Introduction to Game Theory: Infinite Dynamic Games

John C.S. Lui

Department of Computer Science & Engineering
The Chinese University of Hong Kong
www.cse.cuhk.edu.hk/$\sim$cslui

Outline

Why we need a new model of repeated games?

- Consider the prisoners' dilemma, in reality, many crooks do not squeal, how do we explain this?
- Consider the Cournot duopoly, we showed that cartels were unstable, but in real-life, many countries need to make (or enforce) anti-collusion laws. How do we explain this?
- In real-life, decisions may not be made once only, but we make decisions based on what we perceive about the future.
- In the prisoners' dilemma, crooks will not squeal because they are afraid of future retaliation. For cartels, they sustain the collusion by making promises (or threats) about the future.
- Inspired by these observations, we consider situations in which players interact repeatedly.

Stage game

- If a player only needs to make a *single* decision, he is playing an stage game.
- After the stage game is played, the players again find themselves facing the same situation, i.e., the stage game is repeated.
- Taken one stage at a time, the only sensible strategy is to use the Nash equilibrium strategy for each stage game.
- However, if the game is viewed as a whole, the strategy set becomes much richer:
  - players may condition their behavior on the past actions of their opponents, or
  - make threats about what they will do in the future, or
  - collusion.

Exercise

- Consider the following prisoners' dilemma game with cooperation (C) and defection (D):

	C	D
C	3,3	0,5
D	5,0	1,1

- Let say the game is repeated just once so there are two stages. We solve this like any dynamic game by backward induction.
- In the final stage, there is no future interaction, so the payoff to be gained is at this final stage. We choose the best response of playing D. So (D, D) is the NE of this subgame.
- Consider the first stage (the subgame is the whole game). Since payoff is fixed for the final stage, the payoff for the entire game is:

	C	D
C	4,4	1,6
D	6,1	2,2

Exercise: continue

- Note that the pure-strategy set for each player in the entire game is $\boldsymbol{S} = \{CC, CD, DC, DD\}$.
- But because we are only interested in a subgame perfect NE, we only consider two strategies: $\{CD, DD\}$ (since the last stage is fixed).
- Analyzing the above game (previous payoff table), the NE of the entire game is (DD, DD). So the subgame perfect NE for the whole game is to play D in both stages.
- Note that the player cannot induce cooperation:
  - in the first stage by promising to cooperate in the 2nd stage (since they won't);
  - in the first stage by threatening to defect in the 2nd stage since this is what happens anyway.

HW: Exercise 7.2.

Infinite Iterated Prisoners' Dilemma

If the length of the game is infinite, we need the following strategy:

Definition

A stationary strategy is one in which the *rule of choosing an action* is the same in *every stage*. Note that this **does not** imply that the action chosen in each stage will be the same.

Example

Examples of stationary strategy are:

- Play C in every stage.
- Play D in every stage.
- Play C if the other player has never played D and play D otherwise.

Comment

- The payoff for a stationary strategy is the "*infinite sum*" of the payoffs achieved at each stage. Let $r_i(t)$ be the payoff for player i in stage t. The total payoff is $\sum_{t=0}^{\infty} r_i(t)$.
- Unfortunately there is a problem. If both players choose $s_C =$"Play C in every stage", then: $\pi_i(s_C, s_C) = \sum_{t=0}^{\infty} 3 = \infty$.
- If one chooses $s_D =$"Play D in every stage" and other chooses s_C, then: $\pi_1(s_D, s_C) = \pi_2(s_C, s_D) = \sum_{t=0}^{\infty} 5 = \infty$.
- Introduce a discount factor δ $(0 < \delta < 1)$ so the total payoff is: $\sum_{t=0}^{\infty} \delta^t r_i(t)$.
- One can use δ to represent (a) inflation; (b) uncertainty of whether the game will continue, or (c) combination of these.
- Applying, $\pi_i(s_C, s_C) = \sum_{t=0}^{\infty} 3\delta^t = \frac{3}{1-\delta}$. $\pi_1(s_D, s_C) = \pi_2(s_C, s_D) = \sum_{t=0}^{\infty} 5\delta^t = \frac{5}{1-\delta}$.

With discounting δ, can permanent cooperation (e.g., a cartel) be a stable outcome of the infinitely repeated Prisoners' Dilemma?

Definition

A strategy is called a trigger strategy when a change of behavior is triggered by a single defection.

Example of trigger strategy

- Consider a trigger strategy $s_G =$"Start by cooperating and continue to cooperate until the other player defects, then defect forever after".
- If both players adopt s_G, $\pi_i(s_G, s_G) = \sum_{t=0}^{\infty} 3\delta^t = \frac{3}{1-\delta}$.
- But is (s_G, s_G) a Nash equilibrium?

Is (s_G, s_G) a Nash Equilibrium?

- Let's do an informal analysis (formal analysis follows).
- Assume both players are restricted to a pure-strategy set $\boldsymbol{S} = \{s_G, s_C, s_D\}$.
- Suppose player 1 decides to use s_C instead, payoff is: $\pi_1(s_C, s_G) = \pi_2(s_C, s_G) = \frac{3}{1-\delta}$. Same result applies if player 2 adopts s_C, so this will not be better off than (s_G, s_G).
- Assume player 1 adopts s_D, the sequence is:

	$t =$	0	1	2	3	4	5	$\cdots$
player 1	s_D	D	D	D	D	D	D	$\cdots$
player 2	s_G	C	D	D	D	D	D	$\cdots$

For player 1: $\pi_1(s_D, s_G) = 5 + \delta + \delta^2 + \cdots = 5 + \frac{\delta}{1-\delta}$.

- Player 1 cannot do better by switching from s_G to s_D if $\frac{3}{1-\delta} \geq 5 + \frac{\delta}{1-\delta}$. The inequality is satisfied if $\delta \geq 1/2$. So (s_G, s_G) is a NE if $\delta \geq 1/2$.

Exercise

- Consider the iterated Prisoners' Dilemma with pure strategy sets $\boldsymbol{S}_1 = \boldsymbol{S}_2 = \{s_D, s_C, s_T, s_A\}$.
  - The strategy s_T is the famous "*tit-for-tat*": begin with cooperating, then do whatever the other player did in the previous stage.
  - The strategy s_A is the cautious version of the tit-for-tat: begin with defection, then does whatever the other player did in the previous stage.
- What condition does the discount faction δ have to satisfy in order for (s_T, s_T) to be a Nash equilibrium?

Solution

- The payoff of $\pi_1(s_T, s_T) = \frac{3}{1-\delta}$.
- The payoff of $\pi_1(s_C, s_T) = \frac{3}{1-\delta}$, so it is not better off than (s_T, s_T).
- The payoffs of $\pi_1(s_D, s_T) = 5 + \frac{\delta}{1-\delta}$, with $\delta \geq \frac{1}{2}$, (s_T, s_T) is better.
- The payoffs of $\pi_1(s_A, s_T)$ is:

$$\pi_1(s_A, s_T) = 5 + 0 + 5\delta^2 + 0 + 5\delta^4 + \cdots = \frac{5}{1-\delta^2}.$$

When $\delta \geq \frac{3}{4}$, (s_T, s_T) is better.

Homework

- Consider the iterated Prisoners' Dilemma with pure-strategy sets $\boldsymbol{S}_1 = \boldsymbol{S}_2 = \{s_D, s_C, s_G\}$.
- What is the strategic form (or normal form) of the game?
- Find all the Nash equilibria.

Is s_G subgame perfect?

Question: The NE where both players adopt the trigger strategy s_G. Is it a subgame perfect Nash equilibrium strategy?

Analysis

- Since it is an infinite iterated game, at any point in the game, the future of the game (i.e., subgame) is equivalent to the entire game.
- The possible subgames can be classified into four classes:
  - neither player has played D;
  - both players have played D;
  - player 1 used D in the last stage but player 2 did not;
  - player 2 used D in the last stage but player 1 did not;
- Let us analyze them one by one.

Analysis: continue

- **Case (1):** neither player's opponent has played D so the strategy s_G specifies that cooperation should continue until the other player defects (i.e., s_G again). The strategy specified (s_G, s_G) is a NE of the subgame because it is a NE for the entire game.
- **Case (2):** both players have defected so the NE strategy (s_G, s_G) specifies that each player should play D forever. The strategy adopted in this class of subgame (s_D, s_D) is a NE of the subgame since it is a NE of the entire game.

Analysis: continue

- **Case (3):** player 1 used D in the last stage but not player 2.
  - For this case, since player 2 used C, s_G dictates player 1 to play C and player 2 to play D. In summary player 1 will play $C, D, D, \ldots$ while player 2 will play $D, D, D, \ldots$.
  - So (s_G, s_D) is adopted for this subgame.
  - But (s_G, s_D) is not a Nash equilibrium for the subgame because player 1 could get a great playoff by using s_D.
- **Case (4):** similar argument as in Case (3).
- Hence, the NE strategy for the entire game, (s_G, s_G), does not specify that players play a Nash equilibrium in every possible subgame, then (s_G, s_G) is **not** subgame perfect.

Another policy

- Although (s_G, s_G) is not a subgame perfect Nash equilibrium, we can consider the following *similar* strategy which is subgame perfect NE strategy.
- Let $s_g =$"start by cooperating and continue to cooperate until *either* player defects, then defect forever after". The reasons are:
  - player 1 or 2 plays (s_g, s_g) in case 1 and 2 (for case 2, it is actually (s_D, s_D)).
  - player 1 or 2 plays (s_D, s_D) for case 3 and 4.

HW: Exercise 7.6.

Further Analysis

- We showed (s_G, s_G) is a Nash equilibrium of the entire game under the *assumption* the the set of strategies is **finite.**
- Is it possible to allow more strategies?
- Is (s_G, s_G) still a NE if more strategies are allowed?
- If we restrict ourselves to subgame perfect Nash equilibrium, then we need to learn the one-stage deviation principle first.

Definition

A pair of strategies (σ_1, σ_2) satisfies the one-stage deviation condition if neither player can increase their payoff by deviating unilaterally from their strategy in any single stage and returning to the specified strategy thereafter.

Example

- Consider the subgame perfect NE strategy (s_g, s_g): "start by cooperating and continue to cooperate until *either* player defects, then defect forever after". Does this satisfy the one-stage deviation condition?
- At any give stage, the game is in one of the two classes of subgame: (a) either both players have always cooperated, or (b) at least one player has defected.

Analysis

- **case a:** if both players have been cooperated, then s_g specifies cooperation at this stage.
- If either one changes to action D in this stage, then s_g specifies using D forever. The expected future payoff for the player making this change is $5 + \frac{\delta}{1-\delta}$, which is less than the payoff for continued cooperation, $\frac{3}{1-\delta}$, if $\delta > \frac{1}{2}$. So the player will not switch.
- **case b:** if either player has defected in the past, then s_g specifies defection for both players at this stage.
- If either player changes to C in this stage, then s_g still specifies using D forever after. The expected future payoff for the player making this change is $0 + \frac{\delta}{1-\delta}$, which is less than the payoff for following the behavior specified by s_g (by playing D) $\frac{1}{1-\delta}$, provided that $\delta < 1$.
- Thus, the pair (s_g, s_g) satisfies the one-stage deviation condition provided $1/2 < \delta < 1$.

Theorem

A pair of strategies is a subgame perfect Nash equilibrium for a discounted repeated game if and only if it satisfies the one-stage deviation condition.

For proof, please refer to the book.

Exercise

- Consider the following iterated Prisoners' Dilemma:

	Player 2 (C)	Player 2 (D)
Player 1 (C)	4,4	0,5
Player 1 (D)	5,0	1,1

- Let s_P be the strategy: "defect if only one player defected in the previous stage (regardless of which player it was); cooperate if either both players cooperated, or both players defected in the previous stage".
- Use the one-stage principle to find a condition for (s_P, s_P) to be a subgame perfect Nash equilibrium.

Analysis

- Note that s_P depends on the behavior of both players in the *previous* stage. We consider the possible behavior at stage $t-1$ and examine what happens if player 1 deviates from s_P at stage t (since the game is symmetric, we do not need to consider player 2).
- There are three possible cases for behavior at stage $t-1$, they are:
  - Player 1 has used D and player 2 used C in stage $t-1$.
  - Player 1 has used D and player 2 used D in stage $t-1$.
  - Player 1 has used C and player 2 used C in stage $t-1$.

Analysis: **Case 1**

- Strategy s_P dictates player 1 to play $D, C, C, \ldots$ and player 2 to play $D, C, C, \ldots$.
- The total future payoff for player 1 is

$$\pi_1(s_P, s_P) = 1 + \frac{4\delta}{1-\delta}.$$

- Suppose player 1 uses C in stage t and reverts to s_P onwards, let this strategy be s'. The total future payoff for player 1 is

$$\pi_1(s', s_P) = 0 + \delta + \frac{4\delta^2}{1-\delta}.$$

- Player 1 does not benefit from the switch if $\pi_1(s_P, s_P) \geq \pi_1(s', s_P)$, and this is true for all values of δ ($0 \leq \delta \leq 1$).

Analysis: **Case 2 or 3**

- Strategy s_P dictates player 1 to play $C, C, C, \ldots$ and player 2 to play $C, C, C, \ldots$.
- The total future payoff for player 1 is

$$\pi_1(s_P, s_P) = \frac{4}{1-\delta}.$$

- Suppose player 1 uses D in stage t and reverts to s_P onwards, let this strategy be s''. The total future payoff for player 1 is

$$\pi_1(s'', s_P) = 5 + \delta + \frac{4\delta^2}{1-\delta}.$$

- Player 1 does not benefit from the switch if $\pi_1(s_P, s_P) \geq \pi_1(s', s_P)$, which is true if $4 + 4\delta \geq 5 + 3\delta^2$. Or (s_P, s_P) is a subgame perfect NE if $\delta \geq \frac{1}{3}$.

Introduction

- From previous section of the iterated Prisoners' Dilemma, the NE of the static game is (D, D) with the payoff of $(1, 1)$, and this is socially sub-optimal as compare to (C, C) with payoff of $(3, 3)$.
- Common belief: if the NE in a static game is socially sub-optimal, players can always do better if the game is repeated.
- A higher payoff can be achieved (in each stage) by both players as an equilibrium of the *repeated game* if the factor is high enough. Example, by playing s_G or s_g.

Definition

Feasible payoff pairs are pairs of payoffs that can be generated by strategies available to the players.

Definition

Suppose we have a repeated game with discount factor δ. If we interpret it as the probability that the game continues, then the expected number of stages in which the game is played is $T = \frac{1}{1-\delta}$. Suppose two players adopt strategy σ_1 and σ_2 (not necessary NE), the expected payoff to player i is $\pi_i(\sigma_1, \sigma_2)$ and the average payoffs (per stage) is:

$$\frac{1}{T}\pi_i(\sigma_1, \sigma_2) = (1-\delta)\pi_i(\sigma_1, \sigma_2).$$

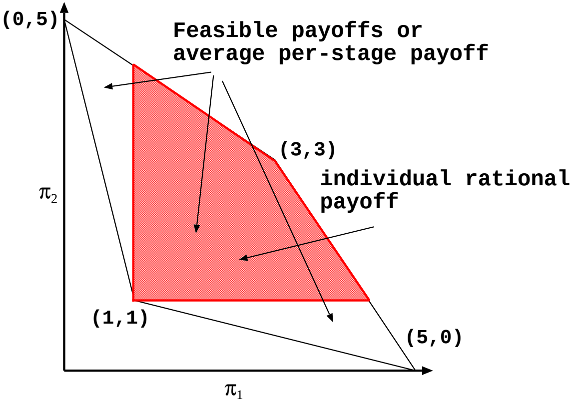
Definition

Individual rational payoff pairs are those average payoffs that exceed the stage Nash equilibrium payoff for both players.

Example

- In the static Prisoners' Dilemma, pairs of payoffs (π_1, π_2) equal to $(1,1)$, $(0,5)$, $(5,0)$ and $(3,3)$ are feasible since they can be generated by some pure strategies.
- Although each player could get a payoff of 0, the payoff pair $(0,0)$ is **not feasible** since there is no strategy pair which generates that payoff pair.
- If player 1 (player 2) uses strategy C with probability p (q), the payoffs are: $(\pi_1, \pi_2) = (1 - p + 4q - pq, 1 - q + 4p - pq)$. Feasible payoff pairs are found by letting $p, q \in [0, 1]$.
- Individual rational payoff pairs are those for which the payoff to each payer is not less than the Nash equilibrium of 1.

Illustration



Theorem

Folk Theorem: let (π_1^*, π_2^*) be a pair of Nash equilibrium payoffs for a stage game and let (v_1, v_2) be a feasible payoff pair when the stage game is repeated. For every individually rational pair (v_1, v_2) (i.e., a pair such that $v_1 > \pi_1^*$ and $v_2 > \pi_2^*$), there exists a $\underline{\delta}$ such that for all $\delta > \underline{\delta}$ there is a subgame perfect Nash equilibrium with payoffs (v_1, v_2).

The Folk's Theorem is the basis as to why collusion or cartel is possible in an infinite stage game.

Proof

- Let (σ_1^*, σ_2^*) be the NE that yields the payoff pair (π_1^*, π_2^*).
- Suppose that the payoff pair (v_1, v_2) is produced by players using action a_1 and a_2 in every stage where $v_1 > \pi_1^*$ and $v_2 > \pi_2^*$ and (v_1, v_2) are pure strategies for player 1 and 2.
- Now consider the following trigger strategy:
  - "Begin by agreeing to use action a_i; continue to use a_i as long as both players use the agreed actions; if any player uses an action other than a_i, then use σ_i^* for in all later stages."
- By construction, any NE involving these strategies will be subgame perfect. So we only need to find the conditions for a NE.

Proof: continue (with (v_1, v_2) are pure strategies)

- Consider another action $a_1^{'}$ such that the payoff of the stage game for player 1 is $\pi_1(a_1^{'}, a_2) > v_1$.
- Then the total payoff for switching to $a_1^{'}$ against a player using the trigger strategy is not greater than

$$\pi_1(a_1^{'}, a_2) + \delta \frac{\pi_1^*}{1-\delta}.$$

- Remember that for the trigger strategy, the payoff of using the trigger strategy is:

$$v_1 + v_1\delta + v_1\delta^2 + \ldots = \frac{v_1}{1-\delta}.$$

- Therefore, it's not beneficial for player 1 to switch to $a_1^{'}$ if $\delta \geq \delta_1$:

$$\delta_1 = \frac{\pi_1(a_1^{'}, a_2) - v_1}{\pi_1(a_1^{'}, a_2) - \pi_1^*}.$$

Proof: continue (with (v_1, v_2) are pure strategies)

- By assumption $\pi_1(a_1^{'}, a_2) > v_1 > \pi_1^*$, we conclude that $0 < \delta_1 < 1$.
- We can use similar argument for player 2 to derive the minimum discount factor δ_2.
- Taking $\underline{\delta} = \max\{\delta_1, \delta_2\}$ completes the proof.

Proof: continue (with randomized strategy)

- Assume the payoff v_i can be achieved by using randomizing strategies σ_i.
- We also assume that there exists a randomizing device whose output is observed by both players, and there is an agreed rule for turning the output of the randomizing device to a choice of action (elaborate).
- Therefore, this implies that these strategies are observable (and not just the outcome).
- If strategies are observable, we can use the previous argument with action a_i and $a_i^{'}$ being replaced by strategies σ_i and $\sigma_i^{'}$.

Stochastic Games

- Under the stochastic game, there is a set of states $\boldsymbol{X}$ with a stage game defined in each state.
- In each state x, player i chooses actions from a set $\boldsymbol{A}_i(x)$.
- One of these stage games is played at each of the discrete time $t = 0, 1, 2, \ldots$.
- Informally,
  - Given the system in state $x \in \boldsymbol{X}$. players choose actions $a_1 \in \boldsymbol{A}_1(x)$ and $a_2 \in \boldsymbol{A}_2(x)$.
  - Player i receives a reward of $r_i(x, a_i, a_{-i})$.
  - The probability that they find state x' in next discrete time is $p(x'|x, a_1, a_2, \ldots, a_n)$.

Definition

A strategy is called a Markov strategy if the behavior of a player at time t depends only on the state x. A pure Markov strategy specifies an action $a(x)$ for each state $x \in \boldsymbol{X}$.

Assumptions

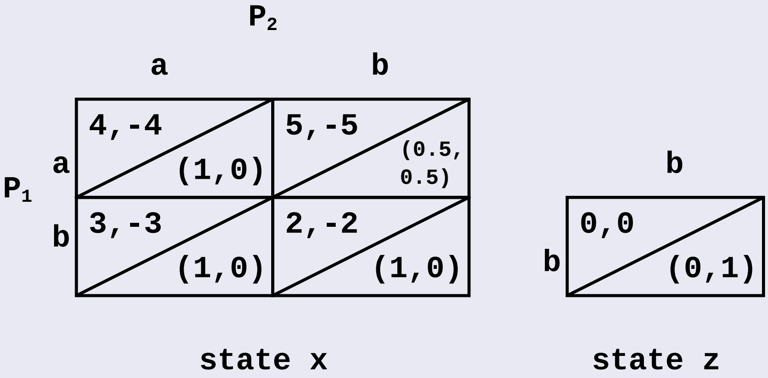
To simplify discussion, we make the following assumptions:

- The length of the game is not know to the players (i.e., infinite horizon).
- The rewards and transitions are time-independents.
- The strategies of interests are Markov.

Example

- The set of states $\boldsymbol{X} = \{x, z\}$.
- In state x, both players can choose action from the set $\boldsymbol{A}_1(x) = \boldsymbol{A}_2(x) = \{a, b\}$.
- The immediate rewards for player 1 are: $r_1(x, a, a) = 4$, $r_1(x, a, b) = 5$, $r_1(x, b, a) = 3$, and $r_1(x, b, b) = 2$.
- It is a zero-sum game, so $r_2(x, a_1, a_2) = -r_1(x, a_1, a_2)$.
- If players choose the action pair $[a, b]$ in state x, then they move to state z with probability $1/2$ and remain in state x with probability $1/2$.
- In state z, they have a single choice set $\boldsymbol{A}(z) = \{b\}$.
- Reward for both players are $r_1(z, b, b) = r_2(z, b, b) = 0$ and state z is an *absorbing state*.

Representation



Payoff

- Consider a game in state x at time t.
- If we know the NE strategies for both players from $t + 1$ onwards, we could calculate the expected future payoffs given that they start from state x. Let $\pi_i^*(x)$ be the expected future payoff for player i starting in state x. (note: the * indicates that these payoffs are derived using the NE strategies for both players).
- At time t, both players would then be playing a single-decision game with payoffs:

$$\pi_i(a_1, a_2) = \left(r_1(x, a_1, a_2) + \delta \sum_{x' \in \boldsymbol{X}} p(x'|x, a_1, a_2)\pi_i^*(x') \right).$$

Payoff: continue

- The payoffs for a Markov-strategy Nash equilibrium are given by the joint solutions of the following pair of equations (one for each state $x \in \boldsymbol{X}$):

$$\pi_1^*(x) = \max_{a_1 \in \boldsymbol{A}_1(x)} \left(r_1(x, a_1, a_2^*) + \delta \sum_{x' \in \boldsymbol{X}} p(x'|x, a_1, a_2^*)\pi_1^*(x') \right),$$
$$\pi_2^*(x) = \max_{a_2 \in \boldsymbol{A}_2(x)} \left(r_2(x, a_1^*, a_2) + \delta \sum_{x' \in \boldsymbol{X}} p(x'|x, a_1^*, a_2)\pi_2^*(x') \right).$$

- In general, solving these equations can be computationally expensive !!

Example

- Given the previous example of stochastic game with $\delta = 2/3$.
- Let v be the expected total future payoff for player 1 of being in state x. Since it is a zero-sum game, the expected total future payoff for player 2 is $-v$. Therefore, in state x, the players are facing the following game:

	a	b
a	$4 + \frac{2}{3}v, -4 - \frac{2}{3}v$	$5 + \frac{1}{3}v, -5 - \frac{1}{3}v$
b	$3 + \frac{2}{3}v, -3 - \frac{2}{3}v$	$2 + \frac{2}{3}v, -2 - \frac{2}{3}v$

- Clearly (b, a) is not a NE for any value of v. The NE for the game in state x will be
  - Play (a, a) if $v < 3$.
  - Play (b, b) if $v > 9$.
  - Play (a, b) if $3 < v < 9$.
- So which is the Markov NE strategy in state x?

Example: continue

- Suppose the players choose (a, a), then $v = 4 + \frac{2}{3}v, \implies v = 12$, which is inconsistent with $v < 3$.
- Suppose the players choose (b, b), then $v = 2 + \frac{2}{3}v, \implies v = 6$, which is inconsistent with $v > 9$.
- Suppose the players choose (a, b), then $v = 5 + \frac{1}{3}v, \implies v = \frac{15}{2}$, which is consistent with $3 < v < 9$.
- So the unique Markov NE has the player using the pair of action (a, a) in the state x.

HW: Exercise 7.8.