Epistemic Logic and Game Theory Ⅰ

Yudai Kubono

JAIST

October 22, 2022

Prisoner's Dilemma

Example 1

Two prisoners, A and B, supposed to have committed a common crime, are being interrogated (取り調べ) in separate rooms. In order to get them to confess, the prosecutor (検事) offers them the following plea bargain (司法取引).

> *If they both confess, they will both be imprisoned for five years.*
>
> *If only one of them confesses, the one who confesses will be released, but the one who keeps silent will be imprisoned for ten years.*
>
> *If they both remain silent, their sentences (判決) will be reduced to two years in prison for lack of evidence.*

A and B have no means of communicating with each other.

	B remains silent	B confesses
A remains silent	(-2,-2)	(-10,0)
A confesses	(0,-10)	(-5,-5)

Figure: A payoff matrix of prisoner's dilemma

The use of epistemic logic in game theory

- Epistemic foundation

  The aim is the understanding of solution concepts (e.g., nash equilibrium) and the creation of new solution concepts by an explicit description of the players' knowledge (Bonanno (2008)).

  Today's topic: Common belief of rationality and dominant strategy equilibrium.
- Description of games with epistemic logic

  The aim is an accurate description of the game under special conditions using epistemic logic (Meier and Schipper (2014)).
- Formalize reasoning in games and analyze the knowledge and conditions needed for decision-making (Kaneko and Suzuki (2002)).

Formalization of Strategic-Form Games

Definition 1

A finite strategic-form game with ordinal payoffs G *is a tuple* $\langle \mathrm{Ag}, \{S_i\}_{i \in \mathrm{Ag}}, \{\pi_i\}_{i \in \mathrm{Ag}} \rangle$, *where:*

- Ag *is a finite set of agents;*
- S_i *is a finite set of* i*'s strategies;*
- $\pi_i : S \to \mathbb{R}$ *is a payoff function, where* $S = \prod_{i \in \mathrm{Ag}} S_i$.

- We label the elements of S_i as $\{s_i^1, \cdots, s_1^n\}$, where $n \in \mathbb{N}$

Syntax

Definition 2

Let $\mathcal{P}$ be a set of strategy symbols $s_i^1, \cdots, s_i^n$ for each $i \in \mathrm{Ag}$ and the forms of $s_i^l \succeq_i s_i^k$ for each $l, k \in \mathbb{N}$ and $i \in \mathrm{Ag}$. The language $\mathcal{L}_{\mathcal{P}}$ is the set of formulas generated by the following grammar:

$$\mathcal{L}_{\mathcal{P}} \ni \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_i\varphi \mid C\varphi,$$

where $p \in \mathcal{P}$. Other logical connectives $\vee$, $\rightarrow$, and $\leftrightarrow$ are defined in the usual manner.

- $s_i^l \succeq_i s_i^k$ means that i's strategy l is better that i's strategy k, that is $\pi_i(s_l, s_{-i}) \geq \pi_i(s_k, s_{-i})$.
- $s_i^l \succ_i s_i^k$ is defined in the usual way.

An Syntactic Model of G

Definition 3

Given a finite strategic-form game with ordinal payoffs G, a syntactic model of G is a tuple $\langle W, \{R_i\}_{i \in \mathrm{Ag}}, \{\sigma_i\}_{i \in \mathrm{Ag}}, V \rangle$, where:

W is a non-empty set of possible worlds (or states);

R_i is a binary relation on W that is serial, transitive, and euclidean;

$\sigma_i : W \to S_i$ is a function satisfying the following property: if $w' \in R_i(w)$ then $\sigma_i(w') = \sigma_i(w)$.

- We denote a set of accessible worlds w' from w as $R_i(w)$, $(\sigma_i(w))_{i \in \mathrm{Ag}}$ as $\sigma(w)$, and a sequence of the players' strategies other than i as σ_{-i}.
- A strategy $\sigma_i(w)$ represents a choice that agent i makes at w.

Satisfaction relations

Definition 4

Let . For each syntactic model M and $w \in W$, the satisfaction relation $\vDash$ is defined as follows:

$$
\begin{aligned}
M, w \vDash s_i^k &\text{ iff } \sigma_i(w) = s_i^k;\\
M, w \vDash (s_i^k \succeq_i s_i^l) &\text{ iff } \pi_i(s_i^k, \sigma_{-i}(w)) \geq \pi_i(s_i^l, \sigma_{-i}(w));\\
M, w \vDash \neg\varphi &\text{ iff } M, w \nvDash \varphi;\\
M, w \vDash \varphi \wedge \psi &\text{ iff } M, w \vDash \varphi \text{ and } M, w \vDash \psi;\\
M, w \vDash B_i\varphi &\text{ iff } M, v \vDash \varphi \text{ for all } v \text{ such that } (w, v) \in R_i;\\
M, w \vDash C\varphi &\text{ iff } M, v \vDash \varphi \text{ for all } v \text{ such that } (w, v) \in (\bigcup_{i \in \mathrm{Ag}} R_i)^+.
\end{aligned}
$$

For any binary relation R, R^+ is the smallest set satisfying the following conditions: $R \subseteq R^+$; for all x, y, z, if $(x, y) \in R^+$ and $(y, z) \in R^+$, then $(x, z) \in R^+$.

Common Belief of Rationality

Definition 5

It is said that i's choice of strategy at w, that is $\sigma_i(w)$, is rational if for every $s_i^n \in S_i$ and all $w' \in R_i(w)$, $\pi_i(\sigma_i(w), \sigma_{-i}(w')) \geq \pi_i(s_i^n, \sigma_{-i}(w'))$.

- The concept of common belief of rationality can be described by the formula, that is $C(\bigwedge_{i \in \mathrm{Ag}} \bigwedge_{k,l \in \mathbb{N}} (s_i^k \to \neg B_i(s_i^l \succeq s_i^k)))$.
- What solutions, which is expressed strategy profiles, correspond to common belief of rationality?

A *Strictly Dominated Strategy* and a *Subgame*

Definition 6

A strategy s_i^k *is a strictly dominated strategy if for every* $s_{-i} \in S_{-i}$*, there exsits* $s_i^l \in S_i$ $\pi_i(s_i^k, s_{-i}) < \pi_i(s_i^l, s_{-i})$*, where* $S_{-i} = \prod_{j \in \mathrm{Ag} \setminus \{i\}} S_j$.

Definition 7

Given $G = \langle \mathrm{Ag}, \{S_i\}_{i \in \mathrm{Ag}}, \{\pi_i\}_{i \in \mathrm{Ag}} \rangle$, $G' = \langle \mathrm{Ag}, \{S'_i\}_{i \in \mathrm{Ag}}, \{\pi'_i\}_{i \in \mathrm{Ag}} \rangle$ *is a subgame of* G*, where* $S'_i \subseteq S_i$ *for every* i*, and* $\pi'_i(s') = \pi_i(s')$ *for every* $s' \in S'$.

Iterated Deletion of Strictly Dominated Strategies (IDSDS)

- IDSDS is one of procedures to solve a game:
  1. Let S_i^0 be S_i and $D_i^0 \subseteq S_i^0$ be a set of i's strategies that is strictly dominated in $G^0 = G$,
  2. For $m \geq 1$, let $S_i^m = S_i^{m-1} \setminus D_i^{m-1}$ and G^m be a subgame of G with S_i^m.
- Let $\prod_{i \in \mathrm{Ag}} S_i^\infty$ denote S^∞, where $S_i^\infty = \bigcap_{m \in \mathbb{N}} S_i^m$.
- If a game has a dominant strategy equilibrium s, $S^\infty = \{s\}$.

Analysis of Common Belief of Rationality and a Strategy Profile

- We obtain the following propositions:

Proposition 1

Let a schema $s_i^k \to \neg B_i(s_i^l \succeq s_i^k)$ *denote* $\mathbf{WR}$.

1. *For all* $w \in W$, *if* $M \vDash \mathbf{WR}$ *then* $\sigma(w) \in S^\infty$;
2. *There exists a syntactic model* M *of* G *that validates* $\mathbf{WR}$ *and is such that (1) for every* $s \in S^\infty$, *there exists a world* w *such that* $w \vDash s$, *and (2) for every* $s \in S$ *and* $w \in W$, *if* $w \vDash s$ *then* $\sigma(w) \in S^\infty$.

- The choices that players make in any model where common belief of rationality holds are identical with the strategy chosen by IDSDS, which is dominant strategy equilibrium.
- A set S^∞ is said to characterize the notion of common belief of rationality.

Conclution

- Formalizing the concept of common belief of rationality with epistemic logic, it showed the concept corresponds to dominant strategy equilibrium.
- It is possible to analyze how to derive the solution concept logically.
- Results of epistemic logic can be used without gaps (e.g., agent communication).

[1] Bonanno, G. (2015) Epistemic Foundations of Game Theory. *Handbook of epistemic logic*, edited van Ditmarsch, H., van Der Hoek, W., Halpern, J. Y., and Kooi, B. College Publications.

[2] Bonanno, G. (2008) A syntactic approach to rationality in games with ordinal payoffs. In *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, pages=59-86.

[3] Meier, M. and Schipper, B. C. (2014) Bayesian games with unawareness and unawareness perfection. *Economic Theory*,56(2):219–249.

[4] Kaneko, M. and Suzuki, N. (2002) Bounded interpersonal inferences and decision making. *Economic Theory*,19(1):63-103.