Online Learning in Games

Rida Laraki and Guillaume Vigeral

CNRS, PSL
IASD, Lecture 5

The Repeated Game

- Consider a two-person repeated game between a decision maker (a player) and Nature (an adversary).
- Their finite action sets are respectively denoted by $\mathcal{A}$ and $\mathcal{B}$ (cardinality A and B)
- Payoffs are defined through some vectorial mapping $g : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$.
- The game is repeated in discrete time, and we denote actions chosen at stage $n \in \mathbb{N}$ by $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$; they induce a payoff $g_n := g(a_n, b_n) \in \mathbb{R}^d$.
- Actions a_n and b_n are chosen as functions of the history $h^{n-1} = (a_1, b_1, \ldots, a_{n-1}, b_{n-1}) \in (\mathcal{A} \times \mathcal{B})^{n-1} =: H_{n-1}$.

Strategies

- A strategy σ of the player is a mapping from $H := \bigcup_{n \in \mathbb{N}} H_n$ to $\Delta(\mathcal{A})$, his the set of *mixed actions*.
- A strategy τ of Nature is a mapping from H to $\Delta(\mathcal{B})$, her set of mixed actions.
- By Kolmogorov's extension theorem, (σ, τ) induces a probability distribution $\mathbb{P}_{\sigma,\tau}$ over $\mathcal{H} = (\mathcal{A} \times \mathcal{B})^{\mathbb{N}}$ endowed with the product topology.

Notations

- Given a closed set $\mathcal{E} \subset \mathbb{R}^d$ endowed with the Euclidean norm.
- $d_{\mathcal{E}}(x) = \inf_{z \in \mathcal{E}}\{\|x - z\|\}$: distance from x to $\mathcal{E}$.
- $\mathcal{E}^\delta = \{z \in \mathbb{R}^d \text{ s.t. } d_{\mathcal{E}}(x) < \delta\}$: δ-open neighbourhood of $\mathcal{E}$.
- $\Pi_{\mathcal{E}}(x) = \{z \in \mathcal{E} \text{ s.t. } \|x - z\| = d_{\mathcal{E}}(x)\}$: projection of x on $\mathcal{E}$.
- $\text{co}\left(\mathcal{E}\right)$: the convex hull of a set $\mathcal{E}$.

$$g(x, y) = \mathbb{E}_{x,y}\left[g(a, b)\right] = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} x^a y^b g(a, b)\,.$$

Notations

- A mapping g on $\mathcal{A} \times \mathcal{B}$ is linearly extended to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ by
$$g(x,y) = \mathbb{E}_{x,y}\Big[g(a,b)\Big] = \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} x^a y^b g(a,b)\,.$$
- The time average of a sequence $s = \{s_m\}_{m\in\mathbb{N}}$ up to stage $n \in \mathbb{N}$ is denoted by $\overline{s}_n := \sum_{m=1}^{n} s_m/n$.
- In particular, $g_n = g(a_n, n_n)$ and $\overline{g}_n := \sum_{m=1}^{n} g_m/n$

Approachability / Excludability

Definition

$\mathcal{E} \subset \mathbb{R}^d$ is approachable by the player if he has a strategy σ ensuring, for every $\varepsilon > 0$, the existence of an integer $N_\varepsilon \in \mathbb{N}$ such that, independently of the strategy τ of Nature,

$$\sup_{n \geq N_\varepsilon} \mathbb{E}_{\sigma,\tau}\left(d_{\mathcal{E}}(\overline{g}_n)\right) \leq \varepsilon \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n \geq N_\varepsilon} d_{\mathcal{E}}(\overline{g}_n) \geq \varepsilon\right) \leq \varepsilon\,. \tag{1}$$

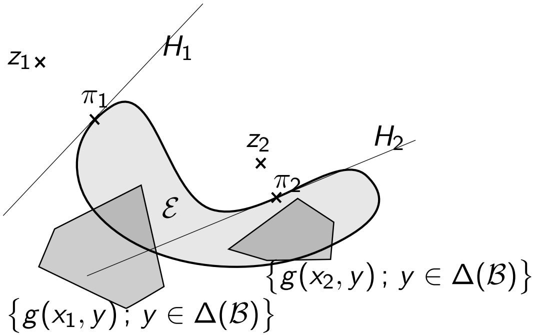
A set $\mathcal{E}$ is excludable by Nature if she can approach the complement of $\mathcal{E}^\delta$ for some $\delta > 0$.

Blackwell Condition

Definition

$\mathcal{E} \subset \mathbb{R}^d$ is a B-set if: $\forall z \in \mathbb{R}^d, \exists \pi \in \Pi_{\mathcal{E}}(z), \ \exists x \in \Delta(\mathcal{A})$ such that

$$\forall y \in \Delta(\mathcal{B}), \quad \langle g(x,y) - \pi, z - \pi \rangle \leq 0$$



Blackwell Theorem

Theorem

If $\mathcal{E}$ is a B-set, then $\mathcal{E}$ is approachable by the player. Moreover, the strategy σ defined by $\sigma(h^n) = x(\overline{g}_n)$ ensures that, for every $\eta > 0$ and against any strategy τ of Nature:

$$\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{E}}(\overline{g}_n)\right] \leq 2\sqrt{\frac{\kappa_0}{n}} \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{m\geq n} d_{\mathcal{E}}(\overline{g}_m) \geq \eta\right) \leq \frac{8}{\eta^2}\frac{\kappa_0}{n},$$

where $\kappa_0 = \sup_{a,b} \|g(a,b)\|^2$.

Proof 1

- Let σ a Blackwell's strategy for $\mathcal{E}$. Define $\delta_n := d_{\mathcal{E}}(\overline{g}_n)$ and denote by π_n the element of $\Pi_{\mathcal{E}}(\overline{g}_n)$ in σ. Then:

$$\delta_{n+1}^2 \leq \left\|\overline{g}_{n+1} - \pi_n\right\|^2 = \left\|\frac{n}{n+1}(\overline{g}_n - \pi_n) + \frac{1}{n+1}(g_{n+1} - \pi_n)\right\|^2$$

$$= \frac{n^2}{(n+1)^2}\delta_n^2 + \frac{1}{(n+1)^2}\left\|g_{n+1} - \pi_n\right\|^2$$

$$+ \frac{2n}{(n+1)^2}\left\langle \overline{g}_n - \pi_n, g_{n+1} - \pi_n \right\rangle .$$

- Conditioning on a history h^n and using Blackwell

$$\mathbb{E}_{\sigma,\tau}\left[\delta_{n+1}^2 \,\middle|\, h^n\right] \leq \frac{n^2}{(n+1)^2}\delta_n^2 + \frac{(2\|g\|_\infty)^2}{(n+1)^2}.$$

Which by induction gives $\mathbb{E}_{\sigma,\tau}[\delta_n^2] \leq 4\kappa_0/n$.

Proof 2

- Define

$$Z_n := \delta_n^2 + \sum_{k=n}^{\infty} \frac{4\|g\|_\infty^2}{(k+1)^2}.$$

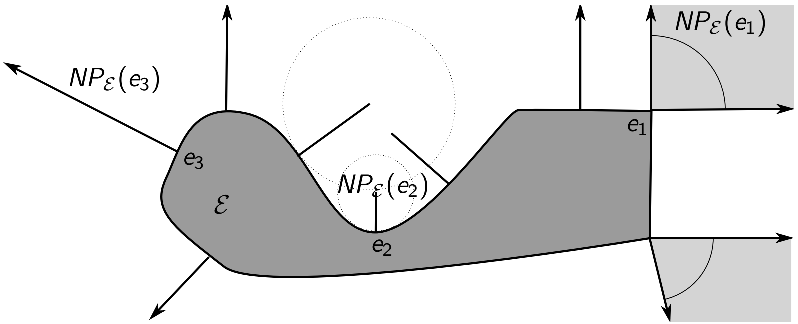
- Z_n is a supermartingale and $\mathbb{E}_{\sigma,\tau}[Z_n] \leq \frac{8\kappa_0}{n}$
- Doobs' inequality implies that

$$\begin{aligned}
\mathbb{P}_{\sigma,\tau}(\exists m \geq n \text{ s.t. } \delta_m^2 \geq \eta^2) &\leq \mathbb{P}_{\sigma,\tau}(\exists m \geq n \text{ s.t. } Z_m \geq \eta^2) \\
&\leq \frac{\mathbb{E}_{\sigma,\tau}[Z_n]}{\eta^2} \\
&\leq \frac{8\kappa_0}{\eta^2 n},
\end{aligned}$$

Definition

The set of proximal normals to some closed set $\mathcal{E} \subset \mathbb{R}^d$ at $e \in \mathcal{E}$ is denoted by $NP_{\mathcal{E}}(e) \subset \mathbb{R}^d$ and is defined by:

$$NP_{\mathcal{E}}(e) = \left\{ p \in \mathbb{R}^d,\ e \in \Pi_{\mathcal{E}}(e+p) \right\} = \left\{ p \in \mathbb{R}^d,\ d_{\mathcal{E}}(e+p) = \|p\| \right\},$$



Proximal Cone Characterization

Theorem

A set $\mathcal{E}$ is a B-set if and only if:

$$\forall\, e \in \mathcal{E},\ \forall p \in NP_{\mathcal{E}}(e), \quad \min_{x \in \Delta(\mathcal{A})} \max_{y \in \Delta(\mathcal{B})} \langle\, p\,, g(x,y) - e \rangle \leq 0. \quad (2)$$

Proof:

- Suppose (2). Normal proximal definition ensures that $\forall z \in \mathbb{R}^d$ and $\forall \pi \in \Pi_{\mathcal{E}}(z)$, $z - \pi \in NP_{\mathcal{E}}(\pi)$. Thus, $\mathcal{E}$ is a B-set.
- Reciprocally, $p \in NP_{\mathcal{E}}(e) \implies \frac{p}{2} \in NP_{\mathcal{E}}(e) \implies \{e\} = \Pi_{\mathcal{E}}(e + p/2)$.
- As a consequence, if $\mathcal{E}$ is a B-set, Equation (2) is satisfied with respect to e and $p/2$, hence with respect to e and p.

Complete characterizations

Theorem

A closed set $\mathcal{E}$ is approachable if and only if it contains a B-set.

Theorem

A closed and convex set $\mathcal{C} \subset \mathbb{R}^d$ is approachable by the player if and only if:

$$\forall\, y \in \Delta(\mathcal{B}),\ \exists\, x \in \Delta(\mathcal{A}), \quad g(x, y) \in \mathcal{C}, \tag{3}$$

Theorem

A convex set is either approachable by P or excludable by N.

Proof 1

- Let $\mathcal{C} \subset \mathbb{R}^d$ be a convex set and $p \in \mathbb{R}^d$ be a normal proximal of $\mathcal{C}$ at some $z \in \mathcal{C}$.
- Condition (3) can be rewritten as

$$\max_{y \in \Delta(\mathcal{B})} \min_{x \in \Delta(\mathcal{A})} \langle\, p \,, g(x,y) - z \rangle \leq 0.$$

- The minmax theorem implies that

$$\min_{x \in \Delta(\mathcal{A})} \max_{y \in \Delta(\mathcal{B})} \langle p, g(x,y) - z \rangle = \max_{y \in \Delta(\mathcal{B})} \min_{x \in \Delta(\mathcal{A})} \langle p, g(x,y) - z \rangle \leq 0, \tag{4}$$

- Thus $\mathcal{C}$ is a B-set and so is approachable by the player.

Proof 2

- If Condition (3) is not satisfied, there exists some $y_0 \in \Delta(\mathcal{B})$ and $\delta > 0$ such that $d_{\mathcal{C}}(g(x, y_0)) \geq \delta$.
- If Nature plays repeatedly accordingly to y_0, then the law of large numbers implies that $\overline{g}_n$ converges to the set

$$\{g(x, y_0),\ x \in \Delta(\mathcal{A})\}$$

- So $\mathcal{C}$ is excludable by Nature.

Extension to infinite action sets

- Compact and convex subset $\mathcal{X} \subset \mathbb{R}^A$ and $\mathcal{U} \subset \left(\mathbb{R}^d\right)^A$.
- At stage n, Nature chooses $U_n = (U_n^a)_{a \in \mathcal{A}} \in \mathcal{U}$.
- The player chooses $x_n = (x_n^1, \ldots, x_n^A) \in \mathcal{X}$.
- This generates the sequence $g_n = x_n.U_n = \sum_{a=1}^A x_n^a U_n^a \in \mathbb{R}^d$.
- **Theorem:** A closed set in $\mathbb{R}^d$ is approachable if and only if it contains a B-set $\mathcal{E}$:

$$\forall\, z \in \mathbb{R}^d,\ \exists\, \pi \in \Pi_{\mathcal{E}}(z), \quad \inf_{x \in \mathcal{X}} \sup_{U \in \mathcal{U}} \langle x.U - \pi, z - \pi \rangle \leq 0\,.$$

- **Theorem:** A closed convex $\mathcal{C} \subset \mathbb{R}^d$ is approachable if and only if:

$$\forall\, U \in \mathcal{U},\ \exists\, x \in \mathcal{X},\ x.U \in \mathcal{C}.$$

External Regret: Hannan 1957

- A two-person repeated game where action spaces of the Player and Nature are $\mathcal{A}$ and $\mathcal{B}$ and $\rho : \mathcal{A} \times \mathcal{B} \to [0, 1]$ is a real payoff mapping, extended linearly to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$
- The choices of actions $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$ generate a regret $r_n \in \mathbb{R}^A$ defined by

$$r_n = r(a_n, b_n) := \Big(\rho(1, b_n) - \rho(a_n, b_n), \ldots, \rho(A, b_n) - \rho(a_n, b_n)\Big) \in \mathbb{R}^A$$

- The regret r_n represents the differences between what the player could have got and what he actually got.
- This defines a vector payoff game between the Player and Nature on $\mathbb{R}^A$.

External Regret: Hannan 1957

Definition

A strategy σ of the player has no external regret if, for all strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,

$$\limsup_{n\to\infty} \max_{a^* \in \mathcal{A}} \rho(a^*, \overline{b}_n) - \overline{\rho}_n \leq 0 \tag{5}$$

where $\overline{b}_n := \sum_{m=1}^n b_m/n$ and $\overline{\rho}_n := \sum_{m=1}^n \rho(a_m, b_m)/n$.

- By linearity of ρ,

$$\overline{r}_n = \Big(\rho(1, \overline{b}_n) - \overline{\rho}_n, \ldots, \rho(A, \overline{b}_n) - \overline{\rho}_n\Big) \in \mathbb{R}^A.$$

- Thus, no external regret $\implies$ player can approach $\mathcal{C} = \mathbb{R}^A_-$ in the vector payoff game.

Hart-Mas-Collel strategy

Given $U \in \mathbb{R}^d$, U^+ stand for the positive part of U, i.e., $U^+ = \left(\max\{U^i, 0\}\right)_{1\le i\le d}$. Similarly, U^- is the negative part of U.

Theorem

The strategy σ defined by playing, at stage $n+1$, proportionally to $\overline{r}_n^+$ (and arbitrarily if $\overline{r}_n^+ = 0$) is **externally consistent**. *Moreover, for every strategy τ of Nature and $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\left[\|\overline{r}_n^+\|_\infty\right] \le \mathbb{E}_{\sigma,\tau}\left[\|\overline{r}_n^+\|_2\right] \le \sqrt{\frac{A}{n}}\ ,$$

and, for every $\eta > 0$,

$$\mathbb{P}_{\sigma,\tau}\left\{\sup_{m\ge n}\|\overline{r}_m^+\| - \sqrt{\tfrac{A}{m}} \ge \eta\right\} \le \tfrac{4A}{\eta^2}\exp\left(-\tfrac{\eta^2 n}{2A}\right).$$

Proof

- We simply have to prove that σ is a Blackwell's approachability strategy of the negative orthant $\mathcal{C} = \mathbb{R}_-^A$ in the game where the vector payoff is $r(a, b)$.
- This is a consequence of the following **geometric property**.

$$\forall x \in \Delta(\mathcal{A}),\ \Big\langle\, x\,, \mathbb{E}_x[r(a,b)]\, \Big\rangle = 0,\ \forall b \in \mathcal{B}.$$

- Because k-th component of $\mathbb{E}_x[r(a,b)]$ is $\rho(k,b) - \rho(x,b)$.
- By construction $x_{n+1} = \sigma(h^n)$ is proportional to $\overline{r}_n^+$, so the **geometric property** implies:

$$\Big\langle\, \overline{r}_n^+, \mathbb{E}_{\sigma,\tau}[r_{n+1}]\, \Big\rangle = 0 \quad \text{thus} \quad \Big\langle\, \overline{r}_n - \overline{r}_n^-\,, \mathbb{E}_{\sigma,\tau}[r_{n+1}] - \overline{r}_n^-\, \Big\rangle = 0$$

because one always has $\langle z^+, z^- \rangle = 0$.

- Since $\overline{r}_n^- = \Pi_{\mathbb{R}_-^A}(\overline{r}_n)$, σ satisfies Blackwell property, hence is an approachability strategy.

Remarks

- The existence of externally consistent strategies is immediate because $\mathbb{R}_-^A$ is obviously a convex approachable set.
- Indeed, for every $y \in \Delta(\mathcal{B})$, there exists $x \in \Delta(\mathcal{A})$ such that $r(x, y) \in \mathbb{R}_-^A$: it suffices to take for x any best response to y.
- The interesting feature of that strategy is its simplicity and the error bound $\sqrt{\frac{A}{n}}$.
- If we follow the exponential weight algorithm, which projects following a non-euclidean distance we obtain:

$$x_{n+1}[a] = \frac{\exp\left(\eta_n \rho(a, \overline{b}_n)\right)}{\sum_{a' \in \mathcal{A}} \exp\left(\eta_n \rho(a', \overline{b}_n)\right)} \text{ where } \eta_n = \sqrt{8n \log(A)}\ ,$$

We obtain the following stronger bound

$$\frac{1}{n}\mathbb{E}_{\sigma,\tau}\left[\sum_{m=1}^{n} \rho(a, b_m) - \rho_m\right] \leq 2\sqrt{\frac{\log(A)}{n}}.$$

Non observation of other players actions

- Suppose that a player observe only the sequence of his realised payoffs $\rho_n(a_n, b_n)$, $n = 1, 2, ...$ but not the actions of the opponents.
- We can let him experiment with probability ε (and the day he experiments, he plays uniformly all his actions).
- From this, the player can form an estimate for it's payoff when he plays some action a.

Non observation of other players actions

- Let $\vec{\varepsilon}_n \in \mathbf{R}^A$ be the random vector s.t. $\vec{\varepsilon}_n(a) = 1$ if at day n the agent experiments and chooses at that day the action $a \in A$, otherwise we let $\vec{\varepsilon}_n(a) = 0$.
- Define $Z_n(a) = \frac{\#A}{\varepsilon}\vec{\varepsilon}_n(a)\rho(a, b_n) - \rho(a, b_n)$.
- Then: $\mathbb{E}Z_n = 0$. As such $\overline{Z}_n = \frac{1}{n}\sum_k^n Z_k \to 0$ a.s.
- But $\overline{Z}_n(a) = \frac{1}{n}\frac{\#A}{\varepsilon}\sum_{k=1}^n \vec{\varepsilon}_k(a)\rho(a, b_k) - \frac{1}{n}\sum_{k=1}^n \rho(a, b_k)$
- Which is equal to the difference between the estimated regret and the real regret.
- Hence, playing the Hart Mas-Collel strategy with the estimated regret, then except a fraction ε of periods, the probability 1, the real regret goes to zero.

Internal Regret

- The player has no internal regret (or his strategy is internally consistent) if he has no external regret on the set of stages where he choses a specific given action.
- Formally, the choices of action $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$ generate, an internal regret R_n which is an $A \times A$-matrix whose rows are null except the a_n-th one which is r_n^T:

$$R_n^{a,a'} = R(a_n, b_n)^{a,a'} := \begin{cases} \rho(a', b_n) - \rho(a_n, b_n) & \text{if } a = a_n \\ 0 & \text{otherwise} \end{cases}.$$

- **Definition:** A strategy σ is internally consistent if, against any strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,

$$\limsup_{n\to\infty} \|\overline{R}_n^+\|_\infty \leq 0$$

Foster & Vohra 1997

Theorem

The strategy σ that dictates to play at stage $n+1$ an invariant measure of $\overline{R}_n^+$ is **internally consistent** *and* $\forall \tau$, $\forall n$, $\forall \eta > 0$,

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\overline{R}_n^+\right\|_\infty\right] \leq \mathbb{E}_{\sigma,\tau}\left[\left\|\overline{R}_n^+\right\|_2\right] \leq \sqrt{\frac{A}{n}},$$

$$\mathbb{P}_{\sigma,\tau}\left\{\sup_{m \geq n} \|\overline{R}_m^+\| - \sqrt{\frac{A}{m}} \geq \eta\right\} \leq \frac{4A}{\eta^2} \exp\left(-\frac{\eta^2 n}{2A}\right)$$

Invariant measures

A probability distribution $\lambda \in \Delta(\{1, \ldots, d\})$ is an invariant measure of a non-negative matrix M of size $d \times d$ if

$$\sum_{k=1}^{d} \lambda^k M^{k,i} = \lambda^i \sum_{k=1}^{d} M^{i,k}, \quad \forall i \in \{1, \ldots, d\} \ ,$$

Existence is a consequence of Perron-Frobenius theorem.

Proof 1

- We prove that σ is a Blackwell's approachability strategy.
- This is a consequence of a **geometric property**:

  Any invariant measure λ of some matrix M with non-negative coefficient satisfies, for any choice of $b \in \mathcal{B}$, $\left\langle M, \mathbb{E}_\lambda[R(a,b)] \right\rangle = 0$.

- Let $U^a := \rho(a,b)$, then the (i,k)-component of $\mathbb{E}_\lambda[R(a,b)]$ is $\lambda^i\left(U^k - U^i\right) = \lambda^i\left(\rho(k,b) - \rho(i,b)\right)$.
- The inner product is equal to $\sum_{i,k} M^{i,k}\lambda^i\left(U^k - U^i\right)$ and the coefficient of U^i in this sum is

$$\sum_k \lambda^k M^{k,i} - \lambda^i \sum_k M^{i,k} = 0$$

  because λ is an invariant measure of M.

Proof 2

- Since x_{n+1} is an invariant measure of $\overline{R}_n^+$, the **geometric property** implies that

$$\left\langle \overline{R}_n^+, \mathbb{E}_{\sigma,\tau}[R_{n+1}] \right\rangle = 0 \quad \text{thus} \quad \left\langle \overline{R}_n - \overline{R}_n^- , \mathbb{E}_{\sigma,\tau}[R_{n+1}] - \overline{R}_n^- \right\rangle = 0 \, .$$

- This proves that σ satisfies Blackwell's property, hence is an approachability strategy.

Swap Regret

- For every mapping $\phi : \mathcal{A} \to \mathcal{A}$ and every family of such mappings $\Phi \subset \{\phi : \mathcal{A} \to \mathcal{A}\}$, define the following quantities:

$$\overline{R^\phi}_n = \frac{1}{n} \sum_{m=1}^{n} \rho(\phi(a_m), b_m) - \rho(a_m, b_m) \text{ and } \overline{R^\Phi}_n = \left(\overline{R^\phi}_n\right)_{\phi \in \Phi} \in \mathbb{R}^{|\Phi|}$$

- A strategy σ has no Φ-regret if, against any strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,

$$\limsup_{n \to \infty} \left\| \overline{R^\Phi}_n^+ \right\|_\infty \leq 0$$

- Finally, given $M \in \mathbb{R}^{|\Phi|}$, let $\Theta^\Phi(M)$ be the $A \times A$-matrix whose (a, a') component is $\Theta^\Phi(M)^{a,a'} = \sum_{\phi : \phi(a) = a'} M^\phi$

Theorem

Let Φ be a family of swap mappings. The strategy σ playing at stage $n+1$ accordingly to any invariant measure of $\Theta^{\Phi}(\overline{R_n^{\Phi}}^{+})$ has no Φ-regret. Moreover, for every strategy τ of Nature and $n \in \mathbb{N}$,

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\overline{R_n^{\Phi}}^{+}\right\|_{\infty}\right] \leq \mathbb{E}_{\sigma,\tau}\left[\left\|\overline{R_n^{\Phi}}^{+}\right\|_{2}\right] \leq \sqrt{\frac{A_{\Phi}}{n}},$$

where

$$A_{\Phi} = \max_{a\in\mathcal{A}}\left|\left\{\phi \in \Phi \text{ s.t. } \phi(a) \neq a\right\}\right|.$$

Also, for every $\eta > 0$,

$$\mathbb{P}_{\sigma,\tau}\left\{\sup_{m\geq n}\|\overline{R_n^{\Phi}}^{+}\| - \sqrt{\frac{A_{\Phi}}{n}} \geq \eta\right\} \leq \frac{4A_{\Phi}}{\eta^2}\exp\left(-\frac{\eta^2 n}{2A_{\Phi}}\right).$$

Proof

- We prove the following geometric property:

  Any invariant measure λ of a matrix $\Theta^{\Phi}(M)$ with non-negative coefficient satisfies, no matter the choice of $b \in \mathcal{B}$, $\left\langle M, \mathbb{E}_{\lambda}[R^{\Phi}(a, b)] \right\rangle = 0$.

- As a consequence, σ is Blackwell's approachability strategy of the negative orthant.

The Repeated Game

- **A repeated game between the Player and Nature**.
- At each stage n, Nature chooses a state of the world ω_n in some finite set Ω.
- The **player predicts** it by choosing a **probability distribution** $p_n \in \Delta(\Omega)$.
- A mixed strategy of Nature is a mapping from the set of finite histories $\cup_{n \in \mathbb{N}} (\Omega \times \Delta(\Omega))^n$ into $\Delta(\Omega)$.
- **A mixed strategy of the player** is a mapping from the set of finite histories $\cup_{n \in \mathbb{N}} (\Omega \times \Delta(\Omega))^n$ into $\Delta(\Delta(\Omega))$.

The Meteorologist Example

- The usual example consists in a meteorologist predicting each day the probability of rain, which corresponds to $\Omega = \{0, 1\}$, where $\omega = 1$ if it rains.
- The law of ω_n, chosen by Nature, can depend on the previous realization $\omega_1, \ldots, \omega_{n-1}$ as well as the past predictions $p_1, \ldots, p_{n-1}$ (and possibly the law of p_n).

Notations

- For $p \in \Delta(\Omega)$ and $\varepsilon > 0$, let $\mathbb{N}_n[p, \varepsilon]$ **be the set of stages where the prediction was ε-close to** p:

$$\mathbb{N}_n[p, \varepsilon] = \Big\{ m \in \{1, \ldots, n\} \text{ s.t. } \|p_m - p\| \leq \varepsilon \Big\},$$

  where $\| \cdot \|$ is the Euclidean norm of $\mathbb{R}^\Omega$.
- We denote by $\overline{\omega}_n[p, \varepsilon] \in \Delta(\Omega)$ **the empirical distribution of states on** $\mathbb{N}_n[p, \varepsilon]$
- We denote by $\overline{p}_n[p, \varepsilon]$ the average prediction on $\mathbb{N}_n[p, \varepsilon]$.

Finite calibration

Definition

A strategy σ of the player is ε-calibrated if for every strategy τ of Nature, and for every $p \in \Delta(\Omega)$,

$$\limsup_{n\to\infty} \frac{|\mathbb{N}_n[p,\varepsilon]|}{n} \left(\left\| \overline{p}_n[p,\varepsilon] - \overline{\omega}_n[p,\varepsilon] \right\| - \varepsilon \right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

A strategy is calibrated if it is ε-calibrated, for every $\varepsilon > 0$.

Foster & Vohra 1997

We assume here that the player only makes predictions in some given grid $\{p[\ell]\ ;\ \ell \in \mathcal{L}\}$ of $\Delta(\Omega)$.

Theorem

There exists a mixed strategy σ calibrated with respect to the grid $\{p[\ell]\,;\ell \in \mathcal{L}\}$, such that, against any strategy τ of Nature,

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{\ell\in\mathcal{L}} \frac{|\mathbb{N}_n[\ell]|}{n}\left(\left\|\overline{\omega}_n[\ell]-p[\ell]\right\|^2 - \min_{k\in\mathcal{L}}\left\|\overline{\omega}_n[\ell]-p[k]\right\|^2\right)\right] \leq \frac{6}{\delta(\mathcal{L})}\sqrt{\frac{\log(L)}{n}}$$

where $\delta(\mathcal{L}) = \inf_{\ell\neq k\in\mathcal{L}}\left\|p[\ell]-p[k]\right\|$ is the diameter of the grid.

Proof

- The proof uses the fact that, for any sequence ω_m and every $\ell, k \in \mathcal{L}$,

$$\sum_{m \in \mathbb{N}_n[\ell]} \frac{\|\omega_m - p[\ell]\|^2}{|\mathbb{N}_n[\ell]|} - \frac{\|\omega_m - p[k]\|^2}{|\mathbb{N}_n[\ell]|}$$

$$= \left\|\overline{\omega}_n[\ell] - p[\ell]\right\|^2 - \left\|\overline{\omega}_n[\ell] - p[k]\right\|^2 .$$

- Consider the **repeated game between Player and Nature**, with action space $\mathcal{L}$ for Player and Ω for Nature and where choices of ℓ and ω generate the **real valued payoff function**

$$\rho(\ell, \omega) = -\|\omega - p[\ell]\|^2$$

- An internally consistent strategy satisfies, by definition,

$$\limsup_{n\to\infty} \sup_{\ell,k} \frac{|\mathbb{N}_n[\ell]|}{n} \left(\sum_{m \in \mathbb{N}_n[\ell]} \frac{\left\|\omega_m - p[\ell]\right\|^2}{|\mathbb{N}_n[\ell]|} - \frac{\left\|\omega_m - p[k]\right\|^2}{|\mathbb{N}_n[\ell]|} \right) \leq 0.$$

Impossibility of deterministic calibration: Oakes and Dawid

- **No ε-deterministic calibrated strategies exist.**
- Let $\Omega = \{0,1\}$ and define the strategy of Nature as follows: given the past history h^n,

$$\text{if } p_{n+1} \geq \frac{1}{2} \text{ then } \omega_{n+1} = 0 \text{ and if } p_{n+1} < \frac{1}{2} \text{ then } \omega_{n+1} = 1.$$

- In words, if the forecaster claims that it will rain with high probability then Nature does not make it rain and if he claims that it will not rain, Nature makes it rain.
- This prevents any deterministic strategies from being ε-calibrated.

Smooth calibration: Foster and Kakade

- Given $F \subset \Delta(\Omega)$, the calibration score can be written as

$$\frac{|\mathbb{N}_n[F]|}{n} \left\| \overline{\omega}_n[F] - \overline{p}_n[F] \right\| = \frac{1}{n} \left\| \sum_{m=1}^{n} \mathbb{1}\{p_m \in F\}(\omega_m - p_m) \right\|$$

- Observe that the mapping $p \mapsto \mathbb{1}\{p \in F\}$ is discontinuous.
- Given some continuous $g : \Delta(\Omega) \to [0, 1]$, consider the following smooth version of the score

$$\frac{1}{n} \left\| \sum_{m=1}^{n} g(p_m)(\omega_m - p_m) \right\|$$

Foster and Kakade

Theorem

There exists a deterministic strategy σ of the player that is smooth calibrated, that is, against any strategy of Nature, for every continuous mapping $g : \Delta(\Omega) \to \mathbb{R}_+$,

$$\limsup_{n\to\infty} \frac{1}{n} \left\| \sum_{m=1}^{n} g(p_m)(\omega_m - p_m) \right\| \leq 0 \ .$$

Proof: Approachability –with activation– in infinite dimension.

Reference

Vianney Perchet (2014). Approachability, Regret and Calibration: Implications and Equivalence. *Journal of Dynamics and Games*, **1**, 181-254.