Généralisation du théorème d'impossibilité d'Arrow et de Gibbard-Satterthwaite.

Théo JAMIN

14 février 2022

Laboratoire angevin de recherche en mathématiques – Université d'Angers

Introduction et plan

Plan

- Modélisation des systèmes électoraux [1].
- Définition des conditions d'admissibilité d'un système de vote.
- Théorème d'impossibilité d'agrégation des préférences individuelles.
- Corollaire : théorèmes d'Arrow et de Gibbard-Satterhwaite.
- Limites du modèle et alternatives.
- Preuve du théorème (pour les plus courageux).

1. *Celle faite par Yu, inspirée de celle d'Arrow.*

La modélisation d’un système de vote

Modélisation

Soient N et $M \geq 3$.

- $\mathscr{V} = \{1, \cdots, N\}$ un ensemble de votants
- $\mathscr{C} = \{c_1, \cdots, c_M\}$ une liste de candidats.

Le vote d'un individu $n \in \mathscr{V}$ est une relation d'ordre totale[2] $\succ_n$ sur $\mathscr{C}$.
Nous noterons l'ensemble des votes possibles par $\mathscr{P}$.
Une urne $\vec{\succ}$ est l'ensemble des N votes $(\succ_1, \cdots, \succ_N)$, $\succ_n \in \mathscr{P}$ $\forall n$.

2. *Les théorème que nous énoncerons plus tard restent vrais même si le vote n'est pas une relation d'ordre total.*

Modélisation

Soit $\mathscr{P}^N$ l'ensemble des urnes possibles obtenues par le votes de N votants pour l'éléction d'un candidat de $\mathscr{C}$.
On modélise un système électoral par

FCS
Une fonction de choix social (abrégée FCS) est une fonction F assignant à toute urne un choix de candidat, i.e.

$$\begin{aligned} F : \mathscr{P}^N &\mapsto \mathscr{C} \\ \vec{\succ} &\mapsto F(\vec{\succ}) \in \mathscr{C} \end{aligned}$$

Conditions d’admissibilité

Conditions d'admissibilité

Conditions que l'on aimerait respecter :

- Toutes les candidats doivent pouvoir être élus.
- Pas de dictature : le système ne doit pas tenir compte du vote d'un individu au détriment des autres.
- Si un votant améliore le rang d'une candidat, cela ne doit jamais le désavantager.
- Ôter un candidat (autre qu'un gagnant) ne doit pas changer le résultat du vote.

Conditions imposées par le modèle :

- Le système de vote doit toujours aboutir.
- Les votes sont des classements des candidats (pas d'intensité des préférences).

Définitions

Définitions

- Dans une urne $\vec{\succcurlyeq}$, le candidat c_i domine le candidat c_j si

$$\forall n \in \mathcal{N}, \; c_i \succ_n c_j$$

On notera $c_i \vec{\succcurlyeq} c_j$.

- Une fonction génératrice de dominance est une fonction

$$T_{i,\cdots,k} : \mathscr{P}^N \mapsto \mathscr{P}^N$$

qui "surclasse" les candidats $c_i, \cdots, c_k$ dans tous les votes d'une urne $\vec{\succcurlyeq}$ sans changer l'ordre interne de ces candidats ni l'ordre des autres.

Exemple

$$\vec{\succ} = \left(\begin{array}{ccc} 1 & 2 & 3 \\ \hline c_i & c_k & c_i \\ c_j & c_j & c_k \\ c_k & c_i & c_j \end{array}\right) \implies T_{ij}(\vec{\succ}) = \left(\begin{array}{ccc} 1 & 2 & 3 \\ \hline c_i & c_j & c_i \\ c_j & c_i & c_j \\ c_k & c_k & c_k \end{array}\right)$$

Modélisation des conditions d'admissibilité

- Une FCS F vérifie le critère éligibilité totale (E) si tous les candidats peuvent être élus, i.e.
$$F(\mathscr{P}^N) = \mathscr{C}$$

- Une FCS F est dite unanime (U) si
$$\forall \vec{\succ} \in \mathscr{P}^N,\ c_i \vec{\succ} c_j \implies F(\vec{\succ}) \neq c_j$$

Modélisation des conditions d'admissibilité

- Soit $\vec{\succ}'$ une urne telle que $\vec{\succ}' = T_{i,j}(\vec{\succ}')$.
  Une FCS F vérifie le critère de Condorcet généralisé (C) si

$$\forall \vec{\succ} \in \mathscr{P}^N \text{ telle que } \vec{\succ}\big|_{c_i,c_j} = \vec{\succ}'\big|_{c_i,c_j},\ F(\vec{\succ}) = c_i \implies F(\vec{\succ}') = c_i$$

En particulier, ce critère permet d'affirmer qu'un gagnant de Condorcet [3] doit être élu.

3. *le candidat qui, comparé tour à tour à chacun des autres candidats, s'avérerait être le candidat préféré.*

Exemple

- (C) :

$$F\left(\begin{array}{ccc} 1 & 2 & 3 \\ \hline c_i & c_k & c_i \\ c_j & c_j & c_k \\ c_k & c_i & c_j \end{array}\right) = c_i \implies F\left(\begin{array}{ccc} 1 & 2 & 3 \\ \hline c_i & c_j & c_i \\ c_j & c_i & c_j \\ c_k & c_k & c_k \end{array}\right) = c_i$$

  En particulier,

$$F(\vec{\succ}) = c_i \overset{(C)}{\implies} F(T_{i,j}(\vec{\succ})) = c_i$$

- (U) :

$$\vec{\succ} = \left(\begin{array}{ccc} 1 & 2 & 3 \\ \hline c_i & c_k & c_i \\ c_j & c_i & c_k \\ c_k & c_j & c_j \end{array}\right) \text{ (i.e. } c_i \mathrel{\vec{\succ}} c_j\text{)} \implies F(\vec{\succ}) \neq c_j$$

Résultat préliminaire

Lemme 1
Une fonction de choix social qui vérfie les critères **d'éligibilité totale** et **de Condorcet généralisé** est **unanime**, i.e.

$$(E) + (C) \implies (U)$$

Théorème central

- Une FCS F est dite dictatoriale (D) s'il existe un individu n tel que

$$\forall \vec{\succ} \in \mathscr{P}^N,\ c_i \vec{\succ}_n c_j \implies F(\vec{\succ}) \neq c_j$$

Théorème
Une fonction de choix social qui vérifie les critères **d'éligibilité totale** et **de Condorcet généralisé** est **dictatoriale**, i.e.

$$(E) + (C) \implies (D)$$

Théorème de Gibbard-Satterthwaite

Vers le théorème de Gibbard-Satterthwaite

On note $Remp_{n \leftarrow \succ'}(\vec{\succ})$ l'urne $\vec{\succ}$ dans laquelle le vote $\succ_n$ de l'individu n est remplacé par le vote $\succ'$.

- Une FCS F est dite non-manipulable ($\cancel{M}$) si $\forall \vec{\succ} \in \mathscr{P}^N$, $\forall \succ' \in \mathscr{P}$,

$$F(Remp_{n \leftarrow \succ'}(\vec{\succ})) \neq F(\vec{\succ}) \implies F(\vec{\succ}) \succ_n F(Remp_{n \leftarrow \succ'}(\vec{\succ}))$$

Cette condition demande qu'un individu n'a pas d'intérêt à mentir sur ses préférences même si cela peut changer l'issue du vote, i.e. être honnête est optimal.

Lemme 2
Une fonction de choix social **non-manipulable** vérifie **le critère de Condorcet**, i.e.

$$(\not{M}) \implies (C)$$

Théorème de Gibbard-Satterwaite

Une fonction de choix social qui vérfie le critère **d'éligibilité totale** et **non-manipulable** est **dictatoriale**, i.e.

$$(E) + (\cancel{M}) \implies (D)$$

Démonstration.

$$(E) + (\cancel{M}) \overset{\text{lemme 2}}{\implies} (E) + (C) \overset{\text{théorème}}{\implies} (D)$$

∎

Théorème d'Arrow

Encore des définitions...

- Une fonction de préférences sociales (abrégée *FPS*) est une fonction P assignant un vote à toute urne

$$\begin{aligned} P : \mathscr{P}^N &\mapsto \mathscr{C} \\ \vec{\succ} &\mapsto P(\vec{\succ}) \in \mathscr{P} \end{aligned}$$

On peut évidemment construire une fonction de choix social à partir d'une *FPS* P en ne gardant que l'élément "maximal". On notera F^P la fonction ainsi construite.

Encore des définitions...

Soit P une fonction de préférences sociales.

- P est dite Arrow-dictatoriale (AD) s'il existe un individu n tel que

$$\forall \vec{\succ} \in \mathscr{P}^N,\ \forall c_i, c_j \in \mathscr{C},\ c_i \succ_n c_j \implies c_i\ P(\vec{\succ})\ c_j$$

- P est dite Arrow-unanime (AU) si

$$\forall \vec{\succ} \in \mathscr{P}^N,\ c_i \vec{\succ} c_j \implies c_i\ P(\vec{\succ})\ c_j$$

- P est dite Arrow-indépendante aux alternatives non pertinentes ($AIAP$) si

$$\forall \vec{\succ}, \vec{\succ}' \in \mathscr{P}^N,\ \vec{\succ}_{i,j} = \vec{\succ}'_{i,j},\ c_i\ P(\vec{\succ})\ c_j \Leftrightarrow c_i\ P(\vec{\succ}')\ c_j$$

Théorème d'Arrow

Une fonction de préférences sociales (Arrow-)**unanime** et (Arrow-)**indépendante aux alternatives non pertinentes** est (Arrow-)**dictatoriale**, i.e.

$$(AU) + (AIAP) \implies (AD)$$

Démonstration.

Soit P une fonction de préférences sociales, on construit F^P une FCS par $F^P(\vec{\succcurlyeq}) = c_i$ tel que $\forall j,\ c_i\ P(\vec{\succcurlyeq})\ c_j$.

On montre

$$(AU) \implies (E),\ (AIAP) \implies (C) \text{ et } (D) \implies (AD)$$

∎

Limites du modèle

Évidemment, ce modèle ne s'applique pas à tous les systèmes de vote. La condition imposée par le modèle de **non-intensité des préférences** est trop restrictive.
Néanmoins, ce que nous apprend cette modélisation, c'est l'impossibilité d'avoir un mode de scrutin intéressant sans tenir compte de l'intensité des préférences (comme la plupart des modes en vigueur aujourd'hui).

Heureusement...

Balinski - Laraki, 2007
Construction d'un mode de scrutin, appelé **jugement majoritaire**, reposant sur la **méthode de la médiane** qui satisfait :

- Universalité [4]
- Unanimité
- Indépendance aux alternatives non pertinentes
- Absence de dictateur
- Monotonie [5]

4. *Le scrutin doit aboutir.*
5. *Déclasser un perdant ne peut pas le faire gagner.*

Principe

- L'électeur attribue à chaque candidat une **mention**[6].
- Pour chaque candidat, on détermine la **mention médiane**.
- Le candidat élu est un candidat qui obtient la meilleure **mention médiane**.

- Méthode de départage :
  On calcule pour chaque candidat à départager
  - Le pourcentage d'électeurs attribuant **strictement plus** que la mention médiane.
  - Le pourcentage d'électeurs attribuant **strictement moins** que la mention médiane.

  La plus grande des 4 valeurs détermine le résultat.

6. *Exemple : très bien, bien, assez bien, passable, insuffisant, à rejeter.*

D’autres solutions

Il existe encore d’autres alternatives à ce système :

- Le scrutin de **Condorcet Randomisé**[7].
- Le **vote par notation**.
- Etc.

7. Système sans intensité des préférences mais non déterministe.

Preuve du théorème central.

Théorème central

Théorème

Une fonction de choix social qui vérifie le critère **d'éligibilité totale** et **de Condorcet généralisé** est **dictatoriale**, i.e.

$$(E) + (C) \implies (D)$$

Preuve : repérage du votant pivot

Soit $\vec{\succcurlyeq}$ tel que $c_i \vec{\succcurlyeq} c_j \vec{\succcurlyeq} \mathscr{C} \setminus \{c_i, c_j\}$ et construisons $\vec{\succcurlyeq}^{(n)}$ en inversant les n-ièmes premières positions des c_i et des c_j, i.e.

	1	$\cdots$	$n-1$	n	$n+1$	$\cdots$	
$\vec{\succcurlyeq}^{(0)}$	c_i	$\cdots$	c_i	c_i	c_i	$\cdots$	c_i
	c_j	$\cdots$	c_j	c_j	c_j	$\cdots$	c_j
$\vec{\succcurlyeq}^{(n)}$	c_j	$\cdots$	c_j	c_j	c_i	$\cdots$	c_i
	c_i	$\cdots$	c_i	c_i	c_j	$\cdots$	c_j
$\vec{\succcurlyeq}^{(N)}$	c_j	$\cdots$	c_j	c_j	c_j	$\cdots$	c_j
	c_i	$\cdots$	c_i	c_i	c_i	$\cdots$	c_i

Par le lemme 1, F est unanime, en particulier $F(\vec{\succcurlyeq}^{(0)}) = c_i$ et $F(\vec{\succcurlyeq}^{(N)}) = c_j$. On note n_{ij} l'individu **pivot**.

Preuve : le pivot est en fait dictateur

Soit maintenant $\vec{\succcurlyeq}'$ une urne telle que :

$$
\begin{array}{c|ccccccc}
 & 1 & \cdots & n_{ij}-1 & n_{ij} & n_{ij}+1 & \cdots & N \\
\hline
\vec{\succcurlyeq}' & c_j & \cdots & c_j & c_i & c_i & \cdots & c_i \\
 & \square & \cdots & \square & & & & \\
 & c_i & \cdots & c_i & c_j & c_j & \cdots & c_j \\
 & \square & \cdots & \square & c_k & c_k & \cdots & c_k
\end{array}
$$

où les $\square$ sont les emplacements possibles pour c_k.

But

On va montrer que $F(\vec{\succcurlyeq}') = c_i$

Preuve : le pivot est en fait dictateur

Soit maintenant $\vec{\succsim}''$ une urne telle que :

	1	$\cdots$	$n_{ij}-1$	n_{ij}	$n_{ij}+1$	$\cdots$	N
$\vec{\succsim}''$	$\square$	$\cdots$	$\square$				
	c_j	$\cdots$	c_j	c_j	c_i	$\cdots$	c_i
	$\square$	$\cdots$	$\square$		$\square$	$\cdots$	$\square$
	c_i	$\cdots$	c_i	c_i	c_j	$\cdots$	c_j
	$\square$	$\cdots$	$\square$	c_k	$\square$	$\cdots$	$\square$

où les $\square$ sont les emplacements possibles pour c_k.

But
On va montrer que $F(\vec{\succsim}'') = c_j$

Références

- **Yu, Ning Neil** : *A one-shot proof of Arrow's and the Gibbard-Satterthwaite theorem*, Economic Theory Bull (2013).
- **Arrow, Kenneth J.** : *Social Choice and Individual Values*, (1951).
- **Gibbard, Allan** : *Manipulation of voting schemes : a general result*, Econometrica (1973).
- **Satterthwaite, Mark Allen** : *Strategy-proofness and Arrow's conditions : Existence and correspondence theorems for voting procedures and social welfare functions*, Journal of Economic Theory (1975).

Merci de votre attention !