ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Κοινωνική Επιλογή, Κανόνες Ψηφοφορίας και Εκτιμητές Μέγιστης Πιθανοφάνειας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΧΡΥΣΤΑΛΛΑΣ ΠΑΥΛΟΥ

Επιβλέπων: Δημήτριος Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΛΟΓΙΚΗΣ ΚΑΙ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ
Αθήνα, Ιούλιος 2014

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Λογικής και Επιστήμης Υπολογιστών

Κοινωνική Επιλογή, Κανόνες Ψηφοφορίας και Εκτιμητές Μέγιστης Πιθανοφάνειας

Διπλωματική Εργασία

της

ΧΡΥΣΤΑΛΛΑΣ ΠΑΥΛΟΥ

Επιβλέπων: Δημήτριος Φωτάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17η Ιουλίου 2014.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
....	....	....
Δημήτριος Φωτάκης	Ευστάθιος Ζάχος	Αριστείδης Παγουρτζής
Επίκουρος Καθηγητής	Καθηγητής	Επίκουρος Καθηγητής
Ε.Μ.Π.	Ε.Μ.Π.	Ε.Μ.Π.

Αθήνα, Ιούλιος 2014

(Υπογραφή)

....

Χρυσταλλα Παυλου
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Ε.Μ.Π.

Ευχαριστίες

Ολοκληρώνοντας αυτή τη διπλωματική εργασία, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα κ. Δημήτριο Φωτάκη ο οποίος με εμπιστεύτηκε για την εκπόνησή της καθώς και για την πολύτιμή του καθοδήγηση και ενθάρρυνση καθ΄ όλη τη διάρκειά της. Θα ήθελα επίσης να ευχαριστήσω τον κ. Ευστάθιο Ζάχο και κ. Αριστείδη Παγουρτζή, για τις γνώσεις και τη σημαντική βοήθεια που μου προσέφεραν όλα αυτά τα χρονια. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για την αγάπη και τη διαρκή στηριξή τους.

Περίληψη

Σκοπός της διπλωματικής αυτής είναι η μελέτη των κανόνων ψηφοφορίας, μίας γενικής μεθόδου που χρησιμοποιείται για τη συνάθροιση ατομικών προτιμήσεων στην περιοχή της κοινωνικής επιλογής. Συγκεκριμένα, έμφαση δίνεται σε μία καλά μελετημένη προσέγγιση που αντιμετωπίζει τους κανόνες ψηφοφορίας ως εκτιμητές μέγιστης πιθανοφάνειας· δεδομένου ότι υπάρχει μία κρυφή αντικειμενική κατάταξη των υποψηφίων σύφωνα με κάποιο μέτρο και ότι κάθε ψήφος αποτελεί μία θορυβώδη εκτίμηση της πραγματικής κατάταξης, ο κανόνας πρέπει να επιστρέψει την κατάταξη που με τη μέγιστη πιθανότητα ταυτίζεται με την αντικειμενική αλήθεια. Ωστόσο, υποστηρίζεται ότι η απαίτηση μέγιστης πιθανοφάνειας είναι περιοριστική και γι΄ αυτό εξετάζεται επίσης ο αριθμός των ψήφων που απαιτεί κάθε κανόνας ώστε να επιστρέψει με μεγάλη πιθανότητα την πραγματική κατάταξη. Τέλος, λαμβάνοντας υπόψη την τάση των ανθρώπων να επηρεάζονται από τις προτιμήσεις των ανθρώπων στον κοινωνικό τους περίγυρο, εξετάζεται εαν η δομή του κοινωνικού δικτύου επηρεάζει τους εκτιμητές μέγιστης πιθανοφάνειας.

Λέξεις Κλειδιά

Κοινωνική Επιλογή, Υπολογιστική Κοινωνική Επιλογή, Κανόνες Ψηφοφορίας, Εκτιμητές Μέγιστης Πιθανοφάνειας, Πολυπλοκότητα Δειγμάτων, Κοινωνικά Δίκτυα

Abstract

In this thesis, voting, a general method of preference aggregation in social choice theory, is studied. In particular, emphasis is given on a well-studied approach that views the voting rules as maximum likelihood estimators; given that there is an underlying true ranking of the candidates according to a quality measure and the fact that every vote is a noisy estimator of the true ranking, the rule must reconstruct the ranking that is most likely to be the ground truth. However, it is argued that the maximum likelihood requirement is restrictive and thus, a generalization of this framework that studies how many votes a rule requires to output with high probability the ground truth is presented as well. Finally, taking into account the fact that voters are clearly influenced by the preferences of the people who are related to them, it is studied whether the social network structure affects the optimal rules under the maximum likelihood approach.

Keywords

Social Choice Theory, Computational Social Choice, Voting Rules, Maximum Likelihood Estimator, Sample Complexity, Social Networks

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Social choice

It is a fact that people evaluate things in their own unique way and hold different opinions and preferences for almost everything. However, in our everyday lives we face many situations where a collective decision based on the individual preferences must be made. Therefore, there is the need to find methods that will combine the individual preferences into a "right" decision and reach a compromise. For example, we often have to select a leader, the representatives of a group or share probably heterogeneous goods to people with different preferences. The best known way to achieve these goals is by voting, an extremely important component of democracy through the years [81].

Social choice theory provides mathematical models for the above situations and involves the design and analysis of voting methods for the aggregation of individual preferences to a "right" joint decision. The mathematical modeling of voting was founded by Marquis de Condorcet and Chevalier de Borda in the 18th century, continued with Dodgson in the 19th century and rose in the 20th century with the influential work of Arrow.

Historically, social choice has been focusing on political elections and referendums. The recent developments in computer science and networks, however, led to the introduction of social choice to low-stakes applications. As social choice can be used to aggregate people's preferences, in a same way can be used to output a joint decision in multiagent systems that are developed in Artificial Intelligence where a number of intelligent, autonomous and self-interested agents interact and collaborate. For example, in a system developed by Ephrati et al. [35], agents vote to decide their next step in a joint plan. Furthermore, principles from traditional social choice theory are also well applicable to problems in network design, recommendation systems, meta-search engines and electronic commerce applications [81].

As these new applications of social choice are usually associated with an extremely large number of alternatives and information, they point out many interesting computational challenges. Specifically, it is investigated whether the joint decision is easily computed while a beneficial strategic behavior is hard to find. Computational social choice is the area that intents to examine computational aspects of voting and other preference aggregation methods and develop new algorithms and methods that confront the emerging challenges. Thus, it is an interdisciplinary field that captures the mutual interaction between economics and computer science; it adds an algorithmic aspect to traditional social choice theory and at the same time includes the application of social choice techniques to decision making in AI [60].

1.2 Outline of this thesis

There are two basic views of voting [24]: under the first view, the purpose of voting is to achieve a compromise among the idiosyncratic preferences of the agents, returning an outcome that best reflects the social "good", while under the second view, the purpose of voting is to reveal an underlying truth which can be either the best alternative or a ranking according to a quality measure. In other words, some candidates are objectively better than others and this is prior and not dependent on the preferences of the voters. In fact, it is assumed that an agent's preference express how he perceives the ground truth, i.e. it is a noisy estimate of the underlying correct outcome.

In this thesis, emphasis will be given on the second approach. Under this approach, a voting rule is better than another rule if it is more likely to return the ground truth. Thus, the selection of the maximum likelihood estimator as the optimal rule follows naturally; a maximum likelihood estimator returns the outcome that maximizes the likelihood of observing the given votes. By the definition of the maximum likelihood estimator, it follows that it depends on the probability of observing a vote given the ground truth i.e. the noise model. Different noise models define the conditional probabilities of observing any vote given the ground truth with an important assumption; it is usually assumed that the agents' votes are conditionally independent given the ground truth. Having the noise model defined, then the maximum likelihood estimator for the noise model can be computed.

A well-studied noise model is the Mallows model which was firstly introduced by Marquis de Condorcet. The basic assumption of this noise model is that each voter can rank correctly and independently any pair of alternatives with probability $p > 1/2$. One may argue that this assumption is quite optimistic and it is not realistic to expect that in practice all votes will follow Mallows model. In fact, different noise models are expected to arise in practice and in an attempt to predict the kind of noise models that may appear, it is usually supposed that under any realistic model, it will be more probable to observe a ranking that is closer to the ground truth according to some distance metric.

As different noise models are expected to emerge and there is exactly one maximum likelihood estimator per noise model, it can be argued that the maximum likelihood requirement is restrictive. That is, a voting rule which is a maximum likelihood estimator for a specific noise model, given votes that follow another model, may present a very bad behavior. Hence, trying to overcome this restriction, a more general setting is examined. Under this setting, it is studied how many samples different rules need in order to reveal the ground truth with high probability, that is, given noisy estimates of the truth, voting rules return a hypothesis of what the truth is. According to the nature of the voting rule and the accuracy with which is required to learn the truth, the number of required samples changes. For example, some rules require logarithmic samples in the number of alternatives while others require exponential number in order to reconstruct the truth with high probability.

Finally, the assumption of independence between the agents' preferences is disputed. Considering the fact that people's opinions and preferences are clearly influenced by the people that are related to, it is examined how the number of required samples changes by taking the social network structure into account. Specifically, preliminary models give an insight into how the dependencies reflecting the social interaction of the individuals affect the optimal rules and they suggest that social network structure should not be ignored.

1.3 Structure of this thesis

The remainder of this thesis is organized as follows: Chapter 2 of the thesis gives an introduction to Social choice theory. In particular, basic concepts and notations are introduced and prominent aggregation methods are presented. Then, important results such as the Arrow's Theorem and the Gibbard-Satterthwaite Theorem are demonstrated. Chapter 3 views the different voting methods under the maximum likelihood approach which aims not only to reach a compromise among the agents but to reveal a supposed underlying ground truth as well. Different settings, such as the model where agents vote to select a "good" set of alternatives and the case where the input votes are given in the form of partial orders are examined. A generalization of the maximum likelihood approach which investigates how many samples different rules need in order to reveal the truth with high probability, is presented in the chapter 4. Furthermore, in chapter 5, the fact that people's preferences are influenced by the people in their social network is considered and it is investigated whether a social network structure among the agents affects the maximum likelihood approach. Finally, in the last chapter, general conclusions and any further future directions are discussed.

Chapter 2

Preliminaries

In this chapter, important definitions and concepts that will be used in the next chapters as well as some important results in social choice theory, will be introduced.

2.1 The basics

The setting considered in Voting Theory is usually formulated as follows. Assume that there is a finite set of agents $N = \{1, ..., n\}$ and a finite set of alternatives (candidates) A, where $|A| = m$. It is supposed that each agent $i \in N$ uses a linear order $<_i$ on A i.e. a ranking of the alternatives to represent his preferences. A linear order is a transitive, antisymmetric and total relation on A and the set of all linear orders on A is denoted with $L(A)$. Then, a preference profile π given by the n agents is a vector $(\sigma_1, ..., \sigma_n) \in L(A)^n$ where σ_i is the preference of the agent $i \in N$. Given a preference profile, the fundamental question in social choice theory is how the preferences of the agents over the set of the alternatives can be aggregated into one collective preference relation. Social welfare functions which are defined formally below, are used for this purpose.

Definition 1. *A* social welfare function (SWF) *is a function* $f\colon L(A)^n \to L(A)$.

However, in some settings the objective is not to find a collective preference relation but to determine the winner alternative. In these cases, social choice functions are used.

Definition 2. *A* social choice function (SCF) *is a function* $f\colon L(A)^n \to A$.

2.2 Common Social Welfare and Social Choice Functions

In this section some common social welfare and social choice functions will be defined. Since the common rules that are presented below are defined to be the maximizers/minimizers of a kind of score, they are often associated with some tie-breaking schemes.

2.2.1 (Positional) Scoring rules

Scoring rules give a particular score to each alternative every time he is ranked in a particular place and the alternatives with the highest total score win. As scoring rules are simple, they are widely used.

Definition 3. *Every* scoring rule *is associated with a score vector* $s = (s_1, s_2, ..., s_m)$ *with* $s_1 \geq s_2 \geq ... \geq s_m$ *and* $s_1 > s_m$*. Each time an alternative is ranked in the* i^{th} *place, he gets* s_i *points. The alternatives with the highest total sum(summed over all agents) are selected.*

Some popular scoring rules are the following.
Borda rule. The score vector for Borda rule is $(m-1, m-2, ..., 0)$ which means that alternative a gets k points from a voter if the voter prefers a to k other alternatives.
Plurality rule. The score vector for the plurality rule is $(1, 0, ..0)$ which means that the total score of an alternative will be the number of the voters that rank him first.
Veto. The score vector for the veto rule is $(1, 1, .., 1, 0)$ which means that the alternative with the highest score will be the alternative that is vetoed in the smallest number of votes.

Example 2.2.1. *Let* $A = \{a, b, c\}$ *the set of alternatives and the following vote profile consisting of four votes:*

1. $a >_i b >_i c$
2. $c >_{ii} a >_{ii} b$
3. $c >_{iii} a >_{iii} b$
4. $b >_{iv} a >_{iv} c$

Under plurality alternative a gets one point, alternative b gets one, and alternative c gets two points, and hence, c is the plurality winner. Under Borda, alternative a gets five(=2+1+1+1) points, alternative b gets three(=1+0+0+2), and alternative c gets four(=0+2+2+0) , thus a is the Borda winner. Under veto, a gets four points, b and c get two points each. Thus, a is the veto winner (it was not vetoed by any voter). Hence, this example illustrates the differences between scoring rules.

2.2.2 Condorcet Extensions

Another important class of rules is the Condorcet extensions that satisfies a compelling criterion suggested by Marquis de Condorcet. During 18^{th} century, Marquis de Condorcet proposed a criterion to select the winner of the elections; he proposed that the winner should be the alternative that beats every other alternative in the pairwise elections. However, the preferences of the majority may be cyclic and hence, there are profiles that do not have a Condorcet winner. This phenomenon is known as *Condorcet Paradox* [9, 22] and it is shown in the following example.

Example 2.2.2. *Let* $A = \{a, b, c\}$ *and three agents with the following preferences:*

1. $a \geq_i b \geq_i c$

2. $b \geq_{ii} c \geq_{ii} a$

3. $c \geq_{iii} a \geq_{iii} b$

The preferences given by majority in this example are cyclic and a Condorcet winner does not exist as a majority of agents prefer a *to* b*, another majority prefers* b *to* c *and a majority prefers* c *to* a*.*

Although a Condorcet winner does not necessarily exist, returning the Condorcet winner whenever he exists is deemed to be extremely important by many social choice theorists and thus, many voting rules were devised so as to satisfy this property.

Definition 4. *An SCF is called* Condorcet extension *if it selects the Condorcet winner whenever he exists.*

The following rules are some SCF which are Condorcet extensions.
Maximin rule. The maximin rule selects the alternative who has the maximum worst pairwise election against any of the other alternatives. Formally, the Maximin winner is the $\arg\max_{a \in A} \min_{b \in A \setminus \{a\}} n_{ab}$ where n_{ab} is the number of voters who prefer a to b. Maximin rule is Condorcet extension as whenever a Condorcet winner exists, he will have a Maximin score $> \frac{n}{2}$ since he is preferred to all other alternatives by the majority of voters. Every other alternative will have a Maximin score$< \frac{n}{2}$ as he loses at the pairwise election with the Condorcet winner. Thus, the maximin rule will select the Condorcet winner whenever he exists.
Ranked pairs. This rule ranks all pairwise elections by the largest strength of victory first to smallest last. First, the number of voters who prefer a to b for each pair $(a, b), a \neq b$ are computed. Then it locks each pair, starting with the one with the largest number of winning votes, and add one pair in turn to a graph as long as they do not create a cycle. The procedure is continued until all pairs of alternatives have been considered. The Ranked pairs winner is the alternative at the top of the ranking. Ranked pairs is a Condorcet extension as whenever a Condorcet winner a exists, the rule will select the pairs (a, b), $\forall b$ prior to (b, a) as they have score larger than $\frac{n}{2}$ and smaller than $\frac{n}{2}$, respectively. Hence, the Ranked pairs winner will be the Condorcet winner.

Although the Kemeny's rule which is presented below is a SWF, when it is used like an SCF, i.e. when it returns the top alternative of the Kemeny's ranking, it is a Condorcet extension.
Kemeny's rule. Kemeny's rule is an SWF that selects the rankings that minimize the disagreements with the pairwise preferences of the agents. A more formal definition of Kemeny' rule uses the Kendall tau distance between two rankings which is defined as: $d_{KT}(\sigma_1, \sigma_2) = |\{(a, b)|((a >_{\sigma_1} b) \wedge (b >_{\sigma_2} a)) \vee ((b >_{\sigma_1} a) \wedge (a >_{\sigma_2} b))\}|$, i.e. it is the number of pairs that the two rankings disagree. Then, Kemeny's rule gives the rankings in $\arg\min_{\sigma \in L(A)} \sum_{1 \leq i \leq n} d_{KT}(\sigma, \sigma_i)$, where n is the number of voters. As the Kemeny's rule returns a ranking that minimizes the total pairwise disagreements with the voters, if the vote profile has a Condorcet winner then

the Kemeny's winner (the alternative at the top of the output ranking) will be the Condorcet winner. Otherwise, let that a is the Condorcet winner and b is another alternative such that Kemeny's rule return a ranking with $b > a$. As a is Condorcet winner, the number of voters that prefer a to b is greater than $\frac{n}{2}$ and the number of voters that prefer b to a is less than $\frac{n}{2}$. Therefore, if $b > a$ is replaced by $a > b$ a ranking with a smaller number of pairwise disagreements is obtained, which is opposed to the definition of Kemeny's rule. In a similar way, for every alternative b in A the Kemeny's rule will return a ranking with $a > b$ and hence, it will output the Condorcet winner.

Example 2.2.3. *The Kendall Tau distance of the rankings* $\sigma_1 = a > b > c$ *and* $\sigma_2 = b > c > a$ *is* $d_{KT}(\sigma_1, \sigma_2) = 2$ *as they disagree on the pairs* $\{(a,b),(a,c)\}$.

Example 2.2.4. *Let* $A = \{a, b, c\}$ *the set of alternatives. Then beginning from the ranking* $a > b > c$*, the following tree graph illustrates how the rankings are formed while the Kendall Tau distance is increasing (moving a layer down translates into increasing the KT distance by one):*

$$\begin{array}{ccc} & a > b > c & \\ \swarrow & & \searrow \\ b > a > c & & a > c > b \\ | & & | \\ b > c > a & & c > a > b \\ | & & | \\ c > b > a & & c > b > a \end{array}$$

Figure 2.1: Kendall Tau distance diagram of the ranking $a > b > c$.

Example 2.2.5. *Suppose that the set of alternatives is* $A = \{a, b, c\}$ *and consider the profile consisting of the following five votes:*

1. *2 votes:* $a > b > c$
2. *2 votes:* $b > c > a$
3. *1 vote:* $a > c > b$

Under the Maximin rule, alternative a gets 3 points, alternative b gets 2 points, and c gets 1 point, thus a is the Maximin winner. Under the Ranked pairs, each pair gets the following points (in decreasing order): $(b, c) : 4$ *points,* $(a, b) : 3$ *points,* $(a, c) : 3$ *points,* $(b, a) : 2$ *points,* $(c, a) : 2$ *points and* $(c, b) : 1$ *point. Hence, the Ranked pairs will output the ranking* $a > b > c$ *and winner the alternative a. Kemeny's rule will output the ranking* $a > b > c$ *and winner the alternative a as each ranking gets the following score:* $a > b > c : 5$ *points,* $a > c > b : 8$ *points,* $b > a > c : 6$ *points,* $b > c > a : 7$ *points,* $c > a > b : 9$ *points,* $c > b > a : 10$ *points. We observe that all three rules output the Condorcet winner a (three out of five voters prefer a to b and tree out of five voters prefer a to c).*

Although there are many common rules that are Condorcet extensions, it is proved that the popular class of scoring rules is disjoint with the class of Condorcet extensions [39].

Theorem 2.2.1. *There are not any scoring rules that are Condorcet extensions. Equivalently, for every scoring rule there is a preference profile that the rule fails to select the Condorcet winner.*

Example 2.2.6. *Suppose that the set of alternatives is $A = \{a, b, c\}$ and the profile consisting of the following seven votes:*

1. *3 votes: $a > b > c$*
2. *2 votes: $b > c > a$*
3. *1 vote: $b > a > c$*
4. *1 vote: $c > a > b$*

In the above profile a is the Condorcet winner since four out of seven voters prefer a to b and four out of seven voters prefer a to c. Let r a scoring voting rule with $s_1 > s_2 > s_3$. Then the score of alternative $a = 3 \cdot s_1 + 2 \cdot s_2 + 2 \cdot s_3$ is smaller than the score of alternative $b = 3 \cdot s_1 + 3 \cdot s_2 + 1 \cdot s_3$. Hence, the scoring rule r will not output the Condorcet winner.

2.2.3 Other rules

Some other prominent rules that are neither scoring rules nor Condorcet extensions, are introduced.
Single transferable vote (STV) rule. Under STV rule, the election procedure consists of $m-1$ rounds. In each round, the alternative that gets the lowest plurality score (the number of times that he is ranked first among the remaining alternatives) is removed. The alternative is also removed from all votes and the remaining alternatives proceed to the next round. The last remaining alternative is the winner. STV rule is widely used in political elections in many countries such as Scotland and Ireland, Australia and India.
Bucklin's rule. Under Bucklin's rule, the score of each alternative is the minimum position k such that the majority of voters rank the alternative among the first k positions. The Bucklin's winner is the alternative with the minimum Bucklin score.
Plurality with runoff. Elections under Plurality with runoff proceed in two rounds. In the first round, all alternatives except the two with the highest plurality score are removed. Then, the winner of the elections is the alternative that is preferred by the majority of voters in the second round. Plurality with runoff rule is used in Iran, France and North Carolina State.

Example 2.2.7. *Assume that the set of the alternatives is $A = \{a, b, c, d\}$ and the following profile consisting of twenty-six votes:*

1. *10 votes: $a > b > c > d$*

2. 7 votes: $d > a > b > c$

3. 6 votes: $c > d > a > b$

4. 3 votes: $b > c > d > a$

Under Bucklin's rule, alternative a is the winner as a gets two points and b,c,d get three points each. Under STV rule, in the first round b drops out, in the second round d drops out, in the third round c is removed and the last remaining is a who is the STV winner. Under Plurality rule with runoff, however, d is the winner as a and d proceed to the second round and d is preferred to a by sixteen out of twenty-six voters.

2.3 Axiomatic Approach

Since many and different rules do exist, a natural question is which voting methods (SWF or SCF) are considered to be "good". When there are only two alternatives, a "good" voting rule is the majority rule; common sense characterizations [52] suggest that alternative a should be preferred to alternative b if the majority of voters prefer a to b. However, when the number of alternatives is larger than three, majority rule cannot consist an SWF since the preferences of majority may lead to cycles. In addition to this, since the agents' preferences are given in the form of rankings, it does not seem obvious which voting method returns the outcome that best reflects the social good. Therefore, to overcome this difficulty, researchers have proposed some desirable properties that "good" SWF and SCF should satisfy and have classified them by the properties they have.

Some of these desirable properties for SWF are defined as follows.

Definition 5. *An SWF f is* unanimous *if strict unanimous agreement is reflected in the social preference relation. That is, if all agents prefer a certain alternative to another, then so must the resulting social preference order. Formally, if alternative a is ranked above b in all rankings $\sigma_1, .., \sigma_n$ in the preference profile π, then a is ranked higher than b in $f(\pi)$.*

Definition 6. *An SWF f is* independent of irrelevant alternatives (IIA) *if the social preference between any alternatives a and b depends only on the agents' preferences between a and b. Formally, for every alternatives $a, b \in A$ and every rankings $\sigma_1, ..., \sigma_n, \sigma'_1, ..., \sigma'_n$, if $\sigma = f(\sigma_1, ...\sigma_n)$ and $\sigma' = f(\sigma'_1, ...\sigma'_n)$ then $a >_{\sigma_i} b \leftrightarrow a >_{\sigma'_i} b$ for all i implies that $a >_{\sigma} b \leftrightarrow a >_{\sigma'} b$.*

Example 2.3.1. *Assume that an SWF f ranks alternative a above alternative b at the profile consisting of the two preferences:*

1. $a \geq_i b \geq_i c$

2. $b \geq_{ii} c \geq_{ii} a$

If f is IIA, then it ranks alternative a above alternative b at all following profiles:

1. $a \geq_i b \geq_i c$, $b \geq_{ii} a \geq_{ii} c$

2. $a \geq_i b \geq_i c$, $b \geq_{ii} c \geq_{ii} a$

3. $a \geq_i b \geq_i c$, $c \geq_{ii} b \geq_{ii} a$

4. $a \geq_i c \geq_i b$, $b \geq_{ii} a \geq_{ii} c$

5. $a \geq_i c \geq_i b$, $b \geq_{ii} c \geq_{ii} a$

6. $a \geq_i b \geq_i b$, $c \geq_{ii} b \geq_{ii} a$

7. $c \geq_i a \geq_i b$, $b \geq_{ii} a \geq_{ii} c$

8. $c \geq_i a \geq_i b$, $b \geq_{ii} c \geq_{ii} a$

9. $c \geq_i a \geq_i b$, $c \geq_{ii} b \geq_{ii} a$

Definition 7. *An SWF is* non-dictatorial *if there is no agent that can force his preference of any pair of alternatives to the resulting social preference, no matter what the preferences of the other agents are. Formally, an SWF f is non-dictatorial if there is no agent i such that for all preference profiles $\pi = (\sigma_1, .., \sigma_n) \in L(A)^n$ implies that $f(\pi) = \sigma_i$.*

Although these properties seem to be quite natural and one would surely want any good voting method to satisfy them, Arrow showed that these natural axiomatic properties are not compatible and there are not any SWF that simultaneously meet the above criteria when the number of alternatives is larger than two.

Theorem 2.3.1 (Arrow, 1951)**.** *There exists no SWF that is simultaneously IIA, unanimous and non-dictatorial whenever $|A| \geq 3$.*

Arrow's impossibility theorem is one of the most influential results in social choice theory and has given the boundaries on what can be achieved in social choice [63]. In particular, it shows that the main concern is to find ways to escape this impossibility result by relaxing or omitting the desired properties.

Such attempt was given by Young [82] who proposed to weaken the IIA requirement as a way out of the impossibility theorem. He proposed to replace IIA with local IIA which requires IIA to hold only for consecutive pairs of alternatives in the social preference. In other words, if two alternatives are in consecutive positions in the social preference, then the one that was ranked higher must win if we delete all other alternatives from the votes.

Various theorists [83] agreed that IIA is a very strong requirement as all SWF that reduce to majority rule when there are only two alternatives fail the IIA requirement. For example, suppose that there are three alternatives a, b, c and the following preference profile :

1. 25% of agents prefer $a > b > c$

2. 40% of agents prefer $b > c > a$

3. 35% of agents prefer $c > a > b$

There are three possible winners:

1. Winner is a. Then, in the preference profile with the above preferences without b, c would be the winner as 75 % prefer c to a.

2. Winner is b. Then, in the preference profile with the above preferences without c, a would be the winner as 60 % prefer a to b.

3. Winner is c. Then, in the preference profile with the above preferences without a, b would be the winner as 65 % prefer b to c.

As a result, the IIA requirement fails for the rules that reduce to majority when there are only two alternatives.

Rather than relaxing the explicit assumptions of the Arrow's theorem, another way out of the impossibility result, is relaxing the implicit assumptions. One of these attempts is presented in the following subsection.

2.3.1 Utilitarian voting

In this subsection, relaxing the requirement for the use of linear orders will be presented as a way out of the Arrow's impossibility result. In contrast with the above rules, under the utilitarian voting a voter gives a score within a permitted scale to each alternative using a utility function that assigns a number(utility) to each alternative. Then, the scores of all voters are summed and the total score of each alternative is calculated. Winner is the alternative with the highest total score. This kind of voting is called cardinal whereas the rules that take as input rankings of the alternatives are called ordinal. Although a cardinal rule implies an ordinal, the inverse does not hold as infinite cardinal mappings imply the same ordinal. For example, giving the utilities 4, 3, 2, 1 to alternatives a, b, c, d respectively is the same as giving 200, 100, 50, 10 which is the same as giving 100, 1, 0.5, 0. They all imply the ordinal ranking $a > b > c > d$.

According to Claude Hilinger [48] there are three conditions that define utilitarian voting:

1. The voting method must have a voting scale (a, b) and the scores given by the voters should be between this scale.

2. The outcome of an election must be based on the total scores of the alternatives.

3. Every voter should be free to assign to each alternative any of the scores permitted by the voting scale.

2.3.1.1 Examples

Some common cardinal rules are defined below.
Approval voting(AV) [11, 13, 12, 72]. The voting scale for this method is $\{0, 1\}$. The main advantage presented is that it is a simple rule and lets voters put the alternatives into two classes; the class with the alternatives they approve and one with the alternatives they reject. However, it is suggested that a voting scale of two values may be too restrictive.
Range voting(RV) [67]. The valid scores can be any real s such that $a \leq s \leq b$ where the voting scale is (a, b). Although range voting allows voters a wide choice of permitted values, the selection of an appropriate voting scale for the voter needs empirical and experimental study.
Evaluative voting(EV) [38, 46, 47]. The only valid scores for this method are $\{-1, 0, 1\}$. This method is between AV and RV and gives the chance to the voter to discriminate the alternatives with three values.

2.3.1.2 Arrow and utilitarian voting

An important advantage of cardinal voting is that it is a way out of the impossibility result by Arrow. Cardinal voting methods that evaluate alternatives by their total utility are not covered by Arrow's theorem (e.g. [67]) and in fact, the conditions of the Arrow's theorem (as they are restated in [70]) are trivially satisfied by the definition of the utilitarian voting.

Specifically:

1. Unanimity: $s_{ik} \geq s_{ij}, \forall i, \forall j \implies s_k \geq s_j, \forall j$, meaning that if all voters give a higher score to alternative k than any other alternative, then he will have the highest total score.

2. Nondictatorship: There exists no voter i, such that $s_{ih} > s_{ik} \implies s_h > s_k$.

3. Transitivity: The ordering of the alternatives implied by their total scores is transitive.

4. Unrestricted Domain: Each voter can give to each alternative any score permitted by the voting scale.

5. Independence of Irrelevant Alternatives: The total score of any alternative does not depend on the scores given to other alternatives.

2.3.2 Social choice functions

Another approach out of the impossibility result is disputing the use of SWF, i.e. the fact that the rule must return a ranking of the alternatives. As it has been mentioned above, in some settings it is not required to return a ranking of the alternatives but it is desired to identify the most desirable alternatives. In these cases SCF are used. Therefore, it is advisable to study some desirable properties that SCF are expected to satisfy.

Definition 8. *An SCF is* resolute *if there is always a unique winner.*

Definition 9. *An SCF satisfies* anonymity *if the outcome of the rule remains the same after renaming the agents. In other words, the rule is insensitive to the names of the agents and ensures the fairness among them.*

Definition 10. *An SCF satisfies* neutrality *if the outcome of the rule is invariant after renaming the alternatives. That means, the rule is insensitive to the names of the alternatives and ensures the fairness among them.*

Definition 11. *An SCF satisfies* homogeneity *if replicating the votes of any profile does not change the outcome of the election.*

Definition 12. *An SCF satisfies* monotonicity *if improving the position of an alternative in a profile without changing the order of the other alternatives cannot worsen the outcome of the rule for the improved alternative, that is, if the alternative was the winner of the election will still be the winner.*

Definition 13. *An SCF is* non-imposing *if for every alternative there is a preference profile that returns him as winner.*

Definition 14. *An SCF f is* non-dictatorial *if there is no agent i such that for all $\pi \in L(A)^n$, $f(\pi) = a$ where a is the agent's i most preferred alternative. In other words, it is non-dictatorial if there is no agent that can force his top preference to be the top alternative.*

Although neutrality and anonymity are two properties that one may expect that every reasonable SCF should satisfy, it is shown that in general (except for some special cases of n [54]) there is no resolute SCF that satisfies both properties.

Theorem 2.3.2. *For resolute SCF, anonymity is not compatible with neutrality.*

Proof. The above theorem can be shown by considering an election with two alternatives $A = \{a, b\}$ and the vote profile consisting of the two following votes:

1. $a >_i b$
2. $b >_{ii} a$

Without loss of generalization, we suppose that winner is the alternative a. Then, we rename the alternatives $a \leftrightarrow b$. If the rule is neutral, then the winner should be b. If the rule satisfies anonymity, then the winner should be a. Hence, we have a contradiction and the rule cannot satisfy both neutrality and anonymity. □

The following table summarizes whether some of the common voting methods mentioned above satisfy the aforementioned axiomatic properties.

	Scoring r.	Maximin	Ranked pairs	STV	Bucklin	Plurality w. runoff
Anonymity	Y	Y	Y	Y	Y	Y
Neutrality	Y	Y	Y	Y	Y	Y
Homogeneity	Y	Y	Y	Y	Y	Y
Monotonicity	Y	Y	Y	N	Y	N

Table 2.1: Properties of common voting rules.

2.4 Manipulation

One of the biggest concerns in social choice theory is that the agents may be motivated not to report their true preferences in order to manipulate the voting method to a more desired outcome. For example, a voter may have alternative a as his most preferred alternative but if he believes that a will get a very few votes, he may vote for one of the popular alternatives instead. In this way, he tries to ensure that his most preferred popular alternative will get to win.

Example 2.4.1. *Assume that there are are three alternatives $A = \{a, b, c\}$ and three voters with the following preferences:*

1. $a >_i b >_i c$
2. $b >_{ii} a >_{ii} c$
3. $c >_{iii} b >_{iii} a$

Suppose that the voting method is plurality with ties broken in favor of c. Then if voter i misreports his preference as $b >_i a >_i c$, then the winner will be b and voter i will achieve a better result for himself as he prefers b to c.

SCF and SWF with a positive behavior should not be vulnerable to manipulation as many fairness issues arise when there is the possibility of manipulation. There is also the doubt whether the social preference responds to the true preferences of the agents or to the probably distorted reported preferences. As a result, many undesirable outcomes can rise and it is very difficult to predict the outcomes of the elections [32, 78] .

Definition 15. *A resolute voting rule f is* strategyproof *if there is no voter i such that there exist preference profiles $\pi = (\sigma_1, ..., \sigma_n)$ and $\pi' = (\sigma'_1, ..., \sigma'_n)$ with $\sigma_j = \sigma'_j$, $\forall j \neq i$ and $f(\pi') >_i f(\pi)$. In other words, a voting rule is* strategyproof *if there is never a beneficial manipulation for any voter under this rule.*

Unfortunately, manipulation cannot be avoided in general, as every single-valued SCF is susceptible to manipulation when there are more than two alternatives [43, 64].

Theorem 2.4.1 (Gibbard - Satterthwaite). *Let f be a non-imposing, strategyproof, resolute social choice function on A, where $|A| \geq 3$, then f is dictatorial.*

According to the Gibbart-Satterthwaite theorem there is no hope for finding a reasonable strategyproof SCF. Therefore, similarly with Arrow's theorem, the above theorem shows that there is the need to find ways to escape the impossibility result by either relaxing or skipping some of its implicit or explicit requirements.

One of the implicit assumptions of the Gibbart-Satterthwaite theorem is that the social choice function is defined for all possible preference profiles. A research direction investigates how to escape the impossibility result by restricting the preference profiles to domains with desirable properties.

An important contribution in this direction was made by Moulin [55] who showed that if the preference profiles are only profiles who have a Condorcet winner, then the SCF that uniquely selects the Condorcet winner is strategyproof. An example of such domain is the single-peaked preferences. Under this domain, it is assumed that there is a linear ordering $<$ of the alternatives and the agents' preferences are supposed to have a single most preferred alternative. As one moves away from the agent's most preferred alternative, the alternatives will become less preferred for that agent. For example, the agents may be voting over the number of new computers that a school needs to buy and let $A = \{10, 15, 20, 25, 30, 40\}$ the set of the alternatives. Then, a voter that thinks that the school needs 20 computers, he would prefer 15 to 10 and 30 to 40.

When the preferences are single-peaked and the number of the agents is odd, a Condorcet winner always exists. In fact, as it was observed by Black [9], if the preferences of the voters are sorted (according to $<$) by their most preferred alternative, then the top alternative of the median($(n+1)/2$)th) voter, is the Condorcet winner. Hence, the median-voter rule i.e. the rule that returns the top alternative of the median voter, is a non-dictatorial, non-imposing, strategyproof SCF as it follows by the result of Moulin [55] mentioned above; a Condorcet extension in preference profiles that have a Condorcet winner, is strategyproof.

While restricting the domain of the preference profiles provides a breakaway of the impossibility result and many positive results arise, it is not realistic to expect that in all settings the preference profiles will fall in the restricted domain. In fact, in many settings the restrictions are not expected to hold. Hence, it is required to find another way to slide over the Gibbard - Satterthwaite result as there is no control over whether the preferences will actually fall in the restricted domain.

Another approach to circumvent the impossibility result is by using SCF with high manipulation complexity. Inspired by Bartholdi et. al [4], resent research[23, 26, 33, 36, 45] examines how to use computational hardness as a barrier against manipulation. While it is desirable to have SCF by which it is easy to determine the winner of the elections, it is also desirable manipulation to be hard to compute. When defining manipulation problem as a computational problem it is usually assumed that there is a manipulator who already knows the votes of all other agents and wants to determine whether he can make a particular alternative to win. Although one may argue that it is not realistic to assume that the manipulator knows all the other votes, any NP-hardness results are stronger as the case where the voter knows all the other votes is just a special case of the problem where he does not know all votes.

For some voting methods, such as Kemeny's rule which has been proved to be NP-hard to compute [3], predicting the winner is too complicated for everyone. Hence, if finding an effective manipulation is computationally hard, manipulation will not be a problem as the agents will prefer to vote truthfully. Indeed, finding an effective manipulation has been proved to be NP-hard for several voting rules such as STV[5] and ranked pairs[77]. In addition, most common voting rules are known to be hard to manipulate when there is a coalition of manipulators which cooperate in order to make a specific alternative win. As the single-manipulator problem is a special case of the coalition problem, the SCF which are NP-hard to manipulate when there is a single manipulator are also NP-hard to manipulate when there is a coalition of manipulations. However, there are also some other common rules such as maximin [77] and Borda [6, 30] which are hard to manipulate in the case of many manipulators.

There is an important drawback, though, of using complexity to avoid manipulation. While complexity provides some protection against manipulation, it is probably not a strong barrier. Computational complexity measures only the worst case which means that it is not very probable to find an effective way of manipulation for all instances, but there still may be some instances of the manipulation problem that are solved effectively. If this holds, then computational complexity provides only a partial protection. For example, it has been proved that when the preferences are single-peaked, then many of the NP-hard manipulation problems become efficiently solvable [16, 37]. Furthermore, some other research results show that some manipulation problems are often easy to compute [25, 57, 58, 71, 75, 76]. Therefore, while it is desirable to prove that a voting method is hard to manipulate for almost all instances, it is under doubt whether it is possible.

2.5 Scoring rules in statistical analysis

In the previous sections, different rules have been presented in the framework of social choice and voting settings; however it is worth mentioning that scoring rules have applications in many areas such as statistical analysis.

In statistical analysis it is often desirable to make forecasts for the future. Specifically, the forecasts must assign probabilities to different values or events [31]. For example, weather forecasters should predict the probability of rain or not rain on the next day and economic agents should give the probability of increase or decrease in unemployment rate. In many cases, there is the need to have precise information about what different agents (forecasters) believe. However, without incentives, there is the danger that agents misreport their beliefs and give noisy probability distributions.

A scoring rule assesses the accuracy of different probabilistic distributions by assigning a score on the predictive distribution based on the event that materializes [44]. Different scoring rules can be used as either a measure of the accuracy of a probabilistic prediction or as a "cost" function. If the scoring rule is used as a cost function, then the purpose is to report probabilistic distributions so as to minimize the expected cost. Otherwise the purpose is to maximize the expected score (reward). In terms of elicitation [42], the role of scoring rules is to encourage agents to report truthfully their beliefs while in terms of evaluation, the role of scoring rules is to evaluate the quality of forecasts and rank different forecast procedures.

2.5.1 Model

The setting that is usually used [7] assumes that there is an assessor A that assesses the probability distribution of n mutually exclusive and collectively exhaustive statements, where $n > 1$. Let an n-vector $\mathbf{p} = (p_1, ..., p_n)$ representing A's private beliefs where p_i is the probability that A believes that statement i is true. These probabilities are the assessors' "true" state of knowledge but are not directly observable. Let A's public response given by $\mathbf{r} = (r_1, ..., r_n)$ where r_i is the stated probability that statement i is true. Then, if the scoring rule R is used, the expected reward of the assessor A is $\bar{\mathrm{R}}(\mathbf{r}|\mathbf{p}) = \mathbb{E}_p[R_i(\mathbf{r})] = \sum_i p_i \cdot R_i(\mathbf{r})$, where R_i is the score received when i is correct. If the aim of each assessor is to maximize the expected reward $\bar{\mathrm{R}}(\mathbf{r}|\mathbf{p})$, then the optimal response is $\mathbf{r}^* = \arg\max_r \bar{\mathrm{R}}(\mathbf{r}|\mathbf{p})$.

2.5.2 Linear Scoring rule

If a linear scoring rule was used [8], then the optimal solution for the assessor would be to report probability 1 for the most probable statement and 0 for the other statements. For example, if he believes that the true probabilities are $\mathbf{p} = (0.85, 0.15)$ then the optimal response is $\mathbf{r} = (1, 0)$ while if he believes that $\mathbf{p} = (0.49, 0.51)$ the optimal response is $\mathbf{r} = (0, 1)$. Thus, it is clear that the linear scoring rule does not motivate assessors to give their true beliefs. However, there are scoring rules that encourage assessors to report truthfully their beliefs and they are defined in the next subsection.

2.5.3 Proper scoring rules

A scoring rule is proper if the highest expected reward or the minimum expected cost is obtained by reporting the true probabilities. A scoring rule is characterized as strictly proper [68, 61, 66, 73] if it is uniquely optimized by the true probabilities. Formally, a strictly proper scoring rule T is a scoring rule such that the assessor maximizes his expected score by setting $\mathbf{r}^* = \mathbf{p}$.

2.5.4 Examples

Some common strictly proper scoring rules are the following:

Quadratic. $Q_i(\mathbf{r}) = 2r_i - \mathbf{r} \cdot \mathbf{r} \in [-1, -1]$.
Spherical. $S_i(\mathbf{r}) = r_i/(\mathbf{r} \cdot \mathbf{r})^{1/2} \in [0, 1]$.
Logarithmic. $L_i(\mathbf{r}) = \ln(r_i) \in (\infty, 0]$.

The possible scores of each rule differ considerably but as any linear transformation of a strictly proper scoring rule is also strictly proper [68], the above rules can be scaled so as to be easily comparable.

Example 2.5.1. *Assume that we have the logarithmic scoring rule and a weather forecaster predicts that with probability 80% will rain and 20% will not rain. If it rains, the score will be* $L(0.8) = \ln(0.8) = -0.22$. *Otherwise, the score will be* $L(0.2) = -1.6$.

2.5.5 Characterizations

As there are many proper scoring rules, a natural question is which are the best scoring rules. In order to answer that, many desirable properties have defined and it is examined which scoring rules satisfy them. The properties are divided into two categories; the ex ante and the ex post properties [74]. The ex ante properties encourage the assessor to report truthfully their beliefs while the ex post properties evaluate the assessor's performance. While all proper scoring rules provide the ex ante proper property, they differ in the ex post properties. Some important ex post properties are defined below.

Definition 16. *A proper scoring rule satisfies* locality *if the value of the scoring rule depends only on probability assigned to the correct event.*

When there are only two events, all scoring rules are local as the probability of the false event can be found as $1 - r_i$. However, logarithmic scoring rule is proved to be the only local rule [66] for any number of events. A local rule has some practical advantages [7, 8]:

1. A local rule can be presented with a two-dimensional chart that shows the score for any assignment while other rules do not have this advantage except in specific assignments such as the uniform distribution.

2. A local rule has the advantage to give higher scores to probability distributions that assign higher probability to the true statement. Other rules may give higher (lower) score to distributions with lower (higher) probability to the true event which in some cases can be perceived as unfair.

3. Different nonlocal rules may generate different rank orderings among assessors for the same set of assessments. According to [7], Quadratic and Spherical often result in extreme ranking differences compared to Logarithmic which always rank assessors according to the probability they assigned to the correct statement.

Another important property is the effectiveness.

Definition 17. Effectiveness *is satisfied by the rules that encourage agents not only to report their true beliefs but also to report distributions close to the truth for some distance metric[40, 41] .*

In other words, scoring rules are sensitive to distance, meaning that if an assessor does not report his true beliefs, he will prefer to report a distribution that is closer to his true beliefs than a distribution with a bigger distance according to some distance metric. Quadratic and Spherical rules satisfy the effectiveness property while Logarithmic is considered [40, 41] not to be effective.

2.5.6 Nonlinear objectives

The proof that the assessors should respond truthfully under a strictly proper scoring rule is based on the assumption that the aim of each assessor is to maximize his expected reward. However, if instead an assessor has a nonlinear utility function over the expected score, then the Quadratic, Spheric and Logarithmic are no longer strictly proper. Clearly, an assessor with nonlinear utility that reports an assignment $\mathbf{r}^* \neq \mathbf{p}$ is being rational and not necessarily dishonest as he tries to maximize his utility. Bickel [7], though, showed that Logarithmic scoring rule has the best behavior as it is the least affected by this.

2.5.7 Social choice - Decision theory

Scoring rules do not only play an important role in social choice but also in statistics and decision theory. Strictly proper scoring rules evaluate different forecasting procedures and at the same time, they motivate forecasters to report their true assessments. However, as different proper rules do exist, different properties are defined in order to compare the rules and classify them in categories; something that is similar with the axiomatic approach of social choice.

Chapter 3

Maximum Likelihood Approach

In social choice, social welfare and social choice functions have two different goals; achieving a compromise/democracy among the preferences of the agents and revealing the truth respectively. The axiomatic approach mentioned in the previous chapter evaluates different rules by the axiomatic properties they satisfy, aiming in this way to find "fair" rules that can lead to a compromise among the agents. The statistical or maximum likelihood approach, on the other hand, seeks to realize the second goal. This different approach views the voting rules as estimators: it is assumed that there is a hidden ground truth (a correct ranking or a correct winner) and that the votes are noisy estimators of that truth, i.e. votes are the agents' different perceptions of what the correct outcome is.

Maximum likelihood approach which was firstly introduced by Marquis de Condordet [22] has been adopted in economics [2, 69] and it has also recently peaked the interest in computational social choice and AI [24, 34, 59, 51, 27, 79] as according to Procaccia et al. [59] and Mao et. al [51], its prerequisites (an underlying truth) are satisfied by the voting in some crowdsourcing and human computation domains. It is notable that not all voting settings have a correct outcome. For example, it may be the case that voters know all the required information for the alternatives and the different votes are due to the personal circumstances of the agents. In other settings, however, there are some alternatives that are objectively better than others according to a quality measure. For instance, voters may have to evaluate the alternatives' quality of a specific property and rank the alternatives according to this measure. An example of this setting is EteRNA a scientific game where players are called to vote stable molecular designs that will be synthesized in the laboratory.

Under this approach, a good voting rule is a rule that will output the ranking that is most likely to be the underlying truth given the noisy votes and consequently, a natural approach is to choose the maximum likelihood estimator. Although one may think that computing the maximum likelihood estimator is just a problem of statistics, finding a maximum likelihood estimator is not just a problem of this area since it is also desirable that a "good" rule under maximum likelihood approach is also "good" under the axiomatic approach. That is, a maximum likelihood estimator should also satisfy some of the traditional social choice axioms.

Formally, when the truth is either a ranking or an alternative, maximum likelihood estimators are defined as below.

Definition 18. *A* maximum likelihood estimator (MLE) *of the underlying ranking for a given preference profile* $\pi = (\sigma_1, \sigma_2, ..., \sigma_n)$ *is the* $\arg\max_{\sigma \in L(A)} Pr[\pi|\sigma]$.

Definition 19. *A* maximum likelihood estimator (MLE) *of the underlying winner for a given preference profile* $\pi = (\sigma_1, \sigma_2, ..., \sigma_n)$ *is the* $\arg\max_{a \in A} Pr[\pi|a]$.

That means the outcome of the estimator maximizes the probability of observing the noisy samples given that it is the ground truth. As it follows from the definition of MLE, computing the MLE depends on the conditional probabilities of observing a profile given a ground truth. These conditional probabilities are defined by the noise models which are studied in the following subsection.

3.1 Noise Models

Under maximum likelihood approach, as the votes are estimators of the ground truth or equivalently the mistakes that voters make in their evaluation of what the ground truth is, it is supposed that they follow a conditional distribution given the correct ranking. Different conditional distributions constitute different noise models. Usual assumptions are that the votes are conditionally independent given the correct ranking and that all votes follow the same distribution.

Mallows model [50] is one of the most popular noise models and was firstly proposed by Marquis de Condorcet. This model assumes that every voter ranks every pair of alternatives correctly with probability $p > \frac{1}{2}$ and incorrectly with probability $1 - p$. Each voter ranks every pair independently and when a pairwise preference creates a cycle in the voter's current ranking, the process is restarted until a full ranking is formed. Two centuries later, Young [82] showed that the Kemeny's rule is the MLE for Mallows model.

Example 3.1.1. *Let* $A = \{a, b, c\}$ *the set of alternatives and consider the following samples from Mallows model:*

1. $a >_1 b >_1 c$
2. $a >_2 c >_2 b$
3. $b >_3 a >_3 c$

Then, the likelihood of each vote given the different possible ground truths when $p = 0.6$ *are given by the Table 3.1. It can be observed that the ranking* $a > b > c$ *maximizes the likelihood to observe the given profile and thus, MLE would output* $a > b > c$.

Given the noise model, i.e. the probability distributions that votes follow, the MLE can be computed according to definitions 18 and 19. As an MLE is a function from the given preference profile to a ranking or an alternative, it constitutes a SWF or a SCF respectively. Therefore, it is natural to examine which of the common voting rules can be interpreted as maximum likelihood estimators as this different view of the rule may contribute in understanding better the voting rule and consequently,

	vote 1	**vote 2**	**vote 3**	**Total**
$a > b > c$	p^3	$p^2 \cdot (1-p)$	$p^2 \cdot (1-p)$	$p^7 \cdot (1-p)^2 = 4.48 \cdot 10^{-3}$
$a > c > b$	$p^2 \cdot (1-p)$	p^3	$p \cdot (1-p)^2$	$p^6 \cdot (1-p)^3 = 2.98 \cdot 10^{-3}$
$b > a > c$	$p^2 \cdot (1-p)$	$p \cdot (1-p)^2$	p^3	$p^6 \cdot (1-p)^3 = 2.98 \cdot 10^{-3}$
$b > c > a$	$p \cdot (1-p)^2$	$(1-p)^3$	$p^2 \cdot (1-p)$	$p^3 \cdot (1-p)^6 = 8.84 \cdot 10^{-4}$
$c > a > b$	$p \cdot (1-p)^2$	$p^2 \cdot (1-p)$	$(1-p)^3$	$p^3 \cdot (1-p)^6 = 8.84 \cdot 10^{-4}$
$c > b > a$	$(1-p)^3$	$p \cdot (1-p)^2$	$p \cdot (1-p)^2$	$p^2 \cdot (1-p)^7 = 5.89 \cdot 10^{-4}$

Table 3.1: Likelihood of each vote in example 3.1.1.

adapt the rule so as to better fit each setting's needs.

In this direction, Conitzer et. al [24] examine for which popular voting rules exists a noise model such that the voting rule is the MLE for that model. Positional scoring rules when the outcome is either an alternative (SCF) or a ranking(SWF) as well as the STV rule when it returns a ranking of the alternatives, are some common voting rules that can be interpreted as MLEs for some noise models. However, not all common voting rules can be interpreted as MLEs and some negative results do exist [24, 34]. A simple criterion introduced by Conitzer et. al [24] to decide if a rule cannot be interpreted as MLE is presented below.

Lemma 3.1.1. *If there exist preference profiles π_1, π_2 such that the rule f produces the same outcome on π_1 and π_2, but a different outcome on $\pi_1 + \pi_2$, then f cannot be an MLE.*

Using the above condition, some popular rules such as Maximin and Bucklin rule are shown that they cannot be MLEs. However, it remains open whether there are rules that are not MLE but the above lemma cannot be used to show this or if the above lemma is a sufficient condition.

Maximum likelihood estimators as defined above cover the cases where the aim is to reveal the underlying truth which can be either the true ranking of the alternatives or the true winner. However, they do not cover the settings where the aim of the agents is to select a subset of good alternatives. Therefore, there was the need for the maximum likelihood approach to be extended to the problem of selecting a set of alternatives that possess special properties.

3.2 Selecting Sets of Alternatives

Procaccia et al. [59] study the problem of selecting a set of alternatives that provides a maximum likelihood estimator for some desirable properties such as containing the best alternative. Specifically, this voting setting has application in areas where given a number of noisy estimators the aim is to select a subset of the alternatives so as to identify the best alternatives.

For example, during the development of a product different designs are suggested. Then, potential customers are called to vote for the designs they prefer and according to the votes, the most popular designs are manufactured. Finally, the best prototype is selected from a group of experts. In this setting, the aim is to manufacture(select) a set of prototypes that contains one of the objectively best designs.

Another example that this research direction matches with is EteRNA, a scientific game. In EteRNA, the players propose different molecular designs and they vote for a specific number of designs that they evaluate to be stable in order to be synthesized in the lab. Then, the designs that collect the larger number of votes are selected to be materialized in the lab. In this setting there is a ground truth; some designs are objectively stable and some are not. The votes of the players can be assumed to be noisy estimates of the ground truth and the objective is to select a set of designs that is most likely to contain a stable design, i.e. one of the best designs of the ground truth.

The model that is used to capture these situations is described below.

3.2.1 Model

Let $A = \{1, 2, .., m\}$ the set of the alternatives and σ^* the underlying ground truth with a_i the alternative that is on the i^{th} position in the true ranking. Given a number of noisy samples of the ground truth, the objective is to select a subset of alternatives that is a maximum likelihood estimator of a "good" subset, i.e. it satisfies one of the three objectives that will be presented below. The samples are assumed that are taken from the noise models defined as follows.

In the *noisy comparisons model*, each pair of alternatives is presented to n voters (probably different voters for each pair) and every voter independently ranks each pair correctly with probability p, where $\frac{1}{2} < p < 1$. Therefore, this model leads to a dataset D where there are n votes for each pair of alternatives. Under noisy comparisons model computing the MLE is NP-hard [14, 20].

Under *Mallows model*, as it was defined above, the probability of observing a ranking σ is $(1-p)^{d_{KT}(\sigma,\sigma^*)} \cdot p^{\binom{m}{2}-d_{KT}(\sigma,\sigma^*)}$, due to the fact that every voter ranks each pair of alternatives correctly with probability p and incorrectly with probability $1-p$. Under normalization the above probability can be written as $Pr[\sigma|\sigma^*] = \frac{\phi^{d_{KT}(\sigma,\sigma^*)}}{Z_\phi^m}$, where $\phi = \frac{1-p}{p} < 1$ and Z_ϕ^m a normalizing constant.

The *noisy choice model* unifies the noisy comparisons model and the Mallows model and the probability of observing a dataset D is $Pr[D|\sigma^*] = \frac{\gamma^{d(\sigma^*,D)}}{Z_\gamma}$, where $d(\sigma^*, D)$ measures the disagreements for every pair of alternatives between the ground truth and the dataset, γ is the level of noise and Z_γ is a normalizing constant. Formally, $d(\sigma^*, D) = \sum_{a,b \in A, a <_{\sigma^*} b} n_{ba}$, where n_{ba} is the number of voters that prefer b to a. The above equation shows that the probability of observing a dataset D given that the ground truth is σ^* decreases exponentially as its distance from the underlying truth is increased.

As it follows, both noisy comparisons and Mallows model assume that every agent compares correctly with probability $p > 1/2$ any pair of alternatives. Noisy comparisons model, though, captures settings where any voter may be asked to rank only some of the pairs of alternatives while Mallows model captures settings where any voter has to report a full ranking by giving his pairwise preferences. Both noisy comparisons and Mallows model reduce to noisy choice model with $\gamma = \phi = \frac{1-p}{p}$. However, noisy choice model can capture more general settings such as the setting where each voter selects some of the alternatives and reports a ranking only of those.

Example 3.2.1. *Let $A = \{a, b, c, d\}$ the set of alternatives and suppose that the true ranking is $\sigma^* = a >_{\sigma^*} b >_{\sigma^*} c >_{\sigma^*} d$. Then for $n = 3$ the following profile could be observed under noisy comparisons model (where each column gives the n votes for the respective pair):*

(a,b)	**(a,c)**	**(a,d)**	**(b,c)**	**(b,d)**	**(c,d)**
$a > b$	$c > a$	$a > d$	$b > c$	$d > b$	$d > c$
$a > b$	$a > c$	$a > d$	$c > b$	$d > b$	$c > d$
$b > a$	$a > c$	$d > a$	$c > b$	$b > d$	$c > d$

Table 3.2: A vote profile under noisy comparisons model.

The probability of observing the above profile is $p^{10} \cdot (1-p)^8$ as each pairwise vote that agrees with the true ranking appears with probability p and each pairwise vote that disagrees with the true ranking appears with probability $1 - p$.

Under Mallows model, for $n = 3$ the following voting profile could be observed:

1. *$a >_1 b >_1> d >_1 c$ which has probability $p^5 \cdot (1-p)^1$*

2. *$b >_2 c >_2> d >_2 a$ which has probability $p^3 \cdot (1-p)^3$*

3. *$b >_3 a >_3> c >_3 d$ which has probability $p^5 \cdot (1-p)^1$*

The probability of observing the above profile is $p^5 \cdot (1-p)^1 \cdot p^3 \cdot (1-p)^3 \cdot p^5 \cdot (1-p)^1 = p^{13} \cdot (1-p)^5$.

3.2.2 Different objectives

Given samples from the above noise models, finding a good subset will be viewed under three different objectives.

The first objective is to select a k-subset of alternatives that is most likely to include the top alternative of the underlying ranking, that means the k alternatives that are most likely to be the best alternative. Formally, the aim is to select a subset $S \subseteq A$ with k alternatives such that $S \in \arg\max_{S \subseteq A, |S|=k} Pr[a_1 \in S|D]$, where a_1 is the top alternative of the underlying ranking. From Bayes rule it is known that $Pr[a|b] \propto Pr[b|a]$ when a follows uniform a priori distribution and therefore, supposing uniform prior over the rankings, the aim is equivalent with selecting a subset $S \in \arg\max_{S \subseteq A, |S|=k} Pr[D|a_1 \in S]$.

The next objective that will be studied is selecting a k-subset that is most likely to contain the k top alternatives of the true ranking, i.e the k alternatives that are most likely to coincide with the top k alternatives. Formally, the aim is to select a subset $S \subseteq A$ with k alternatives such that $S \in \arg\max_{S \subseteq A, |S|=k} Pr[S = \{a_1, a_2, .., a_k\}|D]$.

The last objective aims to select an ordered tuple of k alternatives that is most likely to coincide with the ordered tuple of the first k alternatives in the ground ranking. That is, it extends the second objective as it is not only required to select the k top alternatives but it is required to select them with their true order. Formally, the aim is to select a k-tuple $(s_1, s_2, .., s_k)$ such that $(s_1, s_2, .., s_k) \in \arg\max_{(s_1,s_2,...,s_k)\in A^k} Pr[a_i = s_i, \forall i \in \{1, ..k\}|D]$.

Choosing $k = 1$, under first objective the aim is to select an alternative that maximizes the probability of being the best alternative and under the second and third objective the aim is to select an alternative that is most likely to be on the first position of the true ranking. Hence, the three objectives coincide for $k = 1$.

For the rest of this section, finding an optimal solution to Objective 1, Objective 2 and Objective 3 will be referred as k-Include Top, k-Unordered Set and k-Ordered Tuple respectively. In addition, the notation $\arg\max^k_{a\in A} g(a)$ is used to denote the set of all k-subsets that include the k alternatives with the highest values under the function g.

3.2.3 Computational Complexity

As the three objectives have defined, the complexity of finding optimal solutions to the three respective problems, k-Include Top, k-Unordered Set and k-Ordered Tuple, will be studied.

Theorem 3.2.1. *For any $k \in \{1, .., m\}$ computing k-Ordered Tuple is NP-hard under noisy comparisons.*

Proof. In order to show that the above problem is NP-hard, it is sufficient to show that a known NP-hard problem is polynomially reduced to this problem. In other words, it is sufficient to show that if there is an algorithm that selects a k subset which optimally satisfies objective 3, then there would be an algorithm that can solve the NP-hard problem. The NP-hard problem that will be used is the minimum feedback arcset in unweighted graphs [20]. Given an unweighted tournament (directed graph with one directed edge between every pair of alternatives), the minimum feedback arcset is the smallest subset of edges such that if they are removed from the graph, a DAG is obtained. Then, this DAG corresponds to a ranking that is denoted as minimum feedback ranking.

Assume that there is an algorithm A_1 that solves k-Ordered Tuple. Then, an algorithm A_2 that solves minimum feedback arcset can be constructed using the following lemma.

Lemma 3.2.1. *The k-tuple that the algorithm A_1 outputs on the voting profile D_T is a k-prefix of a minimum feedback ranking in the tournament T, where D_T is constructed with vertices of T as alternatives and with each edge from i to j of T as a vote $i > j$.*

Proof. Let T an unweighted tournament and D_T the voting profile that is constructed from T. Consider $S = (s_1, .., s_k)$ the result of the algorithm A_1 applied on D_T which means that $S \in \arg\max_{(s_1,..,s_k)\in A^k} Pr[D_T|a_i = s_i, \forall i \in \{1, ..k\}]$. It is wanted to show that S is the prefix of a minimum feedback ranking of T. Instead, suppose that S is not a prefix of any minimum feedback ranking and $S' \in A^k$ is another tuple that is a prefix of a minimum feedback ranking of T.

Assume that σ is a ranking obtained from the tournament T. Then, the feedback of this ranking σ equals $d(\sigma, D_T)$, i.e. the number of disagreements between the voting profile and the ranking is the number of edges in tournament T that if they are eliminated, then σ is formed. Moreover, the probability of observing the profile D_T is $Pr[D_T|a_i = s_i, \forall i \in \{1,..k\}] = \sum_{\sigma \in L(A)|\sigma(i)=s_i, \forall i \in \{1,..k\}} Pr[D_T|\sigma] = \sum_{\sigma \in L(A)|\sigma(i)=s_i, \forall i \in \{1,..k\}} p^{\binom{m}{2}-d(\sigma,D_T)} \cdot (1-p)^{d(\sigma,D_T)}$, where the second transition follows the fact that the event $a_i = s_i, \forall i \in \{1,..k\}$ is the union of the disjoint events that one of the rankings $\in \{\sigma \in L(A)|\sigma(i) = s_i\}$ is the ground truth. The third transition follows the definition of the noisy comparisons model.

Suppose that the minimum feedback is f^*. Then there is a ranking with feedback f^* and k-prefix S'. Therefore, $Pr[D_T|a_i = s'_i, \forall i \in \{1,..k\}] \geq p^{\binom{m}{2}-f^*} \cdot (1-p)^{f^*}$. Since S is not a prefix of any minimum feedback ranking, every ranking that has S as prefix has at least feedback $f^* + 1$.

Hence, $Pr[D_T|a_i = s_i, \forall i \in \{1,..k\}] \leq (m-k)! \cdot p^{\binom{m}{2}-f^*-1} \cdot (1-p)^{f^*+1}$, as there are $(m-k)!$ rankings that have $(s_1, .., s_k)$ as a prefix with minimum distance f^*+1. Thus, if $p > \frac{(m-k)!}{1+(m-k)!}$, $Pr[D_T|a_i = s'_i, \forall i \in \{1,..k\}] > Pr[D_T|a_i = s_i, \forall i \in \{1,..k\}]$which is opposed to the assumption that S is an optimum solution to Objective 3. As a result, S is indeed a prefix of a minimum feedback ranking. □

As the result of algorithm A_1 is a k-prefix of a minimum feedback ranking, we can use it to find a minimum feedback ranking. Specifically, given a tournament T and a vote profile D_T the algorithm A_2 can be constructed using A_1 as follows:

1. Apply A_1 on D_T and let ρ_1 the outcome.
2. Construct a tournament T' by removing all vertices of ρ_1 from T and adding k dummy vertices.
3. Add edges from every non-dummy vertex to all dummy vertices in T' and arbitrary edges between the dummy vertices.
4. Construct the voting profile D'_T.
5. Repeat steps 1-4 for $\lceil \frac{m}{k} \rceil$ times.
6. Return the ranking of the first m vertices in $\rho_1\rho_2..\rho_{\lceil \frac{m}{k} \rceil}$, where ρ_i is the result of the i^{th} iteration.

Induction can be used to prove that the first m vertices in $\rho_1\rho_2..\rho_{\lceil \frac{m}{k} \rceil}$ form a minimum feedback ranking. Any suffix of a minimum feedback ranking is a minimum feedback ranking of the tournament restricted on that alternatives. Otherwise, it could be replaced by a smaller feedback ranking and get a different ranking with a smaller total feedback which is opposed to the assumption that it was a minimum feedback ranking from the beginning. Hence, by induction it follows that the first

$m-k$ alternatives in $\rho_2..\rho_{\lceil \frac{m}{k} \rceil}$ form a minimum feedback ranking and concatenating it with ρ_1 we get the minimum feedback. It is also notable that the first m alternatives of $\rho_1..\rho_{\lceil \frac{m}{k} \rceil}$ are the original m alternatives. This is due to the construction of T' as there is an edge from every non-dummy vertex to all the dummy vertices and therefore, the non-dummy vertices are selected as it is higher the probability of the non-dummy vertices being ranked higher than the dummy vertices.

As an algorithm that solves minimum feedback arcset is constructed by using polynomial times the algorithm that solves k-Ordered Tuple, the minimum feedback arcset reduces to k-Ordered Tuple and thus finding an optimal solution to Objective 3 is NP-hard. □

As solving k-Ordered Tuple for $k = 1$ is the same as solving 1-Include Top and 1-Unordered Set the corollary below follows.

Corollary 3.2.1. *For m alternatives and $k = 1$, both k-*Include Top *and k-*Unordered Set *are NP-hard under noisy comparisons.*

The above corollary can be used to extend the NP-hardness of k-Include Top and k-Unordered Set for $k = 1$ to any $k \in \{1, ..., m-1\}$.

Theorem 3.2.2. *k-Include Top with m alternatives and $k \in \{1, ..., m-1\}$ is NP-hard under noisy comparisons.*

Proof. The proof consists of two parts. The first part is to show that k-Include Top with $k = 1$ reduces to k-Include Top with $k \in \{1, .., m/2\}$. Let T_1 an instance of k-Include Top with $k = 1$ and $m - t + 1$ alternatives with $t \in \{1, .., m/2\}$. A T_2 instance can be created with m alternatives and $k = t$ by adding $t - 1$ extra alternatives. In order to complete the voting profile for the pairs with the extra alternatives n preferences $a > b$ are added for every pair a, b such that a is one of the extra alternatives and b is an initial alternative. Arbitrary preferences are also added between each pair a, b with both a and b extra alternatives. It is obvious that the $t-1$ extra alternatives will be selected as they are preferred by all votes to any other initial alternative and hence, the likelihood of being the top alternative is higher. The last alternative of the t-tuble will be one of the initial alternatives and hence, it is the initial alternative with the highest probability of being the top alternative, i.e. he is the solution in T_1. It follows, thus, that k-Include Top with $k \in \{1, .., m/2\}$ is NP-hard.

The second part is to show that k-Include Top with $k = t \in \{1, .., m/2\}$ is equivalent to k-Include Top with $k = m - t$. Let T_1 an instance of k-Include Top with $k = t \in \{1, .., m/2\}$. Then the solution consists of the k alternatives with the highest probability of being the top alternative. Let T_2 an instance of k-Include Top with $k = m - t$ which is constructed from T_1 by reversing all preferences. Then, the $m - t$ alternatives most likely to be the best alternative are the $m - t$ alternatives in T_1 which weren't selected, i.e. the $m - t$ alternatives which were less likely to be the best alternative in T_1. In a same way, an instance of k-Include Top with $k = m - t \geq m/2$ can be reduced to to an instance of k-Include Top with $k \leq m/2$. Therefore, k-Include Top with $k \in \{1, .., m/2\}$ is equivalent to $m - k$-Include Top. As a result, k-Include Top with $k \in \{1, .., m-1\}$ is NP-hard. □

In a similar way, the theorem below follows.

Theorem 3.2.3. *k-Unordered Set is NP-hard for $k \in \{1, .., m-1\}$ under noisy comparisons.*

Except from samples from noisy comparisons, NP-hardness can also be proved for samples from Mallows model using the reduction of computing the Kemeny ranking (which is NP-hard [3]) to the above three objectives.

Theorem 3.2.4. *k-Include Top and k-Unordered Set with $k \in \{1, .., m-1\}$, and k-Ordered Tuple with $k \in \{1, .., m\}$ are NP-hard under Mallows model.*

As Mallows model and noisy comparisons are special cases of noisy choice model, it follows that the three problems k-Include Top, k-Unordered Set and k-Ordered Tuple are NP-hard under noisy choice model.

3.2.4 Finding Solutions

One natural approach to find the optimal solution to Objective 1 would be to compute the MLE ranking and then select the k top alternatives. As Young [82] showed for $k = 1$ when p is close to 1, this method indeed gives an optimal solution. Lemma 1.2 and Theorem 1.1 also show that when p is close to 1 ($p > \frac{(m-k)!}{(m-k)!+1}$) the optimal solution is a k-prefix of the minimum feedback ranking in noisy comparisons or a k-prefix of the Kemeny ranking which are respectively the MLE rankings. However, when p is close to $\frac{1}{2}$ this is not the case. Young showed [82] with an example that for $k = 1$ the optimal solution is given by Borda rule and does not coincide with the top alternative of the MLE ranking. Procaccia et. al [59] extend this case (p close to $\frac{1}{2}$ or equivalently γ close to 1) to any $k \in \{1, .., m-1\}$ using an extended scoring method. Under the extended scoring method the score of any alternative a is given by $sc(a) = \sum_{b \in A \setminus \{a\}} n_{ab}$ where n_{ab} is the number of votes that prefer a to b.

Theorem 3.2.5. *For every n and m there exists $\gamma' < 1$ such that for all $\gamma \geq \gamma'$, the optimal solutions to Objective 1 under the noisy choice model are in $\arg\max_a^k sc(a)$.*

Proof. The probability of observing voting profile D given that the top alternative is $a^* = a$ is the following:

$$Pr[D|a^* = a] = \sum_{\sigma \in L(A) | \sigma(a)=1} Pr[D|\sigma^* = \sigma] = \sum_{\sigma \in L(A) | \sigma(a)=1} \frac{\gamma^{d(\sigma,D)}}{Z_\gamma},$$ where the second transition follows the fact that alternative a is top alternative if one of the rankings that have alternative a as top alternative is the true ranking and the third transition follows the probability distribution of the noisy choice model.

Let $f(a) = Pr[D|a^* = a] \cdot Z_\gamma$. As the optimal solution to Objective 1 is $\arg\max_a^k Pr[D|a^* = a]$, it follows that the optimal solution is also $\arg\max_a^k f(a)$ since $\arg\max_a^k Pr[D|a^* = a] = \arg\max_a^k f(a)$.

Using the inequality $(1-\epsilon)^t \geq 1 - \epsilon \cdot t, \forall t \in \mathbb{N}$ and that $\gamma = 1 - \epsilon, \epsilon \in [0, 1)$ the following comes:

$$f(a) = \sum_{\sigma \in L(A) | \sigma(a)=1} \gamma^{d(\sigma,D)} = \sum_{\sigma \in L(A) | \sigma(a)=1} (1-\epsilon)^{d(\sigma,D)} \geq \hat{f}(a) =$$

$$\sum_{\sigma \in L(A) | \sigma(a)=1} (1 - \epsilon \cdot d(\sigma, D))$$

The difference between $f(a)$ and $\hat{f}(a)$ can be upper bounded using the following inequality:
$|(1-\epsilon)^t - (1-t\cdot\epsilon)| \leq \sum_{i=2}^{t} \binom{t}{i} \cdot \epsilon^i \leq 2^t \cdot \epsilon^2$, where the second transition holds due to the expansion $(1-\epsilon)^t = \sum_{i=0}^{t} \binom{t}{i} \cdot (-\epsilon)^i$ and the last transition follows the property of binomial coefficients $\sum_{i=0}^{t} \binom{t}{i} = 2^t$.

Using the above inequality for $t = d(\sigma, D)$ the following comes:
$f(a) - \hat{f}(a) \leq \sum_{\sigma \in L_(A) | \sigma(a)=1} 2^{d(\sigma,D)} \cdot \epsilon^2 \leq \epsilon^2 \cdot (m-1)! \cdot 2^{n \cdot \binom{m}{2}}$, where the last transition holds as there are $(m-1)!$ rankings that have as top alternative a and $\max d(\sigma, D) = n \cdot \binom{m}{2}$, that is all voters of D disagree with every pair of σ.

Lemma 3.2.2. *For every $a \in A$, $\hat{f}(a) = C_\epsilon + \epsilon \cdot (m-1)! \cdot sc(a)$ where C_ϵ depends only on ϵ.*

Lemma 2.2 shows that as $\hat{f}(a)$ is a linear transformation of $sc(a)$ we get that $\arg\max_a^k \hat{f}(a) = \arg\max_a^k sc(a)$. Hence, in order to complete the proof of theorem 2.4 we have to show that $\arg\max_a^k f(a) \subset \arg\max_a^k \hat{f}(a)$. In order to show that, it is sufficient to show that for every $a, a' \in A$ such that $\hat{f}(a) > \hat{f}(a')$ we have that $f(a) > f(a')$. From lemma 2.2. it follows that if $\hat{f}(a) > \hat{f}(a')$ then $sc(a) \geq sc(a')+1$ and hence, $\hat{f}(a) \geq \hat{f}(a') + \epsilon \cdot (m-1)!$. Thus, $f(a) \geq \hat{f}(a) \geq \hat{f}(a') + \epsilon \cdot (m-1)! \geq f(a') - \epsilon^2 \cdot (m-1)! \cdot 2^{n \cdot \binom{m}{2}} + \epsilon \cdot (m-1)!$, where setting that $\epsilon < 2^{-n \cdot \binom{m}{2}}$ we have that $f(a) > f(a')$ as required. □

Under Mallows model, $sc(a)$ reduces to Borda score as $sc(a) = \sum_{b \in A \setminus \{a\}} n_{ab} = \sum_{i=1}^{n} (m - \sigma_i(a))$, that is the number of alternatives that a beats summed over all votes which equals Borda score. Hence, the optimal solution when given very noisy samples of Mallows model is just picking the k alternatives with the highest Borda scores. This is an extension of Young's result for $k = 1$ to any $1 \leq k \leq m-1$.

Example 3.2.2. *Let $A = \{a, b, c, d\}$ the set of alternatives and the following noisy samples ($p \simeq 1/2$) from Mallows model.*

1. $a > b > c > d$
2. $b > a > d > c$
3. $c > a > b > d$
4. $a > b > d > c$

Then, the Borda scores of every alternative are the following: $sc(a) = 10, sc(b) = 8, sc(c) = 4, sc(d) = 2$. Therefore, for $k = 1$ the candidate most likely to be the top alternative is a.

Furthermore, extended scoring method also gives optimal solutions to Objective 2 when p is close to $\frac{1}{2}$. Thus, the alternatives selected for Objective 1 are not only likely to be the top alternative but as a whole, they are also likely to coincide with the k best alternatives.

Theorem 3.2.6. *For every n and m there exists $\gamma' < 1$ such that for all $\gamma \geq \gamma'$, the optimal solutions to Objective 2 under the noisy choice model are in $\arg\max_a^k sc(a)$.*

Optimal solutions to Objective 3 when p is close to $\frac{1}{2}$ can also be given by an easily computable method; the scoring tuples method. Under this method the score of each k-tuple is computed as follows:

$sc(a_1, a_2, ..., a_k) = \sum_{i=1}^{k} sc(a_i) - d((a_1, a_2, ..., a_k), D)$ where $d((a_1, a_2, ..., a_k), D)$ $= \sum_{1 \leq i \leq j \leq k} n_{a_j a_i}$, that is the number of votes that disagree with the order of the alternatives in the k-tuple. Then, solutions to Objective 3 are the tubles that maximize their score under this method.

Theorem 3.2.7. *For every n and m there exists $\gamma' < 1$ such that for all $\gamma \geq \gamma'$, the optimal solutions to Objective 3 under the noisy choice model are in* $\arg\max_{(a_1, a_2, ..., a_k) \in A^k} sc(a_1, a_2, ..., a_k)$.

While extended scoring and scoring tuples methods are proved to give optimal solutions only when the samples are very noisy (p close to 1/2), simulations show that these methods have good accuracy when p is greater, i.e. when the noise is lower. Thus, these easily computed results can be used to improve the quality and the speed of human computations and affect positively the performance of many human - computer systems.

3.3 Partial Orders

Until now, it has been assumed that every voter can compare any pair of alternatives. However, this is not always the fact. Partial orders are desirable for two important reasons. Voters may not be able to compare a pair of alternatives or it may not be possible and efficient to rank all the alternatives due to the large number of the alternatives. For example, while it is easy and reasonable to compare two cars, it may be difficult to compare a car and a motorcycle and one may just want to declare them as incomparable. In addition, a voter may have different preference criteria which when combined may lead to non-total orders. For example, one may wants a fast and a cheap car, so a 300 km/h car which costs 100000 euros is incomparable to a 180km/h which costs 30000 euros.

Xia et al. [80] extend the previous results in MLE approach to voting with partial orders by introducing the following model.

3.3.1 Model

Let $A = \{a_1, a_2, .., a_m\}$ the set of the alternatives and o^* the underlying ground truth which can be either an alternative or a ranking. Under *pairwise-independent* *model*, votes are drawn conditionally independently given the ground truth and each voter votes independently for every pair according to the following probability distribution:

1. $Pr[a_i > a_j|o^*]$ is the probability of ranking alternative a_i higher than a_j given the ground truth

2. $Pr[a_i < a_j|o^*]$ is the probability of ranking alternative a_i lower than a_j given the ground truth

3. $Pr[a_i \sim a_j|o^*]$ is the probability of alternatives a_i and a_j being incomparable given the ground truth

where $Pr[a_i > a_j|o^*] + Pr[a_i < a_j|o^*] + Pr[a_i \sim a_j|o^*] = 1$.

Due to the independence mentioned above, the probability of observing a vote v is $Pr[v|o^*] = \prod_{1\leq i\leq j\leq m} Pr[v_{ij}|o^*]$, where v_{ij} is the vote restricted only on alternatives i and j. Furthermore, because of the independence among the voters, the probability of observing a profile $\pi = (v_1, .., v_n)$ of n votes is $Pr[\pi|o^*] = \prod_{1\leq i\leq n} Pr[v_i|o^*]$. Then, the objective under MLE approach is to select an outcome that maximizes the probability of observing the vote profile, i.e. $\arg\max_{o\in O} Pr[\pi|o]$, where O can be either $L(A)$ or A.

3.3.2 Pairwise scoring rules

When the samples are taken under pairwise-independent noise model, the MLE is a pairwise scoring rule which is defined as follows.

Definition 20. *A* pairwise scoring function *is a function s: $A \times A \times O \to \mathbb{R}$ with value $s(a, a, o) = 0$ for any $a \in A$ and for any outcome $o \in O$. For any partial order v and any outcome o, $s(v, o)$ can be computed as $s(v, o) = \sum_{(a_i,a_j)\in v} s(a_i, a_j, o)$. Then for any profile $\pi = (v_1, .., v_n)$ of partial orders, $s(\pi, o)$ is the total s score of all partial orders in the profile π, i.e. $s(\pi, o) = \sum_{1\leq i\leq n} s(v_i, o)$.*

Definition 21. *A pairwise scoring rule r_s is defined as $r_s(\pi) = \arg\max_{o\in O} s(\pi, o)$ with the profile of partial orders π as input.*

That is, a pairwise scoring rule for a given profile and pairwise function returns as outcome the linear ranking or alternative that obtains the highest pairwise score for the given input.

The following definitions are required to present the relation between MLEs and pairwise scoring rules.

Definition 22. *A pairwise scoring function s is* weakly neutral *if for any pair of outcomes o,o' there exists a permutation M such that for any pair of alternatives $a_i, a_j \in A$, $s(a_i, a_j, o) = s(M(a_i), M(a_j), o')$.*

That means a pairwise scoring function is weakly neutral if for every pair of outcomes, when the first outcome obtains a score s with a pair of alternatives, then there is another pair of alternatives such that the second outcome gets the same score s. In other words, the score function treats the outcomes in an equal fair way.

Definition 23. *A pairwise noise model is* weakly neutral *if for any pair of outcomes* o,o' *there exists a permutation* M *such that for any pair of alternatives* $a_i, a_j \in A$, $Pr[a_i > a_j|o] = Pr[M(a_i) > M(a_j)|o']$.

That means a pairwise scoring function is weakly neutral if for every pair of outcomes, when given the first outcome there is a probability p of observing $a_i > a_j$, then there is another pair of alternatives such that given the second outcome the probability of observing $a_k > a_l$ is p. In other words, the noise model treats all the outcomes with a same way.

Theorem 3.3.1. *A voting rule is a pairwise scoring rule with a weekly neutral pairwise scoring function if and only if it is the MLE of a weakly neutral pairwise-independent model.*

Proof. Consider a weakly neutral pairwise noise model with probabilities $Pr[a_i > a_j|o]$, $Pr[a_i < a_j|o]$ and $Pr[a_i \sim a_j|o]$, given o the ground truth. Then, a weakly neutral pairwise scoring function can be constructed as follows: $s(a_i, a_j, o) = log(Pr[a_i > a_j|o]) - log(Pr[a_i \sim a_j|o])$ and the score for a vote v , $s(v, o) = \sum_{(a_i,a_j)\in v} log(Pr[a_i > a_j|o]) - \sum_{(a_i,a_j)\in v} log(Pr[a_i \sim a_j|o]) = \sum_{i<j} log(Pr[v_{ij}|o]) - \sum_{i<j} log(Pr[a_i \sim a_j|o]) = log(Pr[v|o]) - \sum_{i<j} logPr([a_i \sim a_j|o])$, where v_{ij} is the vote restricted only on the alternatives i and j.

From the definition of the weakly neutral pairwise model the following holds for any outcomes o and o': $\sum_{i<j} log(Pr[a_i \sim a_j|o]) = \sum_{i<j} log(Pr[a_i \sim a_j|o'])$ (1).

Then an MLE for this noise model and a voting profile $\pi = \{v_1, ..v_n\}$ would be given by the following: $\arg\max_o Pr[\pi|o] = \arg\max_o \sum_j log(Pr[v_j|o]) = \arg\max_o \sum_j s(v_j, o) = \arg\max_o s(\pi, o)$, where the third transition follows the definition of the pairwise score function and (1). Hence, the scoring rule that uses s is equivalent to the MLE of this noise model.

On the other way, consider a pairwise scoring rule r_s where s is a weakly neutral pairwise scoring function. For any $a_i, a_j \in A$ let $b_{i,j}$ a constant such that $2^{s(a_i,a_j,o)+b_{i,j}} + 2^{s(a_j,a_i,o)+b_{i,j}} + 2^{b_{i,j}} = 1$. The existence of $b_{i,j}$ is guaranteed due to the intermediate theorem, since for $b_{i,j} = -\infty$ the left-hand site is $0 < 1$ and for $b_{i,j} = \infty$ the left-hand site is $\infty > 1$.

Then, a pairwise noise model can be constructed with the following probabilities:
$Pr[a_i > a_j|o] = 2^{s(a_i,a_j,o)+b_{i,j}}$
$Pr[a_j > a_i|o] = 2^{s(a_j,a_i,o)+b_{i,j}}$
$Pr[a_i \sim a_j|o] = 2^{b_{i,j}}$

Since s is weakly neutral, the above model is also weakly neutral pairwise model. An MLE for this noise model is $\arg\max_o log(Pr[v|o]) = \arg\max_o \sum_{i<j} log(Pr[v_{i,j}|o])$ $= \arg\max_o(s(v, o) + \sum_{i<j} b_{i,j}) = \arg\max_o s(v, o)$, where the third transition follows the definition of the noise model and the fourth transition follows that $\sum_{i<j} b_{i,j}$ is constant for all outcomes. Hence, the MLE is equivalent to r_s. □

The above theorem shows the relation between MLEs and pairwise scoring rules when the voting profiles consist of partial orders. Another similar result [27], in voting settings with total orders as both input and output, shows that a voting rule is neutral ranking scoring rule if and only if it is an MLE for some noise model.

Ranking scoring rules similar to pairwise scoring rules given the input profile give a score to each possible outcome and select the outcome with the highest ranking.

Chapter 4

Independent samples

In the previous chapter different settings were presented where the aim was to find rules that were maximum likelihood estimators of specific noise models. However, a voting rule that is a maximum likelihood estimator of a noise model, may present a very bad behavior when it is given samples from other noise models. As different noise models are expected to arise in practice, it is argued [18] that MLE requirement is too restrictive. Instead of that, a setting [18] that examines how many votes different rules need in order to reconstruct with high probability the ground truth, will be presented in this chapter. Taking a more normative approach, it is also examined which rules return with probability close to 1 the ground truth when they are given infinite number of samples; a property that intuitively should be satisfied by "good" voting rules.

4.1 Model

Let $A = \{1, 2, .., m\}$ the set of the alternatives and σ^* the underlying ground truth with a_i the alternative that is on the i^{th} position in the true ranking. Given the noise model that the noisy estimators of the ground truth follow, the objective is to find the minimum number of samples required by a voting rule in order to output with high probability the ground truth. The samples are given in the form of full rankings and voting rules that return full rankings as well are used. Formally, the voting rules that will be used in this chapter are defined as follows.

Definition 24. *A* deterministic voting rule *is a function $r : \cup_{n\geq 1} L(A)^n \to L(A)$. A* randomized voting rule *is a function $r : \cup_{n\geq 1} L(A)^n \to D(L(A))$ where $D(L(A))$ is the set of probability distributions over $L(A)$.*

In other words, a randomized voting rule outputs each ranking by following a probability distribution and the notation $Pr[r(\pi) = \sigma]$ gives the probability of returning the ranking σ given the profile π.

Except from Mallows model, noise models parameterized by different distance functions will be used. A distance function d has the following properties for every $\sigma, \sigma', \tau \in L(A)$:

1. $d(\sigma, \sigma') \geq 0$

2. $d(\sigma, \sigma') = 0$ if and only if $\sigma = \sigma'$

3. $d(\sigma, \sigma') = d(\sigma', \sigma)$

4. $d(\sigma, \sigma') \leq d(\sigma, \tau) + d(\tau, \sigma')$

An important property of distance functions is defined below.

Definition 25. *A distance function d is* swap-increasing *if for every σ, σ' and alternatives a, b such that $a >_\sigma b$ and $a >'_\sigma b$, $d(\sigma_{a\leftrightarrow b}, \sigma') \geq d(\sigma, \sigma') + 1$ where the equation holds only when a, b are adjacent in σ'. The ranking $\sigma_{a\leftrightarrow b}$ is obtained by σ when all alternatives except from a, b remain in the same position and a, b are swapped.*

An example of swap-increasing distance functions is the Kendall Tau distance.

Example 4.1.1. *Let $A = \{a, b, c, d\}$ and consider the following rankings: $\sigma_1 = a > c > d > b$, $\sigma' = c > d > a > b$ and $\sigma'' = a > d > b > c$, where all three rankings prefer a to b. According to the Definition 25, $\sigma_{1,a\leftrightarrow b} = b > c > d > a$. Then, the Kendall Tau distances from σ' and σ'' are $d_{KT}(\sigma_1, \sigma') = 2$, $d_{KT}(\sigma_{1,a\leftrightarrow b}, \sigma') = 3$, $d_{KT}(\sigma_1, \sigma'') = 2$, $d_{KT}(\sigma_{1,a\leftrightarrow b}, \sigma'') = 5$. Thus, it holds that $d_{KT}(\sigma_{1,a\leftrightarrow b}, \sigma') = d_{KT}(\sigma_1, \sigma') + 1$ as a and b are adjacent in σ' while in σ'' where a and b are not adjacent, $d_{KT}(\sigma_{1,a\leftrightarrow b}, \sigma'') > d_{KT}(\sigma_1, \sigma') + 1$.*

4.2 Samples required in Mallows model

Firstly, Caragiannis et.al [18] study the minimum number of samples required by different rules in order to reconstruct the underlying ranking with high probability when the samples follow Mallows model.

The probability of a rule retuning a specific ranking when given k samples will be a useful measure to identify the minimum number of samples required for the ground truth. This probability can be computed as $Acc^r(k, \sigma) = \sum_{\pi \in L(A)^k} Pr[\pi|\sigma] \cdot Pr[r(\pi) = \sigma]$, that is the probability for a rule r to return the underlying ranking σ when the number of samples is k. Furthermore, using $Acc^r(k, \sigma)$ two other useful quantities can be defined: $Acc^r(k) = \min_{\sigma \in L(A)} Acc^r(k, \sigma)$ which computes the minimum probability of returning the true ranking, i.e. a rule r will return the ground truth with at least $Acc^r(k)$ probability given k samples no matter what the true ranking is and $N^r(\epsilon) = min\{k | Acc^r(k) \geq 1 - \epsilon\}$ which is the number of samples needed by rule r to return the true ranking with probability at least $1 - \epsilon$.

Naturally, one may expect that Kemeny's rule which is the maximum likelihood estimator for Mallows model, will have a “good” behavior regarding the number of samples that requires to return the true ranking. Indeed, it is proved that Kemeny's rule requires the minimum number of Mallows' samples to return the true ranking. However, the assumption that if a rule is a maximum likelihood estimator, then it needs the minimum number of samples, is not true.

Example 4.2.1. *Let $\{\sigma_1, \sigma_2, \sigma_3\}$ the set of all possible ground truths and $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ the set of all outcomes that can be observed. The following table gives the* *probability of observing the outcome at column j given that the ground truth is the ranking at row i:*

	π_1	π_2	π_3	π_4
σ_1	1/5	1/5	1/5	2/5
σ_2	1/6	1/6	1/6	**1/2**
σ_3	**1/4**	**1/4**	**1/4**	1/4

Table 4.1: The probabilities of observing each outcome given the respective ground truth in the example 4.2.1.

Given only one sample, a maximum likelihood estimator would output the rankings in bold, as they maximize the probability of observing the corresponding profiles. However, if the objective was to return the ground truth with probability at least 1/5 *then the maximum likelihood estimator fails as when the ground truth is σ_1 the probability of returning it is 0. Then, a different rule r that returns the same results with MLE given the outcomes π_1,π_2 and π_4 and σ_1 when π_3 is observed, it returns the ground truth with probability at least* 1/5 *as the probabilities of returning any of the possible rankings are:*

1. $Acc^r(1, \sigma_1) = 1/5$
2. $Acc^r(1, \sigma_2) = 1/2$
3. $Acc^r(1, \sigma_3) = 1/4 + 1/4 + 1/4 = 3/4$

Thus, it is implied that maximum likelihood estimators do not always need the minimum number of samples in order to return the underlying ranking with the desirable probability.

Theorem 4.2.1. *Kemeny's rule needs the least samples to return the true ranking with probability $1 - \epsilon$, i.e. $N^{KEM}(\epsilon) \leq N^r(\epsilon)$ for every voting rule r.*

Proof. In order to show the above theorem, the following two lemmas will be used.

Lemma 4.2.1. $Acc^{KEM}(k, \sigma) = Acc^{KEM}(k, \sigma'), \forall \sigma, \sigma' \in L(A), \forall k \in \mathbb{N}$.

Proof. For all $\sigma, \sigma' \in L(A)$ there are π, π' such that $Pr[\pi|\sigma] = Pr[\pi'|\sigma']$ since Kemeny's rule and Mallows model treat fairly the alternatives(neutrality) and thus, thinking that σ' is a permutation of σ, π' can be found from π by applying the same permutation. As this holds for all σ, σ' and assuming that Kemeny's rule break the ties uniformly, it follows that $\sum\limits_{\pi \in L(A)^k} Pr[\pi|\sigma'] \cdot Pr[KEM(\pi) = \sigma'] = \sum\limits_{\pi \in L(A)^k} Pr[\pi|\sigma] \cdot Pr[KEM(\pi) = \sigma]$. Therefore, $Acc^{KEM}(k, \sigma) = Acc^{KEM}(k, \sigma'), \forall \sigma, \sigma' \in L(A), \forall k \in \mathbb{N}$. □

Lemma 4.2.2. $TotAcc^{KEM}(k) \geq TotAcc^r(k), \forall r, \forall k \in \mathbb{N}$,
where $TotAcc^r(k) = \sum_{\sigma \in L(A)} Acc^r(k, \sigma)$.

Proof. For any rule r and any $k \in \mathbb{N}$,
$TotAcc^r(k) = \sum_{\sigma \in L(A)} Acc^r(k,\sigma) = \sum_{\sigma \in L(A)} \sum_{\pi \in L(A)^k} Pr[\pi|\sigma] \cdot Pr[r(\pi) = \sigma] =$
$\sum_{\pi \in L(A)^k} \sum_{\sigma \in L(A)} Pr[\pi|\sigma] \cdot Pr[r(\pi) = \sigma] \leq \sum_{\pi \in L(A)^k} \sum_{\sigma \in L(A)} Pr[r(\pi) = \sigma] \cdot max_{\sigma' \in L(A)} Pr[\pi|\sigma']$
$= \sum_{\pi \in L(A)^k} max_{\sigma' \in L(A)} Pr[\pi|\sigma'] = \sum_{\pi \in L(A)^k} max_{\sigma' \in L(A)} Pr[\pi|\sigma'] \cdot$
$\cdot \sum_{\sigma \in TIE-KEM(\pi)} \frac{1}{|TIE-KEM(\pi)|} = \sum_{\pi \in L(A)^k} \sum_{\sigma \in TIE-KEM(\pi)} Pr[\pi|\sigma] \cdot Pr[KEM(\pi) =$
$\sigma] = TotAcc^{KEM}(k)$,

where the third transition follows the definition of $Acc^r(k,\sigma)$ and the fourth transition is an exchange in the order of the two sums. Then, the fifth transition follows that $\sum_{\sigma \in L(A)} Pr[r(\pi) = \sigma] = 1$ and the sixth transition holds since Kemeny's rule is an MLE for Mallows model. Therefore, $max_{\sigma' \in L(A)} Pr[\pi|\sigma'] = Pr[\pi|\sigma]$ for every $\sigma \in$ TIE-KEM(π), where TIE-KEM are the results of Kemeny's rule and as the ties are broken uniformly, i.e. $Pr[KEM(\pi) = \sigma] = \frac{1}{|TIE-KEM(\pi)|}$ for every $\sigma \in TIE - KEM(\pi)$. □

Lemmas 4.2.1 and 4.2.2 imply that Kemeny's rule given k samples return the true ranking with the same probability no matter what the true ranking is, and at the same time, the total sum of the probabilities of returning the true ranking (summed over all rankings) is greater or equal than any other rule. Using the above lemmas, Theorem 4.2.1 follows easily. Assuming that $N^{KEM}(\epsilon) = k$, i.e. Kemeny's rule needs at least k samples to return any true ranking with probability at least $1 - \epsilon$, there exists $\sigma \in L(A)$ such that $Acc^{KEM}(k-1,\sigma) < 1-\epsilon$, otherwise by definition $N^{KEM}(\epsilon)$ would be $k-1$. Then, from Lemma 4.2.1 it follows that for every $\sigma' \in L(A)$, $Acc^{KEM}(k-1,\sigma') < 1-\epsilon$. Hence, as there are $m!$ rankings in $L(A)$, $TotAcc^{KEM}(k-1) < m! \cdot (1-\epsilon)$. Lemma 4.2.1 implies that for every rule r, $TotAcc^r(k-1) \leq TotAcc^{KEM}(k-1) < m! \cdot (1-\epsilon)$ and therefore, there should be a $\sigma' \in L(A)$ such that $Acc^r(k-1,\sigma') < 1-\epsilon$. Otherwise, the $TotAcc^r(k-1)$ would be greater than $m! \cdot (1-\epsilon)$. Since there exists σ' such that $Acc^r(k-1,\sigma') < 1-\epsilon$, by the definition of $N^r(\epsilon)$, $N^r(\epsilon) \geq k$ and therefore $N^r(\epsilon) \geq N^{KEM}(\epsilon)$. □

4.3 PM-c rules

As Kemeny's rule has been shown to need the minimum number of samples to return the truth with high probability, it is important to find out how many samples it actually requires. However, instead of examining Kemeny's rule individually, a family of rules, PM-c class, that contains Kemeny's rule is studied since all rules in this family require the same number of samples. The following definitions are required to define the PM-c rules.

Definition 26. *A* pairwise majority (PM) graph *of a vote profile is a graph that has as vertices all the alternatives and there is an edge from alternative a to alternative*

b if the number of voters in the vote profile who prefer a to b is greater than the number of voters who prefer b to a.

When the PM graph is complete and acyclic there is a unique ranking σ such that there is an edge (a, b) if and only if $a >_\sigma b$ (this ranking exists because the PM graph is complete and therefore it provides the result of each pairwise comparison and as it is acyclic there will not be any inconsistencies in the ranking σ). That is, for every pairwise comparison the majority of voters agree with the comparison on σ and it is said that the PM graph reduces to ranking σ.

Example 4.3.1. *Let $A = \{a, b, c\}$ the set of alternatives and consider the following vote profile:*

1. *3 votes: $a > b > c$*
2. *2 votes: $a > c > b$*
3. *2 votes: $c > b > a$*

As the majority of voters prefer a to b(5 votes), a to c(5 votes) and c to b(4 votes), the PM-graph of this voting profile is the following:

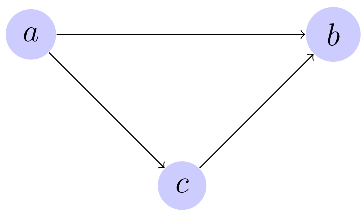

Figure 4.1: The PM-graph of the example 4.3.1.

As we can observe the PM-graph is complete and acyclic and reduces to the ranking $a > c > b$.

Definition 27. *A voting rule is* pairwise majority consistent (PM-c) *if it outputs the ranking σ whenever the PM graph of the profile reduces to σ.*

In other words, a voting rule is PM-c, if whenever a ranking σ exists such that for every pairwise comparison, there is a majority of voters that agree with σ, then the rule outputs σ.

Theorem 4.3.1. *The Kemeny's rule and the ranked pairs method are PM-c.*

Proof. As the Kemeny's rule returns the ranking that minimizes the total pairwise disagreements with the voters, if the PM-graph reduces to a ranking, it has to agree with that ranking. Otherwise, let that (a, b) is an edge of the PM-graph (which reduces to a ranking σ) and that the Kemeny's rule return a ranking σ' with $b >'_\sigma a$. As (a, b) is in PM-graph the number of voters that prefer a to b is greater than $\frac{n}{2}$ and the number of voters that prefer b to a is less than $\frac{n}{2}$ (n is the total number of voters). Therefore, if $b >'_\sigma a$ is replaced by $a >'_\sigma b$ a ranking with a smaller number of pairwise disagreements is obtained, which is opposed to the definition of Kemeny's rule.

As the ranked pairs method sorts the pairwise elections by the largest strength of win to the smallest, if the PM reduces to a ranking, the ranked pairs method will output the same ranking. This is because all edges in the PM graph are voted by more than the half voters and therefore, ranked pairs method will choose all edges in the PM graph before reach the opposite pairs(which are voted by less than the half voters). □

Theorem 4.3.2. *For any $\epsilon > 0$, any PM-c rule needs $O(\log(\frac{m}{\epsilon}))$ samples from Mallows model to return the true ranking with probability at least $1-\epsilon$.*

Proof. To show that a PM-c rule needs $O(\log(\frac{m}{\epsilon}))$ to reconstruct the truth with probability $1-\epsilon$, it is sufficient to show that given $O(\log(\frac{m}{\epsilon}))$ samples the PM-graph reduces to the true ranking σ^* with probability $1-\epsilon$.

The PM-graph of a vote profile will reduce to σ^* if for every pair of alternatives a, b such that $a >_{\sigma*} b$ the number of voters that prefer a to b is greater than the number of voters that prefer b to a. Therefore, if n is the total number of voters, n_{ab} the number of voters that prefer a to b and n_{ba} the number of voters that prefer b to a, it is wanted to have $n_{ab} > n_{ba}$. Since the objective is to return the true ranking with probability at least $1-\epsilon$, the following must hold:
$Pr[\forall a, b \in A, a >_{\sigma^*} b \implies n_{ab} - n_{ba} \geq 1] \geq 1-\epsilon$.

Let a, b a pair of alternatives such that $a >_{\sigma^*} b$ and $\delta_{ab} = p_{a>b} - p_{b<a}$. Then $\delta_{ab} = \mathbb{E}[\frac{n_{ab}-n_{ba}}{n}]$ since by the definition and linearity property of expectation $\mathbb{E}[\frac{n_{ab}-n_{ba}}{n}] = \frac{1}{n} \cdot (n \cdot p_{a>b} - n \cdot p_{b<a}) = p_{a>b} - p_{b<a}$.

The probability of the number of the voters who prefer a to b to be greater than the number of the voters who prefer b to a is :
$Pr[n_{ab} - n_{ba} \leq 0] = Pr[\dfrac{n_{ab}-n_{ba}}{n} \leq 0] \leq Pr[|\dfrac{n_{ab}-n_{ba}}{n} - \mathbb{E}[\dfrac{n_{ab}-n_{ba}}{n}]| \geq \delta_{ab}] \leq 2 \cdot e^{-2\cdot\delta_{ab}^2\cdot n} \leq 2\cdot e^{-2\cdot\delta_{min}^2\cdot n}$, where the third transition holds from Hoeffding's inequality and $\delta_{min} = \min_{a,b\in A: a>_{\sigma*}b} \delta_{ab}$.

Hence, $Pr[\exists a, b \in A, \{(a >_{\sigma^*} b) \wedge (n_{ab} - n_{ba} \leq 0)\}] \leq \dbinom{m}{2} \cdot 2 \cdot e^{-2\cdot\delta_{min}^2\cdot n} \leq m^2 \cdot e^{-2\cdot\delta_{min}^2\cdot n}$, where the second transition holds from the Union Bound. Therefore, in order to return the correct true ranking with probability at least $1-\epsilon$, the probability of returning wrong ranking has to be smaller than ϵ, i.e. $m^2 \cdot e^{-2\cdot\delta_{min}^2\cdot n} \leq \epsilon$ and equivalently $n \geq \dfrac{1}{2\cdot\delta_{min}^2} \cdot \log(\dfrac{m^2}{\epsilon})$.

The number of samples required will be fully defined when δ_{ab} is computed.
$\delta_{ab} = p_{a>b} - p_{b>a} = \displaystyle\sum_{\sigma\in L(A)|a>_\sigma b} Pr[\sigma|\sigma^*] - \sum_{\sigma\in L(A)|b>_\sigma a} Pr[\sigma|\sigma^*] =$

$$\sum_{\sigma\in L(A)|a>_\sigma b} (Pr[\sigma|\sigma^*] - Pr[\sigma_{a\leftrightarrow b}|\sigma^*]) = \sum_{\sigma\in L(A)|a>_\sigma b} \frac{\phi^{d_{KT}(\sigma,\sigma*)} - \phi^{d_{KT}(\sigma_{a\leftrightarrow b},\sigma^*)}}{Z_\phi^m}$$

$\geq \displaystyle\sum_{\sigma\in L(A)|a>_\sigma b} \frac{\phi^{d_{KT}(\sigma,\sigma*)}\cdot(1-\phi)}{Z_\phi^m} = (1-\phi)\cdot p_{a>b} = (1-\phi)\cdot\frac{1+\delta_{ab}}{2}$, where the third transition follows since $\sigma_{a\leftrightarrow b}$ is a bijection and the fifth transition holds since

Kendall Tau distance is swap increasing. The last transition follows by the equalities $\delta_{ab} = p_{a>b} - p_{b>a}$ and $p_{a>b} + p_{b>a} = 1$. Solving the last inequality, $\delta_{ab} \geq \frac{1-\phi}{1+\phi}$ and hence, $\delta_{min} \geq \frac{1-\phi}{1+\phi} = \Omega(1)$ and therefore, the theorem holds. □

Theorem 4.3.3. *For any $\epsilon \in (0, 1/2]$, any voting rule requires $\Omega(\log(\frac{m}{\epsilon}))$ samples from Mallows model to reconstruct the true ranking with probability at least $1 - \epsilon$.*

Proof. Let r a voting rule and assume that $N^r(\epsilon) = n$. Then it is required to show that $n = \Omega(\log(\frac{m}{\epsilon}))$. By the definition of $N^r(\epsilon)$ it follows that $Acc^r(n, \sigma) \geq 1 - \epsilon$ for any $\sigma \in L(A)$ (1). Choosing a ranking σ, let $N(\sigma)$ the set of all rankings in $L(A)$ that have distance 1 from the ranking σ. Therefore for any ranking σ' in $N(\sigma)$ and a vote profile(n votes) the following holds:

$$Pr[\pi|\sigma] = \prod_{i=1}^{n} \frac{\phi^{d_{KT}(\sigma_i,\sigma)}}{Z_\phi^m} \geq \prod_{i=1}^{n} \frac{\phi^{d_{KT}(\sigma_i,\sigma')+1}}{Z_\phi^m} = \phi^n \cdot Pr[\pi|\sigma'] \qquad (2)$$

where the second transition follows the third property of distance functions that is $d(\sigma_i, \sigma) \leq d(\sigma_i, \sigma') + d(\sigma', \sigma) = d(\sigma, \sigma') + 1$.

Then,

$$Acc^r(n, \sigma) = \sum_{\pi \in L(A)^n} Pr[\pi|\sigma] \cdot Pr[r(\pi) = \sigma] = \sum_{\pi \in L(A)^n} Pr[\pi|\sigma] \cdot (1 - Pr[r(\pi) \neq \sigma]) =$$
$$1 - \sum_{\pi \in L(A)^n} Pr[\pi|\sigma] \cdot Pr[r(\pi) \neq \sigma] \leq 1 - \sum_{\pi \in L(A)^n} Pr[\pi|\sigma] \cdot (\sum_{\sigma' \in N(\sigma)} Pr[r(\pi) = \sigma']) \leq$$
$$1 - \sum_{\sigma' \in N(\sigma)} \sum_{\pi \in L(A)^n} \phi^n \cdot Pr[\pi|\sigma'] \cdot Pr[r(\pi) = \sigma'] = 1 - \phi^n \cdot \sum_{\sigma' \in N(\sigma)} Acc^r(n, \sigma') \leq 1 - \phi^n \cdot$$

$(m-1) \cdot (1-\epsilon)$, where the third transition follows the fact that $\sum_{\pi \in L(A)^n} Pr[\pi|\sigma] = 1$ and the fourth transition holds since the events $\{r(\pi) = \sigma'\}$ for any $\sigma' \in N(\sigma)$ are a proper subset of $\{r(\pi) \neq \sigma\}$ and thus, $Pr[r(\pi) \neq \sigma] \geq \sum_{\sigma' \in N(\sigma)} Pr[r(\pi) = \sigma']$. Moreover, the fifth transition follows inequality (2) and the last transition follows inequality (1). Therefore, in order to have $Acc^r(n, \sigma) \geq 1 - \epsilon$ it is required that $\phi^n \cdot (m-1) \cdot (1-\epsilon) \leq \epsilon$ and solving for n, it follows that $\Omega(\log(\frac{m}{\epsilon}))$. □

Theorems 4.3.2 and 4.3.3 indicate the logarithmic number of samples that Kemeny's rule needs to reconstruct the truth and establish its best behavior regarding the required number of samples; since any other rule requires at least logarithmic number of samples, no rule can do better than Kemeny's.

4.4 Scoring rules

While any other rule than Kemeny's needs at least logarithmic number of samples to return the true ranking with high probability, some rules need a significantly larger number than logarithmic. For example, plurality rule will be proved to need at least exponential number of samples.

Since plurality rule considers only the number of times that each alternative appears on the first position, a useful measure in proving the number of samples that plurality rule requires is the probability of an alternative appearing first in a vote.

Lemma 4.4.1. *The probability of the alternative on the position i of the true ranking appearing on the first position, $p_{i,1}$, is $p_{i,1} = \phi^{i-1}/\sum_{j=1}^{m}\phi^{j-1}$ for each alternative i in A.*

Intuitively, the above lemma gives an indication that plurality rule may needs exponential number of samples as the probability of observing the last or the semi-last alternative of the true ranking in the first position is exponentially small, and therefore there is the need of exponential samples to achieve a distinction between them.

Theorem 4.4.1. *For any $\epsilon \in (0, \frac{1}{4}]$, plurality requires $\Omega((\frac{1}{\phi})^m)$ samples from Mallows model to reconstruct the true ranking with probability at least $1-\epsilon$.*

Proof. Assuming that $Acc^{PL}(n) \geq 1-\epsilon$ it is needed to show that $n = \Omega((\frac{1}{\phi})^m)$. As plurality rule considers only the number of times that each alternative appears in the first place, it can be supposed that the rule operates on the vector $v \in A^n$ of the top alternatives of each vote in a profile, instead of operating on the full profile. Then the accuracy of the plurality rule can be written as $Acc^{PL}(n,\sigma) = \sum_{v\in A^n} Pr[v|\sigma]\cdot Pr[PL(v)=\sigma]$, where $Pr[v|\sigma]$ is the sum of the probabilities of observing profiles which have top vote v given the true ranking σ.

Let two rankings $\sigma_1 = (a_1 > a_2 > ... > a_{m-1} > a_m)$ and $\sigma_2 = (a_1 > a_2 > ... > a_m > a_{m-1})$. Then, the $Acc^{PL}(n,\sigma)$ can be split into two parts, considering in the first(denoted as A'^n) the top votes in which the alternatives a_{m-1} and a_m do not appear and in the second the top votes in which at least one of a_{m-1} and a_m appears.

To compute the $Acc^{PL}(n,\sigma_1) + Acc^{PL}(n,\sigma_2)$, calculation will be split for the top votes in A'^n and the top votes in $A^n \backslash A'^n$.

For the top votes in A'^n we have:
$\sum\limits_{v\in A'^n}((Pr[v|\sigma_1]\cdot Pr[PL(v)=\sigma_1]) + (Pr[v|\sigma_2]\cdot Pr[PL(v)=\sigma_2])) = \sum\limits_{v\in A'^n} Pr[v|\sigma_1]\cdot (Pr[PL(v)=\sigma_1] + Pr[PL(v)=\sigma_2]) \leq \sum\limits_{v\in A'^n} Pr[v|\sigma_1] \leq 1$ (1), where the first transition holds since for any $v \in A'^n, Pr[v|\sigma_1] = Pr[v|\sigma_2]$. This holds since σ_2 can be obtained from σ_1 by swapping a_{m-1} and a_m and for any profile π that has top vote v a profile π' can be found with $Pr[\pi|\sigma_1] = Pr[\pi'|\sigma_2]$ by swapping a_{m-1} and a_m. The profile π' also has top vote v as alternatives a_{m-1} and a_m do not appear in the top vote v.

Let $t_{i,j}$ the number of votes in which the alternative a_i appears in the position j. Then, for the top votes in $A^n\backslash A'^n$ we get: $\sum\limits_{v\in A^n\backslash A'^n} Pr[v|\sigma_1]\cdot Pr[PL(v)=\sigma_1] \leq \sum\limits_{v\in A^n\backslash A'^n} Pr[v|\sigma_1] = Pr[(t_{m-1,1}>0)\vee(t_{m,1}>0)] \leq Pr[t_{m-1,1}\geq 0] + Pr[t_{m,1}\geq 0] \leq n\cdot(p_{m-1,1}+p_{m,1})$ (2), where the second transition holds since the probability of observing a top vote in $A^n\backslash A'^n$, i.e. a top vote which contains a_m or a_{m-1}, equals the probability of a_m or a_{m-1} appearing in the first position that is $(t_{m-1,1}\geq 0)\vee(t_{m,1} > 0)$. The third and fourth transition hold due to the union bound and specifically, as $p_{i,1}$ is the probability of the alternative i to appear in the first position then the probability that a_i appears in the first position of at least one vote (which equals the probability of $t_{i,1}$being positive) is at most $n \cdot p_{i,1}$.

In a similar way, it can be proved that $\sum_{v \in A^n \setminus A'^n} Pr[v|\sigma_2] \cdot Pr[PL(v) = \sigma_2] \leq n \cdot (p_{m-1,1} + p_{m,1}) \quad (3)$.

Therefore by adding equations (1),(2) and (3) it follows that:
$Acc^{PL}(n, \sigma_1) + Acc^{PL}(n, \sigma_2) \leq 1 + 2 \cdot n \cdot (p_{m-1,1} + p_{m,1})$.

Considering that $Acc^{PL}(n) \geq 1 - \epsilon$, it follows that
$Acc^{PL}(n, \sigma_1) + Acc^{PL}(n, \sigma_2) \geq 2 \cdot (1 - \epsilon)$ and hence it is needed that
$1 + 2 \cdot n \cdot (p_{m-1,1} + p_{m,1}) \geq 2 \cdot (1 - \epsilon)$.

Then solving by n, the lower bound follows:
$n \geq \dfrac{1 - 2 \cdot \epsilon}{2 \cdot (p_{m-1,1} + p_{m,1})} \geq \dfrac{1}{8 \cdot p_{m-1,1}} = \dfrac{\sum_{j=0}^{m-1} \phi^j}{8 \cdot \phi^{m-2}} \geq \dfrac{1}{8 \cdot \phi^{m-2}}$ and hence,
$n = \Omega(\log(\frac{m}{\epsilon}))$. □

Although plurality is a positional scoring rule, it does not follow that all positional scoring rules need at least exponential number of samples. The following theorem shows that some positional scoring rules need polynomial number of samples.

Theorem 4.4.2. *Consider a positional scoring rule r given by score vector $(a_1, .., a_m)$ and $\beta_i = a_i - a_{i+1}$ for any $i \in \{1, ..., m-1\}$. Let $\beta_{max} = max_{i<m}\beta_i$, $\beta_{min} = min_{i<m}\beta_i$ and $\beta^* = \beta_{max}/\beta_{min}$ with $\beta_{min} > 0$. Then for any $\epsilon > 0$, rule r needs $O((\beta^*)^2 \cdot m^2 \cdot \log(m/\epsilon))$ samples to return the true ranking with probability at least $1 - \epsilon$.*

Thus, Theorems 4.4.1 and 4.4.2 show that rules other than PM-c rules need considerable larger numbers of samples to output the underlying ranking. In particular, plurality needs exponential number of samples and positional scoring rules with positive difference between any pair of consecutive scores (in the score vector) need polynomial samples. Plurality needs at least exponential and not polynomial number of samples as the minimum difference between consecutive scores is zero, i.e. $\beta_{min} = 0$ and therefore, it does not satisfy the conditions of Theorem 4.4.2.

4.5 Generalizations

4.5.1 Infinite samples

While the number of samples that a voting rule requires in order to return the true ranking with high probability gives an indication of the rule's behavior, in this subsection another natural characteristic is examined. Voting rules should be able to return surely the true ranking if they are given infinite number of samples. A voting rule with this property is said to be accurate in the limit, that is a voting rule reproduces with probability 1 the ground truth if it is given infinite many samples.

Since PM-c rules need logarithmic number of samples, they are obviously accurate in the limit. Plurality is also another rule that can be proved to return with probability 1 the true ranking given if it is given infinite samples. In fact, it can be shown that all positional scoring rules satisfy accuracy in the limit.

Instead of examining positional scoring rules individually, a new family of voting rules, PD-c class, that includes scoring rules is studied.

Position dominance graph of a voting profile should be defined in order to define the PD-c rules.

Definition 28. *Given a profile $\pi = (\sigma_1, \sigma_2, ..., \sigma_n) \in L(A)^n$, alternative a and $j \in \{1, .., m-1\}$, $s_j(a)$ is the number of votes in which the alternative a is between the first j positions. For $a, b \in A$ it is said that alternative a dominates alternative b whenever $s_j(a) \geq s_j(b)$ $\forall j, 1 \leq j \leq m-1$. The position dominance graph(PD graph) of the profile π has as vertices all the alternatives and there is an edge (a, b) (directed) if a position dominates b.*

As position dominance is a transitive property PD graph is always acyclic. Whenever PD graph is complete, in a similar way with PM graph, it reduces to a ranking.

Example 4.5.1. *Let $A = \{a, b, c\}$ the set of alternatives and the following vote profile:*

1. *3 votes: $a > b > c$*
2. *3 votes: $a > c > b$*
3. *2 votes: $c > b > a$*

Then the scores of each alternative are $s_1(a) = 6$, $s_2(a) = 6$, $s_1(b) = 0$, $s_2(b) = 5$, $s_1(c) = 2$, $s_2(c) = 5$ and therefore a dominates b and c, and c dominates b. The PD-graph of this voting profile is the following:

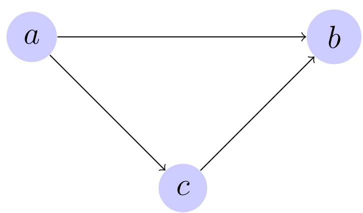

Figure 4.2: The PD-graph of the example 4.5.1.

As we can observe the PD-graph is complete and it reduces to the ranking $a > c > b$.

Definition 29. *A* Position-dominance rule (PD-c) *is a rule that outputs ranking σ, whenever the PD graph reduces to a ranking σ.*

By the definition of PD-c rules, it is demonstrated that the class of PD-c rules includes rules that give higher preference to alternatives that appear at the first positions. Intuitively, one may expect that positional scoring rules are PD-c since the total score of each alternative depends only on the places he appears and the first places are associated with higher scores. Indeed, it can be proved that all positional scoring rules are PD-c.

Theorem 4.5.1. *All positional scoring rules are PD-c.*

Theorem 4.5.2. *PD-c rules are accurate in the limit.*

Proof. In order to show that PD-c rules are accurate in the limit, it is sufficient to show that PD graph reduces to the true ranking when given infinite samples. Given a profile with n samples then $\mathbb{E}[s_j(a_i)] = n \cdot q_{i,j}$ where $q_{i,j}$ is the probability of alternative a_i to appear between the first j positions. As by the definition of Mallows model $q_{i,j} > q_{l,j}$ for any $i < l$ (since it is more probable to observe a ranking that is closest to the ground truth and hence, it is more probable alternative a_i to be ranked higher than a_l as it is ranked higher in the ground truth), given infinite samples $Pr[s_j(a_i) > s_j(a_l)] = 1$ for any j,$1 \leq j \leq m-1$ and $i < l$. Hence, the PD graph will reduce to the true ranking given infinitely many samples. □

Thus, the classes of PM-c and PD-c rules satisfy the natural requirement of returning the ground truth given infinite many samples. Common voting rules such as Kemeny's and ranked pairs which are PM-c and positional scoring rules which are PD-c are all accurate in the limit.

4.5.2 PM-c and PD-c rules

Although the definitions of PM-c and PD-c rules at first glance do not imply any relation between the two classes, in reality the two classes are disjoint. That is, there are not any rules that are both PM-c and PD-c.

Example 4.5.2. *Let $A = \{a, b, c\}$ and a profile π consisting of the following votes:*

(a) 4 votes : $a > b > c$

(b) 2 votes: $b > a > c$

(c) 3 votes: $b > c > a$

(d) 2 votes: $c > a > b$

The PM graph of the above profile reduces to $a > b > c$ and the PD graph reduces to $b > a > c$. Therefore, PM-c rules and PD-c rules output a different ranking for the specific profile. Hence, if there were rules that were both PM-c and PD-c, they would have to output two different rankings for the above profile, which is impossible.

From the above example, the theorem below follows.

Theorem 4.5.3. *There is no rule that is both PM-c and PD-c.*

If one considers that PD-c rules are a generalization of positional scoring rules and PM graphs that reduce to a ranking have a Condorcet winner this impossibility result is not surprising; a previous result [39] mentioned in the introduction states that no positional scoring rule can be Condorcet extension.

4.5.3 Noise models

All the above results consider only samples taken from Mallows model. However, it is unrealistic to expect that all noisy samples would fit Mallows model. Therefore, there is the need to anticipate any reasonable noise models that would arise in practice. Trying to study only realistic noise models, d-Monotonic noise models will be studied as these models satisfy an expected requirement; it is more probable to observe a sample that is closer to the ground truth than a sample with a larger distance.

Definition 30. *Let σ^* the true ranking and d a distance function. A noise model is* d-Monotonic *if for any σ, σ' with $d(\sigma, \sigma^*) < d(\sigma', \sigma^*)$ then, $Pr[\sigma|\sigma^*] > Pr[\sigma'|\sigma^*]$ and for any σ, σ' with $d(\sigma, \sigma^*) = d(\sigma', \sigma^*)$ then, $Pr[\sigma|\sigma^*] = Pr[\sigma'|\sigma^*]$.*

In other words, a noise model is d-monotonic if it is more likely to observe a ranking that is closer to the true ranking according to distance function d. For example, Mallows model is KT-monotonic.

Definition 31. *A voting rule is* d-monotone-robust *if it is accurate in the limit when it is given infinite samples from any d-monotonic noise model.*

In other words, d-monotone-robustness express the requirement that a rule is accurate in the limit in any "realistic" noise model parametrized by distance function d.

4.6 More generalizations

To move to the next step of generalizations, there is the need to use different distance functions from the KT distance. Therefore, it is studied whether PM-c and PD-c rules are monotone-robust to any distance functions other than the KT distance.

4.6.1 Distances and PM-c rules

It is proved that PM-c rules are monotone robust for the Majority-Concentric distances which are defined as follows.

Definition 32. *Considering a distance function d, a ranking $\sigma \in L(A)$ and an integer $k \geq 0$, $N^k_{a>b}(\sigma)$ is the set of the rankings that have distance at most k (according to distance function d) from the ranking σ and have $a > b$. A distance function d is* Majority-Concentric (MC) *if for any $\sigma \in L(A)$ and alternatives a, b with $a >_\sigma b$, $|N^k_{a>b}(\sigma)| \geq |N^k_{b>a}(\sigma)|$ for every $k \geq 0$.*

In other words, a distance function is MC if for each distance k, the majority of rankings that have at most distance k from a specific ranking σ agree with σ on every pair of alternatives.

Lemma 4.6.1. *A distance function is* MC *if and only if for every $\sigma \in L(A)$ and every alternatives a, b with $a >_\sigma b$ there exists a bijection $L_{a>b}(A) \to L_{b>a}(A)$ which is weakly-distance increasing to σ.*

Proof. Assume that d is a MC distance, σ the initial ranking and a pair of alternatives a, b with $a >_\sigma b$. After sorting the rankings with $a > b$ and the rankings with $b > a$ in increasing distance from σ, a bijection $f : L_{a>b} \to L_{b>a}$ is constructed that maps the i^{th} ranking of the first set to the i^{th} ranking of the second set. Assume that f is not weakly-distance increasing. Then, there is a mapping from a ranking with $a > b$ and distance k to a ranking with $b > a$ and distance k' and $k' < k$. Then, the $N^{k'}_{b>a}(\sigma) \geq N^k_{a>b}(\sigma)$ which is opposed to the definition of MC distance.

On the other hand, assume a distance function d, a ranking σ, two alternatives a, b with $a >_\sigma b$, and a weakly-distance increasing bijection $f : L_{a>b} \to L_{b>a}$. Then for any $k \geq 0$, $N^k_{b>a} \subseteq \{f(\tau) | \tau \in N^k_{a>b}\}$. Therefore, $|N^k_{a>b}(\sigma)| \geq |N^k_{b>a}(\sigma)|$. If this holds for each ranking and pair of alternatives, then d is MC. □

Theorem 4.6.1. *All PM-c rules are d-monotone robust if and only if d is MC.*

Proof. Assume that d is MC and consider any d-monotonic noise model G. Assume that σ^* is the true ranking and a, b two alternatives with $a >_{\sigma^*} b$. As d is MC, from the Lemma 2.4 there exists a weakly-distance increasing bijection f from $L_{a>b}$ to $L_{b>a}$. Then, for any $\sigma \in L_{a>b}$ it holds that $d(\sigma, \sigma^*) \leq d(f(\sigma), \sigma^*))$ and therefore by the definition of the d-monotonic noise model, $Pr[\sigma|\sigma^*] \geq Pr[f(\sigma)|\sigma^*]$. However, specifically for σ^*, $d(\sigma^*, \sigma^*) < d(f(\sigma^*), \sigma^*)$ and hence $Pr[\sigma^*|\sigma^*] > Pr[f(\sigma^*)|\sigma^*]$ as only for σ^* it holds that $d(\sigma^*, \sigma^*) = 0$.

Therefore, $Pr[a > b|\sigma^*] = \sum\limits_{\sigma \in L_{a>b}(A)} Pr[\sigma|\sigma^*] > \sum\limits_{\sigma \in L_{a>b}(A)} Pr[f(\sigma)|\sigma^*] =$
$\sum\limits_{\sigma \in L_{b>a}(A)} Pr[\sigma|\sigma^*] = Pr[b > a|\sigma^*]$, where the third transition holds because f is a bijection.

As a result, given infinite samples there will exist the edge from a to b and the PM graph will reduce to the true ranking (since the above holds for any pair a, b). Hence, all PM-c rules will output the true ranking.

On the other hand, consider a distance function d that is not MC. Then, it is needed to show that there exists a PM-c rule that does not surely output the true ranking given infinite samples from some d-monotonic noise model G. As d is not MC, there exists a $\sigma^* \in L(A)$, an integer k and alternatives a, b with $a >_{\sigma^*} b$ such that $|N^k_{a>b}(\sigma^*)| < |N^k_{b>a}(\sigma^*)|$. Let $M = \max_{\sigma \in L(A)} d(\sigma, \sigma*)$ and $T > M$. Then, the noise model G have as probabilities of observing a specific ranking σ, $Pr[\sigma|\sigma*] = w_\sigma / \sum_{\tau \in L(A)} w_\tau$, where if $d(\sigma, \sigma^*) \leq k$ then $w_\sigma = T - d(\sigma, \sigma^*)$ else $w_\sigma = M - d(\sigma, \sigma^*)$. The distribution of G is indeed a probability distribution as $\sum_{\sigma \in L(A)} Pr[\sigma|\sigma^*] = \sum_{\sigma \in L(A)} (w_\sigma / \sum_{\tau \in L(A)} w_\tau) = 1$ and for each $\sigma \in L(A)$, $0 \leq Pr[\sigma|\sigma^*] \leq 1$.

For the above noise model, constant T can be defined so that $Pr[a > b|\sigma^*] < Pr[b > a|\sigma^*]$ which is equivalent to $\sum_{\sigma \in L(A) | a >_\sigma b} w_\sigma < \sum_{\sigma \in L(A) | b >_\sigma a} w_\sigma$.

If $l = |N^k_{a>b}(\sigma^*)|$ then,
$\sum\limits_{\sigma \in L(A) | a >_\sigma b} w_\sigma \leq \sum\limits_{\sigma \in N^k_{a>b}(\sigma^*)} T + \sum\limits_{\sigma \in L(A)^n \setminus N^k_{a>b}(\sigma^*)} M \leq l \cdot T + m! \cdot M$, where for the l rankings that are in $N^k_{a>b}(\sigma^*)$ the maximum possible number of w_σ is T and for the rest rankings the maximum possible number of w_σ is M.

On the other hand, $\sum_{\sigma \in L(A) | b >_\sigma a} w_\sigma \geq \sum_{\sigma \in N^k_{a>b}(\sigma^*)} (T-k) + \sum_{\sigma \in L(A)^n \setminus N^k_{b>a}(\sigma^*)} 0 \geq$ $\geq (l+1) \cdot (T-k)$, where for the at least $l+1$ rankings that are in $N^k_{b>a}(\sigma^*)$ the minimum possible number of w_σ is $T-k$ since their maximum distance from σ^* is k and for the rest rankings the minimum possible number of w_σ is $M - M = 0$ as M is the maximum possible distance of any ranking from σ^*.

Therefore, T should be chosen so as that $l \cdot T + m! \cdot M < (l+1) \cdot (T-k)$ which is $T > (l+1) \cdot k + m! \cdot M$. By choosing a suitable value of T, $Pr[a > b | \sigma^*] < Pr[b > a | \sigma^*]$ and as a result given infinite samples from G, the edge from a to b will not exist and the PM graph will not reduce to σ^*. A PM-c rule that outputs a ranking with $b > a$ when the PM-graph does not reduce to a ranking, will surely not output the true ranking given infinite samples. □

4.6.2 Distances and PD-c rules

It is proved that PD-c rules are monotone robust for the Position-Concentric distances which are defined as follows.

Definition 33. *Considering a distance function d, a ranking $\sigma \in L(A)$, an integer $k \geq 0$ and an integer $0 \leq j \leq m-1$, $S^k_j(\sigma, a)$ is the set of the rankings that have distance at most k (according to distance function d) from the ranking σ and have the alternative a between the first j positions. A distance function d is* Position-Concentric (PC) *if for any $\sigma \in L(A)$, $j \in \{0, ..m-1\}$ and alternatives a, b with $a >_\sigma b$, $|S^k_j(\sigma, a)| \geq |S^k_j(\sigma, b)|$ for every $k \geq 0$ and strict inequality holds for a $k \geq 0$.*

In other words, a distance function is PC if for each distance k and position j, the majority of rankings that have at most distance k from a specific ranking σ rank at the first j votes the alternative a rather than b, where a, b is a pair of alternatives with $a >_\sigma b$.

Lemma 4.6.2. *A distance function is* PC *if and only if for every $\sigma \in L(A)$ and every alternatives a, b with $a >_\sigma b$ there exists a bijection $S_j(a) \rightarrow S_j(b)$ which is distance-increasing (weakly distance increasing and for some ranking the inequality holds) to σ .*

Theorem 4.6.2. *All PD-c rules are d-monotone robust if and only if d is PC.*

A result of the Theorems 4.6.1 and 4.6.2 is that if a distance function d is both PC and MC then all PM-c and PD-c rules are d-monotone robust. On the other hand, if a distance function is not MC(PC), there is a PM-c rule (PD-c rule) that is not d-monote robust.

Corollary 4.6.1. *All PM-c and PD-c rules are d-monotone robust if and only if d is both MC and PC.*

Lemma 4.6.3. *A swap-increasing distance function is both MC and PC.*

Since KT-distance is swap increasing, the corollary below follows.

Corollary 4.6.2. *All PM-c and PD-c rules are KT-monotone robust.*

That is, PM-c and PD- c rules are not only accurate in the limit in Mallows model which is a KT-monotonic model, but in fact they are accurate in the limit for any KT-monotonic model.

In this section, it has been shown that different rules require different number of samples in order to return with high probability the ground truth. Two important classes of rules, PM-c rules which are in some way a generalization of Condorcet extension rules and PD-c rules which are a generalization of positional scoring rules, both reconstruct surely the true ranking when they are given infinite number of samples from Mallows noise model. In an attempt to generalize the results for Mallows model in models that may arise in practice, it is shown that PM-c and PD-c rules are also accurate in the limit when the infinite samples are from any d-monotonic noise model with d MC or PC, respectively.

However, as noise can take unpredictable forms [51], it would be ideal to have a voting rule that is monotone-robust against any monotonic noise model, i.e. it is desirable to have a voting rule that will return almost surely the ground truth given infinite samples from any "realistic" noise model. The attempt to find a robust voting rule will be presented in the next subsection.

4.7 Modal ranking

Caragiannis et. al [19] study which voting rules return with probability close to 1 the ground truth when they are given an extremely large number of samples from any "realistic" noise model; an approach that is natural in crowd-sourcing systems [49] where the aim is to aggregate the preferences of a massive set of agents.

4.7.1 Model

Let $A = \{1, .., m\}$ the set of alternatives and σ^* the underlying true ranking. Given an input profile π, the voting rules that will be examined in this section will be randomized SWF, that are formally functions $f : L(A) \to D(L(A))$, where D(L(A)) is the set of probability distributions over $L(A)$. The objective is to find a voting rule that will be robust against multiple noise models. With the aim to study only "reasonable" noise models that are expected to arise in practise, d-monotonic noise models for any distance function d will be studied.

4.7.2 Classes of rules

Rules from important classes such as PM-c and PD-c rules [18] as well as generalized scoring rules which will be defined below [76], are examined in order to determine robust rules.

The following extra definitions are required for the definition of generalized scoring rules.

Definition 34. *A pair of vectors $y, z \in \mathbb{R}^k$ is* equivalent *if for every $i, j \in \{1, .., k\}$ it holds that $y_i \geq y_j \leftrightarrow z_i \geq z_j$.*

Definition 35. *A function is* compatible *if for every equivalent pair of vectors* $y, z \in R^k$ *it holds that* $g(y) = g(z)$.

Definition 36. *A* generalized scoring rule(GSR) *is given by a pair of functions* (f, g)*, where* $f : L(A) \to \mathbb{R}^k$ *maps every ranking to a k-dimensional vector and a compatible function* $g : \mathbb{R}^k \to D(L(A))$ *maps every k-dimensional vector to a distribution over rankings. Then, given a profile* $\pi = (\sigma_1, .., \sigma_n)$ *the rule outputs* $g(\sum_{i=1}^{n} f(\sigma_i))$.

The above classes contain almost all prominent voting rules such as Kemeny's rule (PM-c and GSR), ranked pairs (PM-c and GSR), positional scoring rules (PD-c and GSR), maximin(PM-c and GSR) and STV(GSR).

A figure that shows the relationship between these classes of voting rules is the following [19]:

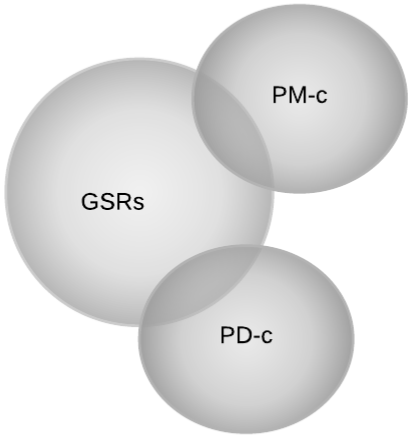

Figure 4.3: Three important families of rules.

Mossel et.al [53] give a geometrical equivalent class with GSR, the hyperplane rules. Given a profile π, x^{π}_{σ} denotes the fraction of times the ranking $\sigma \in L(A)$ appears in π. Then, the point $x^{\pi} = (x^{\pi}_{\sigma})_{\sigma \in L(A)}$ lies in a probability(as $\sum_{\sigma \in L(A)} x^{\pi}_{\sigma} = 1$ and $0 \leq x^{\pi}_{\sigma} \leq 1, \forall \sigma \in L(A)$) simplex $\Delta^{m!}$, i.e the $m!$ rankings in $L(A)$ are used to indicate the $m!$ dimensions of every point in the simplex $\Delta^{m!}$. Assigning weights $w_{\sigma} \in \mathbb{R}$ to every ranking $\sigma \in L(A)$ a hyperplane H can be formed as $H(x) = \sum_{\sigma \in L(A)} w_{\sigma} \cdot x_{\sigma}$ for any point x in $\Delta^{m!}$.

Example 4.7.1. *Let* $A = \{a, b, c\}$ *and* π *the voting profile with the following votes:*

1. $a > b > c$
2. $b > a > c$
3. $b > a > c$
4. $c > b > a$

Then the point x^π *would have the dimensions* $(1, 0, 2, 0, 0, 1)$ *with the dimensions being indicated by the rankings* $a > b > c$, $a > c > b$, $b > a > c$, $b > c > a$, $c > a > b$, $c > b > a$ *respectively.*

Definition 37. *A* hyperplane rule *is given by* $r = (H, g)$*, where* $H = \{H_i\}_{i=1}^{l}$ *is a finite set of hyperplanes, and* $g : \{+, 0, -\}^l \to D(L(A))$ *is a function that takes as input the signs of all the hyperplanes at a point and returns a distribution over rankings. Thus,* $r(\pi) = g(sgn(H(x^\pi)))$*, where* $sgn(H(x^\pi)) = (sgn(H_1(x^\pi)), ..., sgn(H_l(x^\pi)))$ *and* $sgn : \mathbb{R} \to \{+, -, 0\}$ *is the sign function .*

Using this equivalent definition of GSR , the no holes property can be defined. Informally, *a hyperplane rule (GSR)* is *without holes* if whenever it outputs the same ranking (without ties) almost everywhere around a point x^π in the simplex, then the rule outputs the same ranking (without ties) on π. Examples of common rules which are GSR without holes are the Kemeny's rule, STV, the maximin rule, the ranked pairs method and all positional scoring rules. As all prominent rules that are known to be GSR are also GSR without holes, it is believed [19] that without holes property is quite mild and does not restrict the GSR class significantly.

4.7.3 A robust rule in GSR

Caragiannis et. al [19] study a natural voting rule in GSR denoted as modal ranking rule, that selects the ranking that appears the highest number of times in a voting profile, i.e. it selects the most common ranking among the given votes. It is very interesting that this natural rule is proved to be a monotone-robust rule for all d-monotonic noise models given any d distance function. In fact, it is the unique robust rule against any monotonic noise model among a large class of voting rules, GSR rules without holes.

Theorem 4.7.1. *Let* r *be a (possibly) randomized generalized scoring rule without holes. Then,* r *is monotone robust with respect to all distance functions if and only if* r *selects the most common ranking rule with probability 1 where it is unique.*

The above theorem shows that a natural ranking rule, the modal ranking rule is the unique rule among a large class of rules that will output with probability close to 1 the ground truth given infinite samples from any "realistic" (d-monotonic for any distance function "d") noise model.

4.7.4 Robust rules in PM-c and PD-c class

Extending the search for robust rules to the classes of PM-c and PD-c rules, it is proved that the modal ranking rule continues to be the unique robust rule against any monotonic noise model as there is no PM-c or PD-c rule with this property.

Theorem 4.7.2. *For* $m \geq 3$ *alternatives, no PM-c rule or PD-c rule is monotone-robust with respect to all distance functions.*

Therefore, we get that the modal ranking rule, a rule that has passed by in traditional social choice, it can be extremely useful in human-computation and crowdsourcing systems. Indeed, while modal ranking rule does not satisfy many desirable properties such as monotonicity, it is the unique rule among these large classes of voting rules that can output the underlying ranking given a massive number of samples from any “realistic” noise model.

Chapter 5

Voting in social networks

In the previous chapters it has been assumed that the preferences of each voter are independent. In reality, however, voters are clearly influenced by the opinions of people who are related to them, i.e. the people in their social network. As it is acknowledged, social networks play an important role in the individuals' behavior [21] and their structure can be used to explain the ways that people's behavior is correlated. Because of this, social choice in social networks is deemed to be extremely important[62]. While until recently social choice in social networks has received little attention, the emergence of online social networks and the availability of data that reveals these relationships, has led to an increasing research in this direction [10, 15, 17].

In this chapter, voting in social networks, and specifically, voting under the maximum likelihood approach will be presented.

5.1 Maximum Likelihood Approach and Social Networks

Until now, the noise models that were presented assumed that the votes are drawn conditionally independent given the ground truth. However, as it has been explained above, the assumption that the voters' preferences are independent can be easily disputed. Thus, in this section, noise models in which the social network structure affects how the votes are formed will be examined.

5.1.1 A noise model that does not affect the MLE

The first noise model [28] that will be presented takes into account the social network structure but it ends up with an MLE that is not affected by the social network structure.

Let V the set of vertices in the social network graph where each vertex v denotes a voter and $N(v)$ the voter's neighbors, i.e. the voters with whom v is connected. Consider A_v the vote of voter v and $A_{N(v)}$ the vote profile consisting of the votes of all the neighbors-voters of v. Then, supposing that the ground truth is o^*, the probability of observing a vote profile π is $Pr[\pi|o^*] = \prod_{v \in V} f_v(A_v, A_{N(v)}|o^*)$, where

f_v is a function associated with a voter v and it is intended to illustrate the interaction between v and its neighbors given the ground truth.

Moreover, it is assumed that functions f_v take a particular form. Specifically, it is assumed that for any voter v, there exist functions g_v and h_v such that f_v can be factored as $f_v(A_v, A_{N(v)}|o^*) = g_v(A_v|o^*) \cdot h_v(A_v, A_{N(v)})$. That means, there is one factor (g_v) that captures the tendency of v to vote for the correct outcome o^* and one factor(h_v) that captures the tendency of v to vote similarly with his neighbors. As g_v does not depend on $A_{N(v)}$ and h_v does not depend on o^*, the tendency of any voter to vote for the correct outcome is independent of his tendency to agree with his neighbors. For example, in this model, there is still a positive probability for the voter to vote for the correct outcome, even when all the neighbors of a voter vote for the incorrect outcome.

It must be noticed that the model makes no assumption about the form of the votes and the outcome space, i.e. an outcome could be an alternative, a subset of the alternatives or a full ranking and similarly a vote could take these forms, independently of what the outcome space is.

Taking samples of this model, under maximum likelihood approach, the objective is to return an outcome that maximizes the conditional probability of observing the given profile π, i.e. an outcome in $\arg\max_o Pr[\pi|o]$.

Example 5.1.1. *Let A the set of two alternatives and V the set of two voters which in the social network graph are connected via an edge. Suppose that each voter votes for a candidate and the outcome is one of the two alternatives. For each voter $v \in V$ let $g_v(A_v = o|o) = 0.7$ and $g_v(A_v = o'|o) = 0.3$, showing that each voter is more likely to vote for the true winner, where the true winner is denoted with o and the remaining alternative with o'. It is also assumed that $h_v(A_v = o, A_{v'} = o) = 1.142$, $h_v(A_v = o, A_{v'} = o') = 0.762$ indicating that it is more likely that the voters agree with each other, where v is any voter and v' the remaining voter.*

Then, the probability of a voter v to vote for the correct winner is $Pr[A_v = o|o] = Pr[A_v = o, A_{v'} = o|o] + Pr[A_v = o, A_{v'} = o'|o] = 0.7 \cdot 1.142 \cdot 0.7 \cdot 1.142 + 0.7 \cdot 0.762 \cdot 0.3 \cdot 0.762 = 0.761$.

In contrast, assume that the two voters are not connected with each other. Then, for any voter the probability of voting for the correct winner would be $Pr[A_v = o|o] = 0.7$ which is less than the probability of voting for the correct when the two voters are connected.

Thus, in this example it could be concluded that when the voters are connected, they benefit from each other and they are more likely to vote for the true winner.

5.1.2 Computing the MLE

Having defined the noise model, the maximum likelihood estimator can be computed. As it is mentioned above, while the social network has been taken into account, it ends up that the structure of social network does not affect the maximum likelihood estimator and Lemma 5.1.1 follows.

Lemma 5.1.1. *Given the functions f_v, h_v, g_v, the maximum likelihood estimator does not depend on the social network structure, i.e. it does not depend on the* *functions* h_v *and the maximum likelihood estimator of the correct outcome is given by* $\arg\max_o \prod_{v \in V} g_v(A_v|o)$.

Proof. The maximum likelihood estimator of the correct outcome given an input profile π is $\arg\max_o Pr[\pi|o] = \arg\max_o \prod_{v \in V} f_v(A_v, A_{N(V)}|o) =$ $\arg\max_o \prod_{v \in V}(g_v(A_v|o) \cdot h_v(A_v, A_{N(v)})) = \arg\max_o(\prod_{v \in V} g_v(A_v|o) \cdot$ $\prod_{v \in V} h_v(A_v, A_{N(v)})) = \arg\max_o \prod_{v \in V} g_v(A_v|o)$, where the last transition holds since the term $\prod_{v \in V} h_v(A_v, A_{N(v)})$ does not depend on the ground truth. $\square$

From Lemma 5.1.1 it follows that the results derived in the maximum likelihood approach with the standard independence among the voters still hold in this setting. For example, in the case where there are only two alternatives and each voter votes for one alternative, the maximum likelihood estimator of the correct outcome is still the alternative which gets the larger number of votes, i.e. the majority winner.

Thus, one may conclude that modeling the social network structure does not necessarily change the maximum likelihood estimator. However, it would be "naive" to suppose that the social network structure can be ignored. Other models and in particular models that do not assume that the tendency of any voter to vote for the correct outcome is independent of his tendency to agree with his neighbors, may lead to different results. One such model where the social network structure affects the voting rule will be presented in the next subsection.

5.1.3 The Independent Conversations Model

Under the independent conversations model [29], it is assumed that there are two alternatives and every voter talks with all of his neighbors, with each conversation ending up in favor of one of the two alternatives. In other words, for every edge in the social network graph, exactly one conversation takes place and it is supposed that the outcomes of the conversation are independent and identically distributed. Then, according to the results of the conversations that each voter participated, he votes for the alternative that the majority of conservations selected. A more formal definition of this model is given below.

Let V the set of vertices which denote the voters and E the set of edges of a social network graph. Each edge $e = (v, w) \in E$ is associated with a vote A_e which is equal to the true winner with probability $p > 1/2$ and to the other alternative with probability $1 - p$. The edge profile $A_E = \{A_e | e \in E\}$ is not directly observed, but each voter v votes according to the majority of his incident edges that means $A_v = \text{maj}\{A_{(v,w)} | w \in N(v)\}$. Hence, the probability of observing a vote profile $\pi = (A_v)_{v \in V}$ given that the correct alternative is o^* is the sum of the probabilities of all edge profiles A_E that can produce the vote profile π which is,
$Pr[\pi|o^*] = \sum_{A_E : \forall v \in V, A_v = maj\{A_{(v,w)} | w \in N(v)\}} p^{n_{o^*}(A_E)} \cdot (1-p)^{|E| - n_{o^*}(A_E)}$, where $n_{o^*}(A_E)$ is the number of edges associated with o^* when the edge profile is A_E. Then, the maximum likelihood estimator of the outcome is the alternative that maximizes the above expression i.e. the alternative in $\arg\max_o Pr[\pi|o]$.

Example 5.1.2. *Let $A = \{1, -1\}$ the set of alternatives and the social network given by the figure 5.1 [29]. The first graph gives the vote profile where each voter denoted with an open vertex votes for candidate -1 and each voter denoted with a close vertex votes for candidate 1. The other two graphs, give the only two edge profiles which are compatible with the given vote profile, where every open edge is associated with alternative -1 and every close edge is associated with alternative 1. Then the probability of observing any one profile of the two edge profiles given that alternative 1 is the correct winner is $p^5 \cdot (1-p)^4$ as there are four edges that vote for the incorrect alternative and 5 edges that are associated with the correct winner. Hence, the probability of observing the vote profile given that the correct winner is 1 is $Pr[\pi|1] = Pr[A_{E1}|1] + Pr[A_{E2}|1] = 2 \cdot p^5 \cdot (1-p)^4$. Similarly, the probability of observing the vote profile given that the correct winner is -1 is $Pr[\pi| -1] = 2 \cdot p^4 \cdot (1-p)^5$. Therefore, a maximum likelihood estimator of the correct winner would select alternative 1 as he maximizes ($p > 1/2$) the conditional probability of observing the given vote profile.*

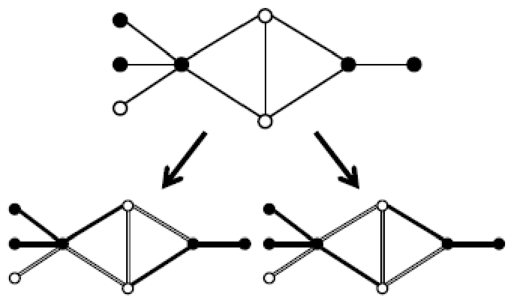

Figure 5.1: A social network structure with a vote profile for the vertices, and the two edge profiles that are consistent with this vote profile.

As the example shows, computing the probability of observing a vote profile is associated with computing the probabilities of observing the compatible edge profiles. However, it has to be noticed that not all vote profiles have compatible edge profiles. For example, the vote profile that consists of two voters that are connected via an edge and the two voters vote for a different alternative, is not consistent with any edge profile. This could be faced by extending the model so that every voter has a small probability of voting against the majority of his incident edges.

While in the Example 5.1.2 enumerating the consistent edge profiles was easy, it has to be examined what happens in general.

5.1.4 Computational Complexity

One may wonder whether there is an easy way to enumerate or just count the different compatible edge profiles. However, theorem 5.1.1 shows that this is not the case and suggests thats it is a hard counting problem.

Theorem 5.1.1. *Computing the probability of observing a vote profile π given the correct outcome o^*, $Pr[\pi|o^*]$, is $\#P-$ hard under the independent conversations model.*

Intuitively, the computational complexity of computing the probability of observing a vote profile indicated by Theorem 5.1.1 is due to the hidden variables (edge profiles) over whose possible values must be summed. Another approach would be to estimate the hidden variables together with the true winner rather than summing over all the hidden variables.

Specifically, under independent conversations model, that means estimating the correct winner o^* with the edge profile A_E^* so as to maximize the probability $Pr[\pi, A_E^*|o^*]$. Considering that $Pr[\pi, A_E|o] = Pr[\pi|A_E, o] \cdot Pr[A_E|o]$, where $Pr[\pi|A_E, o] = Pr[\pi|A_E] = 1$ if A_E is compatible with π and 0 otherwise, then the goal is to find the o^* and a compatible A_E^* with π that maximize the probability $Pr[A_E^*|o^*]$. It is proved that this can be done in polynomial time even under a richer model that is defined below.

5.1.5 Independent weighted conversations model

Independent weighted conversation model [29] is identical to the independent conversations model, except the fact that here there is not a universal probability p of ending up with the correct alternative; different edges have different probabilities $p_e \geq 1/2$ of associating with the correct alternative. The existence of different probabilities p_e associated with different edges stems from the same idea with a previous model [56] where some voters are more skillful and thus, they have a greater probability of voting for the correct alternative.

As it is mentioned above, estimating the edge profile along with the correct winner can be done in polynomial time under the independent weighted conversations model.

Theorem 5.1.2. *An element of* $\arg\max_{(o^*, A_E^*)} Pr[\pi, A_E^*|o^*]$ *can be computed in polynomial time, even in the independent weighted conversations model.*

Until now, three different noise models that take social network structure into account have been presented. While the first noise model ends up with an optimal rule that is not affected by the social network structure, under the independent conversations model and the independent unweighted conversations model, the optimal rule is affected by the social network, showing that it would not be right to assume that social network structure can just be ignored. Although these models try to capture the interaction among the voters, they ignore one important aspect; the time [29]. It could be much more realistic if the gradual evolution of a voter's preference was modeled. Nevertheless, these simple models give a little insight into how a social network may affect the optimal voting rule [29]. However, as it was explained in the previous chapter, finding an optimal rule (MLE) is restrictive [18]. Thus, it is equally important to examine the setting presented in the previous chapter in the social network context.

5.2 Dependent samples

In this section, it will be studied how many samples are needed to reveal the ground truth when the samples are dependent reflecting the social network structure among the agents. Specifically, the model that will be used assumes that the agents are divided into different areas with each area having its own ground truth stemming from a global ground truth. It is interesting to notice that this model indeed follows reality. It is a fact that people living in the same geographical area evaluate many things in a similar way but possibly different from people in other areas. For example, assume that people from different areas are asked to rank the football teams of the country. It is expected to observe that the rankings will differ in the teams associated with the areas that the samples come from, as it is quite natural to overestimate "their" team but they will also have commons in the teams that are independent of the areas. Furthermore, another important reason that justifies the selection of this model is that a geographical area is often associated with people with specific socioeconomic status. For instance, in some areas wealthy people live, while in others, most of the people are unprivileged. Thus, it is expected that different samples will arise when people from different areas are asked to evaluate a number of alternatives, since their different socioeconomic status may make them evaluate some things differently.

The model that captures these situations is defined formally below.

5.2.1 Model

Let $A = \{a_1, .., a_m\}$ the set of alternatives and σ^* the global underlying ground truth with a_i the alternative that is on the i^{th} position in the true ranking. It is assumed that there are k different areas with the i^{th} area having its own underlying ground truth σ_i^*, for any $i \in \{1, .., k\}$. Any ranking σ_i^* has at most l distance from the global ranking σ^*, that means any ranking σ_i^* can be derived from σ^* by at most l swaps. Given that the samples follow the Mallows model, the purpose is to identify how many samples are required to reveal with high probability the true rankings of the k areas.

5.2.2 A lower bound

By Theorem 4.6.1, it follows that taking $O(\log(\frac{m}{\epsilon}))$ from every area. i.e. a total of $O(k \cdot \log(\frac{m}{\epsilon}))$ samples, it would output the underlying rankings with probability $1 - \epsilon$. However, this approach ignores the fact that the rankings of the k areas are derived from the global ranking σ^* and thus, they have many common pairs.

Taking this into account, the first approach that will be examined assumes that $O(\log(\frac{m}{\epsilon}))$ samples are taken from the p^{th} area i.e. σ_p^* is known with high probability and then it is examined how many extra samples from any area $q, q \neq p$ are required in order to output with high probability the σ_q^*. Although these approach seems to take advantage of the fact that the rankings of any two teams have common pairs, it is proved that given the ranking σ_p^*, the number of the required samples is the same with the number of the required samples when none ranking is known.

Theorem 5.2.1. *Given that σ_p^*, $p \in \{1,..,k\}$ is known, then $\Omega(\log(\frac{m}{\epsilon}))$ samples from Mallows model are required to reconstruct the true ranking σ_q^* with probability at least $1-\epsilon$, for any $q \in \{1,..,k\}$.*

Proof. Let r a voting rule and assume that $N^r(\epsilon) = n$. By the definition of $N^r(\epsilon)$ (subsection 4.2), it follows that $Acc^r(n,\sigma) \geq 1-\epsilon$ for any $\sigma \in L(A)$. Since the true rankings of the two areas have a maximum KT-distance l from the ranking σ^*, then the rankings σ_p^* and σ_q^* have KT-distance at most $l' = 2 \cdot l$ as it follows from the following probability of distance functions: $d_{KT}(\sigma_p^*, \sigma_q^*) \leq d_{KT}(\sigma_p, \sigma^*) + d_{KT}(\sigma^*, \sigma_q^*) \leq l + l = 2 \cdot l$. Thus, the possible results of the rule r are restricted in $N^{l'}(\sigma_p^*)$, which is the set of rankings that have l' or smaller distance from σ_p^* . Hence for any $\sigma \in N^{l'}(\sigma_p^*)$, $Pr[\pi|\sigma_q^*] = \prod_{i=1}^{n} \frac{\phi^{d_{KT}(\sigma_i,\sigma_q^*)}}{Z_\phi^m} \geq \prod_{i=1}^{n} \frac{\phi^{d_{KT}(\sigma_i,\sigma)+2\cdot l'}}{Z_\phi^m} = \phi^{2\cdot n\cdot l'} \cdot Pr[\pi|\sigma]$, where the second transition follows the following property of distance functions: $d_{KT}(\sigma_i, \sigma_q^*) \leq d_{KT}(\sigma_i, \sigma) + d_{KT}(\sigma, \sigma_q^*) \leq d_{KT}(\sigma_i, \sigma) + 2 \cdot l'$.

Then, $Acc^r(n, \sigma_q^*) = \sum_{\pi \in L(A)^n} Pr[\pi|\sigma_q^*] \cdot Pr[r(\pi) = \sigma_q^*] =$ $\sum_{\pi \in L(A)^n} Pr[\pi|\sigma_q^*] \cdot (1 - Pr[r(\pi) \neq \sigma_q^*]) = 1 - \sum_{\pi \in L(A)^n} Pr[\pi|\sigma_q^*] \cdot Pr[r(\pi) \neq \sigma_q^*] =$ $1 - \sum_{\pi \in L(A)^n} Pr[\pi|\sigma_q^*] \cdot \sum_{\sigma \in N^{l'}(\sigma_p^*)} Pr[r(\pi) = \sigma] =$ $1 - \sum_{\sigma \in N^{l'}(\sigma_p^*)} \sum_{\pi \in L(A)^n} Pr[\pi|\sigma_q^*] \cdot Pr[r(\pi) = \sigma] \leq$ $1 - \sum_{\sigma \in N^{l'}(\sigma_p^*)} \sum_{\pi \in L(A)^n} \phi^{2\cdot n\cdot l'} \cdot Pr[\pi|\sigma] \cdot Pr[r(\pi) = \sigma] = 1 - \phi^{2\cdot n\cdot l'} \cdot \sum_{\sigma \in N^{l'}(\sigma_p^*)} Acc^r(n,\sigma) \leq$ $1 - \phi^{2\cdot n\cdot l'} \cdot (m-1) \cdot (1-\epsilon)$, where the last transition holds since $m-1 \leq |N^{l'}(\sigma_p^*)| \leq m^{l'}$. Therefore, in order to have $Acc^r(n, \sigma_q^*) \geq 1-\epsilon$ we get that $\phi^{2\cdot n\cdot l'} \cdot (m-1) \cdot (1-\epsilon) \leq \epsilon$ and solving for n we get the bound $\Omega(\log(\frac{m}{\epsilon}))$. ☐

Intuitively, the above approach fails to output the true rankings of the k areas with a smaller number of samples than $O(k \cdot \log(\frac{m}{\epsilon}))$ given the ranking of the p area, as the common pairwise comparisons for all areas will be estimated more than one time, in all areas. In order to overcome this result, the next approach is to take samples from all areas in oder to estimate the pairwise comparisons that all true rankings have in common and then take extra samples from each area in order to estimate the remaining pairs. It is proved that this approach indeed gives better results and the number of samples required is reduced.

Theorem 5.2.2. *Under the approach that the common pairs for all areas are estimated first and then extra samples are taken from each team to estimate the remaining pairs of alternatives, the total number of samples required to estimate all σ_i^* forall $i \in \{1,..,k\}$, is $O(\log(\frac{(m-l'\cdot k)\cdot l'\cdot k}{\epsilon}) + k \cdot \log(\frac{l'\cdot k}{\epsilon}))$.*

Proof. Taking $\frac{n}{k}$ samples from each area it is wanted to define with high probability the pairs that they have in common by selecting the pairs (a,b) which have the

greatest difference in $n_{ab} - n_{ba}$. Hence we want,
$Pr[\forall a, b, \in A, a >_{\sigma_i^*} b, \forall i \in \{1, ..k\} \implies (a,b) \in \arg\max^{x}_{(a,b)\in A^2} n_{ab} - n_{ba}] \geq 1 - \epsilon$, where x is the number of common pairs between all areas and holds that $x \geq \binom{m}{2} - l' \cdot \binom{k}{2}$ since any pair of teams have at most l' pairs of alternatives different and there are $\binom{k}{2}$ different pairs of areas.

Let a and b a pair of alternatives such that $a >_{\sigma_i^*} b$, $\forall i \in \{1, .., k\}$. Then the probability that the difference $n_{ab} - n_{ba}$ is not in the $\arg\max^{x}_{(a,b)\in A^2} n_{ab} - n_{ba}$ is:
$Pr[(a,b) \notin \arg\max^{x}_{(a,b)\in A^2} n_{ab} - n_{ba}] = Pr[\exists c, d \in A, \exists i \in \{1, .., k\}, d >_{\sigma_i^*} c \wedge n_{cd} - n_{dc} \geq n_{ab} - n_{ba}]$ (1)

Let c and d a pair of alternatives such that c is not preferred to d by all areas. Then the probability of difference $y = n_{cd} - n_{dc}$ being larger than $z = n_{ab} - n_{ba}$ is:
$Pr[z - y \leq 0] \leq Pr[|z - y - \mathbb{E}[z - y]| \geq \delta] \leq 2 \cdot e^{-2\cdot\delta^2\cdot n} \leq 2 \cdot e^{-2\cdot\delta^2_{min}\cdot n}$ (2), where $\delta = \mathbb{E}(z - y)$.

Then from (1) and (2) and union bound we have that :
$Pr[(a,b) \notin \arg\max^{x}_{(a,b)\in A^2} n_{ab} - n_{ba}] \leq 2 \cdot l' \cdot \binom{k}{2} \cdot e^{-2\cdot\delta^2_{min}\cdot n}$.

Thus, $Pr[\exists a, b, \in A, a >_{\sigma_i^*} b, \forall i \in \{1, ..k\} \wedge (a,b) \notin \arg\max^{x}_{(a,b)\in A^2} n_{ab} - n_{ba}] \leq (\binom{m}{2} - l \cdot \binom{k}{2}) \cdot l \cdot \binom{k}{2} \cdot 2 \cdot e^{-2\cdot\delta^2_{min}\cdot n} \leq \epsilon$ and solving by n we get that
$n \geq \frac{1}{2\cdot\delta^2_{min}} \cdot \log(\frac{(\binom{m}{2} - l\cdot\binom{k}{2})\cdot l\cdot\binom{k}{2}}{\epsilon})$.

As the common pairs are defined, we want to define the pairs that differ in any area by taking extra samples from each area. Firstly, the extra samples for the first area are computed by selecting the pairs (a, b) such that the number of agents that prefer a to b is larger than the number of samples that prefer b to a.

Hence we want,
$Pr[\forall a, b, \in A, a >_{\sigma_1^*} b \wedge \exists i \in \{1, ..k\}, b >_{\sigma_i^*} a \implies n_{ab} - n_{ba} \geq 1] \geq 1 - \epsilon$.

Let $\delta = p_{a>b} - p_{b>a}$. Then for a and b such that $a >_{\sigma_1^*} b \wedge \exists i \in \{1, ..k\}, b >_{\sigma_i^*} a$ the probability that the number of voters from the first team that prefer a to b is not larger than the number of voters that prefer b to a is:
$Pr[n_{ab} - n_{ba} \leq 0] = Pr[\frac{n_{ab} - n_{ba}}{n_1}] \leq Pr[|\frac{n_{ab} - n_{ba}}{n_1} - \mathbb{E}[\frac{n_{ab} - n_{ba}}{n_1}]| \geq \delta] \leq 2 \cdot e^{-2\cdot\delta^2\cdot n_1} \leq 2 \cdot e^{-2\cdot\delta^2_{min}\cdot n_1}$.

The probability of returning a wrong ranking is:
$Pr[\exists a, b \in A, a >_{\sigma_1^*} b \wedge \exists i \in \{1, ..k\}, b >_{\sigma_i^*} a \wedge n_{ab} - n_{ba} \leq 0] \leq 2 \cdot [\binom{m}{2} - x] \cdot e^{-2\cdot\delta^2_{min}\cdot n_1} \leq 2 \cdot l \cdot \binom{k}{2} \cdot e^{-2\cdot\delta^2_{min}\cdot n_1} \leq \epsilon$ and hence, we have $n_1 \geq \frac{1}{2\cdot\delta^2_{min}} \cdot \log(\frac{l\cdot\binom{k}{2}}{\epsilon})$.
Similarly, we want $n_i \geq \frac{1}{2\cdot\delta^2_{min}} \cdot \log(\frac{l\cdot\binom{k}{2}}{\epsilon})$ samples to define the pairs for each true ranking $i \in \{1, ..k\}$ and thus, the theorem follows. □

Although this approach decreases the number of required samples, it is not optimal as the number of common pairs used is the minimum possible number but in reality the k areas will have a larger number of common pairs. However, it gives an insight of how the social network structure affects the minimum number of samples required to reveal the ground truth; ignoring the social network and the dependencies among the agents would result in a much bigger number of required samples.

Chapter 6

Conclusion

In this thesis, important results in both traditional social choice theory and computational social choice have been presented. In particular, after the introduction to basic concepts and results in social choice theory, emphasis has been given on the maximum likelihood approach. Under the maximum likelihood approach, voting methods, given a number of votes that are assumed to be noisy estimators of the ground truth, aim to reveal the underlying truth which ranks the alternatives according to a quality measure. Different settings and noise models that define the way that different samples can be observed were examined under the maximum likelihood approach and important results were presented.

Specifically, studying the problem of selecting a set of “good” alternatives [59] suggested that the selection of the appropriate voting rule is important as it can significantly affect the performance. Moreover, since in many theoretical and practical applications agents are not able to give a total order of the alternatives, maximum likelihood approach was examined in the case of partial orders [80]. Under this model, pairwise scoring rules play an important role as they are the only maximum likelihood estimators for neutral-pairwise noise models.

However, it was argued that the maximum likelihood estimator requirement is too restrictive and thus, it was studied how many samples different rules need in order to output the underlying ranking with high probability [18]. It was supposed that samples follow the Mallows model but at the same time, some generalizations were made in order to predict some noise models that may arise in practice. A possible future direction would be to extend the analysis of the number of required samples to models where the input votes are not given in the form of total orders but rather, the agents give partial orders or lists with top alternatives. This would be of great importance as in practice the number of required samples translates into the budget that will be consumed to draw the votes and in many real applications, taking total orders as input is unrealistic due to the extremely large number of alternatives [18].

All the aforementioned settings, though, made an assumption that can be easily disputed; it was assumed that the agents' preferences are conditionally independent given the ground truth. Hence, in the last chapter, maximum likelihood approach under the social network context was examined. Some preliminary results, that give an insight of how the social network structure affects both the maximum likelihood estimator and the samples that a rule needs to reconstruct the ground truth with high probability, were presented. It would be interesting to study further these models with the inclusion of some real-world aspects such as the evolution of the voter's preference over the time [28] as well as the extension of the independent conversations model to the case where the number of alternatives is larger than two and the agents give total orders.

Bibliography

[1] Arrow, K.J. 1951. *Social Choice and Individual Values.* New Haven: Cowles Foundation, 2nd edition, 1963.

[2] Bakhkhat, M. and Truchon, M. 2004. *Maximum likelihood approach to vote aggregation with variable probabilities.*. Social Choice and Welfare, 23:161-185.

[3] Bartholdi, J. III, Tovey, C. A. And Trick, M.A. 1989. *Voting schemes for which it can be difficult to tell who won the election.* Social Choice and Welfare, 6(3):157-165.

[4] Bartholdi, J. III, Tovey, C. A. And Trick, M.A. 1989. *The computational difficulty of manipulating an election.* Social Choice and Welfare, 6(3):227-241.

[5] Bartholdi, J. III And Orlin, J.B. 1991. *Single transferable vote resists strategic voting.* Social Choice and Welfare, 8(4):341-354.

[6] Betzler, N., Niedermeier, R. And Woeginger, G.J. 2011. *Unweighted coalitional manipulation under the Borda rule is NP-hard.* In the Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI), 55-60. AAAI Press.

[7] Bickel, E. 2007. *Some comparisons among quadratic, spherical, and logarithmic scoring rules.* Decision Anal. 4(2) 49-65.

[8] Bickel, E. 2010. *Scoring Rules and Decision Analysis Education.* Decision Anal. 7(4) 346-357.

[9] Black, D. 1958. *Theory of Committees and Elections.* Cambridge University Press.

[10] Boldi, P., Bonchi, F., Castillo, C. And Vigna, S. 2009. *Voting in Social Networks.* In the Proceedings of CIKM'09, 777-786.

[11] Brams, S. J. and Fishburn, P. C. 1983. *Approval Voting.* Birkhauser, Boston.

[12] Brams, S. J. and Fishburn, P. C. 2003. *Going from theory to practice: The mixed success of approval voting.* Social Choice and Welfare, 2:5457-474.

[13] Brams, S. J. and Sanver, M. R. 2005. *Critical Strategies Under Approval Voting, Who Gets Ruled In and Ruled Out.* Electoral Studies, 25(2), 287-305.

[14] Braverman, M., And Mossel, E. 2008. *Noisy sorting without resampling.* In proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 268-276.

[15] S. Brânzei, S. And Larson, K. 2009. *Coalitional affinity games and the stability gap.* In the Proceedings of IJCAI'09, 79-84.

[16] Brandt, F., Brill, M., Hemaspaandra, E. And Hemaspaandra, L. 2010. *Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates.* In M. Fox and D. Poole, editors, Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI), 715-722. AAAI Press.

[17] S. Brânzei, S. And Larson, K. 2011. *Social distance games.* In the Proceedings of IJCAI'11, 91-96.

[18] Caragiannis, I., Procaccia, A. D. And Shah, N. 2014. *When do noisy votes reveal the truth?* In the proceedings of the 14th ACM Conference on Electronic Commerce (EC), 143-160.

[19] Caragiannis, I., Procaccia, A. D. And Shah, N. 2014. *Modal Ranking: A Uniquely Robust Voting Rule.* In the proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI) .

[20] Charbit, P., Thomasse, S. And Yeo, A. 2007. *The minimum feedback arcset problem is NP-hard for tournaments.* Combinatorics, Probability and Computing, 16(1):1-4.

[21] Christakis, N. And Fowler, J. 2011. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives.* Back Bay Books.

[22] Condorcet, de M. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Imprimerie Royale.

[23] Conitzer, V. And Sandholm, T. 2003. *Universal voting protocol tweaks to make manipulation hard.* In the Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI), 781-788.

[24] 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.

[25] Conitzer, V. And Sandholm, T. 2006. *Nonexistence of voting rules that are usually hard to manipulate.* In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI), 627-634. AAAI Press.

[26] Conitzer, V., Sandholm, T. And Lang, J. 2007. *When are elections with few candidates hard to manipulate?* Journal of the ACM, 54(3).

[27] Conitzer, V., Matthew, R. And Xia, L. 2009. *Preference functions that score rakings and maximum likelihood estimation.* In IJCAI, 109-115.

[28] Conitzer, V. 2012. *Should social network structure be taken into account in elections?* Mathematical Social Sciences, vol.64, no. 1, 100-102.

[29] Conitzer, V., 2013. *The maximum likelihood approach to voting on social networks.* In the Proceedings of the 51st Annual Allerton Conference on Communication, Control and Computing (Allerton-13), 1482-1487.

[30] Davies, J., Katsirelos, G., Narodytska, N. And Walsh, T. 2011. *Complexity of and algorithms for Borda manipulation.* In Proceedings of the National Conference on Artificial Intelligence (AAAI), 657-662. AAAI Press.

[31] Dawid, A. P. 1984. *Statistical Theory: The Prequential Approach.* Journal of the Royal Statistical Society, Ser. A, 147, 278-292.

[32] Desmedt, Y. And Elkind, E. 2010. *Equilibria of plurality voting with abstentions.* In Proceedings of the ACM Conference on Electronic Commerce (EC), 347-356. ACM Press.

[33] Elkind, E. And Lipmaa, H. 2005. *Hybrid voting protocols and hardness of manipulation.* In Proceedings of the 16th International Symposium on Algorithms and Computation (ISAAC), volume 3827 of Lecture Notes in Computer Science (LNCS), 206-215, Springer-Verlag.

[34] Elkind, E., Faliszewski, P., And Slinko, A. 2010. *On the role of distances in defining voting rules.* In Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge, (TARK). 108-117.

[35] Ephrati, E. and Rosenschein, J. S. 1991. *The Clarke Tax as a Consensus Mechanism Among Automated Agents.* In the Proceedings of the National Conference on Artificial Intelligence (AAAI), 173?178.

[36] Faliszewski, P., Hemaspaandra, E., Hemaspaandra, L. And Rothe, J. 2009. *A richer understanding of the complexity of election systems.* In S. Ravi and S. Shukla, editors, Fundamental Problems in Computing: Essays in Honor of Professor Daniel J. Rosenkrantz, Springer-Verlag.

[37] Faliszewski, P., Hemaspaandra, E., Hemaspaandra, L. And Rothe, J. 2009. *The shield that never was: Societies with single-peaked preferences are more open to manipulation and control.* In the Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK), 118- 127. ACM Press.

[38] Felsenthal, D. S. 1989. *On combining approval with disapproval voting.* Behavioral Science, 34, 53-60.

[39] Fishburn, P.C. 1974. *Paradoxes of voting.* The American Political Science Review 68, 2, 537-546.

[40] Friedman, D. 1979. *An effective scoring rule for probability distributions.* Discussion Paper 164, University of California, Los Angeles, Los Angeles, CA.

[41] Friedman, D. 1983. *An effective scoring rule for probability distributions.* Management Sci. 29(4): 447-454.

[42] Garthwaite, P. H., Kadane, J. B., And O'Hagan, A. 2005. *Statistical Methods for Eliciting Probability Distributions.* Journal of the American Statistical Association, 100, 680-700.

[43] Gibbard, A. 1973. *Manipulation of voting schemes.* Econometrica, 41:587-602.

[44] Gneiting, T. and Raftery, A. E. 2007. *Strictly proper scoring rules, prediction, and estimation.* J. Am. Statist. Ass., 102, 359-378.

[45] Hemaspaandra, E. And Hemaspaandra, L. 2007. *Dichotomy for voting systems.* Journal of Computer and System Sciences, 73(1):73-83.

[46] Hillinger, C. 2004. *On the possibility of democracy and rational collective choice.* Discussion Paper 2004-21, U. Munich.

[47] Hillinger, C. 2004. *Utilitarian collective choice and voting.* Discussion Paper 2004-25, U. Munich.

[48] Hillinger, C. 2005. *The case for utilitarian voting.* Discussion Papers in Econ. No. 653, U. Munich.

[49] Law, E. And von Ahn, L. 2011. *Human Computation.* Morgan & Claypool.

[50] Mallows, C. L., 1957. *Non-null ranking models* . Biometrika 44,114-130.

[51] Mao, A., Procaccia, A. D., And Chen, Y.2013. *Better human computation through principled voting.* In Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI).

[52] May, K. 1952. *A set of independent, necessary and sufficient conditions for simple majority decisions.* Econometrica, 20:680-684.

[53] Mossel, E., Procaccia, A. D. And Racz, M. Z. 2013. *A smooth transition from powerlessness to absolute power.* Journal of Artificial Intelligence Research. 48:923-951.

[54] Moulin, H. 1983. *The Strategy of Social Choice.* North-Holland.

[55] Moulin, H. 1988. *Axioms of Cooperative Decision Making.* Cambridge University Press.

[56] Nitzan, S. And Paroush, J. 1982. *Optimal decision rules in uncertain dichotomous choice situations.* International Economic Review, vol. 23, no. 2, 289-297.

[57] Procaccia, A.D. And Rosenschein, J.S. 2007. *Average-case tractability of manipulation in voting via the fraction of manipulators.* In Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi- Agent Systems (AAMAS), 718-720. AAAI Press.

[58] Procaccia, A.D. And Rosenschein, J.S. 2007. *Junta distributions and the average case complexity of manipulating elections.* Journal of Artificial Intelligence Research, 28:157-181.

[59] Procaccia, A. D., Reddi, S.J., And Shah, N. 2012. *A maximum likelihhod approach for selecting sets of alternatives.* In Proceedings of the 28th Annual Conference on Uncertainty in Artificial Intelligence , (UAI). 695-794.

[60] Procaccia, A. D. 2012. *Computational Voting Theory: Of the Agents, By the Agents, For the Agents.* Ph.D. thesis, The Hebrew University of Jerusalem.

[61] Roby, T. B. 1965. *Belief states: A preliminary empirical study.* Behavioral Sci. 10(3):255-270.

[62] Salehi-Abari, A. And Boutilier, C. 2012. *Empathetic social choice on social networks.* In the Proceedings of COMSOC.

[63] Samuelson, P.A. 1967. *Arrow's mathematical politics.* In S. Hook, editor, Human Values and Economic Policy. New York University Press.

[64] Satterthwaite, M.A. 1975. *Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions.* Journal of Economic Theory, 10:187-217.

[65] Shapley, L. And Grofman, B. 1984. *Optimizing group judgmental accuracy in the presence of interdependencies.*Public Choice, vol. 43, no. 3, 329-343.

[66] Shuford, E. H. Jr., Albert, A., Massengill, E.H. 1966. *Admissible probability measurement procedures.* Psychometrika 31(2):125-145.

[67] Smith, W.D. 2000. *Range voting.* Technical Report 56. NEC Research, Princeton, NJ, USA.

[68] Toda, M. 1963. *Measurement of subjective probability distributions.* ESD-TDR-63-407, Decision Sciences Laboratory, Electronic Systems Division, Air Force Systems Command, United States Air Force, Bedford, MA.

[69] Truchon, M. 2008. *Borda and the maximum likelihood approach to vote aggregation.* Mathematical Social Sciences, 55(1):96-102.

[70] Vickrey, W. 1960. *Utility, strategy, and social decision rules.* Quarterly Journal of Economics, 74, 507-535.

[71] Walsh T. 2009. *Where are the really hard manipulation problems? The phase transition in manipulating the veto rule.* In the Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI), 324-329. AAAI Press.

[72] Weber, R.J. 1995. *Approval voting.* Journal of Economic Perspectives, 9, 1, 39-49.

[73] Winkler, R. L. 1968. *"Good" probability assessors.* J. Appl. Meteorology 7 751-758.

[74] Winkler, R. L. 1968. *Scoring rules and the evaluation of probability assessors.* J. Amer. Statist. Assoc. 64(327) 1073-1078.

[75] Xia, L. And Conitzer, V. 2008. *A sufficient condition for voting rules to be frequently manipulable.* In the Proceedings of the 9th ACM Conference on Electronic Commerce (ACM-EC), 99-108. ACM Press.

[76] Xia, L. And Conitzer, V. 2008. *Generalized scoring rules and the frequency of coalitional manipulability.* In Proceedings of the ACM Conference on Electronic Commerce (EC), 109-118. ACM Press.

[77] Xia, L., Zuckerman, M., Procaccia, A.D., Conitzer, V., And Rosenschein, J.S. 2009. *Complexity of unweighted coalitional manipulation under some common voting rules.* In the Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI), 348-353. AAAI Press.

[78] Xia, L. And Conitzer, V. 2010. *Stackelberg voting games: Computational aspects and paradoxes.* In the Proceedings of the National Conference on Artificial Intelligence (AAAI), 921-926. AAAI Press.

[79] Xia, L., Conitzer, V., And Lang, J. 2010. *Aggregating preferences in multi-issue domains by using maximum likelihood estimators.* In AAMAS, 399-406.

[80] Xia, L. And Conitzer, V. 2011. *A maximum likelihood approach towrds aggregating partial orders.* In Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI). 446-451.

[81] Xia, L. 2011. *Computational Voting Theory: Game-Theoretic and Combinatorial Aspects.* Ph.D. Thesis. Computer Science Department, Duke University, Durham, NC.

[82] Young, H.P. 1988. *Condorcet's theory of voting.* The American Political Science Review, 82(4):1231-1244.

[83] Young, H.P. 1995. *Equity: In Theory and Practice.* Princeton, NJ:Princeton University Press.