Introduction to Mechanism Design

Pramook Khungurn

December 11, 2019

1 Social Choice Theory

- Let A be a finite set of *alternatives*.
- A *preference* $\prec$ on A is an anti-symmetric, transitive, and complete binary relation on A. That is, a preference is a *total order* on A.
- Let L denote the set of all preferences on A.
- We can think of a set of n voters, each with his own preferences over the alternatives. We typically write $\prec_i$ as the preference of Voter i.

  A *preference profile* is a tuple containing preferences of all voters. We typically use π to denote a profile, and we write $\pi = (\prec_1, \prec_2, \dots, \prec_n)$ to show the components of π.
- The set of preferences of all voters is precisely L^n.
- A function $F : L^n \to L$ is called a *social welfare* function.

  A function $f : L^n \to A$ is called a *social choice* function.

1.1 Arrow's Impossibility Theorem

- A social welfare function F is said to satisfies *unanimity* if it satisfies the following condition:

  > Let a and b be any two alternatives in A.
  > Let $\pi = (\prec_1, \prec_2, \dots, \prec_n)$ be any profile.
  > Let $\prec = F(\pi)$.
  >
  > If $a \prec_i b$ for all i, then $a \prec b$.

  That is, if all voters prefer b over a, then the group prefers b over a.
- A social welfare function F is said to satisfies *independence of irrelevant alternatives* (IIA) if it satisfies the following condition:

  > Let a and b be any two alternatives in A.
  > Let $\pi = (\prec_1, \prec_2, \dots, \prec_n)$ and $\pi' = (\prec'_1, \prec'_2, \dots, \prec'_n)$ be any two profiles.
  > Let $\prec = F(\pi)$ and $\prec' = F(\pi')$.
  >
  > If $a \prec_i b \iff a \prec'_i b$, then $a \prec b \iff a \prec' b$.

  That is, the preference between a and b of the group depends only on the preferences between a and b of the voters. Information regarding other alternatives are irrelevant.

- Voter i is said to be a *dictator* of social welfare function F if $F(\prec_1, \prec_2, \ldots, \prec_n) = \prec_i$ for all profiles.

  If F has a dictator, we say that F is a *dictatorship.*

- **Theorem 1.1 (Arrow's Impossibility Theorem).** *Every social welfare function on set A with more than two elements that satisfies unanimity and IIA is a dictatorship.*

  We first prove the following claim.

  Claim 1.2. *Let F be a social welfare function on set A with at least 3 elements that satisfies unanimity and IIA. Then, F satisfies the following* neutrality *condition:*

  > *Let a, b, α, β be any four alternatives in A such that $a \neq b$ and $\alpha \neq \beta$*
  > *Let $\pi = (\prec_1, \prec_2, \ldots, \prec_n)$ and $\pi' = (\prec'_1, \prec'_2, \ldots, \prec'_n)$ be any two profiles.*
  > *Let $\prec = F(\pi)$ and $\prec' = F(\pi')$.*
  >
  > *If $a \prec_i b \iff \alpha \prec'_i \beta$, then $a \prec b \iff \alpha \prec' \beta$.*

  That is, the decision is made the same way for every two alternatives.

  Proof. (Claim) WLOG, we can assume that, in profile π, the group prefers b over a. That is, $a \prec b$.

  Our strategy is to construct a new profile $\pi^* = (\prec^*_1, \prec^*_2, \ldots, \prec^*_n)$ with the following properties:

  1. the preferences between a and b are the same as those in π,
  2. the preferences between α and β are the same as those in π', and
  3. for all i, $\alpha \prec^*_i a$ and $b \prec^*_I \beta$.

  Assuming that we are successful, let $\prec^* = F(\pi^*)$. We have that:

  - By Property 1 and IIA, it is the case that $a \prec^* b$.
  - By Property 3 and unanimity, it is the case that $\alpha \prec^* a$ and $b \prec^* \beta$.
  - Thus, $\alpha \prec^* a \prec^* b \prec^* \beta$. Therefore, $\alpha \prec^* \beta$.
  - Lastly, by Property 2 and IIA, we have that $\alpha \prec' \beta$ as well.

  Hence, the claim is true if we can establishes the existence of π^*.

  To construct π^*, we first assume that $\alpha \neq b$ and $\beta \neq a$, and we will deal with other cases later. We now specify each $\prec^*_i$. We construct $\prec^*_i$ by first setting it to $\prec_i$. Then, we move α and β so that α is just before a (if $\alpha \neq a$) and β is just after b (if $\beta \neq b$). This is done in such a way that preserves the relative order between a and b and that between α and β. More precisely,

  - if $a \prec_i b$ and $\alpha \prec'_i \beta$, then we move α and β so that $\alpha \prec^*_i a \prec^*_i b \prec^*_i \beta$;
  - if $b \prec_i a$ and $\beta \prec'_i \alpha$, then we move α and β so that $b \prec^*_i \beta \prec^*_i \alpha \prec^*_i a$.

  It can be seen that the new profile π^* satisfies all the three properties.

  Now, we turn to cases where $\alpha = b$ or $\beta = a$. Let c be an alternative which is different from a and b. From what we have proved so far, we know that decisions about (a, b) are made in the same way as those about (a, c). These decisions, in turn, are made the same way as those about (b, c), and then (b, a), and then (c, a). We have covered all the cases. □

  Proof. (Arrow's impossibility theorem) Let a and b be two alternatives. We define a set of profiles π^0, π^1, π^2, ..., π^n as follows. In π^i, the first i voters prefers b over a, but the remaining voters prefer a over b. (That is, $a \prec_j b$ if $j \leq i$, and $b \prec_j a$ if $j > i$.)

Let $\prec^i = F(\pi^i)$ for all i. By unanimity, we have that $b \prec^0 a$, but $a \prec^n b$. Hence, there must be a profile i^* such that $b \prec^{i^*-1} a$, but $a \prec^{i^*} b$. We call voter i^* the *pivotal voter.*

We now show that i^* is a dictator. Let $\pi = (\prec_1, \prec_2, \ldots, \prec_n)$ be an arbitrary profile, and let $\prec = F(\pi)$. We shall show that, for any $c, d \in A$, if $c \prec_{i^*} d$, then $c \prec d$.

To do so, let e be an alternative different from c and d. By IIA, moving e around the preferences of any voter without changing the relative order between c and d does not change the group's preference between c and d. We then conduct the following moves:

- for $j < i^*$, we move e so that it is the least preferred alternative in $\prec_j$.
- for $j > i^*$, we move e so that it is the most preferred alternative in $\prec_j$.
- we move e so that $c \prec_i e \prec_i d$.

Note that the relative orders of c and e are the same as those of a and b in π^{i^*-1}. By neutrality, we can conclude that $c \prec e$. Note also that the relative orders of e and d are the same as those of b and a in π^{i^*}. Hence, $e \prec d$. So, we have that $c \prec e \prec d$. $\square$