TTIC 31260 Algorithmic Game Theory

Intro to Social Choice Theory

Your Guide: Avrim Blum

Readings: Chapter 9.2

[Credit to Ariel Procaccia for most of 1st half of the presentation]

Social Choice

- General topic: how to aggregate opinions and preferences
- Theoretical study goes back to 1700s (Condorcet, Borda)
- Even interesting contributions by Charles Dodgson (Lewis Carroll)
- Basically: theory of voting, contests, etc.

The Formal Setup

m

1	2	3
a	a	c
c	c	b
b	b	a

n

- Have a set V of n voters $\{1,2,\ldots,n\}$.
- Have a set A of m alternatives (candidates) $\{a,b,c,\ldots\}$
- Each voter has a ranking over the m alternatives
- Notation: $a \succ_i b$ means that voter i prefers a to b.
- A mapping from a set of n rankings to a single ranking is called a social welfare function, and a mapping from n rankings to a single alternative is called a social choice function or voting rule.

Some voting rules

1	2	3
a	a	c
c	c	b
b	b	a

m

n

Plurality

- Each voter votes for their top choice
- Alternative with the most votes wins
- Used in most elections

Borda Count

- Each voter gives m points to top choice, $m-1$ to next choice, etc.
- Alternative with the most points wins.
- Proposed by Jean-Charles de Borda in 1770 but perhaps goes back to 1400s.
- Used in: Eurovision, Slovenia (partially)

More voting rules

1	2	3
a	a	c
c	c	b
b	b	a

m

n

Veto

- Each voter vetoes (votes against) their lowest choice
- Alternative with the fewest vetoes wins
- Choosing a place to get lunch?

Positional Scoring Rules more generally:

- Defined by a vector $(s_1, \ldots, s_m)$.
- Voter gives s_i points to their ith choice. Alternative with most points wins.
  - Plurality: (1,0,0,…,0)
  - Borda count: $(m, m-1, m-2, \ldots)$
  - Veto: (1,1,…,1,0)

Other kinds of voting rules

1	2	3
a	a	c
c	c	b
b	b	a

m

n

- Terminology: we say that a beats b in a pairwise election if the majority of voters prefer a to b.

Plurality with runoff:

- In the first round, the two alternatives with the highest plurality score survive (the two with the most votes, where each voter just votes for one candidate)
- Second round is a runoff between those two (only needed if there was no strict majority).
- Used in US in some primary elections.

Other kinds of voting rules

Single Transferable Vote (STV):

- $m - 1$ rounds.
- In each round, voters choose their favorite and the candidate with the fewest votes is eliminated.
- Used in several countries and Cambridge, MA.
- Typically implemented by voters giving a ranking and then the rest is done internally in the system.

1	2	3
a	a	c
c	c	b
b	b	a

m

n

STV: EXAMPLE

2 voters	2 voters	1 voter
a	b	c
b	a	d
c	d	b
d	c	a

2 voters	2 voters	1 voter
a	b	c
b	a	b
c	c	a

2 voters	2 voters	1 voter
a	b	b
b	a	a

2 voters	2 voters	1 voter
b	b	b

[Slide from Ariel Procaccia]

Axiomatic approach to analyzing voting methods

Idea:

- Define some reasonable axioms you'd like a voting system to satisfy.
- See which voting systems satisfy them.
- See if it's possible for any voting system to satisfy them.
- Important impossibility results: Arrow's theorem, Gibbard-Satterthwaite theorem.
- GS theorem: When there are $m \geq 3$ alternatives, for any onto, non-dictatorial voting rule, there will exist scenarios where someone would regret voting truthfully.
- For $m = 2$ candidates, plurality is fine. For $m \geq 3$ candidates, "random dictator" rule is incentive-compatible (everyone writes their favorite choice on a piece of paper, put pieces in a hat, select randomly).

Some axioms

Majority consistency: if a majority of voters rank some alternative x first, then x should be the winner.

- Plurality? Y
- Borda count? N
- Plurality with runoff? Y
- Veto? N
- STV? Y

Some axioms

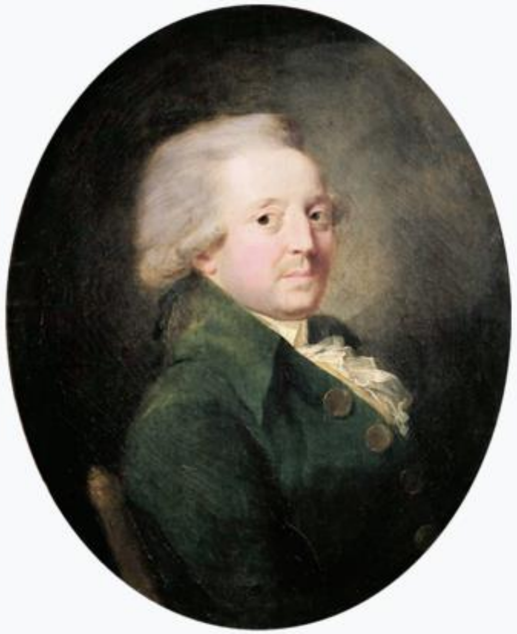
Condorcet consistency: if x beats every other candidate in a pairwise election (x is a Condorcet Winner), then x should win.

Marie Jean Antoine Nicolas de Caritat, Marquis of Condorcet (French: [maʁi ʒɑ̃ ɑ̃twan nikɔla də kaʁita maʁki də kɔ̃dɔʁsɛ]; 17 September 1743 – 29 March 1794), known as **Nicolas de Condorcet**, was a French philosopher and mathematician.[2] His ideas, including support for a free markets, public education, constitutional government, and equal rights for women and people of all races, have been said to embody the ideals of the Age of Enlightenment, of which he has been called the "last witness",[3] and Enlightenment rationalism. A critic of the constitution proposed by Marie-Jean Hérault de Séchelles in 1793, the Convention Nationale — and the Jacobin faction in particular — voted to have Condorcet arrested. He died in prison after a period of hiding from the French Revolutionary authorities.

Early years [edit]

Condorcet was born in Ribemont (in present-day Aisne), descended from the ancient family of Caritat, who took their title from the town of Condorcet in Dauphiné, of which

Nicolas de Condorcet



[Wikipedia]

Some axioms

Condorcet consistency: if x beats every other candidate in a pairwise election (x is a Condorcet Winner), then x should win.

Note: there may not be a Condorcet Winner, a fact known as Condorcet's paradox.

1	2	3
a	b	c
b	c	a
c	a	b

m

n

Some axioms

Condorcet consistency: if x beats every other candidate in a pairwise election (x is a Condorcet Winner), then x should win.

Does this example have a Condorcet Winner? Y

Voting systems satisfying CC:

- Plurality? N
- STV? N
- Borda count? N Veto? N ...
- Hmm....

1	2	3
b	c	d
a	a	a
c	d	b
d	b	c

Some axioms

Condorcet consistency: if x beats every other candidate in a pairwise election (x is a Condorcet Winner), then x should win.

Voting rules that satisfy CC:

Copeland:

- Winner is the candidate that beats the most other candidates in pairwise elections

Maximin:

- Winner is candidate that gets at least an α fraction of the vote in all pairwise elections, for α as large as possible.
- Can you see why this satisfies CC?

1	2	3
b	c	d
a	a	a
c	d	b
d	b	c

Some axioms

Condorcet consistency: if x beats every other candidate in a pairwise election (x is a Condorcet Winner), then x should win.

Voting rules that satisfy CC:

Dodgson:

- If there's a Condorcet winner, then they win.
- If not, find the smallest number of swaps between adjacent pairs (bubble-sort style) needed to produce a Condorcet winner, and have that candidate be the winner.
- Charles Dodgson, "A method of taking votes on more than two issues", 1876.
- Happens to be NP-complete...

1	2	3
b	c	d
a	a	a
c	d	b
d	b	c

A fun example

- Plurality: a
- Borda count: b
- Condorcet winner: c
- STV: d
- Plurality with runoff: e

33 voters	16 voters	3 voters	8 voters	18 voters	22 voters
a	b	c	c	d	e
b	d	d	e	e	c
c	c	b	b	c	b
d	e	a	d	b	d
e	a	e	a	a	a

[Slide from Ariel Procaccia]

Next: Arrow’s theorem and Gibbard-Satterthwaite

Arrow's impossibility theorem

Theorem: Any social welfare function (outputs a ranking) for $m \geq 3$ alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

- Unanimity: for any two alternatives x, y, if all voters rank x above y then the output also ranks x above y.
- IIA: for any two alternatives x, y, if voters modify their rankings but keep their order of x and y unchanged, then the output order of x and y doesn't change. I.e., adding/removing other "dummy" candidates doesn't change whether x beats y.
- Dictatorship: For some i, the output ranking is always i's ranking.

1	2	3
a	b	c
b	c	a
c	a	b

m

n

Intuitive implication: Any reasonable social welfare function will violate IIA.

(Note: we'll be assuming it's a deterministic function)

Arrow's impossibility theorem

Theorem: Any social welfare function (outputs a ranking) for $m \geq 3$ alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

Proof:

- **Lemma 1:** if there is some alternative x such that each voter either ranks x first or ranks x last (they don't have to agree) then the output must have x first or last.
- **Proof:** by contradiction.
  - Suppose there are alternatives y, z such that the output ranks $y \succ x \succ z$.
  - Modify each ranking by putting z first if x was last, or 2nd if x was first. This doesn't affect any relative orders of x and z, so by IIA the output should still rank $x \succ z$.
  - Now, all voters have $z \succ_i y$, so by unanimity, the output should rank $z \succ y$.
  - But we didn't change any relative orders of x and y, so by IIA the output should still rank $y \succ x$, a contradiction.

Arrow's impossibility theorem

Theorem: Any social welfare function (outputs a ranking) for $m \geq 3$ alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

Proof:

- Now, pick some alternative x and consider any set of voters that all rank x last.
- By unanimity, the output must rank x last.
- Now, one at a time, modify each voter's preferences to rank x first, until they all rank x first and (by unanimity) the output must rank x first.
- By Lemma 1, there must have been some voter i such that modifying its preferences caused x to move from last to first.
- We will show that voter i must be a dictator.

Arrow's impossibility theorem

Theorem: Any social welfare function (outputs a ranking) for $m \geq 3$ alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

Proof:

- Modify voters' rankings in an arbitrary way subject to keeping x last/first. Want to show that the output must match i's ranking (will handle other locations for x later).
- By IIA, this didn't change when x moved from last to first in the output order.
- Pick some $y \succ_i z$ in i's ranking (y, z distinct from x).
- Notice that if we move y above x in B, then y must move above x in the output by IIA since it was above x in the output for A. This must also be above z (because x is still above z by IIA).
- So (by IIA), y is above z in the output for A.

A

i
*
y
*
z
x

B

i
x
*
y
*
z

Arrow's impossibility theorem

Theorem: Any social welfare function (outputs a ranking) for $m \geq 3$ alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

Proof:

- Modify voters' rankings in an arbitrary way subject to keeping x last/first. Want to show that the output must match i's ranking (will handle other locations for x later).
- By IIA, this didn't change when x moved from last to first in the output order.
- Now, what if x isn't first/last in the various rankings, will y still be above z in the output? Yes, by IIA.
- So, all that remains is to show that i is also a dictator for pairs involving x.

A	B
i	i
*	x
y	*
*	y
z	*
x	z

Arrow's impossibility theorem

Theorem: Any social welfare function (outputs a ranking) for $m \geq 3$ alternatives that satisfies unanimity and independence of irrelevant alternatives is a dictatorship.

Proof:

- Modify voters' rankings in an arbitrary way subject to keeping x last/first. Want to show that the output must match i's ranking (will handle other locations for x later).
- By IIA, this didn't change when x moved from last to first in the output order.
- Now, what if x isn't first/last in the various rankings, will y still be above z in the output? Yes, by IIA.
- Pick some pair $\{x, y\}$. If we run the above construction using z instead of x, we will find there is some dictator j for the pair $\{x, y\}$.
- But j has to be i since we just saw in moving from A to B that i can impact their order. Done.

A

i
*
y
*
z
x

B

i
x
*
y
*
z

Gibbard-Satterthwaite

Now focus on social choice functions (functions that choose a winner).

f is incentive-compatible if no voter would ever prefer to misrepresent their preferences.

- That is if $x = f(\prec_1, \dots, \prec_i, \dots \prec_n)$ and $y = f(\prec_1, \dots, \prec_i', \dots, \prec_n)$ then it should be the case that $x \succ_i y$.
- Notice we must also have $x \prec_i' y$. This is also called monotonicity. "If you can change the outcome, it must involve raising the new outcome above the old one"

A misrepresentation that leads to a preferred outcome is called a strategic manipulation.

In the case of $m = 2$ candidates, majority voting is incentive-compatible.

Dictatorship: For some voter i, the winner is always i's favorite.

Gibbard-Satterthwaite

Theorem: Any social choice function (picks a winner) for $m \geq 3$ alternatives that is onto and incentive compatible must be a dictatorship.

Proof:

- By contradiction. Suppose f is onto, incentive compatible, and not a dictatorship.
- Plan: construct a social welfare function F that satisfies unanimity, IIA, and non-dictatorship, violating Arrow's impossibility theorem. (So, we're doing a reduction)

Gibbard-Satterthwaite

Theorem: Any social choice function (picks a winner) for $m \geq 3$ alternatives that is onto and incentive compatible must be a dictatorship.

Lemma (set-unanimity): If f is onto and incentive-compatible, and if for some subset of alternatives S, every voter ranks S first (all $x \in S$ above all $y \notin S$ even if voters disagree on orders within S), then f must output some alternative in S.

Proof: Let $\prec_1, \dots, \prec_n$ be the given set of preferences.

- Pick some $x \in S$ and let $\prec'_1, \dots, \prec'_n$ be some preferences s.t. $f(\prec'_1, \dots, \prec'_n) = x$. (by onto)
- Modify $\prec'_1$ to $\prec_1$, $\prec'_2$ to $\prec_2$, etc. If this ever yields prefs s.t. $f(\prec_1, \dots, \prec_i, \prec'_{i+1}, \dots, \prec'_n) = y \notin S$, then this would violate incentive-compatibility.
  - Because it would mean that voter i with true preferences $\prec_i$ would rather misrepresent as $\prec'_i$ when the others are as above.
- So, this yields the lemma.

Gibbard-Satterthwaite

Theorem: Any social choice function (picks a winner) for $m \geq 3$ alternatives that is onto and incentive compatible must be a dictatorship.

Proof:

- Given f, create social welfare function $F(\prec_1, \ldots, \prec_n)$ as follows:
  - For each pair of alternatives x, y: rank them by bringing them to the top of each $\prec_i$ and seeing which of them f would output (must output one of them by Lemma).
  - Need to show this is well-defined (transitive): for any triple $\{x, y, z\}$, one of them should beat the other two.
    - If we bring all three to the top, we know f will output one of them, say x, by Lemma.
    - Then x beats y and z. In particular, suppose for contradiction that z beats x. This means that if one at a time we lower y back to its original location in each voter's ordering, at some point the winner has to switch. This violates IC. (Not monotone)

Gibbard-Satterthwaite

Theorem: Any social choice function (picks a winner) for $m \geq 3$ alternatives that is onto and incentive compatible must be a dictatorship.

Proof:

- Given f, create social welfare function $F(\prec_1, \ldots, \prec_n)$ as follows:
  - For each pair of alternatives x, y: rank them by bringing them to the top of each $\prec_i$ and seeing which of them f would output (must output one of them by Lemma).
- Unanimity: if every voter ranked x above y, then when we bring them to the top, f must output x by set-unanimity with $S = \{x\}$. So, F ranks x above y.
- IIA: ranking of x, y determined by bringing them to top and applying f. If a voter could change this ranking by reordering other alternatives, then f wouldn't be IC.
- Non-dictatorship: Since f is non-dictator, for each i, exist prefs such that f does not pick i's favorite. Say i's favorite is x but f picks y. By monotonicity, moving x, y to top can't change f (nothing got raised above y). So F ranks y above x too.