Voting Criteria

21-301 2018

30 April

1 Evaluating voting methods

In the last session, we learned about different voting methods. In this session, we will focus on the criteria we use to evaluate whether voting methods are doing their jobs; that is, the criteria we use to determine if a voting system is reasonable and fair. We use all notation as in the previous set of notes.

1. **Majority Criterion:** If a candidate gets $> 50\%$ of the first place votes, that candidate should be the winner.

   This criterion is quite straightforward. If a candidate is preferred by more than half of the voters, this candidate by all rights should win the election. No other candidate could be more satisfactory for more people than the first choice of more than 50% of the voters. It is immediately obvious that the plurality method and the runoff methods will satisfy this criterion, since their method of determining the winner is fundamentally based on first place votes. For other methods, the criterion is less clear.

2. **Condorcet Criterion:** If a candidate wins every pairwise comparison, then that candidate should be the winner.

   This should remind you of the Condorcet method for choosing a winner. Of course, it is immediately obvious that if the Condorcet method produces a winner without having to resolve ties, then that candidate wins the election in that method, so clearly the Condorcet method satisfies the Condorcet criterion. However, it is not clear that the other methods will satisfy this criterion; for example, in Hare method, is it possible to eliminate a candidate that would otherwise be a Condorcet winner?

3. **Monotonicity Criterion:** Suppose we run an election with a set of preferences $\{\leq_x \mid x \in X\}$ and yield a ranking $\leq$. Let $A \in \mathcal{C}$, and suppose that a new election is run, with rankings $\leq'_x$ in which the following is true:

   - For all $x \in X$, $\leq'_x = \leq_x$ when restricted to $\mathcal{C}\backslash\{A\}$. That is, the ranking of candidates other than A does not change.
   - For all $x \in X$, if $A \geq_x B$, then $A \geq'_x B$ for all $B \in \mathcal{C}$.

   Essentially, what we are doing here is allowing A to move up in the ranking of individual voters, and changing nothing else. The criterion states that under these changes, if $A \geq B$, then $A \geq' B$. That is, by improving A's standing with individual voters, we can only improve A's standing overall.

   It seems entirely illogical that if someone improves the position of a candidate, that can negatively impact their chance at success. Nonetheless, not all methods will satisfy this criterion. We saw one example of this in last week's work.

4. **Independence of Irrelevant Alternatives (IIA) Criterion:** Suppose $A \geq B$. If another election is held in which some voters change their preference for a different candidate, C, without changing the order in which A and B are ranked, and no other changes are made, then we should still have $A \geq B$.

This criterion can be a bit confusing, but essentially works like this. Suppose $A \geq B$. If we redraw the voting matrix with candidate C moved around, but in such a way that no voter changes their mind about whether they prefer A to B or vice versa, then we should still have $A \geq B$. The position of the other candidates (alternatives) should be irrelevant to how A and B are ranked.

Certainly the plurality method does not satisfy this criterion. Imagine the "spoiler" situation described before. We had the situation of a voting matrix such as the following:

$$\begin{array}{ccc} 8 & 6 & 5 \\ \hline A & B & C \\ B & C & B \\ C & A & A \end{array}$$

Here, clearly A wins the plurality contest. However, if some candidates changed their preferences regarding B and C, A could lose the election. Hence, plurality voting does not satisfy the criterion of irrelevant alternatives.

5. **Pareto Criterion:** If everyone prefers candidate A over candidate B, then candidate A should be ranked above candidate B.

   For example:

$$\begin{array}{ccc} 3 & 10 & 4 \\ \hline A & B & A \\ B & A & C \\ C & C & B \end{array}$$

   Notice that in every column, in every voter's preference, candidate C fares worse than candidate A. Since C is always worse than A, we would want C to be ranked below A overall. Indeed, if C were to be ranked above A, there is an obvious argument that something is wrong with this ranking, since literally everyone says so.

2 How do our voting methods stack up?

For each of the voting methods we discussed last time, consider whether they satisfy the criteria described above. Fill in the chart with Y/N, where a Y indicates that the method does satisfy the criteria, and N indicates that it does not. Some boxes are filled in based on discussions and examples from above. If the answer to a particular cell is no, find a voting matrix that demonstrates a failure to satisfy the criterion under that method. If the answer is yes, sketch a proof.

	Plurality	Runoff	Hare	Borda	Condorcet
Majority Criterion	Y	Y	Y		
Condorcet Criterion					Y
Monotonicity					
Irrelevant Alternatives	*N*				
Pareto					

3 Arrow's Impossibility Theorem

You may have noticed that in the chart in the previous section, there was no voting method that satisfied all of the voting criteria simultaneously. You might, optimistically, think that maybe that's cuz we haven't found the perfect method yet, and that out there, somewhere, is a wonderful voting method that will satisfy all the good rules and be awesome. And, unfortunately, you are wrong.

I'm going to add here one more criterion, so that we can discuss the theorem that says you're wrong. This criterion is

6. **No Dictators Criterion:** No single voter has the power to always determine the group's preference. That is to say, if one voter likes candidate X, but every other voter does not, then candidate X should not win.

None of the voting methods we have yet described have a dictator, but certainly voting methods with a dictator are easy to develop: Take the voting matrix, and then throw it away and ask the one guy.

We also will work with a variant of the Pareto Criterion to prove the theorem. For the proof, we will assume that the voting method doesn't just produce a winner, it produces a second, third, fourth place, etc. That is, we don't just get a winner but we also get a ranking of all the other candidates. The "Weak Pareto Efficiency" criterion states that if A is ranked above B in a head-to-head contest, then A should be ranked above B in the ranking of the candidates. This is sometimes called "transitivity," since it behaves exactly like transitivity in basic ordering. Note that plurality voting can fail transitivity, but some voting methods, like Borda count, will respect it.

Now, we are ready for the statement of Arrow's Impossibility Theorem.

Theorem 1 (Arrow's Theorem.)**.** *Suppose you have a voting system that satisfies transitivity, independence of irrelevant alternatives criterion, and the Pareto criterion. Then your voting system is a dictatorship.*

We will prove Arrow's Theorem in a minute. First, a simpler statement that is perhaps a little easier to digest.

Theorem 2 (Arrow's Theorem: simplified)**.** *Suppose you have a voting system that is not a dictatorship. Then the system cannot satisfy both the Condorcet criterion and the independence of irrelevant alternatives criterion.*

Sometimes this theorem is phrased in popular literature as stating "no voting system is perfectly fair." This is perhaps a simplistic reading of the theorem, since the theorem itself only deals with certain kinds of voting. Indeed, Arrow's Theorem refers to voting systems in which voters rank all the candidates, which as we know do not happen in most real-life situations. Nevertheless, the fundamental reading is basically accurate: if you were to wish to design a voting system that would respect all measures of fairness, you can't. You will necessarily fail at least one of the basic criteria we use to evaluate a system's fairness. By matter of necessity, then, a nonquantitative element must introduce itself: which criteria do we value more highly, and can we design a system that respects those criteria while perhaps violating others?

Before we go to the full proof of Arrow's Theorem, let's do a quick proof of the simplified version. We will do so by looking at a counterexample.

Proof of Theorem 2. Suppose we have a voting system that satisfies both the Condorcet criterion and the independence of irrelevant alternatives criterion. Consider how the voting system would evaluate this vote matrix:

$$\begin{array}{ccc} 1 & 1 & 1 \\ \hline A & C & C \\ B & B & A \\ C & A & B \end{array} \tag{1}$$

Notice that C is certainly a Condorcet winner, and therefore C must win the election. Hence, it must be that our vote ranking places C above A.

By independence of irrelevant alternatives, we should be able to move around candidate B, and not impact the ranking of C relative to A. Hence, it should be that under the following vote matrix, C is still above A:

$$\begin{array}{ccc} 1 & 1 & 1 \\ \hline A & B & C \\ B & C & A \\ C & A & B \end{array} \tag{2}$$

This is immediate, since we have only changed the position of candidate B in the second column. Therefore, A is a nonwinner in matrix (2).

On the other hand, consider the following vote matrix:

$$\begin{array}{ccc} 1 & 1 & 1 \\ \hline B & B & C \\ A & C & A \\ C & A & B \end{array} \tag{3}$$

Notice that here, C is a nonwinner as B is the Condorcet winner. Hence, if we move around candidate A, it should be the case that C is still a nonwinner. Therefore, if we switch the order of candidates A and B in the first column (thereby attaining matrix (2) again), we must have that C is a nonwinner.

Finally, we have one last matrix:

$$\begin{array}{ccc} 1 & 1 & 1 \\ \hline A & B & A \\ B & C & C \\ C & A & B \end{array} \tag{4}$$

In this case, B is a nonwinner, but by switching A and C in the last column, we obtain matrix (2) yet again.

Hence, we have a matrix, (2), under which no candidate is permitted to win! This is inconsistent, and hence our voting system does not exist (unless it is a dictatorship).

□

Now, let us look at the proof of Arrow's Theorem.

Proof of Theorem 1. Let's assume we have a voting system that satisfies transitivity, independence of irrelevant alternatives, and Pareto. We will prove that this system is a dictatorship.

Recall our plan: we will use our voting system not just to select a winner, but to provide a full ranking list. This could be done with any of our existing methods by selecting a winner, then removing that candidate and selecting a winner from the remaining (second place), etc. We will write $X > Y$ to indicate that candidate X appears above candidate Y in the ranking list. Our first observation is as follows:

Claim 1. *Choose some candidate X. If every voter ranks X either best or worst, then the full ranking of candidates puts X either best or worst.*

Note in this claim that not every voter does the same thing with respect to candidate X. Some think he's the best, and some think he's the worst. Nevertheless, if all votes for X are extreme, then X must appear on the extremes of our full ranking.

Proof of Claim 1. Suppose the claim is not true, so that every voter ranks X either best or worst, but X is not best or worst on the full ranking. Then there are other candidates, Y and Z, with $Y > X > Z$.

Make a new voter matrix as follows. If a voter has X at the top of their list, move Z up to second place, so that voter ranks Z above Y. If a voter has X at the bottom of their list, move Z up to first place, so that the voter also ranks Z above Y.

Notice that we have not changed the relationship between Y and X at all for any voter; if X is preferred to Y in the original ranking, it still is, and vice versa. Hence, by independence of irrelevant alternatives, it must be the case that $Y > X$ still holds in the new ranking system.

Likewise, we have not changed the relationship between X and Z for any voter, so $X > Z$ in the new ranking system.

However, we have forced that Z is ranked strictly better than Y for every voter, so we have $Z > Y$ by the Pareto criterion. But now we have violated transitivity, since we cannot have $Y > X > Z > Y$ in a valid ranking. □

Now, select any candidate A. Suppose we have a vote matrix in which every voter ranks A last: call this matrix M. Order the voters as $1, 2, 3, \ldots n$ in any way you like.

By the Pareto criterion, we know that candidate A must be ranked last in our ranking. Let's make a new vote matrix, which we call M_i, by changing the votes of voters $1, 2, \ldots, i$ by moving A from last to first.

Notice that for all of the M_i, we satisfy the conditions of Claim 1, so it must be the case that A is ranked either last or first for every M_i. Moreover, by the time we get to M_n, everyone has A ranked first, so the Pareto criterion says that A should be first in our ranking. Hence, we can select a particular voter, voter j, so that M_j is the first matrix on this list $M_1, M_2, \ldots, M_n$ having A ranked in first place. This voter j has a lot of power: her vote is pivotal in the ranking of candidate A. Let's call her Jen.

We actually can show a lot more. Jen is, in fact, a dictator. We'll write $>_j$ to indicate Jen's personal preferences, and show that our whole ranking system has to look exactly like $>_j$.

Pick any two candidates B, C other than A. Suppose that $B >_j C$, so Jen prefers candidate B to candidate C. Starting from M, let's make a new vote matrix as follows:

- Set B at the top of Jen's list, and A second.
- For voters $1, 2, \ldots, j-1$ (everyone before Jen), put A in first place.

Notice that we haven't changed the relationship between B and C in anyone's voter list. For everyone except Jen, all we did was move A. For Jen, we kept B and C in the same order, since Jen already preferred B to C. In fact, this looks exactly the same as M_j, except that we have moved B to the top of Jen's list. Hence, in this new list, the relationship between B and C should be the same as in M_j, due to independence of irrelevant alternatives. So our goal is to determine the relationship between B and C in this new list. We will do so by going through the chosen candidate A.

Notice that in our new matrix, everyone has the same relationship between B and A as they do in M_{j-1}. Indeed, the only difference between M_{j-1} and the new ranking is in Jen's scores, and she has kept B above A here as in M_{j-1}. Hence, it must be that we preserve the relationship between B and A that we had in M_{j-1}. But we know that in M_{j-1}, candidate A is in last place, so it must be that $B > A$ in the new matrix.

On the other hand, the new matrix preserves the relationship between C and A that we saw in M_j. Hence, it must be that we have the same ordering between C and A. Moreover, A is in first place in M_j, so it must be that $A > C$ in the new matrix.

Therefore, since $B > A$ and $A > C$, we have by transitivity that $B > C$ in the new ranking. By the observation above, the new ranking agrees completely with M_j, so in M_j, we also have $B > C$, just as Jen wanted.

Ok, so Jen's rankings are preserved for every candidate other than candidate A. What about this candidate though? It turns out to be straightforward to consider.

Suppose we chose someone other than A at the beginning, say X, and repeated this whole process. Then we would get a new voter, say Tim, who got to be the dictator, and Tim's preferences would be totally preserved for any candidate other than X. But that means that Tim's preferences are always preserved when it comes to candidate A, so if Tim puts A first, then A will be first. On the other hand, only Jen gets to say if A goes first, which we already showed. Hence, Tim and Jen must actually be the same person!

Therefore, Jen is a dictator, and the ranking system must look exactly like Jen's ranking.

□

4 Some Thinking Exercises

1. Do you think all of the vote criteria laid out in this document are useful? This is a nonquantitative question.
2. What assumptions did we need about our voting system for Arrow's Theorem? Do you think those assumptions are reasonable? Which ones might you consider relaxing?
3. In the simplified version of Arrow's Theorem, we relied on independence of irrelevant alternatives to manipulate our voting matrix. Do you think there's a relaxation of the rules for this criterion that might make a better choice?
4. Given what you found out about independence of irrelevant alternatives in Section 2, do you think it's a reasonable criterion? Why/why not?