they minimizes Bayesian risk under $\mathcal{F}_\varphi^1$ and $\mathcal{F}_\varphi^2$, respectively, for which Kemeny does neither. In addition, neither f_B^1 nor f_B^2 satisfy consistency, which means that they are not positional scoring rules.

Computational Complexity

We consider the following two types of decision problems.

Definition 5 *In the* BETTER BAYESIAN DECISION *problem for a statistical decision-theoretic framework $(\mathcal{M}_\mathcal{C}, \mathcal{D}, L)$ under a prior distribution, we are given $d_1, d_2 \in \mathcal{D}$, and a profile P. We are asked whether $R_B(P, d_1) \leq R_B(P, d_2)$.*

We are also interested in checking whether a given alternative is the optimal decision.

Definition 6 *In the* OPTIMAL BAYESIAN DECISION *problem for a statistical decision-theoretic framework $(\mathcal{M}_\mathcal{C}, \mathcal{D}, L)$ under a prior distribution, we are given $d \in \mathcal{D}$ and a profile P. We are asked whether d minimizes the Bayesian risk $R_B(P, \cdot)$.*

$\mathsf{P}_{||}^{\mathsf{NP}}$ is the class of decision problems that can be computed by a *P* oracle machine with polynomial number of parallel calls to an NP oracle. A decision problem A is $\mathsf{P}_{||}^{\mathsf{NP}}$-hard, if for any $\mathsf{P}_{||}^{\mathsf{NP}}$ problem B, there exists a polynomial-time many-one reduction from B to A. It is known that $\mathsf{P}_{||}^{\mathsf{NP}}$-hard problems are NP-hard.

Theorem 3 *For any φ,* BETTER BAYESIAN DECISION *and* OPTIMAL BAYESIAN DECISION *for $\mathcal{F}_\varphi^1$ under uniform prior are $\mathsf{P}_{||}^{\mathsf{NP}}$-hard.*

Proof: The hardness of both problems is proved by a unified reduction from the KEMENY WINNER problem, which is $\mathsf{P}_{||}^{\mathsf{NP}}$-complete (Hemaspaandra, Spakowski, and Vogel, 2005). In a KEMENY WINNER problem, we are given a profile and an alternative c, and we are asked if c is ranked in the top of at least one $V \in \mathcal{L}(\mathcal{C})$ that minimizes Kendall(P, V).

For any alternative c, the *Kemeny score* of c under $\mathcal{M}_\varphi^1$ is the smallest distance between the profile P and any linear order where c is ranked in the top. We prove that when $\varphi < \frac{1}{m!}$, the Bayesian risk of c is largely determined by the Kemeny score of c:

Lemma 3 *For any $\varphi < \frac{1}{m!}$ and $c, b \in \mathcal{C}$, if the Kemeny score of c is strictly smaller than the Kemeny score of b, then $R_B(P, c) < R_B(P, b)$ for $\mathcal{M}_\varphi^1$.*

Let t be any natural number with $\varphi^t < \frac{1}{m!}$. For any KEMENY WINNER instance (P, c) for alternatives $\mathcal{C}'$, we add

[2] Whether f_B^1 satisfies majority and Condorcet criterion for $\varphi \geq \varphi < \frac{1}{\sqrt{2}}$ is an open question.

two more alternatives $\{a, b\}$ and define a profile P' whose WMG is as shown in Figure 4 using McGarvey's trick (McGarvey, 1953). The WMG of P' contains the WMG(P) as a subgraph, where the weights are 6 times the weights in WMG(P); for all $c' \in \mathcal{C}'$, the weight of $a \to c'$ is 6; for all $c' \in \mathcal{C}' - \{c\}$, the weight of $b \to c'$ is 6; the weight of $c \to b$ is 4 and the weight of $b \to a$ is 2.

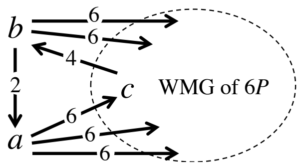

Figure 4: The WMG of P'. $P^* = tP'$.

Then, we let $P^* = tP$, which is t copies of P. It follows that for any $V \in \mathcal{L}(\mathcal{C})$, $\Pr(P^*|V, \varphi) = \Pr(P'|V, \varphi^t)$. By Lemma 3, if an alternative e has the strictly lowest Kemeny score for profile P', then it the unique alternative that minimizes the Bayesian risk for P' and dispersion parameter φ^t, which means that e minimizes the Bayesian risk for P^* and dispersion parameter φ.

Let O denote the set of linear orders over $\mathcal{C}'$ that minimizes the Kendall tau distance from P and let k denote this minimum distance. Choose an arbitrary $V' \in O$. Let $V = [b \succ a \succ V']$. It follows that $\text{Kendall}(P', V) = 4 + 6k$. If there exists $W' \in O$ where c is ranked in the top position, then we let $W = [a \succ c \succ b \succ (V' - \{c\})]$. We have $\text{Kendall}(P', W) = 2 + 6k$. If c is not a Kemeny winner in P, then for any W where d is not ranked in the top position, $\text{Kendall}(P', W) \geq 6 + 6k$. Therefore, a minimizes the Bayesian risk if and only if c is a Kemeny winner in P, and if c does not minimize the Bayesian risk, then b does. Hence BETTER DECISION (checking if a is better than b) and BETTER DECISION (checking if a is the optimal alternative) are $\mathsf{P}_{||}^{\mathsf{NP}}$-hard. □

We note that the OPTIMAL BAYESIAN DECISION for the framework in Theorem 3 is equivalent to checking whether a given alternative c is in $f_B^1(c)$. We do not know whether these problems are $\mathsf{P}_{||}^{\mathsf{NP}}$-complete.

Theorem 4 *For any rational number*[3] φ*,* BETTER BAYESIAN DECISION *and* OPTIMAL BAYESIAN DECISION *for* $\mathcal{F}_\varphi^2$ *under uniform prior are in* P.

The theorem is a corollary of a stronger theorem that provides a closed-form formula for Bayesian risk for $\mathcal{F}_\varphi^2$.[4]

Given a profile P, for any $c, b \in \mathcal{C}$, we let $P(c \succ b)$ denote the number of occurrences of $c \succ b$ in $V \in P$. For any $c, b \in \mathcal{C}$, let $K_{\{c,b\}} = \varphi^{P(c \succ b)} + \varphi^{P(b \succ c)}$.

Theorem 5 *For* $\mathcal{F}_\varphi^2$ *under uniform prior, for any* $c \in \mathcal{C}$*,* $R_B(P, c) = 1 - \prod_{b \neq c} \dfrac{\varphi^{P(b \succ c)}}{K_{\{c,b\}}}$.

Proof sketch: It is equivalent to proving $\sum_{V \in \mathcal{B}_c(\mathcal{C})} \Pr(V|P) = \prod_{b \neq c} \dfrac{\varphi^{P(b \succ c)}}{K_{\{c,b\}}}$. We can show that $\Pr(P) = \sum_{W \in \mathcal{B}_c(\mathcal{C})} \Pr(P|W) \cdot \Pr(W) = \Pr(W) \cdot \frac{1}{Z^n} \cdot \prod_{\{c,b\}} K_{\{c,b\}}$. Then, for any $c \in \mathcal{C}$,

$$
\begin{aligned}
&\sum_{W \in \mathcal{B}_c(\mathcal{C})} \Pr(W|P) = \sum_{W \in \mathcal{B}_c(\mathcal{C})} \Pr(P|W) \cdot \frac{\Pr(W)}{\Pr(P)} \\
=&\frac{\Pr(W)}{\Pr(P)} \cdot \frac{1}{Z^n} \cdot \prod_{b \neq c} \varphi^{P(b \succ c)} \sum_{V' \in \mathcal{B}(\mathcal{C} - \{c\})} \varphi^{\text{Kendall}(P|_{\mathcal{C}_{-c}}, V')} \\
=&\frac{\Pr(W)}{\Pr(P)} \cdot \frac{1}{Z^n} \cdot \prod_{b \neq c} \varphi^{P(b \succ c)} \prod_{b, e \neq c} (\varphi^{P(e \succ b)} + \varphi^{P(b \succ e)}) \\
=&\prod_{b \neq c} \frac{\varphi^{P(b \succ c)}}{K_{\{c,b\}}}
\end{aligned}
$$

□

The comparisons of Kemeny, f_B^1, and f_B^2 are summarized in Table 1. According to the criteria we considered, none of the three outperforms the others. Kemeny does well in normative properties, but does not minimize Bayesian risk under either $\mathcal{F}_\varphi^1$ or $\mathcal{F}_\varphi^2$, and is hard to compute. f_B^1 minimizes the Bayesian risk under $\mathcal{F}_\varphi^1$, but is hard to compute. We would like to highlight f_B^2, which minimizes the Bayesian risk under $\mathcal{F}_\varphi^2$, and more importantly, can be computed in polynomial time despite the similarity between $\mathcal{F}_\varphi^1$ and $\mathcal{F}_\varphi^2$.

Asymptotic Comparisons

In this section, we ask the following question: as the number of voters, $n \to \infty$, what is the probability that Kemeny, f_B^1, and f_B^2 choose different winners?

We show that when the data is generated from $\mathcal{M}_\varphi^1$, all three methods are equal *asymptotically almost surely (a.a.s.)*, that is, they are equal with probability 1 as $n \to \infty$.

Theorem 6 *Let* P_n *denote a profile of* n *votes generated i.i.d. from* $\mathcal{M}_\varphi^1$ *given* $W \in \mathcal{L}_c(\mathcal{C})$*. Then,* $\Pr_{n \to \infty}(\textit{Kemeny}(P_n) = f_B^1(P_n) = f_B^2(P_n) = c) = 1$.

However, when the data are generated from $\mathcal{M}_\varphi^2$, we have a different story.

Theorem 7 *For any* $W \in \mathcal{B}(\mathcal{C})$ *and any* φ*,* $f_B^1(P_n) = \textit{Kemeny}(P_n)$ *a.a.s. as* $n \to \infty$ *and votes in* P_n *are generated i.i.d. from* $\mathcal{M}_\varphi^2$ *given* W*.*

For any $m \geq 5$*, there exists* $W \in \mathcal{B}(\mathcal{C})$ *such that for any* φ*, there exists* $\epsilon > 0$ *such that with probability at least* ϵ*,* $f_B^1(P_n) \neq f_B^2(P_n)$ *and* $\textit{Kemeny}(P_n) \neq f_B^2(P_n)$ *as* $n \to \infty$ *and votes in* P_n *are generated i.i.d. from* $\mathcal{M}_\varphi^2$ *given* W*.*

Proof sketch: The first part is an application of the Central Limit Theorem. For the second part, we sketch a proof for $m = 5$. Other cases can be proved similarly. Let W denote the binary relation as shown in Figure 5.

It can be verified that for all $i \leq 5$, $\Pr(c_i \succ c_{i+1}|W)$ (we let $c_1 = c_6$) are the same and are larger than $1/2$, denoted by p_1; for all $i \leq 5$, $\Pr(c_i \succ c_{i+2}|W)$ are the same and are larger than $1/2$, denoted by p_2. We define a random variable

[3] We require φ to be rational to avoid representational issues.

[4] The formula resembles Young's calculation for three alternatives (Young, 1988). However, it is not clear whether Young's calculation was done for $\mathcal{F}_\varphi^2$.

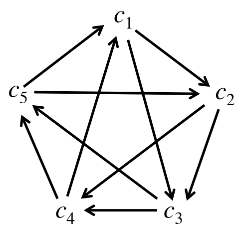

Figure 5: $W \in \mathcal{B}(\mathcal{C})$ for $m = 5$.

$X_{c \succ b}$ for all $c \succ_W b$ such that for any $V \in \mathcal{L}(\mathcal{C})$, if $c \succ_V b$ then $X_{c \succ b} = 1$ otherwise $X_{c \succ b} = -1$.

Lemma 4 *$\{X_{c \succ b} : c \succ_W b\}$ are not linearly correlated.*

Then, it follows from the multivariate Lindeberg-Lévy Central Limit Theorem (CLT) (Greene, 2011, Theorem D.18A) that with positive probability the following hold at the same time in WMG(P_n):

- $0 < w_{P_n}(c_5, c_1) - (2p_1 - 1)n < \sqrt{n}$; $0 < w_{P_n}(c_4, c_1) - (2p_2 - 1)n < \sqrt{n}$.
- $\sqrt{n} < w_{P_n}(c_1, c_2) - (2p_1 - 1)n < 2\sqrt{n}$; $\sqrt{n} < w_{P_n}(c_5, c_2) - (2p_2 - 1)n < 2\sqrt{n}$; $0 < w_{P_n}(c_1, c_3) - (2p_2 - 1)n < \sqrt{n}$.
- For any other $c_i \succ_W c_j$ not mentioned above, $5\sqrt{n} < w_{P_n}(c_i, c_j) - (2\Pr(c_i \succ c_j | W) - 1)n$.

If P_n satisfies all above conditions, then by Theorem 5 $f_B^2(P_n) = \{c_1\}$. Meanwhile, Kemeny$(P_n) = f_B^1(P_n) = \{c_2\}$ with $[c_2 \succ c_3 \succ c_4 \succ c_5 \succ c_1]$ minimizing the total Kendall-tau distance. This shows that $f_B^2(P_n) \neq$ Kemeny(P_n) with non-negligible probability as $n \to \infty$. □

Theorem 6 suggests that, when n is large and the votes are generated from $\mathcal{M}_\varphi^1$, it does not matter much which of f_B^1, f_B^2, and Kemeny we use. Similar observations have been made for other voting rules by Caragiannis, Procaccia, and Shah (2013). On the other hand, Theorem 7 tells us that when the votes are generated from $\mathcal{M}_\varphi^2$, interestingly, for some ground truth parameter f_B^2 is different from the other two with non-negligible probability, and as we will see in the experiments, such probability is quite large.

Experiments

By Theorem 6 and 7, rule f_B^1 and Kemeny are asymptotically equal when the data are generated from $\mathcal{M}_\varphi^1$ or $\mathcal{M}_\varphi^2$. Hence, we focus on the comparison between rule f_B^2 and Kemeny using synthetic data generated from $\mathcal{M}_\varphi^2$ given the binary relation W illustrated in Figure 5.

By Theorem 5, the exact computation of Bayesian risk involves computing $\varphi^{\Omega(n)}$, which is exponentially small for large n since $\varphi < 1$. Hence, we need a special data structure to handle the computation of f_B^2, because a straightforward implementation easily loses precision. In our experiments, we use the following approximation for f_B^2:

Definition 7 *For any $c \in \mathcal{C}$ and profile P, let $s(c, P) = \sum_{b: w_P(b,c) > 0} w_P(b, c)$. Let g be the voting rule such that for any profile P, $g(P) = \arg\min_c s(c, P)$.*

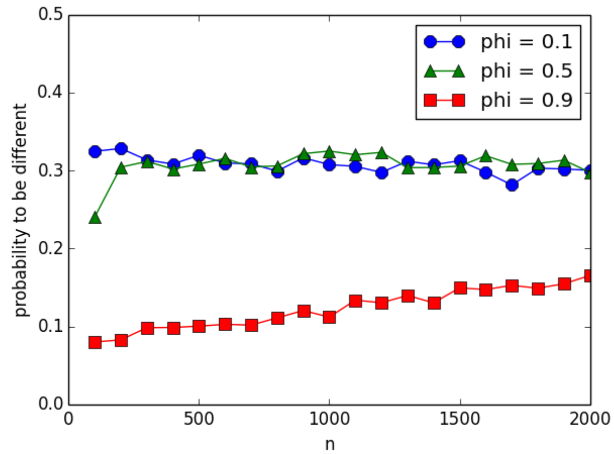

Figure 6: Probability that g is different from Kemeny under $\mathcal{M}_\varphi^2$.

In words, g selects the alternative c with the minimum total weight on the incoming edges in the WMG. By Theorem 5, the Bayesian risk is largely determined by $\varphi^{-s(c,P)}$. Therefore, g is a good approximation of f_B^2 with reasonably large n. Formally, this is stated in the following theorem.

Theorem 8 *For any $W \in \mathcal{B}(\mathcal{C})$ and any φ, $f_B^2(P_n) = g(P_n)$ a.a.s. as $n \to \infty$ and votes in P_n are generated i.i.d. from $\mathcal{M}_\varphi^2$ given W.*

In our experiments, data are generated by $\mathcal{M}_\varphi^2$ given W in Figure 5 for $m = 5$, $n \in \{100, 200, \ldots, 2000\}$, and $\varphi \in \{0.1, 0.5, 0.9\}$. For each setting we generate 3000 profiles, and calculate the percentage for g and Kemeny to be different. The results are shown in Figure 6. We observe that for $\varphi = 0.1$ and 0.5, the probability for $g(P_n) \neq$ Kemeny(P_n) is about 30% for most n in our experiments; when $\varphi = 0.9$, the probability is about 10%. In light of Theorem 8, these results confirm Theorem 7. We have also conducted similar experiments for $\mathcal{M}_\varphi^1$, and found that the g winner is the same as the Kemeny winner in all 10000 randomly generated profiles with $m = 5, n = 100$. This provides a sanity check for Theorem 6.

Conclusions

There are some immediate open questions for future work, including the characterization of the exact computational complexity of f_B^1, the normative properties of g, and the Bayesian risk of f_B^1 and $Kemeny$ under $\mathcal{F}_\varphi^2$.

More generally, it is interesting to study the design and analysis of new voting rules using the proposed statistical decision-theoretic framework under alternative probabilistic models, e.g. random utility models, other loss functions, e.g. a smoother loss function, and other sample spaces including partial orders of a fixed set of k alternatives. We also plan to design and evaluate randomized estimators, and estimators that minimizes the maximum expected loss or the maximum expected regret (Berger, 1985).

References

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2012. Random utility theory for social choice. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 126–134.

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

Caragiannis, I.; Procaccia, A.; and Shah, N. 2013. When do noisy votes reveal the truth? In *Proceedings of the ACM Conference on Electronic Commerce (EC)*.

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

Conitzer, V., and Sandholm, T. 2005. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 145–152.

Conitzer, V.; Rognlie, M.; and Xia, L. 2009. Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, 109–115.

Fishburn, P. C. 1977. Condorcet social choice functions. *SIAM Journal on Applied Mathematics* 33(3):469–489.

Greene, W. H. 2011. *Econometric Analysis*. Prentice Hall, 7th edition.

Hemaspaandra, E.; Spakowski, H.; and Vogel, J. 2005. The complexity of Kemeny elections. *Theoretical Computer Science* 349(3):382–391.

Kemeny, J. 1959. Mathematics without numbers. *Daedalus* 88:575–591.

Lu, T., and Boutilier, C. 2010. The Unavailable Candidate Model: A Decision-theoretic View of Social Choice. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, 263–274.

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

McGarvey, D. C. 1953. A theorem on the construction of voting paradoxes. *Econometrica* 21(4):608–610.

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

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

Thurstone, L. L. 1927. A law of comparative judgement. *Psychological Review* 34(4):273–286.

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

Xia, L.; Conitzer, V.; and Lang, J. 2010. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In *Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 399–406.

Young, H. P., and Levenglick, A. 1978. A consistent extension of Condorcet's election principle. *SIAM Journal of Applied Mathematics* 35(2):285–300.

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