Algorithmic Game Theory, Summer 2018 Lecture 26 (3 pages)

Voting: Gibbard-Saitterthwaite Theorem and Positive Examples

Instructor: Thomas Kesselheim

Today, we come back to the problems of *voting* and *ranking*. Recall that there are n agents, who each have a full preference list over the candidates Γ. We denote agent i's preference by $\succ_i$.

A voting rule computes, given the preferences of all candidates, one candidate. A ranking rule determines a single full ranking of all candidates.

Last time, we have already seen *Arrow's theorem*. We showed that for a ranking rule not to be strategically vulnerable it has to fulfill a property called *independence of irrelevant alternatives* (IIA). Besides, we wanted the ranking rule to fulfill *unanimity*. That is, if all agents' preferences are identical, then this should be the social ranking.

Arrow's theorem states that a ranking rule can fulfill IIA and unanimity only if it is a dictatorship. That is, there is a single fixed agent who determines the entire outcome.

1 Gibbard-Satterthwaite Theorem

There is a similar statement regarding voting rules. In this case, the "truthfulness" property is called *strategy-proofness*. It is defined as follows.

Definition 26.1. *A voting rule is* strategy-proof *if for all preference profiles $\succ_1, \dots, \succ_n$, all agents i, and candidates A and B the following holds. If $A \succ_i B$ and B wins under $\succ_1, \dots, \succ_n$, then A does not win under any false report $\succ_i'$ of agent i.*

So, in words, this definition requires that no agent can enforce an outcome that he likes better by lying. This, once again, is conceptually the same as truthfulness in mechanism design with money.

The following theorem was discovered by Allan Gibbard (1973) and Satterthwaite (1975) independently.

Theorem 26.2. *If a voting rule for three or more candidates is* onto *(that is, every candidate can be elected) and strategy-proof, then it is a dictatorship. That is, there is some agent i such that always agent i's most preferred candidate wins.*

The assumption that the voting rule is onto mirrors unanimity in Arrow's theorem. If a ranking rule fulfills unanimity, then, by definition, it can output every possible ranking, namely if all agents agree on this one. For a voting rule to be onto, a similar property is sufficient: If all agents agree that one candidate is the best, then this candidate gets elected.

We will not prove this theorem because the proof is not very enlightening on the technical level. The general argument is to use Arrow's theorem and to show that a voting rule fulfilling the assumptions of the theorem would be a contradiction to Arrow's theorem.

In more details, such a voting rule could be turned into a ranking rule as follows. The fact whether $A > B$ or not is determined by whether A would win a vote against B if they were both moved to the top of the preference profiles. From the assumptions, we get that this ranking rule fulfills unanimity and IIA. If the voting rule that we started from is not a dictatorship, then the created ranking rule isn't either: For every agent, there are preference profiles, such that some part of the social ranking does not correspond to the individual preference. This is a contradiction, therefore the voting rule cannot exist.

2 Is the Voting Problem Hopeless?

The two impossibility theorems might give the impression that voting is pointless per se because all voting rules are equally bad. This, however, is not really true. We have now adjusted our

expectations, seeing that voting will always to some extent be strategic. The other question, however, is still open: Who should win elections? How to we turn agents' preferences into the choice of one winner.

What should ideal voting or ranking rules fulfill? Let us introduce four examples.

- Anonymity: The identities of the agents do not matter. That is, the outcome on preference profile $(\succ_1, \ldots, \succ_n)$ is the same as on $(\succ_{\pi(1)}, \ldots, \succ_{\pi(n)})$ for any permutation $\pi\colon N \to N$.
- Monotonicity: If one agent moves a candidate up in the ranking, keeping everything else fixed, the candidate does not move down in the combined society's ranking.
- Condorcet winner criterion: If, in pairwise comparisons, one candidate is always preferred to any other candidate in the majority of of all preference profiles, it is the winner of the election. That is, if there is A such that for all B there are more i with $A \succ_i B$ than $B \succ_i A$, then A is the winner of the election.
- Condorcet loser criterion: If for some A for all B there are more i with $B \succ_i A$ than $A \succ_i B$, then A does not win the election.

3 Plurality Voting

Let us come back to our initial example of plurality voting. We choose the candidate who is ranked highest in the plurality of agents' preferences. This voting rule is anonymous and monotone.

Recall the following example, in which C is the winner.

$$35\,\% : A \succ_i B \succ_i C$$
$$25\,\% : B \succ_i A \succ_i C$$
$$40\,\% : C \succ_i A \succ_i B$$

Observe that A is the Condorcet winner. In A vs. B, he gets 75 % of the votes; in A vs. C, he gets 60 % of the votes. So, the Condorcet winner criterion is not fulfilled.

Furthermore, C is the Condorcet loser in this case. We observe that in C vs. A and also in C vs. B he gets 40 % of the votes. That is, the Condorcet loser criterion is not fulfilled either.

4 Borda Count and Positional Voting

A very general voting/ranking approach works as follows. Depending on the position in the preference list, each candidate gets a score from every agent. The social ranking is determined by ordering the candidates by sum of scores. Every vector of $a_1 \geq \ldots \geq a_m$ is such a positional voting rule. Plurality voting is recovered by $a_1 = 1, a_2 = \ldots, a_m = 0$. Again, any of these rules is anonymous and monotone.

Another common rule is the *Borda count*, in which one sets $a_1 = m, a_2 = m-1, \ldots, a_m = 1$. Let us consider our example from above:

$$35\,\% : A \succ_i B \succ_i C$$
$$25\,\% : B \succ_i A \succ_i C$$
$$40\,\% : C \succ_i A \succ_i B$$

If we have $n = 100$ voters, A gets a total score of $35 \times 3 + 25 \times 2 + 40 \times 2 = 235$, B gets a score of $35 \times 2 + 25 \times 3 + 40 \times 1 = 185$, C gets a score of $35 \times 1 + 25 \times 1 + 40 \times 3 = 180$.

Note that equivalently, we could perform pairwise votes between any two candidates and sum up the number of votes each candidate gets in his $m - 1$ pairwise comparisons.

Other voting rules follow the same scheme. For example, the Eurovision Song Contest uses $a_1 = 12, a_2 = 10, a_3, = 8, a_4 = 7, a_5 = 6, a_6 = 5, a_7 = 4, a_8 = 3, a_9 = 2, a_{10} = 1, a_{11} = \ldots = a_m = 0$, where agents correspond to countries participating in the vote.

While in our example above, indeed the Condorcet winner wins and the Condorcet loser loses, this is generally not true for any of these rules. Indeed, no such rule fulfills the Condorcet winner criterion. Consider the following example, in which B wins Borda count, despite the fact that A is the Condorcet winner.

$$60\,\% : A \succ_i B \succ_i C$$
$$40\,\% : B \succ_i C \succ_i A$$

We could also ask at this point whether the Condorcet winner criterion is actually desirable. In this particular example, A is polarizing: Some agents really dislike him, so B might be a reasonable compromise. The reason is that the criterion only makes a binary comparison. It does not matter how much an agent likes one candidate better than the other.

5 Copeland Rule

Yet another approach is to perform pairwise comparisons and to count how many a candidate wins. The *Copeland score* of a candidate is defined to be the difference of how many pairwise comparisons this candidate wins and how many he loses. So, it is an integer in the range of $-(m-1)$ to $m-1$. The candidates are then ranked by decreasing Copeland score.

This rule fulfills the Condorcet winner and loser criteria. A Condorcet winner always has the highest possible score of $m+1$, a Condorcet loser always has score $-(m-1)$, which is the lowest possible score.

6 Outlook

As we have seen, there is a plethora of different voting rules. In each of them, agents sometimes have an incentive to misreport their preferences. Crucially, the counterexamples are often cyclic. That is, there are preferences of the forms $A \succ_i B \succ_i C$, $B \succ_i C \succ_i A$, and $C \succ_i A \succ_i B$. One can show, for example, that there are non-trivial strategy-proof voting rules if preferences have more of a structure. Most prominently, they could correspond to a left-right political spectrum and therefore be not cyclic.

Regarding other properties of voting rules, there are some that are less disputable than others. There is clearly no single best voting rule, simply because there are multiple ways to define who should be the winner.