A Statistical Decision-Theoretic Framework for Social Choice

Hossein Azari Soufiani
Google research
azari@google.com

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

Lirong Xia
CS, Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, 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 estimators 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.

1 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 [8].

In many multi-agent and social choice scenarios the main consideration is to achieve the second objective, and make an objectively correct decision. At the same time, we also want to respect agents' preferences and opinions, and require the voting rule to satisfy well-established normative properties in social choice. For example, when a group of friends vote to choose a restaurant for dinner, perhaps the most important goal is to find an objectively good restaurant, but it is also important to use a good voting rule in the social choice sense.

Such scenarios propose the following new challenge: *How to design a voting rule that has good statistical properties as well as social choice normative properties?*

To tackle this challenge, we develop a general framework that adopts *statistical decision theory* [2] 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 [18], one of the most popular models in social choice. In the 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).

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

Table 1: Kemeny (for winners) vs. Bayesian estimators of variants of the Mallows model to choose *winners*. All results for two Bayesian estimators are new.

$\mathcal{M}^2_\varphi$ is a natural model that captures real-world scenarios where the ground truth may contain cycles, or agents' preferences are cyclic, but they have to report a linear order due to 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 and evaluate them with respect to various normative properties, including *anonymity, neutrality, monotonicity, the majority criterion, the Condorcet criterion* and *consistency*. 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.

We also compare the asymptotic outcomes of the two new rules with the Kemeny rule for winners, which is a natural extension of the *maximum likelihood estimator (MLE)* of $\mathcal{M}^1_\varphi$ proposed by Fishburn [11]. It turns out that when n votes are generated according to $\mathcal{M}^1_\varphi$, all three rules select the same winner asymptotically almost surely (a.a.s.) as $n \to \infty$. When the votes are generated according to $\mathcal{M}^2_\varphi$, the new rule for $\mathcal{M}^1_\varphi$ still selects the same winner as Kemeny a.a.s.; however, for some parameters, the winner selected by the new rule for $\mathcal{M}^1_\varphi$ and Kemeny is different from the winner selected by the new rule for $\mathcal{M}^2_\varphi$ with non-negligible probability. Our experiments on synthetic datasets confirm these observations.

1.1 Related work

Along the second perspective in social choice (to make an objectively correct decision), in addition to Condorcet's statistical approach to social choice [8, 25], most previous work views agents' votes as i.i.d. samples from a statistical model, and computes the MLE to estimate the parameters that maximize the likelihood [9, 10, 24, 23, 1, 22, 6]. 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.

There have also been some proposals to go beyond MLE in social choice. In fact, [25] 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 $\mathcal{M}^1_\varphi$ and $\mathcal{M}^2_\varphi$. Pivato [20] conducted a similar study to Conitzer and Sandholm [9], 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 asked by Conitzer and Sandholm [9] and Pivato [20]: How to design new voting rules given particular loss functions and probabilistic models, and what are their normative properties?

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

The statistical decision-theoretic framework is quite general, allowing considerations such as estimators that minimize the maximum expected loss, or the maximum expected regret [2]. In a different context, focused on uncertainty about the availability of alternatives, Lu and Boutilier [17] adopt a decision-theoretic view of the design of an optimal voting rule. Caragiannis et al. [7] studied on the robustness of social choice mechanisms w.r.t. model uncertainty, and characterized a unique social choice mechanism that is consistent w.r.t. a large class of ranking models.

A number of recent papers in computational social choice took utilitarian and decision-theoretical approaches towards social choice [21, 5, 3, 4]. Most of them evaluate the joint decision w.r.t. agents' *subjective* preferences, for example the sum of agents' subjective utilities (i.e. the *social welfare*). In our framework, the join decision is evaluated objectively w.r.t. the ground truth in the statistical model.

A few papers from machine learning community developed algorithms to compute Bayesian estimators for popular ranking models [15, 16]. To the best of our knowledge, no similar study has not been done for the Mallows model, which is one of the most prominent ranking models in social choice and is our focus of this paper. Moreover, we are not aware of previous work discussing social choice normative properties of Bayesian estimators as we do.

2 Preliminaries

In social choice, we have a set of m alternatives $\mathcal{C} = \{c_1, \ldots, c_m\}$ 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 at 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 $P = \{V_1, \ldots, V_n\}$. 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*) [14, 26] 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, denoted by Kemeny$_{\mathcal{C}}$, was due to [11], 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, Kemeny$_{\mathcal{C}}(P) = \{top(V) : V \in$ 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, $c \in r(P)$, 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 Condorcet criterion then it also satisfies the majority 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 $w_P(a, b) = \#\{V \in P : a \succ_V b\} - \#\{V \in P : b \succ_V a\}$.

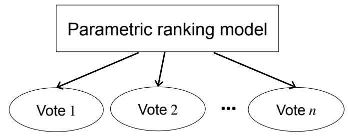

Figure 1: Statistical approach to social choice.

A parametric ranking 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 $\Pr(\cdot|\theta)$.[2]

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 a 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 orders. 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.

The *Mallows model* [18] is defined as follows.

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* $\theta = (W, \varphi)$, $\Pr(P|\theta) = \prod_{V \in P} \frac{1}{Z} \varphi^{\textit{Kendall}(V,W)}$*, where* Z *is the normalization factor with* $Z = \sum_{V \in \mathcal{L}(\mathcal{C})} \varphi^{\textit{Kendall}(V,W)}$.

Statistical decision theory [2] 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.

3 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 $(2^{\mathcal{C}} \setminus \emptyset)$.

[2]This notation should not be taken to mean a conditional distribution over $\mathcal{S}$ unless we are taking a Bayesian point of view.

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* $\Pr(\cdot|(W, \varphi))$ *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* $\Pr(P|W) = \prod_{V \in P} \left(\frac{1}{Z}\varphi^{\mathit{Kendall}(V,W)}\right)$*, where* Z *is the normalization factor such that* $Z = \sum_{V \in \mathcal{B}(\mathcal{C})} \varphi^{\mathit{Kendall}(V,W)}$.

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 [8]. 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 estimators 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 makes it 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 [25]'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.*

The following lemma provides a convenient way to compute the likelihood in $\mathcal{M}^1_\varphi$ and $\mathcal{M}^2_\varphi$ from the WMG.

Lemma 1 *In* $\mathcal{M}^1_\varphi$ *(respectively,* $\mathcal{M}^2_\varphi$*), for any* $W \in \mathcal{L}(\mathcal{C})$ *(respectively,* $W \in \mathcal{B}(\mathcal{C})$*) and any profile* P*,* $\Pr(P|W) \propto \prod_{c \succ_W b} \varphi^{-w_P(c,b)/2}$.

Proof: For any $c \succ_W b$, the number of times $b \succ c$ in P is $(n - w_P(c, b))/2$, which means that $\Pr(P|W) = \varphi^{n^2(n-1)/4} \prod_{c \succ_W b} \varphi^{-w_P(c,b)/2}$. □

4 Normative Properties of Bayesian Estimators

In this section, we compare f^1_B, f^2_B, and the Kemeny rule (for alternatives) w.r.t. various normative properties. We will frequently use the following lemma, whose proof follows directly from Bayes' rule and is omitted due to the space constraint. We recall that $\mathcal{L}_c(\mathcal{C})$ is the set of all linear orders where c is ranked in the top, and $\mathcal{B}_c(\mathcal{C})$ is the set of binary relations in $\mathcal{B}(\mathcal{C})$ where c is ranked in the top.

Lemma 2 *In* $\mathcal{F}^1_\varphi$ *under the uniform prior, for any profile* P *and any* $c, d \in \mathcal{C}$*,* $R_B(P, c) \leq R_B(P, d)$ *if and only if* $\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P|V) \geq \sum_{V \in \mathcal{L}_d(\mathcal{C})} \Pr(P|V)$.

In $\mathcal{F}^2_\varphi$ *under the uniform prior, for any profile* P *and any* $c, d \in \mathcal{C}$*,* $R_B(P, c) \leq R_B(P, d)$ *if and only if* $\sum_{V \in \mathcal{B}_c(\mathcal{C})} \Pr(P|V) \geq \sum_{V \in \mathcal{B}_d(\mathcal{C})} \Pr(P|V)$.

Theorem 1 *For any φ, f_B^1 satisfies anonymity, neutrality, and monotonicity. It does not satisfy majority or the Condorcet criterion for all $\varphi < \frac{1}{\sqrt{2}}$,*[3] *and it does not satisfy consistency.*

Proof: Anonymity and neutrality are obviously satisfied.

Monotonicity. Suppose $c \in f_B^1(P)$. To prove that f_B^1 satisfies monotonicity, it suffices to prove that for any profile P' obtained from P by raising the position of c in one vote, $c \in f_B^1(P')$. We first prove the following lemma for $\mathcal{M}_\varphi^1$.

Lemma 3 *For any $W \in \mathcal{L}_c(\mathcal{C})$, $Pr(P'|W) = Pr(P|W)/\varphi$. For any $W' \in \mathcal{B}_c(\mathcal{L})$, $Pr(P'|W') \leq Pr(P|W')/\varphi$.*

Proof: The first half is proved by noticing that the Kendall$(P', W) =$ Kendall$(P, W) - 1$, and the second half is proved by noticing that Kendall$(P', W') \geq$ Kendall$(P, W') - 1$. □

Therefore, for any $b \neq c$, by Lemmas 2 and 3, we have $\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P'|V) = \varphi \sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P|V) \geq \varphi \sum_{V \in \mathcal{L}_b(\mathcal{C})} \Pr(P|V) \geq \sum_{V \in \mathcal{L}_b(\mathcal{C})} \Pr(P'|V)$, which shows that $c \in f_B^1(P')$.

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, which means that c is the Condorcet winner, but $c \notin f_B^1(P)$.

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

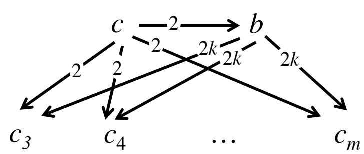

Figure 2: The WMG of the profile P^* where only positive edges are shown.

Lemma 4 $\frac{\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P^*|V)}{\sum_{W \in \mathcal{L}_b(\mathcal{C})} \Pr(P^*|W)} = \frac{1+\varphi^2+\cdots+\varphi^{2(m-2)}}{1+\varphi^{2k}+\cdots+\varphi^{2k(m-2)}} \cdot \varphi^2$

Proof: Let $\mathcal{L}_{-c} = \mathcal{L} - \{c\}$.

$$\begin{aligned} &\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P^*|V) \\ \propto &\varphi^{-m+1} \sum_{V' \in \mathcal{L}(\mathcal{C}_{-c})} \varphi^{\text{Kendall}(P|_{\mathcal{C}_{-c}}, V')} \\ \propto &\varphi^{-m+1} \sum_{t=0}^{m-2} \binom{m-2}{t} t!(m-2-t)!\varphi^{kt}\varphi^{-k(m-2-t)} \qquad (1) \\ \propto &\varphi^{-(m-2)k-m+1} \sum_{t=0}^{m-2} \varphi^{2kt} \end{aligned}$$

In (1), t is the number of alternatives in $\{c_3, \ldots, c_m\}$ ranked below b in V'. There are $\binom{m-2}{t}$ such combinations, for each of which there are $t!$ rankings among alternatives ranked above b and $(m-2-t)!$ rankings among alternatives ranked above t. Notice that there are no edges between alternatives in $\mathcal{C} - \{c, b\}$ in the WMG, which means that for any V' where exactly t alternatives are ranked below b, the probability is proportional to $\varphi^{kt}\varphi^{-k(m-2-t)}$ by Lemma 1.

Similarly, $\sum_{V \in \mathcal{L}_b(\mathcal{C})} \Pr(P^*|V) \propto \varphi^{-k(m-2)+1-(m-2)} \sum_{t=0}^{m-2} \varphi^{2t}$. The lemma follows from these calculations. □

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

Since $\lim_{k,m\to\infty} \frac{1+\varphi^2+\cdots+\varphi^{2(m-2)}}{1+\varphi^{2k}+\cdots+\varphi^{2k(m-2)}} \cdot \varphi^2 = \frac{\varphi^2}{1-\varphi^2}$, for any $\varphi < \frac{1}{\sqrt{2}}$, we can find m and k so that $\frac{\sum_{V\in\mathcal{L}_c(\mathcal{C})}\Pr(P|V)}{\sum_{W\in\mathcal{L}_b(\mathcal{C})}\Pr(P|W)} < 1$, as calculated in Lemma 4. It follows that c is the Condorcet winner in P^* but it does not minimize the Bayesian risk under $\mathcal{M}^1_\varphi$, which means that it is not the winner under f^1_B.

Consistency. We construct an example to show that f^1_B 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.

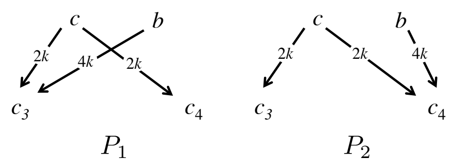

Figure 3: The WMGs of the profile P_1 and P_2. Only positive edges are shown.

We provide the following lemma to compare the Bayesian risk of c vs. d. The proof is similar to the proof of Lemma 4.

Lemma 5 *Let $P \in \{P_1, P_2\}$, $\frac{\sum_{V\in\mathcal{L}_c(\mathcal{C})}\Pr(P|V)}{\sum_{W\in\mathcal{L}_b(\mathcal{C})}\Pr(P|W)} = \frac{3(1+\varphi^{4k})}{2(1+\varphi^{2k}+\varphi^{4k})}$*

Proof: Let $P = P_1$ or P_2.

$$\begin{aligned}&\sum_{V\in\mathcal{L}_c(\mathcal{C})}\Pr(P|V)\\ \propto&\varphi^{-2k}\sum_{V'\in\mathcal{L}(\mathcal{C}_{-c})}\varphi^{\text{Kendall}(P|_{\mathcal{C}_{-c}},V')}\\ \propto&\varphi^{-2k}3(\varphi^{-2k}+\varphi^{2k})\end{aligned}$$

Similarly $\sum_{V\in\mathcal{L}_b(\mathcal{C})}\Pr(P|V) \propto \varphi^{-2k}2(\varphi^{-2k}+1+\varphi^{2k})$. The claim follows from the calculations. □

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 $f^1_B(P_1) = f^1_B(P_2) = \{c\}$. However, it is not hard to verify that $f^1_B(P_1 \cup P_2) = \{c, b\}$, which means that f^1_B is not consistent. This completes the proof of the theorem. □

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

Proof: We will use Theorem 5 in the next section in our proof. Anonymity and neutrality are obvious. The satisfiability of monotonicity has already been proved in the second part of Lemma 3.

Majority and Condorcet criterion. We prove that f^2_B does not satisfy majority or the Condorcet criterion for the same profile P^* as used in the proof of Theorem 1. By Theorem 5, we have:

$$\frac{\sum_{V\in\mathcal{B}_c(\mathcal{C})}\Pr(P^*|V)}{\sum_{W\in\mathcal{B}_b(\mathcal{C})}\Pr(P^*|W)} = (\frac{1+\varphi^{2k}}{1+\varphi^2})^{m-2}\cdot\frac{1+\varphi^{-2}}{1+\varphi^2} \quad (2)$$

For any k, there exits m with (2)< 1, which means that c is not in $f^2_B(P^*)$.

Consistency. We will use the same profiles P_1 and P_2 as in the proof of Theorem 1. For $P = P_1$ or P_2, we have:

$$\frac{\sum_{V\in\mathcal{B}_c(\mathcal{C})}\Pr(P|V)}{\sum_{W\in\mathcal{B}_b(\mathcal{C})}\Pr(P|W)} = \frac{2(1+\varphi^{4k})}{(1+\varphi^{2k})^2} \quad (3)$$

For any k and m, we have that the value of (3) is strictly greater than 1. It is not hard to verify that $f^2_B(P_1) = f^2_B(P_2) = \{c\}$ and $f^2_B(P_1 \cup P_2) = \{c, d\}$, which means that f^2_B is not consistent. □

By Theorem 1 and 2, the two new voting rules do not satisfy as many desired normative properties as the Kemeny rule (for winners). 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.

5 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 polynomial-time many-one reduction from the KEMENY WINNER problem, which was proved to be $\mathsf{P}_{||}^{\mathsf{NP}}$-complete by [13]. 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 $\text{Kendall}(P, V)$.

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 6 *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* $\mathcal{M}_\varphi^1$*.*

Proof: Let k_c and k_b denote the Kemeny scores of c and b, respectively. We have $\sum_{V \in \mathcal{L}_c(\mathcal{C})} \Pr(P|V) > \frac{1}{Z^n}\varphi^{k_c} > \frac{1}{Z^n} m! \varphi^{k_c - 1} \geq \sum_{V \in \mathcal{L}_b(\mathcal{C})} \Pr(P|V)$, which means that $R_B(P, c) < R_B(P, b)$ by Lemma 2. □

We note that φ may be larger than $\frac{1}{m!}$. In our reduction, we will duplicate the input profile so that effectively we are computing the problems for a small φ. Let t be any natural number such that $\varphi^t < \frac{1}{m!}$. For any KEMENY WINNER instance (P, c) for alternatives $\mathcal{C}'$, we add two more alternatives $\{a, b\}$ and define a profile P' whose WMG is as shown in Figure 4 using McGarvey's trick [19]. The WMG of P' contains the $\text{WMG}(P)$ as a subgraph, where the weights are 6 times the weights in $\text{WMG}(P)$; 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.

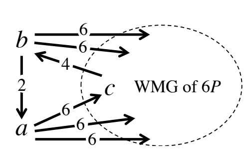

Figure 4: The WMG of P'. $P^* = tP'$.

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 6, 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 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 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, 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 minimizes 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 φ,*[4] 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 proposition that provides a closed-form formula for Bayesian loss for the framework described in Theorem 4.[5]

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 $c \in \mathcal{C}$, $R_B(P, c) = 1 - \prod_{b \neq c} \frac{\varphi^{P(b \succ c)}}{K_{\{c,b\}}}$.*

Proof: It is equivalent to proving $\sum_{V \in \mathcal{B}_c(\mathcal{C})} \Pr(V|P) = \prod_{b \neq c} \frac{\varphi^{P(b \succ c)}}{K_{\{c,b\}}}$. We have:

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

For any $c \in \mathcal{C}$, we have

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

□

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

[5]The formula resembles Young's calculation for three alternatives [25]. However, it is not clear whether Young's calculation was done for $\mathcal{M}_\varphi^2$.

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 $\mathcal{F}_\varphi^2$. This makes f_B^2 a practical voting rule that is also justified by a Mallows-like model.

6 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, $\Pr_{n\to\infty}(\textit{Kemeny}(P_n) = f_B^1(P_n) = f_B^2(P_n) = c) = 1$.*

Proof sketch: It is not hard to see that asymptotically almost surely, for any pair of alternatives $a, b \in \mathcal{C}$, the number of times $a \succ b$ in P_n is $(1 + o(1))n \Pr(a \succ b|W)$. As a corollary of a stronger theorem by [6], as $n \to \infty$, c is the Condorcet winner, which means that $\Pr_{n\to\infty}(\text{Kemeny}(P_n) = c) = 1$.

We now prove a lemma that will be useful for f_B^1 and f_B^2.

Lemma 7 *For any $W \in \mathcal{L}_c(\mathcal{C})$, any alternatives a, b that are different from c, $\Pr(c \succ b|W) > \Pr(a \succ b|W)$.*

Proof: We have $\Pr(c \succ b|W) - \Pr(a \succ b|W) = \Pr(c \succ b \succ a|W) - \Pr(a \succ b \succ c|W)$. For any linear order $V_{c\succ b\succ a}$ where $c \succ b \succ a$, we let $V_{a\succ b\succ c}$ denote the linear order obtained from $V_{c\succ b\succ a}$ by switching the positions of c and a. It follows that $\text{Kendall}(V_{c\succ b\succ a}, W) < \text{Kendall}(V_{a\succ b\succ c}, W)$, which means that $\Pr(c \succ b|W) > \Pr(a \succ b|W)$. □

To prove the theorem for f_B^1, it suffices to prove that for any $b \neq c$ and any $0 < \varphi < 1$, asymptotically almost surely, we have $\sum_{V\in\mathcal{L}_c(\mathcal{C})} \varphi^{\text{Kendall}(P_n,V)} > \sum_{V\in\mathcal{L}_b(\mathcal{C})} \varphi^{\text{Kendall}(P_n,V)}$. For any $V_c \in \mathcal{L}_c(\mathcal{C})$, we let V_b denote the linear order obtained from V_c by exchanging the positions of c and b, which means that $V_b \in \mathcal{L}_b(\mathcal{C})$.

Lemma 8 $\Pr_{n\to\infty}(\textit{Kendall}(P_n, V_c) < \textit{Kendall}(P_n, V_b)) = 1$.

Proof: Given V_c, let $\mathcal{C}'$ denote the set of alternatives between c and b in V_c. We have $\text{Kendall}(P_n, V_b) - \text{Kendall}(P_n, V_c) = \sum_{a\in\mathcal{C}'}[w_{P_n}(a, b) - w_{P_n}(a, c)] + w_{P_n}(c, b) = \sum_{a\in\mathcal{C}'} 2n[\Pr(a \succ b|W) - \Pr(a \succ c|W)] + n(2\Pr(c \succ b|W) - 1) + o(n)$, where we recall that $w_{P_n}(a \succ b) = P_n(a \succ b) - P_n(b \succ a)$. By Lemma 7, for all a that is different from b and c, $\Pr(c \succ a|W) > \Pr(b \succ a|W)$, which means $\Pr(a \succ b|W) - \Pr(a \succ c|W) > 0$. Since c is the Condorcet winner asymptotically almost sure, $\Pr(c \succ b|W) > 1/2$. This proofs the claim. □

By Lemma 8, $\Pr_{n\to\infty}(\forall V_c \in \mathcal{L}_c(\mathcal{C}), \text{Kendall}(P_n, V_c) < \text{Kendall}(P_n, V_d)) = 1$, which means that

$$\Pr_{n\to\infty}\left(\forall V_c \in \mathcal{L}_c(\mathcal{C}), \varphi^{\text{Kendall}(P_n,V_c)} < \varphi^{\text{Kendall}(P_n,V_d)}\right) = 1$$

Hence, $\Pr_{n\to\infty}(\sum_{V\in\mathcal{L}_c(\mathcal{C})} \varphi^{\text{Kendall}(P_n,V)} > \sum_{V\in\mathcal{L}_d(\mathcal{C})} \varphi^{\text{Kendall}(P_n,V)}) = 1$. This proves the theorem for f_B^1.

We use Theorem 5 and Lemma 7 to prove the theorem for f_B^2. We note that $\frac{\varphi^{P_n(b\succ c)}}{K_{\{c,b\}}} = \frac{1}{1+\varphi^{P_n(c\succ b)-P_n(b\succ c)}} = \frac{1}{1+\varphi^{2P_n(c\succ b)-n}}$. By Lemma 7, $\Pr(c \succ b|W) > \Pr(a \succ b|W)$, which means that asymptotically almost surely, we have the following series of reasoning:

(1) $P_n(c \succ b) > P_n(a \succ b)$ for all a, b.

(2) $\frac{1}{1+\varphi^{2P_n(c\succ b)-n}} > \frac{1}{1+\varphi^{2P_n(a\succ b)-n}}$ for all a and b.

(3) $\frac{\varphi^{P_n(b\succ c)}}{K_{\{c,b\}}} > \frac{\varphi^{P_n(b\succ a)}}{K_{\{a,b\}}}$.

(4) For any $a \neq c$, $\prod_{b\neq c} \frac{\varphi^{P_n(b\succ c)}}{K_{\{c,b\}}} > \prod_{b\neq a} \frac{\varphi^{P_n(b\succ a)}}{K_{\{a,b\}}}$.

Finally, applying Theorem 5 to (4), c is the unique winner asymptotically almost surely. This completes the proof of the theorem. □

Theorem 7 *For any $W \in \mathcal{B}(\mathcal{C})$ and any φ, $f_B^1(P_n) =$ 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 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: Due to the Central Limit Theorem, for any $V, W \in \mathcal{B}(\mathcal{C})$, $|\text{Kendall}(P_n, V) - \text{Kendall}(P_n, W)| = \Omega(\sqrt{n})$ a.a.s. By Lemma 1 and Lemma 2, any f_B^1 winner c maximizes $\sum_{V_c \in \mathcal{L}_c(\mathcal{C})} \varphi^{\text{Kendall}(P_n, V_c)} \approx \max_{V_c \in \mathcal{L}_c(\mathcal{C})} \varphi^{\text{Kendall}(P_n, V_c)}$ a.a.s. This means that c is the Kemeny winner a.a.s.

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.

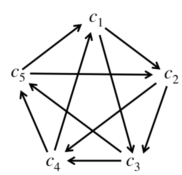

Figure 5: $W \in \mathcal{B}(\mathcal{C})$ for $m = 5$.

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 $X_{c\succ b}$ for any $c \succ_W b$ 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 9 *$\{X_{c\succ b} : c \succ_W b\}$ are not linearly correlated.*

Proof: Suppose for the sake of contradiction $\{X_{c\succ b} : c \succ_W b\}$ are linearly correlated. For any $X_{c\succ b}$ whose coefficient is non-zero, there exists a linear order V where c and b are ranked adjacently. Let V' denote the linear order obtained from V by switching the positions of c and b. We note thta $X_{c\succ b}(V) = -X_{c\succ b}(V')$, and other random variables in $\{X_{c\succ b} : c \succ_W b\}$ take the same values at V and V', this leads to a contradiction. □

Then, it follows from the multivariate Lindeberg-Lévy Central Limit Theorem (CLT) [12, Theorem D.18A] that $\{(\sum_{j=1}^n X_{c\succ b} - pn)/\sqrt{n} : c \succ_W b\}$ converges in distribution to a multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where Σ is the covariance matrix, and is non-singular by Lemma 9. We note that $\sum_{j=1}^n X_{c\succ b} = P_n(c \succ b)$.

Hence, 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 $n \to \infty$, and completes the proof of the theorem. □

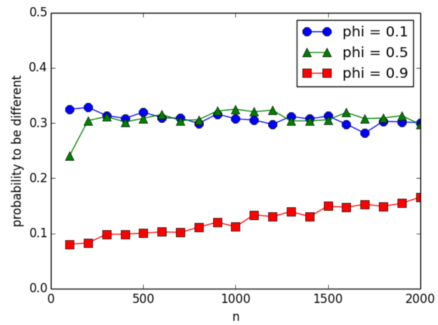

Figure 6: Probability that g is different from Kemeny under $\mathcal{M}_\varphi^2$.

Theorem 6 suggests that, when n is large and the votes are generated from $\mathcal{M}_\varphi^1$, all of f_B^1, f_B^2, and Kemeny will choose the alternative ranked in the top of the ground truth as the winner. Similar observations have been made for other voting rules by [6]. 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 next subsection, we are very confident that such probability is quite large (about 30% for given W shown in Figure 5).

6.1 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, $g(P) = \arg\min_c s(c, P)$.*

In words, g selects the alternative c with the minimum total weight on the incoming edges in the WMG. By Theorem 5, a f_B^2 winner c maximizes $\prod_{b\neq c} \frac{\varphi^{P(b\succ c)}}{K_{\{c,b\}}} = \prod_{b\neq c} \frac{1}{1+\varphi^{w_P(c,b)}}$, which means that c minimizes $\prod_{b\neq c}(1 + \varphi^{w_P(c,b)})$. In our experiments, $\prod_{b\neq c}(1 + \varphi^{w_P(c,b)})$ is $(1 + o(1))\varphi^{\sum_{b:w_P(b,c)>0} w_P(b,c)}$ for reasonably large n. 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 1500 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 $g(P_n) \neq \text{Kemeny}(P_n)$ 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.

7 Conclusions

There are some immediate open questions for future work, including the characterization of the exact computational complexity of f_B^1, and the normative properties of g. 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 [2].

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

[3] Craig Boutilier and Tyler Lu. Probabilistic and Utility-theoretic Models in Social Choice: Challenges for Learning, Elicitation, and Manipulation. In *IJCAI-11 Workshop on Social Choice and Artificial Intelligence*, pages 7–9, 2011.

[4] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. In *ACM Conference on Electronic Commerce*, pages 197–214, Valencia, Spain, 2012.

[5] Ioannis Caragiannis and Ariel D. Procaccia. Voting Almost Maximizes Social Welfare Despite Limited Communication. *Artificial Intelligence*, 175(9–10):1655–1671, 2011.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

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

[13] Edith Hemaspaandra, Holger Spakowski, and Jörg Vogel. The complexity of Kemeny elections. *Theoretical Computer Science*, 349(3):382–391, December 2005.

[14] John Kemeny. Mathematics without numbers. *Daedalus*, 88:575–591, 1959.

[15] Jen-Wei Kuo, Pu-Jen Cheng, and Hsin-Min Wang. Learning to Rank from Bayesian Decision Inference. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 827–836, Hongkong, China, 2009.

[16] Bo Long, Olivier Chapelle, Ya Zhang, Yi Chang, Zhaohui Zheng, and Belle Tseng. Active Learning for Ranking Through Expected Loss Optimization. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 267–274, Geneva, Switzerland, 2010.

[17] Tyler Lu and Craig Boutilier. The Unavailable Candidate Model: A Decision-theoretic View of Social Choice. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, pages 263–274, New York, NY, USA, 2010.

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

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

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

[21] Ariel D. Procaccia and Jeffrey S. Rosenschein. The Distortion of Cardinal Preferences in Voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents*, volume 4149 of *LNAI*, pages 317–331. 2006.

[22] 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.

[23] 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.

[24] Lirong Xia, Vincent Conitzer, and Jérôme Lang. 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)*, pages 399–406, Toronto, Canada, 2010.

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

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