Aggregating Preferences in Multi-Issue Domains by Using Maximum Likelihood Estimators

Lirong Xia
Dept. of Computer Science
Duke University
Durham, NC 27708, USA
lxia@cs.duke.edu

Vincent Conitzer
Dept. of Computer Science
Duke University
Durham, NC 27708, USA
conitzer@cs.duke.edu

Jérôme Lang
LAMSADE
Université Paris-Dauphine
75775 Paris Cedex, France
lang@lamsade.dauphine.fr

ABSTRACT

In this paper, we study a maximum likelihood estimation (MLE) approach to voting when the set of alternatives has a multi-issue structure, and the voters' preferences are represented by CP-nets.

We first consider general multi-issue domains, and study whether and how issue-by-issue voting rules and sequential voting rules can be represented by MLEs. We first show that issue-by-issue voting rules in which each local rule is itself an MLE (resp. a candidate scoring rule) can be represented by MLEs with a weak (resp. strong) decomposability property. Then, we prove two theorems that state that if the noise model satisfies a very weak decomposability property, then no sequential voting rule that satisfies unanimity can be represented by an MLE, unless the number of voters is bounded.

We then consider multi-issue domains in which each issue is binary; for these, we propose a general family of *distance-based noise models*, of which give an axiomatic characterization. We then propose a more specific family of natural distance-based noise models that are parameterized by a threshold. We identify the complexity of winner determination for the corresponding MLE voting rule in the two most important subcases of this framework.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences–Economics; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

General Terms

Algorithms, Economics, Theory

Keywords

Computational social choice, voting in multi-issue domains, maximum likelihood estimator, distance-based models

1. INTRODUCTION

A natural way for agents to make a joint decision when they have possibly conflicting preferences over a set of alternatives is by *voting*. Each agent (voter) is asked to report her preferences, and then a *voting rule* (or *voting correspondence*) selects the winning alternative (or multiple winning alternatives). Mathematically, a voting rule or correspondence is defined as a mapping from the set of possible preference *profiles* to the set of alternatives. Here, a profile is a vector of all the agents' preferences.

In some sense, this means that the agents' preferences are the "causes" of the joint decision. However, there is a different (and almost reversed) point of view: there is a "correct" joint decision, but the agents may have different perceptions (estimates) of what this correct decision is. Thus, the agents' preferences can be viewed as noisy reports on the correct joint decision. Even in this framework, the agents still need to make a joint decision based on their preferences, and it makes sense to choose their best estimate of the correct decision. Given a noise model, one natural approach is to choose the maximum likelihood estimate of the correct decision. The maximum likelihood estimator is a function from profiles to alternatives (more accurately, subsets of alternatives, since there may be ties), and as such is a voting rule (more accurately, a correspondence).

This maximum likelihood approach was first studied by Condorcet [5] for the cases of two and three alternatives. Much later, Young [15] showed that for arbitrary numbers of alternatives, the MLE rule derived from Condorcet's noise model coincides with the Kemeny rule [8]. The approach was further pursued by Drissi and Truchon [6]. More recently, Conitzer and Sandholm [4] studied whether and how common voting correspondences and *preference functions* (that is, mappings that take agents' preferences as input, and output one or more aggregate rankings of the alternatives) can be represented as maximum likelihood estimators. Even more recently, the maximum likelihood approach for preference functions has been investigated in more detail [3]. The related notion of *distance rationalizability* has also received attention in the computational social choice community recently [7].

All of the above work does not assume any structure on the set of alternatives. However, in real life, the set of alternatives often has a multi-issue structure: there are multiple *issues* (or *attributes*), each taking values in its respective domain, and an alternative is characterized by the values that the issues take. For example, consider a situation where the citizens of a country vote to directly determine a government plan, composed of multiple sub-plans for several interrelated issues, such as transportation, environment, and health [2]. Clearly, a voter's preferences for one issue in general depend on the decision taken on the other issues: for example, if a new highway is constructed through a forest, a voter may prefer a nature reserve to be established; but if the highway is not constructed, the voter may prefer that no nature reserve is established.

The number of alternatives in a multi-issue domain is exponential in the number of issues, which makes commonly studied voting methods impractical (for one, they require the agents to rank all the alternatives). One straightforward way to aggregate preferences in multi-issue domains is *issue-by-issue* (a.k.a. *seat-by-seat*) voting,

Cite as: Aggregating Preferences in Multi-Issue Domains by Using Maximum Likelihood Estimators, Lirong Xia, Vincent Conitzer and Jérôme Lang, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. XXX-XXX.
Copyright © 2010, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

which requires that the voters explicitly express their preferences over each issue separately, after which each issue is decided by applying local (issue-wise) voting rules independently. This makes sense if voters' preferences are *separable*, that is, if the preferences of every voter over any issue are independent of the values taken by the other issues. However, if a voter has nonseparable preferences, it is not clear how she should vote in such an issue-by-issue election. Indeed, it is known that natural strategies for voting in such a context can lead to very undesirable results [2, 9].

While in general, a voter's preferences for one issue depend on the decisions taken on other issues, on the other hand, one would not necessarily expect the preferences for one issue to depend on *all* other issues. CP-nets [1] were developed as a natural representation language for capturing such limited dependence among the preferences over multiple issues; they have some obvious similarities to Bayesian networks. Recent work has started to investigate using CP-nets to represent preferences in voting contexts with multiple issues. If there is an order over issues such that every voter's preferences for "later" issues depends only on the decisions made on "earlier" issues, then the voters' CP-nets are acyclic, and a natural approach is to apply issue-wise voting rules *sequentially* [10]. This sequential voting process has a low communication cost, and a low computational cost if each of the local voting rules is easy to compute. While assuming such an order exists is still restrictive, it is much less restrictive than assuming separable preferences (for one, the resulting preference domain is exponentially larger [10]). Recent extensions of sequential voting rules include order-independent sequential voting rules [14], as well as a framework for voting when preferences are modeled by general (that is, not necessarily acyclic) CP-nets [13]; [11] computes the set of possible weak Condorcet winners (called majority-optimal alternatives by [12]) when preferences are modeled by general CP-nets, by first eliminating many alternatives efficiently, and then determining the possible weak Condorcet winners among the remaining alternatives.

In this paper, we combine both research directions: we take an MLE approach to preference aggregation in multi-issue domains, when the voters' preferences are represented by (not necessarily acyclic) CP-nets. Considering the structure of CP-nets, we focus on probabilistic models that are *very weakly decomposable*. That is, given the "correct" winner, a voter's local preferences over an issue are independent from her local preferences over other issues, and as well as from her local preferences over the same issue given a different setting of (at least some of) the other issues.

After reviewing some background, we start with the general case in which the issues are not necessarily binary. The goal here is to investigate when issue-by-issue or sequential voting rules can be modeled as maximum likelihood estimators. When the input profile is separable, we completely characterize the set of all voting correspondences that can be modeled as an MLE for a noise model satisfying a weak decomposability (resp. strong decomposability) property. Then, when the input profile of CP-nets is consistent with a common order over issues, we prove that no sequential voting rule satisfying unanimity can be represented by an MLE, provided the noise model satisfies very weak decomposability. We show that this impossibility result no longer holds if the number of voters is bounded above by a constant.

Then, we move to the special case in which each issue has only two possible values. For such domains, we introduce *distance-based noise models*, in which the local distribution over any issue i under some setting of the other issues depends only on the Hamming distance from this setting to the restriction of the "correct" winner to the issues other than i. We characterize distance-based noise models axiomatically. Then we focus on *distance-based threshold noise models* in which there is a threshold such that if the distance is smaller than the threshold, then a fixed nonuniform local distribution is used, whereas if the distance is at least as large as the threshold, then a uniform local distribution is used. We show that when the threshold is one, it is NP-hard to compute the winner, but that when it is equal to the number of issues, the winner can be computed in polynomial time.

2. TECHNICAL BACKGROUND

Let $\mathcal{X}$ be a finite set of *alternatives* (or *candidates*). A *vote* V is a linear order on $\mathcal{X}$, *i.e.*, a transitive, antisymmetric, and total relation on $\mathcal{X}$. The set of all linear orders on $\mathcal{X}$ is denoted by $L(\mathcal{X})$. An n-voter profile P is a collection of n votes, that is, $P = (V_1, \ldots, V_n)$, where $V_j \in L(\mathcal{X})$ for every $j \leq n$. The set of all profiles on $\mathcal{X}$ is denoted by $P(\mathcal{X})$. A *(voting) rule* $r : P(\mathcal{X}) \to \mathcal{X}$ maps any profile to a single candidate (the winner). A *(voting) correspondence* $c : P(\mathcal{X}) \to 2^{\mathcal{X}}$ maps any profile to a subset of candidates. A *preference function* $f : P(\mathcal{X}) \to 2^{L(\mathcal{X})}$ maps any profile to a set of linear orders over $\mathcal{X}$.

2.1 Maximum likelihood approach to voting

In the maximum likelihood approach to voting, it is assumed that there is a correct winner $d \in \mathcal{X}$, and each vote V is drawn conditionally independently given d, according to a conditional probability distribution $\pi(V|d)$. The independence structure of the noise model is illustrated in Figure 1. The use of this independence structure is standard. Moreover, if conditional independence among votes is not required, then any voting rule can be represented by an MLE for some noise model [4], which trivializes the question.

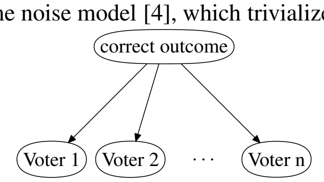


Figure 1: The noise model.

Under this independence assumption, the probability of a profile $P = (V_1, \ldots, V_n)$ given the correct winner d is $\pi(P|d) = \prod_{i=1}^{n} \pi(V_i|d)$. Then, the maximum likelihood estimate of the correct winner is $MLE_\pi(P) = \arg\max_{d \in \mathcal{X}} \pi(P|d)$.

MLE_π is a voting correspondence, as there may be several alternatives d that maximize $\pi(P|d)$. Another model that has been studied assumes that there is a correct *ranking* of the alternatives. Here, the model is defined similarly: given the correct linear order V^*, each vote V is drawn conditionally independently according to $\pi(V|V^*)$. The maximum likelihood estimate is defined as follows.

$$MLE_\pi(P) = \arg\max_{V^* \in L(\mathcal{X})} \prod_{V \in P} \pi(V|V^*)$$

In this paper, we require that all such conditional probabilities to be positive for technical reasons.

Definition 1 ([4]) *A voting rule (correspondence) r is a* maximum likelihood estimator for winners under i.i.d. votes (MLEWIV) *if there exists a noise model π such that for any profile P, we have that $MLE_\pi(P) = r(P)$.*

Definition 2 ([4]) *A preference function f is a* maximum likelihood estimator for rankings under i.i.d. votes (MLERIV) *if there exists a noise model π such that for any profile P, we have that $MLE_\pi(P) = f(P)$.*

Conitzer and Sandholm studied which common voting rules/preference functions are MLEWIVs/MLERIVs [4]. A *candidate scoring correspondence* c is a correspondence defined by a scoring function $s : L(\mathcal{X}) \times \mathcal{X} \to \mathbb{R}$ in the following way: for any profile P, $c(P) = \arg\max_{d \in \mathcal{X}} \sum_{V \in P} s(V, d)$.

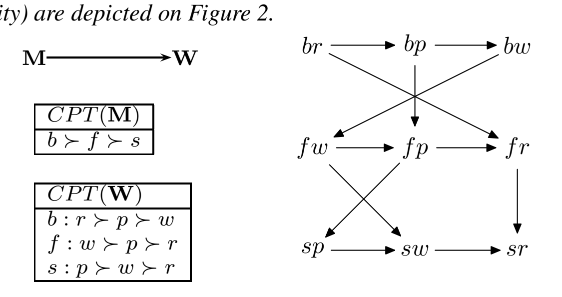
2.2 Voting in multi-issue domains

In this paper, the set of all alternatives $\mathcal{X}$ is a *multi-issue domain*. That is, let $\mathcal{I} = \{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ ($p \geq 2$) be a set of *issues*, where each issue $\mathbf{x}_i$ takes values in a finite *local domain* D_i. The set of alternatives is $\mathcal{X} = D_1 \times \ldots \times D_p$, that is, an alternative is uniquely identified by its values on all issues.[1] A multi-issue domain is *binary* if for every i we have $D_i = \{0_i, 1_i\}$. For any alternative $\vec{d} = (d_1, \ldots, d_p)$ and any issue $\mathbf{x}_i$, we let $\vec{d}|_{\mathbf{x}_i} = d_i$ and $\vec{d}_{-i} = (d_1, \ldots, d_{i-1}, d_{i+1}, \ldots, d_p)$. For any $I \subseteq \{1, \ldots, p\}$, we let $D_I = \prod_{i \in I} D_i$, and $D_{-i} = D_{\{1,\ldots,i-1,i+1,\ldots,p\}}$.

Example 1 *A group must make a joint decision on the dinner menu (the caterer can only serve the same menu to everyone). The menu is composed of two issues: main course (*$\mathbf{M}$*) and wine (*$\mathbf{W}$*). There are three choices for the main course: beef (b), fish (f), or salad (s); and three for the wine: red (r), white (w), or pink (p). The set of alternatives is a multi-issue domain:* $\mathcal{X} = \{b, f, s\} \times \{r, w, p\}$.

CP-nets [1] are a useful language for expressing preferences compactly over multi-issue domains. A CP-net $\mathcal{N}$ over $\mathcal{X}$ consists of two components: (a) a directed graph $G = (\mathcal{I}, E)$ and (b) a set of conditional linear preferences $\succeq^i_{\vec{u}}$ over D_i, for any $i \leq p$ and any setting $\vec{u}$ of the parents of $\mathbf{x}_i$ in G (denoted by $Par_G(\mathbf{x}_i)$). These conditional linear preferences $\succeq^i_{\vec{u}}$ over D_i form the *conditional preference table* for issue $\mathbf{x}_i$, denoted by $CPT(\mathbf{x}_i)$. When G is acyclic, $\mathcal{N}$ is said to be an *acyclic CP-net*. The set of all CP-nets over $\mathcal{X}$ is denoted by CPnet($\mathcal{X}$). A CP-net $\mathcal{N}$ induces the partial order $\succ_{\mathcal{N}}$, defined as the transitive closure of $\{(a_i, \vec{u}, \vec{z}) \succ (b_i, \vec{u}, \vec{z})) \mid i \leq p; \vec{u} \in D_{Par_G(\mathbf{x}_i)}; a_i, b_i \in D_i \text{ s.t. } a_i \succ^i_{\vec{u}} b_i; \vec{z} \in D_{-(Par_G(\mathbf{x}_i) \cup \{\mathbf{x}_i\})}\}$. It is known [1] that if $\mathcal{N}$ is acyclic, then $\succ_{\mathcal{N}}$ is transitive and asymmetric, that is, a strict partial order. (This is not necessarily the case if $\mathcal{N}$ is not acyclic.) For any graph G' on $\mathcal{I}$, a CP-net $\mathcal{N}$ is *compatible* with G' if its graph G is a subgraph of G'.

Example 2 *Let* $\mathcal{X}$ *be the multi-issue domain defined in Example 1. We define a CP-net* $\mathcal{N}$ *as follows:* $\mathbf{M}$ *is the parent of* $\mathbf{W}$*, and the CPTs consist of the following conditional preferences:* $CPT(\mathbf{M}) = \{b \succ f \succ s\}, CPT(\mathbf{W}) = \{b : r \succ p \succ w, f : w \succ p \succ r, s : p \succ w \succ r\}$*, where* $b : r \succ p \succ w$ *is interpreted as follows: "when* $\mathbf{M}$ *is b, then, r is the most preferred value for* $\mathbf{W}$*, p is the second most preferred value, and w is the least preferred value."* $\mathcal{N}$ *and its induced partial order* $\succeq_{\mathcal{N}}$ *(without edges implied by transitivity) are depicted on Figure 2.*



(a) CP-net $\mathcal{N}$. (b) The partial order induced by $\mathcal{N}$.

Figure 2: An acyclic CP-net $\mathcal{N}$ and its induced partial order.

When all issues are binary, a CP-net $\mathcal{N}$ can be visualized as a hypercube with directed edges in p-dimensional space, in the following way. Each vertex is an alternative, and any two adjacent vertices differ in only one component (issue). That is, for any $i \leq p$, and any $\vec{d}_{-i} \in D_{-i}$, there is a directed edge connecting $(0_i, \vec{d}_{-i})$ and $(1_i, \vec{d}_{-i})$, and the direction of the edge is from $(0_i, \vec{d}_{-i})$ to $(1_i, \vec{d}_{-i})$ if and only if $(0_i, \vec{d}_{-i}) \succ_{\mathcal{N}} (1_i, \vec{d}_{-i})$.

[1] In the following, we use vectors, such as $\vec{d}$, to denote alternatives.

Example 3 *Let* $p = 3$ *and let* $\mathcal{N}$ *be a CP-net defined as follows: the directed graph of* $\mathcal{N}$ *has an edge from* $\mathbf{x}_1$ *to* $\mathbf{x}_2$ *and an edge from* $\mathbf{x}_2$ *to* $\mathbf{x}_3$*; the CPTs are* $CPT(\mathbf{x}_1) = \{0_1 \succ 1_1\}$, $CPT(\mathbf{x}_2) = \{0_1 : 0_2 \succ 1_2, 1_1 : 1_2 \succ 0_2\}$, $CPT(\mathbf{x}_3) = \{0_2 : 0_3 \succ 1_3, 1_2 : 1_3 \succ 0_3\}$. $\mathcal{N}$ *is illustrated in Figure 3 (for simplicity, in the figure, a vertex abc represents the alternative* $a_1 b_2 c_3$*, for example, the vertex* 000 *represents the alternative* $0_1 0_2 0_3$*).*

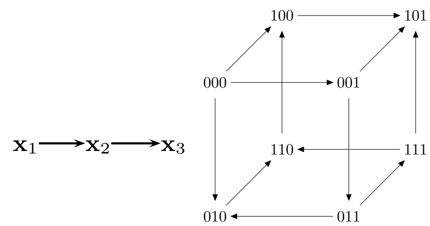


Figure 3: The hypercube representation of the CP-net.

A linear order V *extends* a CP-net $\mathcal{N}$, denoted by $V \sim \mathcal{N}$, if it extends $\succeq_{\mathcal{N}}$. For any setting $\vec{u}$ of $Par_G(\mathbf{x}_i)$, let $V|_{\mathbf{x}_i:\vec{u}}$ and $\mathcal{N}|_{\mathbf{x}_i:\vec{u}}$ denote the the restriction of V (or equivalently, $\mathcal{N}$) to $\mathbf{x}_i$, given $\vec{u}$. That is, $V|_{\mathbf{x}_i:\vec{u}}$ (or $\mathcal{N}|_{\mathbf{x}_i:\vec{u}}$) is the linear order $\succeq^i_{\vec{u}}$.

For any graph G on $\mathcal{I}$, V is *G-legal* if there exists a CP-net $\mathcal{N}$ such that $V \sim \mathcal{N}$ and $\mathcal{N}$ is compatible with G. We say V is *legal* if it is G-legal for some acyclic graph G. A profile is G-legal if all of its votes are G-legal. For any linear order $\mathcal{O}$ on $\mathcal{I}$, we let $G_{\mathcal{O}}$ be the *graph induced by* $\mathcal{O}$—that is, there is an edge $(\mathbf{x}_i, \mathbf{x}_j)$ in $G_{\mathcal{O}}$ if and only if $\mathbf{x}_i >_{\mathcal{O}} \mathbf{x}_j$. For any directed acyclic graph G, a linear order $\mathcal{O}$ can be found such that $G \subseteq G_{\mathcal{O}}$, which means that any G-legal profile is also $G_{\mathcal{O}}$-legal (which we abbreviate as $\mathcal{O}$-legal). For example, let $\mathcal{N}$ be the CP-net defined in Example 2. Any linear order over $\mathcal{X}$ that extends $\succeq_{\mathcal{N}}$ is $G_{(\mathbf{M}>\mathbf{W})}$-legal (or, equivalently, $(\mathbf{M} > \mathbf{W})$-legal). V is *separable* if and only if it extends a CP-net in which there is no edge. Therefore, any separable vote is $\mathcal{O}$-legal for any ordering $\mathcal{O}$ of issues. We emphasize that votes are not always required to be separable or legal in this paper.

In this paper, we fix $\mathcal{O}$ to be $\mathbf{x}_1 > \ldots > \mathbf{x}_p$. Given a collection of *local rules* $(r_1, \ldots, r_p)$ (where for any $i \leq p$, r_i is a voting rule on D_i), the *sequential composition* of $r_1, \ldots, r_p$ w.r.t. $\mathcal{O}$, denoted by $Seq(r_1, \ldots, r_p)$, is defined for all $\mathcal{O}$-legal profiles as follows: $Seq(r_1, \ldots, r_p)(P) = (d_1, \ldots, d_p) \in \mathcal{X}$, where for any $i \leq p$, $d_i = r_i(P|_{\mathbf{x}_i:d_1\ldots d_{i-1}})$. Thus, the winner is selected in p steps, one for each issue, in the following way: in step i, d_i is selected by applying the local rule r_i to the preferences of voters over D_i, conditioned on the values $d_1, \ldots, d_{i-1}$ that have already been determined for issues that precede $\mathbf{x}_i$. $Seq(r_1, \ldots, r_p)$ is well-defined, because for any G-legal profile, the set of winners is the same for all $\mathcal{O}'$ such that $G \subseteq G_{\mathcal{O}'}$ (see [10]). When G has no edges, $Seq(r_1, \ldots, r_p)$ becomes an *issue-by-issue* voting rule. *Sequential composition* of *local correspondences* $c_1, \ldots, c_p$, denoted by $Seq(c_1, \ldots, c_p)$ is defined in a similar way: for any $\mathcal{O}$-legal profile P, $\vec{d} \in Seq(c_1, \ldots, c_p)(P)$ if and only if for any $i \leq p$, we have that $d_i \in c_i(P|_{\mathbf{x}_i:d_1\ldots d_{i-1}})$.

We will focus on voting methods that only use information about voters' preferences that is represented in the CP-nets that those preferences extend. Therefore, we can consider an input profile to be composed of CP-nets instead of linear orders.

3. MULTI-ISSUE DOMAIN NOISE MODELS

In this section, we extend the maximum likelihood estimation approach to multi-issue domains (where $\mathcal{X} = D_1 \times \ldots \times D_p$). For now, we consider the case where there is a correct winner, $\vec{d} \in \mathcal{X}$. Votes are given by CP-nets and are conditionally independent, given $\vec{d}$. The probability of drawing CP-net $\mathcal{N}$ given that the correct winner is $\vec{d}$ is $\pi(\mathcal{N}|\vec{d})$, where π is some noise model. We note that π is a conditional probability distribution over all CP-nets (in contrast to all linear orders in previous studies). Given this noise model, for any profile of CP-nets $P = (\mathcal{N}_1, \dots, \mathcal{N}_n)$, the maximum likelihood estimate of the correct winner is

$$MLE_\pi(P) = \arg\max_{\vec{a}\in\mathcal{X}} \prod_{j=1}^n \pi(\mathcal{N}_j|\vec{a})$$

Again, MLE_π is a voting correspondence.

Even if for all i, $|D_i| = 2$, the number of CP-nets (including cyclic ones) is $2^{p2^{p-1}}$ (2 options for each entry of each CPT, and the CPT of any issue i has 2^{p-1} entries, one for each setting of the issues other than i). Hence, to specify a probability distribution over CP-nets, we will assume some structure in this distribution so that it can be compactly represented. Throughout the paper, we will assume that the local preferences for individual issues (given the setting of the other issues) are drawn conditionally independently, both across issues and across settings of the other issues, given the correct winner. More precisely:

Definition 3 *A noise model is* very weakly decomposable *if for every $\vec{d} \in \mathcal{X}$, every $i \leq p$, and every $\vec{a}_{-i} \in D_{-i}$, there is a probability distribution $\pi_{\vec{d}}^{\vec{a}_{-i}}$ over $L(D_i)$, so that for every $\vec{d} \in \mathcal{X}$ and every $\mathcal{N} \in$ CPnet($\mathcal{X}$), $\pi(\mathcal{N}|\vec{d}) = \prod_{i\leq p, \vec{a}_{-i}\in D_{-i}} \pi_{\vec{d}}^{\vec{a}_{-i}}(\mathcal{N}|_{\mathbf{x}_i:\vec{a}_{-i}})$*

For instance, if $D_i = \{0_i, 1_i, 2_i\}$, $\pi_{\vec{d}}^{\vec{a}_{-i}}(0_i \succ 2_i \succ 1_i)$ is the probability that the CP-net of a given voter contains $\vec{a}_{-i} : 0_i \succ 2_i \succ 1_i$, given that the correct winner is $\vec{d}$. Then, the probability of CP-net $\mathcal{N}$ is the product of the probabilities of all its local preferences $\mathcal{N}|_{\mathbf{x}_i:\vec{a}_{-i}}$ over specific $\mathbf{x}_i$ given specific $\vec{a}_{-i}$ (which contains the setting for $\mathbf{x}_i$'s parents as a sub-vector), when the correct winner is $\vec{d}$. (We will introduce stronger decomposability notions soon.)

Assuming very weak decomposability is reasonable in the sense that a voter's preferences for one issue are not directly linked to her *preferences* for another issue. We note that this is completely different from saying that the voter's preferences for an issue do not depend on the *values* of the other issues. Indeed, the voter's preferences for an issue can, at least in principle, change drastically depending on the values of the other issues. For instance, in Example 1, the event "the voter prefers white to pink to red wine when the main course is fish" is probabilistically independent (conditional on the correct outcome) of the event "the voter prefers beef to salad to fish when the wine is red."

However, we do not want to argue that such a distribution always generates realistic preferences. In fact, with some probability, such a distribution generates cyclic preferences. This is not a problem, in the sense that the purpose of the maximum likelihood approach is to find a natural voting rule that maps profiles to outcomes. The fact that this rule is also defined for cyclic preferences does not hinder its application to acyclic preferences. Similarly, Condorcet's original noise model for the single-issue setting also generates cyclic preferences with some probability, but this does not prevent us from applying the corresponding (Kemeny) rule [8] to acyclic preferences.

Even assuming very weak decomposability, we still need to define exponentially many probabilities. We will now introduce some successive strengthenings of the decomposability notion. First, we introduce *weak decomposability*, which removes the dependence of an issue's local distribution on the settings of the other issues *in the correct winner*.

Definition 4 *A very weakly decomposable noise model π is* weakly decomposable *if for any $i \leq p$, any $\vec{d}_1, \vec{d}_2 \in \mathcal{X}$ such that $\vec{d}_1|_{\mathbf{x}_i} = \vec{d}_2|_{\mathbf{x}_i}$, we must have that for any $\vec{a}_{-i} \in D_{-i}$, $\pi_{\vec{d}_1}^{\vec{a}_{-i}} = \pi_{\vec{d}_2}^{\vec{a}_{-i}}$. Let $WD(\mathcal{X})$ denote the set of correspondences that are the MLE for some weakly decomposable noise model.*

Next, we introduce an even stronger notion, namely *strong decomposability*, which removes all dependence of an issue's distribution on the settings of the other issues. That is, the local distribution only depends on the value of that issue in the correct winner.

Definition 5 *A very weakly decomposable noise model π is* strongly decomposable *if it is weakly decomposable, and for any $i \leq p$, any $\vec{a}_{-i}, \vec{b}_{-i} \in D_{-i}$, any $\vec{d} \in \mathcal{X}$, we must have that $\pi_{\vec{d}}^{\vec{a}_{-i}} = \pi_{\vec{d}}^{\vec{b}_{-i}}$. Let $SD(\mathcal{X})$ denote the set of correspondences that are the MLE for some strongly decomposable noise model.*

4. CHARACTERIZATIONS OF MLE CORRESPONDENCES

It seems that the MLE approaches are quite different from the voting rules that have previously been studied in the context of multi-issue domains, such as issue-by-issue voting and sequential voting. This may imply that the maximum likelihood approach can generate sensible new rules for multi-issue domains. Nevertheless, we may wonder whether previously studied rules also fit under the MLE framework.

In this section, we study whether or not issue-by-issue and sequential voting correspondences can be modeled as the MLEs for very weakly decomposable noise models. We note that voting rules (which always output a unique winner) are a special case of voting correspondences. Therefore, our results easily extend to the case of voting rules. First, we restrict the domain to separable profiles (that is, all votes in the input profile are separable), and characterize the set of all correspondences that can be modeled as the MLEs for strongly/weakly decomposable noise models. **Most of the proofs of the theorems are omitted due to the space constraint.**

Theorem 1 *Over the domain of separable profiles, a voting correspondence c can be modeled as the MLE for a strongly decomposable noise model if and only if c is an issue-by-issue voting correspondence composed of MLEWIVs.*

Theorem 2 *Over the domain of separable profiles, a voting correspondence c can be modeled as the MLE for a weakly decomposable noise model if and only if c is an issue-by-issue voting correspondence composed of candidate scoring correspondences.*

However, for sequential voting correspondences, we have the following negative result. A voting correspondence c satisfies *unanimity* if for any profile P in which each vote ranks an alternative $\vec{d}$ first, we have $r(P) = \{\vec{d}\}$.

Theorem 3 *Let $Seq(c_1, \dots, c_p)$ be a sequential voting correspondence that satisfies unanimity. Over the domain of $\mathcal{O}$-legal profiles, there is no very weakly decomposable noise model such that $Seq(c_1, \dots, c_p)$ is the MLE.*

This theorem tells us that even assuming the weakest conditional independence of the noise model, the voting correspondence defined by the MLE of that noise model is different from any sequential voting correspondence satisfying unanimity. This suggests that the MLE approach gives us new voting rules/correspondences.

However, a connection between MLEs for very weakly decomposable noise models and sequential voting correspondences can be obtained if there is an upper bound on the number of voters. The next theorem states that for any natural number n and any sequential composition of MLEWIVs, there exists a very weakly decomposable noise model such that for any profile of no more than n $\mathcal{O}$-legal votes, the set of winners under the MLE for that noise model is always a subset of the set of winners under the sequential correspondence. That is, if the local correspondences can be justified by a noise model, then, to some extent, so can the sequential voting correspondence that uses these local rules.

Theorem 4 *For any $n \in \mathbb{N}$ and any sequential voting correspondence $Seq(c_1, \ldots, c_p)$ where for each $i \leq p$, c_i is an MLEWIV, there exists a very weakly decomposable noise model π such that for any $\mathcal{O}$-legal profile P composed of no more than n votes, we have that $MLE_\pi(P) \subseteq Seq(c_1, \ldots, c_p)(P)$.*

Proof of Theorem 4: Let r_i be the MLEWIV with the conditional probability distribution $Pr_i(V^i|d_i)$, where $V^i \in L(D_i)$, $d_i \in D_i$. For any $i \leq p$, we let $R^{i,n}_{max} = \max_{P_i, P'_i, d_i, d'_i} \left\{ \frac{Pr_i(P_i|d_i)}{Pr_i(P'_i|d'_i)} \right\}$, where $d_i, d'_i \in D_i$, and P_i and P'_i are profiles with the same number (but no more than n) of linear orders over D_i. We let $R^{i,n}_{min} = 1$ if r_i is the trivial correspondence that always outputs the whole domain; and $R^{i,n}_{min} = \min_{P_i, \vec{d}_i, d'_i} \left\{ \frac{Pr_i(P_i|d_i)}{Pr_i(P_i|d'_i)} : \frac{Pr_i(P_i|d_i)}{Pr_i(P_i|d'_i)} > 1 \right\}$, where $d_i, d'_i \in D_i$, and P_i is a profile of no more than n linear orders over D_i. We note that for any $i \leq p$, any $n \in \mathbb{N}$, we have that $R^{i,n}_{max} \geq R^{i,n}_{min} \geq 1$.

For any $V^i \in L(D_i)$, any $\vec{d} \in \mathcal{X}$, and any $\vec{a}_{-i} \in D_{-1}$, we let

$$\pi^{\vec{a}_{-i}}_{\vec{d}}(V^i) = \begin{cases} Pr_i(V^i|d_i)^{k_i}/Z_i & \text{if } \vec{a}_{-i} = \vec{d}_{-i} \\ \frac{1}{|D_i|!} & \text{otherwise} \end{cases},$$

where $Z_i = \sum_{V^i \in L(D_i)} Pr_i(V^i|d_i)^{k_i}$ is a normalizing factor, and $1 = k_1 > k_2 > \ldots > k_p > 0$ are chosen in the following way: for any $i' < i \leq p$, any $V^i, W^i \in L(D_i)$, and any $d_i, d'_i \in D_i$, if $R^{i,n}_{min} > 1$, then we must have that $(R^{i,n}_{max})^{k_i} < (R^{i',n}_{min})^{k_{i'}/2^{i-i'}}$.

We next prove that for any profile P_{CP} of no more than n CP-nets, we must have that $MLE_\pi(P_{CP}) \subseteq Seq(r_1, \ldots, r_p)(P_{CP})$. For the sake of contradiction, let P_{CP} be a profile of no more than n CP-nets with $MLE_\pi(P_{CP}) \nsubseteq Seq(r_1, \ldots, r_p)(P_{CP})$. Let $\vec{d} \in MLE_\pi(P_{CP})$, and i^* be the number such that there exists $\vec{d}^* \in Seq(r_1, \ldots, r_p)(P_{CP})$ such that for all $i' < i^*$, $d_{i'} = d^*_{i'}$, and $d_{i^*} \notin r_{i^*}(P_{CP}|_{\mathbf{x}_{i^*}:d_1 \ldots d_{i^*-1}})$. Because $r_{i^*}(P_{CP}|_{\mathbf{x}_{i^*}:d_1 \ldots d_{i^*-1}}) \neq D_{i^*}$, we must have that $R^{i^*,n}_{min} > 1$. Because $\vec{d} \in MLE_\pi(P_{CP})$, we must have that $\frac{\pi(P_{CP}|\vec{d})}{\pi(P_{CP}|\vec{d}^*)} \geq 1$. However, we have the following calculation that leads to a contradiction.

$$\begin{aligned} 1 \leq & \frac{\pi(P_{CP}|\vec{d})}{\pi(P_{CP}|\vec{d}^*)} = \frac{\prod_{i=1}^p Pr_i(P_{CP}|_{\mathbf{x}_i:d_1 \ldots d_{i-1}}|d_i)}{\prod_{i=1}^p Pr_i(P_{CP}|_{\mathbf{x}_i:d^*_1 \ldots d^*_{i-1}}|d^*_i)} \\ = & \frac{\prod_{i=i^*}^p Pr_i(P_{CP}|_{\mathbf{x}_i:d_1 \ldots d_{i-1}}|d_i)}{\prod_{i=i^*}^p Pr_i(P_{CP}|_{\mathbf{x}_i:d^*_1 \ldots d^*_{i-1}}|d^*_i)} \\ \leq & \frac{1}{(R^{i^*,n}_{min})^{k_{i^*}}} \cdot \prod_{i=i^*+1}^p (R^{i,n}_{max})^{k_i} \\ < & \frac{1}{(R^{i^*,n}_{min})^{k_{i^*}}} \cdot \prod_{i=i^*+1}^p (R^{i^*,n}_{min})^{k_{i^*}/2^{i-i^*}} < 1 \end{aligned}$$

Therefore, we must have that $MLE_\pi(P) \subseteq Seq(r_1, \ldots, r_p)(P)$ for all profiles P that consist of no more than n CP-nets. □

5. DISTANCE-BASED MODELS

We have shown in the previous section that the MLE approach may give us new voting rules in multi-issue domains. However, assuming very weak decomposability, there are too many (exponentially many) parameters in the noise model, which makes it very hard to implement a rule based on the MLE approach. In this section, we focus on a family of maximum likelihood estimators that are based on noise models defined over binary multi-issue domains (domains composed of binary issues), and that need only a few parameters to be specified. We recall that a CP-net on a binary multi-issue domain corresponds to a directed hypercube in which each edge has a direction representing the local preference. A very weakly decomposable noise model π can be represented by a collection of weighted directed hypercubes, one for each correct winner, in which the weight of each directed edge is the probability of the local preference represented by the directed edge. For any outcome $\vec{d} \in \mathcal{X}$, any issue $\mathbf{x}_i$, any $\vec{e}_{-i} \in D_{-i}$, and any $d_i \neq d'_i \in D_i$, the weight on the directed edge $((\vec{e}_{-i}, d_i), (\vec{e}_{-i}, d'_i))$ of the weighted hypercube corresponding to the correct winner $\vec{d}$ is denoted by $\pi^{\vec{e}_{-i}}_{\vec{d}}(d_i \succ d'_i)$, and represents the probability that a given voter reports the preference $\vec{e}_{-i} : d_i \succ d'_i$ in her CP-net, given that the correct winner is $\vec{d}$.[2] For example, when the correct winner is $0_1 0_2 0_3$, the weight on the directed edge $(0_1 1_2 0_3, 0_1 1_2 1_3)$ is the probability $\pi^{0_1 1_2}_{0_1 0_2 0_3}(0_3 \succ 1_3)$. We now propose and study very weakly decomposable noise models in which the weight of each edge depends only on the Hamming distance between the edge and the correct winner.

For any pair of alternatives $\vec{d}, \vec{d}' \in \mathcal{X}$, the *Hamming distance* between $\vec{d}$ and $\vec{d}'$, denoted by $|\vec{d} - \vec{d}'|$, is the number of components in which $\vec{d}$ is different from $\vec{d}'$, that is, $|\vec{d} - \vec{d}'| = \#\{i \leq p : d_i \neq d'_i\}$. Let $e = (\vec{d}_1, \vec{d}_2)$ be a pair of alternatives such that $|\vec{d}_1 - \vec{d}_2| = 1$ (equivalently, an edge in the hypercube). The distance between e and an alternative $\vec{d} \in \mathcal{X}$, denoted by $|e - \vec{d}|$, is the smaller Hamming distance between $\vec{d}$ and the two ends of e, that is, $|e - \vec{d}| = \min\{|\vec{d}_1 - \vec{d}|, |\vec{d}_2 - \vec{d}|\}$. For example, $|0_1 1_2 0_3 - 0_1 0_2 0_3| = 1$, $|0_1 1_2 1_3 - 0_1 0_2 0_3| = 2$, and $|(0_1 1_2 0_3, 0_1 1_2 1_3) - 0_1 0_2 0_3| = 1$.

We next introduce *distance-based noise models* in which the probability distribution $\pi^{\vec{a}_{-i}}_{\vec{d}}$ only depends on d_i and the Hamming distance between $\vec{a}_{-i}$ and $\vec{d}_{-i}$.

Definition 6 *Let $\mathcal{X}$ be a binary multi-issue domain. For any $\vec{q} = (q_0, \ldots, q_{p-1})$ such that $1 > q_0, \ldots, q_{p-1} > 0$, a* distance-based (noise) model *$\pi_{\vec{q}}$ is a very weakly decomposable noise model such that for any $\vec{d} \in \mathcal{X}$, any $i \leq p$, and any $\vec{a}_{-i} \in D_{-i}$ with $|\vec{a}_{-i} - \vec{d}_{-i}| = k \leq p-1$, we have that $\pi^{\vec{a}_{-i}}_{\vec{d}}(d_i \succ \bar{d}_i) = q_k$.*

The intuition behind the notion of a distance-based model is as follows. First, it is plausible to assume that the "closer" two alternatives are to the correct alternative, the more likely a given voter will order them in the "correct" way, that is, will prefer the one which is closer to the correct alternative. The family of distance-based voting rules is actually more general than this, because we do not impose $q_1 \geq \ldots \geq q_{p-1}$, but we may of course add this restriction if we wish to. Moreover, the choice of the Hamming distance is not necessary, and other intuitive distance-based models can be defined, using other distances – for instance, domain-dependent distances. But, the Hamming distance is a natural starting point (most works in distance-based belief base merging and distance-based belief revision also focus on the Hamming distance).

Given the correct winner $\vec{d}$, a distance-based model $\pi_{\vec{q}}$ can be visualized by the following weighted directed graph built on the hypercube:

• For any undirected edge $e = (\vec{d}_1, \vec{d}_2)$ in the hypercube, where $\vec{d}_1, \vec{d}_2$ differ only on the value assigned to $\mathbf{x}_i$ for some $i \leq p$, if $\vec{d}_1|_{\mathbf{x}_i} = d_i$, then the direction of e is from $\vec{d}_1$ to $\vec{d}_2$; if $\vec{d}_2|_{\mathbf{x}_i} = d_i$, then the direction of e is from $\vec{d}_2$ to $\vec{d}_1$. That is, the direction of the edge is always from the alternative whose $\mathbf{x}_i$ component is the

[2] For every pair of alternatives differing on exactly one issue, there is exactly one weighted edge between them; the direction of the edge only says that we are going further from the correct winner. This will be made more precise after Definition 6.

same as the $\mathbf{x}_i$ component of the correct winner to the other end of the edge.

- For any edge e with $|e - \vec{d}| = l$, the weight of e is q_l.

For example, given that $0_1 0_2 0_3$ is the correct winner, the distance-based model is illustrated in Figure 4.

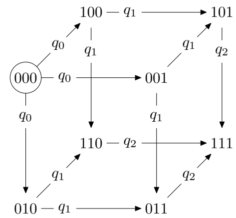


Figure 4: The distance-based model $\pi_{(q_0,q_1,q_2)}$ when the correct winner is 000.

To characterize distance-based models, we first define *inter-issue permutations*. Intuitively, an inter-issue permutation is a permutation that exchanges two issues.

Definition 7 *An* inter-issue permutation *is a permutation $m^+_{i,j}$ or $m^-_{i,j}$ on $D_1 \cup \ldots \cup D_p$, for some $1 \leq i \neq j \leq p$, defined as follows: (1) for any $k \neq i, j$ and any $d_k \in D_k$, $m^+_{i,j}(d_k) = m^-_{i,j}(d_k) = d_k$; (2) $m^+_{i,j}$ exchanges 0_i and 0_j, and exchanges 1_i and 1_j; $m^-_{i,j}$ exchanges 0_i and 1_j, and exchanges 1_i and 0_j. Note that $(m^+_{i,j})^{-1} = m^+_{i,j}$ and $(m^-_{i,j})^{-1} = m^-_{i,j}$.*

Any $m_{i,j} \in \{m^+_{i,j}, m^-_{i,j}\}$ induces a permutation $M_{i,j}$ on the set of all sub-vectors of any $\vec{d} \in \mathcal{X}$ as follows: for any $I \subseteq \mathcal{I}$ and $\vec{d}_I = (d_{i_1}, \ldots, d_{i_{|I|}}) \in D_I$, $M_{i,j}(d_I) = (m_{i,j}(d_{i_1}), \ldots, m_{i,j}(d_{i_{|I|}}))$. For example, let $p = 3$, and $m_{1,2}$ be an inter-issue permutation such that $m_{1,2}(0_1) = 1_2$. Then we have $M_{1,2}(1_1) = 0_2$, $M_{1,2}(0_2) = 1_1$, $M_{1,2}(1_2) = 0_1$; $M_{1,2}(0_1 0_2 0_3) = 1_1 1_2 0_3$, $M_{1,2}(1_1 1_3) = 0_2 1_3$.

We note that since each issue is binary, there are exactly two ways of exchanging issue $\mathbf{x}_i$ and $\mathbf{x}_j$: either map 0_i to 0_j (and 1_i to 1_j), or map 0_i to 1_j (and 1_i to 0_j).

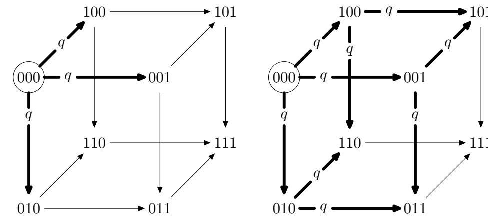
Definition 8 *A very weakly decomposable noise model π satisfies* inter-issue neutrality *if for any $i, j \leq p$, any inter-issue permutation $m_{i,j}$ (which induces $M_{i,j}$), any $i' \leq p$, any $\vec{d} \in \mathcal{X}$, and any $\vec{a}_{-i'} \in D_{-i'}$, we have that $\pi^{\vec{a}_{-i'}}_{\vec{d}}(0_{i'} \succ 1_{i'}) = \pi^{M_{i,j}(\vec{a}_{-i'})}_{M_{i,j}(\vec{d})}(m_{i,j}(0_{i'}) \succ m_{i,j}(1_{i'}))$.*

Thus, the noise model π satisfies inter-issue neutrality if after exchanging any two issues, the resulting noise model is still π. Or equivalently, π is indifferent to the names of the issues as well as the names of the values they take. We next show that the class of distance-based models can be completely characterized as the class of noise models that satisfy very weak decomposability and inter-issue neutrality.

Theorem 5 *Let $\mathcal{X}$ be a binary multi-issue domain. A very weakly decomposable noise model π is a distance-based noise model if and only if it satisfies inter-issue neutrality.*

We are especially interested in a special type of distance-based models in which there exists a threshold $1 \leq k \leq p$ and $q > \frac{1}{2}$, such that for any $i < k$, we have that $q_i = q$, and for any $k \leq i \leq p - 1$, we have that $q_i = \frac{1}{2}$. Such a model is denoted by $\pi_{k,q}$. We call $\pi_{k,q}$ a *distance-based threshold noise model* with threshold k. We say that a noise model π has threshold $k \leq p$ if and only if there exists $q > \frac{1}{2}$ such that $\pi = \pi_{k,q}$. The MLE for a distance-based threshold model $\pi_{k,q}$ is denoted by $MLE_{\pi_{k,q}}$.

Example 4 *Let $p = 3$. $\pi_{1,q}$ and $\pi_{2,q}$ are illustrated in Figure 5 (when the correct winner is 000).*



(a) The threshold is 1. (b) The threshold is 2.

Figure 5: Distance-based threshold models. The weight of the bold edges is $q > \frac{1}{2}$; the weight of all other edges is $\frac{1}{2}$.

The following theorem provides an axiomatic characterization of the set of all noise models that have threshold p, which is the number of issues. This axiomatization is similar to the one in Theorem 5.

Theorem 6 *Let $\mathcal{X}$ be a binary multi-issue domain. A noise model π is a distance-based threshold noise model with threshold p iff π is strongly decomposable and satisfies inter-issue neutrality.*

We next present a direct method for computing winners under the MLE correspondences of distance-based threshold models. For any $1 \leq k \leq p$, any $\vec{d} \in \mathcal{X}$, and any CP-net $\mathcal{N}$, we define the *consistency of degree* k between $\vec{d}$ and $\mathcal{N}$, denoted by $N_k(\vec{d}, \mathcal{N})$, as follows. $N_k(\vec{d}, \mathcal{N})$ is the number of triples $(\vec{a}, \vec{b}, i)$ such that $\vec{a}_{-i} = \vec{b}_{-i}$, $a_i = d_i$, $b_i = \bar{d}_i$, $|(a_i, b_i) - \vec{d}| \leq k - 1$, and $\mathcal{N}$ contains $a_{-i} : d_i \succ \bar{d}_i$. That is, $N_k(\vec{d}, \mathcal{N})$ is the number of local preferences (over any issue $\mathbf{x}_i$, given any $\vec{a}_{-i} \in D_{-i}$) in $\mathcal{N}$ that are $d_i \succ \bar{d}_i$, where the distance between $\vec{d}$ and the edge $((d_i, \vec{a}_{-i}), (\bar{d}_i, \vec{a}_{-i}))$ is at most $k - 1$. For any profile P_{CP} of CP-nets, we let $N_k(\vec{d}, P_{CP}) = \sum_{\mathcal{N} \in P_{CP}} N_k(\vec{d}, \mathcal{N})$.

Theorem 7 *For any $k \leq p$, any $q > \frac{1}{2}$, and any profile P_{CP} of CP-nets, we have that $MLE_{\pi_{k,q}}(P_{CP}) = \arg\max_{\vec{d}} N_k(\vec{d}, P_{CP})$.*

That is, the winner for any profile of CP-nets under any MLE for a distance-based threshold model $\pi_{k,q}$ maximizes the sum of the consistencies of degree k between the winning alternative and all CP-nets in the profile.

Proof of Theorem 7: For any $k \leq p$, any $\vec{d} \in \mathcal{X}$, we let $L_k = \#\{e : |e - \vec{d}| \leq k - 1\}$. That is, L_k is the number of edges in the hypercube whose distance from a given alternative $\vec{d}$ is no more than $k - 1$. For any $\vec{d} \in \mathcal{X}$ and any CP-net $\mathcal{N}$, we have that

$$
\begin{aligned}
&\ln \pi(P_{CP}|\vec{d}) \\
=& \sum_{\mathcal{N} \in P_{CP}} \ln \prod_{i, \vec{a}_{-i} \in D_{-i}} \pi^{\vec{a}_{-i}}_{d_i}(\mathcal{N}|_{\mathbf{x}_i : \vec{a}_{-i}}) \\
=& \sum_{\mathcal{N} \in P_{CP}} (N_k(\vec{d}, \mathcal{N}) \ln q + (L_k - N_k(\vec{d}, \mathcal{N})) \ln(1 - q)) \\
=& \sum_{\mathcal{N} \in P_{CP}} (N_k(\vec{d}, \mathcal{N}) \ln \frac{q}{1 - q} + L_k \ln(1 - q))
\end{aligned}
$$

Therefore, $MLE_{\pi_{k,q}}(P_{CP}) = \arg\max_{\vec{d}} \pi(P_{CP}|\vec{d})$
$= \arg\max_{\vec{d}} \sum_{\mathcal{N} \in P_{CP}} (N_k(\vec{d}, \mathcal{N}) \ln \frac{q}{1-q} + L_k \ln(1 - q))$
$= \arg\max_{\vec{d}} N_k(\vec{d}, P_{CP})$. □

Therefore, we have the following corollary, which states that the winners for any profile under $MLE_{\pi_{k,q}}$ do not depend on q, provided that $q > \frac{1}{2}$.

Corollary 1 *For any $k \leq p$, any $q_1 > \frac{1}{2}, q_2 > \frac{1}{2}$, and any profile P_{CP} of CP-nets, we have $MLE_{\pi_{k,q_1}}(P_{CP}) = MLE_{\pi_{k,q_2}}(P_{CP})$.*

Example 5 *Consider two binary issues $\mathbf{x}_1, \mathbf{x}_2$, and three voters, who report the following CP-nets:*

• $\mathcal{N}_1$ *has an edge from* $\mathbf{x}_1$ *to* $\mathbf{x}_2$*, and the following local preferences:* $\{0_1 \succ 1_1, 0_1 : 0_2 \succ 1_2, 1_1 : 1_2 \succ 0_2\}$.

• $\mathcal{N}_2$ *has an edge from* $\mathbf{x}_1$ *to* $\mathbf{x}_2$ *and an edge from* $\mathbf{x}_2$ *to* $\mathbf{x}_1$*, and the following local preferences:* $\{0_2 : 1_1 \succ 0_1, 1_2 : 0_1 \succ 1_1, 0_1 : 1_2 \succ 0_2, 1_1 : 0_2 \succ 1_2\}$.

• $\mathcal{N}_3$ *has no edge, and the following local preferences:* $\{1_1 \succ 0_1, 1_2 \succ 0_2\}$.

Let $P_{CP} = (\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3)$.

First, consider $k = 1$*. Let us compute* $N_1(1_1 1_2, \mathcal{N}_1)$*. There are two edges whose distance to* $1_1 1_2$ *is 0: one from* $1_1 1_2$ *to* $1_1 0_2$ *and one from* $1_1 1_2$ *to* $0_1 1_2$*. The first one is in the preference relation induced from* $\mathcal{N}_1$*; the second one is not. Therefore,* $N_1(1_1 1_2, \mathcal{N}_1) = 1$*. Similarly, we get* $N_1(1_1 1_2, \mathcal{N}_2) = 0$ *and* $N_1(1_1 1_2, \mathcal{N}_3) = 2$*, henceforth,* $N_1(1_1 1_2, P_{CP}) = 3$*. Similar calculations lead to* $N_1(1_1 0_2, P_{CP}) = 3$*,* $N_1(0_1 1_2, P_{CP}) = 4$ *and* $N_1(0_1 0_2, P_{CP}) = 2$*, hence* $MLE_{\pi_{1,q}}(P_{CP}) = \{0_1 1_2\}$ *(for any value of* $q > \frac{1}{2}$*).*

Now, consider $k = 2$*. Let us compute* $N_1(1_1 1_2, \mathcal{N}_1)$*. Now, we have to consider all four edges, since all of them are at a distance* 0 *or* 1 *to* $1_1 1_2$*. The two edges not considered for the case* $k = 1$ *are the edge from* $0_1 1_2$ *to* $0_1 0_2$ *and one from* $1_1 0_2$ *to* $0_1 1_2$*. In both cases, voter 1 prefers the alternative which is further from* $1_1 1_2$*, therefore,* $N_2(1_1 1_2, \mathcal{N}_1) = 1$*. Similarly, we get* $N_2(1_1 1_2, \mathcal{N}_2) = 2$ *and* $N_2(1_1 1_2, \mathcal{N}_3) = 4$*, henceforth,* $N_2(1_1 1_2, P_{CP}) = 7$*. Similar calculations lead to* $N_2(1_1 0_2, P_{CP}) = 5$*,* $N_2(0_1 1_2, P_{CP}) = 7$ *and* $N_1(0_1 0_2, P_{CP}) = 5$*, hence* $MLE_{\pi_{2,q}}(P_{CP}) = \{0_1 1_2, 1_1 1_2\}$*.*

We next investigate the computational complexity of applying MLE rules with distance-based threshold models. First, we present a polynomial-time algorithm that computes the winners and outputs the winners in a compact way, under $MLE_{\pi_{p,q}}$, where p is the number of issues. This algorithm computes the correct value(s) of each issue separately: for any issue $\mathbf{x}_i$, the algorithm counts the number of tuples $(\vec{a}_{-i}, \mathcal{N})$, where $\vec{a}_{-i} \in D_{-i}$ and $\mathcal{N}$ is a CP-net in the input profile P_{CP}, such that $\mathcal{N}$ contains $a_{-i} : 0_i \succ 1_i$. If there are more tuples $(\vec{a}_{-i}, \mathcal{N})$ in which $\mathcal{N}$ contains $a_{-i} : 0_i \succ 1_i$ than there are tuples in which $\mathcal{N}$ contains $a_{-i} : 1_i \succ 0_i$, then we select 0_i to be the ith component of the winning alternative, and vice versa. We note that the time required to count tuples $(\vec{a}_{-i}, \mathcal{N})$ depends on the size of $\mathcal{N}$. Therefore, even though computing the value for $\mathbf{x}_i$ takes time that is exponential in $|Par_G(\mathbf{x}_i)|$ (the number of parents of $\mathbf{x}_i$ in the directed graph of $\mathcal{N}$), the CPT of $\mathbf{x}_i$ in $\mathcal{N}$ itself is also exponential in $|Par_G(\mathbf{x}_i)|$ (for each setting of $Par_G(\mathbf{x}_i)$, there is an entry in $CPT(\mathbf{x}_i)$). This explains why the algorithm runs in polynomial time.

Algorithm 1 INPUT: $p \in \mathbb{N}$, $\frac{1}{2} < q < 1$, *and a profile of CP-nets* P_{CP} *over a binary domain consisting of p issues.*

1. *For each* $i \le p$*:*

1a. *Let* $S_i = 0$*,* $W_i = \emptyset$*.*

1b. *For each CP-net* $\mathcal{N} \in P_{CP}$*: let* $Par_G(\mathbf{x}_i) = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{p'}}\}$ *be the parents of* $\mathbf{x}_i$ *in the directed graph of* $\mathcal{N}$*. Let l be the number of settings* $\vec{y}$ *of* $Par_G(\mathbf{x}_i)$ *for which* $\mathcal{N}|_{\mathbf{x}_i:\vec{y}} = 0_i \succ 1_i$*. Let* $S_i \leftarrow S_i + l2^{p-p'} - 2^{p-1}$*. Here, p′ is the number of parents of* $\mathbf{x}_i$*, and* $l2^{p-p'} - 2^{p-1}$ *is the number of edges in the CP-net where* $0_i \succ 1_i$*, minus the number of edges where* $1_i \succ 0_i$*.*

1c. *At this point, let* $W_i = \begin{cases} \{0_i\} & \text{if } S_i > 0 \\ \{1_i\} & \text{if } S_i < 0 \\ \{0_i, 1_i\} & \text{if } S_i = 0 \end{cases}$

2.*Output* $W_1 \times \ldots \times W_p$*.*

Proposition 1 *The output of Algorithm 1 is* $MLE_{\pi_{p,q}}(P_{CP})$*, and the algorithm runs in polynomial time.*

The next example shows how to compute the winners under $MLE_{\pi_{p,q}}$ for the profile defined in Example 5.

Example 5, continued *Let us first compute* S_1*. In* $\mathcal{N}_1$ *(respectively,* $\mathcal{N}_1$ *and* $\mathcal{N}_3$*), the table for* x_1 *contributes to 2 edges (respectively, one edge and no edge) from* 0_1 *to* 1_1*, and to no edge (respectively, one edge and two edge) from* 1_1 *to* 0_1*, therefore* $S_1 = (+2) + 0 + (-2) = 0$*. Similarly,* $S_2 = 0 + 0 + (-2) = -2$*. Therefore,* $W_1 = \{0_1, 1_1\}$ *and* $W_2 = \{1_2\}$*, which gives us* $MLE_{\pi_{2,q}}(P_{CP}) = \{0_1 1_2, 1_1 1_2\}$*.*

However, when the threshold is one, computing the winners is NP-hard, and the associated decision problem, namely checking whether there exists an alternative $\vec{d}$ such that $N_1(\vec{d}, P_{CP}) \ge T$, is NP-complete.

Theorem 8 *It is NP-hard to find a winner under* $MLE_{\pi_{1,q}}$*. More precisely, it is NP-complete to decide whether there exists an alternative* $\vec{d}$ *such that* $N_1(\vec{d}, P_{CP}) \ge T$*.*

Proof of Theorem 8: By Theorem 7, the decision problem of finding a winner under $MLE_{\pi_{1,q}}$ is the following: for any profile P that consists of n CP-nets, and any $T \le pn$, we are asked whether or not there exists $\vec{d} \in \mathcal{X}$ such that $N_1(\vec{d}, P) \ge T$.

We prove the NP-hardness by reduction from the decision problem of MAX2SAT. The inputs of an instance of the decision problem of MAX2SAT consists of (1) a set of t atomic propositions $x_1, \ldots, x_t$; (2) a formula $F = c_1 \wedge \ldots \wedge c_m$ represented in *conjunctive normal form*, in which for any $i \le m$, $c_i = l_{i_1} \vee l_{i_2}$, and there exists $j_1, j_2 \le t$ such that l_{i_1} is x_{j_1} or $\neg x_{j_1}$, and l_{i_2} is x_{j_2} or $\neg x_{j_2}$; (3) $T \le m$. We are asked whether or not there exists a valuation $\vec{x}$ for the atomic propositions $x_1, \ldots, x_t$ such that at least T clauses are satisfied under $\vec{x}$.

Given any instance of MAX2SAT, we construct a decision problem instance of computing a winner under $MLE_{\pi_{1,q}}$ as follows.

• Let $\mathcal{X}$ be composed of t issues $\mathbf{x}_1, \ldots, \mathbf{x}_t$.

• Let $T' = 16T - 12m$.

• For any $i \le m$, we let v_{i_1} be the valuation of x_{i_1} under which l_{i_1} is true; let v_{i_2} be the valuation of x_{i_2} under which l_{i_2} is true. For any $j \le t$, we let 0_j corresponds to $\mathbf{x}_j$ being false, and 1_j corresponds to $\mathbf{x}_j$ being true. Then, any valuation of the atomic propositions is uniquely identified by an alternative. We next define six CP-nets as follows:

– $\mathcal{N}_{i,1}$: the DAG of $\mathcal{N}_{i,1}$ has only one directed edge $(\mathbf{x}_{i_1}, \mathbf{x}_{i_2})$. In $\mathcal{N}_{i,1}$, $v_{i_1} \succ \bar{v}_{i_1}$, $v_{i_1} : v_{i_2} \succ \bar{v}_{i_2}$, $\bar{v}_{i_1} : v_{i_2} \succ \bar{v}_{i_2}$, and for any $j \ne i_1$ and $j \ne i_2$, we have that $0_j \succ 1_j$.

– $\mathcal{N}_{i,2}$: the DAG of $\mathcal{N}_{i,2}$ has only one directed edge $(\mathbf{x}_{i_1}, \mathbf{x}_{i_2})$. In $\mathcal{N}_{i,2}$, $v_{i_1} \succ \bar{v}_{i_1}$, $v_{i_1} : \bar{v}_{i_2} \succ v_{i_2}$, $\bar{v}_{i_1} : v_{i_2} \succ \bar{v}_{i_2}$, and for any $j \ne i_1$ and $j \ne i_2$, we have that $0_j \succ 1_j$.

– $\mathcal{N}_{i,3}$: the DAG of $\mathcal{N}_{i,3}$ has only one directed edge $(\mathbf{x}_{i_2}, \mathbf{x}_{i_1})$. In $\mathcal{N}_{i,1}$, $v_{i_2} \succ \bar{v}_{i_2}$, $v_{i_2} : \bar{v}_{i_1} \succ v_{i_1}$, $\bar{v}_{i_2} : v_{i_1} \succ \bar{v}_{i_1}$, and for any $j \ne i_1$ and $j \ne i_2$, we have that $0_j \succ 1_j$.

We next obtain $\mathcal{N}'_{i,1}$, $\mathcal{N}'_{i,2}$, and $\mathcal{N}'_{i,3}$ from $\mathcal{N}_{i,1}$, $\mathcal{N}_{i,2}$, and $\mathcal{N}_{i,3}$, respectively, by letting $1_j \succ 0_j$ for any j with $j \ne i_1$ and $j \ne i_2$. Let $\vec{\mathcal{N}_i} = (\mathcal{N}_{i,1}, \mathcal{N}'_{i,1}, \mathcal{N}_{i,2}, \mathcal{N}'_{i,2}, \mathcal{N}_{i,3}, \mathcal{N}'_{i,3})$. We let the profile of CP-nets be $P_{CP} = (\vec{\mathcal{N}_1}, \ldots, \vec{\mathcal{N}_m})$.

We make the following claim about the number of consistent edges between an alternative $\vec{d}$ and $\vec{\mathcal{N}_i}$.

Claim 1 *For any* $\vec{d} \in \mathcal{X}$ *and any* $i \le m$*,*

$$N_1(\vec{d}, \vec{\mathcal{N}_i}) = \begin{cases} 4 & \text{if } \vec{d}_{i_1} = v_{i_1} \text{ or } d_{i_2} = v_{i_2} \\ -12 & \text{if } \vec{d}_{i_1} = \bar{v}_{i_1} \text{ and } d_{i_2} = \bar{v}_{i_2} \end{cases}$$

Claim 1 states that the number of consistent edges between $\vec{d}$ and $\vec{\mathcal{N}_i}$ within distance 1 is 4 if the clause c_i is true under the valuation represented by $\vec{d}$; otherwise it is -12. For any $\vec{d} \in \mathcal{X}$, we let $T_{\vec{d}}$ denote the number of clauses in $c_1, \ldots, c_m$ that are true under $\vec{d}$. Then, we have that $N_1(\vec{d}, P_{CP}) = 4T_{\vec{d}} - 12(m - T_{\vec{d}}) = 16T_{\vec{d}} - 12m$. It follows from Theorem 7 that for any $q >$

$\frac{1}{2}$, $MLE_{\pi_{1,q}}(P_{CP}) = \arg\max_{\vec{d}} N_1(\vec{d}, P_{CP}) = \arg\max_{\vec{d}} T_{\vec{d}}$. Therefore, a winner of P_{CP} under $MLE_{\pi_{1,q}}$ corresponds to a valuation under which the number of satisfied clauses is maximized; and any valuation that maximizes the number of satisfied clauses corresponds to a winner of P_{CP} under $MLE_{\pi_{1,q}}$. We note that the size of P_{CP} is $O(mt)$. It follows that computing a winner under $MLE_{\pi_{1,q}}$ is NP-hard.

Clearly the decision problem is in NP. Therefore, the decision problem is NP-complete to compute a winner under $MLE_{\pi_{1,q}}$. □

As we have seen (cf. Corollary 1), for a given multi-issue domain composed of p binary issues, there are *exactly* p voting correspondences defined by distance-based threshold models. As far as we know, these voting correspondences are entirely novel, and are tailored especially for multi-issue domains. Now, among these p voting correspondences, two are even more natural and interesting: $MLE_{\pi_{1,q}}$ and $MLE_{\pi_{p,q}}$. $MLE_{\pi_{1,q}}$ proceeds by electing the alternatives which maximize the sum, over all voters, of the number of neighboring alternatives in the voter's hypercube to which she prefers $\vec{x}$. Now, recall that the Borda correspondence can be characterized as the correspondence where candidate x is a winner if it maximizes the sum, over all voters, of the number of candidates the voter prefers to x. Therefore, $MLE_{\pi_{1,q}}$ is somewhat reminiscent of Borda—except, of course, that we do not count all alternatives defeated by $\vec{x}$ but only defeated alternatives that are one of its neighbors in the hypercube. $MLE_{\pi_{p,q}}$ is even more intuitive: for each issue $\mathbf{x}_i$, the winning value maximizes the number of edges (summing over all voters) that are in favor of it, that is, it is somewhat reminiscent of Kemeny.

So, $MLE_{\pi_{1,q}}$ and $MLE_{\pi_{p,q}}$ are genuinely new voting correspondences for multi-issue binary domains, which can be characterized in terms of maximum likelihood estimators and are quite intuitive; lastly, $MLE_{\pi_{p,q}}$ can be computed in polynomial time.

6. CONCLUSION

The central problems in preference aggregation in multi-issue domains are to find practical ways for voters to represent and report their preferences, as well as to find natural and computationally feasible ways of aggregating these reported preferences. In this paper, we considered the maximum likelihood estimation (MLE) approach to voting, and generalized it to multi-issue domains, assuming that the voters' preferences are expressed by CP-nets. We first studied whether issue-by-issue voting rules and sequential voting rules can be represented by the MLE of some noise model. For separable input profiles, we characterized MLEs of strongly/weakly decomposable models as issue-by-issue voting correspondences composed of local MLEWIVs/candidate scoring correspondences. Although we showed that no sequential voting correspondence can be represented as the MLE for a very weakly decomposable model, we did obtain a positive result here under the assumption that the number of voters is bounded above by a constant.

In the case where all issues are binary, we proposed and axiomatized a class of distance-based noise models; then, we focused on a specific subclass of such models, parameterized by a threshold. We identified the computational complexity of winner determination for the two most relevant values of the threshold.

We note that, whereas Section 4 has a non-constructive flavor because we studied existing voting mechanisms and Theorem 3 is an impossibility theorem, quite the opposite is the case for Section 5. Indeed, the MLE principle led us to define genuinely new families of voting rules and correspondences for multi-issue domains. These rules are radically different from the rules that had previously been proposed and studied for these domains. Unlike sequential or issue-by-issue rules, they do not require any domain restriction, and yet their computational complexity is not that bad (NP-complete at worst, and sometimes polynomial in the size of the CP-nets). We believe that these new rules are very promising.

Future research could further investigate the computational aspects of determining the winners for MLE correspondences. For example, in this paper, we characterized the complexity of computing winners under MLEs of distance-based threshold models with thresholds 1 and p (the number of issues). It would be interesting to identify the complexity for other thresholds (however, we conjecture that it is at least NP-hard). More generally, the study of voting in multi-issue domains is still in its infancy. Unlike in the standard (single-issue) case, relatively few rules have been proposed and relatively little is known about social-choice-theoretic properties. We believe that this paper has demonstrated the potential of the maximum likelihood approach to build a theory of social choice in multi-issue domains.

Acknowledgments

Vincent Conitzer and Lirong Xia are supported by NSF under award number IIS-0812113. Vincent Conitzer is supported by an Alfred P. Sloan Research Fellowship, and Lirong Xia is supported by a James B. Duke Fellowship.

7. REFERENCES

[1] C. Boutilier, R. Brafman, C. Domshlak, H. Hoos, and D. Poole. CP-nets: a tool for representing and reasoning with conditional ceteris paribus statements. *JAIR*, 21:135–191, 2004.

[2] S. Brams, D. Kilgour, and W. Zwicker. The paradox of multiple elections. *Social Choice and Welfare*, 15(2):211–236, 1998.

[3] V. Conitzer, M. Rognlie, and L. Xia. Preference functions that score rankings and maximum likelihood estimation. In *Proc. of IJCAI-09*, pages 109–115, 2009.

[4] V. Conitzer and T. Sandholm. Common voting rules as maximum likelihood estimators. In *Proc. of UAI-05*, pages 145–152, 2005.

[5] M. J. A. N. de Caritat (Marquis de Condorcet). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. 1785. Paris: L'Imprimerie Royale.

[6] M. Drissi and M. Truchon. Maximum likelihood approach to vote aggregation with variable probabilities. Technical Report 0211, Departement d'economique, Universite Laval, 2002.

[7] E. Elkind, P. Faliszewski, and A. Slinko. On distance rationalizability of some voting rules. In *Proc. of TARK-09*, pages 108–117, 2009.

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

[9] D. Lacy and E. M. Niou. A problem with referendums. *Journal of Theoretical Politics*, 12(1):5–31, 2000.

[10] J. Lang and L. Xia. Sequential composition of voting rules in multi-issue domains. *Mathematical Social Sciences*, 57(3):304–324, 2009.

[11] M. Li, Q.B. Vo, and R. Kowalczyk. An Efficient Majority-Rule-based Approach for Collective Decision Making with CP-nets. To appear in *KR-10*, 2010.

[12] F. Rossi, K. Venable, and T. Walsh. mCP nets: representing and reasoning with preferences of multiple agents. In *Proc. of AAAI-04*, pages 729–734, 2004.

[13] L. Xia, V. Conitzer, and J. Lang. Voting on multiattribute domains with cyclic preferential dependencies. In *Proc. of AAAI-08*, pages 202–207, 2008.

[14] L. Xia, J. Lang, and M. Ying. Strongly decomposable voting rules on multiattribute domains. In *Proc. of AAAI-07*, pages 776–781, 2007.

[15] H. P. Young. Optimal voting rules. *Journal of Economic Perspectives*, 9(1):51–64, 1995.