Outline

A. Introduction

B. Single Agent Learning

C. Game Theory

D. Multiagent Learning

E. Future Issues and Open Problems

Overview of Game Theory

- Models of Interaction
  - Normal-Form Games
  - Repeated Games
  - Stochastic Games
- Solution Concepts

Normal-Form Games

A normal-form game is a tuple $(n, \mathcal{A}_{1\ldots n}, R_{1\ldots n})$,

- n is the number of players,
- $\mathcal{A}_i$ is the set of actions available to player i
  - $\mathcal{A}$ is the joint action space $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$,
- R_i is player i's payoff function $\mathcal{A} \rightarrow \Re$.

$$R_1 = \begin{array}{cc} & a_2 \\ a_1 & \left(\begin{array}{ccc} & \vdots & \\ \cdots & R_1(a) & \cdots \\ & \vdots & \end{array}\right) \end{array} \qquad R_2 = \begin{array}{cc} & a_2 \\ a_1 & \left(\begin{array}{ccc} & \vdots & \\ \cdots & R_2(a) & \cdots \\ & \vdots & \end{array}\right) \end{array}$$

Example — Rock-Paper-Scissors

- Two players. Each simultaneously picks an action: *Rock*, *Paper*, or *Scissors*.
- The rewards:

Rock	beats	*Scissors*
Scissors	beats	*Paper*
Paper	beats	*Rock*

- The matrices:

$$R_1 = \begin{array}{cc} & \begin{array}{ccc} \text{R} & \text{P} & \text{S} \end{array} \\ \begin{array}{c} \text{R} \\ \text{P} \\ \text{S} \end{array} & \left(\begin{array}{ccc} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{array}\right) \end{array} \qquad R_2 = \begin{array}{cc} & \begin{array}{ccc} \text{R} & \text{P} & \text{S} \end{array} \\ \begin{array}{c} \text{R} \\ \text{P} \\ \text{S} \end{array} & \left(\begin{array}{ccc} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{array}\right) \end{array}$$

More Examples

- Matching Pennies

$$R_1 = \begin{array}{c} \\ H \\ T \end{array}\begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ H \\ T \end{array}\begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}$$

- Coordination Game

$$R_1 = \begin{array}{c} \\ A \\ B \end{array}\begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ A \\ B \end{array}\begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

- Bach or Stravinsky

$$R_1 = \begin{array}{c} \\ B \\ S \end{array}\begin{array}{c} \begin{array}{cc} B & S \end{array} \\ \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ B \\ S \end{array}\begin{array}{c} \begin{array}{cc} B & S \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \end{array}$$

More Examples

- Prisoner's Dilemma

$$R_1 = \begin{array}{c} \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

- Three-Player Matching Pennies

Three-Player Matching Pennies

- Three players. Each simultaneously picks an action: *Heads* or *Tails*.

- The rewards:

Player One	wins by matching	Player Two,
Player Two	wins by matching	Player Three,
Player Three	wins by *not* matching	Player One.

Three-Player Matching Pennies

- The matrices:

$$R_1(\langle \cdot, \cdot, H \rangle) = \begin{array}{c} \\ H \\ T \end{array}\begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \quad R_1(\langle \cdot, \cdot, T \rangle) = \begin{array}{c} \\ H \\ T \end{array}\begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

$$R_2(\langle \cdot, \cdot, H \rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad R_2(\langle \cdot, \cdot, T \rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$R_3(\langle \cdot, \cdot, H \rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \quad R_3(\langle \cdot, \cdot, T \rangle) = \begin{array}{c} H \\ T \end{array}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

Strategies

- What can players do?
  - Pure strategies (a_i): select an action.
  - Mixed strategies (σ_i): select an action according to some probability distribution.

Strategies

- Notation.
  - σ is a joint strategy for all players.

$$R_i(\sigma) = \sum_{a \in \mathcal{A}} \sigma(a) R_i(a)$$

  - σ_{-i} is a joint strategy for all players except i.
  - $\langle \sigma_i, \sigma_{-i} \rangle$ is the joint strategy where i uses strategy σ_i and everyone else σ_{-i}.

Types of Games

- Zero-Sum Games (a.k.a. constant-sum games)

$$R_1 + R_2 = 0$$

  Examples: Rock-paper-scissors, matching pennies.

- Team Games

$$\forall i, j \qquad R_i = R_j$$

  Examples: Coordination game.

- General-Sum Games (a.k.a. all games)
  Examples: Bach or Stravinsky, three-player matching pennies, prisoner's dilemma

Repeated Games

- You can't learn if you only play a game once.
- Repeatedly playing a game raises new questions.
  - How many times? Is this common knowledge?

    Finite Horizon   Infinite Horizon

  - Trading off present and future reward?

    $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} r_t$ $\sum_{t=1}^{\infty} \gamma^t r_t$

    Average Reward   Discounted Reward

Repeated Games — Strategies

- What can players do?
  - Strategies can depend on the history of play.

$$\sigma_i : \mathcal{H} \rightarrow PD(\mathcal{A}_i) \qquad \text{where} \quad \mathcal{H} = \bigcup_{n=0}^{\infty} \mathcal{A}^n$$

  - Markov strategies a.k.a. stationary strategies

$$\forall a^{1\ldots n} \in \mathcal{A} \qquad \sigma_i(a^1, \ldots, a^n) = \sigma(a^n)$$

  - k-Markov strategies

$$\forall a_{1\ldots n} \in \mathcal{A} \qquad \sigma_i(a_1, \ldots, a_n) = \sigma(a_{n-k}, \ldots, a_n)$$

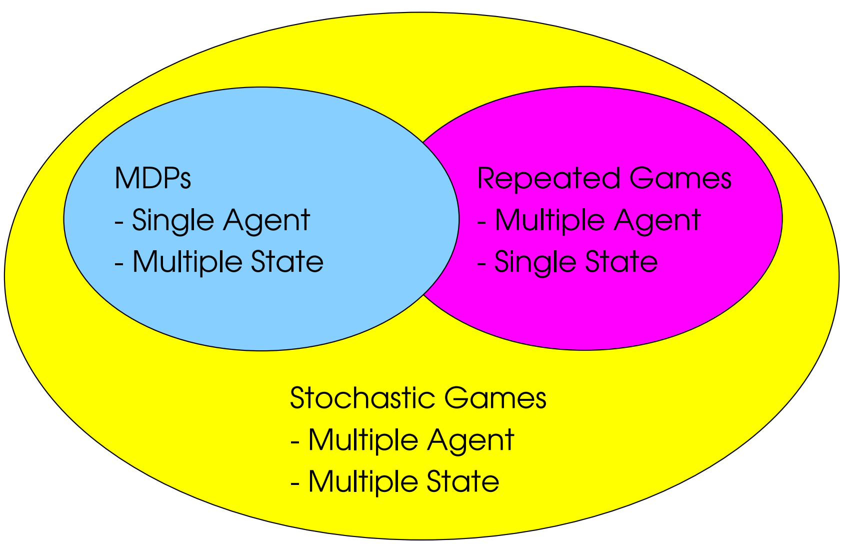
Repeated Games — Examples

- Iterated Prisoner's Dilemma

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \mathrm{C} & \mathrm{D} \end{array} \\ \begin{array}{c} \mathrm{C} \\ \mathrm{D} \end{array} & \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c c} & \begin{array}{cc} \mathrm{C} & \mathrm{D} \end{array} \\ \begin{array}{c} \mathrm{C} \\ \mathrm{D} \end{array} & \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

  - The single most examined repeated game!
  - Repeated play can justify behavior that is not rational in the one-shot game.
  - Tit-for-Tat (TFT)
    - ∗ Play opponent's last action (C on round 1).
    - ∗ A 1-Markov strategy.

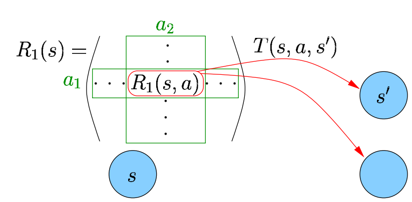
Stochastic Games



Stochastic Games — Definition

A stochastic game is a tuple $(n, \mathcal{S}, \mathcal{A}_{1\ldots n}, T, R_{1\ldots n})$,

- n is the number of agents,
- $\mathcal{S}$ is the set of states,
- $\mathcal{A}_i$ is the set of actions available to agent i,
  - $\mathcal{A}$ is the joint action space $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$,
- T is the transition function $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$,
- R_i is the reward function for the ith agent $\mathcal{S} \times \mathcal{A} \rightarrow \Re$.



Stochastic Games — Policies

- What can players do?
  - Policies depend on history and the current state.

$$\pi_i : \mathcal{H} \times \mathcal{S} \to PD(\mathcal{A}_i) \qquad \text{where} \quad \mathcal{H} = \bigcup_{n=0}^{\infty} (\mathcal{S} \times \mathcal{A})^n$$

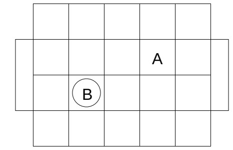
  - Markov polices a.k.a. stationary policies

$$\forall h, h' \in \mathcal{H}\ \forall s \in \mathcal{S} \qquad \pi_i(h, s) = \pi(h', s)$$

  - Focus on learning Markov policies, but the learning itself is a non-Markovian policy.

Example — Soccer

(Littman, 1994)



- Players: Two.
- States: Player positions and ball possession (780).
- Actions: N, S, E, W, Hold (5).
- Transitions:
  - Simultaneous action selection, random execution.
  - Collision could change ball possession.
- Rewards: Ball enters a goal.

Example — Goofspiel

- Players hands and the deck have cards $1 \ldots n$.
- Card from the deck is bid on secretly.
- Highest card played gets points equal to the card from the deck.
- Both players discard the cards bid.
- Repeat for all n deck cards.

n	$\lvert S \rvert$	$\lvert S \times A \rvert$	SIZEOF(π or Q)	V(det)	V(random)
4	692	15150	~ 59KB	-2	-2.5
8	3×10^6	1×10^7	~ 47MB	-20	-10.5
13	1×10^{11}	7×10^{11}	~ 2.5TB	-65	-28

Stochastic Games — Facts

- If $n = 1$, it is an MDP.
- If $\lvert S \rvert = 1$, it is a repeated game.
- If the other players play a stationary policy, it is an MDP to the remaining player.

$$\hat{T}(s, a_i, s') = \sum_{a_{-i} \in \mathcal{A}_{-i}} \pi_{-i}(s, a) T(s, \langle a_i, a_{-i} \rangle, s')$$

  - The interesting case, then, is when the other agents are not stationary, i.e., are learning.

Overview of Game Theory

- Models of Interaction
- Solution Concepts

Normal Form Games	Repeated/Stochastic Games
– Dominance	– Nash Equilibria
– Minimax	– Universally Consistent
– Pareto Efficiency	
– Nash Equilibria	
– Correlated Equilibria	

Dominance

- An action is strictly dominated if another action is always better, i.e,

$$\exists a'_i \in \mathcal{A}_i \; \forall a_{-i} \in \mathcal{A}_{-i} \qquad R_i(\langle a'_i, a_{-i}\rangle) > R_i(\langle a_i, a_{-i}\rangle).$$

- Consider prisoner's dilemma.

$$R_1 = \begin{array}{c} \\ \text{C} \\ \text{D} \end{array} \begin{array}{c} \text{C} \quad \text{D} \\ \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{C} \\ \text{D} \end{array} \begin{array}{c} \text{C} \quad \text{D} \\ \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

  – For both players, D dominates C.

Iterated Dominance

- Actions may be dominated by mixed strategies.

$$R_1 = \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \text{D} \quad \text{E} \\ \begin{pmatrix} 1 & 1 \\ 4 & 0 \\ 0 & 4 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \text{D} \quad \text{E} \\ \begin{pmatrix} 4 & 0 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \end{array}$$

- If strictly dominated actions should not be played. . .

$$R_1 = \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \text{D} \quad \text{E} \\ \begin{pmatrix} 1 & 1 \\ 4 & 0 \\ 0 & 4 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \text{D} \quad \text{E} \\ \begin{pmatrix} 4 & 0 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \end{array}$$

- This game is said to be dominance solvable.

Minimax

- Consider matching pennies.

$$R_1 = \begin{array}{c} \\ \text{H} \\ \text{T} \end{array} \begin{array}{c} \text{H} \quad \text{T} \\ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c} \\ \text{H} \\ \text{T} \end{array} \begin{array}{c} \text{H} \quad \text{T} \\ \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}$$

- Q: What do we do when the world is out to get us?
  A: Make sure it can't.
- Play strategy with the best worst-case outcome.

$$\operatorname*{argmax}_{\sigma_i \in \Delta(\mathcal{A}_i)} \; \min_{a_{-i} \in \mathcal{A}_{-i}} \; R_i(\langle \sigma_i, \sigma_{-i}\rangle)$$

- Minimax optimal strategy.

Minimax

- Back to matching pennies.

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \text{H} & \text{T} \end{array} \\ \begin{array}{c} \text{H} \\ \text{T} \end{array} & \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \sigma_1^*$$

- Consider Bach or Stravinsky.

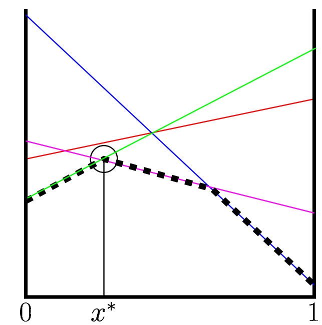
$$R_1 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \qquad \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = \sigma_1^*$$

- Minimax optimal guarantees the saftey value.
- Minimax optimal never plays dominated strategies.

Minimax — Linear Programming

- Minimax optimal strategies via linear programming.

$$\operatorname*{argmax}_{\sigma_i \in \Delta(\mathcal{A}_i)} \min_{a_{-i} \in \mathcal{A}_{-i}} R_i(\langle \sigma_i, \sigma_{-i} \rangle)$$



Pareto Efficiency

- A joint strategy is Pareto efficient if no joint strategy is better for all players, i.e.,

$$\forall a' \in \mathcal{A}\ \exists i \in 1, \ldots, n \qquad R_i(a) \geq R_i(a')$$

- In zero-sum games, all strategies are Pareto efficient.

Pareto Efficiency

- Consider prisoner's dilemma.

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \text{C} & \text{D} \end{array} \\ \begin{array}{c} \text{C} \\ \text{D} \end{array} & \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c c} & \begin{array}{cc} \text{C} & \text{D} \end{array} \\ \begin{array}{c} \text{C} \\ \text{D} \end{array} & \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

  - $\langle D, D \rangle$ is not Pareto efficient.

- Consider Bach or Stravinsky.

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \end{array}$$

  - $\langle B, B \rangle$ and $\langle S, S \rangle$ are Pareto efficient.

Nash Equilibria

- What action should we play if there are no dominated actions?
- Optimal action depends on actions of other players.
- A best response set is the set of all strategies that are optimal given the strategies of the other players.

$$\mathrm{BR}_i(\sigma_{-i}) = \{\sigma_i \quad | \quad \forall \sigma'_i \quad R_i(\langle \sigma_i, \sigma_{-i} \rangle) \geq R_i(\langle \sigma'_i, \sigma_{-i} \rangle)\}$$

- A Nash equilibrium is a joint strategy, where all players are playing best responses to each other.

$$\forall i \in \{1 \ldots n\} \qquad \sigma_i \in \mathrm{BR}_i(\sigma_{-i})$$

Nash Equilibria

- A Nash equilibrium is a joint strategy, where all players are playing best responses to each other.

$$\forall i \in \{1 \ldots n\} \qquad \sigma_i \in \mathrm{BR}_i(\sigma_{-i})$$

- Since each player is playing a best response, no player can gain by unilaterally deviating.
- Dominance solvable games have obvious equilibria.
  - Strictly dominated actions are never best responses.
  - Prisoner's dilemma has a single Nash equilibrium.

Examples of Nash Equilibria

- Consider the coordination game.

$$R_1 = \begin{array}{cc} & \begin{array}{cc} \mathsf{A} & \mathsf{B} \end{array} \\ \begin{array}{c} \mathsf{A} \\ \mathsf{B} \end{array} & \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{cc} & \begin{array}{cc} \mathsf{A} & \mathsf{B} \end{array} \\ \begin{array}{c} \mathsf{A} \\ \mathsf{B} \end{array} & \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array}$$

- Consider Bach or Stravinsky.

$$R_1 = \begin{array}{cc} & \begin{array}{cc} \mathsf{B} & \mathsf{S} \end{array} \\ \begin{array}{c} \mathsf{B} \\ \mathsf{S} \end{array} & \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{cc} & \begin{array}{cc} \mathsf{B} & \mathsf{S} \end{array} \\ \begin{array}{c} \mathsf{B} \\ \mathsf{S} \end{array} & \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{2} \end{pmatrix} \end{array}$$

Examples of Nash Equilibria

- Consider matching pennies.

$$R_1 = \begin{array}{cc} & \begin{array}{cc} \mathsf{H} & \mathsf{T} \end{array} \\ \begin{array}{c} \mathsf{H} \\ \mathsf{T} \end{array} & \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{cc} & \begin{array}{cc} \mathsf{H} & \mathsf{T} \end{array} \\ \begin{array}{c} \mathsf{H} \\ \mathsf{T} \end{array} & \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \end{array}$$

  - No pure strategy Nash equilibria. Mixed strategies?

$$\mathrm{BR}_1\Big(\langle 1/2, 1/2 \rangle\Big) = \{\sigma_1\}$$

  - Corresponds to the minimax strategy.

Existence of Nash Equilibria

- All finite normal-form games have at least one Nash equilibrium. (Nash, 1950)
- In zero-sum games...
  - Equilibria all have the same value and are interchangeable.

$$\langle \sigma_1, \sigma_2 \rangle, \langle \sigma_1', \sigma_2' \rangle \text{ are Nash } \Rightarrow \langle \sigma_1, \sigma_2' \rangle \text{ is Nash.}$$

  - Equilibria correspond to minimax optimal strategies.

Computing Nash Equilibria

- The exact complexity of computing a Nash equilibrium is an open problem. (Papadimitriou, 2001)
- Likely to be NP-hard. (Conitzer & Sandholm, 2003)
- Lemke-Howson Algorithm.
- For two-player games, bilinear programming solution.

Fictitious Play

(Brown, 1949; Robinson 1951)

- An iterative procedure for computing an equilibrium.

1. Initialize $C_i(a_i \in \mathcal{A}_i)$, which counts the number of times player i chooses action a_i.
2. Repeat.

   (a) Choose $a_i \in BR(C_{-i})$.
   (b) Increment $C_i(a_i)$.

Fictitious Play

(Fudenberg & Levine, 1998)

- If C_i converges, then what it converges to is a Nash equilibrium.
- When does C_i converge?
  - Two-player, two-action games.
  - Dominance solvable games.
  - Zero-sum games.
- This could be turned into a learning rule.

Correlated Equilibria

- Is there a way to be fair in Bach or Stravinsky?

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{2} \end{pmatrix} \end{array}$$

  - Suppose we wanted to both go to Bach or both go to Stravinsky with equal probability?
  - We want to correlate our action selection.

$$\begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \end{array} \quad \text{but not} \quad \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix} \end{array}$$

Correlated Equilibria

- Assume a shared randmoizer (e.g., a coin flip) exists.
- Define a new concept of equilibrium.
  - Let σ be a probability distribution over *joint actions*.
  - Each player observes their own action in a joint action sampled from σ.
  - σ is a correlated equilibrium if no player can gain by deviating from their prescribed action.

$$\forall i \quad a_i \in \mathrm{BR}_i(\sigma_{-i}|\sigma, a_i)$$

Correlated Equilibria

- Back to Bach or Stravinsky.

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} \boxed{2} & 0 \\ 0 & \boxed{1} \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{2} \end{pmatrix} \end{array}$$

$$\sigma = \begin{array}{c c} & \begin{array}{cc} \text{B} & \text{S} \end{array} \\ \begin{array}{c} \text{B} \\ \text{S} \end{array} & \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \end{array}$$

- All Nash equilibria are correlated equilibria.
- All mixtures of Nash are correlated equilibria.

Overview of Game Theory

- Models of Interaction
- Solution Concepts

Normal Form Games
- Dominance
- Minimax
- Pareto Efficiency
- Nash Equilibria
- Correlated Equilibria

Repeated/Stochastic Games
- Nash Equilibria
- Universally Consistent

Nash Equilibria in Repeated Games

- Obviously, Markov strategy equilibria exist.
- Consider iterated prisoner's dilemma and TFT.

$$R_1 = \begin{array}{c c} & \begin{array}{cc} \mathsf{C} & \mathsf{D} \end{array} \\ \begin{array}{c} \mathsf{C} \\ \mathsf{D} \end{array} & \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} \end{array} \qquad R_2 = \begin{array}{c c} & \begin{array}{cc} \mathsf{C} & \mathsf{D} \end{array} \\ \begin{array}{c} \mathsf{C} \\ \mathsf{D} \end{array} & \begin{pmatrix} 3 & 4 \\ 0 & 1 \end{pmatrix} \end{array}$$

  - With average reward, what's a best response?
    * Always D has a value of 1.
    * D then C has a value of 2.5
    * Always C and TFT have a value of 3.
  - Hence, both players following TFT is Nash.

Nash Equilibria in Repeated Games

- The TFT equilibria is strictly preferred to all Markov strategy equilibria.
- The TFT strategy plays a dominated action.
- TFT uses a threat to enforce compliance.
- TFT is not a special case.

Nash Equilibria in Repeated Games

Folk Theorem. For any repeated game with average reward, every *feasible* and *enforceable* vector of payoffs for the players can be achieved by some Nash equilibrium strategy. (Osborne & Rubinstein, 1994)

- A payoff vector is *feasible* if it is a linear combination of individual action payoffs.
- A payoff vector is *enforceable* if all players get at least their minimax value.

Nash Equilibria in Repeated Games

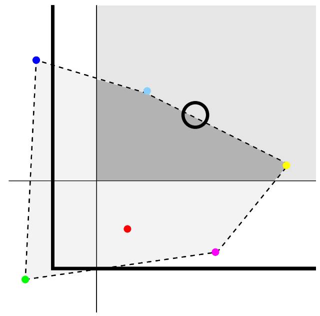
Folk Theorem. For any repeated game with average reward, every *feasible* and *enforceable* vector of payoffs for the players can be achieved by some Nash equilibrium strategy. (Osborne & Rubinstein, 1994)

- Players' follow a deterministic sequence of play that achieves the payoff vector.
- Any deviation is punished.
- The threat keeps players from deviating as in TFT.

Computing Repeated Game Equilibria

(Littman & Stone, 2003)

- Polynomial time algorithm for finding a Nash equilibrium in a repeated game.
  - Find a feasible and enforceable payoff vector.
  - Construct a strategy that punishes deviance.



Universally Consistent

- A.k.a. Hannan consistent, regret minimizing.
- For a history $h = a^1, a^2, \ldots, a^n \in \mathcal{A}$, define regret for player i,

$$\text{Regret}_i(h) = \left(\max_{a_i \in \mathcal{A}_i} \sum_{t=1}^{n} R(\langle a_i, a_{-i}^t \rangle) \right) - \sum_{t=1}^{n} R_i(a^t)$$

i.e., the difference between the reward that could have been received by a stationary strategy and the actual reward received.

Universally Consistent

- A strategy σ_i is universally consistent if for any $\epsilon > 0$ there exists a T such that for all σ_{-i} and $t > T$,

$$\Pr\left[\frac{\text{Regret}_i\,(a^1, \ldots, a^t)}{t} > \epsilon \;\middle|\; \langle \sigma_i, \sigma_{-i} \rangle \right] < \epsilon$$

  i.e., with high probability the average regret is low for all strategies of the other players.

- If regret is zero, then must be getting at least the minimax value.

Nash Equilibria in Stochastic Games

- Consider Markov policies.
- A best response set is the set of all Markov policies that are optimal given the other players' policies.

$$\mathrm{BR}_i(\pi_{-i}) = \left\{ \pi_i \;\middle|\; \begin{array}{l} \forall \pi_i' \; \forall s \in \mathcal{S} \\ V_i^{\langle \pi_i, \pi_{-i} \rangle}(s) \geq V_i^{\langle \pi_i', \pi_{-i} \rangle}(s) \end{array} \right\}$$

- A Nash equilibrium is a joint policy, where all players are playing best responses to each other.

$$\forall i \in \{1 \ldots n\} \qquad \pi_i \in \mathrm{BR}_i(\pi_{-i})$$

Nash Equilibria in Stochastic Games

- All discounted reward and zero-sum average reward stochastic games have at least one Nash equilibrium. (Shapley, 1953; Fink, 1964)

- Stochastic games are the general model.

- Nash equilibria in stochastic games has certainly received the most attention.