CMSC 474, Introduction to Game Theory

17. Repeated Games

Mohammad T. Hajiaghayi
University of Maryland
Repeated Games

- Used by game theorists, economists, social and behavioral scientists as highly simplified models of various real-world situations
Finitely Repeated Games

- In repeated games, some game $G$ is played multiple times by the same set of agents.
  - $G$ is called the **stage game**
    - Usually (but not always) a normal-form game
  - Each occurrence of $G$ is called an **iteration**, **round**, or **stage**
- Usually each agent knows what all the agents did in the previous iterations, but not what they’re doing in the current iteration.
  - Thus, **imperfect information** with **perfect recall**
- Usually each agent’s payoff function is additive

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>3, 3</td>
<td>0, 5</td>
</tr>
<tr>
<td>$D$</td>
<td>5, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

**Prisoner’s Dilemma:**

**Iterated Prisoner’s Dilemma, 2 iterations:**

Agent 1: Round 1: $C$, Round 2: $D$

Agent 2: Round 1: $C$, Round 2: $C$

Total payoff: $3 + 5 = 5$, $3 + 0 = 3$
The repeated game has a much bigger strategy space than the stage game.

One kind of strategy is a stationary strategy:

- Use the same strategy in every stage game

More generally, an agent’s play at each stage may depend on what happened in previous iterations.
Examples

Some well-known IPD strategies:

- **AllC**: always cooperate
- **AllD**: always defect
- **Grim**: cooperate until the other agent defects, then defect forever
- **Tit-for-Tat (TFT)**: on 1st move, cooperate. On nth move, repeat the other agent’s (n−1)th move
- **Tit-for-Two-Tats (TFTT)**: like TFT, but only retaliates if the other agent defects twice in a row
- **Tester**: defect on round 1. If the other agent retaliates, play TFT. Otherwise, alternately cooperate and defect
- **Pavlov**: on 1st round, cooperate. Thereafter, win => use same action on next round; lose => switch to the other action ("win" means 3 or 5 points, "lose" means 0 or 1 point)
Backward Induction

- If the number of iterations is finite and all players know what it is, we can use backward induction to find a subgame-perfect equilibrium.

- This time it’s simpler than game-tree search.
  - Regardless of what move you make, the next state is always the same.
    - Another instance of the stage game.
  - The only difference is how many points you’ve accumulated so far.

- First calculate the SPE actions for round $n$ (the last iteration).

- Then for round $j = n-1, n-2, \ldots, 1$,
  - Common knowledge of rationality $\Rightarrow$ everyone will play their SPE actions after round $j$.
  - Construct a payoff matrix showing what the cumulative payoffs will be from round $j$ onward.
  - From this, calculate what the SPE actions will be at round $j$. 
Example

- $n$ repetitions of the Prisoner’s Dilemma
- Round $n$ (the last round)
  - SPE profile is $(D,D)$; each player gets 1
- Case $j = n-1$:
  - If everyone plays their SPE actions after round $j$, then
    - Cumulative payoffs = 1 + payoffs at round $j$
    - SPE actions at round $j$ are $(D,D)$; each player gets 2
- Case $j = n-2$:
  - If everyone plays SPE actions after round $j$, then
    - Cumulative payoffs = 2 + payoffs at round $j$
    - SPE actions at round $j$ are $(D,D)$; each player gets 3
    ...
- The SPE is to play $(D,D)$ on every round
- As in the Centipede game, there are both empirical and theoretical criticisms
Two-Player Zero-Sum Repeated Games

- In a two-player zero-sum repeated game, the SPE is for every player to play a minimax strategy at every round

- Your minimax strategy is best for you if the other agents also use their minimax strategies

- In some cases, the other agents won’t use those strategies
  - If you can predict their actions accurately, you may be able to do much better than the minimax strategy would do

- Why won’t the other agents use their minimax strategies?
  - Because they may be trying to predict your actions too
Roshambo (Rock, Paper, Scissors)

<table>
<thead>
<tr>
<th></th>
<th>A₁</th>
<th>Rock</th>
<th>Paper</th>
<th>Scissors</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rock</td>
<td>0, 0</td>
<td>–1, 1</td>
<td>1, –1</td>
<td></td>
</tr>
<tr>
<td>Paper</td>
<td>1, –1</td>
<td>0, 0</td>
<td>–1, 1</td>
<td></td>
</tr>
<tr>
<td>Scissors</td>
<td>–1, 1</td>
<td>1, –1</td>
<td>0, 0</td>
<td></td>
</tr>
</tbody>
</table>

- Nash equilibrium for the stage game:
  - choose randomly, \( P=1/3 \) for each move
- Nash equilibrium for the repeated game:
  - \( always \) choose randomly, \( P=1/3 \) for each move
- Expected payoff = 0
Roshambo (Rock, Paper, Scissors)

<table>
<thead>
<tr>
<th></th>
<th>A₁</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A₂</td>
<td>Rock</td>
<td>Paper</td>
<td>Scissors</td>
</tr>
<tr>
<td>Rock</td>
<td>0, 0</td>
<td>−1, 1</td>
<td>1, −1</td>
</tr>
<tr>
<td>Paper</td>
<td>1, −1</td>
<td>0, 0</td>
<td>−1, 1</td>
</tr>
<tr>
<td>Scissors</td>
<td>−1, 1</td>
<td>1, −1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

- 1999 international roshambo programming competition
  - www.cs.ualberta.ca/~darse/rsbpc1.html
  
  - Round-robin tournament:
    - 55 programs, 1000 iterations for each pair of programs
    - Lowest possible score = −55000; highest possible score = 55000
  
  - Average over 25 tournaments:
    - Lowest score (Cheesebot): −36006
    - Highest score (Iocaine Powder): 13038

  - http://www.veoh.com/watch/e1077915X5GNatn
Infinitely Repeated Games

- An infinitely repeated game in extensive form would be an infinite tree
  - Payoffs can’t be attached to any terminal nodes
- Let $r_i^{(1)}, r_i^{(2)}, \ldots$ be an infinite sequence of payoffs for agent $i$
  - the sum usually is infinite, so it can’t be $i$’s payoff
- Two common ways around this problem:
  1. **Average reward**: average over the first $k$ iterations; let $k \to \infty$
     $$\lim_{k \to \infty} \sum_{j=1}^{k} r_i^{(j)} / k$$
  2. **Future discounted reward**: $\sum_{j=1}^{\infty} \beta^{j-1} r_i^{(j)}$
     - $\beta \in [0,1)$ is a constant called the *discount factor*
     - Two possible interpretations:
       1. The agent cares more about the present than the future
       2. At each round, the game ends with probability $1 - \beta$
Nash Equilibria

- What are the Nash Equilibria in an infinitely repeated game?
  - Often many more than if the game were finitely repeated
  - Infinitely many Nash equilibria for the infinitely repeated prisoner’s dilemma

- There’s a “folk theorem” that tells what the possible equilibrium payoffs are in repeated games, if we use average rewards

- First we need some definitions …
Feasible Payoff Profiles

- A payoff profile \( \mathbf{r} = (r_1, r_2, \ldots, r_n) \) is **feasible** if it is a convex rational combination of \( G \)'s possible outcomes
  - i.e., for every action profile \( \mathbf{a}_j \) there is a rational nonnegative number \( c_j \) such that \( \sum_j c_j = 1 \) and \( \sum_j c_j \mathbf{u}(\mathbf{a}_j) = \mathbf{r} \)

- Intuitive meaning:
  - There’s a finite sequence of action profiles for which the average reward profile is \( \mathbf{r} \)

- Example: in the Prisoner’s Dilemma,
  \[
  \begin{align*}
  \mathbf{u}(C,C) &= (3,3) & \mathbf{u}(C,D) &= (0,5) \\
  \mathbf{u}(D,C) &= (5,0) & \mathbf{u}(D,D) &= (1,1)
  \end{align*}
  \]
  - \( \frac{1}{4} \mathbf{u}(C,C) + \frac{1}{2} \mathbf{u}(C,D) + \frac{1}{4} \mathbf{u}(D,C) + 0 \mathbf{u}(D,D) = (8/4, 13/4) \)
  - so \( (2, 13/4) \) is feasible
  - \( (5,5) \) isn’t feasible; no convex combination can produce it
  - \( (\pi/2, \pi/2) \) isn’t feasible; no **rational** convex combination can produce it

Keep repeating this sequence:

<table>
<thead>
<tr>
<th>Agent 1</th>
<th>Agent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>
Enforceable Payoff Profiles

- A payoff profile $\mathbf{r} = (r_1, \ldots, r_n)$ is **enforceable** if for each $i$,
  - $r_i \geq \text{player } i^{\prime}\text{'s minimax value in } G$

- Intuitive meaning:
  - If $i$ deviates from the sequence of action profiles that produces $\mathbf{r}$, the other agents can punish $i$ by reducing $i^{\prime}\text{'s average reward to } \leq i^{\prime}\text{'s minimax value}$

- The other agents can do this by using **grim trigger** strategies:
  - Generalization of the Grim strategy
    - If any agent $i$ deviates from the sequence of actions it is supposed to perform, then the other agents punish $i$ forever by playing their minimax strategies against $i$
The Theorem

Theorem: If $G$ is infinitely repeated game with average rewards, then

- If there’s a Nash equilibrium with payoff profile $r$, then $r$ is enforceable
- If $r$ is both feasible and enforceable, then there’s a Nash equilibrium with payoff profile $r$

Summary of the proof:

- **Part 1:** Use the definitions of minimax and best-response to show that in every equilibrium, each agent $i$’s average payoff $\geq i$’s minimax value
- **Part 2:** Show how to construct a Nash equilibrium that gives each agent $i$ an average payoff $r_i$
  - The agents are grim-trigger strategies that cycle in lock-step through a sequence of game outcomes $r^{(1)}, r^{(2)}, ..., r^{(n)}$ such that $r = u(r^{(1)}) + u(r^{(2)}) + ... + u(r^{(n)})$
  - No agent can do better by deviating, because the others will punish it $\Rightarrow$ Nash equilibrium
Iterated Prisoner’s Dilemma

- For a finitely iterated game with a large number of iterations, the practical effect can be roughly the same as if it were infinite.

- E.g., the Iterated Prisoner’s Dilemma

- Widely used to study the emergence of cooperative behavior among agents
  - e.g., Axelrod (1984), *The Evolution of Cooperation*

- Axelrod ran a famous set of tournaments
  - People contributed strategies encoded as computer programs
  - Axelrod played them against each other

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3, 3</td>
<td>0, 5</td>
</tr>
<tr>
<td>D</td>
<td>5, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

If I defect now, he might punish me by defecting next time.
**TFT with Other Agents**

- In Axelrod’s tournaments, TFT usually did best
  - It could establish and maintain cooperations with many other agents
  - It could prevent malicious agents from taking advantage of it

<table>
<thead>
<tr>
<th>TFT</th>
<th>AllC, TFT, TFTT, Grim, or Pavlov</th>
<th>TFT</th>
<th>AllD</th>
<th>TFT Tester</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>

...
Example:

- A real-world example of the IPD, described in Axelrod’s book:
  - World War I trench warfare

- Incentive to cooperate:
  - If I attack the other side, then they’ll retaliate and I’ll get hurt
  - If I don’t attack, maybe they won’t either

- Result: evolution of cooperation
  - Although the two infantries were supposed to be enemies, they avoided attacking each other
Summary

- Topics covered:
  - Finitely repeated games
  - Infinitely repeated games
  - Evolution of cooperation