

Minimization of regrets and convergence to sets of equilibria

Gilles Stoltz

CNRS – École normale supérieure – HEC Paris

Two procedures to ensure convergence

Convergence results are of two kinds.

- 1 Players independently minimize some notion of **regret** (the strategy to do so is up to them). If it happens that they **all** minimize **simultaneously** a given notion, then some convergence takes place.
- 2 Players must play a **common** given **strategy**; if they simultaneously do so, then some convergence takes place.

Minimization of regrets

In the case of a general K -players game,

- if all players minimize their **external** regrets, then the empirical distributions of their action profiles converge to the **Hannan set**;
- if all players minimize their **internal** regrets, then the empirical distributions of their action profiles converge to the set of **correlated equilibria**.

In the case of a zero-sum 2-players game,

- if the two players minimize their **external** regrets, then the couples of the **marginals** of the empirical distributions of their action profiles converge to the **set of minimax** (= Nash) equilibria.



Procedures to be played simultaneously

If all players play the **regret-matching** procedure of Hart and Mas-Colell, then the empirical distributions of their action profiles converge to the set of **correlated equilibria**.

If all players play the **regret-testing** procedure of Foster and Young, see also Germano and Lugosi, then the mixed action profiles themselves converge to the set of **Nash equilibria**, for almost all games.

A base one-shot game is repeated

K players play a game.

Player k takes actions $I_1^{(k)}, I_2^{(k)}, \dots$ from a **finite** set $\mathcal{X}^{(k)} = \{1, \dots, N_k\}$.

We denote $\mathcal{X} = \mathcal{X}^{(1)} \times \dots \times \mathcal{X}^{(K)}$ the set of **action profiles**.

The payoff function for player k is $r^{(k)} : \mathcal{X} \rightarrow [0, 1]$.

We assume that **full monitoring** is possible, i.e., all players observe the action profile

$$I_t = \left(I_t^{(1)}, \dots, I_t^{(K)} \right)$$

at the end of each round t .

To extend the results to the case of **bandit** feedback, when some players are only informed with their own payoff $r^{(k)}(I_t)$, it suffices to perform **estimation**.



Strategies for the players

A (randomized) strategy $\sigma^{(k)}$ for player k is a **sequence of functions**.

The t -th of them, associates

- to the past action profiles I_s , for $s = 1, \dots, t - 1$,
- a probability distribution $\mathbf{p}_t^{(k)} = (p_{1,t}^{(k)}, \dots, p_{N_k,t}^{(k)})$ on the set $\mathcal{X}^{(k)} = \{1, \dots, N_k\}$ of actions.

The played action $I_t^{(k)}$ is drawn at random according to $\mathbf{p}_t^{(k)}$.

Convergence of ... to sets of equilibria

We define

- the empirical distributions of **action profiles**,

$$\bar{\pi}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t} \in \Delta(\mathcal{X}) ;$$

- the couples of the **marginals** of the empirical distributions of the action profiles (in case of a 2-player game),

$$\left(\bar{\mathbf{p}}_n^{(1)}, \bar{\mathbf{p}}_n^{(2)} \right) \quad \text{where}$$

$$\bar{\mathbf{p}}_n^{(1)} = \frac{1}{n} \sum_{t=1}^n \delta_{I_t^{(1)}} \quad \text{and} \quad \bar{\mathbf{p}}_n^{(2)} = \frac{1}{n} \sum_{t=1}^n \delta_{I_t^{(2)}} ;$$

- the **mixed action profiles**,

$$\mathbf{p}_t^{(1)} \otimes \dots \otimes \mathbf{p}_t^{(K)} \in \Delta_{\text{ind}}(\mathcal{X}) = \Delta(\mathcal{X}^{(1)}) \times \dots \times \Delta(\mathcal{X}^{(K)}) .$$



Convergence of some empirical quantities to ... (1)

$(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(K)}) \in \Delta_{\text{ind}}(\mathcal{X})$ is a Nash equilibrium if for all k and all $\mathbf{q}^{(k)} \in \Delta(\mathcal{X}^{(k)})$,

$$r^{(k)}(\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(k)}, \dots, \mathbf{p}^{(K)}) \geq r^{(k)}(\mathbf{p}^{(1)}, \dots, \mathbf{q}^{(k)}, \dots, \mathbf{p}^{(K)})$$

where we linearly extended $r^{(k)}$.

In case of a **zero-sum 2-players game**, i.e., $K = 2$ and $r^{(1)} = -r^{(2)}$, these inequalities take a simple form and the Nash equilibria are called minimax equilibria:

$(\mathbf{p}^{(1)}, \mathbf{p}^{(2)})$ is a minimax equilibrium if

$$\sup_{\mathbf{q}^{(1)}} r^{(1)}(\mathbf{q}^{(1)}, \mathbf{p}^{(2)}) \leq v = r^{(1)}(\mathbf{p}^{(1)}, \mathbf{p}^{(2)}) \leq \inf_{\mathbf{q}^{(2)}} r^{(1)}(\mathbf{p}^{(1)}, \mathbf{q}^{(2)}) .$$

The **value** of the game v is independent of the chosen equilibrium.



Convergence of some empirical quantities to ... (2)

The set of **correlated equilibria** is formed by the following joint distributions,

$$\mathcal{C} = \left\{ \pi \in \Delta(X) : \quad \forall K \text{ and } \varphi^{(k)} \right. \\ \left. \sum_i \pi(i) r^{(k)}(i^{(k)}, i^{(-k)}) \geq \sum_i \pi(i) r^{(k)}(\varphi^{(k)}(i^{(k)}), i^{(-k)}) \right\}$$

where $i \in \mathcal{X}$ is denoted by

$$i = (i^{(1)}, \dots, i^{(k)}, \dots, i^{(K)}) = (i^{(k)}, i^{(-k)}) .$$

Note:

It suffices to consider $\varphi^{(k)}$ that only differ from identity in one point.

Convergence of some empirical quantities to ... (3)

The **Hannan set** is formed by the following joint distributions,

$$\mathcal{H} = \left\{ \pi \in \Delta(\mathcal{X}) : \forall K \text{ and } j \in \mathcal{X}^{(k)} \right. \\ \left. \sum_i \pi(i) r^{(k)}(i^{(k)}, i^{(-k)}) \geq \sum_i \pi(i) r^{(k)}(j, i^{(-k)}) \right\}$$

where $i \in \mathcal{X}$ is denoted by

$$i = (i^{(1)}, \dots, i^{(k)}, \dots, i^{(K)}) = (i^{(k)}, i^{(-k)}) .$$

Note:

The Hannan set will correspond to **external regret** and the set of correlated equilibria to **internal regret**.

Summary

The sets \mathcal{N} of Nash equilibria, \mathcal{C} of correlated equilibria, and the Hannan set \mathcal{H} are nested,

$$\mathcal{N} \subseteq \mathcal{C} \subseteq \mathcal{H}$$

We will see the following convergence results,

- $\bar{\pi}_n \rightarrow \mathcal{H}$;
- $\bar{\pi}_n \rightarrow \mathcal{C}$;
- $(\bar{\mathbf{p}}_n^{(1)}, \bar{\mathbf{p}}_n^{(2)}) \rightarrow \mathcal{N}$ in case of a zero-sum 2-players game;
- $\mathbf{p}_t^{(1)} \otimes \dots \otimes \mathbf{p}_t^{(K)} \rightarrow \mathcal{N}$ for almost all games.

The first three convergences can be ensured in an **efficient way**.

External regret

We have seen last time how players can minimize their **external regrets**,

$$R_n^{(k)} = \max_{j=1, \dots, N_k} \sum_{t=1, \dots, n} r^{(k)}(j, I_t^{(-k)}) - \sum_{t=1, \dots, n} r^{(k)}(I_t^{(k)}, I_t^{(-k)}) .$$

If $R_n^{(k)} = o(n)$ a.s. for all k , then

$$\bar{\pi}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t} \rightarrow \mathcal{H} \quad \text{a.s.}$$

Actually, each $\bar{\pi}_n$ is an ε_n -**Hannan equilibrium**, where $\varepsilon_n = \max R_n^{(k)} / n$.

Zero-sum two-players games

In this case, $r^{(1)} = -r^{(2)} = r$.

If the first player minimizes his external regret, then his **average payoff** is at least the **value v** of the base one-shot game,

$$\liminf_{n \rightarrow \infty} \frac{\sum_{t=1}^n r(I_t)}{n} \geq v = \min_{\mathbf{q}} \max_{\mathbf{p}} r(\mathbf{p}, \mathbf{q}) .$$

Proof: We denote by $\bar{\mathbf{q}}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t^{(2)}}$ the empirical distribution of plays of the opponent player,

$$\begin{aligned} \liminf \frac{\sum_{t=1}^n r(I_t)}{n} &\geq \liminf \max_{j=1, \dots, N_1} \frac{\sum_{t=1}^n r(j, I_t^{(2)})}{n} = \liminf \max_{j=1, \dots, N} r(j, \bar{\mathbf{q}}_n) \\ &= \liminf \max_{\mathbf{p}} r(\mathbf{p}, \bar{\mathbf{q}}_n) \geq \min_{\mathbf{q}} \max_{\mathbf{p}} r(\mathbf{p}, \mathbf{q}) = v . \end{aligned}$$

If **both** players use such strategies, then $(\sum_{t=1}^n r(I_t))/n \rightarrow v$.



Zero-sum two-players games

As a **consequence**, if both players minimize their external regrets, denoting by

$$\bar{\mathbf{p}}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t^{(1)}} \quad \text{and} \quad \bar{\mathbf{q}}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t^{(2)}}$$

the **marginals** of the empirical distributions of action profiles, we have

$$(\bar{\mathbf{p}}_n, \bar{\mathbf{q}}_n) \rightarrow \mathcal{N} .$$

Proof: We get equalities in the proof of the previous slide, and by symmetry, since $r^{(1)} = -r^{(2)} = r$:

$$\lim_{n \rightarrow \infty} \max_{\mathbf{p}} r(\mathbf{p}, \bar{\mathbf{q}}_n) = v = \lim_{n \rightarrow \infty} \min_{\mathbf{q}} r(\bar{\mathbf{p}}_n, \mathbf{q})$$

and thus by a sandwich argument,

$$r(\bar{\mathbf{p}}_n, \bar{\mathbf{q}}_n) \longrightarrow v .$$



Internal regret

It studies how **constant replacements** of one action by another one perform,

$$S_n^{(k)} = \max_{i \neq j} \widehat{X}_{n,i \rightarrow j}^{(k)} \quad \text{where}$$

$$\widehat{X}_{n,i \rightarrow j}^{(k)} = \sum_{t=1, \dots, n} \mathbb{I}_{\{I_t^{(k)}=i\}} \left(r^{(k)}(i, I_t^{(-k)}) - r^{(k)}(j, I_t^{(-k)}) \right).$$

If $S_n^{(k)} = o(n)$ a.s. for all k , then

$$\bar{\pi}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t} \rightarrow \mathcal{C} \quad \text{a.s.}$$

Actually, each $\bar{\pi}_n$ is an ε_n -**Hannan equilibrium**, where $\varepsilon_n = N \max S_n^{(k)} / n$.

A simple trick to minimize internal regret

The trick is described in Stoltz and Lugosi '05. A related trick was found independently by Blum and Mansour '07. We simply use the exponentially weighted average predictor on the **pseudo-actions** $i \rightarrow j$ and solve a **fixed-point** equation.

More precisely, we choose at rounds $t \geq 2$, the distribution $\mathbf{p}_t^{(k)}$ such that

$$\mathbf{p}_t^{(k)} = \sum_{i \neq j} \frac{\exp\left(\eta \widehat{X}_{t-1, i \rightarrow j}^{(k)}\right)}{\sum_{i' \neq j'} \exp\left(\eta \widehat{X}_{t-1, i' \rightarrow j'}^{(k)}\right)} \mathbf{p}_{t, i \rightarrow j}^{(k)}$$

where $\mathbf{p}_{t, i \rightarrow j}^{(k)}$ is the same as $\mathbf{p}_t^{(k)}$ except for the i -th component, which equals 0, and for the j -th component.

Since there are $N(N - 1)$ pseudo-predictors, the bound on **internal** regret can be deduced from the usual one on **external** regret and is of the order of $\sqrt{n \ln N}$ in expectation.



Regret-matching (Hart and Mas-Colell)

It relies on a parameter $\mu > 0$ and consists, for player k , in choosing $\mathbf{p}_t^{(k)}$ given by

$$p_{j,t}^{(k)} = \begin{cases} \frac{1}{\mu} \left(X_{t-1,i \rightarrow j}^{(k)} \right)^+ & \text{if } j \neq i \\ 1 - \frac{1}{\mu} \sum_{j' \neq i} \left(X_{t-1,i \rightarrow j'}^{(k)} \right)^+ & \text{if } j = i \end{cases}$$

where we denoted $i = I_{t-1}^{(k)}$.

In some sense, at every stage, one tests whether some **internal regret** is suffered but does not minimize it.

If **all players** use this procedure (and μ is large enough), then

$$\bar{\pi}_n = \frac{1}{n} \sum_{t=1}^n \delta_{I_t} \rightarrow \mathcal{C} \quad \text{a.s. .}$$

Note: No quantification possible this time.



Regret-testing (Foster and Young, Germano and Lugosi)

This procedure basically works by **random search**.

It basically works as follows, for player k .

- 1 Before round 1: Choose a distribution on $\mathcal{X}^{(k)}$ **at random**.
- 2 Rounds 1 to $m - 1$: Play randomly according to it (a fixed number $m - 1$ of times).
- 3 Round m : **Test** whether some external regret is suffered (up to a tolerance factor ρ); play $I_m = 1$ or $I_m = 2$ to **signal** the result of the test.
- 4 If all players signalled that everything is fine, they **stick** to the last randomly chosen distribution forever.
- 5 Otherwise, **start again**.



Regret-testing (Foster and Young, Germano and Lugosi)

This amounts to random search: by **pure chance**, a Nash equilibrium is achieved at some point.

By setting m and ρ adaptively and with some other modifications, one can prove that for almost all games,

$$\mathbf{p}_t^{(1)} \otimes \dots \otimes \mathbf{p}_t^{(K)} \rightarrow \mathcal{N} .$$

Note: The procedure is **unefficient**. For instance, in case a pure Nash equilibrium exists, convergence is reached in a time amount proportional to the exponential complexity of finding it analytically.

