Lecture 16

Mechanism Design

May 20, 2005
Lecturer: Anna R. Karlin
Notes: Atri Rudra

Today we will talk about mechanism design which is also called Incentive Engineering. In a typical problem in this area we need to satisfy certain goals in the presence of multiple self-interested parties with private information. This is also sometimes called the "private value optimization problem".

16.1 The Basics

We begin with some definitions and notations. Every agent i has *type* $\theta_i \in \Theta_i$. Unless otherwise mentioned n would be the total number of agents. The set of all possible outcomes is denoted by $\mathcal{O}$. Agent i has a utility function for every type $\theta_i \in \Theta_i$ and outcome $o \in \mathcal{O}$ and is denoted by $u_i(\theta_i, o)$. The set of strategies available to agent i is denoted by Σ_i.

A *mechanism* is a function $f : \Sigma_1 \times \Sigma_2 \cdots \times \Sigma_n \to \mathcal{O}$. One can think of the mechanism f has a black box to which every agent i submits a strategy $s_i \in \Sigma_i$ depending on its type θ_i. The mechanism then outputs $f(s_1, s_2, \cdots, s_n)$ as the outcome. We stress that this is a game of incomplete information in that the agents do not know the payoff matrix (for example, in an auction setting, the value $u_i(\cdot)$ is private and agent i does not know $u_j(\cdot)$ for $j \neq i$). We next talk about solution concepts which are commonly used in mechanism design.

16.2 Solution Concepts

16.2.1 Dominant Strategy Equilibrium

This implies every agent has a strategy which is "best" irrespective what strategy the other agents play.

16.2.2 Bayesian Nash Equilibrium

In this setting a common prior knowledge of the distribution on the types of agents is assumed. More formally, it is assumed that everyone knows

$$F(\Theta) = (F_1(\Theta_1), F_2(\Theta_2), \cdots, F_n(\Theta_n))$$

where $F_i(\cdot)$ is the distribution for type for agent i. It is assumed that $F_i(\cdot)$ has all the "nice" properties like it is continuous and differentiable. Each agent has a strategy (function) $s_i \in \Sigma_i$. In this setting each agent i gets a type θ_i which is drawn from the distribution $F_i(\cdot)$. The agent then submits $s_i(\theta_i)$ to the mechanism. In this setting agent i knows θ_i and the distribution $F_{-i} = F_1 \times \cdots F_{i-1} \times F_{i+1} \cdots \times F_n$ and nothing else.

Definition 16.1. A *strategy profile* $(s_1(\theta_1), s_2(\theta_2), \cdots, s_n(\theta_n))$ is in *Bayesian Nash Equilibrium* for mechanism f if for all i, for all $\theta_i \in \Theta_i$ and for all $s'_i \in \Sigma_i$

$$\mathbb{E}_{\theta_{-i}}[u_i(f(s_i(\theta_i), s_{-i}(\theta_{-i})), \theta_i)] \geq \mathbb{E}_{\theta_{-i}}[u_i(f(s'_i(\theta_i), s_{-i}(\theta_{-i})), \theta_i)]$$

As an example consider the *First Price Auction*. Let X_i denote the random variable which denotes the value for the (single) item for bidder (agent) i. θ_i is the outcome of sampling from X_i. Also assume that all X_is are i.i.d from distribution function $F(\cdot)$. Define a new random variable $Y = \max(X_1, X_2 \cdots, X_{n-1})$. Assume that Y is distributed according to the distribution $G(\cdot)$ and let the corresponding density function be denoted by $g(\cdot)$.

Claim 1. The Bayesian Nash Equilibrium for the first price auction is given by

$$s_i(\theta_i) = \mathbb{E}[Y | Y \leq \theta_i]$$

Proof. With a slight abuse of notation (as in a continuous probability distribution, the probability of a single point is not defined) it is easy to see that

$$\mathbb{E}[Y | Y \leq \theta_i] = \int y \Pr[Y = y | Y \leq \theta_i] dy$$

More formally, it is

$$\mathbb{E}[Y | Y \leq \theta_i] = \int_0^{\theta_i} y \cdot \frac{g(y)}{G(\theta_i)} dy$$

It is easy to see that this expectation is continuous and increasing (note that $\int_0^{\theta_i} g(y)dy = G(\theta_i)$). For the rest of the proof we will drop the subscript i because all the strategies are symmetric (as all the X_is are i.i.d.).

We now have to argue that $s(\theta) = \mathbb{E}[Y | Y \leq \theta]$ is the Bayesian Nash strategy. So assume not and let the bidder on getting a type θ bid $s(\theta')$ (note that as $s(\cdot)$ is an increasing and continuous function such a $\theta' \neq \theta$ exists). Now the bidder's (expected) payoff on bidding $s(\theta')$ (instead of $s(\theta)$) is given by

$$G(\theta')[\theta - s(\theta')]$$

The first term is the probability that the bidder would win on bidding $s(\theta')$ and the second term is the payoff if she wins. Using the definition of $s(\cdot)$ and expanding we get that the payoff is

$$\begin{aligned}
& G(\theta')[\theta - \int_0^{\theta'} y \cdot \frac{g(y)}{G(\theta')} dy] \\
= \; & G(\theta')\theta - \int_0^{\theta'} y g(y) dy \\
= \; & G(\theta')\theta - yG(y) \mid_0^{\theta'} + \int_0^{\theta'} G(y) dy \\
= \; & G(\theta')(\theta - \theta') + \int_0^{\theta'} G(y) dy
\end{aligned}$$

If $\theta' > \theta$ then the payoff can be written as

$$\int_0^{\theta} G(y)dy + [\int_{\theta}^{\theta'} G(y)dy - G(\theta')(\theta' - \theta)]$$

Note that the second term is upper bounded by 0. If on the other hand $\theta' < \theta$ the payoff is given by

$$\int_0^{\theta} G(y)dy + [-\int_{\theta'}^{\theta} G(y)dy + G(\theta')(\theta - \theta')$$

Again the second term is upper bounded by 0. Thus, the payoff is maximized at $\theta' = \theta$ which proves the claim. $\square$

16.2.3 Nash equilibrium

We covered this in the first lecture.

16.3 Social Choice Function

Assume that there is some function which is "desirable" given by

$$f : \Theta_1 \times \Theta_2 \cdots \times \Theta_n \to \mathcal{O}$$

The *Mechanism Design Problem* is to design the game such that the outcome in the equilibrium is the same as f. In more detail, note that the mechanism now defines the strategies Σ_i. The outcome of the mechanism is denoted by

$$g : \Sigma_1 \times \Sigma_2 \cdots \times \Sigma_n \to \mathcal{O}$$

Also every agent i receives a payment which is denoted by the following function

$$p_i : \Sigma_1 \times \Sigma_2 \cdots \Sigma_n \to \mathbb{R}^{\geq 0}$$

We say that the mechanism implements the social choice function if the outcome computed in the equilibrium is f.

We now outline two 'standard' assumptions in this area

1. The agents have quasi-linear preferences, that is, $u_i(o, \theta_i) = v_i(o, \theta_i) + p_i$.
2. Agents are *risk neutral* that is, every agent's goal is to maximize the (expected) utility (as opposed to say just making sure that the utility is above some fixed value).

16.4 Revelation Principle

A *Direct Revelation Mechanism* is one where the stragey set Σ_i is to reveal an element of Θ_i. A *truthful mechanism* is where in the equilibrium, the agents reveal their true type.

Theorem 16.1 (Revelation Principle). *Any mechanism can be transformed into an equivalent truthful direct revelation mechanism that implements the same social choice function and the same payment.*

Proof. Let the given mechanism implement the social choice function f and let the equilibrium strategy for agent i be $s_i(\cdot)$. The equivalent truthful direct revelation mechanism f' is defined as follows, $f'(\theta_1, \theta_2, \cdots, \theta_n) = f(s_1(\theta_1), s_2(\theta_2), \cdots, f_n(\theta_n))$. It is easy to see that f' satisfies all the required properties. □

Let us consider a couple of examples to illustrate the direct revelation principle.

Example 16.1. Consider the ascending auction (also known as the *English auction*) for a single item. Here the auctioneer keep on increasing the price till there is just one bidder (or agent) who is left. The last bidder is declared the winner and pays the current price. It is not too hard to see that the equivalent direct revelation mechanism is the *second price auction*.

Example 16.2. Consider the *Dutch auction* where the auctioneer starts with a high price and keeps on decreasing the price till one bidder accepts the price and is declared the winner. He pays the current price. The equivalent direct revelation mechanism is the *first price auction*.

The Revelation principle allows us to just consider truthful mechanism– this can be useful in some context like proving impossibility results. There are some issues though–

1. Bidders can be averse to revealing information (consider for example the ISP routing game where ISPs do not want to reveal their true information).

2. Sometimes it is not easy for the agents to determine their valuation exactly.

3. Agents' types can be complicated. The classical example is the *Combinatorial Auctions* where there are N items and the types are subsets of items. Thus, $|\Theta_i|$ is exponential which is prohibitively large. In real life people do care about these kinds of auction– for example FCC uses one to auction the spectrum.

16.5 Truthful Mechanisms with Dominant Strategies

In this section, we consider the following questions–

Question. What social function f can be truthfully implemented (in the dominant strategy sense)?

We first look at the question as to what f can be implemented without payment. The answer as the following theorem shows is basically nothing.

Theorem 16.2 (Gibbard-Sattherwaite). *If* $|\mathcal{O}| \geq 3$ *and* f *is truthfully implementable then* f *is a dictatorship– that is the outcome just depends on what one single bidder decides to do.*

The next obvious question is what can be one with payment and this brings us to the classical Vickery-Clark-Groves (or VCG) mechanisms.

16.5.1 Vickery Clark Groves Mechanism

The VCG mechanism implements the following choice function

$$f(\vec{\theta}) = \text{argmax}_{o \in \mathcal{O}} \sum_i v_i(o, \theta_i)$$

where the utility of agent i is defined by

$$u_i(o, \theta_i) = v_i(o, \theta_i) + p_i$$

After the "optimal solution" o^* is picked, the payments are fixed as follows

$$p_i(\vec{\theta}) = \sum_{j \neq i} v_j(o^*, \theta_j) - h_i(\theta_{-i})$$

Note that the function h_i does not depend on what the type of agent i is. The "standard" function for $h_i(\cdot)$ is the following– $h_i(\theta_{-i}) = \sum_{j \neq i} v_j((o_{-i})^*, \theta_j)$ where $(o_{-i})^*$ is the optimal solution when agent i is not present. It is not too hard to verify the following claim.

Claim 2. The VCG mechanism is truthful.