Lecture 8: Learning Models

1. Learning and Experimentation

n-player finite normal form game G (or normal form representation of extensive form game) - "stage game"

population of N agents, divided into n nonempty classes ($\Longrightarrow$ $N \geq n$)

game played repeatedly

In each round, one agent is chosen at random from each class (with full support). Chosen agents play the game in the respective round.

Learning models characterized by

finite set of all possible states $\Omega = \{\omega_1, \omega_2, \omega_3, ...\}$

definition of the state depends on the learning model, e.g.

each state ω determines for the stratgies used by the players in the m rounds, $m \geq 1$.

definition of state such that Ω is finite.

Transition matrix (or Markov chain) P: element p_{ij} gives probability, that state i is followed by state j in next period. Due to learning agents change strategies depending on the choices of all players during the last m rounds
$\Longrightarrow$

P depends on the type of learning (e.g. imitation learning - see below)

Experimentation (mistakes): On top of the learning process characterized by P, every player experiments (makes mistakes) with a common probability ϵ. In case of experimentation a player chooses any of his pure strategies, according to some pre-specified full support probability distribution.

$\lim_{\epsilon \to o} P(\epsilon) = P$

Notation

Take any Markov chain M. A positive probability path from state ω_1 to state ω_k is a finite chain of states $(\omega_1, \omega_2, \omega_3, ...\omega_k)$ such that $m_{i,i+1} > 0$ for all $i = 1, 2..k-1$.

A Markov chain M is irreducible, if there exists a positive probability path from any state i to any state j

Obviously, the Markov process $P(\epsilon)$ is irreducible for $\epsilon > 0$, even if $p_{i,j} = 0$ for some states i, j.

2. Long Run Predictions

Definition: A nonempty set of states, A, is absorbing with respect to a Markov chain M, iff
i) $m_{i,j} = 0$ for all $\omega_i \in A$, $\omega_j \notin A$
ii) for all $\omega_i \in A$ there exists a positive probability path from state ω_i to any other state state $\omega_k \in A$

"A is absorbing, if it is a minimal subset of states which, once entered, is never abandoned."

If M is irreducible, the only absorbing set is Ω.

From now on: absorbing set are always meant to be absorbing sets of Markov learning P without experimentation.

Definition: For any Markov chain M a probability distribution over states, $\mu \in \Delta(\Omega)$, is an invariant distribution if $\mu \cdot M = \mu$.

Result: For every absorbing set A of a Markov chain M there is an unique invariant distribution, which has support A.

$\Longrightarrow$ For any $\epsilon > 0$, there is an unique invariant distribution $\mu(P(\epsilon))$.

From now on: invariant distribution always refers to the Markov chain with experimentation, $P(\epsilon)$.

Result: If M is irreducible, then for any state ω_i, the probability $\mu(\omega_i)$ of the unique invariant distribution gives the portion of periods, in which the process is at ω_i in the long run.

Definition: The limit invariant distribution μ^* is given by

$$\mu^* = \lim_{\epsilon \to o} \mu(P(\epsilon))$$

Definition: A state ω is stochastically stable, if it is in the support of μ^*.

Because of previous result: A state, which is not stochastically stable, is in the long run observed only in a negligible portion of periods, if experimentation becomes rare.

"Stochastically stable states do not die out in the long run."

How to characterize stochastically stable states?

Result: Only states which belong to an absorbing set of the dynamics without experimentation can be stochastically stable.

Definition: Given two absorbing sets A and B, let $c(A, B) > 0$ (referred to as the transition cost from A to B) denote the minimal number of experiments necessary for a positive probability path from an element of A into an element of B.

Definition: Let A be an absorbing set.

i) The Radius of A is given by $R(A) = min\{c(A, B)|B$ is an absorbing set, $B \neq A\}$.

ii) The Coradius of A is given by $CR(A) = max\{c(B, A)|B$ is an absorbing set, $B \neq A\}$

"The Radius measures the minimal transition costs to get out of A into another absorbing set B. The Coradius measures the maximum transition costs to get from any absorbing set B to A."

Theorem: (Ellison 2000)

(i) If $R(A) \geq CR(A)$, all the states in A are stochastically stable.

(ii) If $R(A) > CR(A)$, the only stochastically stable states are those in A.

(iii) If the states in an absorbing set B are stochastically stable and $R(A) = c(B, A)$, the states in A are also stochastically stable.

Theorem allows for many games with stochastic learning models implying a Markov process a simple way to find stochastically stable states.

But: The conditions on radius and coradius are sufficient, but not always necessary - in general not all stochastically stable states are characterized.

3. Imitation Learning

Each agent is "programmed" to play a particular pure strategy, and at the beginning of each round some of the chosen agents (i.e. players) get the opportunity to revise their strategies.

Players imitate most successful other players

Whom to imitate?

symmetric game with a single population:

players choose the strategy the most successful players of the last round

players calculate the average payoff of all the strategies used during e.g the last 3 rounds, and chooses the most successful strategy

etc.

asymmetric game:

Player i imitates the most successful player or strategy of his own class

i imitates the strategy of the most successful player who has played "the same role" as i e.g. in the last 5 rounds.

i calculates the average payoff of all the strategies used by players of his own role during e.g the last 4 rounds, and imitates the most successful strategy

Game such that the set of players can be partitioned into types (e.g. buyers and sellers), where each player of same type has same payoff-function.

Players imitate most successful player of own type during e.g. the last 2 periods

Players pick the strategy which was on average the most successful for players of the own type during the last e.g. 4 rounds

etc.

Ties broken randomly with full support

Not each player can revise at the beginning of an round: random revision possibilities.

Independent revision probabilities: For each player the revision possibility is determined by an independent draw with an exogenous probability $0 < \rho < 1$.

Non-simultaneous learning: In each period, only one player is allowed to revise his strategy, and each player is equally likely to get the revision opportunity.

Non-simultaneous learning within types: In each period, only one player per type is (randomly) allowed to revise his strategy. Each agent is equally likely to get the revision possibility

A more general form of revision probabilities will be discussed in lecture 10.

The game, imitation learning, the specification of the revision probabilities, and random player picking determine P.

P determines the stochastically stable states.

3.1. Imitation Learning and Walrasian Behavior (Vega-Redondo 1997)

Cournot game

n identical firms

symmetric game with $N = n$ ($\Longrightarrow$ choosing players is trivial)

each firm i chooses simultaneously quantity q_i from a common finite set of possible quantities S.

$C(q)$ denotes the common, weakly convex, differentiable cost function.

Inverse demand function $\pi(Q)$, with $Q = \sum_N q_i$. $\pi(Q)$ strictly decreasing.

Walrasian quantity q_W defined by:

$$\pi(nq_W)q_W - C(q_W) \geq \pi(nq_W)q - C(q) \text{ for all } q > 0,\ q \neq q_W$$

If q_W exists, it is unique.

$Q > nq_W,\ q_j > q_W \implies \pi(Q) < C'(q_W) \leq C'(q_j) \implies$
$\pi(Q)q_W - C(q_W) > \pi(Q)q_j - C(q_j)$

$Q < nq_W,\ q_j < q_W \implies \pi(Q) > C'(q_W) \geq C'(q_j) \implies$
$\pi(Q)q_W - C(q_W) > \pi(Q)q_j - C(q_j)$

To ensure existence of the Walrasian quantity

$$\pi(0) - C'(0) > 0,\quad \lim_{Q\to\infty} \pi(Q) - C'(q) < 0$$

$q_W \in S$.

game played repeatedly

if a firms can revise its quantity, it imitates the firm with the highest profit in the last round (ties broken randomly).

$\Longrightarrow$ state is a distribution of feasible quantities over players

independent revision probabilities $0 < \rho < 1$ such that a firm gets revision opportunity.

$\Longrightarrow$ Markov process P

experimentation with probability $\epsilon > 0$. In case of experimentation, firm chooses quantity at random, with full support probability distribution.

$\Longrightarrow$ $p_{ij}(\epsilon) > 0$ for all states i, j

ω_W: state where all firms choose q_W

Obviously, $\{\omega_W\}$ is an absorbing set.

Proposition:

i) $R(\{\omega_W\}) > 1$

ii) $CR(\{\omega_W\}) = 1$.

Proof: i) Take state ω_W, and assume that only firm i experiments and chooses $q_i < q_W$. This implies that, $Q_{t+1} < nq_W$ and $\pi(Q_{t+1}) > C'(q_W) \geq C'(q_i)$. Hence $\pi(Q_{t+1})q_W - C(q_W) > \pi(Q_{t+1})q_i - C(q_i)$ - without further experimentation, no other firm will follow $i's$ lead to switch away from q_W. Symmetric case for $q_i > q_W$.

ii) $CR(\{\omega_W\}) > 0$, since any state where all firms choose same q is absorbing.

Take any absorbing set B. Because of the proof of i) B must consist of states where more than one firm does not choose q_W. Assume that one of these firms, denoted by i, experiments in period t and ends up at q_W. The overall quantity in period t is then given by $Q_t = \sum_{j \in N \setminus i} q_j + q_W$. The set of firms with highest profits in t is denoted by D.

4 cases

1) $i \in D$. In this case there exists strictly positive probability that all firms switch to q_W in the next period $t+1$ - no additional experimentation necessary for get to $\{\omega_W\}$.

2) For all $j \in D$ it holds that $q_j > q_W$. There exists strictly positive probability that in the next period all firms but i can revise. All of them will switch to any $q_j > q_W$, implying that $Q_{t+1} > nq_W$ and $\pi(Q_{t+1}) < C'(q_W) \leq C'(q_j)$. Hence, $\pi(Q_{t+1})q_W - C(q_W) > \pi(Q_{t+1})q_j - C(q_j)$ - in period $t+1$ all firms are worse off then i. Hence, there exists a strictly positive probability that all firms switch to q_W in period $t+2$ - no additional experimentation necessary for get to $\{\omega_W\}$.

3) For all $j \in D$ it holds that $q_j < q_W$. Similar to the previous case.

4) $\exists j, j' \in D$ with $q_j < q_W < q_{j'}$. There exists a strictly positve probability that every firm but i imitates $j' \Longrightarrow Q_{t+1} > nq_W$. And by the same argument as in case 2) there exists a strictly positive probaility that everybody chooses q_W in period $t+2$.

Hence, in each case 1 experimentation is necessary to get from any state and hence from any absorbing set to $\{\omega_W\}$.■

Theorem: The only stochastically stable state is where every firm chooses the Walrasian quantity.

Proof: follows immediately from the previous proposition and Ellison's result.

4. Best Reply with One-Period Memory

s_{-i}: strategy combination of all players but i

set of best reply strategies of player i against s_{-i}:

$$BR_i(s_{-i}) = \{s_i \in S_i : u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}) \text{ for all } s_i' \in S_i\}$$

with u denoting the payoff.

Best reply learning with one period memory: If player i is allowed to revise his strategy in period t, then he plays a (possibly) mixed strategy with support $BR_i(s_{-i}^{t-1})$.

$\Longrightarrow$ Markov process with $m = 1$

Revision possibilities

always: original formulation of best reply model. Problem: cycles

independent inertia

etc.

4.1. Cournot Oligopoly and Best Reply (Huck et al 1999)

n identical firms

symmetric game, $N = n$

each firm i chooses simulataneously quantity q_i from a common finite set of possible quantities S, with grid $\delta > 0$. $q_{max} = 100$.

$C(q_i) = q_i$

Inverse demand function $\pi(Q) = max\{101 - Q, 0\}$, with $Q = \sum_{i \in N} q_i$.

For continuous strategy set $[0, 100]$

Best reply: $BR_i(Q_{-i}) = 50 - \frac{Q_{-i}}{2}$, with $Q_{-i} = \sum_{j \in N \setminus i} q_j$

Cournot equilibrium: $q_c = \frac{100}{n-1}$

$q_c \in S$

best reply learning with independent inertia

Result: The only absorbing set of this learning process consists of the state ω_c, where all firms choose the Cournot equilibrium quantity.

Hence, ω_c is the only stochastically stable state - "global convergence" to ω_c.

5. Best Reply with Finite Memory and Random Sampling (Young 1993)

n-player finite normal form game G (or normal form representation of extensive form game) - "stage game"

n players divided into T types, where every player belonging to same type has same payoff-function. $|\tau|$ number of players of type τ.

population of N agents, distributed over types, and for each type τ at least $|\tau|$ agents.

in each period t, $|\tau|$ per type τ chosen at random, and the chosen agents play the game. random choice with full support.

each player (i.e. chosen agent) chooses his strategy the following way:

a) full support random choice of k rounds of the last m repetitions ($k < m$). These k rounds are observed.

observed repetitions chosen independently for each agent.

before round m, random "hypothetical" history.

b) each player plays best response against the empirical distribution of the k rounds he observes. Identity of players in the observed k rounds is ignored, and ties are broken randomly (with full support).

$\Longrightarrow$ Markov process P with states defined by the strategy profiles played in the last m rounds.

on top of best reply, experimentation - with probability ϵ each player chooses randomly a strategy according to a full support probability distribution

$\Longrightarrow$ for $\epsilon > 0$, $P(\epsilon)$ irreducible - from each state there exists a positive probability path to each other state.

$\Longrightarrow$ Ellison's Theorem applicable to this learning model

5.1. The Evolution of Conventions (Young 1993)

symmetric coordination game

		U	D	
U	0	0	b	a
D	a	b	1	1

with $a < 0$, $b < 1$ $\Longrightarrow$ two strict NE, (U, U) and (D, D)

only one type of players, $N \geq 2$

q: probability, that the other player plays U

Best response

$$BR(q) = \begin{array}{c} U, \text{ if } q > \frac{(1-b)}{(1-a-b)} \\ \text{anything, if } q = \frac{(1-b)}{(1-a-b)} \\ D, \text{ if } q < \frac{(1-b)}{(1-a-b)} \end{array}$$

$|U|$: the number of times a player has chosen U during last m periods $(0 \leq |U| \leq 2m)$

Obviously, $2m - |U| = |D|$

$\frac{|U|}{2k} < \frac{(1-b)}{(1-a-b)} \implies$ for every possible sampling, players will choose D in next round, because even if the observed k rounds are such that all choices of U are observed, best response against observation is D.

$\frac{|D|}{2k} < 1 - \frac{(1-b)}{(1-a-b)} \implies$ for every possible sampling, players will choose U in next round, because even if the observed k rounds are such that all choices of D are observed, best response against observation is U.

In between cases: strictly positive probability of choices of D as well as of U.

$\implies$ 2 absorbing sets

singleton containing state ω_D, where the players have chosen always D during the last m rounds.

singleton with state ω_U, here the players have chosen always U during the last m rounds.

Only two absorbing sets A and $B \implies R(A) = CR(B)$

$$R(\omega_D) = min\ Z \in \mathbb{N} : Z > 2k\frac{(1-b)}{(1-a-b)}$$

$$R(\omega_U) = min\ Z \in \mathbb{N} : Z > 2k\left(1 - \frac{(1-b)}{(1-a-b)}\right)$$

3 Cases:

a) Only ω_U is stochastically stable; If and only if

i) $R(\omega_D) < R(\omega_U) \iff 2k\frac{(1-b)}{(1-a-b)} < 2k\left(1 - \frac{(1-b)}{(1-a-b)}\right)$
$\iff 1 < -a + b$

ii) $\exists\ Z \in \mathbb{N}$: $2k\frac{(1-b)}{(1-a-b)} < Z < 2k\left(1 - \frac{(1-b)}{(1-a-b)}\right)$

If i) holds, (U, U) is the risk-dominant equilibrium.

ii) because of integer problems: For given a, b fulfilling i), the ii) is fulfilled for high enough k.

b) Only ω_D is stochastically stable; If and only if

i) $R(\omega_U) < R(\omega_D) \iff 2k\left(1 - \frac{(1-b)}{(1-a-b)}\right) < 2k\frac{(1-b)}{(1-a-b)} \iff 1 > -a+b$

ii) $\exists\ Z \in \mathbb{N}$: $2k\left(1 - \frac{(1-b)}{(1-a-b)}\right) < Z < 2k\frac{(1-b)}{(1-a-b)}$

In this case pareto-dominant equilibrium (D,D) is also risk-dominant.

c) ω_D and ω_U are stochastically stable; iff either:

i) $1 = -a+b$

ii) there exists no $Z \in \mathbb{N}$ such that

$$\min\left\{2k\frac{(1-b)}{(1-a-b)}, 2k\left(1 - \frac{(1-b)}{(1-a-b)}\right)\right\}$$
$$< \quad Z < \max\left\{2k\frac{(1-b)}{(1-a-b)}, 2k\left(1 - \frac{(1-b)}{(1-a-b)}\right)\right\}$$

Hence, in this game best reply learning leads to coordination on the risk dominant equilibrium, when integer problems are disregarded.