Game Theory

Rationalizability

Tetsuya Hoshino February 11, 2022

A player is rational if she maximizes her (expected) payoff given her belief about opponents' play.[1] Assume that all players' rationality is **common knowledge**; that is, all players are rational, they know they are rational, they know they know are rational, and so on. *What are all the strategies that they could potentially play based only on this assumption?*

1 Correlated Rationalizability

Example 1. Consider the following normal-form game:

	L	R
U	$1, 1$	$-2, 0$
D	$-2, 0$	$0, 0$

Table 1: all pure strategies are rationalizable

For player 1, U is rationalizable in the sense that if player 1 believes that player 2 will play L then playing U is rational. Similarly, D is also rationalizable for player 1. Moreover, both L and R are rationalizable for player 2. □

1.1 Never-Best Responses

Beliefs about Opponents' Strategies In a normal-form game G, let $\mu_{-i} \in \Delta(A_{-i})$ be player i's **belief** about players $-i$' strategies. Given this belief μ_{-i}, if player i plays a strategy $\sigma_i \in \Sigma_i$ then his expected payoff is:

$$u_i(\sigma_i, \mu_{-i}) = \sum_{a \in A} u_i(a)\sigma_i(a_i)\mu_{-i}(a_{-i}).$$

Remark 1. Note that $\Delta(\prod_{j \neq i} A_j) \neq \prod_{j \neq i} \Delta(A_j)$. The left-hand side is the set of *correlated* mixed strategies of players $-i$, while the right-hand side is the set of *independent* mixed strategies of players $-i$. □

Remark 2. The beliefs herein are often called "conjectures." We will use terminology "beliefs" but the reader should not confuse the "beliefs" here with "beliefs" used in incomplete-information games. □

[1] Rationality is often defined differently in other contexts.

Never-Best Responses The concept of rationalizability is independently defined by Bernheim (1984) and Pearce (1984).

Definition 1. In a normal-form game G, player i's strategy $\sigma_i \in \Sigma_i$ is a **never-best response** if for her every belief $\mu_{-i} \in \Delta(A_{-i})$, there exists some strategy $\sigma_i' \in \Sigma_i$ such that:

$$u_i(\sigma_i', \mu_{-i}) > u_i(\sigma_i, \mu_{-i}).$$

1.2 Rationalizable Strategies

Rationalizable Pure Strategies Analogous to the iterated deletion of strictly dominated strategies, we iteratively delete all players' never-best responses. At each step of the deletion, we ask "what are actions that could potentially be played by rational players?" Then, each player will conclude that no (rational) player will ever play pure strategies that are a never-best response. Since each player will expect that no (rational) player will play such pure strategies with positive probabilities. Furthermore, it will be common knowledge that players arrive at this conclusion, so that it justifies the deletion of these pure strategies from the game. The iteration proceeds until no further pure strategies can be deleted.

Definition 2. In a normal-form game G, for each $i \in I$ and each $k \in \mathbb{N}$, let $\mathrm{CR}_i^0 = A_i$ and

$$\mathrm{CR}_i^k = \mathrm{CR}_i^{k-1} \setminus \underbrace{\left\{a_i \in \mathrm{CR}_i^{k-1} : \forall\, \mu_{-i} \in \Delta(\mathrm{CR}_{-i}^{k-1}) \quad \exists\, \sigma_i' \in \Delta(\mathrm{CR}_i^{k-1}) \quad u_i(\sigma_i', \mu_{-i}) > u_i(a_i, \mu_{-i})\right\}}_{\text{pure strategies that are never-best responses}}.$$

Let player i's set of **correlated rationalizable pure strategies** be such that

$$\mathrm{CR}_i^\infty = \bigcap_{k=0}^{\infty} \mathrm{CR}_i^k.$$

We often call them rationalizable strategies by omitting "correlated."

From this definition, it follows that:

$$a_i \in \mathrm{CR}_i^\infty \quad \Longrightarrow \quad \exists\, \mu_{-i} \in \Delta(\mathrm{CR}_{-i}^\infty) \quad \forall\, \sigma_i' \in \Delta(\mathrm{CR}_i^\infty) \quad u_i(a_i, \mu_{-i}) \geq u_i(\sigma_i', \mu_{-i}).$$

Rationalizable Mixed Strategies We consider the set of mixed rationalizable strategies. One may jump to the (wrong) conclusion that it is merely the set $\Delta(\mathrm{CR}_i^\infty)$ of all distributions over CR_i^∞, but this is not true in general. That is, mixed strategies in CR_i^∞ are not always rationalizable.

Example 2. Consider the following normal-form game:

	L	R
U	$1, 0$	$-2, 0$
M	$-2, 0$	$1, 0$
D	$0, 0$	$0, 0$

Table 2: a mixed strategy in $\Delta(\text{CR}_i^\infty)$ is not necessarily correlated rationalizable

For each player, all pure strategies are rationalizable: $\text{CR}_i^\infty = A_i$. For example, D is rationalizable, because it is optimal for player 1's belief that player 2 plays L and R with equal probabilities. However, player 1's mixed strategy $\frac{1}{2}U \oplus \frac{1}{2}M \in \Delta(\text{CR}_i^\infty)$ yields payoff $-\frac{1}{2}$ regardless of player 2's strategy; that is, $\frac{1}{2}U \oplus \frac{1}{2}M$ is strictly dominated by D and is not rationalizable. $\square$

2 Correlated Rationalizability versus Iterated Strict Dominance

Correlated rationalizability and iterated strict dominance ask complementary questions to each other. While the concept of iterated strict dominance deletes all strategies that a player *will never* play under common knowledge of rationality, the concept of rationalizability identifies all strategies that a player *could potentially* play under common knowledge of rationality. Then, the question that naturally arises is: How are these two concepts related to each other?

Never-Best Response $\Leftrightarrow$ Strict Dominance Because rationalizability is based on the notion of never-best responses and iterated strict dominance on the notion of strict dominance, it is natural to examine the relationship between the two notions.

Theorem 1. *In a finite normal-form game G, player i's strategy $\sigma_i \in \Sigma_i$ is a never-best response if and only if it is strictly dominated.*

Proof. We show the "only if" part, as the "if" part is immediate. We show the contrapositive: If σ_i is not strictly dominated then it is not a never-best response. Suppose that σ_i is not strictly dominated. That is, there exists no $\sigma_i' \in \Sigma_i$ such that for each $a_{-i} \in A_{-i}$, $u_i(\sigma_i', a_{-i}) > u_i(\sigma_i, a_{-i})$. There are, in total, $n = \prod_{j \neq i} |A_j|$ possible action profiles for players $-i$, which we enumerate by $a_{-i}^1, a_{-i}^2, \ldots, a_{-i}^n$. To use the Separating Hyperplane Theorem, we define a set Y by

$$Y = \left\{ \begin{pmatrix} u_i(\sigma_i', a_{-i}^1) - u_i(\sigma_i, a_{-i}^1) \\ \vdots \\ u_i(\sigma_i', a_{-i}^n) - u_i(\sigma_i, a_{-i}^n) \end{pmatrix} : \sigma_i' \in \Delta(A_i) \right\} \subset \mathbb{R}^n.$$

Since Y is a non-empty, convex set and $Y \cap \mathbb{R}_{++}^n = \varnothing$, it follows from the Separating Hyperplane Theorem that there exist some $c \in \mathbb{R}$ and some $v \in \mathbb{R}^n \setminus \{0\}$ such that for each $x \in \mathbb{R}_{++}^n$ and each $y \in Y$,

$$v \cdot x > c \geq v \cdot y.$$

It is immediate that $v \in \mathbb{R}^n_+ \setminus \{0\}$.[2] Letting $\bar{v} = \sum_l v_l > 0$, we define a vector $\mu_{-i} = v/\bar{v}$. Since all elements of μ_{-i} are non-negative and the sum of them is one, it follows that $\mu_{-i} \in \Delta(A_{-i})$. Then, $\mu_{-i} \cdot x \geq c/\bar{v} \geq \mu_{-i} \cdot y$. Since $\mu_{-i} \cdot x > 0$, it follows that $c \leq 0$. Hence, $0 \geq \mu_{-i} \cdot y$ for each $y \in Y$. That is, for each σ'_i,

$$\sum_{l=1}^{n} u_i\Big(\sigma'_i, a^l_{-i}\Big)\mu_{-i}\Big(a^l_{-i}\Big) \leq \sum_{l=1}^{n} u_i\Big(\sigma_i, a^l_{-i}\Big)\mu_{-i}\Big(a^l_{-i}\Big),$$

or equivalently $u_i(\sigma'_i, \mu_{-i}) \leq u_i(\sigma_i, \mu_{-i})$. That is, σ_i is not a never-best response. ■

Remark 3. Pearce's (1984) proof is based on Nash's Existence Theorem and the Minimax Theorem. The Minimax Theorem itself can be proven by (a corollary of) the Separating Hyperplane Theorem. □

Remark 4. From Theorem 1, it follows that a strategy that is weakly dominated but not strictly dominated is a best response to some belief. Indeed, a weakly dominated strategy may be played in a Nash equilibrium. □

Correlated Rationalizability ⇔ Iterated Strict Dominance Recall that ND_i^∞ denotes the set of player i's pure strategies that survive iterated deletion of strictly dominated strategies.

Corollary 1. *In a finite normal-form game G, $\mathrm{ND}_i^\infty = \mathrm{CR}_i^\infty$ for each $i \in I$,*

Proof. By definition, $\mathrm{ND}_i^0 = \mathrm{CR}_i^0 = A_i$. By Theorem 1, $\mathrm{ND}_i^k = \mathrm{CR}_i^k$ for each $k \in \mathbb{N}$. Hence, $\mathrm{ND}_i^\infty = \mathrm{CR}_i^\infty$. ■

3 Correlated Rationalizability versus Correlated Equilibrium

Next, we compare correlated rationalizability with correlated equilibrium.

Theorem 3. *In a finite normal-form game G, let $\mu \in \Delta(A)$ be a direct correlated equilibrium. Then, for each $i \in I$ and each $a_i \in A_i$, if $\mu(\{a_i\} \times A_{-i}) > 0$ then $a_i \in \mathrm{CR}_i^\infty$.*

Proof. It suffices to show that for each $k \in \mathbb{N} \cup \{0\}$, if a correlated equilibrium μ assigns non-zero probability on player j playing action a_j then $a_j \in \mathrm{CR}_j^k$. We prove this claim by

[2] To see $v \in \mathbb{R}^n_+$, suppose that $v = (v_1, v_2, \ldots, v_n) \notin \mathbb{R}^n_+$. Then, there is some $l \in \{1, 2, \ldots, n\}$ such that $v_l < 0$. Then, $v \cdot x$ is arbitrarily small and thus less than c if we take $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n_{++}$ such that x_l is large enough and x_k is close to zero, but this contradicts inequality $v \cdot x > c$. It must be that $v \in \mathbb{R}^n_+$. Since $v \neq 0$ by the Separating Hyperplane Theorem, it follows that $v \in \mathbb{R}^n_+ \setminus \{0\}$.

Recap (Separating Hyperplane Theorem):

Theorem 2. *Let $X, Y \subset \mathbb{R}^n$ be two disjoint non-empty convex subsets for $n \in \mathbb{N}$. Then, there exist some $c \in \mathbb{R}$ and some $v \in \mathbb{R}^n \setminus \{0\}$ such that for each $x \in X$ and each $y \in Y$,*

$$v \cdot x > c \geq v \cdot y.$$

That is, the hyperplane $\{z \in \mathbb{R}^n : v \cdot z = c\}$, with the normal vector v, separates X and Y.

This theorem is illustrated in Figure 1. It is intuitive, in the two-dimensional case, that there exists a line that separates the two disjoint non-empty convex sets $X, Y \subset \mathbb{R}^2$.

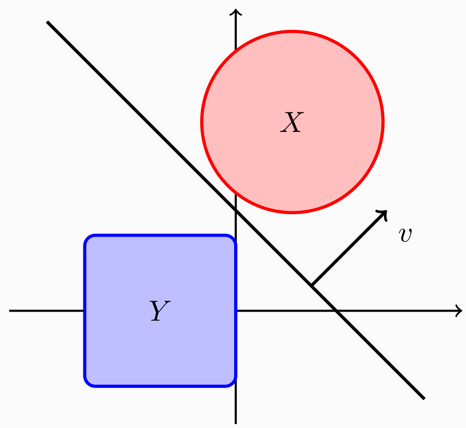

Figure 1: the Separating Hyperplane Theorem

Remark 5. When should we try to use the Separating Hyperplane Theorem? When you want to show the existence of some non-negative vector—such as probabilities or prices—it is worth trying to use the Separating Hyperplane Theorem. In the proof of the Second Welfare Theorem, for example, we use the Separating Hyperplane Theorem to prove the existence of a positive price vector to support a Pareto-efficient allocation as a quasi-Walrasian equilibrium. In the proof of Theorem 1, we use the Separating Hyperplane Theorem to show the existence of a belief, which is a non-negative vector with length n. □

induction. First, the claim is obvious for $k = 0$. Second, we suppose that the claim is true for $k = K$ and then show that it is true for $k = K + 1$. For each $a'_{-i} \in A_{-i}$,

$$\mu_{-i}(a'_{-i}) = \frac{\mu(a_i, a'_{-i})}{\sum_{a''_{-i} \in A_{-i}} \mu(a_i, a''_{-i})}.$$

By the induction hypothesis, $\mu_{-i} \in \Delta(\mathrm{CR}_i^K)$. By the definition of correlated equilibrium, action a_i is a best response to μ_{-i}. That is, $a_i \in \mathrm{CR}_i^{K+1}$. Hence, $a_i \in \bigcap_{k=0}^{\infty} \mathrm{CR}_i^k = \mathrm{CR}_i^{\infty}$. ■

4 Independent Rationalizability*

In the concept of correlated rationalizability, player i's belief allows for correlation between players $-i$' strategies. Indeed, her belief μ_{-i} is defined on $\Delta(\mathrm{CR}_{-i}^{k-1}) = \Delta(\prod_{j \neq i} \mathrm{CR}_j^{k-1})$. In a relevant concept of independent rationalizability, her belief no longer allows for correlation between the opponents' strategies.

Definition 3. In a normal-form game G, let $\mathrm{IR}_i^0 = A_i$ and for each $k \in \mathbb{N}$,

$$\mathrm{IR}_i^k = \mathrm{IR}_i^{k-1} \setminus \Big\{ a_i \in \mathrm{IR}_i^{k-1} : \forall\, \mu_{-i} \in \prod_{j \neq i} \Big(\Delta\Big(\mathrm{IR}_j^{k-1}\Big)\Big) \quad \exists\, \sigma_i' \in \Delta\Big(\mathrm{IR}_i^{k-1}\Big) \quad u_i(\sigma_i', \mu_{-i}) > u_i(a_i, \mu_{-i}) \Big\}.$$

We define player i's set of **independent rationalizable pure strategies** by

$$\mathrm{IR}_i^\infty = \bigcap_{k=0}^{\infty} \mathrm{IR}_i^k.$$

Remark 6. It is immediate that $\mathrm{IR}_i^\infty \subset \mathrm{CR}_i^\infty$ in general. If there are two players, it is obvious that $\mathrm{IR}_i^\infty = \mathrm{CR}_i^\infty$. In contrast, if there are more than two players, it is often the case that $\mathrm{IR}_i^\infty \neq \mathrm{CR}_i^\infty$. For such a game, see Osborne & Rubinstein (1994, Figure 58.1 together with the discussion). $\square$

References

Bernheim, B. D. (1984). Rationalizable strategic behavior. *Econometrica*, *52*(4), 1007–1028.

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory.* MIT Press.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica*, *52*(4), 1029–1050.