Game Theory and Algorithms*
Lecture 15: Elections and Arrow's Theorem

May 3, 2011

Summary: In an election, each voter (player) gives a ranking of the candidates (alternatives), and we have some rule (e.g., plurality vote) to combine these individual votes into a group decision on which candidate wins. An important main result we will cover today is Arrow's theorem: there are several reasonable properties of a voting system that cannot all hold simultaneously.

1 Voting Systems; Condorcet Winners

Typical elections around the world use a *plurality voting* system: each voter casts exactly one vote, and the candidate who gets the largest number of votes is declared the winner. However, other voting systems are sometimes used. This week we introduce the general setting of *voting systems*. We will introduce several concrete systems, and talk about properties that some have (or not).

To motivate our first definitions, consider the following example. Suppose there is an election and there are three candidates: A, B and Z. When we run the election using the standard plurality voting system, the votes are:

30% vote for A, 30% vote for B, 40% vote for Z.

Thus, Z wins the election. However, suppose we ask the voters for a little more information:

30% like A the most and B second, 30% like B the most and A second,
and 40% like Z the most and B second.

The following concept's namesake was an 18th-century French mathematician and politician.

Definition 1. Given a list of voters' preferences, a candidate X is a *Condorcet winner* if, for every other candidate Y, more than half of the voters prefer X to Y.

* Lecture Notes for a course given by David Pritchard at EPFL, Lausanne.

Therefore, in the example above, candidate B is a Condorcet winner: 70% of voters prefer B to A and 60% of voters prefer B to Z. This differs from A, the winner in plurality voting! The above type of example is not rare in three-party elections: the US presidential election between Gore-Bush-Nader in 2000 and the Canadian 2011 elections had roughly this form. (These real-life votes are further complicated by electoral districts which amalgamate votes, and which we won't discuss.)

We frame a variety of different voting systems as follows.

Definition 2. Given a finite set C of *candidates*, a *(strict linear) order* O is a permutation of the candidates, from best to worst. Whenever candidate A appears before B in the order, we write $A >_O B$ and interpret A as preferred to B.

Unless we say otherwise, we'll assume voters have this type of ordering for simplicity. Generally, we relax the output of the voting mechanism to allow ties:

Definition 3. A *weak linear order* O is similar but allows ties. In detail, $A \geq_O B$ and $B \geq_O C$ imply $A \geq_O C$; as well, $A =_O B$ and $B =_O C$ imply $A =_O C$. This is isomorphic to giving a real number to each candidate and comparing preference by these values.

We'll mention a third "quasi-transitive" type of order later.

Definition 4. A *voting system* is a function which takes n inputs — an order on candidates from each voter — and gives one output, which is another order on candidates. Unless we say otherwise, the inputs are strict orders, and the output is a weak order.

Thus so far, we have seen

- One voting system: plurality voting, which assigns one point to each voter's top choice, and orders the candidates by points.
- One property of voting systems: a system satisfies the *Condorcet criterion* if, whenever a Condorcet winner exists, the voting system makes them the winner.
- This voting system (plurality) does not have this property (Condorcet criterion).

Remark: In the literature, election system design is sometimes referred to as "mechanism design without money," capturing the idea that they must select an alternative based on group inputs, but without bids or payments.

1.1 Condorcet Cycles

If there is no Condorcet winner, (and assuming no pairwise ties) then there must be a *Condorcet cycle*: candidates $A_1, A_2, \ldots, A_k$ so that a majority of voters prefer A_1 to A_2, a majority of winners prefer A_2 to A_3, et cetera, and finally a majority of voters prefer A_k to A_1. For example, if

$$\text{voter 1: } A > B > C\text{; voter 2: } B > C > A\text{; voter 3: } C > A > B$$

then we have a Condorcet cycle since the majority (2/3) prefer A to B, B to C, and C to A. The moral of the story:

Pairwise "majority choices" made by society may not be transitively consistent.

1.1.1 Non-transitive dice

A factoid related to the existence of Condorcet cycles is the following dice game. We have three dice:

Die A : gives 1, 6, or 8, each with probability $\frac{1}{3}$.

Die B : gives 3, 5, or 7, each with probability $\frac{1}{3}$.

Die C : gives 2, 4, or 9, each with probability $\frac{1}{3}$.

Suppose we play the following game: you pick any die you like; then I pick a different die. We roll the dice and whoever rolls the higher number wins \$1 from the other player. I'll let you switch to any die (even mine) later if you like, and take another die. Is this a good game to play? In other words: for each pair of dice, calculate how often one beats the other one.

With a stretch, we can even imagine a connection to elections. Suppose party A, if elected, promises to perform one of three actions a_1, a_6, a_8, each with probability 1/3. Party B, if elected, does the same to a_3, a_5, a_7 with independent probabilities of 1/3 each, and similarly for C. You prefer action a_9 to a_8, etc. Which of the three parties will you vote for? (Your "preference over lotteries" is what you need to determine here.)

1.2 The Borda Count Method

Jean Charles de Borda, another 18th-century French mathematician, popularized the following method:

- Let k be the number of candidates.
- For each voter's ordering, we give their top choice k points, their second choice $k-1$ points, and so on, with their bottom choice receiving 1 point.
- We add up the points among all voters, and rank the candidates by the number of points they receive.

This system is fairly simple and has a lot of nice properties. It does not satisfy the Condorcet criterion: roughly, the problem is that the Condorcet criterion compares candidates by absolute preference whereas Borda awards points based on relative preference. A concrete example: if three voters prefer $A > B > C$ and two voters prefer $B > C > A$, then A is the Condorcet (pairwise) winner, but B wins the Borda Count (A gets $3 \cdot 3 + 2 \cdot 1 = 11$ points, B gets 12 points, and C gets 7 points).

Exercise. (a) Show that the Borda Count satifies the *Condorcet loser criterion*: if a candidate A strictly loses in pairwise comparison with every other candidate, then A is not the Borda winner. (b) Using part (a), show that Borda *never strictly opposes all pairwise comparisons*: there is always at least one pair of candidates A, B such that at least half of voters prefer A to B, and such that A gets at least as many points as B under Borda.

Definition 5. Generalizing the plurality vote and the Borda count, in a *point-based* election system with point vector $c = (c_1, c_2, \dots, c_k)$, each voter contributes c_1 points to their favourite candidate, c_2 points to their 2nd-favourite, and so on, with c_k points to their least-favourite. The output is to rank candidates by total points. We require that $c_1 \geq c_2 \geq \cdots \geq c_k$ and to exclude trivial systems, we require $c_1 \neq c_k$.

We have seen two examples: plurality has $c = (1, 0, \dots, 0)$ and Borda Count has $c = (k, k - 1, \dots, 1)$.

Exercise. Following up on the previous exercise: if c determines a point-based voting system, and c satisfies the *Condorcet loser criterion*, then prove c is essentially equivalent to the Borda count, in the sense that $c_1, c_2, \dots$ is an arithmetic sequence. (Hint: get a contradiction by focusing on a candidate who will violate the Condorcet loser criterion, and "average out" the other candidates.)

More strongly, a theorem of Saari (mentioned on [1, p. 51]) shows that Borda is the unique point-based system which never strictly opposes all pairwise comparisons.

2 Arrow's Theorem

Arrow's theorem (from 1951) considers several reasonable-looking properties of voting systems, and shows that it is impossible to satisfy them all simultaneously.

- A voting system is *unanimity-respecting* if, when all voters vote with the same preference list, the output of the system is this same list.
- Voter i is a *dictator* if the output of the system always equals voter i's list. If no dictator exists, the system is *non-dictatorial*.
- A voting system is *pairwise independent* if the output's ranking of every pair of candidates only depends on the input's n rankings of that pair of candidates.

Example 6 (Borda Count is not pairwise independent)**.** Take three candidates A, B, C, and five voters with preferences

$$\text{two voters prefer } A > B > C \text{ and three prefer } B > A > C.$$

On this input, the Borda Count gives $B > A > C$ (with points $13 > 12 > 5$). Now suppose the first two voters swap their preference between B and C, giving

$$\text{now, two voters prefer } A > C > B \text{ and three prefer } B > A > C.$$

Now the Borda Count gives $A > B > C$ (with points $12 > 11 > 7$). The relative input preferences of A and B did not change, but the relative output preference of A and B changed. So pairwise independence did not hold.

In fact, any reasonable system violates pairwise independence once there are more than 2 candidates:

Theorem 7 (Arrow)**.** *For three or more candidates, there is no unanimity-respecting, non-dictatorial, pairwise independent voting system.*

Proof. We prove the theorem only in the case of three candidates A, B, and C, and in a slightly weaker setting where we require the voting mechanism to output a *strict* order (no ties allowed). No major new ideas are needed to prove the stronger version.

We use a proof by contradiction, so assume a mechanism exists with all three properties. We will extensively use pairwise independence, rephrased in the following way: if I tell you all n voters' preferences between 2 candidates X and Y, then the output's preference between X and Y is uniquely determined.

Voter i making a *critical change between X and Y* is defined as follows: there exists some fixed inputs for all voters other than i, so that the X-Y output comparison when i adds their vote with $X > Y$ is different from the X-Y output comparison when i adds their vote with $Y > X$. Let P_{XY} be the set of all voters who can make a critical change.

Next, we claim that each critical change set P_{XY} is nonempty. Notice that if all voters prefer A to B then the output also prefers A to B (since it is unanimity-respecting) and vice-versa for preferring B to A. Imagine taking an input where all voters prefer A to B; then swap the n inputs from $A > B$ one at a time to $B > A$. At the start the output preferred A to B and at the end it preferred B to A. Thus at some point changing a single input affected the output; this is a *critical change*, as needed.

Suppose that $P_{AB} = P_{BC} = P_{CA} = \{i\}$ for some single voter i. Then we claim that i is a dictator, which is not hard to verify.

Hence, the set $P_{AB} \cup P_{AC} \cup P_{BC}$ has at least two voters. Without loss of generality, we have voter $1 \in P_{AB}$ and voter $2 \in P_{AC}$. By the definition of a critical change, there is a collection of $n-1$ pairwise A-B rankings for voters $\neq 1$ such that the A-B output depends on player 1's ranking of A and B; and there is a similar collection of $n-1$ pairwise A-C rankings for voters $\neq 2$. An important idea is that these collections can be combined while keeping the pairwise flexibility for voters 1 and 2:

- For voters $\neq 1, 2$ we can combine any A-B ranking and any A-C ranking into a linear order, e.g., $A > B$ and $A < C$ can be combined to $A > B > C$.
- For voter 1 (and similarly for voter 2), any fixed A-C ranking can be extended to two linear orders which differ only in the pairwise ranking of A and B. E.g., $A > C$ can be extended to $A > B > C$ or $B > A > C$.

So from now on, we think of all voters $\neq 1, 2$ as fixed, and we only allow "one bit" of flexibility to voters 1 and 2. So we'll consider 2×2 possible outcomes. This "one bit" of freedom means that voter 1 can either force the mechanism to rank A above B, or A below B, and similarly for voter 2.

Crucially, observe that in our 4 outcomes, the input rankings of B and C never change. So again by pairwise independence, all 4 outcomes give rise to the same pairwise ordering of B and C, without loss of generality $B > C$. The killer idea is that we can now force the output of the mechanism to be cyclic: voter 1's bit of freedom may be used to force the mechanism to rank A above B, while voter 2's bit of freedom may be used independently to force the mechanism to rank C above A. Since the output cannot be a ranking with all of $B > C, A > B, C > A$ holding, we are done. □

Remarks. Out of the three conditions, you should strongly suspect the last condition, pairwise independence, as being the main culprit in this impossibility theorem. Specifically, given the possibility of Condorcet cycles as shown previously, it is less surprising that pairs must interfere with each other in some way; Arrow's theorem basically says that we can't "squish out" the Condorcet cycles in one place without causing another problem somewhere else in the mechanism.

Arrow's theorem has a stronger relative, the Gibbard-Satterthwaite theorem, which roughly says that there is no strategy-proof voting system. One interpretation is that no voting system is perfect. Many decent voting systems exist, but different decent voting systems will give different outputs on the same inputs.

Exercise. Prove the stronger version of Arrow's theorem where we have 3 or more candidates, and where the output of the mechanism is allowed to be a weak order.

An interesting, and somewhat different, extension of Arrow's theorem pertains to *quasi-linear orders* which satisfy the following:

- Each pair A, B has $A > B, A < B$ or $A = B$;
- $A > B$ and $B > C$ imply $A > C$.

However, unlike weak linear orders, $A = B, B = C$ need not imply $A = C$. So $A = B > C = A$ is allowed, as well as $A = B > C = D > A$.

Definition 8. An *oligarchy* for voter set $R \subset \{1, \ldots, n\}$ is the following voting system: all votes by players outside of R are ignored; for each pair X, Y of candidates, if all players in R agree on the pairwise ranking of X and Y then we take this as the output, and otherwise the output ranks them pairwise $X = Y$.

Claim 9. *An oligarchy outputs a quasi-linear order, and satisfies the conditions of Arrow's theorem when* $|R| > 1$.

Theorem 10 (Generalized Arrow)**.** *Every voting system with a quasi-linear output, which is pairwise-independent and unanimity-respecting, is an oligarchy.*

In [1, p. 163] the origins of this theorem are attributed to Gibbard.

References

[1] D. G. Saari. *Decisions and Elections.* Cambridge University Press, 2001.