Fair Divsion in Theory and Practice

Ron Cytron (Computer Science)
Maggie Penn (Political Science)

Lecture 6-b: Arrow's Theorem

Arrow's Theorem

The general question: Given a collection of individuals with heterogeneous preferences over a set of alternatives, how might we aggregate their different preferences in order to derive a social preference over (or choice from) our set?

Beginning notation and definitions:

A set of alternatives X, with elements $a, b, c, d, w, x, y, z, ...$

A *binary relation*, used to compare any two elements of X

- $\geq$
- "Is older than"
- "Is preferred by a majority of people to"

We'll deal with a type of binary relation called a *weak preference relation*, $\succeq$ and $\succeq_i$

$\succeq$ is the weak social preference relation, and $\succeq_i$ is individual i's weak preference relation

- $x \succeq_i y$ implies that x is at least as good as y, for person i
- $x \succeq y$ implies that x is at least as good as y for the group

$\succeq$ will ultimately stem from group decision-making procedure

Given a weak preference relation $\succeq$ we can define two related binary relations: $\succ$ and $\sim$

- $\succ$ is *strict preference*: $x \succ y \Leftrightarrow x \succeq y$ and $\neg[y \succeq x]$
- $\sim$ is *indifference*: $x \sim y \Leftrightarrow x \succeq y$ and $y \succeq x$

Assumptions about individual preferences

- We assume that a person's preference relation is a weak order: transitive, reflexive and complete
- Transitive: $x \succeq_i y$ and $y \succeq_i z \Rightarrow x \succeq_i z$
- Reflexive: $x \succeq_i x$
- Complete: for $x \neq y$, $x \succeq_i y$ or $y \succeq_i x$ or both

- $\mathcal{R}$ is the set of all possible weak orders over X, so that $\succeq_i \in \mathcal{R}$

Example: $X = \{a, b, c\}$. Then

$$\mathcal{R} = \left\{ \begin{array}{cc} a \succ b \succ c, & a \succ b \sim c \\ a \succ c \succ b, & b \succ a \sim c \\ b \succ a \succ c, & c \succ a \sim b \\ b \succ c \succ a, & b \sim c \succ a \\ c \succ b \succ a, & a \sim c \succ b \\ c \succ a \succ b, & a \sim b \succ c \\ \multicolumn{2}{c}{a \sim b \sim c} \end{array} \right\}$$

- If people are never indifferent between alternatives, then preferences are a *strict order*, $\succ_i$ (transitive, complete and asymmetric)
- $\mathcal{P}$ is the set of all strict orders, with $\succ_i \in \mathcal{P}$
- In above example (with three alternatives)

$$\mathcal{P} = \left\{ \begin{array}{c} a \succ b \succ c \\ a \succ c \succ b \\ b \succ a \succ c \\ b \succ c \succ a \\ c \succ b \succ a \\ c \succ a \succ b \end{array} \right\}$$

- A *preference profile*, ρ, is an n-tuple of weak orders: $\rho = (\succeq_1, ..., \succeq_n)$
- ρ describes the preferences of all individuals within society
- The set of all preference profiles is $\mathcal{R}^n$, with $\rho \in \mathcal{R}^n$
- $\mathcal{R}^n$ is the Cartesian product of $\mathcal{R}$ with itself n times, or $\mathcal{R} \times \mathcal{R} \times ... \times \mathcal{R} = \{(\succeq_1, \succeq_2, ..., \succeq_n) : \succeq_i \in \mathcal{R} \text{ for all } i\}$

For any $\rho \in \mathcal{R}^n$ and $S \subseteq X$, let $\rho|_S$ be the restriction of ρ to the set S

Example: Let

$$\rho = \begin{pmatrix} a \succ b \succ c \\ a \succ c \succ b \\ b \succ a \succ c \end{pmatrix}$$

Then

$$\rho|_{\{a,c\}} = \begin{pmatrix} a \succ c \\ a \succ c \\ a \succ c \end{pmatrix}$$

Last, let $\mathcal{B}$ be the set of all reflexive and complete binary relations on X (but not necessarily transitive)

A *Preference Aggregation Rule* is a mapping, $F : \mathcal{R}^n \rightarrow \mathcal{B}$

In other words, a preference aggregation rule takes in a preference profile and generates a social preference relation (that is not guaranteed to be transitive); we'll call this binary relation $\succeq$, with asymmetric part $\succ$.

A little more notation

- Sometimes we will use "R" to refer to $\succeq$ (as in "xRy")
- Sometimes we will use "P" to refer to $\succ$ (as in "xPy")
- $P(x, y; \rho)$ is the set of people who strictly prefer x to y at profile ρ
- Formally, $P(x, y; \rho) = \{i \in N : xP_iy\}$

Aggregation rule examples

- Simple majority: $\forall x, y \in X$, xPy if and only if $|P(x, y; \rho)| > \frac{n}{2}$
- i is a dictator: $\forall x, y \in X$, xPy if and only if xP_iy
- Borda count: Let $r_i(x)$ be the ordinal ranking of x in i's preference ordering; then xPy if and only if $\sum_{i \in N} r_i(x) < \sum_{i \in N} r_i(y)$
- Plurality: xPy if and only if $|\{i \in N : r_i(x) = 1\}| > |\{i \in N : r_i(y) = 1\}|$

Unrestricted domain

Recall that an aggregation rule is a mapping, $F : \mathcal{R}^n \to \mathcal{B}$

The fact that the domain of F is all of $\mathcal{R}^n$ is referred to as the condition of unrestricted domain (or universal domain)

It implies that *all* preference profiles are possible–in other words, the only restriction we place on individual preferences is that they be transitive

Four criteria—Aggregation rule F is:

- *nondictatorial* if $\nexists i \in N$ such that $\forall \rho \in \mathcal{R}^n$ and for any $x, y \in X$, $x \succ_i y$ implies $x \succ y$

- In words: There is no person ("i") so that for EVERY possible profile and EVERY possible pair x, y, if i strictly prefers x to y then x is strictly socially preferred to y

- *weakly Paretian* if for every $\rho \in \mathcal{R}^n$ and any $x, y \in X$, $x \succ_i y$ $\forall i \in N$ implies $x \succ y$

- In words: For every pair, if EVERY person strictly prefers x to y then x is strictly (socially) preferred to y

- *independent of irrelevant alternatives* if for every $\rho, \rho' \in \mathcal{R}^n$ and for any $x, y \in X$, if $\rho|_{\{x,y\}} = \rho'|_{\{x,y\}}$ then $x \succeq y$ if and only if $x \succeq' y$

- In words: If ρ and ρ' look the same on x, y then $\succeq$ and $\succeq'$ produce the same x, y social ranking

- *transitive*, if for all $\rho \in \mathcal{R}^n$, $\succeq$ is transitive

- In words: If x is weakly (socially) preferred to y and y is weakly (socially) preferred to z, then x is weakly (socially) preferred to z

Example: Plurality violates IIA and weak Pareto

Consider this 2-player preference profile

$$x \succ z \succ y \succ w \ \text{ and } \ y \succ x \succ z \succ w$$

Under plurality we get a transitive social ranking

$$x \sim y \succ z \sim w$$

$$x \succ z \succ y \succ w \quad \text{and} \quad y \succ x \succ z \succ w$$

$$\text{Social ranking: } x \sim y \succ z \sim w$$

Plurality in this example violates weak Pareto: for all $i \in N$, zP_iw, but zIw socially (it's not the case that zPw)

Now consider the following two profiles:

$$\rho_1 : x \succ_1 z \succ_1 y \succ_1 w \;\text{ and }\; y \succ_2 x \succ_2 z \succ_2 w$$

Social ranking: $x \sim y \succ z \sim w$

$$\rho_2 : z \succ_1 x \succ_1 y \succ_1 w \;\text{ and }\; y \succ_2 x \succ_2 z \succ_2 w$$

Social ranking: $z \sim y \succ x \sim w$

Clearly $\rho_1|_{\{x,y\}} = \rho_2|_{\{x,y\}}$; both have $x \succ_1 y$ and $y \succ_2 x$

But socially, xIy under ρ_1 and yPx under ρ_2; IIA is violated

Arrow's General Possibility / Impossibility Theorem:

If an aggregation rule is transitive, weakly Paretian and independent of irrelevant alternatives, then it is dictatorial

Unrestricted domain is also essential, but is incorporated into our definition of an aggregation rule

We will illustrate with an example: 2 people, 3 alternatives, and strict preferences

	xyz	xzy	yxz	yzx	zxy	zyx
xyz						
xzy						
yxz						
yzx						
zxy						
zyx						

First, we fill in what we can using the Pareto condition

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xy, xz	xz, yz	yz	xy	
xzy	xz, xy	xzy	xz		xy, zy	zy
yxz	xz, yz	xz	yxz	yz, yx		yx
yzx	yz		yz, yx	yzx	zx	yx, zx
zxy	xy	xy, zy		zx	zxy	zx, zy
zyx		zy	yx	yx, zx	zx, zy	zyx

That only got us so far in filling out the table. Now we make a tie-breaking assumption in ONE cell. Assume that $x \succ y$ in cell (1, 4). With only this assumption we can now fill in the rest of the table.

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xy, xz	xz, yz	yz	xy	
xzy	xz, xy	xzy	xz		xy, zy	zy
yxz	xz, yz	xz	yxz	yz, yx		yx
yzx	yz		yz, yx	yzx	zx	yx, zx
zxy	xy	xy, zy		zx	zxy	zx, zy
zyx		zy	yx	yx, zx	zx, zy	zyx

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xy, xz	xz, yz	xy, yz	xy	
xzy	xz, xy	xzy	xz		xy, zy	zy
yxz	xz, yz	xz	yxz	yz, yx		yx
yzx	yz		yz, yx	yzx	zx	yx, zx
zxy	xy	xy, zy		zx	zxy	zx, zy
zyx		zy	yx	yx, zx	zx, zy	zyx

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xy, xz	xz, yz	xyz	xy	
xzy	xz, xy	xzy	xz		xy, zy	zy
yxz	xz, yz	xz	yxz	yz, yx		yx
yzx	yz		yz, yx	yzx	zx	yx, zx
zxy	xy	xy, zy		zx	zxy	zx, zy
zyx		zy	yx	yx, zx	zx, zy	zyx

Using the rankings we generated through the Pareto assumption and our assumption that $x \succ y$ in cell (1, 4) we can start applying IIA. We'll begin with the (x, y) social rankings. We'll also simultaneously apply transitivity if we can; so if we have "xy, yz" in a cell we'll automatically write it as "xyz".

First, we can say that $x \succ y$ in cells (1, 3), (1, 6), (2, 3), (2, 4), (2, 6), (5, 3), (5, 4), and (5, 6), because in all of these cells Player 1 prefers x to y and Player 2 prefers y to x.

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xy, xz	xyz	xyz	xy	xy
xzy	xz, xy	xzy	xz, xy	xy	xy, zy	xy, zy
yxz	xz, yz	xz	yxz	yz, yx		yx
yzx	yz		yz, yx	yzx	zx	yx, zx
zxy	xy	xy, zy	xy	zxy	zxy	zxy
zyx		zy	yx	yx, zx	zx, zy	zyx

Similarly, we will now determine as many (x, z) and (z, y) rankings as possible. In the previous step we used cell (1, 4) as our starting point. In this step we'll use cells (1, 4) and (5, 4) as our starting points.

In cell (1, 4) we know by transitivity that $x \succ z$. Thus, whenever Player 1 has preference $x \succ_1 z$ and Player 2 has $z \succ_2 x$, then socially $x \succ z$. This occurs in cells (1, 5), (1, 6), (2, 4), (2, 5), (2, 6), (3, 4), (3, 5), and (3, 6).

In cell (5, 4) we know by transitivity that $z \succ y$. Thus, whenever Player 1 has preference $z \succ_1 y$ and Player 2 has $y \succ_2 z$, then socially $z \succ y$. This occurs in cells (2, 1), (2, 3), (2, 4), (5, 1), (5, 3), (6, 1), (6, 3), and (6, 4).

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xy, xz	xyz	xyz	xy, xz	xy, xz
xzy	xzy	xzy	xzy	xz, zy	xzy	xzy
yxz	xz, yz	xz	yxz	yxz	xz	yxz
yzx	yz		yz, yx	yzx	zx	yx, zx
zxy	xy, zy	xy, zy	xy, zy	zxy	zxy	zxy
zyx	zy	zy	zyx	zyx	zx, zy	zyx

Now we will determine as many (y, z) and (z, x) rankings as possible. Again, we'll do this exactly the same way as in the previous step. In this step we'll use cells (3, 6) and (6, 3) as our starting points.

In cell (3, 6) we know by transitivity that $y \succ z$. Thus, whenever Player 1 has preference $y \succ_1 z$ and Player 2 has $z \succ_2 y$, then socially $y \succ z$. This occurs in cells (1, 2), (1, 5), (1, 6), (3, 2), (3, 5), (4, 2), (4, 5), (4, 6).

Similarly, in cell (6, 3) we know by transitivity that $z \succ x$. Thus, whenever Player 1 has preference $z \succ_1 x$ and Player 2 has $x \succ_2 z$, then socially $z \succ x$. This occurs in cells (4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (5, 3), (6, 1), (6, 2)

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xyz	xyz	xyz	xyz	xyz
xzy	xzy	xzy	xzy	xzy	xzy	xzy
yxz	xz, yz	xz, yz	yxz	yxz	xz, yz	yxz
yzx	yzx	yz, zx	yzx	yzx	yzx	yzx
zxy	zxy	zxy	zxy	zxy	zxy	zxy
zyx	zy, zx	zy, zx	zyx	zyx	zx, zy	zyx

We're almost done; we've applied IIA to (x, y), (x, z), (z, y), (y, z) and (z, x) pairs. All that is left is to apply IIA to the remaining (y, x) pairs whose rankings are still unspecified.

In this step we'll use cell (4, 5) as our starting point. In this cell we know, by transitivity, that $y \succ x$. Thus, whenever Player 1 has preference $y \succ_1 x$ and Player 2 has $x \succ_2 y$, then socially $y \succ x$. This occurs in cells (3, 1), (3, 2), (3, 5), (4, 1), (4, 2), (6, 1), (6, 2), (6, 5)

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	xyz	xyz	xyz	xyz	xyz	xyz
xzy	xzy	xzy	xzy	xzy	xzy	xzy
yxz	yxz	yxz	yxz	yxz	yxz	yxz
yzx	yzx	yzx	yzx	yzx	yzx	yzx
zxy	zxy	zxy	zxy	zxy	zxy	zxy
zyx	zyx	zyx	zyx	zyx	zyx	zyx

We've shown that the consequence of assuming $x \succ y$ in cell (1, 4) and requiring our aggregation rule F to satisfy transitivity, Pareto, IIA, and unrestricted domain is to make Player 1 a "dictator".

If we had assumed that $y \succ x$ in cell (1, 4), then (by the symmetry of the example) Player 2 would have turned out to be the dictator.

Question for you to think about on your own: What problem would have occurred if we had assumed that $x \sim y$ in cell (1, 4)?