Mechanism Design

Christoph Schottmüller

Outline

Bayesian implementation

- so far: dominant strategy equilibrium
  - problem 1: Gibbard-Satterthwaite theorem for general preferences
  - problem 2: budget balance problem even with quasilinear preferences
- from now on: Bayesian Nash equilibrium instead of dominant strategy equilibrium
  - upside: weaker concept $\rightarrow$ more social choice functions can be implemented
  - downside: weaker concept $\rightarrow$ less confident that players adhere to it in reality
  - ϕ becomes relevant
  - same problems/examples as before

Bayesian implementation

- I agents
- set of alternatives X
- nature draws types from the probability distribution ϕ
- type $\theta_i \in \Theta_i$ is private information of i
- given the distribution ϕ each agent uses Bayes' rule to form a belief over other agents' types

$$prob(\theta_i, \theta_{-i} | \theta_i) = \frac{\phi(\theta_i, \theta_{-i})}{\sum_{\tilde{\theta}_{-i} \in \Theta_{-i}} \phi(\theta_i, \tilde{\theta}_{-i})}$$

Bayesian Nash equilibrium

Definition (Bayesian Nash equilibrium)

The strategy profile $s^* = (s_1^*, \ldots, s_I^*)$ is a Bayesian Nash equilibrium of the mechanism $\Gamma = (S_1, \ldots, S_I, g)$ if, for all i and all $\theta_i \in \Theta_i$, $E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(g(\hat{s}_i, s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right]$ for all $\hat{s}_i \in S_i$.

- roughly: a strategy profile is a BNE if no type of no player can increase his expected payoff by deviating

Bayesian implementation

Definition (Bayesian implementation)

The mechanism $\Gamma = (S_1, \ldots, S_I, g)$ implements the social choice function f in Bayesian Nash equilibrium if there is a Bayesian Nash equilibrium $s^* = (s_1^*, \ldots, s_I^*)$ of Γ such that $f(\theta) = g(s^*(\theta))$ for all $\theta \in \Theta$.

- roughly: there has to be a BNE such that in this BNE the desired outcome results for every possible type vector

Truthful Bayesian implementation

- recall direct mechanism: every player is asked for his type and the outcome is $f(announcement)$

Definition (Truthful Bayesian implementation)

The social choice function f is truthfully implementable (or incentive compatible) in Bayesian Nash equilibrium if $s_i^*(\theta_i) = \theta_i$ is a Bayesian Nash equilbrium in the direct revelation $\Gamma = (\Theta_1, \ldots, \Theta_I, f)$. That is, for all i and all θ_i $E_{\theta_{-i}}\left[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i\right]$ for all $\hat{\theta}_i \in \Theta_i$.

- roughly: no type of no player can get a higher expected payoff by announcing a wrong type in the direct mechanism

Example

Example (A public project)

- 2 agents value a public project either with 2 or 4 ($\Theta_1 = \Theta_2 = \{2, 4\}$)
- $\phi(2, 2) = \phi(2, 4) = \phi(4, 2) = \phi(4, 4) = 1/4$
- costs of project are c=5
- social choice function f:
    - undertake project if $\theta_1 + \theta_2 \geq c$
    - split costs equally if both have valuation 4
    - low (high) valuation agent pays 2 (3) if project is undertaken and valuations are not the same
- is f truthfully implementable in dominant strategy equilibrium?
- is f truthfully implementable in Bayesian Nash eq.?

Revelation principle for Bayesian Nash equilibrium

Theorem (Revelation principle for Bayesian Nash equilibrium)

Suppose there exists a mechanism $\Gamma = (S_1, \ldots, S_I, g)$ *that implements* f *in Bayesian Nash equilibrium. Then* f *is truthfully implementable in Bayesian Nash equilibrium.*

- very roughly: if some mechanism works, then the direct mechanism works as well

Revelation principle: idea and implication

- idea:
  - same as before:
    - tell me your type and I will play the equilibrium strategy of this type in $\Gamma = (S_1, \ldots, S_I, g)$
    - as deviating was not profitable in the equilibrium of $\Gamma = (S_1, \ldots, S_I, g)$, telling the true type is optimal given that the other players tell their true type
- implication:
  - we can concentrate on direct mechanisms:
    - if the direct mechanism is not incentive compatible, then there is no mechanism that can implement f

Revelation principle: Proof I

- say $\Gamma = (S_1, \ldots, S_I, g)$ implements f in BNE
- then there is a BNE $s^* = (s_1^*, \ldots, s_I^*)$ such that $f(\theta) = g(s^*(\theta))$ for all $\theta \in \Theta$
- this implies that for all i and θ_i

$$E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(g(\hat{s}_i, s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right]$$

for all $\hat{s}_i \in S_i$

- in particular for all i and θ_i

$$E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right]$$

for all $\hat{\theta}_i \in \Theta_i$

Revelation principle: Proof II

- which is equivalent to:
  for all i and θ_i

  $$E_{\theta_{-i}}\left[u_i(f(\theta_i,\theta_{-i}),\theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(f(\hat{\theta}_i,\theta_{-i}),\theta_i)|\theta_i\right]$$

  for all $\hat{\theta}_i \in \Theta_i$. (because $f(\theta) = g(s^*(\theta))$)
- Hence, the direct mechanism is incentive compatible. □

Towards expected externality mechanism

Example (Provision of a public good or: Should CPH host Olympics?)

- decision whether to provide a public good (host or not)
- costs of public good (hosting Olympics) are c
- I agents (citizens)
- agent i values the good θ_i which is his private information and possibly negative
- agent i has utility $t_i + \theta_i + \bar{m}_i$ if the good is provided and $t_i + \bar{m}_i$ if it is not
- t_i is transfer to i (additional tax); most likely $t_i \leq 0$

Towards expected externality mechanism

- Is there a mechanism such that...
  - ... we host Olympics if and only if $\sum_i \theta_i \geq c$
  - ... the budget is balanced: $\sum_i -t_i = c$ if host and $\sum_i t_i = 0$ otherwise
  - ... the mechanism is incentive compatible in Bayesian Nash equilibrium.
- note:
  - Clarke-Groves mechanism satisfies 1 and 3 (even in dominant strategies) but did not satisfy 2!
  - we can focus on incentive compatible, direct revelation mechanisms because of the revelation principle

Expected externality mechanism: preliminaries

- similar trick as in Clarke-Groves:
  - design transfers that make each agent an *expected* welfare maximizer $\rightarrow$ incentive to announce type truthfully
- Assumption: The agents' types are statistically independent.
  - nature draws type of agent i from distribution ϕ_i and there is no correlation between the draws of any agents i and j
- Simple case: c=0
- notation: define the efficient choice function $k^*(\theta)$ as

$$k^*(\theta) = \begin{cases} 1 & \text{if } \sum_{i=1}^{I} \theta_i \geq c = 0 \\ 0 & \text{else} \end{cases}$$

Expected externality mechanism ($c = 0$)

$$t_i(\theta) = E_{\tilde{\theta}_{-i}}\left[\sum_{j \neq i} \tilde{\theta}_j k^*(\theta_i, \tilde{\theta}_{-i})\right] + h_i(\theta_{-i})$$

where h_i is an arbitrary function of the other agents' types that we will specify later (in a clever way that will give us budget balance)

- note:
  - only expected value of other agents features in transfers
  - actual types/announcements of other players are only relevant for h_i
  - the first term ($E_{\tilde{\theta}_{-i}}[\,]$) depends only on own type not on other agents' type

Expected externality mechanism ($c = 0$)

Theorem (Expected externality mechanism)

The social choice function $f(\theta) = (k^*(\theta), t_1, \ldots, t_I)$ is Bayesian incentive compatible.

- *Proof:*
  - In the direct mechanism, we have to show: If all other agents announce their true type, announcing my true type is optimal.
  - Agent i of type θ_i gets the following expected utility if he announces $\hat{\theta}_i$ and the other agents announce truthfully $\hat{\theta}_{-i} = \theta_{-i}$:

Expected externality mechanism ($c = 0$)

$$
\begin{aligned}
U_i(\hat{\theta}_i, \theta_i) &= E_{\theta_{-i}} \left\{ t_i(\hat{\theta}_i, \theta_{-i}) + \theta_i k^*(\hat{\theta}_i, \theta_{-i}) + \bar{m}_i \right\} \\
&= E_{\theta_{-i}} \left\{ \theta_i k^*(\hat{\theta}_i, \theta_{-i}) + E_{\tilde{\theta}_{-i}} \left[\sum_{j \neq i} \tilde{\theta}_j k^*(\hat{\theta}_i, \tilde{\theta}_{-i}) \right] \right. \\
&\quad \left. + h_i(\theta_{-i}) + \bar{m}_i \right\} \\
&= E_{\theta_{-i}} \left[\theta_i k^*(\hat{\theta}_i, \theta_{-i}) + \sum_{j \neq i} \theta_j k^*(\hat{\theta}_i, \theta_{-i}) + h_i(\theta_{-i}) \right] + \bar{m}_i
\end{aligned}
$$

- note that $\hat{\theta}_i$ influences $U_i(\hat{\theta}_i, \theta_i)$ only through k^*
- $U_i(\hat{\theta}_i, \theta_i)$ is maximal if $k^*(\hat{\theta}_i, \theta_{-i}) = 1$ if and only if $E_{\theta_{-i}} \left[\theta_i + \sum_{j \neq i} \theta_j \right] \geq 0$
- agent i can ensure this by announcing his true type! □

Expected externality mechanism ($c = 0$)

- we found a mechanism that is incentive compatible and undertakes the project only if it is efficient
- what about budget balance?
  - we choose h_i such that the mechanism satisfies budget balance denote by $\chi_i(\theta_i)$ the first part of t_i:

$$\chi_i(\theta_i) = E_{\tilde{\theta}_{-i}}\left[\sum_{j \neq i} \tilde{\theta}_j k^*(\theta_i, \tilde{\theta}_{-i})\right]$$

- set $h_i(\theta_{-i})$ as

$$h_i(\theta_{-i}) = -\left(\frac{1}{I-1}\right)\sum_{j \neq i} \chi_j(\theta_j)$$

- agent i receives $\chi_i(\theta_i)$ and this amount is paid for by the other agents (in equal shares)

Expected externality mechanism: comments

- *expected externality*:
- agent i gets transfers χ_i equal to the expected benefit of the other agents
- if he changes his type announcement the change in χ_i reflects the expected externality this change imposes on the other agents

Expected externality mechanism: extensions

- for $c > 0$ a similar trick as in the Clarke-Groves mechanism works leading to transfers

$$t_i(\theta) = E_{\tilde{\theta}_{-i}}\left[\sum_{j \neq i}\left(\tilde{\theta}_j - \frac{c}{I}\right) k^*(\theta_i, \tilde{\theta}_{-i})\right] - k^*(\theta_i, \theta_{-i})\frac{c}{I} + h_i(\theta_{-i})$$

- as in the Clarke-Groves mechanism it can be easily generalized to several possible levels of, say, a public good (k can be more than just 0 and 1) and utility functions

$$u_i(k, t_i, \theta_i) = v(k, \theta_i) + t_i + \bar{m}_i$$

leading to transfers

$$t_i(\theta) = E_{\tilde{\theta}_{-i}}\left[\sum_{j \neq i}\left(v(k^*(\theta_i, \tilde{\theta}_{-i}), \tilde{\theta}_j) - \frac{c(k^*(\theta_i, \tilde{\theta}_{-i}))}{I}\right)\right] - k^*(\theta_i, \theta_{-i})\frac{c(k^*(\theta_i, \theta_{-i})}{I} + h_i(\theta_{-i})$$

Example

Example (Hosting the Olympics)

- group 1 values hosting the Olympics with value $\theta_1 = 5$
- group 2 dislikes the Olympics (and all the noise that comes with it) either much $\theta_2 = -6$ (with prob $1/3$) or a little $\theta_2 = -4$ (with prob $2/3$)
- costs are 0
- what are the "expected externality transfers" in this example?

Problem: Participation

- in the previous example:
  group 1 had (in expectation) to pay more to group 2 than it actually valued the Olympics
- group 1 would prefer *not to participate* in the mechanism (e.g. forget about Olympics)
- this is a general problem in the expected externality mechanism
- government might be able to force agents to participate
- in private settings (trade, auctions etc.) forcing agents to participate is impossible
- next lecture:
  Is there an efficient mechanism if participation is voluntary?

Expected externality mechanism: Take aways and economics

if people

- have quasilinear preferences
- play Bayesian Nash equilibrium
- can be forced to participate

we can achieve efficient, budget balanced allocation

- coercion and government vs. private actors
- what if people can ”walk away”?

Teaser

You are on a fleemarket and you are interested in a certain item (say an old copy of MWG). The seller has a certain valuation c for the book (which probably stems from his expectation when and at what price he can sell it if he does not sell it to you). You value the book at 150kr. It would be efficient if you got the book if and only if $c \leq 150kr$.
Will the bargaining between you and the seller be efficient?
Can there be an efficient mechanism?
(Why can in this setting the expected externality mechanism *not* be used to ensure efficiency?)

Review questions

- What is the main advantage and disadvantage of Bayesian Nash equilibrium implementation compared to dominant strategy implementation?
- State the revelation principle for Bayesian Nash equilibrium implementation and give the intuition behind the result!
- What is the main motivation for the expected externality mechanism?
- In what sense, does the expected externality mechanism extend the Clarke-Groves mechanism?

Exercises

- There are two agents and one indivisible good. An alternative $x \in X$ specifies who gets the good and which transfer payments the agents make. The transfer payments have to add up to 0 (that is the costs of the good are zero). The valuations of the two agents are independently and uniformly distributed on $[0, 1]$. The efficient allocation rule k^* gives the good to the agent with the highest valuation (you can ignore that agents might have the same valuation). Can you compute the expected externality transfers?
- MWG 23.D.1