Evolution & Learning in Games
Econ 243B

Jean-Paul Carvalho

Lecture 5:
Learning Protocols

Learning Protocol

The learning (or evolutionary) protocol at the individual level determines the *change* in the evolutionary dynamic at the population level at any point in time, i.e. the differential equation.

- Solve the initial value problem whenever possible, to get the full trajectory for the population state at any time.

Otherwise:

- Analyze asymptotics using Lyapunov functions.
- Local behavior using linear approximation.

Boundedly Rational Learning

- **Learning protocols** considered here exhibit two forms of *bounded rationality*:

  1. ***Inertia***: agents do not consider revising their strategies at every point in time.

  2. ***Myopia***: agents' choices depend on current behavior and payoffs, not on beliefs about the future trajectory of play.

- In large populations, interactions tend to be *anonymous*—considerations such as *reputation* and other sources of forward-looking behavior in repeated games are far less relevant here.

- Each agent is 'small', so its behavior has very little affect on the trajectory of play.

Learning Protocols

- Let $F : X \to \mathbb{R}^n$ be a population game with pure strategy sets $(S^1, \ldots S^P)$, one for each population $p \in \mathcal{P}$.
- Each population is large but <u>finite</u>.

Definition. A learning protocol ρ^p is a map $\rho^p : \mathbb{R}^{n^p} \times X^p \to \mathbb{R}_+^{n^p \times n^p}$. The scalar $\rho_{ij}^p(\pi^p, x^p)$ is called the *conditional switch rate* from strategy $i \in S^p$ to strategy $j \in S^p$ given payoff vector π^p and population state x^p.

- A population game F, a population size N and a revision protocol $\rho = (\rho^1, ...\rho^P)$ defines a continuous-time evolutionary process on the state space $\mathcal{X}^N$.
- In the single population case, $\mathcal{X}^N = X \bigcap \frac{1}{N}\mathbb{Z}^n = \{x \in X : Nx \in \mathbb{Z}^n\}$, i.e. a discrete grid embedded in the original state space X.

Revision Opportunities

- Let each agent be equipped with a (Poisson) alarm clock. When the alarm clock rings the agent has one opportunity to revise its strategy.
  - The time between rings is independent across agents and follows a rate 1 exponential distribution.
  - Then the number of rings during time interval $[0, t]$ follows a Poisson distribution with mean t.
  - Strategy choices are made independently of the clocks' rings.
- An i player in population p faced with a revision opportunity switches to strategy j with prob. $\rho^p_{ij}(\pi^p, x^p)$, which is a function only of the current payoff vector π^p and the current strategy distribution x^p in population p (alone).

Mean Dynamic

- When each agent uses such a revision protocol, the state x follows a stochastic process $\{X_t^N\}$ on the state space $\mathcal{X}^N$.
- We shall now derive a <u>deterministic</u> process—the **mean dynamic**—which describes the expected motion of $\{X_t^N\}$.
- We shall show that the mean dynamic is a good approximation to $\{X_t^N\}$ when the time horizon is finite and the population size large.

Mean Dynamic

- Recall that the mean number of rings of one agent's alarm clock during time interval $[0, t]$ is t.

- Given the current state x, the number of revision opportunities received by agents currently playing i over the next dt time units (dt small) is approximately:

$$Nx_i dt.$$

  This is an approximation because x_i may change between time 0 and dt, but this change is likely to be small if dt is small.

Mean Dynamic

For the moment, drop population (p) notation:

- Therefore, the number of switches $i \to j$ in the next dt time units is approximately:

$$Nx_i \rho_{ij} dt.$$

- Hence, the expected change in the use of strategy i is:

$$N\left(\sum_{j \in S} x_j \rho_{ji} - x_i \sum_{j \in S} \rho_{ij}\right) dt,$$

i.e. the ‘inflow’ minus the ‘outflow’.

Mean Dynamic

- Dividing by N (to get the expected change in the proportion of agents) and eliminating the time differential dt yields:

$$\dot{x}_i = \sum_{j \in S} x_j \rho_{ji} - x_i \sum_{j \in S} \rho_{ij},$$

  the **mean dynamic** corresponding to the revision protocol ρ and population game F.

- For the general multipopulation case we write:

$$\dot{x}_i^p = \sum_{j \in S^p} x_j^p \rho_{ji}^p - x_i^p \sum_{j \in S^p} \rho_{ij}^p.$$

Classes of Learning Protocols

1. **Imitative Protocols**: a strategy must be present in the population to be learned, e.g. imitation of more successful strategies.

2. **Direct Protocols**: full knowledge of strategy set, e.g. myopic best responses.

We shall now consider examples of the two, focussing on the single population case.

Imitative Protocols & Dynamics

- Imitative protocols take the following form:

$$\rho_{ij}(\pi, x) = x_j r_{ij}(\pi, x).$$

- That is, when faced with a revision opportunity an i player is randomly exposed to an opponent and observes his strategy, say j. He then switches to j with probability proportional to r_{ij}.

- Again, for any agent to switch to strategy j, x_j must be greater than zero.

Examples

(a) Reinforcement Learning

- A revising agent abandons current strategy with prob. linearly decreasing in current payoff, then imitates the strategy of a randomly chosen opponent:

$$\rho_{ij}(\pi, x) = (K - \pi_i)x_j,$$

where K can be thought of as an *aspiration level*; K must be sufficiently large for switch rates to always be positive.

Minimal information: agents need not even know they are in a game.

Examples

- The mean dynamic is:

$$\begin{aligned}
\dot{x}_i &= \sum_{j \in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j \in S} \rho_{ij}\big(F(x), x\big) \\
&= \sum_{j \in S} x_j \big(K - F_j(x)\big) x_i - x_i \big(K - F_i(x)\big) \\
&= x_i \left(K - \sum_{j \in S} x_j F_j(x) - K + F_i(x) \right) \\
&\equiv x_i \big(F_i(x) - \overline{F}(x)\big).
\end{aligned}$$

This is the **replicator dynamic**.

Examples

(b) Imitation of Success

- A revising agent is exposed to a randomly chosen opponent and imitates its strategy, say j with prob. linearly increasing in the current payoff to strategy j:

$$\rho_{ij}(\pi, x) = x_j(\pi_j - K),$$

where again K can be thought of as an *aspiration level*; K must be smaller than any feasible payoff for switch rates to always be positive.

Examples

- The mean dynamic is:

$$
\begin{aligned}
\dot{x}_i &= \sum_{j\in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j\in S} \rho_{ij}\big(F(x), x\big) \\
&= \sum_{j\in S} x_j x_i \big(F_i(x) - K\big) - x_i \sum_{j\in S} x_j \big(F_j(x) - K\big) \\
&= x_i\big(F_i(x) - K\big) + x_i K - x_i \sum_{j\in S} x_j F_j(x) \\
&\equiv x_i\big(F_i(x) - \overline{F}(x)\big).
\end{aligned}
$$

Once again, the replicator dynamic.

Examples

(c) Pairwise Proportional Imitation

- An agent is paired at random with an opponent, and imitates the opponent only if the opponent's payoff is higher than its own, doing so with prob. proportional to the payoff difference:

$$\rho_{ij}(\pi, x) = x_j \left[\pi_j - \pi_i\right]_+.$$

Examples

- This generates the following mean dynamic:

$$
\begin{aligned}
\dot{x}_i &= \sum_{j\in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j \in S} \rho_{ij}\big(F(x), x\big) \\
&= \sum_{j\in S} x_j x_i \big[F_i(x) - F_j(x)\big]_+ - x_i \sum_{j\in S} x_j \big[F_j(x) - F_i(x)\big]_+ \\
&= x_i \sum_{j\in S} x_j \big[F_i(x) - F_j(x)\big] \\
&= x_i \Big(F_i(x) - \sum_{j\in S} x_j F_j(x)\Big) \\
&\equiv x_i \big(F_i(x) - \overline{F}(x)\big).
\end{aligned}
$$

  Yet again, the replicator dynamic.

- There are of course imitative protocols that do not generate the replicator dynamic.

Direct Protocols & Dynamics

- Direct Protocols are ones in which agents can switch to a strategy directly, without having to observe the strategy of another player.
- This requires <u>awareness</u> of the full set of available strategies S.

Examples

(a) Best Response

- The best response protocol is given by:

$$\rho_{ij}(\pi, x) = \sigma(\pi) = M(\pi) \equiv \operatorname*{arg\,max}_{y \in X} y'\pi,$$

where $M(\pi)$ is the set of mixed strategies that place mass only on pure strategies optimal under payoff vector π.

Examples

- The mean dynamic is:

$$\begin{aligned}
\dot{x}_i &= \sum_{j\in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j\in S} \rho_{ij}\big(F(x), x\big) \\
&= \sum_{j\in S} x_j \sigma_i\big(F(x)\big) - x_i \sum_{j\in S} \sigma_j\big(F(x)\big) \\
&= \sigma_i\big(F(x)\big) \sum_{j\in S} x_j - x_i \\
&= \sigma_i\big(F(x)\big) - x_i.
\end{aligned}$$

- This can be expressed as:

$$\dot{x} = \sigma\big(F(x)\big) - x.$$

Examples

(b) Logit Choice

- Suppose choices are made according to the logit choice protocol:

$$\rho_{ij}(\pi) = \frac{exp(\eta^{-1}\pi_j)}{\sum_{k\in S} exp(\eta^{-1}\pi_k)},$$

where η is the noise level.

- As $\eta \to \infty$, the choice probabilities tend to be uniform.
- As $\eta \to 0$, choices tend to best responses.

Examples

- The mean dynamic is:

$$
\begin{aligned}
\dot{x}_i &= \sum_{j \in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j \in S} \rho_{ij}\big(F(x), x\big) \\
&= \sum_{j \in S} x_j \frac{exp(\eta^{-1} F_i(x))}{\sum_{k \in S} exp(\eta^{-1} F_k(x))} - x_i \sum_{j \in S} \frac{exp(\eta^{-1} F_j(x))}{\sum_{k \in S} exp(\eta^{-1} F_k(x))} \\
&= \frac{exp(\eta^{-1} F_i(x))}{\sum_{k \in S} exp(\eta^{-1} F_k(x))} - x_i.
\end{aligned}
$$

This is the **logit dynamic** with noise level η.

Examples

(c) Comparison to the Average Payoff

- A revising agent switches to j if its payoff is above the average payoff; in this case, switching to it with prob. proportional to that strategy's excess payoff (payoff above average):

$$\rho_{ij}(\pi) = \left[\pi_j - \sum_{k \in S} x_k \pi_k\right]_+ .$$

Examples

- The mean dynamic is:

$$
\begin{aligned}
\dot{x}_i &= \sum_{j\in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j\in S} \rho_{ij}\big(F(x), x\big) \\
&= \sum_{j\in S} x_j \Big[F_i(x) - \sum_{k\in S} x_k F_k(x)\Big]_+ - x_i \sum_{j\in S} \Big[F_j(x) - \sum_{k\in S} x_k F_k(x)\Big]_+ \\
&= \Big[F_i(x) - \sum_{k\in S} x_k F_k(x)\Big]_+ - x_i \sum_{j\in S} \Big[F_j(x) - \sum_{k\in S} x_k F_k(x)_k\Big]_+,
\end{aligned}
$$

which is called the **Brown–von Neumann–Nash dynamic**.

Examples

(d) Pairwise Comparisons

- A revising agent switches to strategy j if its payoff is higher than the payoff to i; in this case, switches to it with probability proportional to the difference between the two payoffs:

$$\rho_{ij}(\pi) = \left[\pi_j - \pi_i\right]_+ .$$

Examples

- The mean dynamic is:

$$\begin{aligned}\dot{x}_i &= \sum_{j\in S} x_j \rho_{ji}\big(F(x), x\big) - x_i \sum_{j\in S} \rho_{ij}\big(F(x), x\big) \\ &= \sum_{j\in S} x_j \big[F_i(x) - F_j(x)\big]_+ - x_i \sum_{j\in S} \big[F_j(x) - F_i(x)\big]_+\end{aligned}$$

which is called the **Smith dynamic**.

Evolutionary Dynamics—Some General results

Definition. Fix an open set $O \subseteq \mathbb{R}^n$. The function $f : O \to \mathbb{R}^m$ is *Lipschitz continuous* if there exists a scalar K such that:

$$|f(x) - f(y)| \leq K|x - y| \qquad \text{for all } x, y \in O.$$

For example:

- The continuous but not (everywhere) differentiable function $f(x) = |x|$ is Lipschitz continuous.
- The continuous but not (everywhere) differentiable function $f(x) = 0$ for $x < 0$ and $f(x) = \sqrt{x}$ for $x \geq 0$ is <u>not</u> Lipschitz continuous.
  - Its right-hand slope at $x = 0$ is $+\infty$ and hence no K can be found that meets the above condition.

Evolutionary Dynamics—Some General results

Theorem. If F is Lipschitz continuous, then for any evolutionary dynamic and from any initial condition $\xi \in X$, there exists at least one trajectory of the dynamic $\{x_t\}_{t\geq 0}$ with $x_0 = \xi$. In addition, for every trajectory, $x_t \in X$ for all $t > 0$. That is, the set of trajectories exhibits *existence* and *forward invariance*.

- However, this does not guarantee uniqueness.
- Note: uniqueness requires that for each initial state $\xi \in X$, there exists exactly one solution $\{x_t\}_{t\geq 0}$ to the system of differential equations with $x_0 = \xi$.

Evolutionary Dynamics—Some General results

- Let the dynamic be characterized by the vector field $V_F : X \to \mathbb{R}^n$, where $\dot{x} = V_F(x)$.

Theorem. Suppose that V_F is Lipschitz continuous and $V_F(x)$ is contained in the tangent cone $TX(x)$, that is the set of directions of motion that do not point out of X. Then the set of trajectories of the evolutionary dynamic exhibits existence, forward invariance, *uniqueness* and *Lipschitz continuity*.

- The latter states that for each t, $x_t = x_t(\xi)$ is a Lipschitz continuous function of ξ.

Stochastic Approximation

Recall that a population game F, a revision protocol ρ, and a population size N define a Markov process $\{X_t^N\}$ on the state space $\mathcal{X}^N$.

Theorem Let $\{\{X_t^N\}\}_{N=N_0}^{\infty}$ be a sequence of continuous-time Markov processes on the state space $\mathcal{X}^N$.

Suppose that V_F is Lipschitz continuous. Let the initial conditions $X_0^N = x_0^N$ converge to state $x_0 \in X$ as $N \to \infty$, and let $\{x_t\}_{t \geq 0}$ be the solution to the mean dynamic starting from x_0.

Then for all $T < \infty$ and $\varepsilon > 0$,

$$\lim_{N \to \infty} \mathbb{P}\left(\sup_{t \in [0,T]} |X_t^N - x_t| < \varepsilon \right) = 1.$$

Conditions

- Lipschitz continuity guarantees unique solutions to the mean dynamic, but does not apply to the best response dynamic.
- Still, deterministic approximation results for the best response dynamic are available.
- More importantly, this is a finite-horizon result.