UPenn NETS 412: Algorithmic Game Theory March 15, 2018

Lecture 12

Lecturer: Bo Waggoner *Scribe: Bo Waggoner*

Voting and Social Choice

Following the theme of the last few lectures, we will continue to consider algorithms that take in preferences from a group of agents and output some decisions or assignments for those agents. Specifically, we will suppose there is some choice to be made, for example, a candidate must be elected to a political office. We would like a *preference aggregation* or *voting* scheme to do so in a reasonable and fair way.

Notation and setting. We have n agents $1, \dots, n$ and m *candidates* $1, \dots, m$. To avoid confusion, we will use i, i', etc. to refer to agents and j, j', etc. to refer to candidates.

Each agent i has a preference order $\succ_i$ where $j \succ_i j'$ if agent i prefers candidate j to j'. For this class, we will assume that the preference orders are *strict*, i.e. there are no ties.

A *social welfare function* g takes as input the preferences of the agents, $(\succ_1, \dots, \succ_n)$, and outputs a ranking or strict preference order $\succ$.

A *social choice function* (or "voting rule") f takes as input the preferences of the agents, $(\succ_1, \dots, \succ_n)$, and outputs a winning candidate $j \in \{1, \dots, m\}$. For the sake of time, we will mostly focus on social choice functions rather than social welfare functions, but both are widely studied.

Pareto optimality. The first, basic requirement we might think of for a voting rule is that we should not pick an "obviously bad" candidate. That is, there should not be some other candidate that *everyone* prefers to the selected one.

Definition 1 *Given a set of preferences $\succ_1, \dots, \succ_n$, candidate j* ***Pareto dominates*** *j' if, for all agents i, $j \succ_i j'$. Candidate j is* ***Pareto optimal*** *if there is no j' that Pareto dominates it. A social choice function f satisfies* ***Pareto optimality*** *if it always selects a Pareto optimal candidate.*

However, Pareto optimality is not enough. Consider the voting rule that always selects voter 1's favorite candidate, regardless of what all the other voters say. This rule is Pareto optimal — why? — but not so representative of the group preferences. In fact, it is *dictatorial*, which is what we call any voting rule that just follows the preference of a single voter.

Condorcet's criterion and paradox. One of the first researchers into voting was the Marquis de Condorcet in the late 1700s. Given a set of preferences $\succ_1, \dots, \succ_n$ and two candidates j, j', let

$$N_{j>j'} = |\{i : j \succ_i j'\}|$$

that is, $N_{j>j'}$ is the number of voters who prefer j to j'.

Definition 2 *Given a set of preferences $\succ_1, \dots, \succ_n$, candidate j is a* ***Condorcet winner*** *if, for all other candidates j', $N_{j>j'} > \frac{n}{2}$. A social choice function f satisfies the* ***Condorcet criterion*** *if, whenever a Condorcet winner exists, f always selects one.*

In other words, a Condorcet winner would beat every other candidate in a pairwise election, if they were the only two candidates. If there is a Condorcet winner, then it makes sense to select it. If we picked anyone else, a majority of voters would rather switch back to the Condorcet winner.

Unfortunately, voting systems used in practice don't always satisfy the Condorcet criterion. Consider the majority vote on 3 or more candidates; one can find likely examples from US presidential elections in which a 3rd party candidate "spoiled" the election for a presumptive Condorcet winner.

Worse, a Condorcet winner need not always exist. Imagine 50 people prefer $A \succ B \succ C$, and 50 people prefer $B \succ C \succ A$, and 50 people prefer $C \succ A \succ B$. So if we look at pairwise comparisons, A beats B, and B beats C, but C beats A.

This is an important point: Even if each voter is rational and decisive, the group's preferences can exhibit inconsistencies and non-transitivity.

Impossibilities. Unfortunately, Arrow showed[1] that satisfying even very weak axioms for rationality is not always possible.

Let $R = (\succ_1, \dots, \succ_n)$ be the list of voters' rankings, and for any subset of candidates $S \subseteq \{1, \dots, m\}$, let R_S be the same rankings restricted only to candidates in S. In other words, we throw all other candidates except the set S out of the rankings.

Definition 3 *A social choice function f satisfies the* ***weak axiom of revealed preferences (WARP)*** *if, whenever $j = f(R)$, for any $S \subseteq \{1, \dots, m\}$ with $j \in S$, we have $j \in f(R_S)$.*

This axiom says that, if the voters elect candidate j, then they would still elect j if there were fewer competing candidates. In other words, if we throw out some losing candidates from consideration and re-run the election, j should still win.

This might seem like a natural and reasonable axiom for a voting rule to satisfy. Unfortunately, however, we have a basic impossibility result:

Theorem 4 (Arrow, 1950s) *Any social choice function f that satisfies Pareto optimality and WARP must be dictatorial.*

(Proof omitted.)

This is interpreted as an impossibility because it says the nice conditions of WARP and Pareto optimality cannot be satisfied, except by dictatorial rules which aren't considered to be so great. There is also the more famous "Arrow impossibility theorem" which has a similar result but applies to social welfare functions rather than social choice functions, with a different axiom than WARP.

So far, we have implicitly assumed everyone reports truthfully. But what if we also want the rule to be incentive-compatible? Recall that *dominant strategy incentive compatible* means that each voter is always best-responding by truthfully reporting her true preference order, regardless of what the others do.

Theorem 5 (Gibbard-Satterthwaite, 1970s) *Any social choice function f that satisfies Pareto optimality and is dominant strategy incentive compatible must be dictatorial.*

(Proof omitted.)

Note that both of these results only hold if $m \geq 3$. If $m = 2$, then the majority vote rule satisfies just about every nice axiom you can think of (though many axioms, like WARP, aren't very meaningful with only two candidates).

The maximum-likelihood approach. One can view voting as a statistical method: each voter has a piece of information, and we wish to aggregate these into a final prediction or estimate of the best candidate or the best ranking. To formalize this, suppose there is a true underlying correct ranking $\succ_*$ and everyone draws a preference order randomly, independently and identically distributed from a distribution that depends on $\succ_*$. Can we aggregate the votes and learn $\succ_*$?

One noise model, informally studied by Condorcet, is the following (called the Mallows model of noise). Let $p \in (0.5, 1.0)$.

[1]This is a version of the more famous Arrow impossibility theorem. The famous one applies to functions that generate a ranking, whereas this one applies to functions that produce a single winner.

1. There is a true underlying ranking $\succ_*$.

2. To generate $\succ_i$, consider each pairwise comparison j vs j'. With probability p, let $\succ_i$ match $\succ_*$; that is, $j \succ_i j'$ if and only if $j \succ_* j'$. Otherwise, with probability $1-p$, let $\succ_i$ disagree with $\succ_*$ on this pair.

3. Construct $\succ_i$ as above for all pairs. If this produces a valid strict preference ordering on the candidates $\succ_i$, then stop and output it. Otherwise, start over.

The interpretation of this process is that each pair of candidates j, j' has some "true" underlying comparison, e.g. $j \succ_* j'$, and each voter gets that comparison correct with probability p and incorrect with probability $1-p$. The voter correctly ranks this pair with probability $p > 0.5$, otherwise, they incorrectly rank j above j^*. Create a ranking by repeating this for all pairwise comparisons (start over if you do not obtain a valid ranking).

Mathematically, what is the probability of generating a given ranking $\succ_i$, given true underlying ranking $\succ_*$? If $\succ_i$ and $\succ_*$ agree on a_i pairs of candidates and disagree on the remaining $\binom{m}{2} - a_i$ pairs, then the probability is proportional to $p^{a_i}(1-p)^{\binom{m}{2}-a_i}$. (There is a normalizing constant, but it's the same for all $\succ_i$, so let's not worry about it.)

Now the question is: given a set of votes $(\succ_1, \ldots, \succ_n)$, what ranking best explains them, i.e. what underlying ranking maximizes the probability of seeing these votes? This is called the "maximum likelihood estimator".

First, in seemingly unrelated news, the *Kemeny* social welfare function picks the ranking of candidates that agrees with the most total pairwise comparisons of all the voters. That is, for every possible ranking of $\{1, \ldots, m\}$, give that ranking a point for each voter and pairwise comparison j, j' the ranking agrees with. Pick the ranking with the most points.

Theorem 6 (Young 1995) *The Kemeny rule outputs a ranking that is the maximum-likelihood estimator for the Mallows noise model described above.*

Proof For a given ranking $\succ$, we compute the likelihood (or probability) of seeing the list of rankings $(\succ_1, \ldots, \succ_n)$ assuming that $\succ$ is the true underlying ranking. Then we pick the $\succ$ that maximizes this likelihood.

As derived above, the probability of i's vote, given that $\succ$ is the true ranking, is proportional to $p^{a_i}(1-p)^{\binom{m}{2}-a_i}$, where a_i is the number of pairwise agreements between $\succ$ and $\succ_i$. Let $A = \sum_{i=1}^n a_i$.

So the probability of a whole set of votes, if the underlying ranking were $\succ$, is (note we use $\propto$ for "proportional to")

$$\begin{aligned}
\Pr[\succ_1, \ldots, \succ_n] &= \prod_{i=1}^n \Pr[\succ_i] \\
&\propto \prod_{i=1}^n p^{a_i}(1-p)^{b_i} \\
&= p^{\sum_{i=1}^n a_i}(1-p)^{\sum_{i=1}^n b_i} \\
&= p^A (1-p)^{n\binom{m}{2}-A}.
\end{aligned}$$

Now, this probability is increasing in A, because $p > 0.5$, so it is higher the more total pairwise agreements between $\succ$ and the rankings $(\succ_1, \ldots, \succ_n)$. So the ranking $\succ$ with the highest total number of pairwise agreements maximizes the likelihood of these votes. But this is exactly the Kemeny rule. ■

What's cool is that the maximum-likelihood approach led us to a known voting rule that is already known to be nice from an axiomatic standpoint. In fact:

Theorem 7 *When using the Kemeny rule to produce a ranking, and then taking the top candidate as the winner, one satisfies the Condorcet criterion.*

Proof Suppose j is preferred to j' by more than half the voters, for all j'. Consider any ranking produced by Kemeny. If j were not first in the ranking, we could move j up one spot (suppose we swap it with some j'). This would cause more new pairwise agreements than disagreements, so the new ranking would have a better "Kemeny score"; a contradiction. ■

Unfortunately, there is the following problem with the Kemeny rule:

Theorem 8 (Bartholdi, Tovey, Trick 1989) *The Kemeny rule is NP-hard to compute (as the number of candidates grows).*

However, you'll show on the homework that just finding a Condorcet winner is not computationally difficult.