Online Learning in Games

Rida Laraki and Guillaume Vigeral

CNRS, PSL

IASD, Lecture 6

Contents

Example 1

	L	R
T	$3, 1$	$-10, -10$
B	$0, 0$	$1, 3$

Example 1

	L	R
T	$3, 1$	$-10, -10$
B	$0, 0$	$1, 3$

- **Without communication**, There are two pure, efficient and disymmetrical Nash equilibrium payoffs $(3, 1)$ and $(1, 3)$, and one mixed and innefficient Nash equilibrium payoff $(\frac{3}{14}, \frac{3}{14})$.

Example 1

	L	R
T	$3, 1$	$-10, -10$
B	$0, 0$	$1, 3$

- **Without communication**, There are two pure, efficient and disymmetrical Nash equilibrium payoffs $(3, 1)$ and $(1, 3)$, and one mixed and innefficient Nash equilibrium payoff $(\frac{3}{14}, \frac{3}{14})$.
- **With communication**, The use of a public coin allows us to get a symmetrical and efficient outcome as an equilibrium payoff : if the coin shows “heads”, player 1 plays T, player 2 plays L, the outcome is $(3, 1)$ and if the coin shows “tails”, player 1 plays B, player 2 plays R, inducing $(1, 3)$.

Example 1

	L	R
T	$3, 1$	$-10, -10$
B	$0, 0$	$1, 3$

- **Without communication**, There are two pure, efficient and disymmetrical Nash equilibrium payoffs $(3, 1)$ and $(1, 3)$, and one mixed and innefficient Nash equilibrium payoff $(\frac{3}{14}, \frac{3}{14})$.
- **With communication**, The use of a public coin allows us to get a symmetrical and efficient outcome as an equilibrium payoff : if the coin shows "heads", player 1 plays T, player 2 plays L, the outcome is $(3, 1)$ and if the coin shows "tails", player 1 plays B, player 2 plays R, inducing $(1, 3)$.
- Facing such a plan, no deviation is profitable. The expected profit of this correlated equilibrium is $(2, 2)$.

Example 1

	L	R
T	$3, 1$	$-10, -10$
B	$0, 0$	$1, 3$

$Q =$

$1/2$	0
0	$1/2$

- **Without communication**, There are two pure, efficient and disymmetrical Nash equilibrium payoffs $(3, 1)$ and $(1, 3)$, and one mixed and innefficient Nash equilibrium payoff $(\frac{3}{14}, \frac{3}{14})$.
- **With communication**, The use of a public coin allows us to get a symmetrical and efficient outcome as an equilibrium payoff : if the coin shows "heads", player 1 plays T, player 2 plays L, the outcome is $(3, 1)$ and if the coin shows "tails", player 1 plays B, player 2 plays R, inducing $(1, 3)$.
- Facing such a plan, no deviation is profitable. The expected profit of this correlated equilibrium is $(2, 2)$.
- This device induces the distribution $Q \in \Delta\left(\{T, L\} \times \{L, R\}\right)$

Example 2

	L	R
T	2, 7	6, 6
B	0, 0	7, 2

Example 2

	L	R
T	2, 7	6, 6
B	0, 0	7, 2

- **Without communication :** three equilibria with payoffs respectively $(2, 7)$, $(7, 2)$ and $(\frac{14}{3}, \frac{14}{3})$.

Example 2

	L	R
T	2, 7	6, 6
B	0, 0	7, 2

- **Without communication :** three equilibria with payoffs respectively $(2, 7)$, $(7, 2)$ and $(\frac{14}{3}, \frac{14}{3})$.
- **With communication :** players can obtain any point in the convex hull of : $(2, 7)$, $(7, 2)$ and $(\frac{14}{3}, \frac{14}{3})$.

Example 2 with mediation

	L	R
T	$2,7$	$6,6$
B	$0,0$	$7,2$

Example 2 with mediation

	L	R
T	2, 7	6, 6
B	0, 0	7, 2

- **With mediation :** suppose a benevolent mediator selects a message m in the set (X, Y, Z) uniformly (proba $(1/3, 1/3, 1/3)$).

Example 2 with mediation

	L	R
T	2, 7	6, 6
B	0, 0	7, 2

- **With mediation :** suppose a benevolent mediator selects a message m in the set (X, Y, Z) uniformly (proba $(1/3, 1/3, 1/3)$).
- The mediator send privately to each player a signal.

Example 2 with mediation

	L	R
T	$2,7$	$6,6$
B	$0,0$	$7,2$

- **With mediation :** suppose a benevolent mediator selects a message m in the set (X, Y, Z) uniformly (proba $(1/3, 1/3, 1/3)$).
- The mediator send privately to each player a signal.
- Player 1 receives signal a if $m \in \{X, Y\}$ and signal b if $m = Z$.
- Player 2 receives signal α if $m = X$ and signal β if $m \in \{Y, Z\}$.

Example 2

- Player 1 receives a if $m \in \{X, Y\}$ and b if $m = Z$.
- Player 2 receives α if $m = X$ and β if $m \in \{Y, Z\}$.

Example 2

- Player 1 receives a if $m \in \{X, Y\}$ and b if $m = Z$.
- Player 2 receives α if $m = X$ and β if $m \in \{Y, Z\}$.
- Let player 1 plays T if the signal is a and B otherwise.

Example 2

- Player 1 receives a if $m \in \{X, Y\}$ and b if $m = Z$.
- Player 2 receives α if $m = X$ and β if $m \in \{Y, Z\}$.
- Let player 1 plays T if the signal is a and B otherwise.
- Let player 2 plays L if the signal is α and R otherwise.

Example 2

- Player 1 receives a if $m \in \{X, Y\}$ and b if $m = Z$.
- Player 2 receives α if $m = X$ and β if $m \in \{Y, Z\}$.
- Let player 1 plays T if the signal is a and B otherwise.
- Let player 2 plays L if the signal is α and R otherwise.
- The induced distribution is.

$1/3$	$1/3$
0	$1/3$

Example 2

- Player 1 receives a if $m \in \{X, Y\}$ and b if $m = Z$.
- Player 2 receives α if $m = X$ and β if $m \in \{Y, Z\}$.
- Let player 1 plays T if the signal is a and B otherwise.
- Let player 2 plays L if the signal is α and R otherwise.
- The induced distribution is.

$1/3$	$1/3$
0	$1/3$

- A correlated equilibrium : no player has incentive to deviate.

Example 2

- Player 1 receives *a* if $m \in \{X, Y\}$ and *b* if $m = Z$.
- Player 2 receives α if $m = X$ and β if $m \in \{Y, Z\}$.
- Let player 1 plays T if the signal is a and B otherwise.
- Let player 2 plays L if the signal is α and R otherwise.
- The induced distribution is.

$1/3$	$1/3$
0	$1/3$

- A correlated equilibrium : no player has incentive to deviate.
- The corresponding outcome $(5, 5)$ Pareto dominates the set of symmetrical Nash outcomes.

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

- **A probability space** $(\Omega, \mathcal{C}, P)$.

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

- **A probability space** $(\Omega, \mathcal{C}, P)$.
- **A family of measurable maps** θ^i from $(\Omega, \mathcal{C})$ to $(A^i, \mathcal{A}^i)$ (measurable set of signals of player i).

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

- **A probability space** $(\Omega, \mathcal{C}, P)$.
- **A family of measurable maps** θ^i from $(\Omega, \mathcal{C})$ to $(A^i, \mathcal{A}^i)$ (measurable set of signals of player i).

Definition

Let G, defined by $g : S = \Pi_{i \in I} S^i \to \mathbf{R}^n$, be a strategic game. G ***extended* by $\mathcal{I}$, denoted** $[G, \mathcal{I}]$, is the 2 stages game :

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

- **A probability space** $(\Omega, \mathcal{C}, P)$.
- **A family of measurable maps** θ^i from $(\Omega, \mathcal{C})$ to $(A^i, \mathcal{A}^i)$ (measurable set of signals of player i).

Definition

Let G, defined by $g : S = \Pi_{i \in I} S^i \to \mathbf{R}^n$, be a strategic game. G ***extended*** **by** $\mathcal{I}$**, denoted** $[G, \mathcal{I}]$, is the 2 stages game :

- **Stage 0** : the random variable ω is selected according to the law P and the signal $\theta^i(\omega)$ is sent privately to player i.

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

- **A probability space** $(\Omega, \mathcal{C}, P)$.
- **A family of measurable maps** θ^i from $(\Omega, \mathcal{C})$ to $(A^i, \mathcal{A}^i)$ (measurable set of signals of player i).

Definition

Let G, defined by $g : S = \Pi_{i \in I} S^i \to \mathbf{R}^n$, be a strategic game. G ***extended* by** $\mathcal{I}$**, denoted** $[G, \mathcal{I}]$, is the 2 stages game :

- **Stage 0** : the random variable ω is selected according to the law P and the signal $\theta^i(\omega)$ is sent privately to player i.
- **Stage 1 :** the players play in the game G.

Information structures and extended games

Definition

An **information structure** $\mathcal{I}$ is defined by :

- **A probability space** $(\Omega, \mathcal{C}, P)$.
- **A family of measurable maps** θ^i from $(\Omega, \mathcal{C})$ to $(A^i, \mathcal{A}^i)$ (measurable set of signals of player i).

Definition

Let G, defined by $g : S = \Pi_{i \in I} S^i \to \mathbf{R}^n$, be a strategic game. G ***extended*** **by** $\mathcal{I}$**, denoted** $[G, \mathcal{I}]$, is the 2 stages game :

- **Stage 0** : the random variable ω is selected according to the law P and the signal $\theta^i(\omega)$ is sent privately to player i.
- **Stage 1 :** the players play in the game G.

A strategy σ^i of player i is a measurable map $(A^i, \mathcal{A}^i) \to (S^i, \mathcal{S}^i)$.

Correlated Equilibrium

The payoff corresponding to a profile σ is

$$\gamma[G,\mathcal{I}](\sigma) = \int_\Omega g(\sigma(\omega))P(\mathrm{d}\omega).$$

A correlated equilibrium of G is a Nash equilibrium of $[G,\mathcal{I}]$.

Correlated Equilibrium

The payoff corresponding to a profile σ is

$$\gamma[G,\mathcal{I}](\sigma) = \int_\Omega g(\sigma(\omega))P(\mathrm{d}\omega).$$

A correlated equilibrium of G is a Nash equilibrium of $[G,\mathcal{I}]$.

- For each signal a^i, player i forms a belief $\sigma_{-i}(a^i) \in \Delta(S^{-i})$ about the strategy profil of the opponents.

Correlated Equilibrium

The payoff corresponding to a profile σ is

$$\gamma[G,\mathcal{I}](\sigma) = \int_\Omega g(\sigma(\omega))P(\mathrm{d}\omega).$$

A correlated equilibrium of G is a Nash equilibrium of $[G,\mathcal{I}]$.

- For each signal a^i, player i forms a belief $\sigma_{-i}(a^i) \in \Delta(S^{-i})$ about the strategy profil of the opponents.
- At equilibrium, $\sigma_i(a^i)$ is a best reply against $\sigma_{-i}(a^i)$.

Correlated Equilibrium Distributions

A profil σ of strategies in $[G, \mathcal{I}]$ maps the probability P on Ω to an image probability $Q(\sigma)$ on S

$$Q(\sigma)(s) = \sum_{\omega} p(\omega) \prod_{i} \sigma_i(\theta_i(\omega))(s_i)$$

Correlated Equilibrium Distributions

A profil σ of strategies in $[G, \mathcal{I}]$ maps the probability P on Ω to an image probability $Q(\sigma)$ on S

$$Q(\sigma)(s) = \sum_{\omega} p(\omega) \prod_{i} \sigma_i(\theta_i(\omega))(s_i)$$

$\mathrm{CED}(G)$ is the set of *correlated equilibrium distributions* in G :

$$\mathrm{CED}(G) = \bigcup_{\mathcal{I},\sigma} \{Q(\sigma);\ \sigma \text{ equilibrium in } [G, \mathcal{I}]\}.$$

Canonical correlation

A **canonical information structure** $\mathcal{I}$ for G corresponds to the framework where :

Canonical correlation

A **canonical information structure** $\mathcal{I}$ for G corresponds to the framework where :

- the underlying space is $\Omega = S$.

Canonical correlation

A **canonical information structure** $\mathcal{I}$ for G corresponds to the framework where :

- the underlying space is $\Omega = S$.
- the signal space of player i is $A^i = S^i$.

Canonical correlation

A **canonical information structure** $\mathcal{I}$ for G corresponds to the framework where :

- the underlying space is $\Omega = S$.
- the signal space of player i is $A^i = S^i$.
- the signaling function, $\theta^i : S \to S^i$ is $\theta^i(s) = s^i$ for all $i \in I$.

Canonical correlation

A **canonical information structure** $\mathcal{I}$ for G corresponds to the framework where :

- the underlying space is $\Omega = S$.
- the signal space of player i is $A^i = S^i$.
- the signaling function, $\theta^i : S \to S^i$ is $\theta^i(s) = s^i$ for all $i \in I$.

A **canonical correlated equilibrium** (CCE) is a Nash equilibrium of the game G extended by a canonical information structure $\mathcal{I}$ and where the equilibrium strategies are given by

$$\sigma^i(\omega) = \sigma^i(s) = \sigma^i(s^i) = s^i.$$

Canonical correlation

A **canonical information structure** $\mathcal{I}$ for G corresponds to the framework where :

- the underlying space is $\Omega = S$.
- the signal space of player i is $A^i = S^i$.
- the signaling function, $\theta^i : S \to S^i$ is $\theta^i(s) = s^i$ for all $i \in I$.

A **canonical correlated equilibrium** (CCE) is a Nash equilibrium of the game G extended by a canonical information structure $\mathcal{I}$ and where the equilibrium strategies are given by

$$\sigma^i(\omega) = \sigma^i(s) = \sigma^i(s^i) = s^i.$$

Interpretation : mediator selects $s = (s_1, ..., s_n) \in S$ using Q, informs each player i about his own recommended action s_i. At equilibrium, each player has interest to follow the recommendation and $Q = P$.

Result 1

Theorem

Let σ be an equilibrium of $[G, \mathcal{I}]$ and $Q = Q(\sigma)$ the induced distribution on S. Then Q is a canonical correlated equilibrium distribution :

CCED (G)= CED (G)

Result 1

Theorem

Let σ be an equilibrium of $[G, \mathcal{I}]$ and $Q = Q(\sigma)$ the induced distribution on S. Then Q is a canonical correlated equilibrium distribution :

CCED (G)= CED (G)

Proof : Let the mediator gives to each player i less information : the action s^i to play versus the signal a^i such that $\sigma^i(a^i) = s^i$.

Result 1

Theorem

Let σ be an equilibrium of $[G, \mathcal{I}]$ and $Q = Q(\sigma)$ the induced distribution on S. Then Q is a canonical correlated equilibrium distribution :

CCED (G)= CED (G)

Proof : Let the mediator gives to each player i less information : the action s^i to play versus the signal a^i such that $\sigma^i(a^i) = s^i$.

- For each signal a^i such that $\sigma^i(a^i) = s^i$, player i forms a belief $\sigma_{-i}(a^i) \in \Delta(S^{-i})$ about the strategy profile of the other players.

Result 1

Theorem

Let σ be an equilibrium of $[G, \mathcal{I}]$ and $Q = Q(\sigma)$ the induced distribution on S. Then Q is a canonical correlated equilibrium distribution :

CCED (G)= CED (G)

Proof : Let the mediator gives to each player i less information : the action s^i to play versus the signal a^i such that $\sigma^i(a^i) = s^i$.

- For each signal a^i such that $\sigma^i(a^i) = s^i$, player i forms a belief $\sigma_{-i}(a^i) \in \Delta(S^{-i})$ about the strategy profile of the other players.
- At equilibrium, s^i is a best reply to $\sigma_{-i}(a^i)$.

Result 1

Theorem

Let σ be an equilibrium of $[G, \mathcal{I}]$ and $Q = Q(\sigma)$ the induced distribution on S. Then Q is a canonical correlated equilibrium distribution :

CCED (G)= CED (G)

Proof : Let the mediator gives to each player i less information : the action s^i to play versus the signal a^i such that $\sigma^i(a^i) = s^i$.

- For each signal a^i such that $\sigma^i(a^i) = s^i$, player i forms a belief $\sigma_{-i}(a^i) \in \Delta(S^{-i})$ about the strategy profile of the other players.
- At equilibrium, s^i is a best reply to $\sigma_{-i}(a^i)$.
- If player i's information is reduced to s^i, his belief γ^{-i} is a convex combinaison of $\sigma_{-i}(a^i)$.

Result 1

Theorem

Let σ be an equilibrium of $[G, \mathcal{I}]$ and $Q = Q(\sigma)$ the induced distribution on S. Then Q is a canonical correlated equilibrium distribution :

CCED (G)= CED (G)

Proof : Let the mediator gives to each player i less information : the action s^i to play versus the signal a^i such that $\sigma^i(a^i) = s^i$.

- For each signal a^i such that $\sigma^i(a^i) = s^i$, player i forms a belief $\sigma_{-i}(a^i) \in \Delta(S^{-i})$ about the strategy profile of the other players.
- At equilibrium, s^i is a best reply to $\sigma_{-i}(a^i)$.
- If player i's information is reduced to s^i, his belief γ^{-i} is a convex combinaison of $\sigma_{-i}(a^i)$.
- By convexity of BR^i over $\Delta(S^{-i})$, s^i remains a best response given to γ^{-i} .

Characterization

Theorem

$Q \in DEC(G)$ if and only if : $\forall s^i, t^i \in S^i, \forall i = 1, ..., n$:

$$\sum_{s^{-i} \in S^{-i}} [G^i(s^i, s^{-i}) - G^i(t^i, s^{-i})] Q(s^i, s^{-i}) \geq 0.$$

Characterization

Theorem

$Q \in DEC(G)$ if and only if : $\forall s^i, t^i \in S^i, \forall i = 1, ..., n$:

$$\sum_{s^{-i} \in S^{-i}} [G^i(s^i, s^{-i}) - G^i(t^i, s^{-i})] Q(s^i, s^{-i}) \geq 0.$$

Proof :

Characterization

Theorem

$Q \in DEC(G)$ if and only if : $\forall s^i, t^i \in S^i, \forall i = 1, ..., n$:

$$\sum_{s^{-i} \in S^{-i}} [G^i(s^i, s^{-i}) - G^i(t^i, s^{-i})] Q(s^i, s^{-i}) \geq 0.$$

Proof :

- If s^i is announced to i (i.e. $Q^i(s^i) = \sum_{t^{-i}} Q(s^i, t^{-i}) > 0$), at equilibrium **player i must be a best reply agianst the conditional distribution on S^{-i} given his information s^i**,

$$Q(s_{-i}|s^i) = \frac{Q(s^i, s^{-i})}{\sum_{t_{-i}} Q(s^i, t^{-i})} = \frac{Q(s^i, s^{-i})}{Q(s^i)}$$

Corollary

The set of correlated equilibrium distributions is a polytope *(e.g. is the convex hull of finitely many points).*

Corollary

The set of correlated equilibrium distributions is a polytope *(e.g. is the convex hull of finitely many points).*

Proof : It is a subset of $\Delta(S)$ which is defined by finitely many weak linear inequalities.

Corollary

The set of correlated equilibrium distributions is a polytope *(e.g. is the convex hull of finitely many points).*

Proof : It is a subset of $\Delta(S)$ which is defined by finitely many weak linear inequalities.

Remarks :

Corollary

The set of correlated equilibrium distributions is a polytope *(e.g. is the convex hull of finitely many points).*

Proof : It is a subset of $\Delta(S)$ which is defined by finitely many weak linear inequalities.

Remarks :

- Every Nash Equilibrium is a correlated equilibrium.

Corollary

The set of correlated equilibrium distributions is a polytope *(e.g. is the convex hull of finitely many points).*

Proof : It is a subset of $\Delta(S)$ which is defined by finitely many weak linear inequalities.

Remarks :

- Every Nash Equilibrium is a correlated equilibrium.
- There is an elementary proof for existence of a correlated equilibrium (Hart and Mas-Collel).

Corollary

The set of correlated equilibrium distributions is a polytope *(e.g. is the convex hull of finitely many points).*

Proof : It is a subset of $\Delta(S)$ which is defined by finitely many weak linear inequalities.

Remarks :

- Every Nash Equilibrium is a correlated equilibrium.
- There is an elementary proof for existence of a correlated equilibrium (Hart and Mas-Collel).
- There are correlated equilibria outside the convex envelop of Nash equilibria.

Correlated equilibrium via minmax : Hart & Mas-Collel

Let G be a finite strategic two-player game with strategy sets S^1 and S^2 and payoff $g : S = S^1 \times S^2 \longrightarrow \mathbf{R}^2$.

Consider the game Γ which is a **two-player finite zero-sum** game with the strategy set S **for the max player**, the strategy set $L = (S^1)^2 \cup (S^2)^2$ **for the min player** and payoff function γ :

$$\gamma(s; t^i, u^i) = (g^i(t^i, s^{-i}) - g^i(u^i, s^{-i}))\mathbf{1}_{\{t^i = s^i\}}.$$

- By minmax theorem, Γ **has a value** v **& optimal strategies.**
- **Claim :** $v = 0$ and $Q \in \Delta(S)$ is optimal for the max player iff Q is a CCED of G.

Correlated equilibrium distribution via minmax

- Let $\pi \in \Delta(L)$. Define ρ^1, a transition probability on S^1, by

$$\begin{aligned} \rho^1(t^1;u^1) &= \pi(t^1,u^1), \quad \text{if } t^1 \neq u^1, \\ \rho^1(t^1;t^1) &= 1 - \sum_{u^1 \neq t^1} \pi(t^1,u^1). \end{aligned}$$

  Let now μ^1 be a probability on S^1 invariant under ρ^1 :

$$\mu^1(t^1) = \sum_{u^1} \mu^1(u^1)\rho(u^1;t^1).$$

  Define ρ^2 and μ^2 similarly and let $\mu = \mu^1 \times \mu^2$.

Correlated equilibrium distribution via minmax

- We can show that the payoff $\gamma(\mu;\pi)$ can be decomposed into terms of the form

$$\sum_{t^1} \mu^1(t^1) \sum_{u^1} \rho(t^1;u^1)(g^1(t^1,\cdot) - g^1(u^1,\cdot))$$

which implies hat

$$\forall \pi \in \Delta(L), \exists \phi \in \Delta(S) \text{ satisfying } \gamma(\phi,\pi) \geq 0.$$

- This implies existence of a CED for G.
- The construction clearly extends to I players.

Contents

What is learning ?

- **Different notions of equilibria** (Nash, correlated, coarse)

What is learning ?

- **Different notions of equilibria** (Nash, correlated, coarse)
- **Different Information structures** : do players play observe or not the actions of the opponents ? do they observe/know the payoff function of their opponents ?

What is learning ?

- **Different notions of equilibria** (Nash, correlated, coarse)
- **Different Information structures** : do players play observe or not the actions of the opponents ? do they observe/know the payoff function of their opponents ?
- Do the players **learn** the equilibrium, or do they **converge to play** an equilibrium.

What is learning ?

- **Different notions of equilibria** (Nash, correlated, coarse)
- **Different Information structures** : do players play observe or not the actions of the opponents ? do they observe/know the payoff function of their opponents ?
- Do the players **learn** the equilibrium, or do they **converge to play** an equilibrium.
- Convergence in which sense ? and in which class of games ?

Locking for a robust learning procedure

- Players $\{1, ..., n\}$, with action sets S^i, and payoff function $g^i : S \to [0, 1]$

Locking for a robust learning procedure

- Players $\{1, ..., n\}$, with action sets S^i, and payoff function $g^i : S \to [0, 1]$
- We will define a learning procedure for player 1 such that

- He does not know the set of adversaries
- Nor does he know their set of actions
- Only his own payoff function

Locking for a robust learning procedure

- Players $\{1, ..., n\}$, with action sets S^i, and payoff function $g^i : S \to [0, 1]$
- We will define a learning procedure for player 1 such that

  - He does not know the set of adversaries
  - Nor does he know their set of actions
  - Only his own payoff function

- At each step $t \in N$, P1 chooses $s_t^1 \in S^1$, the other players choose $s_t^{-1} \in \prod_{i \neq 1} S^i$
- P1 does not observe s_t^{-1} but only :

$$\text{the vector } U_t = G^1(\cdot, s_t^{-1}) \in [0, 1]^d$$

where $d = |S^1|$

Locking for a robust learning procedure

- Players $\{1, ..., n\}$, with action sets S^i, and payoff function $g^i : S \to [0, 1]$
- We will define a learning procedure for player 1 such that

  - He does not know the set of adversaries
  - Nor does he know their set of actions
  - Only his own payoff function

- At each step $t \in N$, P1 chooses $s_t^1 \in S^1$, the other players choose $s_t^{-1} \in \prod_{i \neq 1} S^i$
- P1 does not observe s_t^{-1} but only :

$$\text{the vector } U_t = G^1(\cdot, s_t^{-1}) \in [0, 1]^d$$

where $d = |S^1|$

- P1 chooses the next action s_{t+1}^1

The learning model

- A repeated game between a player (action set $\{1, ..., d\}$) against Nature (or aversary).
- **The set of actions of Nature is** $\mathcal{U} \subset [0, 1]^d$.

The learning model

- A repeated game between a player (action set $\{1, ..., d\}$) against Nature (or aversary).
- **The set of actions of Nature is** $\mathcal{U} \subset [0, 1]^d$.
- At each step $t \in N$, nature chooses a vecteur $U_t \in \mathcal{U}$ and J1 chooses (at random) a component $s_t \in \{1, .., d\}$.
- The payoff of J1 is $U_t^{s_t}$.

The learning model

- A repeated game between a player (action set $\{1, ..., d\}$) against Nature (or aversary).
- **The set of actions of Nature is** $\mathcal{U} \subset [0, 1]^d$.
- At each step $t \in N$, nature chooses a vecteur $U_t \in \mathcal{U}$ and J1 chooses (at random) a component $s_t \in \{1, .., d\}$.
- The payoff of J1 is $U_t^{s_t}$.
- We want the strategy of the player to be good against **all possible** strategies/behavior/objectives of nature.

Looking for a strategy without external regret

- $$\limsup_{t\to\infty} \sup_{k\in\{1,..,d\}} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$$

Looking for a strategy without external regret

- $$\limsup_{t\to\infty} \sup_{k\in\{1,..,d\}} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s} - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$$
- $\forall s \in \{1, .., d\}$, $$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s} - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$$

Looking for a strategy without external regret

- $$\limsup_{t\to\infty} \sup_{k\in\{1,..,d\}} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$$
- $\forall s \in \{1,..,d\}$, $$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$$
- **Asymptotically :** each component of $\frac{1}{t}\sum_{\tau=1}^{t} U_\tau$ must be smaller than $\frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau}$, or

$$\frac{1}{t}\sum_{\tau=1}^{t} U_\tau - \left(\frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau}\right) \mathbf{1} \in \mathbb{R}_-^d$$

Looking for a strategy without external regret

- $\limsup_{t\to\infty} \sup_{k\in\{1,..,d\}} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$
- $\forall s \in \{1,..,d\}$, $\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$
- **Asymptotically :** each component of $\frac{1}{t}\sum_{\tau=1}^{t} U_\tau$ must be smaller than $\frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau}$, or

$$\frac{1}{t}\sum_{\tau=1}^{t} U_\tau - \left(\frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau}\right) \mathbf{1} \in \mathbb{R}^d_-$$

- $R_\tau = U_\tau - U_\tau^{s_\tau}\mathbf{1}$, $\bar{R}_t = \frac{1}{t}\sum_{\tau=1}^{t} R_\tau$ converges to $\mathbb{R}^d_-$....

Looking for a strategy without external regret

- $\limsup_{t\to\infty} \sup_{k\in\{1,..,d\}} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$
- $\forall s \in \{1, .., d\}$, $\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^s - \frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau} \leq 0$
- **Asymptotically :** each component of $\frac{1}{t}\sum_{\tau=1}^{t} U_\tau$ must be smaller than $\frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau}$, or

$$\frac{1}{t}\sum_{\tau=1}^{t} U_\tau - \left(\frac{1}{t}\sum_{\tau=1}^{t} U_\tau^{s_\tau}\right) \mathbf{1} \in \mathbb{R}_-^d$$

- $R_\tau = U_\tau - U_\tau^{s_\tau}\mathbf{1}$, $\bar{R}_t = \frac{1}{t}\sum_{\tau=1}^{t} R_\tau$ converges to $\mathbb{R}_-^d$....

Regret can be minimized using Blackwell approchabily

Blackwell Approachability

- $R_T \in \mathbb{R}^d$, $\mathcal{C} \subset \mathbb{R}^d$, $\Pi(\cdot)$ projection on $\mathcal{C}$

Blackwell Approachability

- $R_\tau \in \mathbb{R}^d$, $\mathcal{C} \subset \mathbb{R}^d$, $\Pi(\cdot)$ projection on $\mathcal{C}$
- Suppose there is $x_{t+1} \in \Delta(S^1)$ such that for any action of Nature,
$$\left\langle \mathbb{E}[R_{t+1}] - \Pi(\bar{R}_t) \; ; \bar{R}_t - \Pi(\bar{R}_t) \right\rangle \leq 0$$

Blackwell Approachability

- $R_\tau \in \mathbb{R}^d$, $\mathcal{C} \subset \mathbb{R}^d$, $\Pi(\cdot)$ projection on $\mathcal{C}$
- Suppose there is $x_{t+1} \in \Delta(S^1)$ such that for any action of Nature,
$$\left\langle \mathbb{E}[R_{t+1}] - \Pi(\bar{R}_t) \; ; \bar{R}_t - \Pi(\bar{R}_t) \right\rangle \leq 0$$
- Then, playing x_{t+1} at stage $t+1$ guarantee
$$\mathbb{E}\left[d_{\mathcal{C}}\left(\bar{R}_t\right)\right] \leq \frac{2\|R_\tau\|_\infty}{\sqrt{t}}$$

Blackwell Approachability

- $R_\tau \in \mathbb{R}^d$, $\mathcal{C} \subset \mathbb{R}^d$, $\Pi(\cdot)$ projection on $\mathcal{C}$
- Suppose there is $x_{t+1} \in \Delta(S^1)$ such that for any action of Nature,
$$\left\langle \mathbb{E}[R_{t+1}] - \Pi(\bar{R}_t) \; ; \bar{R}_t - \Pi(\bar{R}_t) \right\rangle \leq 0$$
- Then, playing x_{t+1} at stage $t+1$ guarantee
$$\mathbb{E}\left[d_{\mathcal{C}}\left(\bar{R}_t\right)\right] \leq \frac{2\|R_\tau\|_\infty}{\sqrt{t}}$$

Yes ! with $x_{t+1} = \dfrac{\bar{R}_t^+}{\left\|\bar{R}_t^+\right\|_1} \in \Delta(\{1, .., d\})$

- where $\bar{R}_t^+ = \max\left\{\bar{R}_t^s, 0\right\}_{s\in\{1,..,d\}}$

Speed of convergence

- And from Blackwell Approachability we deduce that :

$$\mathbb{E}[\bar{r}_t] = \mathbb{E}\left[\left\|\bar{R}_t^+\right\|_\infty\right] \leq \mathbb{E}\left[\left\|\bar{R}_t^+\right\|_2\right] \leq 2\sqrt{\frac{d}{t}}$$

Speed of convergence

- And from Blackwell Approachability we deduce that :

$$\mathbb{E}[\bar{r}_t] = \mathbb{E}\left[\|\bar{R}_t^+\|_\infty\right] \le \mathbb{E}\left[\|\bar{R}_t^+\|_2\right] \le 2\sqrt{\frac{d}{t}}$$

- And using concentration inequalities we can conclude that :

$$\mathbb{P}\left(\|\bar{R}_t^+\|_2 - 2\sqrt{\frac{d}{t}} \ge \varepsilon\right) \le \exp\left(-\frac{n\varepsilon^2}{16d}\right)$$

Speed of convergence

- And from Blackwell Approachability we deduce that :

$$\mathbb{E}[\bar{r}_t] = \mathbb{E}\left[\|\bar{R}_t^+\|_\infty\right] \leq \mathbb{E}\left[\|\bar{R}_t^+\|_2\right] \leq 2\sqrt{\frac{d}{t}}$$

- And using concentration inequalities we can conclude that :

$$\mathbb{P}\left(\|\bar{R}_t^+\|_2 - 2\sqrt{\frac{d}{t}} \geq \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{16d}\right)$$

- There is another strategy (using a Riemannien metric to project) that guarantees :

$$\mathbb{E}[\bar{r}_t] = \mathbb{E}\left[\|\bar{R}_t^+\|_\infty\right] \leq 2\sqrt{\frac{\log(d)}{t}}$$

Speed of convergence

- And from Blackwell Approachability we deduce that :

$$\mathbb{E}[\bar{r}_t] = \mathbb{E}\left[||\bar{R}_t^+||_\infty\right] \leq \mathbb{E}\left[||\bar{R}_t^+||_2\right] \leq 2\sqrt{\frac{d}{t}}$$

- And using concentration inequalities we can conclude that :

$$\mathbb{P}\left(||\bar{R}_t^+||_2 - 2\sqrt{\frac{d}{t}} \geq \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{16d}\right)$$

- There is another strategy (using a Riemannien metric to project) that guarantees :

$$\mathbb{E}[\bar{r}_t] = \mathbb{E}\left[||\bar{R}_t^+||_\infty\right] \leq 2\sqrt{\frac{\log(d)}{t}}$$

- If the player observes only its stage payoff but not its vector of all possible payoffs, he is able to minimize the external regret by experimenting an ε fraction of time.

What if all players minimise the external regret ?

If in a two player game, both players minimize the external regret, then : $\bar{q}_t = \frac{1}{t}\sum_{\tau=1}^{t}\delta_{(s_\tau,\sigma_\tau)} \in \Delta(S^1 \times S^2)$, the empirical distribution of the couple of actions, converges to the set $\mathcal{H} \subset \Delta(S^1 \times S^2)$ of probability distributions q such that :

$$\forall t^1 \in S^1 \quad \sum_{(s^1,s^2)} q(s^1,s^2)G^1(s^1,s^2) \geq \sum_{(s^1,s^2)} q(s^1,s^2)G^1(t^1,s^2)$$

$$\text{et} \;\; \forall t^2 \in S^2 \quad \sum_{(s^1,s^2)} q(s^1,s^2)G^2(s^1,s^2) \geq \sum_{(s^1,s^2)} q(s^1,s^2)G^2(s^1,t^2)$$

What if all players minimise the external regret ?

If in a two player game, both players minimize the external regret, then : $\bar{q}_t = \frac{1}{t}\sum_{\tau=1}^{t}\delta_{(s_\tau,\sigma_\tau)} \in \Delta(S^1 \times S^2)$, the empirical distribution of the couple of actions, converges to the set $\mathcal{H} \subset \Delta(S^1 \times S^2)$ of probability distributions q such that :

$$\forall t^1 \in S^1 \quad \sum_{(s^1,s^2)} q(s^1,s^2)G^1(s^1,s^2) \geq \sum_{(s^1,s^2)} q(s^1,s^2)G^1(t^1,s^2)$$

$$\text{et} \quad \forall t^2 \in S^2 \quad \sum_{(s^1,s^2)} q(s^1,s^2)G^2(s^1,s^2) \geq \sum_{(s^1,s^2)} q(s^1,s^2)G^2(s^1,t^2)$$

The Hannan set $\mathcal{H}$ is a polytope, which contains the set of correlated equilibria (and so the set of Nash equilibria).

What if all players minimise the external regret ?

If in a two player game, both players minimize the external regret, then : $\bar{q}_t = \frac{1}{t}\sum_{\tau=1}^{t}\delta_{(s_\tau,\sigma_\tau)} \in \Delta(S^1 \times S^2)$, the empirical distribution of the couple of actions, converges to the set $\mathcal{H} \subset \Delta(S^1 \times S^2)$ of probability distributions q such that :

$$\forall t^1 \in S^1 \quad \sum_{(s^1,s^2)} q(s^1,s^2)G^1(s^1,s^2) \geq \sum_{(s^1,s^2)} q(s^1,s^2)G^1(t^1,s^2)$$

$$\text{et } \forall t^2 \in S^2 \quad \sum_{(s^1,s^2)} q(s^1,s^2)G^2(s^1,s^2) \geq \sum_{(s^1,s^2)} q(s^1,s^2)G^2(s^1,t^2)$$

The Hannan set $\mathcal{H}$ is a polytope, which contains the set of correlated equilibria (and so the set of Nash equilibria).

The result extends to n-player games.

Minimizing regret on a zero-sum game

In a zero-sum game, if players play a non-regret strategy then $(\bar{x}_t, \bar{y}_t) = \left(\frac{1}{t}\sum_{\tau=1}^{t}\delta_{s_\tau}, \frac{1}{t}\sum_{\tau=1}^{t}\delta_{\sigma_\tau}\right)$, the couple of empirical action profiles converges to the set of optimal strategies.

Minimizing regret on a zero-sum game

In a zero-sum game, if players play a non-regret strategy then $(\bar{x}_t, \bar{y}_t) = \left(\frac{1}{t}\sum_{\tau=1}^{t} \delta_{s_\tau}, \frac{1}{t}\sum_{\tau=1}^{t} \delta_{\sigma_\tau}\right)$, the couple of empirical action profiles converges to the set of optimal strategies.

Proof. Asymptotically, player 1 has no external regret. Thus :

$$\frac{1}{t}\sum_{\tau=1}^{t} g(s_\tau, \sigma_\tau) \geq \max_{s^1 \in S^1} g(s^1, \bar{y}_t) = \max_{x \in \Delta(S^1)} g(x, \bar{y}_t) \geq v$$

Minimizing regret on a zero-sum game

In a zero-sum game, if players play a non-regret strategy then $(\bar{x}_t, \bar{y}_t) = \left(\frac{1}{t}\sum_{\tau=1}^{t} \delta_{s_\tau}, \frac{1}{t}\sum_{\tau=1}^{t} \delta_{\sigma_\tau}\right)$, the couple of empirical action profiles converges to the set of optimal strategies.

Proof. Asymptotically, player 1 has no external regret. Thus :

$$\frac{1}{t}\sum_{\tau=1}^{t} g(s_\tau, \sigma_\tau) \geq \max_{s^1 \in S^1} g(s^1, \bar{y}_t) = \max_{x \in \Delta(S^1)} g(x, \bar{y}_t) \geq v$$

Player 2 has no external regret too :

$$\frac{1}{t}\sum_{\tau=1}^{t} g(s_\tau, \sigma_\tau) \leq \min_{\sigma \in S^2} g(\bar{x}_t, \sigma) = \min_{y \in \Delta(S^2)} g(\bar{x}_t, y) \leq v$$

Internal Regret

Consider the case of a player against nature.

- Periods where a player used a strategy s :
  $N_t(s^i) = \{\tau \leq t \; ; \; s_\tau = s\}$

Internal Regret

Consider the case of a player against nature.

- Periods where a player used a strategy s : $N_t(s^i) = \{\tau \leq t \; ; \; s_\tau = s\}$
- It was the best thing to do : for all $s' \in \{1, ..., d\}$,

$$\frac{1}{|N_t(s)|} \sum_{\tau \in N_t(s)} U_\tau^s \geq \frac{1}{|N_t(s)|} \sum_{\tau \in N_t(s)} U_\tau^{s'}$$

Internal Regret

Consider the case of a player against nature.

- Periods where a player used a strategy s : $N_t(s^i) = \{\tau \leq t\ ;\ s_\tau = s\}$
- It was the best thing to do : for all $s' \in \{1, ..., d\}$,

$$\frac{1}{|N_t(s)|} \sum_{\tau \in N_t(s)} U_\tau^s \geq \frac{1}{|N_t(s)|} \sum_{\tau \in N_t(s)} U_\tau^{s'}$$

- or that this action was almost unused : $\frac{|N_t(s)|}{t} \to 0$

Internal Regret

Consider the case of a player against nature.

- Periods where a player used a strategy s : $N_t(s^i) = \{\tau \leq t \; ; \; s_\tau = s\}$
- It was the best thing to do : for all $s' \in \{1, ..., d\}$,

$$\frac{1}{|N_t(s)|} \sum_{\tau \in N_t(s)} U_\tau^s \geq \frac{1}{|N_t(s)|} \sum_{\tau \in N_t(s)} U_\tau^{s'}$$

- or that this action was almost unused : $\frac{|N_t(s)|}{t} \to 0$

The player has no internal regret if for all U_τ, a.s. :

$$\limsup_{t \to \infty} \max_{s,s'} \frac{1}{t} \sum_{\tau \in N_t(s)} U_\tau^{s'} - \frac{1}{t} \sum_{\tau \in N_t(s)} U_\tau^s \leq 0$$

Reformulation

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau\in N_t(1)} U_\tau^1 - \sum_{\tau\in N_t(1)} U_\tau^1, & \cdots & , \sum_{\tau\in N_t(1)} U_\tau^d - \sum_{\tau\in N_t(1)} U_\tau^1 \\ \ldots, & \cdots & , \ldots \\ \sum_{\tau\in N_t(d)} U_\tau^1 - \sum_{\tau\in N_t(d)} U_\tau^d, & \cdots & , \sum_{\tau\in N_t(d)} U_\tau^d - \sum_{\tau\in N_t(1)} U_\tau^d \end{pmatrix}$$

Reformulation

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau \in N_t(1)} U_\tau^1 - \sum_{\tau \in N_t(1)} U_\tau^1, & \dots & , \sum_{\tau \in N_t(1)} U_\tau^d - \sum_{\tau \in N_t(1)} U_\tau^1 \\ \dots, & \dots & , \dots \\ \sum_{\tau \in N_t(d)} U_\tau^1 - \sum_{\tau \in N_t(d)} U_\tau^d, & \dots & , \sum_{\tau \in N_t(d)} U_\tau^d - \sum_{\tau \in N_t(1)} U_\tau^d \end{pmatrix}$$

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau \in N_t(1)} U_\tau - \sum_{\tau \in N_t(1)} U_\tau^1 \mathbf{1} \\ \dots \\ \sum_{\tau \in N_t(d)} U_\tau - \sum_{\tau \in N_t(d)} U_\tau^d \mathbf{1} \end{pmatrix}$$

Reformulation

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau\in N_t(1)} U^1_\tau - \sum_{\tau\in N_t(1)} U^1_\tau, & \ldots & , \sum_{\tau\in N_t(1)} U^d_\tau - \sum_{\tau\in N_t(1)} U^1_\tau \\ \ldots, & \ldots & , \ldots \\ \sum_{\tau\in N_t(d)} U^1_\tau - \sum_{\tau\in N_t(d)} U^d_\tau, & \ldots & , \sum_{\tau\in N_t(d)} U^d_\tau - \sum_{\tau\in N_t(1)} U^d_\tau \end{pmatrix}$$

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau\in N_t(1)} U_\tau - \sum_{\tau\in N_t(1)} U^1_\tau \mathbf{1} \\ \ldots \\ \sum_{\tau\in N_t(d)} U_\tau - \sum_{\tau\in N_t(d)} U^d_\tau \mathbf{1} \end{pmatrix} = \begin{pmatrix} \frac{|N_t(1)|}{t}\bar{r}_t(1) \\ \ldots \\ \frac{|N_t(d)|}{t}\bar{r}_t(d) \end{pmatrix}$$

Reformulation

$$\bar{\mathcal{R}}_t = \frac{1}{t} \begin{pmatrix} \sum_{\tau \in N_t(1)} U_\tau^1 - \sum_{\tau \in N_t(1)} U_\tau^1, & \dots & , \sum_{\tau \in N_t(1)} U_\tau^d - \sum_{\tau \in N_t(1)} U_\tau^1 \\ \dots, & \dots & , \dots \\ \sum_{\tau \in N_t(d)} U_\tau^1 - \sum_{\tau \in N_t(d)} U_\tau^d, & \dots & , \sum_{\tau \in N_t(d)} U_\tau^d - \sum_{\tau \in N_t(1)} U_\tau^d \end{pmatrix}$$

$$\bar{\mathcal{R}}_t = \frac{1}{t} \begin{pmatrix} \sum_{\tau \in N_t(1)} U_\tau - \sum_{\tau \in N_t(1)} U_\tau^1 \mathbf{1} \\ \dots \\ \sum_{\tau \in N_t(d)} U_\tau - \sum_{\tau \in N_t(d)} U_\tau^d \mathbf{1} \end{pmatrix} = \begin{pmatrix} \frac{|N_t(1)|}{t} \bar{r}_t(1) \\ \dots \\ \frac{|N_t(d)|}{t} \bar{r}_t(d) \end{pmatrix}$$

$$\mathcal{R}_\tau = \begin{pmatrix} \mathbb{1}_{s_\tau = 1} \; r_\tau \\ \dots \\ \mathbb{1}_{s_\tau = d} \; r_\tau \end{pmatrix}$$

Reformulation

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau\in N_t(1)} U^1_\tau - \sum_{\tau\in N_t(1)} U^1_\tau, & \dots & , \sum_{\tau\in N_t(1)} U^d_\tau - \sum_{\tau\in N_t(1)} U^1_\tau \\ \dots, & \dots & , \dots \\ \sum_{\tau\in N_t(d)} U^1_\tau - \sum_{\tau\in N_t(d)} U^d_\tau, & \dots & , \sum_{\tau\in N_t(d)} U^d_\tau - \sum_{\tau\in N_t(1)} U^d_\tau \end{pmatrix}$$

$$\bar{\mathcal{R}}_t = \frac{1}{t}\begin{pmatrix} \sum_{\tau\in N_t(1)} U_\tau - \sum_{\tau\in N_t(1)} U^1_\tau \, \mathbf{1} \\ \dots \\ \sum_{\tau\in N_t(d)} U_\tau - \sum_{\tau\in N_t(d)} U^d_\tau \, \mathbf{1} \end{pmatrix} = \begin{pmatrix} \frac{|N_t(1)|}{t}\bar{r}_t(1) \\ \dots \\ \frac{|N_t(d)|}{t}\bar{r}_t(d) \end{pmatrix}$$

$$\mathcal{R}_\tau = \begin{pmatrix} \mathbb{1}_{s_\tau=1}\, r_\tau \\ \dots \\ \mathbb{1}_{s_\tau=d}\, r_\tau \end{pmatrix}$$

We want that $\bar{\mathcal{R}}_t = \frac{1}{t}\sum_{\tau=1}^t \mathcal{R}_\tau$ converges vers $\mathbb{R}^{d^2}_-$ ...
Blackwell approachability

Blackwell condition

Internal regret can be minimized if there exists x_{t+1} such that for all payoff vector U_{t+1} chosen by nature

$$\left\langle \mathbb{E}[\mathcal{R}_{t+1}] - \bar{\mathcal{R}}_t^-, \bar{\mathcal{R}}_t^+ \right\rangle \leq 0$$

Blackwell condition

Internal regret can be minimized if there exists x_{t+1} such that for all payoff vector U_{t+1} chosen by nature

$$\left\langle \mathbb{E}[\mathcal{R}_{t+1}] - \bar{\mathcal{R}}_t^-, \bar{\mathcal{R}}_t^+ \right\rangle \leq 0$$

Yes just take x_{t+1} to be an invariante measure of $\bar{\mathcal{R}}_t^+$!

Collectively minimisation of internal regret

[Foster-Vohra, Hart-MasCollel] If all players minimize their internal regret then the empirical distribution of the action profiles converges to the set of canonical correlated equilibrium distributions.

Collectively minimisation of internal regret

[Foster-Vohra, Hart-MasCollel] If all players minimize their internal regret then the empirical distribution of the action profiles converges to the set of canonical correlated equilibrium distributions.

Proof. Let Q^* be an accumulation point of the empirical distribution of actions

$$Q(s) = \lim_{t_k \to \infty} \left(\frac{1}{t_k} \#\{1 \leq m \leq t_k ; s_m = s\} \right)$$

By the internal non regret condition, we must have :

$$\sum_{s^{-i} \in S^{-i}} [G^i(k, s^{-i}) - G^i(\ell, s^{-i})] Q(k, s^{-i}) \geq 0$$

Which is exactly the correlated equilibrium condition.

Playing a best response w.r.t. calibrated strategies

Theorem [Foster-Vohra]

If all players play at each period a best response with respect to a belief generated by a calibrated strategy, the empirical distribution of the action profiles converges to the set of correlated equilibrium distributions.

Playing a best response w.r.t. calibrated strategies

Theorem [Foster-Vohra]

If all players play at each period a best response with respect to a belief generated by a calibrated strategy, the empirical distribution of the action profiles converges to the set of correlated equilibrium distributions.

Theorem [Kakade-Foster]

If all players play at each period a best response with respect to a belief generated by a smooth calibrated strategy then in most of the periods they play close to a Nash equilibrium.