Week 5

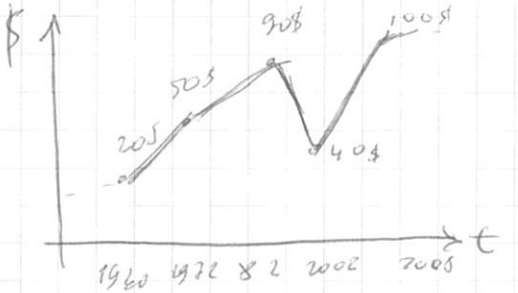
Repeated Games

players play repeatedly over time

many (most) interactions occur more than once

- firms
- political alliances
- friends
- workers

eg: OPEC prices



repeated Prisoners' dilemma →

- need to easily observe each other's plays and react quickly to punish undesired behavior
- patient players who value the long run (care about the future)
- stable set of players

Infinitely Repeated Games

Utility

⇓

infinite sequence of utilities

How to write it?

- can't do it in extensive form
- can't use sum — can be not bounded

① Can use average:

Given an infinite sequence of payoffs r_1, r_2 for player i,

avg. reward of i is

$$\lim_{k \to \infty} \sum_{j=1}^{k} \frac{r_j}{k}$$

② Discounted utility

payoffs are weighted (multiplied by some kind of a discount factor)

⇓

~~weight one factor one payoff to other, etc~~

different payoffs matter differently

Given an inf sequence of payoffs $r_1, r_2, \ldots$ for player i

and discount factor β $(0 \leq \beta < 1)$

i's <u>future discounted reward</u> is

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

(β can be ~~seen~~ as "an interest rate")

interpretation:

1. Agents care more about nearest well-being
2. With probability $1-\beta$ game may finish

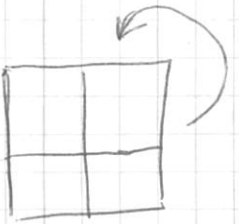
Stochastic Games

Generalization of repeated games.

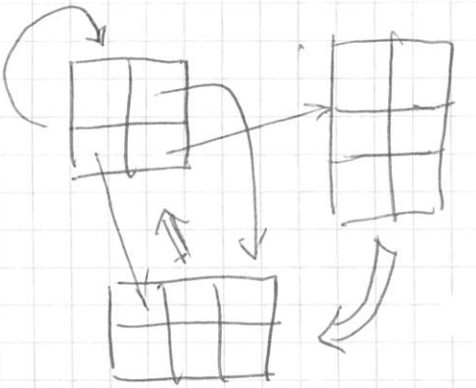
- play game from the same set of games
- game played at any iteration depends on the previous game, and actions taken

Informal Visualization

Repeated Game



Stochastic Game



Definition

A Stochastic Game is a tuple (Q, N, A, P, R)

where

- Q - finite set of states
- N - finite set of players
- $A = A_1 \times \ldots \times A_n$ - finite set of actions
  A_i - available to player i
- $P: Q \times A \times Q \to [0,1]$ - transition probability function
- $P(q, a, \hat{q})$ - probability of transition from state q to $\hat{q}$ after action a

(reward)

- $R = r_1, \ldots, r_n$ where $r_i: Q \times A \to \mathbb{R}$ payoff function

It generalizes MDP (Markov Decision Process)

MDP - single-agent stochastic game

Learning in Repeated Games

As one player tries to learn, so do others

- Fictitious play (Model-based Learning)

initially - method for computing NE

each player maintains explicit beliefs about the other players

- initialize beliefs about opponent's strategies
- each turn
  - play a BR to the assumed strategy
  - observe actual play and update beliefs accordingly

Formally

For $\forall a \in A$ let $w(a)$ be the number of times the opponent has played action a

$$\sigma(a) = \frac{w(a)}{\sum_{a' \in A} w(a')}$$ ← proportional to ~~stat~~ actual plays

- tie breaking

Consider Matching Pennies

↓
Won't converge, but there's a certain Balance.

but empirical ~~points~~ frequencies will converge to NE.

Theorem.

if the empirical distribution of each player's strategies converges in fictional play,

then it converges to a Nash Eq.

Theorem

Each of the following are sufficient condition 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 potential game
- the game is $2 \times n$ and has generic payoffs

don't care for now

- No-regret Learning (doesn't model, rather adapts)

the regret an agent experiences at time t for not having played $s \in S$

$$R^t(s) = \alpha^t - \alpha^t(s)$$

α^t: payoff he actually got at time t

$\alpha^t(s)$: payoff he would've got if had played s

No-regret rule

a learning rule exhibits no regret if for any pure strategy of the agent s it holds that

$$Pr([\liminf R^t(s)] \leq 0) = 1$$

(if player shows no regret)

Regret Matching

look at the regret you've experienced so far and pick a pure strategy in proportion to this regret

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

σ_i^{t+1} probability that agent i plays s at time $t+1$

So it converges to equilibrium for finite games.

Equilibria of Infinitely Repeated Games

Pure Strategy – action on every ~~action~~ stage

(given you remember everything! (history etc))

⇓

infinite set

Some strategies for PD

1. Tit-for-tat

   - start out cooperating
   - if opponent defects, defect next round
   - then go back to cooperation

2. Trigger

   - start out cooperating
   - if opponent ever defects, it will defect forever

As now we have infinite-size game (# of strategies $= \infty$)
we can't compute NE in usual way

But we can have infinite number of NE

Idea we can characterize a set of payoffs that are achievable under equilibrium, without having to enumerate the equilibria

(Which payoff vectors will achieve eq.)

$G = (N, A, u)$ n-player game

$r = (r_1 \dots r_n)$ ← utilities
average reward cases

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

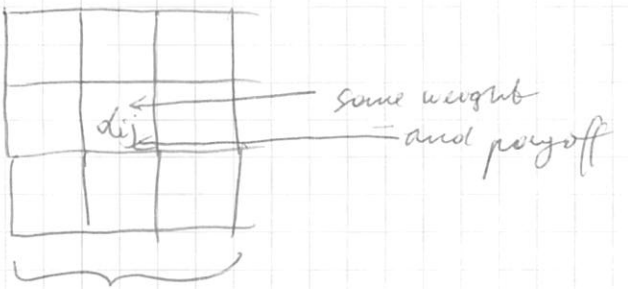
value i will get if others want to hurt i as much as they can

A payoff profile r is <u>enforcable</u> if $r_i \geq v_i$
↓ for everybody
↑ if everybody's payoff at least their minmax value

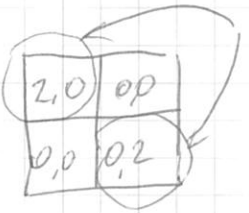
A payoff profile r is <u>feasible</u> if
there exist rational, non-negative values α_a
such that for all i we can express r_i as

$$\sum_{a \in A} \alpha_a u_i(a), \text{ with } \sum_{a \in A} \alpha_a = 1$$

if it's actually possible to achieve



Feasibility – T_i for player i
(it's possible to have this payoff)



(1, 1) - feasible

Folk Theorem

Consider any n-player G game and any payoff vector $(r_1, \ldots, r_n)$

① if r_i is the payoff in any NE of G, then for each player i
r_i is enforceable

(i.e. greater or equal to his/her minimax value)

② If r is both feasible and enforceable, then r is the payoff in some NE of G

So, feasibility and enforceability — things you need to find NE

⇓

As long as you meet these 2 conditions, you have a NE

Proof

① Payoff in NE ⇒ enforceable

pf Suppose r is not enforceable, i.e. $r_i < v_i$ for some i

Then consider a deviation to b_i, where b_i - his best response action

(since $r_i < v_i$, there must be a better strategy)

⇓

So i cannot receive $r_i < v_i$ in any NE

(contradiction)

p2 Feasible and Enforceable ⇒ NE

as α are rational, we ~~can divide it on~~ γ can write it as $\frac{\beta_a}{\gamma}$

$$r_i = \sum_{a \in A} \left(\frac{\beta_a}{\gamma}\right) u_i(a)$$

where β_a and γ – non-negative integers

and $\gamma = \sum_{a \in A} \beta_a$

$\gamma = 7$

	D	E	F
A	AA 2 DD	0	0
B	0	B 1 E	BB 2 FF
C	0	CC 2 EE	0

cycle for 1:

(AABBBCC)

for 2

(DDEEFFEE)

Strategies that cycle through all outcomes $a \in A$ with cycles of len γ, each cycle repeating action a exactly β_a times

Let (a^t) be such a sequence of outcomes

Let's define a trigger strategy s_i for i:

if nobody deviates then s_i plays a_i^t at period of t

However, if one of them deviates, (let it be j) then s_i will play $(P_{-j})_i$, where

(P_{-j}) - minimal value of v_j (minmax strategy)

First, observe that if everybody plays s_i, then i receives payoff of r_i (by construction of f...)

Second, it's a NE. If smb deviates at some point, then, forever after, pl. will receive his minmax payoff, rendering the deviation unprofitable

Discounted Repeated Games

- The future is uncertain, we often motivated by what happens today
- Will people punish me if I misbehave today?
  - is it in their interest?
  - Do I care? (about the future)

How people will react? (temporary gain?

Stage Game : (N, A, u)

Discount factor

$\beta_1 = \beta_2$, $\beta_i \in [0, 1]$

$\underbrace{a^1, \ldots, a^t}$

payoff of from a play

$\Rightarrow$ going to weak them by exp. decreasing function

$$\sum_t \beta_i^t u_i(a^t)$$

Histories of length t:

$$H^t = \{h^t : h^t = (a^1, \ldots, a^t) \in A^t\}$$

↑ what everybody did on period of time t

All finite histories:

$$H = \bigcup_t H^t$$

A strategy: $S_i : H \to \Delta(A_i)$

play based on history

Prisoners' Dilemma

$A_i = \{C, D\}$

History: $(C,C)\ (C,D),\ (D,D)$

A Strategy for period 4 would specify what a player would do after seeing

Subgame perfection

Subgame starts at a particular t' and contains everything what remains

Subgame perfection: take t', play NE, and in will be NE for ever on

No matter what is the history, if all stop and start playing NE, it will be a subgame perfection.

Prisoners' Dilemma

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

(given nobody has defected in past)

if cooperate:

$$3 + \beta 3 + \beta^2 3 + \ldots = \frac{3}{1-\beta}$$

if Defect:

$$5 + \beta 1 + \beta^2 1 \ldots = 5 + \beta \frac{1}{1-\beta}$$

5: deviate once

$\beta 1 + \beta^2 1 \ldots$: opponent deviates for ever

Difference:

$$\frac{\beta}{1-\beta} 2 - 2$$

↑ we want it to be positive

$$\beta \frac{2}{1-\beta} - 2$$

or $\beta \geq (1-\beta)$, $\beta \geq 1/2$.

So in order not to defect, people have to care about tomorrow at least as half as much as today!

So they sustain cooperation if $\beta \geq \frac{1}{2}$

	C	C	← Let's make
C	3,3	0,10	D more
C	10,0	1,1	attractive

Cooperate : $\frac{3}{1-\beta}$

Deviate : $10 + \beta \frac{1}{1-\beta}$

Difference : $\beta \frac{2}{1-\beta} - 7 \geq 0$, $\beta \geq \frac{7}{9}$

↗

trade off between punishment tomorrow ~~but~~ and payoff today

Basic Logic

- sustain by punishment

Folk Theorem for Discounted Repeated Games

Consider a finite normal form game $G = (N, A, u)$

Let $a = (a_1 \dots a_n)$ – a NE of G

if $a' = (a_1' \dots, a_n')$, such that

$$u_i(a') > u_i(a) \text{ for all } i$$

(there is a better strategy)

then there exists a discount factor $\beta < 1$ such that

$$\text{if } \beta_i > \beta \text{ for all } i$$

(take any game, find a NE, find a better strategy which you want to sustain, and make high enough discount factor to make it sustainable, as in prev. Prs. Dil examples)

then there exists a subgame perfect eq. of the inf. repetition of G that has a'

played in every period on the equilibrium path.

Maximum gain from deviating

$$M = \max_{i, a''} u_i(a_i'', a_{-i}') - u_i(a')$$

$$m = \min_i u_i(a') - u_i(a) \quad \text{- min per-period from future}$$

if deviate, the maximum possible punishment

net gain is $M = m \frac{\beta_i}{1-\beta_i}$

so $\beta_i \geq \frac{M}{M+m}$ for all i

<u>So</u>

- players can condition future play on past action

  ⇒ allows them to react on things

- brings in many equilibria

- Need:
  - be able to observe and react
  - sufficient value to the future
    high enough discount factor

Questions

② Consider a Repeated Game

- p – prob that game continues
- $1-p$ – game finishes

G. starts in $t_0 = 1$
if t odd, both play L
if t even, both play R

1 \ 2	L	R
L	3, 3	-1, 4
R	4, -1	1, 1

What is the expected total future payoff (from t_0) for each player

it's

$$3 + 1p + 3p^2 + 1p^3 + \dots$$

because

- in odd periods both get 1, even = 3
- (p^i) – probability that i^{th} period is reached

⑥ Consider Rock-Paper-Scissors game

	R	P	S
R	0,0	-1,1	1,-1
P	1,-1	0,0	-1,1
S	-1,1	1,-1	0,0

How many elements are there in H^2

↑ the set of histories of two plays of the Game?

1. H^1: has 9 elements

$$(RR)(RP)(RS)\dots(SP)(SS)$$

2. then H^2 has 9^2:

of the form (h^1, h^2)

↑ values of H^1

(7)

1 \ 2	M	H	max	min
M	3, 0	1, 2	2	2
H	2, 1	0, 3	3	
max	3	1		
min	1			

maximin of 1st is 1

maximin of 2nd is 2

thus

(0, 3)
(3, 0) not enforceable, as they
(2, 1) give lower payoff than maximin value