Game Theory - Week 5

Mojtaba Tefagh

Sharif University of Technology

mtefagh@sharif.edu

December 12, 2022

Overview

- Repeated games
- Infinitely Repeated Games: Utility
- Stochastic Games
- Learning in Repeated Games
- Equilibria of Infinitely Repeated Games
- Discounted Repeated Games
- A Folk Theorem for Discounted Repeated Games

Repeated game

- Many (most?) interactions occur more than once:
  - Firms in a marketplace
  - Political alliances
  - Friends (favor exchange...)
  - Workers (team production...)

Repeated game

- OPEC: Oil Prices
  - 20$/bbl or less from 1930-1973 (2008 dollars)
  - 50$/bbl by 1976
  - 90$/bbl by 1982
  - 40$/bbl or less from 1986 to 2002
  - 100$/bbl by late 2008 ...

Repeated game

- Cooperative Behavior: Cartel is much like a repeated Prisoner's Dilemma

Repeated game

- Cooperative Behavior: Cartel is much like a repeated Prisoner's Dilemma
  - Need to easily observe each other's plays and react (quickly) to punish undesired behavior

Repeated game

- Cooperative Behavior: Cartel is much like a repeated Prisoner's Dilemma
  - Need to easily observe each other's plays and react (quickly) to punish undesired behavior
  - Need patient players who value the long run (wars don't help!)

Repeated game

- Cooperative Behavior: Cartel is much like a repeated Prisoner's Dilemma
  - Need to easily observe each other's plays and react (quickly) to punish undesired behavior
  - Need patient players who value the long run (wars don't help!)
  - Need a stable set of players and some stationarity helps
    - constantly changing sources of production can hurt, but growing demand can help ...

Infinitely Repeated Games

What is a player's utility for playing an infinitely repeated game?

- Can we write it in extensive form?

Infinitely Repeated Games

What is a player's utility for playing an infinitely repeated game?

- Can we write it in extensive form?
- The sum of payoffs in the stage game?

Definition

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

Definition

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

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

Discounted reward Definition

Definition

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

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

- Two equivalent interpretations of the discount factor:
  1. the agent cares more about his well-being in the near term than in the long term

Discounted reward Definition

Definition

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

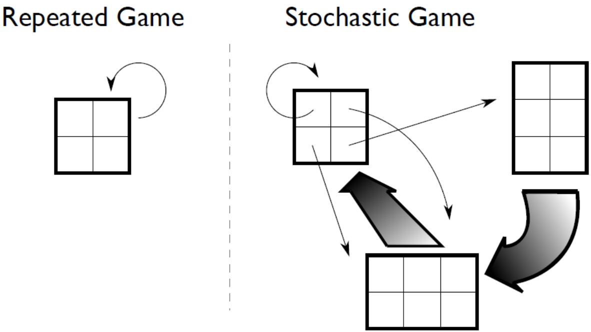
$$\sum_{j=1}^{\infty} \beta^j r_j$$

- Two equivalent interpretations of 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.

Stochastic Games- Introduction

- What if we didn't always repeat back to the same stage game?
- A stochastic game is a generalization of **repeated games**
  - agents repeatedly play games from a set of normal-form games
  - the game played at any iteration depends on the previous game played and on the actions taken by all agents in that game

Stochastic Games- Visualization



An informal visualization of the difference between repeated and stochastic games.

Stochastic Games- Formal Definition

Definition

A repeated games is a tuple (Q, N, A, P, R), where

- Q is a finite set of states,
- N is a finite set of n players,
- $A = A_1 \times \cdots \times A_n$, where A_i is a finite set of actions availableto player i,
- $P : Q \times A \times Q \to [0, 1]$ is the transition probability function; $P(q, \mathfrak{a}, \hat{q})$ is the probability of transitioning from state q to state $\hat{q}$ after joint action $\mathfrak{a}$, and
- $R = r_1, \ldots, r_n$, where $r_i : Q \times A \to \mathbb{R}$ is a real-valued payoff function for player i.

Stochastic Games- Remarks

- This definition assumes strategy space is the same in all games
  - otherwise just more notation
- Also generalizes MDP (Markov Decision Process)
  - i.e. MDP is a single-agent stochastic game

Stochastic Games- Analysis

Can do analysis as with repeated games.

- limit average reward
- future discount reward

Introduction

- We will cover two types of learning in repeated games.
  - Fictitious Play
  - No-regret Learning
- In general Learning in Game Theory is a rich subject with many facets we will not be covering.

Fictitious Play

- Initially proposed as a method for computing Nash equilibrium.
- Each player maintains explicit belief about the other players.
  - Initialize beliefs about the opponent's strategies.
  - Each turn:
    - Play a best response to the assessed strategy of the opponent.
    - Observe the opponent's actual play and update beliefs accordingly.

Fictitious Play

Formally

- Maintain counts of opponents actions
  - For every $\mathrm{a} \in \mathrm{A}$ let $\omega(\mathrm{a})$ be the number of times the opponent has player action a.
  - Can be initialized to non-zero starting values.
- Assess opponent's strategy using these counts:

$$\sigma(\mathrm{a}) = \frac{\omega(\mathrm{a})}{\sum_{\mathrm{a}' \in A} \omega(\mathrm{a}')}$$

- (pure strategy) best respond to this assessed strategy.
  - Break ties somehow.

Fictitious Play

Example using matching pennies

Round	1's action	2's action	1's beliefs	2's beliefs
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- Convergence

Theorem

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

Fictitious Play- Convergence

Theorem

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

Theorem

Each of the following are a 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* $2 \times \mathrm{n}$ *and has generic payoffs.*

No-regret Learning- Definitions

Definition (Regret)

The regret an agent experiences at time t for not having played s is $R^t(s) = max(\alpha^t(s) - \alpha^t, 0)$.

No-regret Learning- Definitions

Definition (Regret)

The regret an agent experiences at time t for not having played s is $R^t(s) = max(\alpha^t(s) - \alpha^t, 0)$.

Definition (No-regret learning rule)

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

No-regret Learning- Regret Matching

- Example learning rule that exhibits no regret: **Regret Matching**.

No-regret Learning- Regret Matching

- Example learning rule that exhibits no regret: **Regret Matching**.

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

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

  where $\sigma_i^{t+1}(s)$ is the probability that agent i plays pure strategy s at time $\mathrm{t}+1$.

- Converges to a correlated equilibrium for finite games.

Strategy Space

- What is a pure strategy in an infinitely-repeated game?

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.
  - **Trigger**: Start out cooperating. If the opponent ever defects, defect forever.

Nash Equilibria

- With an infinite number of pure strategies, 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!
- We can characterize a set of **payoffs** that are achievable under equilibrium, without having to enumerate the equilibria.

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 $\sum_{a \in A} \alpha_a = 1$

- feasible: 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 $\Longrightarrow$ enforceable

Part 1: Suppose r is not enforceable, i.e. $r_i < v_i$ for some i.

Folk Theorem (Part 1)

Payoff in Nash $\Longrightarrow$ 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 strategy of other players given the current history h.

Folk Theorem (Part 1)

Payoff in Nash $\Longrightarrow$ 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 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 $\Longrightarrow$ Nash

Part 2: Since r is a feasible payoff profile and the α's are rational, we can write it as $\mathrm{r_i} = \sum_{\mathrm{a} \in \mathrm{A}} (\frac{\beta_\alpha}{\gamma}) \mathrm{u_i(a)}$, where β_α and γ are non-negative integers and $\gamma = \sum_{\mathrm{a} \in \mathrm{A}} \beta_\alpha$.

Folk Theorem (Part 2)

Feasible and enforceable $\Longrightarrow$ Nash

Part 2: Since r is a feasible payoff profile and the α's are rational, we can write it as $\mathrm{r_i} = \sum_{\mathrm{a} \in \mathrm{A}} (\frac{\beta_\alpha}{\gamma}) \mathrm{u_i(a)}$, where β_α and γ are non-negative integers and $\gamma = \sum_{\mathrm{a} \in \mathrm{A}} \beta_\alpha$.
We're going to construct a strategy profile that will cycle through all outcomes $\mathrm{a} \in \mathrm{A}$ of G with cycles of length γ, each cycle repeating action a exactly β_α times. Let ($\mathrm{a^t}$) be such a sequence of outcomes.

Folk Theorem (Part 2)

Feasible and enforceable $\Longrightarrow$ Nash

Part 2: Since r is a feasible payoff profile and the α's are rational, we can write it as $r_i = \sum_{a \in A} (\frac{\beta_\alpha}{\gamma}) u_i(a)$, where β_α and γ are non-negative integers and $\gamma = \sum_{a \in A} \beta_\alpha$.
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_α 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.

Folk Theorem (Part 2)

Feasible and enforceable $\Longrightarrow$ Nash

Part 2: Since r is a feasible payoff profile and the α's are rational, we can write it as $r_i = \sum_{a \in A}(\frac{\beta_\alpha}{\gamma})u_i(a)$, where β_α and γ are non-negative integers and $\gamma = \sum_{a \in A} \beta_\alpha$.
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_α 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}), where (p_{-j}) is a solution to the minimization problem in the definition of v_j .

Folk Theorem (Part 2)

Feasible and enforceable $\Longrightarrow$ Nash

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 minmax payoff $v_j \leq r_j$, rendering the deviation unprofitable.

Discounted Repeated Games

- The future is uncertain, we are often motivated by what happens today
- Tradeoffs of today and the future are important in how I will behave today
- Will people punish me if I misbehave today?
  - Is it in their interest?
  - Do I care?

Discounted Repeated Games

- Stage game: $(\mathrm{N}, \mathrm{A}, \mathrm{u})$
- Discount factors: $\beta_1, \ldots, \beta_n, \beta_i \in [0, 1]$
- Assume a common discount factor for now: $\beta_i = \beta$ for all i
- Payoff from a play of actions $\mathrm{a}^1, \ldots, \mathrm{a}^{\mathrm{t}}, \cdots$:

$$\sum_{\mathrm{t}} \beta_{\mathrm{i}}^{\mathrm{t}} \mathrm{u}_{\mathrm{i}}(\mathrm{a}^{\mathrm{t}})$$

Histories

- Histories of length $t : H^t = \{h^t : h^t = (a^1, \ldots, a^t) \in A^t\}$
- All finite histories: $H = \bigcup_t H^t$
- A strategy: $s_i : H \to \Delta(A_i)$

Prisoners Dilemma

- $A_i = \{C, D\}$
- A history for three periods: $(C, C), (C, D), (D, D)$
- A strategy for period 4 would specify what a player would do after seeing $(C, C), (C, D), (D, D)$ played in the first three periods ...

Subgame Perfection

- Profile of strategies that are Nash in every subgame
- So, a Nash equilibrium following every possible history
- Repeatedly playing a Nash equilibrium of the stage game is always a subgame perfect equilibrium of the repeated game (Stop and check this!)

Repeated Prisoner's Dilemma

- Cooperate as long as everyone has in the past
- Both players defect forever after if anyone ever deviates: **Grim Trigger**

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

Repeated Prisoner's Dilemma

- Let's check that nobody wants to deviate if everyone has cooperated in the past:
- Cooperate: $3 + \beta 3 + \beta^2 3 + \beta^3 3 \cdots = \frac{3}{1-\beta}$
- Defect: $5 + \beta 1 + \beta^2 1 + \beta^3 1 \cdots = 5 + \beta \frac{1}{1-\beta}$

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

Repeated Prisoner's Dilemma

- Let's check that nobody wants to deviate if everyone has cooperated in the past:
- Cooperate: $3 + \beta 3 + \beta^2 3 + \beta^3 3 \cdots = \frac{3}{1-\beta}$
- Defect: $5 + \beta 1 + \beta^2 1 + \beta^3 1 \cdots = 5 + \beta \frac{1}{1-\beta}$
- Difference: $-2 + \beta 2 + \beta^2 2 + \beta^3 2 \cdots = \beta \frac{2}{1-\beta} - 2$
- Difference is nonnegative if $\beta \frac{2}{1-\beta} - 2 \geq 0$ or $\beta \geq (1-\beta)$, so $\beta \geq \frac{1}{2}$
- Need to care about tomorrow at least half as much as today!

Repeated Prisoner's Dilemma

- What if we make defection more attractive:

	C	D
C	3,3	0,10
D	10,0	1,1

Repeated Prisoner's Dilemma

- Let's check that nobody wants to deviate if everyone has cooperated in the past:
- Cooperate: $3 + \beta 3 + \beta^2 3 + \beta^3 3 \cdots = \frac{3}{1-\beta}$
- Defect: $10 + \beta 1 + \beta^2 1 + \beta^3 1 \cdots = 10 + \beta \frac{1}{1-\beta}$
- Difference: $-7 + \beta 2 + \beta^2 2 + \beta^3 2 \cdots = \beta \frac{2}{1-\beta} - 7$
- Difference is nonnegative if $\beta \frac{2}{1-\beta} - 7 \geq 0$ or $2\beta \geq 7(1-\beta)$, so $\beta \geq \frac{7}{9}$
- Need to care about tomorrow at least 7/9 as much as today!

Discounted Repeated Games

- Basic logic:
  - Play something with relatively high payoffs, and if anyone deviates
  - Punish by resorting to something that
    - has lower payoffs (at least for that player)
    - and is credible: it is an equilibrium in the subgame.

A (Simple) Folk Theorem for Discounted Repeated Games

- Consider a finite normal form game $\mathrm{G} = (\mathrm{N}, \mathrm{A}, \mathrm{u})$.
- Let $\mathrm{a} = (\mathrm{a}_1, ..., \mathrm{a}_\mathrm{n})$ be a Nash equilibrium of the stage game G
- If $\mathrm{a} = (\mathrm{a}_1, \ldots, \mathrm{a}_\mathrm{n}^{'})$ is such that $\mathrm{u_i(a)} > \mathrm{u_i(a)}$ for all i, then there exists a discount factor $\beta < 1$, such that if $\beta_i \geq \beta$ for all i, then there exists a subgame perfect equilibrium of the infinite repetition of G that has $\mathrm{a}^{'}$ played in every period on the equilibrium path.

A (Simple) Folk Theorem for Discounted Repeated Games

- Outline of the Proof:
- Play a' as long as everyone has in the past.
- If any player ever deviates, then play a forever after (Grim Trigger).
- Check that this is a subgame perfect equilibrium for high enough discount factors:

A (Simple) Folk Theorem for Discounted Repeated Games

- Check that this is a subgame perfect equilibrium for high enough discount factors:

A (Simple) Folk Theorem for Discounted Repeated Games

- Check that this is a subgame perfect equilibrium for high enough discount factors:
  - Playing a forever if anyone has deviated is a Nash equilibrium in any such subgame.

A (Simple) Folk Theorem for Discounted Repeated Games

- Check that this is a subgame perfect equilibrium for high enough discount factors:
  - Playing a forever if anyone has deviated is a Nash equilibrium in any such subgame.
  - Will someone gain by deviating from a if nobody has in the past?
  - Maximum gain from deviating is$M = \max_{i,a_i''} u_i(a_i'', a_i') - u_i(a')$

A (Simple) Folk Theorem for Discounted Repeated Games

- Check that this is a subgame perfect equilibrium for high enough discount factors:
  - Playing a forever if anyone has deviated is a Nash equilibrium in any such subgame.
  - Will someone gain by deviating from a if nobody has in the past?
  - Maximum gain from deviating is$M = \max_{i,a_i''} u_i(a_i'', a_i') - u_i(a')$
  - minimum per-period loss from future punishment is $m = \min_i u_i(a') - u_i(a)$ (why this?)

A (Simple) Folk Theorem for Discounted Repeated Games

- Check that this is a subgame perfect equilibrium for high enough discount factors:
  - Playing a forever if anyone has deviated is a Nash equilibrium in any such subgame.
  - Will someone gain by deviating from a if nobody has in the past?
  - Maximum gain from deviating is$M = \max_{i,a_i''} u_i(a_i'', a_i') - u_i(a')$
  - minimum per-period loss from future punishment is $m = \min_i u_i(a') - u_i(a)$ (why this?)
  - If deviate, then given other players' strategies, the maximum possible net gain is $M - m\frac{\beta_i}{1-\beta_i}$

A (Simple) Folk Theorem for Discounted Repeated Games

- Check that this is a subgame perfect equilibrium for high enough discount factors:
  - Playing a forever if anyone has deviated is a Nash equilibrium in any such subgame.
  - Will someone gain by deviating from a if nobody has in the past?
  - Maximum gain from deviating is$M = \max_{i,a_i^{''}} u_i(a_i^{''}, a_i^{'}) - u_i(a^{'})$
  - minimum per-period loss from future punishment is $m = \min_i u_i(a^{'}) - u_i(a)$ (why this?)
  - If deviate, then given other players' strategies, the maximum possible net gain is $M - m\frac{\beta_i}{1-\beta_i}$
  - Deviation is not beneficial if $\frac{M}{m} \leq \frac{\beta_i}{1-\beta_i}$ or $\beta_i \geq \frac{M}{M+m}$for all i.

Repeated Prisoner's Dilemma

- More complicated play: something to think about

	C	D
C	3,3	0,10
D	10,0	1,1

Repeated Prisoner's Dilemma

- Players can condition future play on past actions
- Brings in many(!) equilibria: Folk Theorems
- Need key ingredients
  - Some (fast enough) observation about how others behave
  - Sufficient value to the future (limit of the means - extreme value) or high enough discount factor