No regret dynamics

$G = (P, A, C)$ of n players

$\hookrightarrow$ cost function : $-1 \cdot$ utility

$C \in \langle -1, 1 \rangle$ – if not, we can scale...

Play G T-times, the player $i \sim$ i-th agent chooses $\underline{p^t}$ as his mixed strategy at step t.

$\hookrightarrow$ prob. distr. on A_i obtained by PW alg. with average regret $\leq \varepsilon$

Losses of i are captured by C_i

for $a_i \in A_i$: $\ell^t(a_i) = \underline{E_{a_{-i}^t \sim p_{-i}^t}} \left[C_i(a_i, \underline{a_{-i}^t}) \right]$

$\hookrightarrow$ all other play according to the PW alg. strategy

Application:

The expected cost $C_2(s) = x^T M y$, $M_{ij} = -C_1(a_i, b_j) = C_2(a_i, b_j) \in \langle -1, 1 \rangle$

Thm: Modern proof of minimax

1) $\max\limits_{x \in S_1} \min\limits_{y \in S_2} x^T M y \leq \min\limits_{y \in S_2} \max\limits_{x \in S_1} x^T M y$

Dk: For player 1 it is only worse to go first.

second player can make it worse for the first player

but here the first can choose just based on the S_2 choice.

2) $\max\limits_{x \in S_1} \min\limits_{y \in S_2} x^T M y \geq \min\limits_{y \in S_2} \max\limits_{x \in S_1} x^T M y$

Dk: Apply no regret dynamics with average regret ε. $\left(T \geq 4 \log(\max(m, n)) / \varepsilon^2\right)$

For $t = 1 \dots T$, let $p^t =$ mixed strategy used by 1

$q^t =$ 2

Define average strategies $\bar{x} = \frac{1}{T} \sum\limits_{t=1}^{T} p^t$ for player 1

$\bar{y} = \frac{1}{T} \sum\limits_{t=1}^{T} q^t$ for player 2

Average cost of 2 $= v = \frac{1}{T} \sum\limits_{t=1}^{T} (p^t)^T M q^t$

Assume that player 1 plays a_i pure strategy with $p(a_i)=1$

$\Rightarrow$ mixed strategy vector $e_i = (0,\dots,0,\underset{i}{1},0,\dots,0)$

Now payoff of 1 against $\bar{y}$ $\approx$ $\underset{\text{by definition of } \bar{y}}{e_i^T M \bar{y} = \frac{1}{T}} \overbrace{\sum_{t=1}^{T} e_i^T M q^t}^{= -L_i^T} \le \frac{1}{T} \overbrace{\sum_{t=1}^{T} (p^t)^T M q^t}^{= -L_{A_i}^T} + \varepsilon$

because average regret $\le \varepsilon$

But we can observe: $L_{A_i}^T \le L_i^T + \varepsilon^T$

$\forall x \in S_1$ is convex combination of some pure strategy $e_1 \dots e_m$, also linear,

therefore the inequality holds even for mixed strategy vectors.

also: $\Rightarrow \forall x \in S_1: x^T M \bar{y} \le v + \varepsilon$ (*)

By symmetry: $\forall y \in S_2: \bar{x}^T M y \ge v - \varepsilon$ (**)

We want to prove $\max_{x \in A_1} \min_{y \in A_2} x^T M y \ge \min_{y \in S_2} (\bar{x})^T M y \underset{(**)}{\ge}$

$\underset{(**)}{\ge} v - \varepsilon \underset{(*)}{\ge} \max_{x \in A_1} x^T M \bar{y} - 2\varepsilon \ge \min_{y \in S_2} \max_{x \in S_1} x^T M y - 2\varepsilon$

works for any x

arbitr. number chosen.

With $T \to \infty$, we have $\varepsilon \to 0$ $\boxtimes$

Coarse correlated equilibria (CCE)

Again a prob. distr. p on A, if $\forall i \in A\ \forall a_i \in A_i$:

$$\sum_{a \in A} C_i(a)\, p(a) \leq \sum_{a \in A} C_i(a_i', a_{-i})\, p(a)$$

$-i$ does not want to deviate in expectation (before he knows a_i that is suggested to him)

CCE are more general that the true CE.

$\Rightarrow$ therefore every CE is CCE.

For $\varepsilon > 0$, ε-CCE is a prob. distr. p on A such that:

$\forall i \in P,\ \forall a_i' \in A_i$:

$$\sum_{a \in A} C_i(a)\, p(a) = \sum_{a \in A} C_i(a_i', a_{-i})\, p(a) + \varepsilon$$

Thm: $\forall G = (P, A, C),\ \forall \varepsilon > 0,\ \forall T = T(\varepsilon)$:

After T steps of no-regret dynamics, with average regret $\leq \varepsilon$,

set $p^t = \prod_{i=1}^{n} p_i^t$, $p = \frac{1}{T} \sum_{t=1}^{T} p^t$

The p is ε-CCE.

Dk: We want to show that p is ε-CCE.

That is: $\forall i \in P\ \forall a_i' \in A_i$: $\mathbb{E}_{a \sim p}[C_i(a)] \leq \mathbb{E}_{a \sim p}[C_i(a_i', a_{-i})] + \varepsilon$

definition of p

$$\Rightarrow \mathbb{E}_{a \sim p}[C_i(a)] = \frac{1}{T} \underbrace{\sum_{t=1}^{T} \mathbb{E}_{a \sim p^t}[C_i(a)]}_{\text{cumulative loss of } i \text{ in no regret dynamics}} \Big/ T$$

$$\mathbb{E}_{a \sim p}[C_i(a_i', a_{-i})] = \frac{1}{T} \underbrace{\sum_{t=1}^{T} \mathbb{E}_{a \sim p^t}[C_i(a_i', a_{-i})]}_{\text{cumulative loss of playing } a_i' \text{ in no regret dynamics}} \Big/ T$$

$\wedge | \rightarrow$ by average regret $\leq \varepsilon$

$+\varepsilon$

$\boxtimes$