DEVOIR MAISON FACULTATIF

Le théorème d'Arrow

Problème — Un groupe d'individus doit établir un classement collectif de plusieurs options (des candidats, des projets, etc.), alors que chaque individu a son propre classement. En 1785, Condorcet a remarqué que le vote à la majorité peut produire un classement collectif incohérent. En 1951, Kenneth Arrow (prix Nobel d'économie 1972) a démontré que, dès qu'il y a au moins trois options, aucune règle ne peut satisfaire à la fois trois exigences qui semblent pourtant très naturelles, dont l'absence de **dictateur**. L'objet de ce devoir est de démontrer ce théorème.

Cadre. Dans tout le problème, on fixe deux entiers $n \geq 1$ et $m \geq 3$.

- Les **individus** sont numérotés de 1 à n : on note $I = [\![1; n]\!]$.
- Les **options** forment un ensemble fini A à m éléments.

Classements. Classer les options, c'est les ranger de la préférée à la moins appréciée, sans ex æquo. On code un classement par une relation binaire $\succ$ sur A, et l'on lit « $a \succ b$ » : « a est préféré à b ». Pour que cette relation décrive bien un classement, on impose qu'elle soit :

- **irréflexive** : $\forall a \in A,\ \ a \not\succ a$;
- **transitive** : $\forall a, b, c \in A,\ \ (a \succ b \text{ et } b \succ c) \implies a \succ c$;
- **totale** : $\forall a, b \in A,\ \ a \neq b \implies (a \succ b \text{ ou } b \succ a)$.

Une telle relation s'appelle un **ordre total strict** sur A ; on note $\mathcal{L}(A)$ l'ensemble de ces relations.

Exemple. Si $A = \{a, b, c\}$, le classement « a, puis c, puis b » est la relation $\succ$ pour laquelle $a \succ c$, $c \succ b$ et $a \succ b$, et rien d'autre. On l'écrit simplement $a \succ c \succ b$. Plus généralement, $a_1 \succ a_2 \succ \cdots \succ a_k$ signifie $a_1 \succ a_2,\ \ldots,\ a_{k-1} \succ a_k$.

Profils. Chaque individu $i \in I$ a son propre classement, noté $\succ_i$. Un **profil** est la donnée des classements de tous les individus : $P = (\succ_1, \ldots, \succ_n) \in \mathcal{L}(A)^n$. Lorsqu'on considère un second profil P', ses classements sont notés $\succ'_1, \ldots, \succ'_n$.

Règles d'agrégation. Une **règle d'agrégation** est une application $F : \mathcal{L}(A)^n \longrightarrow \mathcal{L}(A)$: à chaque profil P, elle associe un classement collectif, que l'on note $\succ_P$ plutôt que $F(P)$. Ainsi, « $a \succ_P b$ » signifie « le groupe préfère a à b ».

> Il serait plus naturel d'autoriser les ex æquo. Par simplicité, on impose un classement strict, quitte à départager les égalités par un ordre fixé à l'avance entre les options (l'ordre alphabétique, par exemple) ; avec de nombreux individus, une égalité parfaite est de toute façon rare.

Trois exigences naturelles. On considère les trois propriétés suivantes, qu'une règle d'agrégation peut vérifier ou non.

- **(U) Unanimité.** Une règle d'agrégation vérifie **(U)** si, chaque fois que tous les individus préfèrent a à b, le groupe aussi :
$$\forall P \in \mathcal{L}(A)^n,\ \ \forall a, b \in A,\ \ \Big(\forall i \in I,\ \ a \succ_i b\Big) \implies a \succ_P b.$$
- **(IIA) Indépendance des options non pertinentes.** Une règle d'agrégation vérifie **(IIA)** si, pour savoir si le groupe préfère a à b, il suffit de savoir, pour chaque individu, s'il préfère a à b : la place des autres options dans les classements n'intervient pas. Autrement dit, si chaque individu compare a et b de la même façon dans deux profils P et P', alors le groupe compare aussi a et b de la même façon dans ces deux profils :
$$\forall P, P' \in \mathcal{L}(A)^n,\ \ \forall a, b \in A,\ \ \Big(\forall i \in I,\ \ a \succ_i b \iff a \succ'_i b\Big) \implies \Big(a \succ_P b \iff a \succ_{P'} b\Big).$$
Exemple. Si l'individu 1 passe du classement $a \succ_1 b \succ_1 c$ au classement $c \succ'_1 a \succ'_1 b$ (les autres individus ne changeant rien), il préfère toujours a à b : d'après **(IIA)**, ce changement ne modifie pas la comparaison collective de a et b.
- **(ND) Absence de dictateur.** On dit qu'un individu $d \in I$ est un **dictateur** pour une règle d'agrégation si le classement collectif qu'elle produit est toujours le sien :
$$\forall P \in \mathcal{L}(A)^n,\ \ \succ_P = \succ_d,$$
où $\succ_d$ désigne le classement de l'individu d dans le profil P. Une règle d'agrégation vérifie **(ND)** si elle n'admet aucun dictateur :
$$\forall d \in I,\ \ \exists P \in \mathcal{L}(A)^n,\ \ \succ_P \neq \succ_d .$$

Résultat admis. On peut toujours construire un classement qui place des options choisies en tête, dans un ordre choisi. Précisément : si $k \in [\![1; m]\!]$ et si $a_1, \ldots, a_k \in A$ sont deux à deux distincts, il existe $\succ \in \mathcal{L}(A)$ tel que
$$a_1 \succ a_2 \succ \cdots \succ a_k \qquad \text{et} \qquad \forall x \in A \setminus \{a_1, \ldots, a_k\},\ \ a_k \succ x.$$
Un tel classement est noté $a_1 \succ \cdots \succ a_k \succ \cdots$: les options $a_1, \ldots, a_k$ sont en tête, dans cet ordre, et les autres suivent dans un ordre quelconque. Par exemple, si $A = \{a, b, c, d\}$, la notation $b \succ a \succ \cdots$ désigne l'un des deux classements $b \succ a \succ c \succ d$ ou $b \succ a \succ d \succ c$.

Théorème d'Arrow. *On suppose $m \geq 3$. Aucune règle d'agrégation ne vérifie à la fois **(U)**, **(IIA)** et **(ND)**.*

Autrement dit, dès qu'il y a au moins trois options, une règle qui respecte **(U)** et **(IIA)** admet forcément un dictateur : c'est sous cette forme que l'on démontre le théorème dans la partie II.

Partie I. Préliminaires et paradoxe de Condorcet

1) Soient $\succ \in \mathcal{L}(A)$ et $a, b \in A$. Montrer que si $a \succ b$, alors $a \neq b$ et $b \not\succ a$ (on dit que $\succ$ est **asymétrique**). En déduire que si $a \neq b$, une et une seule des deux relations $a \succ b$ et $b \succ a$ est vraie.

2) **Dictatures.** Pour $d \in I$, on note F_d la règle d'agrégation qui renvoie toujours le classement de l'individu d :

$$F_d : \mathcal{L}(A)^n \longrightarrow \mathcal{L}(A), \quad (\succ_1, \dots, \succ_n) \longmapsto \succ_d .$$

a) Montrer que F_d vérifie **(U)** et **(IIA)**.

b) Soient F une règle d'agrégation et $d \in I$. Montrer que d est un dictateur pour F si et seulement si

$$\forall P \in \mathcal{L}(A)^n, \quad \forall a, b \in A, \quad a \succ_d b \implies a \succ_P b.$$

Autrement dit, pour montrer que d est un dictateur, il suffit de montrer que le groupe suit l'avis de d chaque fois que d préfère une option à une autre.

3) **Le paradoxe de Condorcet.** Pour un profil P et $x, y \in A$, on note $N_P(x, y)$ le nombre d'individus qui préfèrent x à y :

$$N_P(x, y) = \mathrm{Card}\left(\{i \in I \mid x \succ_i y\}\right).$$

On définit la **relation majoritaire** M_P sur A par : $x \, M_P \, y$ si une majorité stricte d'individus préfère x à y, c'est-à-dire

$$x \, M_P \, y \iff N_P(x, y) > \frac{n}{2}.$$

On suppose $n = 3$ (trois individus), on fixe trois options distinctes a, b, c, et l'on considère un profil P dans lequel les trois individus placent a, b, c en tête de la façon suivante (s'il y a d'autres options, elles sont classées après, dans un ordre quelconque) :

$$\begin{aligned} &\text{individu 1 :} \quad a \succ b \succ c \succ \cdots \\ &\text{individu 2 :} \quad b \succ c \succ a \succ \cdots \\ &\text{individu 3 :} \quad c \succ a \succ b \succ \cdots \end{aligned}$$

Montrer que $a \, M_P \, b$, $b \, M_P \, c$ et $c \, M_P \, a$. En déduire que M_P n'est pas transitive, puis que l'application $P \longmapsto M_P$ n'est pas une règle d'agrégation.

Partie II. Démonstration du théorème d'Arrow

Dans toute cette partie, on fixe une règle d'agrégation F qui vérifie **(U)** et **(IIA)**, et l'on va montrer qu'elle admet un dictateur. L'idée est d'étudier les groupes d'individus capables d'imposer leur avis à tous les autres.
Une partie S de I, c'est-à-dire un groupe d'individus (éventuellement vide), est appelée une **coalition**. Soient S une coalition et $a, b \in A$ deux options distinctes.

- On dit qu'un profil P **sépare** S **selon** (a, b) si tous les membres de S préfèrent a à b et si tous les autres individus préfèrent b à a :

$$\forall i \in S, \quad a \succ_i b \qquad \text{et} \qquad \forall i \in I \setminus S, \quad b \succ_i a.$$

- On dit que S est **décisive pour** (a, b) si, pour tout profil P qui sépare S selon (a, b), on a $a \succ_P b$. Autrement dit, chaque fois que les membres de S préfèrent a à b alors que tous les autres pensent le contraire, c'est l'avis de S qui l'emporte dans le classement collectif.
- On dit que S est **décisive** si elle est décisive pour tout couple (a, b) d'options distinctes. On note $\mathcal{D}$ l'ensemble des coalitions décisives : c'est un ensemble de parties de I.

Exemple. Avec $n = 3$ et $S = \{1, 2\}$, tout profil dans lequel les individus 1 et 2 préfèrent a à b et l'individu 3 préfère b à a sépare S selon (a, b), quelle que soit la place des autres options.
Attention à l'ordre. Être décisive pour (a, b) et être décisive pour (b, a) sont deux propriétés différentes.
Plan. La Question 6) montre qu'une coalition décisive pour un seul couple l'est pour tous ; les questions suivantes étudient l'ensemble $\mathcal{D}$; enfin, la Question 10) en extrait une coalition décisive réduite à un seul individu, et la Question 11) montre que cet individu est un dictateur.

4) Dans cette question, S est une coalition et $a, b \in A$ sont deux options distinctes.

a) Justifier qu'il existe au moins un profil qui sépare S selon (a, b).
Indication : on pourra utiliser le résultat admis.

b) Montrer que S est décisive pour (a, b) si et seulement s'il existe *au moins un* profil P qui sépare S selon (a, b) et tel que $a \succ_P b$.
Ainsi, pour montrer qu'une coalition est décisive pour (a, b), il suffit de trouver un seul profil qui la sépare selon (a, b) et dans lequel le groupe préfère a à b.
*Indication : pour le sens réciproque, si P et P' séparent tous les deux S selon (a, b), comparer, pour chaque individu i, les assertions « $a \succ_i b$ » et « $a \succ'_i b$ », puis utiliser **(IIA)**.*

5) Montrer que la coalition I, formée de tous les individus, est décisive. Montrer ensuite que la coalition vide $\varnothing$ n'est décisive pour aucun couple (a,b) d'options distinctes ; en particulier, $\varnothing \notin \mathcal{D}$.
Indication : utiliser ***(U)****.*

6) **Propagation.** On montre dans cette question qu'une coalition décisive pour un couple d'options est décisive pour tous les couples. Soient S une coalition et $a,b \in A$ deux options distinctes telles que S soit décisive pour (a,b). Comme $m \geq 3$, on peut choisir une troisième option $c \in A \setminus \{a,b\}$.

a) On considère un profil P (il en existe d'après le résultat admis) dans lequel les classements sont les suivants :

$$\begin{array}{ll} \text{membres de } S : & a \succ b \succ c \succ \cdots \\ \text{individus de } I \setminus S : & b \succ c \succ a \succ \cdots \end{array}$$

Montrer que $a \succ_P b$ et $b \succ_P c$, en déduire que $a \succ_P c$, puis que S est décisive pour (a,c).
On admet que l'on montre de même, avec un autre profil bien choisi, que S est décisive pour (c,b). Ce raisonnement s'applique à n'importe quel couple pour lequel S est décisive et à n'importe quelle troisième option ; on retiendra donc la règle suivante.
(R) Pour toutes options u,v,w deux à deux distinctes, si S est décisive pour (u,v), alors :
- **(R1)** S est décisive pour (u,w) ;
- **(R2)** S est décisive pour (w,v).

b) En déduire que $S \in \mathcal{D}$, c'est-à-dire que S est décisive pour tout couple (x,y) d'options distinctes.
Indication : appliquer plusieurs fois la règle ***(R)****, en montrant successivement que S est décisive pour les couples (a,z) avec $z \neq a$, puis pour les couples (x,y) avec $x \neq y$, $x \neq a$ et $y \neq a$, et enfin pour les couples (x,a) avec $x \neq a$.*

7) **Complémentaire.** Soit S une coalition. Montrer que $S \in \mathcal{D}$ ou $I \setminus S \in \mathcal{D}$: de la coalition S et de la coalition formée de tous les autres individus, l'une au moins est décisive.
Indication : fixer deux options distinctes a,b et un profil P qui sépare S selon (a,b), puis distinguer les cas $a \succ_P b$ et $b \succ_P a$.

8) **Intersection.** Soient $S,T \in \mathcal{D}$ deux coalitions décisives et $a,b,c \in A$ trois options distinctes. Chaque individu appartient à exactement un des quatre ensembles $S \cap T$, $S \setminus T$, $T \setminus S$ et $I \setminus (S \cup T)$ (certains peuvent être vides). On considère un profil P (il en existe d'après le résultat admis) dans lequel les classements sont les suivants :

$$\begin{array}{ll} \text{individus de } S \cap T : & a \succ b \succ c \succ \cdots \\ \text{individus de } S \setminus T : & c \succ a \succ b \succ \cdots \\ \text{individus de } T \setminus S : & b \succ c \succ a \succ \cdots \\ \text{individus de } I \setminus (S \cup T) : & c \succ b \succ a \succ \cdots \end{array}$$

a) Montrer que $a \succ_P b$ et que $b \succ_P c$.
Indication : déterminer quels individus préfèrent a à b, puis quels individus préfèrent b à c.

b) En déduire que $S \cap T \in \mathcal{D}$.
Indication : déterminer quels individus préfèrent a à c, puis utiliser la Question 4)b) et la Question 6)b).

9) **Sur-ensembles.** Soient $S \in \mathcal{D}$ et T une coalition telle que $S \subset T$. Montrer que $T \in \mathcal{D}$: une coalition qui contient une coalition décisive est elle-même décisive.
Indication : raisonner par l'absurde, et utiliser la Question 7) puis la Question 8)b).

10) **Une coalition décisive réduite à un individu.**

a) Justifier qu'il existe une coalition décisive S_0 de cardinal minimal, c'est-à-dire telle que $\mathrm{Card}(S_0) \leqslant \mathrm{Card}(S)$ pour toute coalition $S \in \mathcal{D}$. Justifier ensuite que $S_0 \neq \varnothing$.

b) On suppose dans cette question que $\mathrm{Card}(S_0) \geq 2$. On peut alors écrire $S_0 = S_1 \cup S_2$, où S_1 et S_2 sont deux coalitions non vides et disjointes (par exemple $S_1 = \{i\}$ pour un $i \in S_0$, et $S_2 = S_0 \setminus \{i\}$). Montrer que $S_1 \in \mathcal{D}$ ou $S_2 \in \mathcal{D}$, puis aboutir à une contradiction.
Indication : si $S_1 \notin \mathcal{D}$, utiliser la Question 7) puis la Question 8)b).

c) En déduire qu'il existe un individu $d \in I$ tel que $\{d\} \in \mathcal{D}$.

11) Soit d l'individu obtenu à la Question 10)c). Montrer que d est un dictateur pour F.
Indication : d'après la Question 2)b), il suffit de montrer que $a \succ_P b$ pour tout profil P et toutes options a,b telles que $a \succ_d b$. On pourra considérer l'ensemble $T = \{i \in I \mid a \succ_i b\}$ des individus qui préfèrent a à b dans le profil P.

12) **Bilan.** Déterminer toutes les règles d'agrégation qui vérifient **(U)** et **(IIA)**. Combien y en a-t-il ? On pourra supposer $n \geq 2$, le cas $n = 1$ étant immédiat.
Indication : on vérifiera que les règles F_d de la Question 2) sont deux à deux distinctes.
Remarques.
- L'hypothèse $m \geq 3$ est indispensable : pour deux options et un nombre impair $n \geq 3$ d'individus, le vote à la majorité vérifie **(U)**, **(IIA)** et **(ND)**.
- La méthode de Borda (chaque individu donne $m-1$ points à son option préférée, $m-2$ à la suivante, …, 0 à la dernière, les ex æquo étant départagés par ordre alphabétique) vérifie **(U)** et, pour $n \geq 2$, **(ND)** : d'après le théorème d'Arrow, elle ne vérifie donc pas **(IIA)**.

Partie III. Bonus : une échappatoire

Le théorème d'Arrow porte sur des règles définies sur *tous* les profils. Dans la pratique, certains profils sont peu plausibles : si les candidats se répartissent sur un axe gauche-droite, un électeur a un candidat favori, et apprécie d'autant moins un candidat qu'il en est éloigné. On montre dans cette partie que, si l'on se limite à de tels profils, le paradoxe de Condorcet de la Question 3) ne peut plus se produire : une option bat toutes les autres dans un duel à la majorité. On suppose dans cette partie que n est impair et que les options sont les entiers $1, 2, \dots, m$: $A = [\![1; m]\!]$. On les imagine placées de gauche à droite sur un axe, comme des candidats rangés de la gauche à la droite de l'échiquier politique.

Classements unimodaux. Soit $p \in A$. On dit qu'un classement $\succ \in \mathcal{L}(A)$ est **unimodal de pic** p si

$$p \succ p-1 \succ p-2 \succ \cdots \succ 1 \qquad \text{et} \qquad p \succ p+1 \succ p+2 \succ \cdots \succ m$$

(si $p = 1$ ou $p = m$, l'une des deux chaînes se réduit à p). Autrement dit, l'option préférée est p ; parmi les options situées à gauche de p, on préfère celle qui est la plus proche de p, et de même parmi les options situées à droite de p. En revanche, on n'impose rien pour comparer une option située à gauche de p et une option située à droite de p.

Exemple. Pour $m = 5$, le classement $3 \succ 4 \succ 2 \succ 1 \succ 5$ est unimodal de pic 3 : on a bien $3 \succ 2 \succ 1$ et $3 \succ 4 \succ 5$. Le classement $3 \succ 1 \succ 2 \succ 4 \succ 5$ n'est pas unimodal : son option préférée est 3, mais 1 y est préféré à 2, qui est pourtant plus proche de 3.

On pourra vérifier que, quelle que soit la façon de placer a, b et c sur l'axe, le profil de la Question 3) contient un classement qui n'est pas unimodal.

Pic médian. On considère un profil P dans lequel, pour tout $i \in I$, le classement $\succ_i$ est unimodal ; on note p_i son pic. On range les nombres $p_1, \dots, p_n$ dans l'ordre croissant, en répétant une valeur autant de fois qu'elle apparaît : on obtient $q_1 \leqslant q_2 \leqslant \dots \leqslant q_n$. Le terme du milieu de cette liste, $p^* = q_{(n+1)/2}$, est appelé le **pic médian**.

Exemple. Si $n = 5$ et si les pics sont $p_1 = 4$, $p_2 = 2$, $p_3 = 4$, $p_4 = 1$ et $p_5 = 5$, la liste rangée est $1, 2, 4, 4, 5$, donc $p^* = q_3 = 4$.

13) Le pic médian gagne tous ses duels. On montre dans cette question que $p^* \, M_P \, x$ pour toute option $x \neq p^*$, où M_P est la relation majoritaire définie à la Question 3) : dans un duel à la majorité, l'option p^* bat chacune des autres.

a) Montrer qu'au moins $\dfrac{n+1}{2}$ individus ont un pic supérieur ou égal à p^*, et qu'au moins $\dfrac{n+1}{2}$ individus ont un pic inférieur ou égal à p^*.

b) Soit $x \in A$ tel que $x < p^*$. Montrer que tout individu i dont le pic vérifie $p_i \geq p^*$ préfère p^* à x. En déduire que $p^* \, M_P \, x$.

c) Montrer de même que $p^* \, M_P \, x$ pour toute option $x > p^*$.

d) Conclure : que peut-on dire du vote à la majorité lorsque toutes les préférences sont unimodales ? Pourquoi cela ne contredit-il pas le théorème d'Arrow ?

e) Montrer que la relation M_P est un ordre total strict sur A : lorsque toutes les préférences sont unimodales, le vote à la majorité classe toutes les options, sans cycle.
Indication : pour la transitivité, si $x \, M_P \, y$ et $y \, M_P \, z$, raisonner par l'absurde, ne garder que les trois options x, y, z (rangées dans l'ordre croissant et renumérotées $1, 2, 3$), vérifier que les classements restent unimodaux, et appliquer ce qui précède.