Utilitarisme et théorie de l'utilité espérée: les deux théorèmes d'Harsanyi

Cyril Hédoin

Séminaire de spécialité "Economie normative et équité"

Master 2 – Parcours recherche 2014-15

Introduction : la démarche axiomatique en économie normative

- Question du statut de la connaissance en économie normative et en philosophie politique et morale
- Existent-ils des **faits** moraux et éthiques ?
  - Ex : « L'alternative A est plus inégalitaire que l'alternative B » est-il un jugement de fait ou un jugement de valeur ?
- Comment « tester » les propositions normatives, éthiques et morales ?

- Deux approches générales sur le plan méthodologique:
  - L'approche « **constructive** »
  - L'approche **axiomatique**
- L'approche constructive consiste à raisonner à partir d'un problème de décision hypothétique ou d'une expérience de pensée pour y appliquer des intuitions et des raisonnements moraux. Ex: le théorème de l'observateur impartial d'Harsanyi
- L'approche axiomatique consiste à déterminer quels sont les principes (axiomes) nécessaires et/ou suffisants pour qu'un raisonnement ou un jugement moral soit valide. Ex: le théorème d'agrégation d'Harsanyi

- La démarche axiomatique est dominante en économie normative, en particulier dans le cadre de la théorie du choix social
- Elle invite à plusieurs types de questionnements :
  - Quels sont les axiomes que doit respecter une doctrine morale pour être valide ?
  - Dans quelle mesure différents axiomes sont-ils compatibles entre eux ?
  - Dans quelle mesure un axiome donné est-il justifiable ?
- **L'intérêt principal de la démarche axiomatique est qu'elle permet de déterminer les implications et la cohérence de nos intuitions morales**

Plan

1. L'utilitarisme en économie et en philosophie

2. Les deux théorèmes utilitaristes d'Harsanyi
   a. Le théorème de l'observateur impartial
   b. Le théorème d'agrégation

3. Discussion

1. L'utilitarisme en économie et en philosophie

- L'utilitarisme est une doctrine politique et morale développée au 18ème siècle par Jeremy Bentham (1748 - 1832)
- Principe de la minimisation des peines et de la maximisation des plaisirs
- L'utilitarisme benthamien est ancré dans une psychologie hédoniste et repose sur l'hypothèse que l'utilité est une quantité mesurable
- John Stuart Mill ajoute à l'utilitarisme benthamien une distinction entre différents types de plaisirs

- Les premiers économistes marginalistes vont reprendre le concept d'utilité dans son sens Benthamien : une quantité cardinalement mesurable et comparable
- Rappel : une mesure $u(.)$ est
  - Ordinale : $u(x) \geq u(y)$ ssi $v(x) \geq v(y)$ avec v toute transformation monotone et positive de u
  - Cardinale à intervalle : $u(x) \geq u(y)$ ssi $v(x) \geq v(y)$ avec v toute transformation affine positive de u, i.e. $v = au + b$ avec $a > 0$
  - Cardinale à ratio : $u(x) \geq u(y)$ ssi $v(x) \geq v(y)$ avec v toute transformation linéaire positive de u, i.e. $v = au$ avec $a > 0$
- Avec Pareto, les économistes vont remettre en cause l'idée d'utilité cardinale; avec Robbins, ils vont rejeter la possibilité de comparaisons interpersonnelles d'utilité

- En économie, l'utilité n'est plus définie comme une quantité mais comme une représentation de la satisfaction des préférences
  - Ecrire $u(x) > u(y)$ signifie que l'alternative *x* est strictement préférée à l'alternative *y*
- La théorie du consommateur (par exemple) repose uniquement sur la notion d'utilité ordinale
- Problème : avec des utilités ordinales et non comparables, l'utilitarisme n'a plus de sens → Faire la somme des utilités dans une population suppose que :
  - Les **différences** entre niveaux d'utilités sont significatives (nécessite au minimum une cardinalité par intervalle)
  - Les différences de niveaux d'utilités sont comparables d'une personne à l'autre (le paramètre *a* dans la transformation affine est commune à tous les individus)

- Harsanyi réintroduit l'utilitarisme via la **théorie de l'utilité espérée** (TUE)
- Selon la TUE, un agent rationnel choisit l'alternative *x* telle que

$$(1) \qquad \max_x u(x) = \sum_s p(s)u(x; s)$$

avec $p(.)$ une mesure probabiliste sur l'ensemble *S* des états de nature et *s* un état de nature

- Le théorème de l'utilité espérée montre que le choix d'un agent rationnel peut être **représenté** par (1) si ses préférences sont notamment :
  - Complètes
  - Transitives
  - Indépendantes sur le plan des états de nature
- Vocabulaire : on appelle *x* un **prospect** ou une **loterie** et chaque (*x*; *s*) un **résultat** ou une **conséquence** (note : un résultat est une loterie « dégénérée »)
- La fonction $u(.)$ est unique pour toute transformation affine positive; elle est cardinale à intervalle

- Exemple de prospect : Bob doit choisir entre faire des études de gestion (G) et des études de philosophie (P);
  - S'il choisit P et si la conjoncture est bonne (état *s*1), il aura un poste de professeur de philosophie
  - S'il choisit P et si la conjoncture est mauvaise (état *s*2), il sera chômeur
  - S'il choisit G, il aura un travail ennuyeux peu importe la conjoncture
  - On suppose que $p(s1) = p(s2) = ½$
- Formellement

	***s*1**	***s*2**
Prospect D	$u(D; s1) = 5$	$u(D; s2) = 5$
Prospect P	$u(P; s1) = 9$	$u(P; s2) = 0$

2. Les deux théorèmes d'Harsanyi

- L'utilitarisme développé par Harsanyi est dit ***preference-based*** → les utilités représentent les préférences des agents concernant les prospects
- Il s'agit de montrer que si les agents sont rationnels au sens de la TUE, alors le choix collectif sera utilitariste; i.e. l'alternative sociale choisie sera celle qui maximise la somme (pondérée ou non) des utilités individuelles
- Vocabulaire : on appelle **alternative sociale** un prospect de type matriciel

	$s1$	$s2$
Ann	$u_{Ann}(.; s1)$	$u_{Ann}(.; s2)$
Bob	$u_{Bob}(.; s1)$	$u_{Bob}(.; s2)$

Le théorème de l'observateur impartial

- Le théorème de l'observateur impartial (TOI) relève d'une approche constructive
- Il consiste à déterminer quelle alternative sociale sera préférée par un individu rationnel placé **sous un voile d'ignorance**
- Sous voile d'ignorance, l'individu ignore :
  - Sa position sociale et les caractéristiques associées (niveau de revenu, état de santé, etc.)
  - Son identité personnelle et les caractéristiques associées (préférences, histoire personnelle, etc.)
- Formellement, chaque alternative sociale est un prospect matriciel avec une probabilité *p* associée à différentes positions sociales et une probabilité *q* associée à différentes identités sociales

- Ex : soit une société avec trois individus (Ann, Bob et Chris) et trois états *s*1, *s*2 et *s*3 :
  - Dans *s*1, Ann est riche et est bonne santé, Bob est riche et est en mauvaise santé, Chris est pauvre et en mauvaise santé
  - Dans *s*2, Ann est pauvre et est en mauvaise santé, Bob est riche et en bonne santé, Chris est riche et en mauvaise santé
  - Dans *s*3, Ann est riche et est en mauvaise santé, Bob est pauvre et est en mauvaise santé, Chris est riche et en bonne santé
- On suppose qu' Ann accorde plus d'importance à la richesse qu'à sa santé, l'inverse pour Bob, et que Chris accorde autant d'importance aux deux (i.e. leurs préférences diffèrent)

	*s*1	*s*2	*s*3
Ann	10	0	6
Bob	4	10	0
Chris	0	5	10

- Si les agents sont rationnels au sens de la TUE, on peut simplifier en ignorant les états de nature et en s'intéressant directement à l'utilité espérée de chaque personne (on suppose que les états sont equi-probables)…

	Ann	Bob	Chris
u	16/3	14/3	5

- … et on compare avec d'autres prospects possibles, par exemple

	Ann	Bob	Chris
u	10	4	0

- Selon Harsanyi, derrière le voile d'ignorance, l'observateur doit donner la même probabilité au fait de revêtir chaque identité possible → **Impartialité**
- Quelle sera l'alternative sociale choisie par un observateur rationnel ? **Celle qui maximise son utilité moyenne**

$$(2)\ \max_x U(x) = 1/n \sum_j u_j(x)$$

- Cela suppose que l'observateur puisse comparer les utilités des différents personnes → L'observateur s'appuie sur ses **préférences étendues** : « je préfère être Bob lorsqu'il est pauvre et en bonne santé qu'être Ann lorsqu'elle est pauvre et en bonne santé »
- Le choix de l'observateur impartial est un choix classique entre plusieurs prospects « en ligne » où les identités sociales remplacent les états de nature

Le théorème d'agrégation

- Le théorème d'agrégation relève de la démarche axiomatique
  - On pose un certains nombre de conditions (axiomes) et on détermine quelle fonction de bien-être social les satisfait
- On suppose ici une population de *n* personnes et un « dictateur bienveillant » dont les préférences sur les alternatives sociales dépendent des préférences des personnes

- Soit ΔX l'ensemble des prospects possibles
- Les axiomes sont les suivants :

A1 : Les *n* membres de la population ont des préférences qui satisfont la TUE → chaque individu *i* a une fonction $u_i(.)$ définie sur ΔX

A2 : Le dictateur bienveillant *k* a des préférences qui satisfont la TUE → *k* a une fonction U_K définie sur ΔX

A3 : La relation entre les préférences des *n* membres et celles du dictateur *k* satisfait le principe de Pareto *ex ante* : si $u_i(x) > u_i(y)$ pour un individu *i* et $u_j(x) \geq u_j(y)$ pour le reste des individus *j*, alors $U_K(x) > U_k(y)$

- **Théorème** : si les axiomes A1, A2 et A3 sont satisfaits, alors les préférences du dictateur *k* peuvent être représentées par la fonction suivante :

$$(3)\ U_k(x) = \sum_i \alpha_i u_i(x)$$

- (3) correspond à **l'utilitarisme pondéré**
- Selon Harsanyi, si les utilités sont comparables, l'impartialité implique que α_i est unique ; i.e. (3) satisfait une condition d'anonymat et devient

$$(4) \quad U_k(x) = \sum_i u_i(x)$$

- (4) satisfait par ailleurs une condition de « séparabilité croisée », i.e. on a à la fois
  - $U_k(.) = f[u_1(.), \ldots, u_n(.)]$
  - $U_k(.) = g[U_k(., s1), \ldots, U_k(., sm)]$
- La séparabilité croisée implique la séparabilité additive (Gorman) :

$$(5)\ U_k(x) = \sum_i \sum_j p(sj) u_i(x, sj)$$

- Autrement dit, pour obtenir l'utilité totale, on peut soit **additionner les utilités des individus par état puis additionner l'utilité des états**, soit **additionner les utilités des états par individu, puis additionner les utilités des individus**

3. Discussion

- Les deux théorèmes d'Harsanyi font un lien entre rationalité (au sens de la TUE) et utilitarisme
- L'implication semble être **qu'un agent rationnel et impartial est nécessairement utilitariste**
- Corolaire : la rationalité et l'impartialité semblent impliquer que l'égalité n'est pas une valeur moralement pertinente
- Harsanyi renouvelle les fondements de l'utilitarisme en éliminant sa dimension hédoniste et psychologique

- Les théorèmes d'Harsanyi peuvent toutefois être critiqués de nombreuses manières…
- … notamment au niveau de leurs supposées implications utilitaristes
- On peut distinguer deux types de critiques :
  - Une critique qui concerne les deux théorèmes simultanément
  - Des critiques qui portent sur l'un ou l'autre théorème

La critique sur les implications utilitaristes des théorèmes

- Selon A. Sen, (2) et (4) ne sont pas utilitaristes car les fonctions $u_i(.)$ ne sont en réalité **pas cardinales**
- On peut légitimement appliquer n'importe quelle transformation monotone positive à u_i …
- … et dans ce cas la forme additive de U_k disparait.
- Ex : si $u_{Ann} = 9$, $u_{Bob} = 4$ et $u_{Chris} = 1$ et que l'on applique la transformation $v_i = \sqrt{u_i}$, alors

$$U_k = 3^2 + 2^2 + 1^2$$

- L'utilité totale n'est plus la somme des utilités, **mais la somme des carrés des utilités** → Relation non linéaire entre utilité individuelle et utilité sociale

Les critiques à l'encontre du TOI

- Plusieurs critiques plus ou moins fortes :
  - Critique de l'hypothèse d'équi-probabilité (cf. textes pour la semaine prochaine)
  - Critique de la notion de « préférences étendues »
  - Critique de la mesure des utilités (attitudes face au risque de l'observateur vs attitudes face au risque des membres de la population)
  - Critique de la notion de voile d'ignorance

Les critiques à l'encontre du TA

- Les critiques portent essentiellement sur la pertinence des axiomes :
  - Critique du principe de Pareto *ex ante*
  - Critique de la pertinence de l'axiome d'indépendance de la TUE sur le plan moral
  - Critique de la pertinence de l'axiome de transitivité de la TUE

Pareto *ex ante* ?

- Quel est le prospect socialement préférable ?

	s1	*s2*
Ann	9	11
Bob	11	9

ou

	s1	*s2*
Ann	16	5
Bob	5	16

Indépendance par rapport aux états de nature ?

- Quel est le prospect socialement préférable ?

	s1	*s2*
Ann	1	0
Bob	0	1

ou

	s1	*s2*
Ann	1	1
Bob	0	0