The Maximum Likelihood Approach to Voting on Social Networks

Vincent Conitzer

Departments of Computer Science and Economics
Duke University, Durham, NC 27708, USA
conitzer@cs.duke.edu

Abstract

One view of voting is that voters have inherently different preferences – *de gustibus non est disputandum* – and that voting is merely a method for reaching a reasonable compromise solution. An alternative view is that some of the alternatives really are better in an objective sense, and by voting over the alternatives we hope to be more likely to reach the correct outcome. In this latter view, we can see the votes as noisy estimates of the truth. Specifying a probabilistic noise model gives us a natural "optimal" voting rule for determining the outcome based on the votes, namely, the function that takes the votes as input and produces the outcome that maximizes the likelihood of these votes as output.

We will first review some of the work on the maximum likelihood approach to voting. Most of this work supposes that, conditional on the correct outcome, votes are independent. In reality, however, voters are clearly influenced by the opinions of those close to them. How should we model the effects of the social network, and what does this imply for the maximum likelihood approach? We will first review an earlier result (Conitzer 2012) that states that, under certain assumptions, the social network structure should not affect the voting rule. We then consider a new model under which this is not true, and prove that computing the probability of the votes given the correct outcome is #P-hard under this model. On the other hand, if the goal is to simultaneously also give a point estimate of the hidden variables in the model, then the optimization problem can be solved in polynomial time.

Introduction

Not all voting settings are created equal. In some, none of the alternatives are *inherently* the "wrong" or "right" choice. Each voter finds some alternatives more palatable than others, and this is fundamentally due to each voter's personal preferences. In particular, it may be the case that the voters know everything there is to know about the situation and that they understand perfectly how the different personal circumstances in which other voters find themselves lead them to have different preferences. The objective is merely to reach an outcome that most voters find acceptable. In other settings, however, the situation is different, and we may suppose that some alternatives are *inherently* better than others. For example, the voters may genuinely be trying to evaluate each alternative's absolute quality (according to some measure), and the reason that they vote differently is not that they have different personal interests in the matter, but rather that it is difficult to evaluate an alternative's quality and the voters simply differ in their assessments.

Of course, both of these are descriptions of extremes, and in many voting settings we find elements of both. A standard political election clearly involves the voters having different personal interests in the outcome and voting accordingly. Meanwhile, however, there is typically also uncertainty about the candidates. For example, would a given candidate be willing to work long hours if elected? Even if all voters agree that such a good work ethic is important and desirable, they may have different assessments of which candidate would work longer hours. However, in spite of the fact that these two views of voting often blur together, it is conceptually useful to consider the extremes. In this paper, we will consider the second extreme, where one alternative is inherently better than the other, and any disagreement among the voters is merely due to different assessments of this absolute quality. We will be interested in the optimal design of the voting rule in this context. That is, given the votes, how should the outcome of the election be chosen to maximize the probability that the correct one is chosen?

The specific approach that we will consider is the *maximum likelihood estimation* approach to the design of voting rules. Its basic idea is that voters obtain noisy estimates of the alternatives' qualities and vote accordingly; and, given a precise probabilistic model for how the alternatives' true qualities lead to these noisy observations, we can attempt to estimate the correct outcome of the election as the one that maximizes the likelihood of the observed votes.

This approach dates back all the way to Condorcet (Condorcet 1785), but there has been a recent resurgence of interest in it, especially from people in the computational social choice community. It involves techniques from optimization and reasoning under uncertainty. Most of this work so far has assumed that (conditional on the correct outcome) votes are independent. In this paper, we will discuss a new model in which the social network structure among the agents affects how votes are formed.

Background

In this section, we first review the maximum likelihood approach to voting in settings where votes are drawn independently (conditional on the correct outcome). Then, we discuss an existing model (Conitzer 2012) in which the votes are not independent—the social network structure plays a role—but nevertheless this social network structure does not end up affecting the optimal voting rule.

The Maximum Likelihood Approach to Voting

The approach is easiest to describe in the context where there are only two alternatives—call them -1 and 1. One of these two is the "correct" (or "better") alternative, but it is not directly observed which one. Let $c \in \{-1, 1\}$ denote the correct alternative (and $-c$ the other alternative). In the most basic model, the voters' votes are i.i.d., with each voter voting for c with some fixed probability $p > 0.5$ and for $-c$ with the remaining probability $1 - p$. Hence, voter v's vote A_v represents a noisy estimate of the truth. Conditional on -1 (resp. 1) being the correct alternative, a specific vote profile A_V (a vector of votes, one for each voter) that has n_{-1} votes for -1 and n_1 votes for 1 has probability $P(A_V|c = -1) = p^{n_{-1}}(1 - p)^{n_1}$ (resp. $P(A_V|c = 1) = p^{n_1}(1 - p)^{n_{-1}}$). We would like to pick the winner of our election—equivalently, our estimate $\hat{c}$ of c—to be the one that maximizes this likelihood. Because $p > 1 - p$, we should pick $\hat{c} = -1$ (resp. $\hat{c} = 1$) if $n_{-1} > n_1$ (resp. $n_1 > n_{-1}$). That is, we should simply pick the *majority winner*, the one that receives more votes (with ties broken arbitrarily). The focus on maximizing the likelihood of the observed profile can also be justified as follows: if we *a priori* believe that either alternative is equally likely to be the correct one, then the alternative that maximizes the likelihood of the observed votes is also the *maximum a posteriori* estimate of the correct alternative. This is because by Bayes' rule we have $P(c = \hat{c}|A_V) = P(A_V|c = \hat{c})P(c = \hat{c})/P(A_V)$. Because $P(c = \hat{c})$ (due to the equally-likely-a-priori assumption) and $P(A_V)$ do not depend on $\hat{c}$, choosing $\hat{c}$ to maximize $P(A_V|c = \hat{c})$ is equivalent to choosing it to maximize $P(c = \hat{c}|A_V)$. The approach can be generalized to settings where there is a potentially different probability p_i for every voter i, resulting in a weighted majority rule (Nitzan and Paroush 1982; Shapley and Grofman 1984).

In many settings, there are more than two alternatives. How should the above model be extended to such a setting? First, it is natural to presume now that there is a correct *ranking* of all the alternatives. Moreover, each voter's noisy estimate is now also a ranking of the alternatives. This fits well with the theory of voting more generally, where a vote is often assumed to rank all alternatives. But what is the probability distribution over such votes given the correct ranking? Condorcet (Condorcet 1785) attempted to give such a model, and this was later made more precise by Young (Young 1988; 1995), who showed that the optimal voting rule for the resulting noise model is one proposed by Kemeny (Kemeny 1959). A number of articles have since been devoted to the study of different noise models and the optimal rules to which they lead (Drissi-Bakhkhat and Truchon 2004; Conitzer and Sandholm 2005; Truchon 2008;

Conitzer, Rognlie, and Xia 2009). Elkind and Slinko (Elkind and Slinko to appear) give a recent overview. In this paper, however, we will restrict our attention to settings with two alternatives.

Review: A Model Where Social Network Structure Plays a Role but Does Not Matter

We next discuss a model (Conitzer 2012) in which votes are not independent (even when conditioning on the correct outcome). Instead, the voters are the vertices V of a social network. The probability of a profile A_V of votes (where A_v is the vote of voter v) is $\prod_{v \in V} f_v(A_v, A_{N(v)}|c)$, where $N(v)$ is the set of voters that are neighbors of v. Hence, f_v is intended to capture the interaction between v and its neighbors (and the truth, i.e., the correct alternative). Then, there is the further assumption in the model that for each $v \in V$, f factors as $f_v(A_v, A_{N(v)}|c) = g_v(A_v|c)h_v(A_v, A_{N(v)})$. In words, there is one factor representing that the voter is more likely to vote for the correct alternative, and another factor representing the interaction between a voter and its neighbors (which does not depend on the correct alternative). Under this assumption, it can be shown that the social network structure, represented by the functions h_v, does not affect the maximum likelihood rule. This is for the simple reason that

$$\arg\max_{\hat{c}} \prod_{v \in V} f_v(A_v, A_{N(v)}|\hat{c})$$

$$= \arg\max_{\hat{c}} \prod_{v \in V} g_v(A_v|\hat{c})h_v(A_v, A_{N(v)})$$

$$= \arg\max_{\hat{c}} \prod_{v \in V} g_v(A_v|\hat{c})$$

A New Model That Takes Social Network Structure into Account

The conclusion from the above model that we can simply ignore social network structure should, of course, be taken with a grain of salt. Other models will lead to different conclusions. The most interesting aspect of the above result, in my opinion, is that it is at least not *obvious* how social network structure should affect the voting rule. In the below, we will consider a new model that leads to a different conclusion. Like the model considered above, this new model undeniably leaves out many important aspects of how agents form their votes in social networks. Perhaps most notably, this new model still does not include a temporal component for modeling the gradual evolution of opinion. The objective here, rather, is to give a simple model that helps to illustrate which phenomena we are likely to encounter as we move to more complex models.

The Independent Conversations Model

The model can be described as follows. Every voter will have a conversation with each of her neighbors, so that there is exactly one conversation per edge. This conversation will turn out in favor of one of the two alternatives. The outcomes of the conversations are i.i.d., and every voter
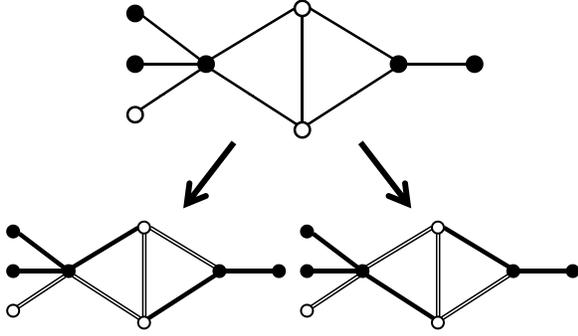
Figure 1: A social network with a vote profile for the vertices, and the two edge profiles that are consistent with this vote profile. Open vertices indicate votes for -1 and closed vertices indicate votes for 1. Similarly, for the edge profile graphs, open edges are associated with -1 and closed edges are associated with 1.

votes according to the majority of the outcomes of the conversations in which she participated. More specifically:

Definition 1 (Independent conversations model). *For simplicity, assume that every voter has an odd number of neighbors. Associated with every edge $e = (v, w)$ is a random alternative A_e, which is equal to the correct winner c with probability $p > 0.5$ and to the other alternative $-c$ with probability $1 - p$. The edge profile $A_E = (A_e)_{e \in E}$ is not directly observed, but each vertex votes according to the majority of its incident edges, i.e., $A_v = maj\{A_{(v,w)}\}_{w \in N(v)}$. Hence, the probability of observing the profile $A_V = (A_v)_{v \in V}$ given that the correct alternative is $\hat{c}$ is the sum of the probabilities of the edge profiles A_E that are consistent with A_V, that is,*

$$P(A_V|\hat{c}) = \sum_{\substack{A_E \,:\, for\ all\ v \in V, \\ A_v = maj\{A_{(v,w)}\}_{w \in N(v)}}} p^{n(\hat{c}, A_E)}(1-p)^{|E|-n(\hat{c}, A_E)}$$

where $n(\hat{c}, A_E)$ is the number of edges associated with $\hat{c}$ in A_E. The maximum likelihood alternative $\hat{c}$, then, is the one maximizing this expression.

Example 1. *Figure 1 illustrates the model. At the top, it shows the social network and the votes cast by the voters (vertices). Below, it shows the (only) two different edge profiles (ways of labeling the edges) that would result in this vote profile. If 1 is the correct alternative, each of these edge profiles has probability $p^5(1-p)^4$. Hence, $P(A_V|1) = 2p^5(1-p)^4$. (Here, $P(A_V|1)$ is shorthand for $P(A_V|c = 1)$.) Similarly, $P(A_V|-1) = 2p^4(1-p)^5$. It follows that 1 maximizes the likelihood.*

Note that under this model, it is possible that *no* edge profiles are consistent with the observed votes—for example, if two vertices that are each other's only neighbor vote differently. To address this, it would be straightforward to extend the model so that a vertex has some small probability
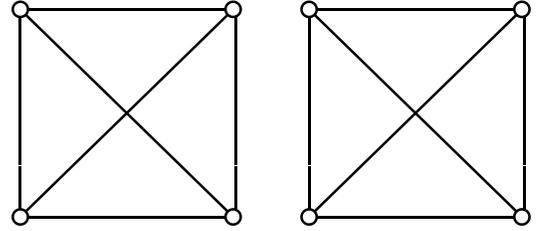

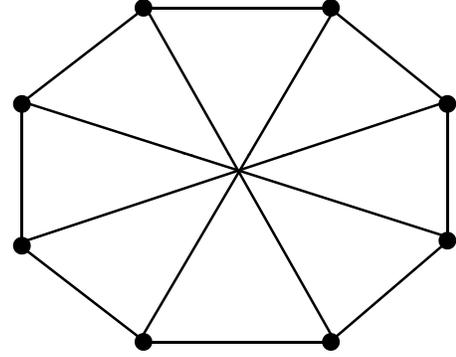Figure 2: A graph consisting of two cliques of size four. Open vertices indicate votes for -1.

Figure 3: A "wheel" graph (note there is no vertex in the middle). Closed vertices indicate votes for 1.

of voting against the majority of its incident edges. Since the model is anyway supposed to be illustrative rather than comprehensive, for the sake of simplicity, we will not flesh out this extension here.

The outcome in Example 1 was perhaps not surprising. Let us now consider an example where it is not immediately obvious which alternative should win.

Example 2. *Consider the graph in Figure 2, consisting of two cliques of size 4. Suppose all these vertices vote for -1. Because every vertex has 3 incident edges, an edge profile is consistent with this profile $A_{V\,cliques}$ if and only if every vertex has at most one incident edge that is associated with 1— that is, the edges associated with 1 constitute a matching. In this graph, there is 1 matching with 0 edges, there are 12 with 1 edge, 42 with 2 edges, 36 with 3 edges, and 9 with 4 edges. Thus, if -1 is in fact the correct alternative, then the probability of $A_{V\,cliques}$ is $P(A_{V\,cliques}|-1) = p^{12} + 12p^{11}(1-p) + 42p^{10}(1-p)^2 + 36p^9(1-p)^3 + 9p^8(1-p)^4$. For $p = 0.6$, we obtain $P(A_{V\,cliques}|-1) = 8.73 \cdot 10^{-2}$. On the other hand, if 1 is the correct alternative, then the probability of $A_{V\,cliques}$ is $P(A_{V\,cliques}|1) = (1-p)^{12} + 12(1-p)^{11}p + 42(1-p)^{10}p^2 + 36(1-p)^9p^3 + 9(1-p)^8p^4$. For $p = 0.6$, we obtain $P(A_{V\,cliques}|1) = 4.71 \cdot 10^{-3}$.*

Now consider the graph in Figure 3. Suppose all these vertices vote for 1. Again, because every vertex has 3 in-

cident edges, an edge profile is consistent with this profile $A_{V^{wheel}}$ if and only if the edges associated with -1 constitute a matching. In this graph, there is 1 matching with 0 edges, there are 12 with 1 edge, 42 with 2 edges, 44 with 3 edges, and 7 with 4 edges. Using similar expressions as above, we obtain that for $p = 0.6$, $P(A_{V^{wheel}}|-1) = 4.99 \cdot 10^{-3}$ and $P(A_{V^{wheel}}|1) = 9.16 \cdot 10^{-2}$.

Now consider taking the union of the two graphs. Because there are no edges between the two graphs, they are drawn independently, so $P(A_{V^{union}}|c) = P(A_{V^{cliques}}|c)P(A_{V^{wheel}}|c)$. Thus we get, for $p = 0.6$, that $P(A_{V^{union}}|-1) = 8.73 \cdot 10^{-2} \cdot 4.99 \cdot 10^{-3} = 4.36 \cdot 10^{-4}$ and $P(A_{V^{union}}|1) = 4.71 \cdot 10^{-3} \cdot 9.16 \cdot 10^{-2} = 4.31 \cdot 10^{-4}$. Hence, -1 would be the maximum likelihood estimate winner in this case. Intriguingly, if we change p to a value of 0.9 (but keep everything else the same), then the calculations become $P(A_{V^{union}}|-1) = 8.20 \cdot 10^{-1} \cdot 8.15 \cdot 10^{-8} = 6.68 \cdot 10^{-8}$ and $P(A_{V^{union}}|1) = 8.88 \cdot 10^{-8} \cdot 8.23 \cdot 10^{-1} = 7.31 \cdot 10^{-8}$ so that then, 1 wins! Intuition can be given for this as follows: as p goes to 1, one of the two graphs (the "correct" voters) will have probability close to 1, and so the likelihood is primarily determined by the other graph (the "incorrect" voters). The probability of the latter graph is dominated by the terms with the largest matching (4 edges), because those have the fewest factors $(1 - p)$. Because the two-cliques graph has more of those matchings, it is better to make it the "incorrect" graph—it is easier to explain why it might have voted so incorrectly.

Computational Hardness

In the examples above, we calculated probabilities by enumerating, or at least counting, the various edge profiles that could give rise to the observed vote profiles. Can we scale this approach to large graphs? For example, is there a scheme that always allows us to count the various types of edge profiles quickly? Or, if this is not possible, perhaps there is a shortcut that allows us to compute the relevant probabilities without any counting? Unfortunately, the following complexity result suggests that our problem is fundamentally a hard counting problem.

Theorem 1. *Computing $P(A_V|\hat{c})$ is #P-hard under the independent conversations model.*

Proof. We reduce from the following problem: given a bipartite graph G' with n' vertices on each side, how many perfect matchings does it have? This problem is #P-complete (Valiant 1979). Given this bipartite graph $G' = (V', E')$, we construct a social network $G = (V, E)$ and votes A_V over the alternatives -1 and 1 as follows. For each vertex v' on the left side of V', in G we construct a vertex $v_{v'}$, which votes for 1, and another $n_{v'} - 1$ vertices $v_{v'}^1, \ldots, v_{v'}^{n_{v'}-1}$ (where $n_{v'}$ is the number of neighbors of v' in G'; w.l.o.g., $n_{v'} \geq 1$), which all vote for 1 as well. For each vertex w' on the right side of V', in G we construct a vertex $v_{w'}$, which votes for -1, and another $n_{w'} + 1$ vertices $v_{w'}^1, \ldots, v_{w'}^{n_{w'}+1}$, which all vote for 1 except for two (say, $v_{w'}^1$ and $v_{w'}^2$ vote for -1—again, w.l.o.g., $n_{w'} \geq 1$).
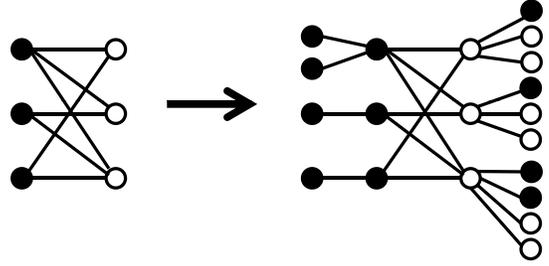

Figure 4: An illustration of the reduction. On the left-hand side, there is an instance of the problem of counting the number of perfect matchings in a bipartite graph. On the right-hand side, there is an instance of the problem of computing $P(A_V|\hat{c})$. Here, again, open vertices indicate votes for -1, and closed vertices indicate votes for 1.

We construct an edge from every $v_{v'}^i$ (resp. every $v_{w'}^i$) to its corresponding $v_{v'}$ (resp. $v_{w'}$). Moreover, for every edge $(v', w') \in E'$, we construct an edge $(v_{v'}, v_{w'}) \in E$ (call these the "original" edges). Figure 4 illustrates the reduction.

Which edge profiles $\hat{A}_E$ are consistent with the profile A_V? First, we note that every $v_{v'}^i$ (resp. every $v_{w'}^i$) has only a single edge. The alternative with which that edge is associated must thus coincide with the alternative chosen by $v_{v'}^i$ (resp. $v_{w'}^i$). For each left-hand-side $v_{v'} \in V$, this immediately gives $n_{v'} - 1$ edges incident to $v_{v'}$ that are associated with 1. Because $v_{v'}$ votes for 1 and has $2n_{v'} - 1$ incident edges in G, *at least one of the original edges incident to $v_{v'}$ must be associated with 1 as well.* On the other hand, for each right-hand-side $v_{w'} \in V$, $v_{w'}$ has only two non-original incident edges that are associated with -1. Because $v_{w'}$ votes for -1 and has $2n_{w'} + 1$ incident edges in G, at least $n_{w'} - 1$ of the original edges incident to $v_{w'}$ must be associated with -1—or, equivalently, *at most one of the original edges incident to $v_{w'}$ may be associated with 1.* It follows that for $\hat{A}_E$ to be consistent with A_V, every $v_{v'}$ must have *exactly* one incident original edge associated with 1 (or otherwise some $v_{w'}$ would have more than one), and similarly every $v_{w'}$ must have *exactly* one incident original edge associated with 1 (or otherwise some $v_{v'}$ would have less than one). That is, $\hat{A}_E$ is consistent with A_V if and only if the original edges associated with 1 constitute a perfect matching (and the non-original edges take their required associated alternative). Therefore, the number of consistent edge profiles is equal to the number of perfect matchings in the original bipartite graph. Every one of these edge profiles has the same probability (because they all have the same number of edges associated with 1), and therefore the probability $P(A_V|\hat{c})$ is proportional to the number of matchings in the original bipartite graph. □

Estimating the Correct Alternative and the Edge Profile Together Is Easy

As should be clear from the above proof, the computational hardness of computing the relevant probabilities is due to the hidden variables (A_E) over whose possible values we must sum. It seems such computational hardness is likely to occur for many other models that involve hidden variables. However, another approach is to compute an estimate not only of the correct winner (summing over all the hidden variables in the process), but rather to estimate the correct winner *together with* the hidden variables, so that we need not sum over the latter, but rather have a point estimate of them. In the model considered here, this corresponds to estimating the correct winner c together with the edge profile A_E, so as to maximize $P(A_V, \hat{A}_E | \hat{c})$. Note that $P(A_V, A_E | c) = P(A_V | A_E, c) P(A_E | c)$, where $P(A_V | A_E, c) = P(A_V | A_E)$ is 1 if A_E is consistent with A_V and 0 otherwise. Therefore, the goal is to find $\hat{c}$ and $\hat{A}_E$ to maximize $P(\hat{A}_E | \hat{c})$, under the constraint that $\hat{A}_E$ is consistent with A_V.[1] It turns out that it is actually possible to do so in polynomial time. In fact, we can do so even under the following slightly richer model, in which different edges can have different probabilities of being associated with the correct alternative.

Definition 2 (Independent weighted conversations model). *This model is identical to the independent conversations model from Definition 1, except here, rather than a single universal value p, there is a separate probability $p_e \geq 1/2$ associated with each edge, which is the probability that that edge will be associated with the correct alternative.*

Theorem 2. *An element of* $\arg\max_{(\hat{c}, \hat{A}_E)} P(A_V, \hat{A}_E | \hat{c})$ *can be computed in polynomial time, even in the independent weighted conversations model.*

Proof. As discussed above, the goal is to find $\hat{c}$ and $\hat{A}_E$ to maximize $P(\hat{A}_E | \hat{c})$, under the constraint that $\hat{A}_E$ is consistent with A_V. For each possibility for $\hat{c}$, we proceed by reducing the problem to a maximum weighted b-matching problem (for a discussion of this problem and further references, see, for example, Penn and Tennenholtz (Penn and Tennenholtz 2005)). This is a generalized weighted matching problem in which each vertex v has a lower bound l_v and an upper bound u_v on how many of its incident edges may be chosen in the matching, and the edges have weights. We wish to maximize the total weight of the chosen edges. (In the general version of the problem, it is possible to choose an edge more than once, i.e., associate an unrestricted nonnegative integer with each edge; moreover, lower and upper bounds (capacities) on the number of times each edge can be chosen may be provided. We will only need the case where each edge can be chosen at most once.)

In the reduction from our problem to the b-matching problem, we keep the graph the same. Choosing an edge in the b-matching problem will correspond to associating it with our current estimate of the correct alternative $\hat{c}$ (and not choosing it will correspond to associating it with our current estimate of the incorrect alternative $-\hat{c}$). Accordingly, for a vertex v that is voting for $\hat{c}$ and has n_v neighbors, we set $l_v = (n_v + 1)/2$ (recall that n_v is odd by assumption) and $u_v = n_v$. For a vertex v that is voting for $-\hat{c}$, we set $l_v = 0$ and $u_v = (n_v - 1)/2$. These conditions are necessary and sufficient for $\hat{A}_E$ to be consistent with A_V. Furthermore, we set the weight of edge e to $w_e = \log p_e - \log(1 - p_e)$.[2] This results in a total objective value of $\sum_{\text{chosen } e}(\log p_e - \log(1 - p_e)) = -\sum_{e \in E} \log(1 - p_e) + \sum_{\text{chosen } e} \log p_e + \sum_{\text{not chosen } e} \log(1 - p_e)$ that we seek to maximize. Because $-\sum_{e \in E} \log(1 - p_e)$ is a constant, this is equivalent to maximizing $\sum_{\text{chosen } e} \log p_e + \sum_{\text{not chosen } e} \log(1 - p_e)$. But by exponentiation, this is equivalent to maximizing $(\prod_{\text{chosen } e} p_e)(\prod_{\text{not chosen } e}(1 - p_e)) = P(\hat{A}_E | \hat{c})$, as required. $\square$

The fact that estimating the hidden variables together with the correct alternative can be done in polynomial time may not generalize to other models. Nevertheless, it seems likely that in many models, it will still be *easier* than summing over all the hidden variables; it may be, for example, the difference between solving an NP-hard problem and a #P-hard one.

Conclusion

Most of the work on interpreting voting rules as maximum likelihood estimators so far has assumed that voters' votes are drawn independently (conditional on the correct outcome). While this assumption results in some nice characterizations, in many contexts it is clearly unrealistic. Often, voters have the opportunity to discuss the alternatives with neighboring voters before casting their votes, and this will affect their votes. On the other hand, it is not straightforward to create a noise model that captures all the aspects of how conversations with other voters affect vote formation. A model that truly achieves this would presumably be quite baroque. Nevertheless, simple models, which undoubtedly leave out many real-world aspects of social vote formation, can nevertheless provide insight into how an election organizer should take social network structure into account in the design of the voting rule (when it is possible to do so).

In this paper, we first reviewed basic existing results on MLE voting rules, paying particular attention to a result that, for a particular type of noise model that does take social

---

[1] It should be pointed out here that there is no guarantee that this will result in the same estimate $\hat{c}$ as in the previous case where we sum over all edge profiles. A similar issue occurs in the standard model (with conditionally independent votes) when there are more than two alternatives. There, we have to choose whether to estimate the entire correct ranking, or merely the correct top alternative. What we obtain in the latter case is not guaranteed to be equal to the top alternative in our estimate of the correct ranking. For further discussion, see Elkind and Slinko (Elkind and Slinko to appear).

[2] In the case of independent voters with different probabilities p_i of choosing the correct outcome, Nitzan and Paroush (Nitzan and Paroush 1982) and Shapley and Grofman (Shapley and Grofman 1984) also use a weight $\log p_i - \log(1 - p_i)$.

network structure into account, the optimal voting rule does not depend on this network structure. Then, to illustrate how the same result may not hold under different models that take social network structure into account, we considered a new noise model—the independent conversations model—in which for every edge, there is a conversation that settles on one of the two alternatives. The outcomes of these conversations are independent (conditional on the correct outcome), and a voter votes according to the majority of the outcomes of the conversations in which she participated. We illustrated the model with some examples. We showed that computing the conditional probability of a vote profile in this model is in fact #P-hard, by reduction from the problem of computing the number of perfect matchings in a bipartite graph. Intuitively, this hardness is due to the fact that we need to sum over all possible edge profiles (combinations of outcomes for the edges). Indeed, we then showed that if the goal is to find a probability maximizing estimate of *both* the correct outcome and the edge profile, this can be done in polynomial time using matching techniques. The prevalence of matching techniques in these results is intriguing.

There can be little doubt, though, that the independent conversations model leaves out many real-world aspects of how votes are socially formed. The most obvious gap is that time still plays no role in the model. It would better match reality to explicitly model the evolution of a voter's opinion over time. One interesting approach would be to try to integrate DeGroot-style models of opinion formation ((DeGroot 1974); for a recent article with further references, see (Dandekar, Goel, and Lee 2013)) with the maximum likelihood approach to voting.

Acknowledgments

References

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

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

Conitzer, V. 2012. Should social network structure be taken into account in elections? *Mathematical Social Sciences* 64(1):100–102. Special Issue on Computational Foundations of Social Choice.

Dandekar, P.; Goel, A.; and Lee, D. T. 2013. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* 110(15):5791–5796.

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

DeGroot, M. H. 1974. Reaching a consensus. *Journal of the American Statistical Association* 69(345):118–121.

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

Elkind, E., and Slinko, A. Rationalizations of voting rules. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*. To appear.

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

Nitzan, S., and Paroush, J. 1982. Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review* 23(2):289–297.

Penn, M., and Tennenholtz, M. 2005. On multi-object auctions and matching theory: Algorithmic aspects. In Golumbic, M. C., and Hartman, I. B.-A., eds., *Graph Theory, Combinatorics and Algorithms: Interdisciplinary Applications*, Operations Research/Computer Science Interfaces Series. Springer. chapter 7, 173–188.

Shapley, L., and Grofman, B. 1984. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice* 43(3):329–343.

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

Valiant, L. 1979. The complexity of computing the permanent. *Theoretical Computer Science* 8:189–201.

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

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