Computational Social Choice 2020

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

[http://www.illc.uva.nl/~ulle/teaching/comsoc/2020/]

Plan for Today

We first are going to review a few more voting rules and remark on some surprising shortcomings of what look like reasonable rules.

To help us choose a good voting rule from amongst the plentitude of options, we then discuss two types of *characterisation results:*

- *normative characterisation:* using the axiomatic method to single out voting rules that satisfy certain intuitively appealing properties
- *epistemic characterisation:* using probabilistic modelling to single out voting rules most likely to recover the "truly" best alternative

For full details see Zwicker (2016) and Elkind and Slinko (2016).

W.S. Zwicker. Introduction to the Theory of Voting. In F. Brandt et al. (eds.), *Handbook of Computational Social Choice*. CUP, 2016.

E. Elkind and A. Slinko. Rationalizations of Voting Rules. In F. Brandt et al. (eds.), *Handbook of Computational Social Choice*. CUP, 2016.

Preview: Some Axioms

We are going to use these *axioms* to highlight certain shortcomings of some of the voting rules we have seen and are going to see:

- *Participation Principle:* It should be in the best interest of voters to participate: voting truthfully should be no worse than abstaining.
- *Pareto Principle:* There should be no alternative that every voter strictly prefers to the alternative selected by the voting rule.
- *Condorcet Principle:* If there is an alternative that is preferred to every other alternative by a majority of voters, then it should win.

Reminder: The Model

Fix a finite set $A = \{a, b, c, \ldots\}$ of *alternatives*, with $|A| = m \geqslant 2$.

Let $\mathcal{L}(A)$ denote the set of all strict linear orders R on A. We use elements of $\mathcal{L}(A)$ to model (true) *preferences* and (declared) *ballots*.

Each member i of a finite set $N = \{1, \ldots, n\}$ of *voters* supplies us with a ballot R_i, giving rise to a *profile* $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathcal{L}(A)^n$.

A *voting rule* (or *social choice function*) for N and A selects (ideally) one or (in case of a tie) more winners for every such profile:

$$F : \mathcal{L}(A)^n \to 2^A \setminus \{\emptyset\}$$

If $|F(\boldsymbol{R})| = 1$ for all profiles $\boldsymbol{R}$, then F is called *resolute*.

Reminder: Some Voting Rules

So far we saw the following voting rules:

- Positional scoring rules: Borda, plurality, antiplurality, k-approval
- Based on majority graph: Copeland, Slater
- Based on weighted majority graph: Kemeny, ranked-pairs, (Borda)
- Plurality with runoff (generalisation to follow)

Runoff Methods: Single Transferable Vote & Variants

STV (used, e.g., in Australia) works in stages:

- If some alternative is top for an *absolute majority*, then it wins.
- Otherwise, the alternative ranked at the top by the fewest voters (the plurality loser) gets *eliminated* from the race.
- Votes for eliminated alternatives get *transferred:* delete removed alternatives from ballots and "shift" rankings (i.e., if your 1st choice got eliminated, then your 2nd choice becomes 1st).

Various options for how to deal with *ties* during elimination.

In practice, voters need not be required to rank all alternatives (non-ranked alternatives are assumed to be ranked lowest).

For three alternatives, STV and *plurality with runoff* coincide.

<u>Variants:</u> *Coombs*, *Baldwin*, *Nanson* (different elimination criteria)

The No-Show Paradox

Under plurality with runoff (and thus under STV), it may be better to abstain than to participate and vote for your favourite alternative!

25 voters: $a \succ b \succ c$
46 voters: $c \succ a \succ b$
24 voters: $b \succ c \succ a$

Given these voter preferences, b gets eliminated in the first round, and c beats a 70:25 in the runoff.

Now suppose two voters from the first group abstain:

23 voters: $a \succ b \succ c$
46 voters: $c \succ a \succ b$
24 voters: $b \succ c \succ a$

a gets eliminated, and b beats c 47:46 in the runoff.

P.C. Fishburn and S.J Brams. Paradoxes of Preferential Voting. *Mathematics Magazine*, 1983.

Cup Rules via Voting Trees

We can define a voting rule via a *binary tree*, with the alternatives labelling the leaves, and an alternative progressing to a parent node if it beats its sibling in a *majority contest*.

Two examples for such *cup rules* and a possible profile of ballots:

```
(1)

        o
       / \
      o   c
     / \
    a   b
```

```
(2)      o
        / \
       /   \
      o     o
     / \   / \
    a   b b   c
```

$\mathtt{a} \succ \mathtt{b} \succ \mathtt{c}$

$\mathtt{b} \succ \mathtt{c} \succ \mathtt{a}$

$\mathtt{c} \succ \mathtt{a} \succ \mathtt{b}$

Rule (1): c wins

Rule (2): a wins

Cup Rules and the Pareto Principle

The (weak) *Pareto Principle* requires that we should never elect an alternative that is strictly dominated in *every* voter's ballot.

Cup rules do not always satisfy this most basic principle!

```
        o
       / \
      o   d
     / \
    o   a
   / \
  b   c
```

Consider this profile with three voters:

Ann: $\mathtt{a} \succ \mathtt{b} \succ \mathtt{c} \succ \mathtt{d}$

Bob: $\mathtt{b} \succ \mathtt{c} \succ \mathtt{d} \succ \mathtt{a}$

Cindy: $\mathtt{c} \succ \mathtt{d} \succ \mathtt{a} \succ \mathtt{b}$

d *wins!* (despite being dominated by c)

What happened? Note how this "embeds" the Condorcet Paradox, with every occurrence of c being replaced by $\mathtt{c} \succ \mathtt{d}$ ...

Condorcet Extensions

An alternative that beats every other alternative in pairwise majority contests is called a *Condorcet winner*. Sometimes there is no CW.

The *Condorcet Principle* says that, if it exists, only the CW should win. Voting rules that satisfy this principle are called *Condorcet extensions*.

Exercise: *Show that Copeland, Slater, Kemeny, and cup rules are CEs.*

Two further Condorcet extensions:

- *Young:* Elect alternative x that minimises the number of voters we need to remove before x becomes the Condorcet winner.
- *Dodgson:* Elect alternative x that minimises the number of swaps of adjacent alternatives in the profile we need to perform before x becomes the Condorcet winner. (Note difference to Kemeny!)

Trivia: Dodgson is also known as Lewis Carroll (*Alice in Wonderland*).

Positional Scoring Rules and the Condorcet Principle

Consider this example with three alternatives and seven voters:

3 voters: $a \succ b \succ c$
2 voters: $b \succ c \succ a$
1 voter: $b \succ a \succ c$
1 voter: $c \succ a \succ b$

So a is the *Condorcet winner:* a beats both b and c (with 4 out of 7). But any *positional scoring rule* makes b win (because $s_1 \geqslant s_2 \geqslant s_3$):

$$\begin{aligned} a&: \quad 3 \cdot s_1 + 2 \cdot s_2 + 2 \cdot s_3 \\ b&: \quad 3 \cdot s_1 + 3 \cdot s_2 + 1 \cdot s_3 \\ c&: \quad 1 \cdot s_1 + 2 \cdot s_2 + 4 \cdot s_3 \end{aligned}$$

Thus, *no positional scoring rule* for three (or more) alternatives can possibly satisfy the *Condorcet Principle*.

Fishburn's Classification

Can classify voting rules on the basis of the *information* they require. The best known such classification is due to Fishburn (1977):

- *C1:* Winners can be computed from the *majority graph* alone.

  Examples: Copeland, Slater

- *C2:* Winners can be computed from the *weighted majority graph* (but not from the majority graph alone).

  Examples: Kemeny, ranked-pairs, Borda

- *C3:* All other voting rules. Examples: Young, Dodgson, STV

Remark: Fishburn originally intended this for Condorcet extensions only, but the concept also applies to all other voting rules.

P.C. Fishburn. Condorcet Social Choice Functions. *SIAM Journal on Applied Mathematics*, 1977.

Nonstandard Ballots

We defined voting rules over profiles of strict linear orders (even if some rules, e.g., plurality, don't use all information). Other options:

- *Approval voting:* You can approve of any subset of the alternatives. The alternative with the most approvals wins.
- *Even-and-equal cumulative voting:* You vote as for AV, but 1 point gets split evenly amongst the alternatives you approve.
- *Range voting:* You vote by dividing 100 points amongst the alternatives as you see fit (as long every share is an integer).
- *Majority judgment:* You award a grade to each of the alternatives ("excellent", "good", etc.). Highest median grade wins.

The most important of these is approval voting.

Remark: *k-approval* and *approval voting* are very different rules!

Normative Characterisation: The Axiomatic Method

So many different voting rules! *How do you choose?*

One approach is to use the *axiomatic method* to identify voting rules of *normative* appeal. We will see one example for a classical result.

Axioms: Anonymity and Neutrality

Two basic fairness requirements for a voting rule F:

- F is *anonymous* if $F(R_1, \ldots, R_n) = F(R_{\pi(1)}, \ldots, R_{\pi(n)})$ for any profile $(R_1, \ldots, R_n)$ and any permutation $\pi : N \to N$.
- F is *neutral* if $F(\pi(\boldsymbol{R})) = \pi(F(\boldsymbol{R}))$ for any profile $\boldsymbol{R}$ and any permutation $\pi : A \to A$ (with π extended to profiles and sets of alternatives in the natural manner).

Thus: A is symmetry w.r.t. voters. N is symmetry w.r.t. alternatives.

Axiom: Positive Responsiveness

Notation: Write $N^{\boldsymbol{R}}_{x \succ y} = \{i \in N \mid (x, y) \in R_i\}$ for the set of voters who rank alternative x above alternative y in profile $\boldsymbol{R}$.

A (not necessarily resolute) voting rule satisfies *positive responsiveness* if, whenever some voter raises a (possibly tied) winner $x^\star$ in her ballot, then $x^\star$ will become the *unique* winner. Formally:

> F is *positively responsive* if $x^\star \in F(\boldsymbol{R})$ implies $\{x^\star\} = F(\boldsymbol{R'})$ for any alternative $x^\star$ and any two *distinct* profiles $\boldsymbol{R}$ and $\boldsymbol{R'}$ s.t. $N^{\boldsymbol{R}}_{x^\star \succ y} \subseteq N^{\boldsymbol{R}'}_{x^\star \succ y}$ and $N^{\boldsymbol{R}}_{y \succ z} = N^{\boldsymbol{R'}}_{y \succ z}$ for all $y, z \in A \setminus \{x^\star\}$.

Thus, this is a monotonicity requirement (we'll see others later on).

May's Theorem

When there are only *two alternatives*, then all the voting rules we have seen coincide with the *simple majority rule*. Good news:

Theorem 1 (May, 1952) *A voting rule for two alternatives satisfies anonymity, neutrality, and positive responsiveness if and only if that rule is the simple majority rule.*

This provides a good justification for using this rule (arguing in favour of "majority" directly is harder than arguing for anonymity etc.).

K.O. May. A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions. *Econometrica*, 1952.

Proof Sketch

Clearly, the simple majority rule satisfies all three properties. ✓

Now for the other direction:

Assume the number of voters is *odd* (other case: similar) $\rightsquigarrow$ no ties.

There are two possible ballots: $a \succ b$ and $b \succ a$.
Anonymity $\rightsquigarrow$ only *number of ballots* of each type matters.

Consider all possible profiles $\boldsymbol{R}$. Distinguish two cases:

- Whenever $|N^{\boldsymbol{R}}_{a \succ b}| = |N^{\boldsymbol{R}}_{b \succ a}| + 1$, then only a wins.
  By *PR*, a wins whenever $|N^{\boldsymbol{R}}_{a \succ b}| > |N^{\boldsymbol{R}}_{b \succ a}|$. By *neutrality*, b wins otherwise. But this is just what the simple majority rule does. ✓
- There exist a profile $\boldsymbol{R}$ with $|N^{\boldsymbol{R}}_{a \succ b}| = |N^{\boldsymbol{R}}_{b \succ a}| + 1$, yet b wins.
  Suppose one a-voter switches to b, yielding $\boldsymbol{R}'$. By *PR*, now only b wins. But now $|N^{\boldsymbol{R}'}_{b \succ a}| = |N^{\boldsymbol{R}'}_{a \succ b}| + 1$, which is symmetric to the earlier situation, so by *neutrality* a should win. Contradiction. ✓

Epistemic Characterisation: Voting as Truth-Tracking

An alternative interpretation of "voting":

- There exists an objectively "correct" ranking of the alternatives.
- The voters want to identify the correct ranking (or winner), but cannot tell with certainty which ranking is correct. Their ballots reflect what they believe to be true.
- We want to estimate the most likely ranking (or winner), given the ballots we observe. *Can we use a voting rule to do this?*

Example: Ballots as Noisy Signals

Consider the following scenario:

- two alternatives: a and b
- either $a \succ b$ or $b \succ a$ (we don't know which and have no priors)
- 20 voters/experts with probability 75% each of getting it right

Now suppose we observe that 12/20 voters say $a \succ b$.
What can we infer, given this observation (let's call it E)?

- Probability for E to happen *in case $a \succ b$* is correct:

$$P(E \mid a \succ b) \quad = \quad \binom{20}{12} \cdot 0.75^{12} \cdot 0.25^{8}$$

- Probability for E to happen *in case $b \succ a$* is correct:

$$P(E \mid b \succ a) \quad = \quad \binom{20}{8} \cdot 0.75^{8} \cdot 0.25^{12}$$

Thus: $P(E \mid a \succ b) \, / \, P(E \mid b \succ a) \; = \; 0.75^4/0.25^4 \; = \; 81.$
From Bayes and assuming uniform priors $[P(a \succ b) = P(b \succ a)]$:
Given E, a being better is 81 times as likely as b being better.

The Condorcet Jury Theorem

For the case of two alternatives, the simple majority rule is the best choice also under the truth-tracking perspective:

Theorem 2 (Condorcet, 1785) *Suppose a jury of n voters need to select the better of two alternatives and each voter independently makes the correct decision with the same probability $p > \frac{1}{2}$. Then the probability that the simple majority rule returns the correct decision increases monotonically in n and approaches 1 as n goes to infinity.*

Proof sketch: By the law of large numbers, the number of voters making the correct choice approaches $p \cdot n > \frac{1}{2} \cdot n$. ✓

For modern expositions see Nitzan (2010) and Young (1995).

Writings of the Marquis de Condorcet. In I. McLean and A. Urken (eds.), *Classics of Social Choice*, University of Michigan Press, 1995.

S. Nitzan. *Collective Preference and Choice*. Cambridge University Press, 2010.

H.P. Young. Optimal Voting Rules. *Journal of Economic Perspectives*, 1995.

Characterising Voting Rules via Noise Models

For n alternatives, Young (1995) shows that, if the probability of a voter to rank a given pair correctly is $p > \frac{1}{2}$, then the voting rule selecting the most likely winner coincides with the *Kemeny rule*.

Conitzer and Sandholm (2005) ask a general question:

> F*or a given voting rule F, can we design a "noise model" such that F is a maximum likelihood estimator for the winner?*

H.P. Young. Optimal Voting Rules. *Journal of Economic Perspectives*, 1995.

V. Conitzer and T. Sandholm. Common Voting Rules as Maximum Likelihood Estimators. UAI-2005.

E. Elkind and A. Slinko. Rationalizations of Voting Rules. In F. Brandt et al. (eds.), *Handbook of Computational Social Choice*. CUP, 2016.

The Borda Rule as a Maximum Likelihood Estimator

It *is* possible for the Borda rule:

Proposition 3 (Conitzer and Sandholm, 2005) *If each voter independently ranks the true winner at position k with probability $\frac{2^{m-k}}{2^m-1}$, then the maximum likelihood estimator is the Borda rule.*

<u>Proof:</u> Let $r_i(x)$ be the position at which voter i ranks alternative x.

Probability to observe the actual ballot profile if x is the true winner:

$$\frac{\prod_{i\in N} 2^{m-r_i(x)}}{(2^m-1)^n} = \frac{2^{\sum_{i\in N} m-r_i(x)}}{(2^m-1)^n} = \frac{2^{\mathrm{BordaScore}(x)}}{(2^m-1)^n}$$

Hence, x has maximal likelihood of being the true winner <u>iff</u> x has a maximal Borda score. ✓

V. Conitzer and T. Sandholm. Common Voting Rules as Maximum Likelihood Estimators. UAI-2005.

Summary

We have by now seen a very large number of voting rules:

- they explore different *intuitions* about how voting "should" work and they seem to sometimes suffer from *counterintuitive problems*
- they differ in view of the *profile information* they require
- they differ in view of their *computational requirements*

We then saw two approaches to characterising a voting rule:

- as the only rule satisfying certain *axioms* (normative desiderata)
- as optimally *tracking the truth* (under some probabilistic model)

What next? More applications of the axiomatic method.