Noncooperative Game Theory

J. Leite (adapted from Kevin Leyton-Brown)

Section 1

Infinitely Repeated Games

Infinitely Repeated Games

- What is a player's utility for playing infinitely repeated games?
  - Can we write the game in extensive form?
    - It's an infinite tree!
    - Payoffs cannot be attached to terminal nodes.
  - Can we define the payoffs as the sum of the payoffs in the stage games?
    - The sum could be unbounded. They would be infinite in general!

Two canonical alternatives:

Definition (Average Reward)

Given an infinite sequence of payoffs $r_1, r_2, \ldots$ for player i, the average reward of i is

$$\lim_{k \to \infty} \sum_{j=1}^{k} \frac{r_j}{k}.$$

Discounted reward

Definition (Future Discounted Reward)

Given an infinite sequence of payoffs $r_1, r_2, \ldots$ for player i and discount factor β with $0 < \beta < 1$, i's future discounted reward is

$$\sum_{j=1}^{\infty} \beta^j r_j.$$

- Interpreting the discount factor:
  1. the agent cares more about his well-being in the near term than in the long term
  2. the agent cares about the future just as much as the present, but with probability $1 - \beta$ the game will end in any given round.
- The analysis of the game is the same under both perspectives.

Section 2

Equilibria in Infinitely Repeated Games

Strategy Space

- What is a pure strategy in an infinitely-repeated game?
  - a choice of action at every decision point
  - here, that means an action at every stage game
  - ...which is an infinite number of actions!
- Some famous strategies (repeated PD):
  - Tit-for-tat: Start out cooperating. If the opponent defected, defect in the next round. Then go back to cooperation.
  - Grim Trigger: Start out cooperating. If the opponent ever defects, defect forever.

Indefinite Repetition – Grim Trigger

- Defect in every period is a sub-game perfect equilibrium strategy.
  - No possible gain if the other player is defecting regardless of history.
- After any history where someone has defected, then this is an equilibrium.
- Let's check that it is a best response to cooperate if nobody had defected yet...

Indefinite Repetition – Grim Trigger

- Let's check that it is a best response to cooperate if nobody had defected yet...
- Recall that $u(C, C) = (3, 3)$, $u(D, D) = (1, 1)$, $u(C, D) = (0, 5)$ and $u(D, C) = (5, 0)$.
- Starting from some period where have cooperated in past:
  - If Cooperate: $3 + 3\beta + 3\beta^2 + 3\beta^3 \ldots = 3/(1 - \beta)$
  - If Defect: $5 + \beta + \beta^2 + \beta^3 \ldots = 5 + \beta/(1 - \beta)$
- Gain from defection: 2 today
- Loss from defection: $2\beta + 2\beta^2 + 2\beta^3 \ldots = 2\beta/(1 - \beta)$
- Whether or not to cooperate depends on β. Cooperate if $2 \leq 2\beta/(1 - \beta)$. If $\beta \geq 1/2$, then cooperate is a best response.
- Sustain cooperation by threatening to resort to permanent defection if cooperation fails. The threat is <u>credible</u>: it is an equilibrium.
- If there is a high enough weight on future interactions, then can sustain behaviors that were not possible in the long run.
- Repetition changes the interaction – can respond to others actions, and that provides different incentives.

Indefinite Repetition – Grim Trigger

- Probability p that the game continues next period, probability $1 - p$ that it ends (same analysis as with future discount β).
- $u(C, C) = (3, 3)$, $u(D, D) = (1, 1)$, $u(C, D) = (0, \mathbf{7})$ and $u(D, C) = (\mathbf{7}, 0)$.
- Starting from some period where have cooperated in past: $3 + 3p + 3p^2 + 3p^3 \ldots$
- If defect: $7 + p + p^2 + p^3 \ldots$
- Gain from defection: 4 today
- Loss from defection: $2p + 2p^2 + 2p^3 \ldots = 2p/(1 - p)$
- If $p \geq 2/3$, then cooperate is a best response.
- Harder to sustain cooperation as one-time gains from deviation increase.
- Easier to sustain cooperation as probability to continue increases.

Nash Equilibria

- With an infinite number of pure stratagies, what can we say about Nash equilibria?
  - we won't be able to construct an induced normal form and then appeal to Nash's theorem to say that an equilibrium exists
  - Nash's theorem only applies to finite games
- Furthermore, with an infinite number of strategies, there could be an infinite number of pure-strategy equilibria!
- It turns out we can characterize a set of payoffs that are achievable under equilibrium, without having to enumerate the equilibria.

Section 3

Folk Theorem

Folk Theorem

- Called folk theorem(s) since they were part of the folklore without being formally proven.
- They come in many versions.
- We will look at infinite repeated games with average reward.
- Versions also exist for discounted rewards as well as finitely repeated games, repeated games of incomplete information, and more.

Folk Theorem – Rough Intuition

- Consider any stage (one-shot) game
- Any $n-1$ agents can hold the remaining agent down to its minmax value in that one-shot game.
- So, they can enforce any strategy profile that guarantees that agent at least that minmax value, by threatening to play the minmax strategy forever if it ever revels.
- This is a ‘grim” (or “grim trigger”) strategy. Prisoner’s Dilemma example: “I’ll start by so cooperating but if you ever defect I’ll defect forever from then on”).
- Imagine every $n-1$ agents doing it to the remaining one.

Definitions

- Consider any n-player game $G = (N, A, u)$ and any payoff vector $r = (r_1, r_2, \ldots, r_n)$.
- Let $v_i = \min_{s_{-i} \in S_{-i}} \max_{s_i \in S_i} u_i(s_{-i}, s_i)$.
  - i's minmax value: the amount of utility i can get when $-i$ play a minmax strategy against him

Definition

A payoff profile r is enforceable if $r_i \geq v_i$.

Definition

A payoff profile r is feasible if there exist rational, non-negative values α_a such that for all i, we can express r_i as $\sum_{a \in A} \alpha_a u_i(a)$, with $\sum_{a \in A} \alpha_a = 1$.

- a payoff profile is feasible if it is a convex, rational combination of the outcomes in G.

Folk Theorem

Theorem (Folk Theorem)

Consider any n-player game G and any payoff vector $(r_1, r_2, \ldots, r_n)$.

1. *If r is the payoff in any Nash equilibrium of the infinitely repeated G with average rewards, then for each player i, r_i is enforceable.*
2. *If r is both feasible and enforceable, then r is the payoff in some Nash equilibrium of the infinitely repeated G with average rewards.*

Folk Theorem (Part 1)

Payoff in Nash $\Rightarrow$ enforceable

Part 1: Suppose r is not enforceable, i.e. $r_i < v_i$ for some i. Then consider a deviation of this player i to $b_i(s_{-i}(h))$ for any history h of the repeated game, where b_i is any best-response action in the stage game and $s_{-i}(h)$ is the equilibrium strategy of other players given the current history h. By definition of a minmax strategy, player i will receive a payoff of at least v_i in every stage game if he adopts this strategy, and so i's average reward is also at least v_i. Thus i cannot receive the payoff $r_i < v_i$ in any Nash equilibrium.

Folk Theorem (Part 2)

Feasible and enforceable ⇒ Nash

Part 2: Since r is a feasible payoff profile, we can write it as $r_i = \sum_{a \in A} \left(\frac{\beta_a}{\gamma}\right) u_i(a)$, where β_a and γ are non-negative integers[1] and $\gamma = \sum_{a \in A} \beta_a$. We're going to construct a strategy profile that will cycle through all outcomes $a \in A$ of G with cycles of length γ, each cycle repeating action a exactly β_a times. Let (a^t) be such a sequence of outcomes. Let's define a strategy s_i of player i to be a trigger version of playing (a^t): if nobody deviates, then s_i plays a_i^t in period t. However, if there was a period t' in which some player $j \neq i$ deviated, then s_i will play $(p_{-j})_i$, where (p_{-j}) is a solution to the minimization problem in the definition of v_j.

First observe that if everybody plays according to s_i, then, by construction, player i receives average payoff of r_i (look at averages over periods of length γ). Second, this strategy profile is a Nash equilibrium. Suppose everybody plays according to s_i, and player j deviates at some point. Then, forever after, player j will receive his $\min\max$ payoff $v_j \leq r_j$, rendering the deviation unprofitable.

[1] Recall that α_a were required to be rational. So we can take γ to be their common denominator.

Section 4

Bounded Rationality: repeated games played by automata

Bounded Rationality

- So far, we assumed players can engage in arbitrarily deep reasoning and mutual modelling, regardless of their complexity.
- Even in the relatively uncontroversial case of two-player zero-sum games, this is a questionable stance in practice
  - otherwise, for example, there would be no point in playing chess
- In the finitely repeated version of the Prisoner's Dilemma, each player's dominant strategy is to Defect.
- In reality, when people play, we typically observe a significant amount of cooperation.
- What might model this fact?

Finite-state automata

- A strategy for k repetitions of an m-action game is thus a specification of $\frac{m^k-1}{m-1}$ different actions.
- A naive encoding of a strategy as a table mapping each possible history to an action can be extremely inefficient.
  - the strategy of choosing D in every round can be represented using just the single-stage strategy D
  - the Tit-for-Tat strategy can be represented simply by specifying that the player mimic what his opponent did in the previous round
- One representation that exploits this structure is the finite-state automaton

Finite-state automata

Definition (Automaton)

Given a game $G = (N, A, u)$ that will be played repeatedly, an automaton M_i for player i is a four-tuple $(Q_i, q_i^0, \delta_i, f_i)$, where:

- Q_i is a set of states
- q_i^0 is the start state
- $\delta_i : Q_i \times A \mapsto Q_i$ is a transition function mapping the current state and an action profile to a new state; and
- $f_i : Q_i \mapsto A_i$ is a strategy function associating with every state an action for player i

Finite-state automata

- The machine begins in the start state q_i^0 , and in the first round plays the action given by $f_i(q_i^0)$.
- Using the transition function and the actions played by the other players in the first round, it then transitions automatically to the new state $\delta_i(q_i^0, a_1, ..., a_n)$ before the beginning of round 2
- It then plays the action $f_i(\delta_i(q_i^0, a_1, ..., a_n))$ in round two, and so on
- The current strategy and state at round t are given by:

$$
\begin{aligned}
a_i^t &= f_i(q_i^t) \\
q_i^{t+1} &= \delta_i(q_i^t, a_1^t, ..., a_n^t)
\end{aligned}
$$

Strategies for the Prisoner's Dilemma

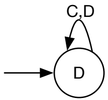

Figure: An automaton representing the Always Defect strategy

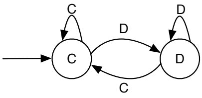

Figure: An automaton representing the Tit-for-Tat strategy

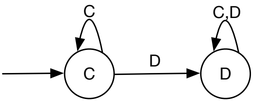

Figure: An automaton representing the Trigger strategy

Strategies represented as finite state automata

- Intuitively, automata with fewer states represent simpler strategies.
- One way to bound the rationality of the player is by limiting the number of states in the automaton
- Placing severe restrictions on the number of states induces an equilibrium in which cooperation always occurs.
- Additionally, it also causes the always-defect equilibrium to disappear. Why?
  - This equilibrium in a finitely repeated Prisoners Dilemma game depends on the assumption that each player can use backward induction.
  - In order to perform backward induction in a k-period repeated game, each player needs to keep track of at least k distinct states: one state to represent the choice of strategy in each repetition of the game.

Section 5

Learning in Repeated Games

Learning in Repeated Games

- We will cover two types of learning in repeated games
  - Fictitious Play
  - No-regret Learning
- Many other facets of learning in Game Theory are left out...

Fictitious Play

Definition (Fictitious Play)

At each time step, each player best responds to the empirical frequency of play of their opponent.

- Example using matching pennies
  - assuming a starting point – or fictitious past – with frequencies
    - $(1.5, 2)$ for player 2 (i.e. player 1's beliefs about player 2)
    - $(2, 1.5)$ for player 1 (i.e. player 2's beliefs about player 1).

Round	1's action	2's action	1's belief	2' belief
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Fictitious Play

Theorem (Convergence)

If the empirical distribution of each player's strategies converges in fictitious play, then it converges to Nash equilibrium.

Theorem

Each of the following are sufficient conditions for the empirical frequencies of play to converge in fictitious play

- *The game is zero sum;*
- *The game is solvable by iterated elimination of strictly dominated strategies;*
- *The game is a potential game*
- *The game is* $2 \times n$ *and has generic payoffs*

No-regret Learning

Definition (Regret)

The regret an agent experiences at time t for not having played s is $R^t(s) = \alpha^t - \alpha^t(s)$.

Definition (No-regret learning rule)

A learning rule exhibits no-regret if for any pure strategy of the agent s it holds that $Pr([\liminf R^t(s)] \leq 0) = 1$.

No-regret Learning

Definition (Regret Matching)

At each time step, each action is chosen with probability proportional to its regret. That is,

$$\sigma_i^{t+1}(s) = \frac{R^t(s)}{\sum_{s' \in S_i} R^t(s')}$$

- Exhibits no regret.
- Converges to a correlated equilibrium for finite games