Decision Making

in Robots and Autonomous Agents

Repeated, Stochastic and Bayesian Games

Subramanian Ramamoorthy
School of Informatics

26 February, 2013

Repeated Game

- 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?

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

Average Reward

$$\sum_{t=1}^{\infty} \gamma^t r_t$$

Discounted Reward

Repeated Game - 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 Game - Examples

- Iterated Prisoner's Dilemma

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

  - 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.

Well Known IPD Strategies

- AllC/D: always cooperate/defect
- Grim: cooperate until the other agent defects, then defect forever
- Tit-for-Tat (TFT): on 1^{st} move, cooperate. On n^{th} move, repeat
- the other agent's $(n-1)^{th}$ move
- Tit-for-Two-Tats (TFTT): like TFT, but only only retaliates if the other agent defects twice
- Tester: defect on round 1. If the other agent retaliates, play TFT. Otherwise, alternately C/D
- Pavlov: on 1st round, cooperate. Thereafter, win => use same action next; lose => switch

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
⋮	⋮

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

Nash Equilibria – Repeated Game

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

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

  - 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 – Repeated Game

- 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 – Repeated Game

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 – Repeated Game

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.

Equilibria by 'Learning' – 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.

Minimax by Regret Minimization

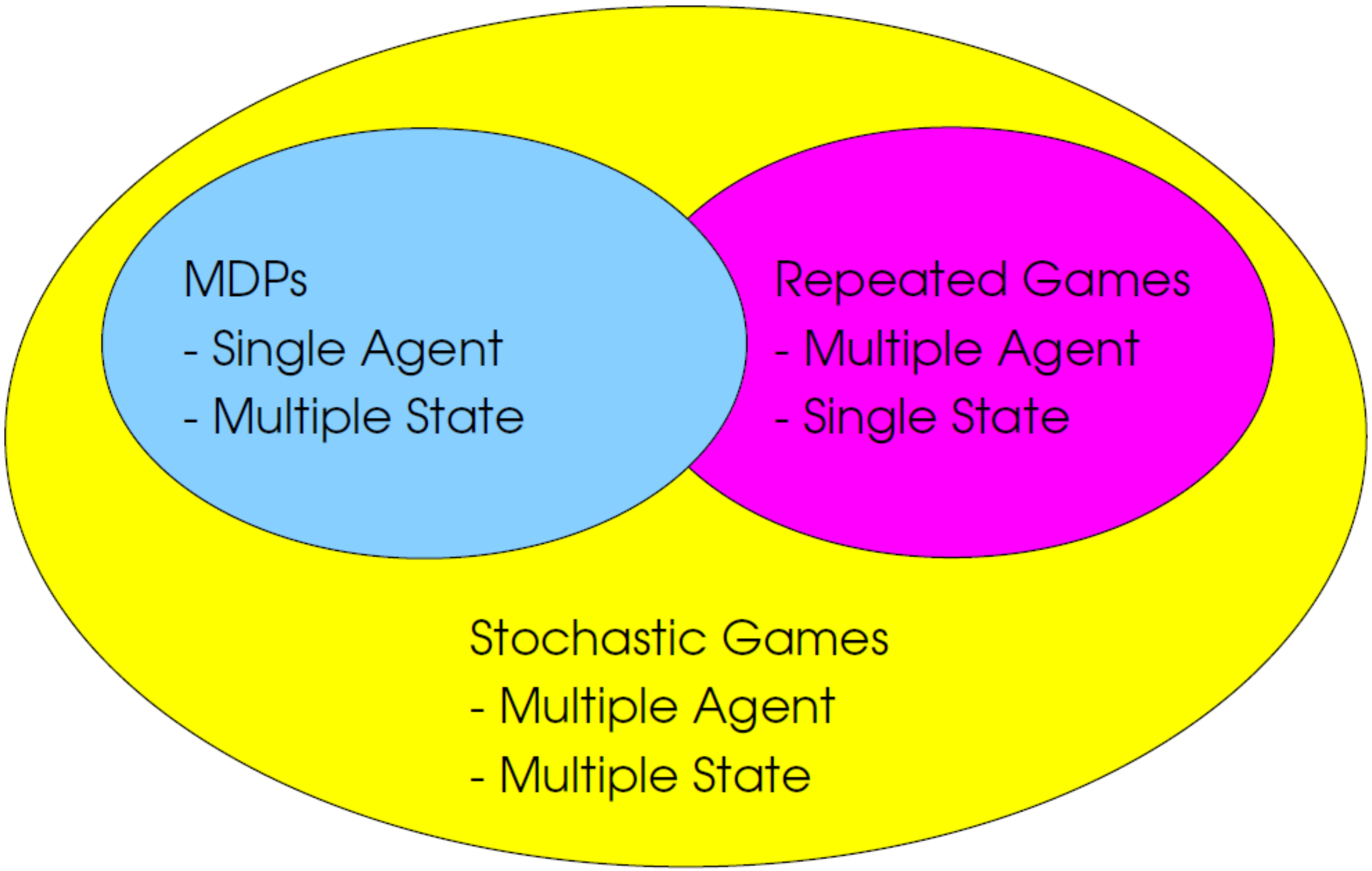
- A strategy σ_i is universally consistent if for any $\epsilon > 0$ there exists a T such that for all σ_{-i} and $t > T$,

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

  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.

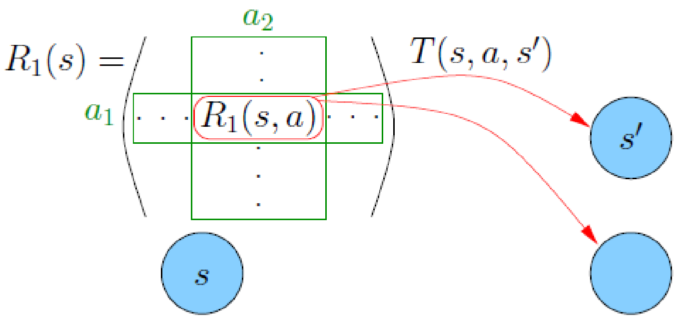
Stochastic Games



Stochastic Game - Setup

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 Game - Policies

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

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

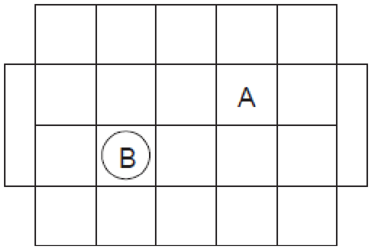
  - Markov polices a.k.a. stationary policies

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

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

Stochastic Game - Example

(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.

Stochastic Game - Remarks

- If $n = 1$, it is an MDP.
- If $|S| = 1$, 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.

Nash Equilibria – Stochastic Game

- 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 – Stochastic Game

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

Incomplete Information

- So far, we assumed that everything relevant about the game being played is common knowledge to all the players:
  - the number of players
  - the actions available to each
  - the payoff vector associated with each action vector
- True even for imperfect-information games
  - The actual moves aren't common knowledge, but the game is
- We'll now consider games of **incomplete (*not imperfect*) *information***
  - Players are uncertain about the game being played

Incomplete Information

	L	R
T	$100,\ a$	$1-\epsilon, b$
B	$2,\ c$	$1,\ d$

- Consider the payoff matrix shown here
  - ε is a small positive constant; Agent 1 knows its value
- Agent 1 doesn't know the values of *a, b, c, d*
  - Thus the matrix represents a ***set of games***
  - Agent 1 doesn't know which of these games is the one being played
- Agent 1 seeks strategy that works despite lack of knowledge

- If Agent 1 thinks Agent 2 is malicious, then Agent 1 might want to play a maxmin, or "safety level," strategy
  - minimum payoff of T is 1–ε
  - minimum payoff of B is 1
- So agent 1's maxmin strategy is B

Regret

- Suppose Agent 1 doesn't think Agent 2 is malicious
- Agent 1 might reason as follows:
  - If Agent 2 plays *R,* then *1's* strategy changes 1's payoff by only a small amount
    - Payoff is 1 or 1–ε;
    - Agent 1's difference is only ε
  - If Agent 2 plays *L,* then 1's strategy changes 1's payoff by a much bigger amount
    - Either 100 or 2, difference is 98
  - If Agent 1 chooses *T,* this will minimize *1's* worst-case **regret**
    - Maximum difference between the payoff of the chosen action and the payoff of the other action

	L	R
T	$100,\ a$	$1-\epsilon, b$
B	$2,\ c$	$1,\ d$

Minimax Regret

- Suppose i plays action a_i and the other agents play action profile a_{-i}
- i's regret: amount i lost by playing a_i instead of i's best response to a_{-i}

$$\text{regret}(a_i, \mathbf{a}_{-i}) = \left[\max_{a_i' \in A_i} u_i\left(a_i', \mathbf{a}_{-i}\right)\right] - u_i\left(a_i, \mathbf{a}_{-i}\right)$$

- i doesn't know what a_{-i} will be, but can consider worst case:
  - maximum regret for a_i , maximized over every possible a_{-i}

$$\max_{\mathbf{a}_{-i} \in \mathbf{A}_{-i}} \text{regret}(a_i, \mathbf{a}_{-i}) = \max_{\mathbf{a}_{-i} \in \mathbf{A}_{-i}} \left(\left[\max_{a_i' \in A_i} u_i\left(a_i', \mathbf{a}_{-i}\right)\right] - u_i\left(a_i, \mathbf{a}_{-i}\right)\right)$$

Minimax Regret

- **Minimax regret action**: an action with the smallest maximum regret

$$\underset{a_i \in A_i}{\operatorname{argmin}} \max_{\mathbf{a}_{-i} \in \mathbf{A}_{-i}} \operatorname{regret}(a_i) = \underset{a_i \in A_i}{\operatorname{argmin}} \max_{\mathbf{a}_{-i} \in \mathbf{A}_{-i}} \left(\left[\max_{a_i' \in A_i} u_i(a_i', \mathbf{a}_{-i}) \right] - u_i(a_i, \mathbf{a}_{-i}) \right)$$

- Can extend to a solution concept
  - All agents play minimax regret actions
  - This is one way to deal with the incompleteness, but often we can do more with the representation

Bayesian Games

- In the previous example, we knew the set ***G*** of all possible games, but didn't know anything about which game in ***G***
  - Enough information to put a probability distribution over games
- A **Bayesian Game** is a class of games ***G*** that satisfies two fundamental conditions
- ***Condition 1:***
  - The games in ***G*** have the same number of agents, and the same strategy space (set of possible strategies) for each agent. The only difference is in the payoffs of the strategies.
- This condition isn't very restrictive
  - Other types of uncertainty can be reduced to the above, by reformulating the problem

An Example

- Suppose we don't know whether player 2 only has strategies L and R, or also an additional strategy C:

Game G_1

	L	R
U	1, 1	1, 3
D	0, 5	1, 13

Game G_2

	L	C	R
U	1, 1	0, 2	1, 3
D	0, 5	2, 8	1, 13

- If player 2 doesn't have strategy C, this is equivalent to having a strategy C that's strictly dominated by other strategies:
  - Nash equilibria for G_1' are the same as for G_1

Game G_1'

	L	C	R
U	1, 1	0, −100	1, 3
D	0, 5	2, −100	1, 13

  - Problem is reduced to whether C's payoffs are those of G_1' or G_2

Bayesian Games

Condition 2 (common prior):

- The probability distribution over the games in ***G*** is **common knowledge** (i.e., known to all the agents)

- So a Bayesian game defines
  - the uncertainties of agents about the game being played,
  - what each agent believes the other agents believe about the game being played
- The beliefs of the different agents are posterior probabilities
  - Combine the common prior distribution with individual “private signals” (what’s “revealed” to the individual players)
- The common-prior assumption rules out whole families of games
  - But it greatly simplifies the theory, so most work in game theory uses it

The Bayesian Game Model

A Bayesian game consists of

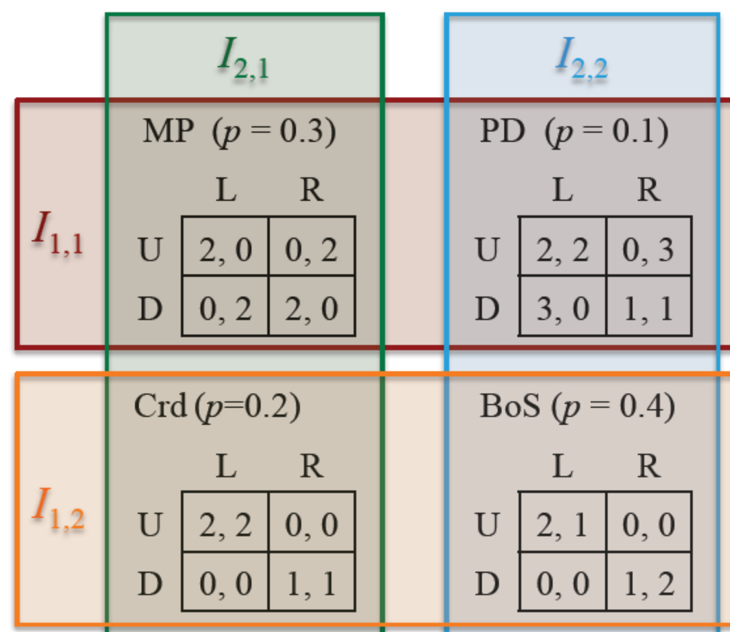
- a set of games that differ only in their payoffs
- a common (known to all players) prior distribution over them
- for each agent, a partition structure (set of information sets) over the games

Bayesian Game: Information Sets Defn.

A Bayesian game is a 4-tuple (N,G,P,I)

- *N is a set of agents*
- ***G** is a set of N-agent games*
- For every agent i, every game in ***G*** has the same strategy space
- *P is a **common prior*** over ***G***
  - ***common:*** common knowledge (known to all the agents)
  - ***prior:*** probability before learning any additional information
- $I = (I_1, \ldots, I_N)$ is a tuple of partitions of **G**, one for each agent (information sets)

G = {Matching Pennies (MP),
Prisoner's Dilemma (PD),
Coordination (Crd),
Battle of the Sexes (BoS)}



Example

- Suppose the randomly chosen game is MP
- Agent 1's information set is $I_{1,1}$
  - 1 knows it's MP or PD
  - 1 can infer **posterior probabilities** for each

$$\Pr[MP|I_{1,1}] = \frac{\Pr[MP]}{\Pr[MP]+\Pr[PD]} = \frac{0.3}{0.3+0.1} = \frac{3}{4}$$

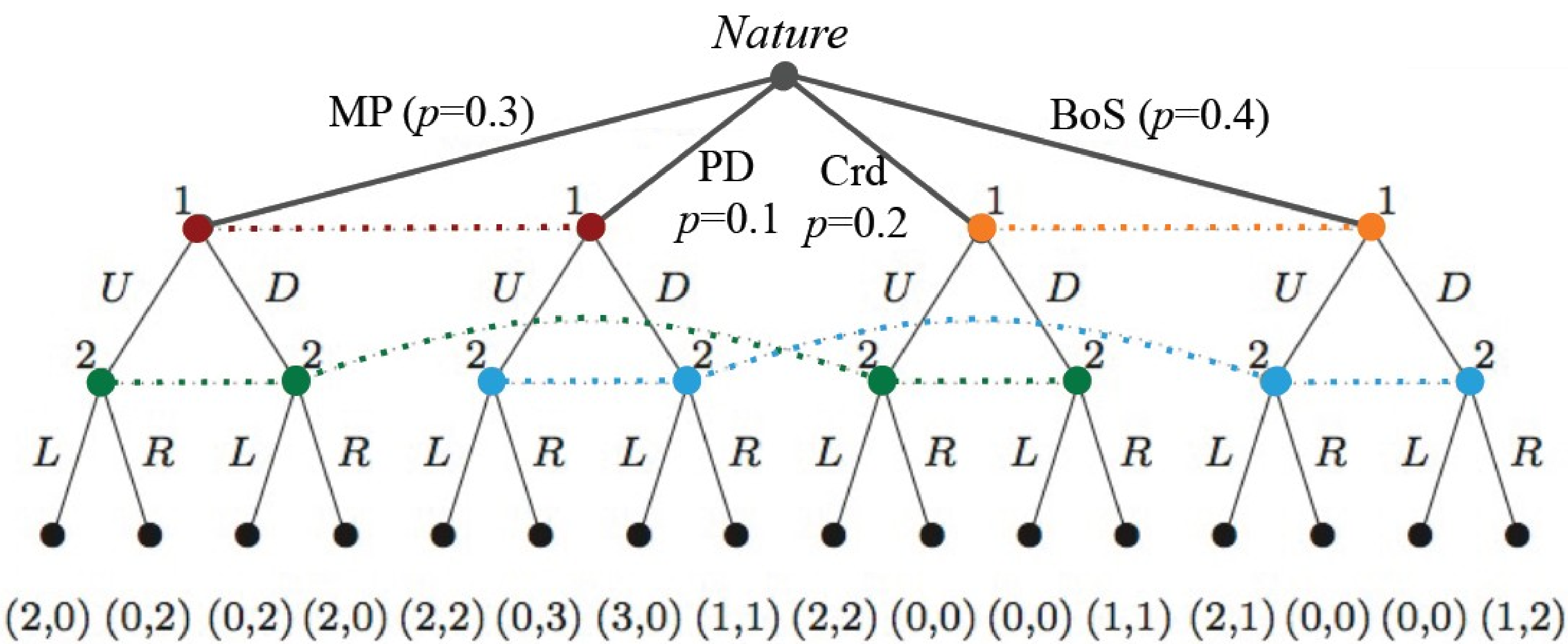
$$\Pr[PD|I_{1,1}] = \frac{\Pr[PD]}{\Pr[MP]+\Pr[PD]} = \frac{0.1}{0.3+0.1} = \frac{1}{4}$$

- Agent 2's information set is $I_{2,1}$

$$\Pr[MP|I_{2,1}] = \frac{\Pr[MP]}{\Pr[MP]+\Pr[CrD]} = \frac{0.3}{0.3+0.2} = \frac{3}{5}$$

$$\Pr[Crd|I_{2,1}] = \frac{\Pr[Crd]}{\Pr[MP]+\Pr[CrD]} = \frac{0.2}{0.3+0.2} = \frac{2}{5}$$

Another Interpretation: Extensive Form



Epistemic Types

- We can assume the only thing players are uncertain about is the game's utility function
- Thus we can define uncertainty directly over a game's utility function

Definition : a **Bayesian game** is a tuple (N, A, Θ, p, u) where:

N is a set of agents;

$A = A_1 \times \ldots \times A_n$, where A_i is the set of actions available to player i ;

$\Theta = \Theta_1 \times \ldots \times \Theta_n$, where Θ_i is the type space of player i ;

$p : \Theta \rightarrow [0, 1]$ is a common prior over types; and

$u = (u_1, \ldots, u_n)$, where $u_i : A \times \Theta \rightarrow \Re$ is the utility function for player i

- All this is common knowledge; each agent knows its own type

Types

An agent's **type** consists of all the information it has that isn't common knowledge, e.g.,

- The agent's actual payoff function
- The agent's beliefs about other agents' payoffs,
- The agent's beliefs about *their* beliefs about his own payoff
- Any other higher-order beliefs

Strategies

Similar to what we had in imperfect-information games:

- A pure strategy for player i maps each of i's types to an action
  - what i would play if i had that type
- A mixed strategy s_i is a probability distribution over pure strategies

$$s_i(a_i|\theta_j) = Pr[i \text{ plays action } a_j \mid i\text{'s type is } \theta_j]$$

- Many kinds of expected utility: *ex post, ex interim, and ex ante*
  - Depend on what we know about the players' types

Expected Utility

If we know every agent's type (i.e., the type profile $\boldsymbol{\theta}$)

agent i's ***ex post*** expected utility:

$$\boxed{EU_i(\mathbf{s},\boldsymbol{\theta})} = \sum_{\mathbf{a}} \Pr[\mathbf{a} \mid \mathbf{s},\boldsymbol{\theta}]\; u_i(\mathbf{a},\boldsymbol{\theta}) = \sum_{\mathbf{a}} \left(\prod_{j \in N} s_j(a_j \mid \theta_j) \right) u_i(\mathbf{a},\boldsymbol{\theta})$$

If we only know the common prior

agent i's ***ex ante***
expected utility:

$$EU_i(\mathbf{s}) = \sum_{\boldsymbol{\theta}} \Pr[\boldsymbol{\theta}]\, \boxed{EU_i(\mathbf{s},\boldsymbol{\theta})} = \sum_{\theta_i} \Pr[\theta_i]\; EU_i(\mathbf{s},\theta_i)$$

If we know the type θ_i of one agent i, but not the other agents' types

i's ***ex interim***
expected utility:

$$EU_i(\mathbf{s},\theta_i) = \sum_{\boldsymbol{\theta}_{-i}} \Pr[\boldsymbol{\theta}_{-i} \mid \theta_i]\, \boxed{EU_i(\mathbf{s},(\theta_i,\boldsymbol{\theta}_{-i}))}$$

Bayes-Nash Equilibrium

Given a strategy profile $\mathbf{s}_{-i}$, a **best response** for agent i is a strategy s_i such that

$$s_i \in \underset{s'_i}{\arg\max}(EU_i\,(s'_i\,, \mathbf{s}_{-i}))$$

Above, the set notation is because more than one strategy may produce the same expected utility

A **Bayes-Nash** equilibrium is a strategy profile **s** such that for every s_i in **s**, s_i is a best response to $\mathbf{s}_{-i}$

Just like the definition of a Nash equilibrium, except that we're using Bayesian-game strategies

Computing Bayes-Nash Equilibria

- The idea is to construct a payoff matrix for the entire Bayesian game, and find equilibria on that matrix

	$\theta_{2,1}$	$\theta_{2,2}$
$\theta_{1,1}$	MP ($p = 0.3$)	PD ($p = 0.1$)
$\theta_{1,2}$	Crd (p=0.2)	BoS ($p = 0.4$)

MP ($p = 0.3$)

	L	R
U	2, 0	0, 2
D	0, 2	2, 0

PD ($p = 0.1$)

	L	R
U	2, 2	0, 3
D	3, 0	1, 1

Crd (p=0.2)

	L	R
U	2, 2	0, 0
D	0, 0	1, 1

BoS ($p = 0.4$)

	L	R
U	2, 1	0, 0
D	0, 0	1, 2

Write each of the pure strategies as a list of actions, one for each type:

Agent 1's pure strategies:

- UU: U if type $\theta_{1,1}$, U if type $\theta_{1,2}$
- UD: U if type $\theta_{1,1}$, D if type $\theta_{1,2}$
- DU: D if type $\theta_{1,1}$, U if type $\theta_{1,2}$
- DD: D if type $\theta_{1,1}$, D if type $\theta_{1,2}$

Agent 2's pure strategies:

- LL: L if type $\theta_{2,1}$, L if type $\theta_{2,2}$
- LR: L if type $\theta_{2,1}$, R if type $\theta_{2,2}$
- RL: R if type $\theta_{2,1}$, L if type $\theta_{2,2}$
- RR: R if type $\theta_{2,1}$, R if type $\theta_{2,2}$

Computing Bayes-Nash Equilibria

Compute *ex ante* expected utility for each pure-strategy profile:

$$
\begin{aligned}
EU_2(UU,LL) &= \sum_{\theta} \Pr[\theta]\ u_2(U,L,\theta) \\
&= \Pr[\theta_{1,1},\theta_{2,1}]u_2(U,L,\theta_{1,1},\theta_{2,1}) \\
&\quad + \Pr[\theta_{1,1},\theta_{2,2}]u_2(U,L,\theta_{1,1},\theta_{2,2}) \\
&\quad + \Pr[\theta_{1,2},\theta_{2,1}]u_2(U,L,\theta_{1,2},\theta_{2,1}) \\
&\quad + \Pr[\theta_{1,2},\theta_{2,2}]u_2(U,L,\theta_{1,2},\theta_{2,2}) \\
&= 0.3(0) + 0.1(2) + 0.2(2) + 0.4(1) \\
&= 1
\end{aligned}
$$

$\theta_{2,1}$ | $\theta_{2,2}$

$\theta_{1,1}$

MP ($p = 0.3$)

	L	R
U	2, 0	0, 2
D	0, 2	2, 0

PD ($p = 0.1$)

	L	R
U	2, 2	0, 3
D	3, 0	1, 1

$\theta_{1,2}$

Crd (p=0.2)

	L	R
U	2, 2	0, 0
D	0, 0	1, 1

BoS ($p = 0.4$)

	L	R
U	2, 1	0, 0
D	0, 0	1, 2

Computing Bayes-Nash Equilibria

- Put all of the *ex ante* expected utilities into a payoff matrix

  e.g., $EU_2(UU,LL) = 1$

- Now we can compute best responses and Nash equilibria

$\theta_{2,1}$ / $\theta_{1,1}$: MP ($p = 0.3$)

	L	R
U	2, 0	0, 2
D	0, 2	2, 0

$\theta_{2,2}$ / $\theta_{1,1}$: PD ($p = 0.1$)

	L	R
U	2, 2	0, 3
D	3, 0	1, 1

$\theta_{2,1}$ / $\theta_{1,2}$: Crd (p=0.2)

	L	R
U	2, 2	0, 0
D	0, 0	1, 1

$\theta_{2,2}$ / $\theta_{1,2}$: BoS ($p = 0.4$)

	L	R
U	2, 1	0, 0
D	0, 0	1, 2

	LL	LR	RL	RR
UU	2, 1	1, 0.7	1, 1.2	0, 0.9
UD	0.8, 0.2	1, 1.1	0.4, 1	0.6, 1.9
DU	1.5, 1.4	0.5, 1.1	1.7, 0.4	0.7, 0.1
DD	0.3, 0.6	0.5, 1.5	1.1, 0.2	1.3, 1.1

Computing Bayes Nash Equilibria

- Suppose we learn agent 1's type is $\theta_{1,1}$

  Recompute the payoff matrix using the posterior probabilities

  $\Pr[MP|\theta_{1,1}] = ¾$, $\Pr[PD|\theta_{1,1}] = ¼$
- $u_2(UU,LL|\theta_{1,1}) = ¾\,(0) + ¼\,(2) = 0.5$

$\theta_{2,1}$ $\theta_{2,2}$

$\theta_{1,1}$

MP ($p = 0.3$)

	L	R
U	2, 0	0, 2
D	0, 2	2, 0

PD ($p = 0.1$)

	L	R
U	2, 2	0, 3
D	3, 0	1, 1

- ***Ex interim*** payoff matrix when agent 1's type is $\theta_{1,1}$
- Can't use this to compute equilibria, because $\theta_{1,1}$ isn't common knowledge

	LL	*LR*	*RL*	*RR*
UU	2, 0.5	1.5, 0.75	0.5, 2	0, 2.25
UD	2, 0.5	1.5, 0.75	0.5, 2	0, 2.25
DU	0.75, 1.5	0.25, 1.75	2.25, 0	1.75, 0.25
DD	0.75, 1.5	0.25, 1.75	2.25, 0	1.75, 0.25

and so on…

Acknowledgements

Slides are adapted from:

- Tutorial at IJCAI 2003 by Prof Peter Stone, University of Texas
- Game Theory lectures by Prof. Dana Nau, University of Maryland