Setup

ADV

columns j

ALG

rows i

game matrix M

$M(i,j)$ cost of ALG for (row i, col j).

Each round t:

- ALG chooses distribution p_t over rows
  ADV — " — q_t over columns

  (ALG goes first, or simultaneous)

- row & column realized: $i \sim p_t$, $j \sim q_t$.

- ALG suffers cost $M(i_t, j_t)$

- ALG observes $M(i_t, j_t)$ and

  ... nothing else $\Rightarrow$ "bandit feedback"

  ... $M(i, j_t)$ for all rows j $\Rightarrow$ "full feedback
  e.g. if M is known and j_t is revealed.

  ... or anything in between $\Rightarrow$ "partial feedback".

ALG faces adaptive adversary
has regret $\bar{R}(T) = R(T)/T$ vs adaptive adv.
Assume $\bar{R}(T) \longrightarrow 0$ as $T \to \infty$.

Zero-Sum games

(1) $M(p,q) = \mathbb{E}_{i\sim p,\, j\sim q}\, M(p,q)$

$p \in \Delta(\text{rows})$
$q \in \Delta(\text{cols})$

$$v^* = \min_p \underbrace{\max_q M(p,q)}_{f(p)}$$ "minimax value".

$\left.\begin{array}{l} f(p) \text{ continuous} \\ \Delta(R) \text{ closed \& bdd} \end{array}\right\} \Rightarrow \exists\ p^* \in \arg\min_p f(p)$ "minimax strategy":

$M(p^*, q) \le v^* \quad (\forall q).$

(2) Arbitrary ADV

$$\begin{cases} C_t(i) = M(i, j_t) \\ \text{cost}(ALG) = \frac{1}{T}\sum_t \underbrace{\mathbb{E}[C_t(i_t)]}_{M(p_t,\, q_t)} \\ \bar{R}(T) = R(T)/T \end{cases}$$

Lm 1

$$\text{cost}(ALG) \le \underbrace{\min_i \text{cost}(i)}_{\text{cost}^*} + \bar{R}(T) \le v^* + \bar{R}(T)$$

regret

pf

$$\text{cost}^* \le \text{cost}(p^*) = \frac{1}{T}\sum_t \underbrace{M(p^*, q_t)}_{\le v^*} \le v^*.$$

(3) Best-responce ADV: $q_t \in \arg\max_q M(p_t, q)$

$\bar{p} = \frac{1}{T}\sum_t p_t$

Clm 2 $\quad M(\bar{p}, q) \le \text{cost}(ALG), \ \forall q$

pf

$$M(\bar{p}, q) = \frac{1}{T}\sum_t \underbrace{M(p_t, q)}_{\le M(p_t, q_t) \text{ by best-responce}} \le \frac{1}{T}\sum_t M(p_t, q_t) = \text{cost}(ALG)\ \square.$$

Lm 3 $\quad M(\bar{p}, q) \le v^* + \bar{R}(T), \quad \forall q.$

pf

$$M(\bar{p}, q) \underset{\text{Clm 2}}{\le} \text{cost}(ALG) \underset{\text{Lm 1}}{\le} v^* + \bar{R}(T)$$

$\bar{p}$: ε-approx minimax strategy, $\quad \varepsilon = \bar{R}(T)$

4 $$\min_p \underbrace{\max_q M(p,q)}_{f(p)} = \max_q \underbrace{\min_p M(p,q)}_{h(q)} \qquad (*)$$

pf: $\geq$ $$M(p,q) \geq \min_{p'} M(p',q)$$

$$\max_q M(p,q) \geq \max_q \min_{p'} M(p',q) \quad \rightarrow (*)$$

pf: $\leq$ Consider ALG vs ADV s.t.

$$f(\bar{p}) \leq \text{cost}(ALG) + \bar{R}(T), \quad (\forall T). \quad (**)$$

(e.g., best-responce ADV by Lm 2).

$$\text{cost}^* = \min_p \underbrace{\tfrac{1}{T} \textstyle\sum_t M(p, q_t)}_{M(p,\bar{q})} = h(\bar{q}), \qquad \bar{q} = \tfrac{1}{T} \textstyle\sum_t q_t$$

$$\begin{aligned}
\min_p f(p) &\leq f(\bar{p}) \\
&\leq \text{cost}(ALG) + \bar{R}(T) && \text{by } (**) \\
&\leq \text{cost}^* + 2\bar{R}(T) && \text{by Lm 1.} \\
&\leq h(\bar{q}) + 2\bar{R}(T) \\
&\leq \max_q h(q) + 2\bar{R}(T).
\end{aligned}$$

take $T \to \infty$. □

$q^* \in \arg\max_q h(q)$ exists for same reason as p^*.

Interpretation: (p^*, q^*) Nash eq.: p^* best responce to q^* & vice versa

(5) ALG vs ALG'. (rewards vs costs).

$$\begin{cases} \text{cost}(ALG) \le \text{cost}^* + \bar{R}(T) \\ \text{rew}(ALG') \le \text{rew}^* + \bar{R}(T) \\ \text{cost}(ALG) = \text{rew}(ALG') \end{cases}$$

$$\text{cost}^* = \min_p \underbrace{\tfrac{1}{T} \textstyle\sum_t M(p, q_t)}_{M(p,\bar{q})} = h(\bar{q})$$

$$\text{rew}^* = \max_q \underbrace{\tfrac{1}{T} \textstyle\sum_t M(p_t, q)}_{M(\bar{p}, q)} = f(\bar{p}).$$

$$\underbrace{f(\bar{p}) - \bar{R}(T)}_{v^* - \bar{R}(T) \le} \le \text{cost}(ALG) \le \underbrace{h(\bar{q}) + \bar{R}(T)}_{\le v^* + \bar{R}(T)} \quad (***)$$

$\Rightarrow$ can use to prove (4)

$$\begin{cases} v^* = \min_p f(p) \le f(\bar{p}) \\ h(\bar{q}) \le \max_q h(q) = v^* \end{cases}$$

used (4).

$\cancel{v^* - \bar{R}(T) \le \text{cost}(ALG) \le v^* + \bar{R}(T)}$

Lm
$$\left| \text{cost}(ALG) - v^* \right| \le \bar{R}(T).$$

Lm
$$\begin{cases} f(\bar{p}) \le v^* + 2\bar{R}(T) & \text{"}\varepsilon^*\text{-approx minimax strategy"} \\ h(\bar{q}) \ge v^* - 2\bar{R}(T) & \text{"}\varepsilon\text{-approx maximin strategy"} \end{cases}$$

$(\bar{p}, \bar{q})$: ε-approx. MNE, $\varepsilon = 2\bar{R}(T)$.

Interpretation:

- ALG vs ALG' $\longrightarrow$ MNE. (face value)
- algo to approximate MNE.
- natural dynamics, ppl can plausibly arrive @ MNE.

NB: starting from $(***)$, pf works for any (ALG, ADV) s.t. $(**)$.
in particular, for (ALG, Best Response).

(6) Arbitrary games, ALG vs ALG'

$$\begin{cases}\sigma_t = p_t \times q_t \\ \sigma = \frac{1}{T}\sum_t \sigma_t\end{cases}$$ distribution over (i,j) "outcomes".

$$U_\sigma \triangleq \mathbb{E}_{(i,j)\sim\sigma} M(i,j) = \frac{1}{T}\sum_t \underbrace{\mathbb{E}_{(i,j)\sim\sigma_t} M(i,j)}_{\mathbb{E}[c_t(i_t)]}$$

$$= \text{cost}(ALG).$$

$$U_\sigma(i) \triangleq \mathbb{E}_{j\sim\bar{q}} M(i,j) = \frac{1}{T}\sum_t \underbrace{\mathbb{E}_{j\sim q_t} M(i,j)}_{\mathbb{E}[c_t(i)]}$$

$$= \text{cost}(i)$$

$$U_\sigma \le U_\sigma(i) + \bar{R}(T).$$

[σ [approx] coarse correlated eq. (CCE).

Same for ALG'.

Interpretation: [coordinator suggests σ
each agent must commit before σ is realized
→ no incentive to deviate.

Extends to N agents.

NB: what if agent can look at ~~realization~~ "his" recommendation?
Stronger notion of eq.,
needs stronger notion of regret. "internal regret".

Benchmark: $ALG_f = \sum_t c_t(f(a_t))$, $f: A \to A$.

$$\sup_f ALG_f$$