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

Iterated Prisoner's Dilemma

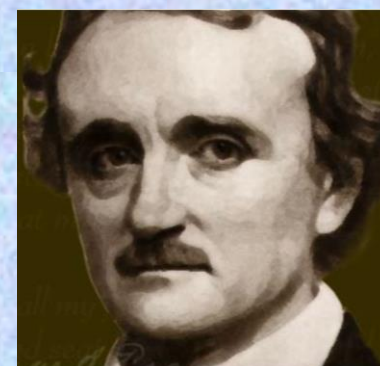
Roshambo

Repeated Ultimatum Game

Iterated Battle of the Sexes

Iterated Chicken Game

Repeated Stag Hunt



Repeated Matching Pennies

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

Prisoner's Dilemma:

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

Iterated Prisoner's Dilemma, 2 iterations:

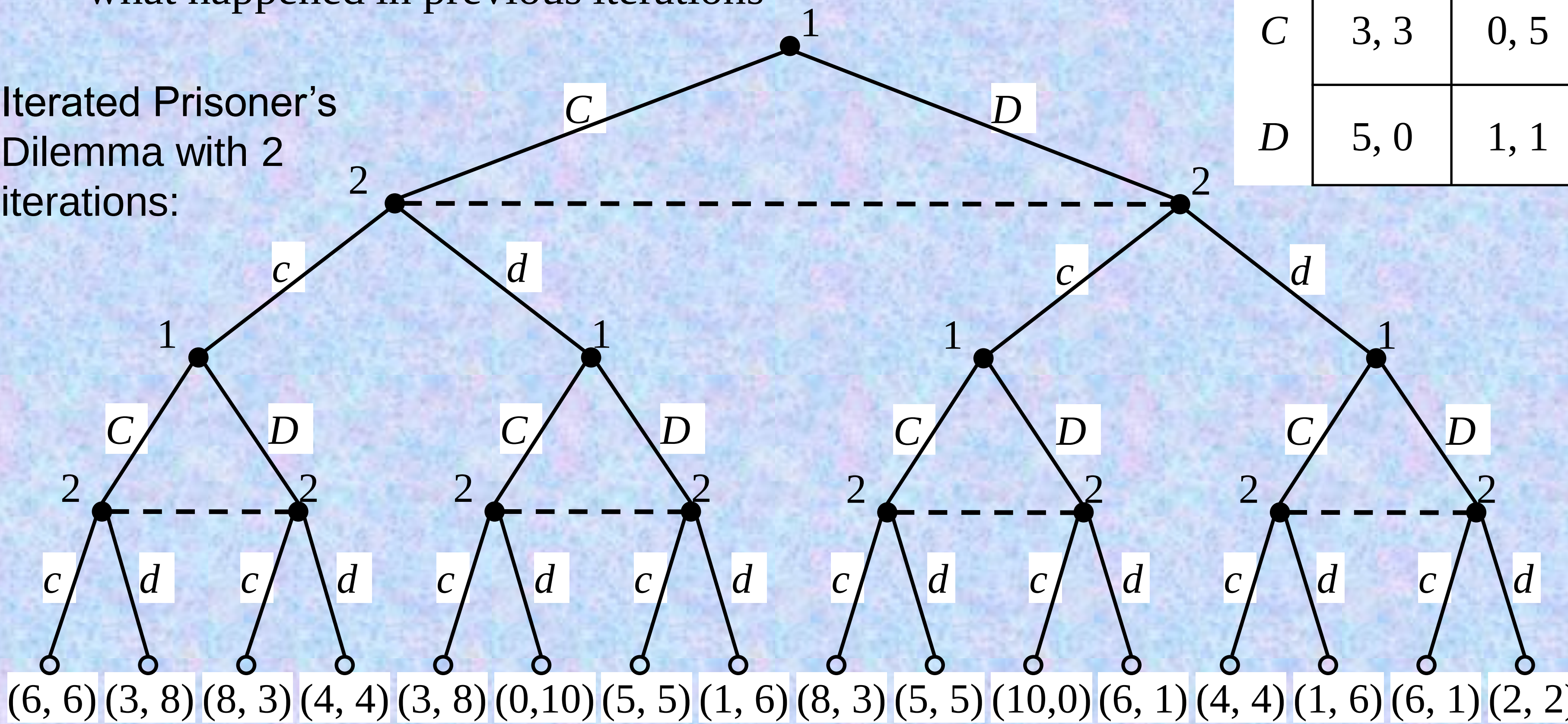

	Agent 1:	*Agent 2*:
Round 1:	C	C
Round 2:	D	C
Total payoff:	3+5 = 5	3+0 = 3

Strategies

- 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

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

Iterated Prisoner's Dilemma with 2 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 n^{th} 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)

AllC, Grim, TFT, or Pavlov	*AllC, Grim, TFT, or Pavlov*
C	C
C	C
C	C
C	C
C	C
⋮	⋮

TFT	*Tester*
C	*D*
D	C
C	C
C	C
C	C
C	C
C	C
⋮	⋮

TFTT	*Tester*
C	*D*
C	C
C	C
C	*D*
C	C
C	*D*
C	C
⋮	⋮

TFT or Grim	*AllD*
C	*D*
D	*D*
D	*D*
D	*D*
D	*D*
D	*D*
D	*D*
⋮	⋮

Pavlov	*AllD*
C	*D*
D	*D*
C	*D*
D	*D*
C	*D*
D	*D*
C	*D*
⋮	⋮

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 → 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

n	*C*	*D*
C	3, 3	0, 5
D	5, 0	1, 1

n–1	*C*	*D*
C	4, 4	1, 6
D	6, 1	2, 2

n–2	*C*	*D*
C	5, 5	2, 7
D	7, 2	3, 3

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)

A_2 \ A_1	Rock	Paper	Scissors
Rock	0, 0	–1, 1	1, –1
Paper	1, –1	0, 0	–1, 1
Scissors	–1, 1	1, –1	0, 0

- 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)

A_2 \ A_1	Rock	Paper	Scissors
Rock	0, 0	–1, 1	1, –1
Paper	1, –1	0, 0	–1, 1
Scissors	–1, 1	1, –1	0, 0

- 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**: $\sum_{j=1}^{\infty} \beta^j r_i^{(j)}$

   - $\beta \in [0,1)$ 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 $\Sigma_j\, c_j = 1$ and $\Sigma_j\, c_j\, \mathbf{u}(\mathbf{a}_j) = \mathbf{r}$
- Intuitive meaning:
  - There's a finite sequence of action profiles for which the average reward profile is $\mathbf{r}$

Keep repeating this sequence:

Agent 1	*Agent 2*
C	C
C	*D*
C	*D*
D	C

- Example: in the Prisoner's Dilemma,

$$\mathbf{u}(C,C) = (3,3) \qquad \mathbf{u}(C,D) = (0,5)$$

$$\mathbf{u}(D,C) = (5,0) \qquad \mathbf{u}(D,D) = (1,1)$$

  - ¼ $\mathbf{u}$(C,C) + ½ $\mathbf{u}$(C,D) + ¼ $\mathbf{u}$(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
  - (π/2, π/2) isn't feasible; no **rational** convex combination can produce it

Enforceable Payoff Profiles

- A payoff profile $\mathbf{r} = (r_1, \ldots, r_n)$ is **enforceable** if for each i,
  - $r_i \geq$ player i'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's average reward to $\leq i$'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

	Agent 1	*Agent 2*	
	C	C	
	C	*D*	
	C	*D*	
	D	C	
	C	C	
	C	*D*	
	C	*D*	
	D	C	
	C	$\boxed{D}$	deviate
	D	…	
	D	…	
punish	*D*	…	
	D	…	
	⋮	⋮	

The Theorem

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

- If there's a Nash equilibrium with payoff profile $\mathbf{r}$, then $\mathbf{r}$ is enforceable
- If $\mathbf{r}$ is both feasible and enforceable, then there's a Nash equilibrium with payoff profile $\mathbf{r}$

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 $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \ldots, \mathbf{r}^{(n)}$ such that $\mathbf{r} = \mathbf{u}(\mathbf{r}^{(1)}) + \mathbf{u}(r^{(2)}) + \ldots + \mathbf{u}(r^{(n)})$
    - No agent can do better by deviating, because the others will punish it => 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

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

- 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

If I defect now, he might punish me by defecting next time

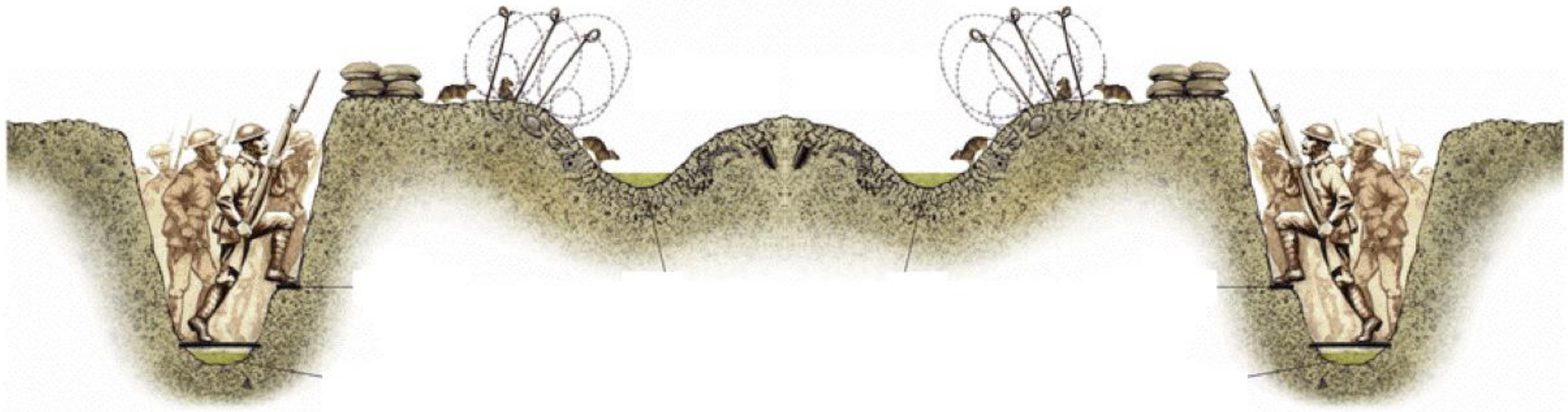
TFT with Other Agents

- 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

TFT	*AllC, TFT, TFTT, Grim, or Pavlov*	*TFT*	*AllD*	*TFT*	*Tester*
C	C	C	*D*	C	*D*
C	C	*D*	*D*	*D*	C
C	C	*D*	*D*	C	C
C	C	*D*	*D*	C	C
C	C	*D*	*D*	C	C
C	C	*D*	*D*	C	C
C	C	*D*	*D*	C	C
⋮	⋮	⋮	⋮	⋮	⋮

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