Towards less manipulable voting systems

François Durand, Fabien Mathieu, Ludovic Noirie

Inria / Alcatel-Lucent Bell Labs France

Inria's Junior Seminar

March 21, 2013

Introduction: voting systems

Context:

- Origins in **politics**.
- Applications in any situation of **collective choice**.

Questions:

- Is there a natural way to select a **reasonable winner**?
- Can we **trust** the electors?
- If not, is it possible to design a voting system that is **resistant to manipulation**, i.e. to tactical voting?

Plan

In quest for a reasonable winner

Presentation of manipulability

Minimizing manipulability (my stuff)

Plan

A simplified framework

n **electors**.

- Agents that can make decisions.
- Can be persons or machines (servers for routing, etc.).

m **candidates** named A, B, C, ...

- Mutually exclusive options about a question.
- Can be persons, routes...

Each elector i has a **binary relation of preferences** r_i over the candidates. Example of i's preferences: A $\sim$ B $\succ$ D $\succ$ C.

Voting system $f : (r_1, \dots, r_n) \rightarrow v \in \{\text{A}, \text{B}, \text{C}...\}$.

2 candidates: May's theorem

Plurality:

- Each elector votes for one candidate.
- The candidate with most votes gets elected.

May's theorem (1952): plurality is the only anonymous, neutral and positively responsive voting system for 2 candidates.

3 candidates or more: Condorcet winner

Independance of irrelevant alternatives (IIA): if we remove one of the losing candidates, the winner should remain the same.

E.g. if C wins, then she must win any electoral duel versus A or B. If it is the case, we say that she's a **Condorcet winner**.

Example:

	Electors		
	40	35	25
Preferences	A	B	C
	C	C	A
	B	A	B

"Majority matrix":

	A	B	C	Victories
A	■			0
B		■		0
C			■	0

3 candidates or more: Condorcet winner

Independance of irrelevant alternatives (IIA): if we remove one of the losing candidates, the winner should remain the same.

E.g. if C wins, then she must win any electoral duel versus A or B. If it is the case, we say that she's a **Condorcet winner**.

Example:

	Electors		
	40	35	25
Preferences	A	B	C
	C	C	A
	B	A	B

"Majority matrix":

	A	B	C	Victories
A		**65**		1
B	35			0
C				0

3 candidates or more: Condorcet winner

Independance of irrelevant alternatives (IIA): if we remove one of the losing candidates, the winner should remain the same.

E.g. if C wins, then she must win any electoral duel versus A or B. If it is the case, we say that she's a **Condorcet winner**.

Example:

	Electors		
	40	35	25
Preferences	A	B	C
	C	C	A
	B	A	B

"Majority matrix":

	A	B	C	Victories
A		**65**	40	1
B	35			0
C	**60**			1

3 candidates or more: Condorcet winner

Independance of irrelevant alternatives (IIA): if we remove one of the losing candidates, the winner should remain the same.

E.g. if C wins, then she must win any electoral duel versus A or B. If it is the case, we say that she's a **Condorcet winner**.

Example:

	Electors		
	40	35	25
Preferences	A	B	C
	C	C	A
	B	A	B

"Majority matrix":

	A	B	C	Victories
A		**65**	40	1
B	35		35	0
C	**60**	**65**		2

3 candidates or more: Condorcet winner

Independance of irrelevant alternatives (IIA): if we remove one of the losing candidates, the winner should remain the same.

E.g. if C wins, then she must win any electoral duel versus A or B. If it is the case, we say that she's a **Condorcet winner**.

Example:

	Electors		
	40	35	25
Preferences	A	B	C
	C	C	A
	B	A	B

"Majority matrix":

	A	B	C	Victories
A		**65**	40	1
B	35		35	0
C	**60**	**65**		2

If we want to extend Plurality for $m \geq 3$ and respect IIA, then C must be elected. We then say that our voting system respects **Condorcet criterion**.

3 candidates or more: Condorcet's paradox

Example:

	Electors		
	40	35	25
Preferences	A	B	C
	B	C	A
	C	A	B

"Majority matrix":

	A	B	C	Victories
A		**65**	40	1
B	35		**75**	1
C	**60**	25		1

Condorcet's paradox (1785): A defeats B, B defeats C and C defeats A.

It's not possible to extend Plurality for $m \geq 3$ while respecting IIA (independance of irrelevant alternatives).

Arrow's theorem

We would like a voting system with the following properties.

- **Non-dictatorship:** there is not one elector who always decides alone.
- **Unanimity:** whenever all electors prefers A to B, candidate B cannot get elected.
- **Independance of irrelevant alternatives (IIA):** if we remove one of the losing candidates, the winner should remain the same.

Arrow's theorem

We would like a voting system with the following properties.

- **Non-dictatorship:** there is not one elector who always decides alone.
- **Unanimity:** whenever all electors prefers A to B, candidate B cannot get elected.
- **Independance of irrelevant alternatives (IIA):** if we remove one of the losing candidates, the winner should remain the same.

Arrow's theorem (1951): for $m \geq 3$ candidates, such a voting system does not exist.

$\Rightarrow$ For $m \geq 3$ candidates, there is no "natural", canonical way to agregate binary relations of preferences from several electors in order to choose a winning candidate.

Extending the framework of preferences

Old framework: binary relations of preferences only.

Example of extended framework:

- Each elector has a **utility vector** about the candidates, e.g. $(10, 10, 0, 2)$.
- This utility vector induces a binary relation of preferences over the candidates, e.g. $\text{A} \sim \text{B} \succ \text{D} \succ \text{C}$.

Extending the framework of preferences

Old framework: binary relations of preferences only.

Example of extended framework:

- Each elector has a **utility vector** about the candidates, e.g. $(10, 10, 0, 2)$.
- This utility vector induces a binary relation of preferences over the candidates, e.g. $\mathrm{A} \sim \mathrm{B} \succ \mathrm{D} \succ \mathrm{C}$.

General extended framework:

- Each elector i has a **state** $\omega_i \in \Omega_i$,
- This state contains enough information so that we can extract her **binary relation of preferences** $r_i = R_i(\omega_i)$.

Voting system $f : (\omega_1, \ldots, \omega_n) \to v \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \ldots\}$.

Escaping Arrow's theorem

Example: range voting.

- Each elector gives her utility vector, that is, a note for each candidate.
- The candidate with highest average (or median) wins.

This voting system is **non-dictatorial**, **unanimous** and **independant of irrelevant alternatives**... and infinitely many other voting systems are too!

So... have we won? Have we found a voting system that is fully satisfying?

Plan

Manipulability: an example

Voting system: plurality.

	Electors		
	40	35	25
Preferences	A	B	C
	B	C	A
	C	A	B
Sincere ballot	**A**	B	C

Manipulability: an example

Voting system: plurality.

	Electors		
	40	35	25
Preferences	A	B	C
	B	C	A
	C	A	B
Sincere ballot	**A**	B	C
Manipulation	A	**C**	**C**

We say that this situation is **manipulable** for this voting system:

- A subset of electors, by casting a tactical ballot, may change the result to a candidate that they prefer.
- I.e., sincere voting is not a **strong Nash equilibrium**.

Why would manipulability be a problem?

If electors do not vote sincerely, the collective decision relies on **false information**.

If electors do vote sincerely, they may be frustrated and find the system nonsensical, since a non-sincere ballot, **misrepresenting** their preferences, would have **defended these preferences better**.

Gibbard's theorem

Gibbard's theorem (1973): for any non-dictatorial voting system with at least 3 eligible candidates, there exists a situation that is manipulable by one elector.

I.e.: this situation is not even a **weak Nash equilibrium**.

This theorem is true in the general framework of preferences!
In fact, is is also true in an even more general framework...

Manipulability rate

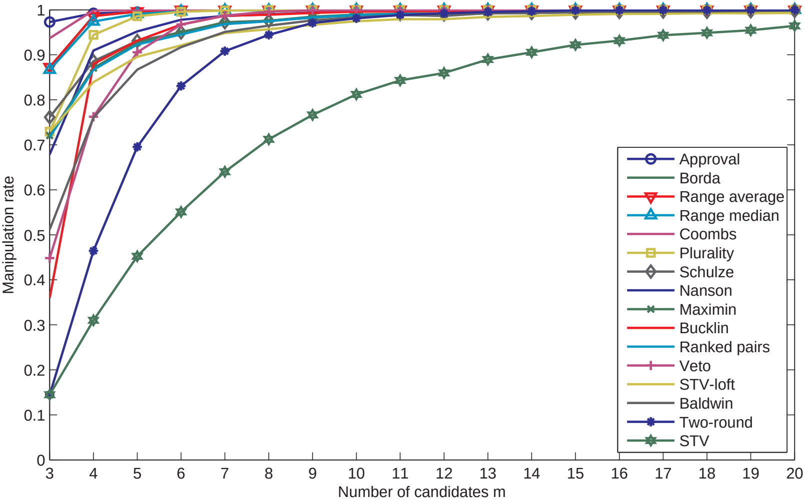
We draw a situation (preferences of the population) according to a **probability measure** P.

Manipulability rate: what is the probability that this situation is manipulable for voting system f?

$$\rho_P(f) \in [0,1].$$

Manipulability is quite frequent

P = "Uniform spherical culture", $n = 33$ electors



Plan

In quest for a reasonable winner

Presentation of manipulability

Minimizing manipulability (my stuff)

Condorcification theorem

Take a voting system f (with reasonable properties).

We define its **condorcification** f^c like this:

- Whenever there is a Condorcet winner, designate her;
- Otherwise, use f.

Then f^c **is at most as manipulable as** f:

- Any situation manipulable for f^c is manipulable for f;
- In particular, $\rho_P(f^c) \leq \rho_P(f)$ for any probability measure P.

Note: for all classical voting systems that fail Condorcet criterion, we proved that their condorcification is **strictly less manipulable**.

Slicing theorem

Take a probability measure P used to draw population preferences (with reasonable properties).

Take a voting system f. This system may depend on the whole information about **electors's states** (utilities, etc.), not only on **binary relations of preferences**.

Then there exists a voting system f' that:

- **Depends on binary relations of preferences only**,
- Is **at most as manipulable as** f for P, in the sense that $\rho_P(f') \leq \rho_P(f)$.

Furthermore, if f respects Condorcet criterion, there exists such a f' that does too.

Existence of an optimal voting system

Take a probability measure P used to draw population preferences (with reasonable properties).

Then there exists a voting system:

- whose **manipulation rate is minimal** (among system with reasonable properties),
- that **depends only on binary relations of preferences**,
- that **respects Condorcet criterion**.

Future work

So, we can restrict the search for an optimal voting system in the class of those that:

- **depend only on binary relations of preferences**,
- **respect Condorcet criterion**.

For a fixed pair (n, m), there is a finite number of such voting systems: $\sim m^{m!^n}$.

For the moment, we can give this optimal voting system explicitly for **very small** values: $n \leq 5$ electors and $m \leq 4$ candidates...

Conclusion

- For voting systems that rely on binary relations of preferences only, there is no canonical way to choose the winner (Arrow's theorem).
- For $m \geq 3$ candidates, all non-dictatorial voting systems are subject to manipulation (Gibbard's theorem).
- We can limit manipulability by **condorcification** and **slicing**.
- There exists an **optimal voting system** that depends only on binary relations of preferences and respects Condorcet criterion.

Thanks for your attention! Questions?