Game Theory & Learning: LECTURE 10 (CONT'D)

Theorem *Suppose* $U^{0\infty}(S) = \{\bar{x}\}$.

$$\|x(t) - \bar{x}\| \to 0$$

iff $x(t)$ *is consistent with sophisticated learning.*

Stochastic Learning Processes

We now allow the players to experiment as we will now assume that each user may not know his own pay-off function. See Freudenberg & Kreps (1988).

Game consists of alternations among

- *Exploration*: Every strategy is experimented with equi-probability.
- *Exploitation*: Good strategies —based on exploration— are played.

At each date t, player n conducts an experiment with probability ϵ_{nt} in an attempt to learn its best play.

1. **Independence:** Decision to experiment is independent of other players' decisions.
2. **Rare:** $\epsilon_{nt} \to 0$ as $t \to \infty$.
3. **Infinitely Often:** $\sum_t \epsilon_{nt} = \infty$.

$\{t(k, \omega)\}$ = Subsequence of dates at which player n conducts no experiment.

ω = Realization of the players' randomized choices.

Thus the interval $[0, t]$ consists of experiment dates $P_n(x_n, t)$ and play dates $M_n(x_n, t)$. Write $M(t)$ to denote the expected total number of experiments.

- $\Pi(x_n, t)$ = Total Pay off received with $M_n(x_n, t)$
- $\Pi(y_n, t)$ = Total Pay off received with $M_n(y_n, t)$

Game Theory & Learning: LECTURE 10 (CONT'D)

Claim:

Let T = set of strategy profiles.

$$\forall_{z \in T}\ \pi_n(x_n, z_{-n}) < \pi_n(y_n, z_{-n}) - 2\epsilon$$

$\Rightarrow$ For large t

$$\Pi(x_n, t) < \Pi(y_n, t) - \epsilon M(t)/|S_n|.$$

$$\begin{aligned} & E[\Pi(x_n, \tau+1) - \Pi(y_n, \tau+1)|T] - \Pi(x_n, \tau) + \Pi(y_n, \tau) \\ &= \epsilon_{n,\tau+1}/|S_n|\ E[\pi_n(x_n, x_{-n}(\tau+1)) - \pi_n(y_n, x_{-n}(\tau+1))] \\ &< -2\epsilon \cdot \epsilon_{n,\tau+1}/|S_n|. \end{aligned}$$

Taking expectations

$$\begin{aligned} & E[\Pi(x_n, \tau+1) - \Pi(y_n, \tau+1)|T] \\ &= E[\Pi(x_n, \tau) - \Pi(y_n, \tau)|T] - 2\epsilon \cdot \epsilon_{n,\tau+1}/|S_n|. \end{aligned}$$

and then telescoping,

$$E[\Pi(x_n, t) - \Pi(y_n, t)] < -2\epsilon/|S_n| \sum \epsilon_{n,t} = -2\epsilon M(t)/|S_n|.$$

Let $\Delta > 2|\pi_n|$.

$$\text{Var}[\Pi(x_n, t) - \Pi(y_n, t)] \leq -2\Delta^2 \epsilon M(t)/|S_n|.$$

Thus $\Pi(x_n, t) - \Pi(y_n, t)] + \epsilon M(t)/|S_n|$ converges to $-\infty$ and hence represents a super-martingale.

In other words, x_n dominates y_n then the player n will discover this fact eventually by repeated experiments.

Theorem *For any finite strategy game Γ, the sequence $\{x_n(t(k, \omega))\}$ constructed as described above is consistent with adaptive learning a.s.(almost surely).*