Introduction to Learning Equilibria

Vianney Perchet

ENS Paris Saclay & Criteo AI Lab

Summer School "Strategy and Dynamics in Games"

Outline

I) External Regret

II) Internal Regret

III) Impossibility Results

IV) Query Complexity

External Regret

1 player vs Nature

- Action Set $\quad a \in A \quad\quad b \in B$
- Payoff: $\quad g(a,b) \xrightarrow{\text{mixed}} g(x,y)$
- Repeated $\quad t \in \mathbb{N} \quad\quad g_t := g(x_t, y_t)$

$$\boxed{\text{REGRET}: \quad \max_{x \in \Delta(A)} \frac{1}{T}\sum_{t=1}^{T} g(x, y_t) - \frac{1}{T}\sum_{t=1}^{T} g_t}$$

Vectorial Formulation

$$R_T \leq 0 \iff \forall a \in A, \ \frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) \geq \frac{1}{T}\sum_{t=1}^{T} g(a, y_t)$$

$$\iff \forall a \in A, \ \frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) - g(a, y_t) \geq 0$$

$$\iff \frac{1}{T}\sum_{t=1}^{T} \Big[g(x_t, y_t) - g(a, y_t) \Big]_{a \in A} \in \mathbb{R}_+^A$$

$$\iff \frac{1}{T}\sum_{t=1}^{T} V(x_t, y_t) =: \overline{V}_T \in \mathbb{R}_+^A$$

Regret via Approachability

$$V(x,y) = [g(x,y) - g(a,y)]_{a \in A} \in \mathbb{R}^A$$

$$\boxed{\text{Approaching } \mathcal{C} = \mathbb{R}^A_+ \iff \text{No Regret}}$$

$\mathcal{C}$ app. iff $\forall y, \exists x_y : V(x_y, y) \in \mathcal{C}$

$$a^*_y = \arg\max g(a,y) \Rightarrow V(a^*_y, y) \in \mathcal{C}$$

Reminder on Approachability

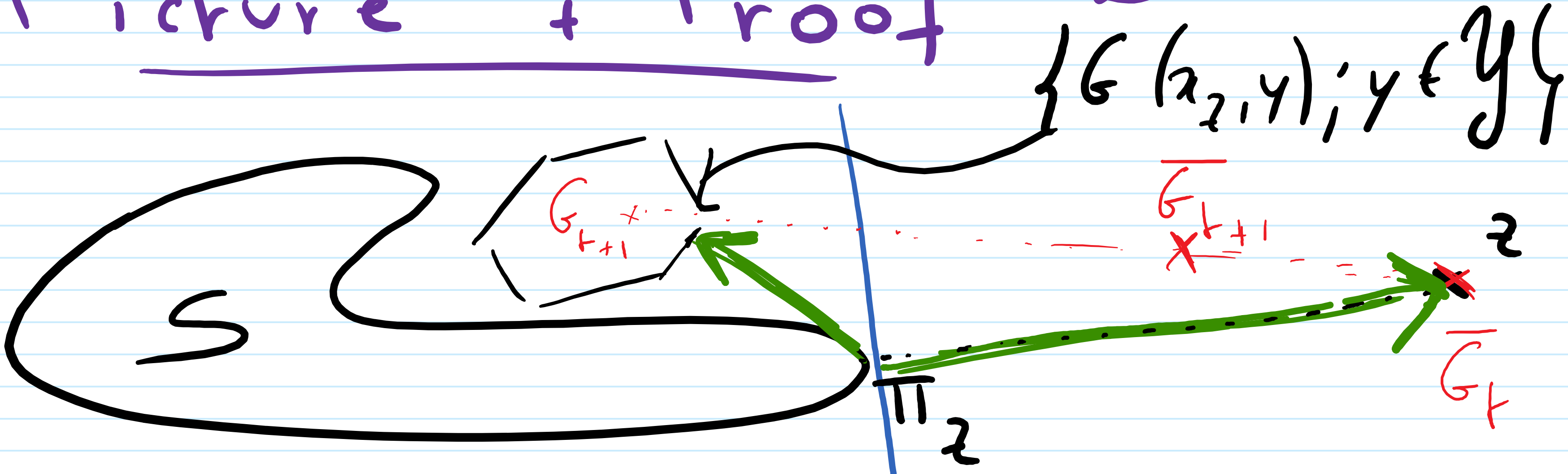
- $x \in \mathcal{X}$, $y \in \mathcal{Y}$ $\Rightarrow$ $G(x,y) \in \mathbb{R}^d$
  Target set $S \subset \mathbb{R}^d$

- If $\forall z \in \mathbb{R}^d$, $\exists$ projection Π_z.
  action $x_z \in \mathcal{X}$ s.t. $\forall y \in \mathcal{Y}$
  $$\langle z - \Pi_z , G(x_z, y) - \Pi_z \rangle \leq 0$$

- Approachability strat: $x_{t+1} = x_{\bar{G}_t}$

$$\boxed{d(\bar{G}_t, S) \leq \frac{2\,\|G\|}{\sqrt{t}}}$$

Picture + Proof 1/2



"$\bar G_{t+1}$ is closer to S than $\bar G_t$"

$$d^2(\bar G_{t+1}, S) \text{ “}\leq\text{” } d^2(\bar G_t, S)$$

$$\|\bar G_{t+1} - \Pi_{\bar G_{t+1}}\|^2 \leq \|\bar G_t - \Pi_{\bar G_t}\|^2$$

$$\leq \|\bar G_{t+1} - \Pi_{\bar G_t}\|^2 \leq$$

Proof 2/2

$$\|\bar G_{t+1} - \Pi_{\bar G_t}\|^2 = \left\| \frac{t}{t+1}\bar G_t + \frac{1}{t+1} G_{t+1} - \Pi_{\bar G_t}\right\|^2$$

$$= \left(\frac{t}{t+1}\right)^2 \|\bar G_t - \Pi_{\bar G_t}\|^2 + \left(\frac{1}{t+1}\right)^2 \underbrace{\|G_{t+1} - \Pi_{\bar G_t}\|^2}_{\le \text{ constant } M}$$

$$+ 2\,\frac{t}{t+1}\cdot\frac{1}{t+1}\underbrace{\langle \bar G_t - \Pi_{\bar G_t}, G_{t+1} - \Pi_{\bar G_t}\rangle}_{\le 0 \text{ in expectation}}$$

$$(t+1)^2\, \mathbb{E}\left[\|\bar G_{t+1} - \Pi_{\bar G_{t+1}}\|^2\right] \le t^2\, \mathbb{E}\left[\|\bar G_t - \Pi_{\bar G_t}\|^2\right] + M$$

$$\Longrightarrow \boxed{\mathbb{E}\left[\|\bar G_{t+1} - \Pi_{\bar G_{t+1}}\|^2\right] < \frac{M}{t+1}}$$

Consequence : Existence !

- No Regret $\iff$ $\overline{V_T} \in \mathbb{R}_+^A$ with

$$V(x,y) = \left[g(x,y) - g(a,y) \right]_{a \in A}$$

- Given $y \in \Delta(B)$, $x_y = \operatorname{argmax}\, g(\cdot, y)$

$$V(x_y, y) \in \mathbb{R}_+^A$$

- There exist "no-regret strategies"

Explicit Strategy: Regret Matching

- $z^+ = (\max\{0, z_a\})_{a\in A}$; $z^- = (\min\{0, z_a\})_{a\in A}$
- $x_{t+1} \propto \bar{V}_T^{\,+} \in \Delta(A)$ "regret matching"
- $\langle V(x_{t+1}, y) - \Pi_{\bar{V}_T}, \bar{V}_T - \Pi_{\bar{V}_T} \rangle$

$$\Pi_{\bar{V}_T} = \bar{V}_T^+ \qquad \bar{V}_T - \Pi_{\bar{V}_T} = \bar{V}_T^- \,!$$

$$\propto \langle V(x_{t+1}, y), x_{t+1} \rangle = \langle (g(x_{t+1}, y) - g(a, y))_{a\in A}, x_{t+1} \rangle$$

$$= g(x_{t+1}, y) - g(x_{t+1}, y) = \underline{\underline{0}}$$

Let's Recap

- Regret: $R_T = \max_{a \in \Delta(A)} \frac{1}{T} \sum_{t=1}^{T} g(a, y_t) - \frac{1}{T} \sum_{t=1}^{T} g_t$
- Regret Matching: $x_{t+1} \propto \left(\frac{1}{T} \sum_{t=1}^{T} g(a, y_t) - g_t \right)_a^+$
- Convergence: $\mathbb{E}[R_T] \leq \frac{2\|g\|}{\sqrt{T}}$

What if all players minimize Regret

Learning in Games: The setup

* n players: $g_i(a_i, a_{-i}) \in [0,1]$
* All minimize regret (Nature $= -i$)

Convergence of

- joint empirical distributions

$$\frac{1}{T}\sum_{t=1}^{T}(a_{1,t}, \dots, a_{n,t}) \in \Delta(A_1 \times \dots \times A_n)$$

- Indpt empirical distributions

$$\left(\frac{1}{T}\sum_{t=1}^{T} a_{1,t}; \frac{1}{T}\sum_{t=1}^{T} a_{2,t}, \dots, \frac{1}{T}\sum_{t=1}^{T} a_{n,t}\right) \in \Delta(A_1) \times \dots \times \Delta(A_n)$$

Hannan vs Nash

- Hannan set of a game $\mathcal{H} = \bigcap_{i=1}^{n} H_i$

$H_i = \{ q \in \Delta(A_1 \times \dots \times A_n) ;$
$\mathbb{E}_q [g_i(a_i, a_{-i})] \geq \mathbb{E}_q [g_i(a^*, a_{-i})], \ \forall a^* \in A_i \}$

- i has no regret $\Rightarrow$ joint emp. dist $\in H_i$

If all players minimize regret, joint empirical dist. converges to $\mathcal{H}$

- Nash $\subset$ Hannan ; Hannan $\not\subset$ Nash

The 0-sum case

* 2 players $\quad g_1 = g = -g_2$
* Minimize Regret $\Longrightarrow$

the set of

| Independant Emp. distr converges to optimal strat. |

Proof: $R_T^1 = \max_{x \in \Delta(A)} g(x, \overline{y}_T) - \frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) \leq 0 \quad (+o(1))$

$R_T^2 = \frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) - \min_{y \in \Delta(B)} g(\overline{x}_T, y) \leq 0 \quad (+o(1))$

- $V \leq \max_{x \in \Delta(A)} g(x, \overline{y}_T) \leq \frac{1}{T}\sum_{t=1}^{T} g(x_t, y_t) \leq \min_{y \in \Delta(B)} g(\overline{x}_T, y) \leq V$

$\Rightarrow \quad V = \max_{x \in \Delta(A)} g(x, \overline{y}_\infty) = \min_{y \in \Delta(B)} g(x_\infty, y) = V$

"LP = Min.max = Regret"

* Sorin's course : Minmax $\subseteq$ LP
* Adding - removing constraint : LP $\subseteq$ Minmax
* ℓ regret algo $\Rightarrow$ solve minmax

  Minmax $\subseteq$ Regret
* Regret $\subseteq$ Approachability $\subseteq$ Minmax

Internal Regret

Regret too weak for Nash convergence.

Refined concept: "Internal" Regret

"I have no regret of playing action $a \in A$"

$$N_T(a) = \{ t \leq T, \ a_t = a \}$$

$$R_T^{int} = \max_{a \in A} \left[\frac{1}{T} \max_{a^* \in A} \sum_{t \in N_T(a)} g(a^*, y_t) - \sum_{t \in N_T(a)} g(a, y_t) \right]$$

$R_T^{int} \leq 0 \Rightarrow$ "When a was played, it was the best action... or it was almost never played"

Vectorial Formulation

$$R_T^{int} \leq 0 \iff \forall a \in A, \forall a' \in A, \frac{1}{T} \sum_{t \in N_T(a)} g(a, y_t) - g(a', y_t) \geqslant 0$$

$$\iff \forall a \in A, \forall a' \in A, \frac{1}{T} \sum_{t=1}^{T} \left(g(a, y_t) - g(a', y_t)\right) \mathbb{1}_{\{a_t = a\}} \geqslant 0$$

$$\iff \frac{1}{T} \sum_{t=1}^{T} \left[\left(g(a, y_t) - g(a', y_t)\right) \mathbb{1}_{\{a_t = a\}} \right]_{(a, a' \in A)} \in \mathbb{R}_+^{A \times A}$$

$$\iff \frac{1}{T} \sum_{t=1}^{T} U(a_t, y_t) \in \mathbb{R}_+^{A \times A}$$

with $U(\underline{a}, y) = \left[\left(g(a, y) - g(a', y)\right) \mathbb{1}_{\{a = \underline{a}\}} \right]_{a, a'} \in \mathbb{R}^{A \times A}$

Existence!

Approachability: $\forall y \in \Delta(B),\ \exists x \in \Delta(A)$ s.t. $U(x,y) \in \mathbb{R}_+^{A \times A}$

Given $y \in \Delta(B)$, $a^* = \operatorname{argmax} g(a,y)$

$$\Rightarrow g(a^*,y) - g(a',y) \geq 0$$

$$\Rightarrow \left(g(a,y) - g(a',y)\right) \mathbb{1}\{a = a^*\} \geq 0$$

ok

There exist strategies without internal regret!

Existence easy! Explicit strategies?

Explicit Strategies

Need $\langle U(x_{t+1}, y) - \underbrace{\Pi_{\mathbb{R}_+^{A\times A}}(\bar{U}_t)}_{=\bar{U}_t^+}, \underbrace{\bar{U}_t - \Pi_{\mathbb{R}_+^{A\times A}}(\bar{U}_t)}_{=\bar{U}_t^-} \rangle \leq 0, \ \forall y \in \Delta(B)$

Or $\langle U(x_{t+1}, y), \bar{U}_t^- \rangle = 0, \ \forall y \in \Delta(B)$

$\Leftrightarrow \sum_{a\in A} x_{t+1}(a) \langle U(a, y), \bar{U}_t^- \rangle = 0, \ \forall y \in \Delta(B)$

$\Leftrightarrow \sum_{a\in A} x_{t+1}(a) \sum_{a'\in A} \left(g(a,y) - g(a',y)\right) \cdot \bar{U}_{a,a'} = 0, \ \forall y \in \Delta(B)$

$\Leftrightarrow \sum_{a\in A} g(a,y)\, x_{t+1}(a) \sum_{a'\in A} \bar{U}_{a,a'} = \sum_{a'\in A} g(a',y) \sum_{a\in A} x_{t+1}(a) \bar{U}_{a,a'}$

$\Leftarrow \sum_{a\in A} g(a,y)\, x_{t+1}(a) \sum_{a'\in A} \bar{U}_{a,a'} = \sum_{a\in A} g(a,y) \sum_{a'\in A} x_{t+1}(a') \bar{U}_{a',a}$

Perron-Frobenius $\Rightarrow$ Explicit Strat!

Need: $x_{t+1} \in \Delta(A)$ s.t. $x_{t+1}(a) \sum_{a' \in A} \bar{U}_{a,a'} = \sum_{a' \in A} x_{t+1}(a') \bar{U}_{a',a}$

Existence because of Perron Frobenius:

$$x_{t+1}(a) \sum_{a' \neq a} \bar{U}_{a,a'} = \sum_{a' \neq a} x_{t+1}(a') \bar{U}_{a',a}$$

Let $M = \max_{a \in A} \sum_{a' \neq a} \bar{U}_{a',a}$

$$\Leftrightarrow \quad x_{t+1}(a)\left(\cancel{\sum_{a' \neq a} \bar{U}_{a,a'}} + M - \cancel{\sum_{a' \neq a} \bar{U}_{a,a'}}\right) = \sum_{a' \neq a} x_{t+1}(a') \bar{U}_{a',a} + x_{t+1}(a)\left(M - \sum_{a' \neq a} \bar{U}_{a',a}\right)$$

$$\Leftrightarrow \quad x_{t+1}(a) = \sum_{a \neq a} x_{t+1}(a') \frac{U_{a',a}}{M} + x_{t+1}(a)\left(1 - \frac{\sum_{a' \neq a} \bar{U}_{a',a}}{M}\right)$$

$\Leftrightarrow \quad x_{t+1} = \mathcal{U} x_{t+1}$ with $\mathcal{U}$ stochastic matrix

What about computations?

Given stochastic matrix $\mathcal{U} \in \mathbb{R}_+^{A \times A}$ find eigenvector associated to eigenvalue 1.

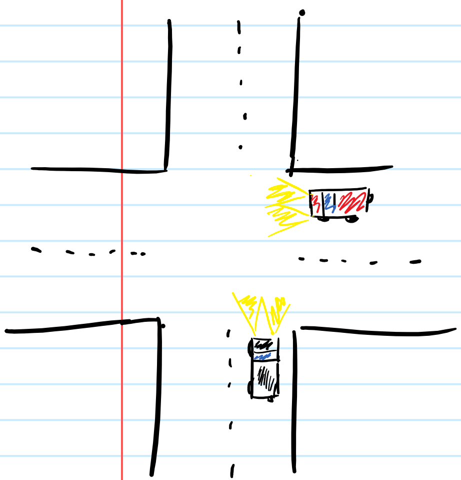
Linear Programming!

$\Rightarrow$ Solve it via Regret minimization!

What if all players minimize Internal Regret?

~~Corollated~~ Correlated!! Equilibria: Examples

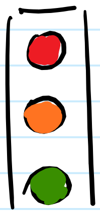
crossing roads



	S	G
S	(0,0)	(1,3)
G	(3,1)	(☠,☠)

& Nash (1,3) or (3,1)

with corollating devices



⇝ (2,2)

	A	D
D	(1,4)	(3,3)
A	(0,0)	(4,1)

Nash
- (1,4)
- (4,1)
- (2,2)

corollated $\left(\frac{8}{3}, \frac{8}{3}\right)$

	1/3	1/3
	0	1/3

not product distribution.

If told D, P_1 infers $\left(\frac{1}{2}, \frac{1}{2}\right)$

⇒ best to play D

Corollated Equilibria: Theory

* $q \in \Delta(A_1 \times A_2 \times \dots A_n)$ (not product) s.t.

For all $i \in \{1, \dots, n\}$ and all $a_i \in A_i$

$$\forall a \in A_i, \sum_{a_{-i} \in A_{-i}} g_i(a_i, a_{-i})\, q(a_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} g_i(a, a_{-i})\, q(a_i, a_{-i})$$

* Nash equilibrium is a Corollated eq.
* Set of corollated eq: finite inequalities linear

$\Rightarrow$ It is a non-empty polytope!

Learning Corollated Equilibria

* n players minimize internal regret

* $$\frac{1}{T}\sum_{t\in N_T(a_i)} g_i(a_i, a_{-i,t}) \geq \frac{1}{T}\sum_{t\in N_T(a_i)} g_i(a, a_{-i,t})$$

$$\iff \sum_{a_{-i}\in A_{-i}} \bar{q}_T(a_i, a_{-i})\, g_i(a_i, a_{-i}) \geq \sum_{a_{-i}\in A_{-i}} \bar{q}_T(a_i, a_{-i})\, g_i(a, a_{-i})$$

$\implies$ $\bar{q}_\infty$ is a corollated equilibrium

If all players minimize internal regret
They are learning corollated eq

Impossibility Results

⇒ Minimize Regret: * Learn Hannan (weak)
* optimal strat. in 0-sum

⇒ Minimize Internal regret ⇒ Corollated eq.

What about Nash? Impossible!

"There is no uncoupled ("independant") Strategies such that players can always learn Nash equlibria in any game"

Possibility Results

* Learning Eq can be learned independently in some class of games (0-sum, potential games...)
* They can be learned without computing with "coupled" strategies.

Basic ideas: - Play mixed actions "at random" for several time steps (a "block")
- Check independently, on a block that best responding.

Learning vs Computing: Complexity

* Computing Nash eq is PPAD complete
  easier than finding fixed point $f:[0,1]^3 \to [0,1]^3$

* "Cannot" have a "simple" learning algo
* Corollated eq. are "simple": LP.
  ↳ can exist learning solutions

⇒ "How many samples to find ε-equilibrium?"

"Improved Rates for Regret"

UNILATERAL

- Reg. Matching: $R_T^{(int)} \leq 2\sqrt{\frac{A}{T}} \Rightarrow O(A/\varepsilon^2)$ samples

- EXP algo: play a with proba $\dfrac{e^{\eta \sum_{t=1}^{T} g(a, y_t)}}{\sum_{a' \in A} e^{\eta \sum_{t=1}^{T} g(a', y_t)}}$

$$\boxed{R_T \leq \sqrt{\frac{\log(A)}{2T}}} \quad (\text{small variant for } R_T^{int})$$

$\hookrightarrow$ only $O\left(\frac{\log(A)}{\varepsilon^2}\right)$ samples

MULTI

- Variant of EXP by all players
  Only $O\left(\frac{\log(A)}{\varepsilon}\right)$ samples.

ε-Nash Sample Complexity

* n players with m actions each.

Complexity of finding ε-Nash eq?

$$C = 32\,\frac{\log(n) + \log(m) + \log(\frac{1}{\varepsilon}) + \log(\frac{1}{\delta}) + \log(8)}{\varepsilon^2}$$

An ε-Nash eq can be found with proba at least $1-\delta$ with less than

$$\mathrm{poly}(mn)\,\min\{m^{nC}, C^{nm}\} \text{ samples}$$

Argument 1: There exist "sparse" ε-Nash

* Start from $(x_1^*, \dots, x_n^*)$ Nash eq.

  get T iid samples from it.

  Empirical distribution $(\bar{x}_{1,T}, \dots, \bar{x}_{n,T})$ εNash?

  Yes: Proba of each action is $\frac{z}{T}$ ← some integer

* Evaluate how $\mathbb{E}_{(\bar{x}_{1,T}, \dots, \bar{x}_{n,T})}[f]$ converges to

  $\mathbb{E}_{(x_1^*, \dots, x_n^*)}[f]$ for any function f.

* "Concentration inequalities"

  $$\left|\mathbb{E}_{x^*}[f] - \mathbb{E}_{\bar{x}_T}[f]\right| \geq \varepsilon \text{ w.p less than } \frac{4e^{-\frac{\varepsilon^2 T}{8}}}{\varepsilon}$$

"Concentration Inequalities"

$X_1, X_2, \dots, X_T$ iid in $[0,1]$, $E[X_t] = \mu$, $\mathrm{var}(X_t) = \sigma^2$

* Law of Large Numbers $\frac{1}{T}\sum_{t=1}^{T} X_t = \bar{X}_T \xrightarrow{a.s.} \mu$
* Central Limit Th $\bar{X}_T \sim N(\mu, \frac{\sigma^2}{T})$
* Concentration $\mathbb{P}(\bar{X}_t - \mu \geq \varepsilon) \leq e^{-2T\varepsilon^2}$

$$\Rightarrow \mathbb{P}\left(\left|\mathbb{E}_x[f] - \frac{1}{T}\sum_{t=1}^{T} f(a_t)\right| \geq \varepsilon\right) \leq 2e^{-2T\varepsilon^2}$$

but $\frac{1}{T}\sum_{t=1}^{T}(a_{1,t}, \dots, a_{n,t})$ is **not** a product dist.

still $\mathbb{P}\left(\left|\mathbb{E}_{x^*}[f] - \mathbb{E}_{\bar{x}_T}[f]\right| \geq \varepsilon\right) \leq \frac{4e^{-\frac{\varepsilon^2 T}{8}}}{\varepsilon}$

Sparsity Argument - The end

- $8 \frac{\log(\frac{4}{\varepsilon}) + \log(\frac{1}{\delta})}{\varepsilon^2}$ sample to ε-estimate 1 fct.
- Check ε-Nash eq, estimate nm fct up to $\frac{\varepsilon}{2}$

$$\Rightarrow 8 \frac{1}{(\frac{\varepsilon}{2})^2} \left(\log\left(\frac{4}{\frac{\varepsilon}{2}}\right) + \log\left(\frac{nm}{\delta}\right) \right) \text{ samples}$$

$$C = \frac{32}{\varepsilon^2} \left(\log(n) + \log(m) + \log\left(\frac{1}{\delta}\right) + \log\left(\frac{1}{\varepsilon}\right) + \log(8) \right)$$

- With high proba $(1-\delta)$, exists ε-Nash with "proba denominator" equal to T

Argument 2: Brute force!

* ε-Nash with proba denominator equal to c

Test all possibilities

→ Put c balls in m urns: At most m^c

→ Put m balls in c urns: At most c^m

⇒ $\min\{(m^c)^n ; (c^m)^n\}$ mixed actions to check

↳ each one has a constant cost mn

That's all Folks !

Some references

. The excellent book of N. Cesa-Bianchi and G. Lugosi : "Prediction, Learning and Games"

. My Survey (J. Dynamics and Games) "Approachability, Calibration and Regret: Implications and Equivalences"

. Y. Babichenko paper "Empirical Distributions of Equilibrium Play and its testing Applications"