G5212: Game Theory

Mark Dean

Spring 2017

Rationalizability

- So far I have cheated you (a bit) in my description of dominance
- I motivated this section by describing rationality as
  - A player has some belief over the actions of others
  - Plays a strategy which is a best response to that belief
- But then I started talking about players never playing a strategy which is strictly dominated
- Are these the same thing?
- i.e. are the following sets equivalent?
  - The set of strategies that are strictly dominated
  - The set of strategies that cannot be rationalized as a best response?

Never-Best Response

	L_p	R_{1-p}
T	$2, 0$	$-1, 1$
M	$0, 10$	$0, 0$
B	$-1, -6$	$2, 0$

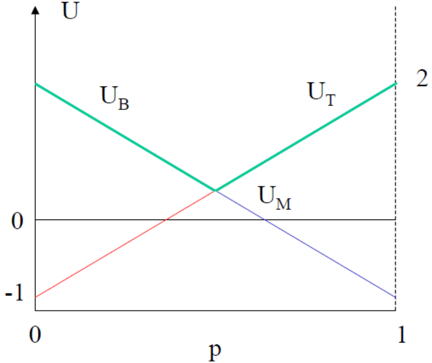
$\longrightarrow\longrightarrow\longrightarrow$

Player 1's payoff from T,M,B

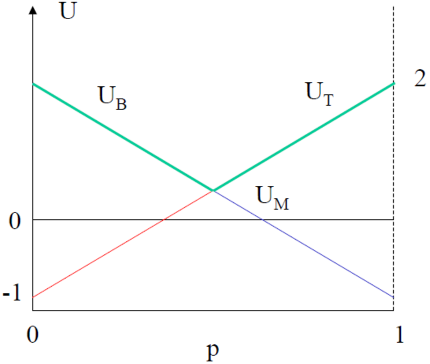
$U_T = 2p - (1-p) = 3p - 1$

$U_M = 0$

$U_B = -p + 2(1-p) = 2 - 3p$



Never-Best Response



- $U_M(p) < \max\{U_T(p), U_B(p)\}$ for any $p \in [0,1]$, i.e., M is never a best response to any belief about his opponent's play.

Never-Best Response

Definition

A strategy $s_i \in S_i$ is **a best response against a belief** $\mu_i \in \Delta(S_{-i})$ if $u_i(s_i, \mu_i) \geq u_i(s_i', \mu_i)$ for any $s_i' \in S_i$.

Note (1) belief μ_i could be correlated: $\Delta(S_{-i}) \neq \prod_{j \neq i} \Delta(S_j)$;
(2) $u_i(s_i, \mu_i) = \sum_{s_{-i} \in S_{-i}} u_i(s_i, s_{-i}) \mu_i(s_{-i})$.

Definition

A strategy $s_i \in S_i$ is a **never-best response** if it is not a best response against any belief $\mu_i \in \Delta(S_{-i})$.

- Is the set of strategies which are never best responses equal to the set of strategies which are strictly dominated?
  - It is obvious that strategies which are strictly dominated are never best responses
  - But is it true that strategies which are never best responses must be strictly dominated

Never-Best Response and Strict Dominance

- Yes!

Theorem

(Pearce 1984) A strategy s_i^* is a never-best response if and only if it is strictly dominated.

- Note correlation here is important!
  - If we do not assume that beliefs can be correlated, there are strategies which are never best responses but are **not** strictly dominated
  - As you will see for homework
- Proof will come later in the course

Rationalizability

- So that deals with rationality
- What about common knowledge of rationality?
- This requires a bit more
  - Player 1 is rational
  - Player 2 thinks 1 is rational
  - Player 1 thinks 2 thinks 1 is rational...
- So the beliefs that rationalize the choice of player 1 must themselves be rational (in some sense)
  - As must the actions underlying those beliefs
  - And so on....

Rationalizability

- We say that a strategy s_i^* is rationalizable if the following is true
  - There is a set of beliefs μ_i^1 such that s_i^* is a best response
  - For every player j and s_j in the support of μ_i^1, there exists a belief μ_j^2 such that s_j is a best response
  - For every player k and s_k in the support of μ_j^2, there exists a belief μ_k^3 such that s_k is a best response
  - And so on
- Again we allow for correlated beliefs

Rationalizability

Definition

A strategy s_i^* is rationalizable if there exists

- A sequence of sets $\{X_j^t\}_{t=1}^{\infty}$ for each player j with $\varnothing \neq X_j^t \subset S_j$
- A belief μ_i^1 whose support is a subset of X_{-i}^1
- For each $j \in N$, t and action $s_j \in X_j^t$, a belief $\mu_j^t(s_j)$ whose support is in X_{-j}^{t+1}

such that

- s_i^* is a best response to μ_i^1
- s_j is the best response to $\mu_j^t(s_j)$ for every t, j and $s_j \in X_j^t$
- $s_j \in X_j^t$ if and only if, for some player k there exists a strategy $s_k \in X_k^{t-1}$ such that s_j is in the support of $\mu_k^t(s_k)$

Rationalizability

- It turns out that we can describe this in a somewhat more compact way

Rationalizability

Theorem

A strategy s_i^* is rationalizable if and only if there exists a set $Z_j \subset S_j$ for each player j such that

- $s_i^* \in Z_i$
- *For each j and $s_j \in Z_j$ there is a belief $\mu_j(s_j)$ to which s_j is a best response and whose support is in Z_{-j}*

Proof.

[Proof (Sketch)]

1. (If) Define $\mu_j^t(s_j) = \mu_j(s_j)$, X_j^1 as the support of $\mu_i(s_i^*)$ and $X_j^t = s_j$ such that for some player k there exists a strategy $s_k \in X_k^{t-1}$ such that s_j is in the support of $\mu_k^t(s_k)$
2. (Only if) Define $Z_i = \cup_{t=1}^{\infty} X_i^t \cup \{s_i^*\}$ and $Z_j = \cup_{t=1}^{\infty} X_j^t$

□

Matching Pennies (Again!)

Example

Matching Pennies

			Bob	
			H	T
Anne	H		$+1, -1$	$-1, +1$
	T		$-1, +1$	$+1, -1$

- Example: can we rationalize $\{H, H\}$ as an outcome?
  - A thinks B will play Heads and so plays Heads
  - A thinks B thinks A will play Tails, justifying B playing Heads
  - A thinks B thinks A thinks B will play Tails, justifying A playing Tails
  - A thinks B thinks A thinks B thinks A will play Heads, justifying B playing Tails

Matching Pennies (Again!)

Example

Matching Pennies

		Bob	
		H	T
Anne	H	$+1, -1$	$-1, +1$
	T	$-1, +1$	$+1, -1$

- Example: can we rationalize $\{H, H\}$ as an outcome?
- Formally
  - $Z_A = Z_B = H, T$
  - $\mu_A(H) = 1.H$, $\mu_A(T) = 1.T$
  - $\mu_B(H) = 1.T$, $\mu_B(T) = 1.H$

- Note that we are not, at this stage, demanding that beliefs be consistent

Rationalizability and IDSDS

- We have seen that there is a link between Never Best Response and Strict Dominance
- Is there an equivalent link for rationalizable strategies?
- Yes!

Theorem

The set of rationalizable outcomes = the set of outcomes that survive iterated deletion of strictly dominant strategies.

Proof.

See Osborne and Rubinstein Proposition 61.2 □