CS698W: Game Theory and Collective Choice **Jul-Nov 2017**

Lecture 16: September 6, 2017

Lecturer: Swaprava Nath *Scribe(s): Sandipan Mandal*

16.1 Recap

In the last class we continued our discussion on *mechanism design*. We first discussed *social choice functions (SCF)*, utility functions for SCFs and their types, defined *mechanism design (MD)* and its types. Then we went on to discuss *weakly dominant strategy* w.r.t. mechanism and conditions under which a SCF is implemented in dominant startegy. It was followed by discussion on *strategyproofness* or *dominant strategy incentive compatibility (DSIC)* and *revelation principle*.

Before going forward lets revisit the definition of DSIC as it will be used later to illustrate a subtle point w.r.t. Bayesian incentive compatiblity. The definition is as follows:

Definition 16.1 *A direct mechanism* $\langle \Theta, f \rangle$ *is* strategyproof *or* dominant strategy incentive compatible (DSIC) *if*

$$u_i(f(\theta_i, \tilde{\theta}_{-i}), \theta_i) \geq u_i(f(\theta_i', \tilde{\theta}_{-i}), \theta_i),\ \forall \theta_i, \theta_i' \in \Theta_i, \forall \tilde{\theta}_{-i} \in \Theta_{-i}, \forall i \in N. \tag{16.1}$$

This definition states that irrespective of whether others are reporting their types to the central authority truthfully or not, for any agent$i$, reporting her type truthfully is a weakly dominant strategy.

16.2 Relating Mechanism Design to Bayesian Games

Suppose the types are generated from a common prior P and type of any player i, θ_i is revealed only to the respective players. We can consider the mechanism design scenario in a Bayesian game setup, where the Bayesian game is as follows.

$$\langle N, (A_i)_{i \in N}, (\Theta_i)_{i \in N}, P, (\Gamma_\theta)_{\theta \in \Theta} \rangle.$$

- N = Set of Players $\{1, 2, , \ldots, n\}$
- M_i = Message space of Player i (Corresponding to Action set in Bayesian Games)
- Θ_i = Set of type for player i
- P = Common prior over $\bigtimes_{i \in N} \Theta_i$
- $\Gamma_\theta = \langle N, (A_i)_{i \in N}, (u_i(\theta))_{i \in N} \rangle$

Consider the function $m_i : \Theta_i \to M_i$. We can observe that given any type, the function m_i maps to a member of the message space. It is equivalent to pure strategies of players in a normal Bayesian game. Intuitively every player sends some message to central authority to maximize its utility which are equivalent to strategy a player might have taken in the case of Bayesian games.

16.3 SCF implemementedin Bayesian Equilibrium

Definition 16.2 *A mechanism $\langle M, g\rangle$ implements SCF f in Bayesian equilibrium if the following two conditions hold.*

1. $\exists(m_1, m_2, \ldots, m_n)$ *s.t.* $m_i(\theta_i)$ *maximizes the* ex-interim *utility of agent* i, $\forall \theta_i \in \Theta_i, \forall i \in N$, *i.e,*

$$\mathbb{E}_{\theta_{-i}|\theta_i}[u_i(g(m_i(\theta_i), m_{-i}(\theta_{-i})), \theta_i)] \geq \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(g(m_i', m_{-i}(\theta_{-i})), \theta_i)], \ \forall \theta_i \in \Theta_i, \forall i \in N \qquad (16.2)$$

2. $g(m_i(\theta_i), m_{-i}(\theta_{-i})) = f(\theta_i, \theta_{-i})$

Observation: If SCF f is implementable in dominant strategies then f will be implementable in Bayesian equilibrium.

16.4 Bayesian Incentive Compatibility (BIC)

Definition 16.3 *A direct mechanism $\langle \Theta, f\rangle$ is Bayesian incentive compatible if*

$$\mathbb{E}_{\theta_{-i}|\theta_i}[u_i(f(\theta_i, \theta_{-i}), \theta_i)] \geq \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(f(\theta_i', \theta_{-i}), \theta_i)], \ \forall \theta_i, \theta_i' \in \Theta_i, \forall i \in N. \qquad (16.3)$$

Observation. From Equations 16.1 and 16.3, we can observe that the condition for a direct mechanism to be DSIC is required to hold for all $\theta_{-i} \in \Theta_{-i}$ but for it to be BIC this condition is not required because we are doing weighted average over all possible θ_{-i} in some sense.

16.5 Revelation Principle for BI SCFs

Theorem 16.4 *If a Social Choice Function f is implementable in Bayesian Equilibrium then f is Bayesian Incentive Compatible.*

The proof of the above theorem is similar to the proof of Revelation Principle for DSI SCFs and hence left as an exercise.

16.6 Arrovian Social Welfare Function (SWF)

> Even before considering strategyproofness, how can we aggregate the individual preferences into a social preference?

To further understand the implication of this question consider the following example. Suppose three friends want to watch a movie every day for a week. Seven movies are chosen − they are the alternatives. But everyone wants to watch their favorite movie first and does not want to watch an already watched movie. Hence, everyone has a different preference order over the movies. However the three friends want to watch the movies *together*. So what is the best order or watching the movies together? In other words, how can we have a social ordering of these alternatives? The question leads us to study of *Arrovian Social Welfare Functions* (ASWF).

Definition 16.5 *Arrovian Social Welfare Function takes as input individual preferences of different agents outputs a social preference for all the agents.*

Notation

- Set of alternatives $A = \{a_1, \ldots, a_m\}$.
- Agents $N = \{1, 2, \ldots, n\}$.
- aR_ib: Alternative a is atleast as good as b for agent i.
- Set of all possible ordering $\mathcal{R}$.

Properties of R_i (ordering)

- *Completeness*: For every $a, b \in A$ either aR_ib or bR_ia
- *Reflexivity*: $\forall a \in A$, aR_ia
- *Transitivity*: If aR_ib, bR_ic then aR_ic

Lets divide the relation R_i into two parts P_i (asymmetric part) and I_i (symmetric part) i.e.,

- aP_ib: Alternative a is strictly better than b for i.
- aI_ib: Alternative a is indifferent to b for i.

Definition 16.6 *An ordering R_i is* linear *if aR_ib and bR_ia, then $a = b$, i.e., indifference is not allowed.*

16.6.1 Arrovian Social Welfare Function Representation

Using the notations we have discussed have Arrovian Social Welfare Function F is represented as $F : \mathcal{R}^n \to \mathcal{R}$. Since $F(R)$ is an ordering over the alternatives like R_i we can split it into two parts $\hat{F}(R)$(asymmetric part) and $\bar{F}(R)$(symmetric part) where $R = (R_1, \ldots, R_n)$.

16.6.2 Desirable Properties for Arrovian SWF

Definition 16.7 *A Social Welfare Function F satisfies* weak Pareto *(WP) if*

$$\forall a, b \in A, \quad [aP_ib, \ \forall i \in N] \implies [a\hat{F}(R)b].$$

Definition 16.8 *A Social Welfare Function F satisfies* strong Pareto *if*

$$\forall a, b \in A, \quad [aR_ib, \ \forall i \in N, \ \exists j, \ aP_jb] \implies [a\hat{F}(R)b].$$

Clearly, strong Pareto $\implies$ weak Pareto.