Topics on strategic learning

Sylvain Sorin

IMJ-PRG
Sorbonne Université - UPMC
sylvain.sorin@imj-prg.fr

Network, population and congestion games
IHP
April 16 -17, 2019

General presentation

Common features : multistage interaction / stationarity

Difference with repeated games

Large family of models:

- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures

Applications to games

Alternative approaches

Link with dynamical systems

Common features : multistage interaction / stationarity
Difference with repeated games
Large family of models:
- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures
Applications to games
Alternative approaches
Link with dynamical systems

General presentation

Common features : multistage interaction / stationarity
Difference with repeated games
Large family of models:
- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures
Applications to games
Alternative approaches
Link with dynamical systems

General presentation

Common features : multistage interaction / stationarity
Difference with repeated games
Large family of models:
- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures
Applications to games
Alternative approaches
Link with dynamical systems

General presentation

Common features : multistage interaction / stationarity
Difference with repeated games
Large family of models:
- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures
Applications to games
Alternative approaches
Link with dynamical systems

General presentation

Common features : multistage interaction / stationarity
Difference with repeated games
Large family of models:
- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures
Applications to games
Alternative approaches
Link with dynamical systems

General presentation

Common features : multistage interaction / stationarity
Difference with repeated games
Large family of models:
- information at each stage
- knowledge of the environment
- hypotheses on unknown parameters and rationality

Unilateral procedures
Applications to games
Alternative approaches
Link with dynamical systems

Basic references

CESA-BIANCHI N. AND G. LUGOSI (2006) *Prediction, Learning and Games*, Cambridge University Press.

FUDENBERG D. AND D.K. LEVINE (1998) *Theory of Learning in Games*, M.I.T. Press.

FUDENBERG D. AND DAVID K. LEVINE (2009) *Learning and equilibrium*, Annual Review of Economics, **1**, 385-420.

HART S. AND A. MAS COLELL (2013) *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, World Scientific Publishing.

PERCHET V. (2014) Approachability, regret and calibration: implications and equivalences, *Journal of Dynamics and Games*, **1**, 181-254.

YOUNG P. (1998) *Theory of Learning in Games*, M.I.T. Press.

Unilateral procedures

Framework: agent acting in discrete time and facing an unknown environment.
At each stage n:
Choice: k_n in a finite set K
Observation: reward vector $U_n \in \mathcal{U} = [-1, 1]^K$
Payoff: the k_n^{th} component, $\omega_n = U_n^{k_n}$.
History at stage n: $h_{n-1} = \{k_1, U_1, ..., k_{n-1}, U_{n-1}\} \in H_{n-1}$.
A strategy of the player is a map σ from $H = \cup_{m=0}^{+\infty} H_m$ to $\Delta(K)$ (set of probabilities on K).

External regret

The (external) regret given $k \in K$ and $U \in \mathcal{U} \subset \mathbb{R}^K$ is the vector $R(k, U) \in \mathbb{R}^K$ defined by:

$$R(k, U)^\ell = U^\ell - U^k, \ \ell \in K.$$

Regret at stage n = $R_n = R(k_n, U_n)$:

$$R_n^\ell = U_n^\ell - \omega_n, \ \ell \in K.$$

Average regret vector at stage n, $\overline{R}_n = \frac{1}{n} \sum_{m=1}^{n} R_m$:

$$\overline{R}_n^\ell = \overline{U}_n^\ell - \overline{\omega}_n, \ \ell \in K.$$

Compare the actual (average) payoff to the payoff corresponding to the choice of a constant component, see Hannan (1957), Foster and Vohra (1999), Fudenberg and Levine (1995).

Definition

A strategy σ satisfies external consistency (or exhibits no external regret) if, for every process $\{U_m\} \in \mathcal{U}$:

$$\max_{k \in K} [\overline{R}_n^k]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

or, equivalently $\qquad \sum_{m=1}^n (U_m^k - \omega_m) \leq o(n), \qquad \forall k \in K.$

Internal regret

The internal regret given (k, U) is the $K \times K$ matrix $S(k, U)$ with components: $S^{j\ell}(k, U) = (U^\ell - U^j)\, \mathbf{I}_{\{j=k\}}$.
The evaluation at stage n is $S_n = S(k_n, U_n)$ so that:

$$S_n^{k\ell} = \begin{cases} U_n^\ell - U_n^k & \text{for } k = k_n \\ 0 & \text{otherwise.} \end{cases}$$

Average internal regret matrix:

$$\overline{S}_n^{k\ell} = \frac{1}{n} \sum_{m=1, k_m=k}^{n} (U_m^\ell - U_m^k)$$

Comparison for each component k, of the average payoff obtained on the dates where k was played, to the payoff for an alternative choice ℓ.
See Foster and Vohra (1999), Fudenberg and Levine (1999).

Definition

A strategy σ satisfies internal consistency (or exhibits no internal regret) if, for every process $\{U_m\} \in \mathcal{U}$ and every couple k, ℓ:

$$[\overline{S}_n^{k\ell}]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

The proof of existence of a strategy satisfying EC or IC will rely on approachability theory

Deterministic approachability: geometry

All the results are due to Blackwell (1956).

We describe the basic geometric principle that sustains the approachability property.

$x_1, x_2, \ldots$ is a sequence in $\boldsymbol{R}^K$, uniformly bounded: $\|x_n\|^2 \leq L$.
$\overline{x}_n$ the average of the first n elements in the sequence:
$\overline{x}_n = \frac{1}{n}\sum_{m=1}^{n} x_m$.

Given $C \subset \boldsymbol{R}^K$ closed, $\Pi_C(x)$ is a closest point to x in C.
(If C is convex, it is the projection of x on C.)
$d(x, C) = \|x - \Pi_C(x)\|$ is the distance from x to C.

Theorem (The geometric principle)

Suppose that $\{x_n\}$ *satisfies:*

$$\langle x_{n+1} - \Pi_C(\overline{x}_n), \overline{x}_n - \Pi_C(\overline{x}_n)\rangle \leq 0, \tag{1}$$

then $d(\overline{x}_n, C)$ *converges to* 0*.*

Proof

Let $y_n = \Pi_C(\overline{x}_n)$ and $d_n^2 = \|\overline{x}_n - y_n\|^2$. Then:

$$\begin{aligned}
d_{n+1}^2 &= \|\overline{x}_{n+1} - y_{n+1}\|^2 \\
&\leq \|\overline{x}_{n+1} - y_n\|^2 \\
&= \|\frac{1}{n+1}(x_{n+1} - y_n) + \frac{n}{n+1}(\overline{x}_n - y_n)\|^2 \\
&= (\frac{1}{n+1})^2\|x_{n+1} - y_n\|^2 + (\frac{n}{n+1})^2\|\overline{x}_n - y_n\|^2 \\
&\quad +2\frac{n}{(n+1)^2}\langle x_{n+1} - y_n, \overline{x}_n - y_n\rangle
\end{aligned}$$

Theorem (The geometric principle)

Suppose that $\{x_n\}$ satisfies:

$$\langle x_{n+1} - \Pi_C(\overline{x}_n), \overline{x}_n - \Pi_C(\overline{x}_n)\rangle \leq 0, \quad (1)$$

then $d(\overline{x}_n, C)$ converges to 0.

Proof

Let $y_n = \Pi_C(\overline{x}_n)$ and $d_n^2 = \|\overline{x}_n - y_n\|^2$. Then:

$$\begin{aligned}
d_{n+1}^2 &= \|\overline{x}_{n+1} - y_{n+1}\|^2 \\
&\leq \|\overline{x}_{n+1} - y_n\|^2 \\
&= \|\frac{1}{n+1}(x_{n+1} - y_n) + \frac{n}{n+1}(\overline{x}_n - y_n)\|^2 \\
&= (\frac{1}{n+1})^2\|x_{n+1} - y_n\|^2 + (\frac{n}{n+1})^2\|\overline{x}_n - y_n\|^2 \\
&\quad +2\frac{n}{(n+1)^2}\langle x_{n+1} - y_n, \overline{x}_n - y_n\rangle
\end{aligned}$$

Since $\|x_{n+1} - y_n\|^2 \leq 4L$, we obtain:

$$d_{n+1}^2 \leq (\frac{n}{n+1})^2 d_n^2 + (\frac{1}{n+1})^2 \, 4L \tag{2}$$

so that, by induction:

$$d_n^2 \leq \frac{4L}{n}.$$

∎

The framework is as follows:
A is a $I \times J$ matrix with coefficients in $\boldsymbol{R}^K$.
At each stage n, Player 1 (resp. Player 2) chooses a move i_n in I (resp. j_n in J).
The corresponding vector payoff, $g_n = A_{i_n j_n} \in \boldsymbol{R}^K$ is then announced.
$\overline{g}_n = \frac{1}{n}[\sum_{m=1}^n g_m]$
$L = \max_{i \in I, j \in J, k \in K} |A_{ij}^k|$.

Definitions

A set C in $\boldsymbol{R}^K$ is approachable by Player 1 if for any $\varepsilon > 0$ there exists a strategy σ and N such that, for any strategy τ of Player 2 and any $n \geq N$:

$$E_{\sigma,\tau}(d_n) \leq \varepsilon$$

where d_n is the euclidean distance $d(\overline{g}_n, C)$.
A set C in $\boldsymbol{R}^K$ is excludable by Player 1 if for some $\delta > 0$, the set $C^c_\delta = \{z; d(z, C) \geq \delta\}$ is approachable by him.

Given x in $X = \Delta(I)$, define $[xA] = co \{\sum_i x_i A_{ij};\ j \in J\}$, and similarly $[Ay]$, for y in $Y = \Delta(J)$.
If Player 1 uses x, his expected payoff will be in $[xA]$, whatever being the move of player 2.

Definition

A closed set C in $\boldsymbol{R}^K$ is a **B**-set for Player 1 if:
for any $z \notin C$, there exists a closest point $w = w(z)$ in C to z and a mixed move $x = x(z)$ in X, such that the hyperplane trough w orthogonal to the segment $[wz]$ separates z from $[xA]$.

$$\langle z - w, u - w \rangle \leq 0, \forall u \in [xA].$$

Theorem

Let C be a **B***-set for Player 1.*
Then C is approachable by that player.
Explicitly, a strategy satisfying $\sigma(h_{n+1}) = x(\overline{g}_n)$, whenever $\overline{g}_n \notin C$, gives:

$$E_{\sigma\tau}(d_n) \;\leq\; \frac{2L}{\sqrt{n}}, \quad \forall\tau$$

and d_n converges $P_{\sigma\tau}$ a.s. to 0, more precisely:

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{8L}{\varepsilon N}$$

Proof

Let Player 1 use a strategy σ as above. Denote $w_n = w(\overline{g}_n)$.

Theorem

*Let C be a **B**-set for Player 1.*
Then C is approachable by that player.
Explicitly, a strategy satisfying $\sigma(h_{n+1}) = x(\overline{g}_n)$, whenever $\overline{g}_n \notin C$, gives:

$$E_{\sigma\tau}(d_n) \;\le\; \frac{2L}{\sqrt{n}}, \quad \forall\tau$$

and d_n converges $P_{\sigma\tau}$ a.s. to 0, more precisely:

$$P(\exists n \ge N; d_n^2 \ge \varepsilon) \le \frac{8L}{\varepsilon N}$$

Proof

Let Player 1 use a strategy σ as above. Denote $w_n = w(\overline{g}_n)$.

The property of $x(\overline{g}_n)$ implies that:

$$\langle E(g_{n+1}|h_n) - w_n, \overline{g}_n - w_n\rangle) \leq 0$$

since $E(g_{n+1}|h_n)$ belongs to $[x(\overline{g}_n)A]$.
Hence the previous equation in the deterministic case:

$$d_{n+1}^2 \leq (\frac{n}{n+1})^2\ d_n^2 + (\frac{1}{n+1})^2 \|x_{n+1} - y_n\|^2,$$

gives here by taking conditional expectation with respect to the history h_n:

$$\mathrm{E}(d_{n+1}^2|h_n) \leq (\frac{n}{n+1})^2\ d_n^2 + (\frac{1}{n+1})^2 4L \tag{3}$$

So that we obtain, like in (2):

$$\mathrm{E}(d_{n+1}^2) \leq (\frac{n}{n+1})^2 \, \mathrm{E}(d_n^2) + (\frac{1}{n+1})^2 \, 4L$$

and by induction:

$$\mathrm{E}(d_n^2) \leq \frac{4L}{n}.$$

This gives in particular the convergence in probability of d_n to 0. Let $W_n = d_n^2 + 4L \sum_{m=n+1}^{\infty} \frac{1}{m^2}$. Then from (3):

$$E(W_{n+1} | h_n) \leq W_n$$

thus W_n is a positive supermartingale hence converges P *a.s.* to 0. More precisely Doob's maximal inequality gives :

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{E(W_N)}{\varepsilon} \leq \frac{8L}{\varepsilon N}.$$

■

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2, by playing y i.i.d.

In fact this condition is also sufficient for convex sets.
This provides a simple criteria for approachability of convex sets.

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2, by playing y i.i.d.

In fact this condition is also sufficient for convex sets.
This provides a simple criteria for approachability of convex sets.

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2, by playing y i.i.d.

In fact this condition is also sufficient for convex sets.
This provides a simple criteria for approachability of convex sets.

Convex case

Theorem

Assume C closed and convex in $\mathbb{R}^K$.
C is a **B**-*set for Player 1 iff*

$$(*) \qquad [Ay] \cap C \neq \emptyset, \qquad \forall y \in Y.$$

In particular a set is approachable iff it is a **B**-*set.*

The proof follows from the minmax theorem.

Approachability implies EC

The on-line decision problem with choice set K defines a game where the vector payoff is the regret in $\boldsymbol{R}^K$.
We prove the existence of a strategy satisfying EC by showing that the negative orthant $D = \boldsymbol{R}^K_-$ is approachable by the sequence of average regret $\{\overline{R}_n\}$.

Lemma

$\forall x \in \Delta(K), \forall U \in \mathcal{U}$:

$$\langle x, \mathrm{E}_x[R(., U)] \rangle = 0.$$

Proof

$$\mathrm{E}_x[R(., U)] = \sum_{k \in K} x_k \, R(k, U) = \sum_{k \in K} x_k (U - U^k \mathbf{1}) = U - \langle x, U \rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathrm{E}_x[R(., U)] \rangle = 0$.

Approachability implies EC

The on-line decision problem with choice set K defines a game where the vector payoff is the regret in $\boldsymbol{R}^K$.
We prove the existence of a strategy satisfying EC by showing that the negative orthant $D = \boldsymbol{R}_-^K$ is approachable by the sequence of average regret $\{\overline{R}_n\}$.

Lemma

$\forall x \in \Delta(K), \forall U \in \mathcal{U}$:

$$\langle x, \mathrm{E}_x[R(., U)]\rangle = 0.$$

Proof

$$\mathrm{E}_x[R(., U)] = \sum_{k \in K} x_k \, R(k, U) = \sum_{k \in K} x_k (U - U^k \mathbf{1}) = U - \langle x, U\rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathrm{E}_x[R(., U)]\rangle = 0$.

Approachability implies EC

The on-line decision problem with choice set K defines a game where the vector payoff is the regret in $\boldsymbol{R}^K$.

We prove the existence of a strategy satisfying EC by showing that the negative orthant $D = \boldsymbol{R}_-^K$ is approachable by the sequence of average regret $\{\overline{R}_n\}$.

Lemma

$\forall x \in \Delta(K), \forall U \in \mathcal{U}$:

$$\langle x, \mathrm{E}_x[R(., U)] \rangle = 0.$$

Proof

$$\mathrm{E}_x[R(., U)] = \sum_{k \in K} x_k \; R(k, U) = \sum_{k \in K} x_k (U - U^k \mathbf{1}) = U - \langle x, U \rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathrm{E}_x[R(., U)] \rangle = 0$. ∎

$\overline{R}_n$ is the average regret at stage n, $\overline{R}_n^+$ the non negative components.
Define, if $\overline{R}_n^+ \neq 0$, $\sigma(h_n)$ to be proportional to this vector.
Claim

$$\langle \mathrm{E}(R_{n+1}|h_n) - \Pi_D(\overline{R}_n), \overline{R}_n - \Pi_D(\overline{R}_n)\rangle = 0$$

i) $\langle \Pi_D(\overline{R}_n), \overline{R}_n - \Pi_D(\overline{R}_n)\rangle = 0$
ii)

$$\begin{aligned}
\langle \mathrm{E}(R_{n+1}|h_n), \overline{R}_n - \Pi_D(\overline{R}_n)\rangle &= \langle \mathrm{E}(R_{n+1}|h_n), \overline{R}_n^+\rangle \\
&\div \langle \mathrm{E}(R_{n+1}|h_n), \sigma(h_n)\rangle \\
&= \langle \mathrm{E}_x[R(., U_{n+1})], x\rangle, \quad \text{for } x = \sigma(h_n) \\
&= 0
\end{aligned}$$

Thus the **B** condition is satisfied, so D is *approachable* hence $d(\overline{R}_n, \boldsymbol{R}_-^K)$ goes to 0 and $\max_{k\in K}[\overline{R}_n^k]^+ \longrightarrow 0$.

Approachability implies IC

Given a $K \times K$ real matrix A with nonnegative coefficients, let $Inv[A]$ be the non-empty set of invariant measures for A, namely vectors $\mu \in \Delta(K)$ satisfying:

$$\sum_{k \in K} \mu^k A^{k\ell} = \mu^\ell \sum_{k \in K} A^{\ell k} \qquad \forall \ell \in K.$$

(The existence follows from the existence of an invariant measure for a Markov chain - which is itself a consequence of the minmax theorem).

Approachability implies IC

Given a $K \times K$ real matrix A with nonnegative coefficients, let $Inv[A]$ be the non-empty set of invariant measures for A, namely vectors $\mu \in \Delta(K)$ satisfying:

$$\sum_{k \in K} \mu^k A^{k\ell} = \mu^\ell \sum_{k \in K} A^{\ell k} \qquad \forall \ell \in K.$$

(The existence follows from the existence of an invariant measure for a Markov chain - which is itself a consequence of the minmax theorem).

Lemma

Given $A \in \boldsymbol{R}_+^{K^2}$*, let* $\mu \in Inv[A]$ *then:*

$$\langle A, \mathrm{E}_\mu(S(., U)) \rangle = 0, \quad \forall U \in \mathcal{U}.$$

Proof

$$\langle A, \mathrm{E}_\mu(S(., U)) \rangle = \sum_{k,\ell} A^{k\ell} \mu^k (U^\ell - U^k)$$

and the coefficient of each U^ℓ is

$$\sum_{k \in K} \mu^k A^{k\ell} - \mu^\ell \sum_{k \in K} A^{\ell k} = 0$$

∎

To prove the existence of a strategy satisfying internal consistency, we show that $\Delta = \boldsymbol{R}_{-}^{K\times K}$ is approachable by the sequence of internal regret $\{S_n\}$.

Define, if $A = \overline{S}_n^{+} \neq 0$, $\sigma(h_n)$ to be an invariant measure of A.

Claim:

$$\langle \mathsf{E}(S_{n+1}|h_n) - \Pi_\Delta(\overline{S}_n), \overline{S}_n - \Pi_\Delta(\overline{S}_n)\rangle = 0$$

since again $\langle \Pi_\Delta(\overline{S}_n), \overline{S}_n - \Pi_\Delta(\overline{S}_n)\rangle = 0$ and

$$\begin{aligned}
\langle \mathsf{E}(S_{n+1}|h_n), \overline{S}_n - \Pi_\Delta(\overline{S}_n)\rangle &= \langle \mathsf{E}(S_{n+1}|h_n), \overline{S}_n^{+}\rangle \\
&= \langle \mathsf{E}(S_{n+1}|h_n), A\rangle \\
&= \langle \mathsf{E}_\mu[S(., U_{n+1})], A\rangle, \quad \text{for } \mu = \sigma(h_n) \\
&= 0
\end{aligned}$$

Thus Δ is approachable, hence $\max_{k,\ell}[\overline{S}_n^{k,\ell}]^{+} \longrightarrow 0$. ∎

To prove the existence of a strategy satisfying internal consistency, we show that $\Delta = \boldsymbol{R}_-^{K\times K}$ is approachable by the sequence of internal regret $\{S_n\}$.
Define, if $A = \overline{S}_n^+ \neq 0$, $\sigma(h_n)$ to be an invariant measure of A.
Claim:

$$\langle \mathsf{E}(S_{n+1}|h_n) - \Pi_\Delta(\overline{S}_n), \overline{S}_n - \Pi_\Delta(\overline{S}_n)\rangle = 0$$

since again $\langle \Pi_\Delta(\overline{S}_n), \overline{S}_n - \Pi_\Delta(\overline{S}_n)\rangle = 0$ and

$$\begin{aligned}
\langle \mathsf{E}(S_{n+1}|h_n), \overline{S}_n - \Pi_\Delta(\overline{S}_n)\rangle &= \langle \mathsf{E}(S_{n+1}|h_n), \overline{S}_n^+\rangle \\
&= \langle \mathsf{E}(S_{n+1}|h_n), A\rangle \\
&= \langle \mathsf{E}_\mu[S(., U_{n+1})], A\rangle, \quad \text{for } \mu = \sigma(h_n) \\
&= 0
\end{aligned}$$

Thus Δ is approachable, hence $\max_{k,\ell}[\overline{S}_n^{k,\ell}]^+ \longrightarrow 0$. ■

Calibrating

Consider a sequence X_m with values in a finite set Ω that one wants to predict.
Any deterministic prediction algorithm ϕ_m - where the loss is measured by $\|X_m - \phi_m\|$ - will have a worst loss 1 and any random predictor a loss at least $1/2$ (take $X_m = 1$ iff $\phi_m(1) \leq 1/2$).
Introduce a finite discretization V of the set $D = \Delta(\Omega)$ and consider a predictor acting in V with the following interpretation: "$\phi_m = v$" means that the anticipated probability that $X_m = \omega$ (or $X_m^\omega = 1$) is v^ω.
Definition:
ϕ is ε-calibrated if, for any $v \in V$:

$$\lim_{n \to +\infty} \frac{1}{n} \| \sum_{\{m \leq n, \phi_m = v\}} (X_m - v) \| \leq \varepsilon$$

If the proportion of stages where v is predicted does not vanish, the average value of X_m on these dates is close to v.

Let B_n^v be the set of stages before n where v is announced, let N_n^v be its cardinal and $\overline{X}_n(v)$ the empirical average of X_m on these stages.

Then the condition writes:

$$\lim_{n\to+\infty} \frac{N_n^v}{n} \|\overline{X}_n(v) - v\| \leq \varepsilon, \qquad \forall v \in V.$$

From internal consistency to calibrating

Foster and Vohra (1997)
Consider the online process where the choice set of the forecaster is V and the outcome given v and X_m is:

$$U_m^v = \|X_m - v\|^2$$

(where we use the L^2 norm).
Given an internal consistent procedure ϕ one obtains (the outcome is here a loss)

$$\frac{1}{n} \sum_{m \in B_n^v} (U_m^v - U_m^w) \leq o(n), \qquad \forall w \in V,$$

This is:

$$\frac{1}{n}\sum_{m\in B_n^v}(\|X_m - v\|^2 - \|X_m - w\|^2) \leq o(n), \qquad \forall w \in V,$$

and is equal to :

$$\frac{N_n^v}{n}(\|\overline{X}_n(v) - v\|^2 - \|\overline{X}_n(v) - w\|^2) \leq o(n), \qquad \forall w \in V.$$

In particular by chosing a point w closest to $\overline{X}_n(v)$

$$\frac{N_n^v}{n}(\|\overline{X}_n(v) - v\|^2) \leq \delta^2 + o(n)$$

where δ is the L^2 mesh of V, from which calibration follows.

Foster and Vohra (1997)
We use calibrating to prove approachability of convex sets.

Assume that C satisfies: $\forall\, y \in Y, \exists\, x \in X$ such that $x\, A\, y \in C$.

Consider a δ-grid of Y defined by $\{y_v, v \in V\}$.
A stage is of type v if player 1 predicts y_v and then plays a mixed move x_v such that $x_v\, A\, y_v \in C$.
By using a calibrated procedure, the average move of player 2 on the stages of type v will be δ close to y_v.
By a martingale argument the average payoff will then be ε close to $x_v\, A\, y_v$ for δ small enough and n large enough.
Finally the total average payoff is a convex combination of such amounts hence is close to C by convexity.

From calibrating to approachability

Foster and Vohra (1997)
We use calibrating to prove approachability of convex sets.

Assume that C satisfies: $\forall\, y \in Y, \exists\, x \in X$ such that $x\,A\,y \in C$.

Consider a δ-grid of Y defined by $\{y_v, v \in V\}$.
A stage is of type v if player 1 predicts y_v and then plays a mixed move x_v such that $x_v\,A\,y_v \in C$.
By using a calibrated procedure, the average move of player 2 on the stages of type v will be δ close to y_v.
By a martingale argument the average payoff will then be ε close to $x_v\,A\,y_v$ for δ small enough and n large enough.
Finally the total average payoff is a convex combination of such amounts hence is close to C by convexity.

1. Conditional expectation

Regret at stage n that the player wants to control:

$$\sum_{m=1}^{n} U_m^k - \omega_m, \quad k \in K$$

where $\omega_m = U_m^{k_m}$ is the random payoff at stage m.
Let $x_m \in \Delta(K)$ be the strategy of the player at stage m, then

$$\mathrm{E}(\omega_m | h_{m-1}) = \langle U_m, x_m \rangle$$

so that $\omega_m - \langle U_m, x_m \rangle$ is a bounded martingale difference.
Hoeffding-Azuma's concentration inequality for a process $\{Z_n\}$ of martingale differences with $|Z_n| \leq L$ states that:

$$\mathrm{P}\{|\overline{Z}_n| \geq \varepsilon\} \leq 2 \exp(-\frac{n\,\varepsilon^2}{2L^2})$$

Hence the average difference between the payoff and its conditional expectation is controlled.

Thus we consider quantities of the form:

$$\sum_{m=1}^{n} U_m^k - \langle U_m, x_m \rangle, \quad k \in K.$$

or equivalently, because of the linearity:

$$\sum_{m=1}^{n} \langle U_m, x \rangle - \langle U_m, x_m \rangle, \quad x \in \Delta(K).$$

hence EC writes:

$$\sum_{m=1}^{n} \langle U_m, x \rangle - \langle U_m, x_m \rangle \leq o(n), \quad x \in \Delta(K)$$

Similarly IC becomes:

$$\sum_{m=1}^{n} x_m^i [U_m^j - U_m^i] \leq o(n), \quad \forall i, j \in K.$$

2. Procedures in law

Assume that the actual move k_n is not observed and define a pseudo-process $\tilde{R}$ defined through the conditional expected regret:

$$R_n = U_n - \omega_n \mathbf{1}, \qquad \tilde{R}_n = U_n - \langle U_n, x_n \rangle \mathbf{1}$$

and introduce the associated strategy $\tilde{\sigma}$.
Then consistency holds both for the pseudo and the realized processes under $\tilde{\sigma}$.

3. Experts

External consistency can be considered as a robustness property of σ facing a given finite family of "external" experts using procedures $\phi \in \Phi$:

$$\lim \frac{1}{n}[\sum_{m=0}^{n} \langle \phi_m - x_m, U_m \rangle] \leq 0, \quad \forall \phi \in \Phi.$$

The typical case corresponds to a constant choice : $\phi = k$ and $\Phi = K$.
In general "k" will be the (random) move of expert k, that the player follows with probability x_m^k at stage m.
U_m^k is then the payoff to expert k at stage m.
Internal consistency corresponds to experts adjusting their behavior to the one of the predictor.

4. From external to internal consistency

Stoltz and Lugosi (2005)
Consider a family $\psi^{ij}, (i,j) \in K \times K$ of experts and θ an algorithm that satisfies external consistency with respect to this family.
Define σ inductively as follows.
Given some element $p \in \Delta(K)$, let $p(ij)$ be the vector obtained by adding p^i to the j^{th} component of p.
Let $q_{n+1}(p)$ be the distribution induced by θ at stage $n+1$ given the history h_n and the behavior $\psi^{ij}(h_n) = p(ij)$ of the experts.
Assume that the map $p \mapsto q_{n+1}(p)$ is continuous and let $\overline{p}_{n+1}$ be a fixed point which defines $\sigma(h_n) = x_{n+1}$.

The fact that σ is an incarnation of θ implies that it performs well facing any ψ^{ij} hence

$$[\sum_{m=0}^{n} \langle \psi^{ij}_m - x_m, U_m \rangle] \leq o(n), \qquad \forall i, j$$

which is

$$[\sum_{m=0}^{n} \langle \overline{p}(ij)_m - \overline{p}_m, U_m \rangle] \leq o(n), \qquad \forall i, j$$

hence

$$[\sum_{m=0}^{n} \overline{p}^i_m (U^j_m - U^i_m)] \leq o(n), \qquad \forall i, j$$

and this is the internal consistency condition.

Blum and Mansour (2007)

Consider K parallel algorithms $\{\phi[k]\}$ having no external regret, that generates each a (row) vector $q[k] \in \Delta(K)$ then define σ by the invariant measure p with $p = pQ$. Given the outcome $U \in \mathbb{R}^K$, add $p^k U$ to the entry of algorithm $\phi[k]$. Expressing the fact that $\phi[k]$ satisfies no external regret gives, at stage m, for all $j \in K$

$$[\sum_{m=0}^{n} p_m^k U_m^j - \langle q[k]_m, p_m^k U_m \rangle] \leq o(n)$$

Note that $\sum_k \langle q[k]_m, p_m^k U_m \rangle = \sum_k \langle p_m^k q[k]_m, U_m \rangle = \langle p_m, U_m \rangle$, hence by summing over k, for any function $M : K \mapsto K$, corresponding to a perturbation of σ with $j = M(k)$ the difference between the performances of σ_M and σ will satisfy:

$$[\sum_{m=0}^{n} \sum_{k} p_m^k U_m^{M(k)} - \langle p_m, U_m \rangle] \leq o(n).$$

This is the internal consistency for “swap experts”.

5. Bandit framework

This is the case where given the move k and the vector U the only information to the predictor is the realization $\omega = U^k$ (the vector U is not announced).

Define the pseudo regret vector at each stage n by:

$$\hat{U}_n^k = \frac{\omega_n}{\sigma_n^k} \mathbf{1}_{\{k_n = k\}}$$

and note that it is an unbiased estimator of the true regret.

To keep the outcome bounded one may have to perturb the strategy but same asymptotic properties hold.

(Auer, Cesa-Bianchi, Freund, Shapire, 2002)

For recent advances, see Bubeck and Cesa-Bianchi (2012), chapter 5.

6. Imperfect monitoring

At each stage n, given a profile of moves (i_n, j_n), a signal s_n in a finite set S with law $M(i_n, j_n)$ is sent to player 1 and this is his only information.

Given $y \in Y = \Delta(J)$, $m(y) \in \Delta(S)^I = \{\, M(i, y), i \in I\}$ is the "flag" induced by y.

$$d(\mu) = \max_{x\in\Delta(I)} \min_{y\in\Delta(J); m(y)=\mu} G(x, y).$$

Note that in general best replies are not pure.

Given a n-stage play the average flag is $\overline{\mu}_n$, where $\mu_r = m(j_r)$ (hence also $m(\overline{y}_n)$) and the external regret is then

$$r_n = d(\overline{\mu}_n) - \overline{G}_n$$

Cesa-Bianchi, Lugosi and Stoltz (2006), Lehrer and Solan (2007), Lugosi, Mannor and Stoltz (2008), Perchet (2009)

The vector of internal regret is defined similarly.
Let $A_n[\ell]$ be the set of stages before n where player 1 uses $x[\ell]$ and $N_n[\ell]$ its cardinality. $\overline{\mu}_n[\ell]$ resp. $\overline{G}_n[\ell]$, are the corresponding average flag resp. payoff. Then:

$$R_n[\ell] = d(\overline{\mu}_n[\ell]) - \overline{G}_n[\ell], \qquad \ell \in L$$

and define $\varepsilon-$internal consistency as:

$$\limsup_{n\to+\infty} \frac{N_n[\ell]}{n}(R_n[\ell] - \varepsilon)^+ \to 0, \qquad \forall \ell \in L.$$

The main result in this framework is the existence of $\varepsilon-$internal consistent strategies, Perchet (2009)

Approachability

Perchet V. (2011a)
Let

$$P(x,\mu) = \{G(x,y); m(y) = \mu, y \in Y\} \subset \mathbb{R}^d$$

be the set of payoffs compatible with the strategy $x \in X$ and the flag μ.

Proposition

A closed convex set $C \subset \mathbb{R}^d$ is approachable (by player 1) if and only if

$$\forall \mu \in m(Y), \exists x \in X \qquad \textit{such that} \qquad P(x,\mu) \subset C.$$

Note that this is exactly Blackwell's condition in the full monitoring case.

Global procedures

Application to games

Let $\mathcal{G}$ be a finite game in strategic form.
Finitely many players $i \in I$.
S^i : finite moves set of player i, $S = \prod_i S^i$,
$Z = \Delta(S)$ set of probabilities on S (correlated moves).
Repeated interaction in discrete time
At each stage the players observe the actions of their opponents.
We want evaluate the joint impact on the play of the prescribed behavior of the players.
Study the procedure from the view point of player 1
$S^1 = K, X = \Delta(K)$ (mixed moves of player 1),
$L = \prod_{i \neq 1} S^i$, and $Y = \Delta(L)$ (correlated moves of player 1's opponents) hence $Z = \Delta(K \times L)$.
$F : S \to R$ denotes the payoff function of player 1
and its linear extension to Z.

Application to games

Let $\mathcal{G}$ be a finite game in strategic form.
Finitely many players $i \in I$.
S^i : finite moves set of player i, $S = \prod_i S^i$,
$Z = \Delta(S)$ set of probabilities on S (correlated moves).
Repeated interaction in discrete time
At each stage the players observe the actions of their opponents.
We want evaluate the joint impact on the play of the prescribed behavior of the players.
Study the procedure from the view point of player 1
$S^1 = K, X = \Delta(K)$ (mixed moves of player 1),
$L = \prod_{i \neq 1} S^i$, and $Y = \Delta(L)$ (correlated moves of player 1's opponents) hence $Z = \Delta(K \times L)$.
$F : S \to \mathbb{R}$ denotes the payoff function of player 1
and its linear extension to Z.

External consistency and Hannan's set

H^1 (for Hannan's set) is the set of correlated moves in $z \in Z$ satisfying:

$$F(k, z^{-1}) \leq F(z), \forall k \in K$$

where z^{-1} stands for the marginal of z on L.
(Player 1 compares his payoff using a given move k to his actual payoff at z, assuming the other players' behavior, z^{-1}, given.)

Linearity (and stationarity) allow to deduce from property on the payoffs, property on the moves.

Consider the on line problem corresponding to the repeated game with outcome vector at stage m given by $\{F(k, \ell_m\}_{k \in K}\}$, where ℓ_m is the profile of moves of his opponents, and define the empirical average distribution of moves

$$z_n = \frac{1}{n} \sum_{m=1}^{n} (k_m, \ell_m) \in Z$$

Proposition

If Player 1 follows some external consistent procedure, z_n converges a.s. to the Hannan set H^1.

Proof

The proof is straightforward due to the linearity of the payoff.
The external consistency property is

$$\frac{1}{n}\sum_{m=1}^{n} F(k,\ell_m) - \frac{1}{n}\sum_{m=1}^{n} F(k_m,\ell_m) \leq o(n) \qquad \forall k \in K$$

which gives:

$$F(k, \frac{1}{n}\sum_{m=1}^{n} \ell_m)) - F(\frac{1}{n}\sum_{m=1}^{n}(k_m,\ell_m)) \leq o(n) \qquad \forall k \in K$$

and this expression is:

$$F(k, z_n^{-1}) - F(z_n) \leq o(n) \qquad \forall k \in K.$$

∎

One defines similarly H^i for each player and $H = \cap_i H^i$ which is the global Hannan's set.

Proposition

If all players follow some external consistent procedure, the empirical distribution of moves converges a.s. to the Hannan set H.

Note that no coordination is required.

In the case of a zero-sum game one has, for $z \in H$ with marginals z^1, z^2:

$$f(z) \geq f(s^1, z^2), \quad \forall s^1 \in S^1$$

and the opposite inequality for the other player hence the marginals z^1, z^2 are optimal strategies and $f(z)$ is equal to the value.

One defines similarly H^i for each player and $H = \cap_i H^i$ which is the global Hannan's set.

Proposition

If all players follow some external consistent procedure, the empirical distribution of moves converges a.s. to the Hannan set H.

Note that no coordination is required.

In the case of a zero-sum game one has, for $z \in H$ with marginals z^1, z^2:

$$f(z) \geq f(s^1, z^2), \quad \forall s^1 \in S^1$$

and the opposite inequality for the other player hence the marginals z^1, z^2 are optimal strategies and $f(z)$ is equal to the value.

One defines similarly H^i for each player and $H = \cap_i H^i$ which is the global Hannan's set.

Proposition

If all players follow some external consistent procedure, the empirical distribution of moves converges a.s. to the Hannan set H.

Note that no coordination is required.

In the case of a zero-sum game one has, for $z \in H$ with marginals z^1, z^2:

$$f(z) \geq f(s^1, z^2), \quad \forall s^1 \in S^1$$

and the opposite inequality for the other player hence the marginals z^1, z^2 are optimal strategies and $f(z)$ is equal to the value.

Example: for the zero-sum game

0	1	-1
-1	0	1
1	-1	0

the distribution

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

is in the Hannan set.

Internal consistency and correlated equilibria

A correlated equilibrium of $\mathcal{G}$ is a Nash equilibrium of the game extended by an information structure $\mathcal{I}$ given by:
- a probability space $(\Omega, \mathcal{A}, P)$
- a family of measurable maps θ^i from $(\Omega, \mathcal{A})$ to A^i (set of signals for player i).
A profile σ of strategies in $[\mathcal{G}, \mathcal{I}]$ maps the initial probability P on Ω to a probability $Q(\sigma)$ on S.
$CED(\mathcal{G})$ is the set of equilibrium correlated distributions in $\mathcal{G}$ $= \{Q(\sigma), \sigma$ equilibrium in $[\mathcal{G}, \mathcal{I}]\}$.

Theorem (Aumann, 1974)

$Q \in CED(\mathcal{G})$ *iff*

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

Back to the repeated game framework we still consider only player 1 and denote by F his payoff.
Given $z = (z_s)_{s \in S} \in Z$, introduce the family of comparison payoffs, testing k against j defined by:

$$C(j,k)(z) = \sum_{\ell \in L} [F(k,\ell) - F(j,\ell)] z_{(j,\ell)} \qquad j,k \in K.$$

Define :

$$C^1 = \{z \in Z; C(j,k)(z) \leq 0, \forall j,k \in K\}.$$

Proposition

If Player 1 follows some internal consistency procedure, the empirical distribution of moves converges a.s. to the set C^1.

Proof
The internal consistency property is

$$\frac{1}{n}\sum_{1\leq m\leq n, k_m=k}[F(j,\ell_m)-F(k_m,\ell_m)] \leq o(n) \qquad \forall k,j\in K$$

which gives:

$$\sum_{\ell\in L}[F(j,\ell)-F(k,\ell)]z_n(k,\ell) \leq o(n) \qquad \forall k,j\in K$$

■

Recall that the set of correlated equilibrium distribution of the game $\{\mathcal{G}\}$ is defined by

$$C = \{z \in Z; \sum_{\ell \in S^{-i}} [F^i(k,\ell) - F^i(j,\ell)] z_{(j,\ell)} \leq 0, \quad \forall j,k \in S^i, \forall i \in I\}.$$

so that

$$C = \cap_{i \in I} C_i$$

Thus we obtain:

Proposition

If each player follows some internal consistency procedure, the empirical distribution of moves converges a.s. to the set of correlated equilibria.

Note that this provides a proof of existence of correlated equilibrium through the existence of internally consistent procedures.

From calibrating to correlated equilibrium

Foster and Vohra (1997)

Consider the case where Player 1 is forecasting the behavior of his opponents (a profile in L).

Given a precision level δ, Player 1 predicts points in a δ-grid V of $\Delta(L)$ and plays a pure best reply k to his forecast.

If the forecast is calibrated the empirical distribution of the moves of the opponents will be close to v, on each set of stages of the form $\{m; v_m = v \in \Delta(L)\}$, hence the action chosen by Player 1, k, will be almost a best reply to the frequency near v.

If z is the average empirical distribution, the conditional distribution $z|k$ on L will correspond to a convex combination of distributions v to which k is best reply, hence k will still be (approximate) best reply to $z|k$: hence z is (approximately) in C^1.

If all players use calibrated strategies the empirical average frequency converges to C.

From calibrating to correlated equilibrium

Foster and Vohra (1997)
Consider the case where Player 1 is forecasting the behavior of his opponents (a profile in L).
Given a precision level δ, Player 1 predicts points in a δ-grid V of $\Delta(L)$ and plays a pure best reply k to his forecast.
If the forecast is calibrated the empirical distribution of the moves of the opponents will be close to v, on each set of stages of the form $\{m; v_m = v \in \Delta(L)\}$, hence the action chosen by Player 1, k, will be almost a best reply to the frequency near v.
If z is the average empirical distribution, the conditional distribution $z|k$ on L will correspond to a convex combination of distributions v to which k is best reply, hence k will still be (approximate) best reply to $z|k$: hence z is (approximately) in C^1.
If all players use calibrated strategies the empirical average frequency converges to C.

No convergence to Nash

There is no uncoupled deterministic smooth dynamic that converges to Nash equilibrium in all finite 2-person games: Hart and Mas-Colell (2003).
Similarly there are no learning process with finite memory such that the stage behavior will converge to Nash equilibrium: Hart and Mas-Colell (2005).
Similar results were obtained for MAD dynamics, Hofbauer and Swinkels (1995).
See also Foster and Young (2001) On the impossibility of predicting.
Young (2002) On the limits to rational learning .

Alternative approaches

Hypothesis testing

Procedures that corresponds to a random search of an equilibrium profile.

First approach: prediction of the behavior of the opponents and hypothesis testing, Foster and Young (2003).
Each player state variable has 3 components:
- the empirical frequency of the moves of the opponent during the last s periods
- an hypothesis on this variable
- a counting variable relevant to the mode of the player.

If the hypothesis is rejected, the player chooses a new one at random. Then for specific choices of the parameters convergence in probability to Nash equilibria will occur.

Hypothesis testing

Procedures that corresponds to a random search of an equilibrium profile.

First approach: prediction of the behavior of the opponents and hypothesis testing, Foster and Young (2003).
Each player state variable has 3 components:
- the empirical frequency of the moves of the opponent during the last s periods
- an hypothesis on this variable
- a counting variable relevant to the mode of the player.
If the hypothesis is rejected, the player chooses a new one at random. Then for specific choices of the parameters convergence in probability to Nash equilibria will occur.

A more direct process can be described as follows:
Consider a δ-discretization of the set of mixed strategies $X = \prod_i X^i = \prod_i \Delta(S^i)$ denoted by $\{x_v; v \in V\}$.
Given the payoff function G and $\varepsilon > 0$ at least one of the x_v is for δ small enough an ε-equilibrium of G.
Each player i plays by large blocks L an i.i.d. strategy x^i_v while occasionally testing all his moves in S^i. Given a tolerance bound $\eta > 0$, if one move s^i gives more than the average payoff the block + η, he chooses at random a new point in the grid. Otherwise he keeps playing x^i_v for another block.
Conditions such that the proportion of blocks played with $\{v \in V^*; x_v \ \varepsilon\text{-equilibrium of } G\}$ approaches 1.

This model has been proposed by Foster and Young (2006) and improved by Germano and Lugosi (2007).
Note that this strategy is radically uncoupled, in the sense that not only it does not depend on the payoff function of the opponents but it does not depend on the knowledge of their moves. It is simply a function of the realized payoffs of the player.
Characteristics of this procedure are:
inertia (keep playing if there are small variations)
search (with positive probability experiment)

Coordination is in the choice of the parameters

Pradelski and Young (2012)

Discrete time fictitious play

Consider a finite game with I players having pure strategy sets S^i and mixed strategy sets $X^i = \Delta(S^i)$. The payoff fonction is F from $S = \prod_i S^i$ to $\boldsymbol{R}^I$.
The game is played repeatedly in discrete time and the moves are announced.
Given an n-stage history $h_n = (x_1 = \{x_1^i\}_{i \in I}, x_2, ..., x_n) \in S^n$, the move x_{n+1}^i of player i at stage $n+1$ is a best reply to the "time average moves" of her opponents.

$$x_{n+1}^i \in BR^i(\overline{x}_n^{-i}) \tag{4}$$

where BR^i is the best reply correspondence of player i, from $\Delta(S^{-i})$ to X^i, with $S^{-i} = \prod_{j \neq i} S^i$.

The stage difference is expressed as

$$\overline{x}^i_{n+1} - \overline{x}^i_n = \frac{x^i_{n+1} - \overline{x}^i_n}{n+1}$$

so that (4) can also be written as :

$$\overline{x}^i_{n+1} - \overline{x}^i_n \in \frac{1}{(n+1)}[BR^i(\overline{x}^{-i}_n) - \overline{x}^i_n]. \tag{5}$$

Definition

Brown (1949, 1951)
A sequence $\{x_n\}$ of moves in S satisfies discrete fictitious play (DFP) if (5) holds.

Remark. x^i_n does not appear explicitely any more in (5): the natural state variable of the process is the empirical average $\overline{x}^i_n \in X^i$.

Continuous fictitious play and best reply dynamics

The continuous (formal) counterpart of the above difference inclusion is the differential inclusion:

$$\dot{X}_t^i \in \frac{1}{t}[BR^i(X_t^{-i}) - X_t^i]. \tag{6}$$

The change of time $Z_s = X_{e^s}$ leads to

$$\dot{Z}_s^i \in [BR^i(Z_s^{-i}) - Z_s^i] \tag{7}$$

called continuous best reply (CBR) and studied by Gilboa and Matsui (1991).
One can deduce properties of the initial discrete time process from the analysis of the continuous time counterpart, Harris (1998), Hofbauer and Sorin (2006), Benaim, Hofbauer and Sorin (2005)

Main results

Convergence of $\overline{x}_n$ to the set of Nash equilibria:
- zero-sum games, Robinson (1951) (and convergence of the average realized payoff to the value)
- potential games, Monderer and Shapley (1996)
There exists G such that $\forall s^i, t^i, s^{-i} \in {S^i}^2 \times S^{-i}, \forall i \in I$:

$$F^i(s^i, s^{-i}) - F^i(t^i, s^{-i}) = G(s^i, s^{-i}) - G(t^i, s^{-i}).$$

- no unilateral good properties (but smooth FP does, Fudenberg and Levine (1995))
- no convergence in general: Shapley triangle (1964)

Bayesian approaches

Rational behavior with uncertainty

Merging

Blackwell and Dubins (1962)

$\{X_n\}, n \in N$, random process with values in a finite set Ω

P true distribution, Q belief distribution

Assume P absolutely continuous wtrt Q

Then Q merges to P:

$$\sup_{A \in \mathcal{F}_\infty} |P(A|\mathcal{F}_n) - Q(A|\mathcal{F}_n)| \to 0$$

Rational behavior with uncertainty

Merging

Blackwell and Dubins (1962)

$\{X_n\}, n \in N$, random process with values in a finite set Ω

P true distribution, Q belief distribution

Assume P absolutely continuous wtrt Q

Then Q merges to P:

$$\sup_{A \in \mathcal{F}_\infty} |P(A|\mathcal{F}_n) - Q(A|\mathcal{F}_n)| \to 0$$

Rational behavior with uncertainty

Merging
Blackwell and Dubins (1962)
$\{X_n\}, n \in \boldsymbol{N}$, random process with values in a finite set Ω
P true distribution, Q belief distribution
Assume P absolutely continuous wtrt Q
Then Q merges to P:

$$\sup_{A \in \mathcal{F}_\infty} |P(A|\mathcal{F}_n) - Q(A|\mathcal{F}_n)| \to 0$$

Kalai and Lehrer (1993)
Repeated game
Each player i has a belief on the behavior of the opponents and plays a best reply. This induces a probability Q^i on plays.
If the true probability on plays P is absolutely continuous wrt each Q^i, the players will eventually play like an approximate equilibria.

Application:
games with incomplete information
reputation effects

Related topics:
weak merging
grain of truth
speed of cv

Kalai and Lehrer (1993)
Repeated game
Each player i has a belief on the behavior of the opponents and plays a best reply. This induces a probability Q^i on plays.
If the true probability on plays P is absolutely continuous wrt each Q^i, the players will eventually play like an approximate equilibria.

Application:
games with incomplete information
reputation effects

Related topics:
weak merging
grain of truth
speed of cv

Kalai and Lehrer (1993)
Repeated game
Each player i has a belief on the behavior of the opponents and plays a best reply. This induces a probability Q^i on plays.
If the true probability on plays P is absolutely continuous wrt each Q^i, the players will eventually play like an approximate equilibria.

Application:
games with incomplete information
reputation effects

Related topics:
weak merging
grain of truth
speed of cv

Revision of beliefs and consensus

Aumann (1976) Agreing to disagree
Consider a random parameter and n players with private information:
if the beliefs are common knowledge they must be the same.
Geanakoplos and Polemarchakis (1982)
Explicit process of beliefs revision.

Random parameter $\omega \in \Omega$ each player receives a private signal correlated to ω and plays as a function of her information.
repeated play, observation of the other players moves and revision of the beliefs
Main questions :
convergence of the beliefs: consensus
exhaustivity of the information

Revision of beliefs and consensus

Aumann (1976) Agreing to disagree

Consider a random parameter and n players with private information:

if the beliefs are common knowledge they must be the same.

Geanakoplos and Polemarchakis (1982)

Explicit process of beliefs revision.

Random parameter $\omega \in \Omega$ each player receives a private signal correlated to ω and plays as a function of her information.

repeated play, observation of the other players moves and revision of the beliefs

Main questions :

convergence of the beliefs: consensus

exhaustivity of the information

Revision of beliefs and consensus

Aumann (1976) Agreing to disagree

Consider a random parameter and n players with private information:

if the beliefs are common knowledge they must be the same.

Geanakoplos and Polemarchakis (1982)

Explicit process of beliefs revision.

Random parameter $\omega \in \Omega$ each player receives a private signal correlated to ω and plays as a function of her information.

repeated play, observation of the other players moves and revision of the beliefs

Main questions :

convergence of the beliefs: consensus

exhaustivity of the information

Revision of beliefs and consensus

Aumann (1976) Agreing to disagree

Consider a random parameter and n players with private information:

if the beliefs are common knowledge they must be the same.

Geanakoplos and Polemarchakis (1982)

Explicit process of beliefs revision.

Random parameter $\omega \in \Omega$ each player receives a private signal correlated to ω and plays as a function of her information.

repeated play, observation of the other players moves and revision of the beliefs

Main questions :

convergence of the beliefs: consensus

exhaustivity of the information

Example:
a) line network
2 neighbors with the same signal will not change
b) royal family
false belief will invade

Rosenberg, Solan and Vieille (2009)
Properties of social network
Mossel, Sly and Tamuz (2015)

Example:
a) line network
2 neighbors with the same signal will not change
b) royal family
false belief will invade

Rosenberg, Solan and Vieille (2009)
Properties of social network
Mossel, Sly and Tamuz (2015)

Example:
a) line network
2 neighbors with the same signal will not change
b) royal family
false belief will invade

Rosenberg, Solan and Vieille (2009)
Properties of social network
Mossel, Sly and Tamuz (2015)

Example:
a) line network
2 neighbors with the same signal will not change
b) royal family
false belief will invade

Rosenberg, Solan and Vieille (2009)
Properties of social network
Mossel, Sly and Tamuz (2015)

Informational casacade

Social learning and herd behavior
Random parameter $\omega \in \Omega$ each player receives a private signal correlated to ω and plays once in turn as a function of his information: behavior of the predecessors and private signal

Example

perturbation : with probability p_n player n has no access to the previous performances
evaluation: probability r_n of good prediction at stage n
Peres, Racz, Sly and Stuhl (2018)

Questions:
accuracy of the beliefs
speed of convergence

Informational casacade
Social learning and herd behavior
Random parameter $\omega \in \Omega$ each player receives a private signal correlated to ω and plays once in turn as a function of his information: behavior of the predecessors and private signal

Example

perturbation : with probability p_n player n has no access to the previous performances
evaluation: probability r_n of good prediction at stage n
Peres, Racz, Sly and Stuhl (2018)

Questions:
accuracy of the beliefs
speed of convergence

Learning in extensive games

Fudenberg and Levine and alii

Rationality and equilibrium
evolution of non equilibrium behavior
from passive to active learning
Conjectural equilibrium
Self confirming equilibrium
Selten's horse
experimentation to obtain information on the strategy used by the other players
or to have an impact on their behavior via cascade
knowledge of the game structure
rationality hypotheses
selection of equilibria

Fudenberg and Levine and alii

Rationality and equilibrium
evolution of non equilibrium behavior
from passive to active learning
Conjectural equilibrium
Self confirming equilibrium
Selten's horse
experimentation to obtain information on the strategy used by the other players
or to have an impact on their behavior via cascade
knowledge of the game structure
rationality hypotheses
selection of equilibria

Learning in extensive games

Fudenberg and Levine and alii

Rationality and equilibrium
evolution of non equilibrium behavior
from passive to active learning
Conjectural equilibrium
Self confirming equilibrium
Selten's horse
experimentation to obtain information on the strategy used by the other players
or to have an impact on their behavior via cascade
knowledge of the game structure
rationality hypotheses
selection of equilibria

Fudenberg and Levine and alii

Rationality and equilibrium
evolution of non equilibrium behavior
from passive to active learning
Conjectural equilibrium
Self confirming equilibrium
Selten's horse
experimentation to obtain information on the strategy used by the other players
or to have an impact on their behavior via cascade
knowledge of the game structure
rationality hypotheses
selection of equilibria

Link with dynamical systems

Continuous time dynamics

Hofbauer and Sigmund (1998) *Evolutionary Games and Population Dynamics*, Cambridge U.P.
Sandholm (2010) *Population Games and Evolutionary Dynamics*, M.I.T Press.

Consider games where equilibria $x \in X$ are characterized via variational inequalities (Sorin and Wang, 2016)

$$\langle \Phi(x), x - y \rangle = \sum_{i \in I} \langle \Phi^i(x), x^i - y^i \rangle \geq 0, \quad \forall y \in X. \tag{8}$$

where $X^i \subset \boldsymbol{R}^n$, compact, convex is the strategy set of player i,
$X = \prod_i X^i$.
Φ^i is a map from X to X^i.
Examples are:
- finite games
- $\mathcal{C}^1$ concave games
- population games

Potential games are such that there exist P from X to $\boldsymbol{R}$ with:

$$\langle \Phi^i(x) - \nabla_i P(x), x - y \rangle = 0, \qquad y \in X$$

Dissipative games satisfy:

$$\langle \Phi(x) - \Phi(y), x - y \rangle \leq 0$$

Consider games where equilibria $x \in X$ are characterized via variational inequalities (Sorin and Wang, 2016)

$$\langle \Phi(x), x - y \rangle = \sum_{i \in I} \langle \Phi^i(x), x^i - y^i \rangle \geq 0, \quad \forall y \in X. \qquad (8)$$

where $X^i \subset \boldsymbol{R}^n$, compact, convex is the strategy set of player i,
$X = \prod_i X^i$.
Φ^i is a map from X to X^i.
Examples are:
- finite games
- $\mathcal{C}^1$ concave games
- population games

Potential games are such that there exist P from X to $\boldsymbol{R}$ with:

$$\langle \Phi^i(x) - \nabla_i P(x), x - y \rangle = 0, \qquad y \in X$$

Dissipative games satisfy:

$$\langle \Phi(x) - \Phi(y), x - y \rangle \leq 0$$

Consider games where equilibria $x \in X$ are characterized via variational inequalities (Sorin and Wang, 2016)

$$\langle \Phi(x), x - y \rangle = \sum_{i \in I} \langle \Phi^i(x), x^i - y^i \rangle \geq 0, \quad \forall y \in X. \tag{8}$$

where $X^i \subset \boldsymbol{R}^n$, compact, convex is the strategy set of player i, $X = \prod_i X^i$.
Φ^i is a map from X to X^i.
Examples are:
- finite games
- $\mathcal{C}^1$ concave games
- population games

Potential games are such that there exist P from X to $\boldsymbol{R}$ with:

$$\langle \Phi^i(x) - \nabla_i P(x), x - y \rangle = 0, \qquad y \in X$$

Dissipative games satisfy:

$$\langle \Phi(x) - \Phi(y), x - y \rangle \leq 0$$

Consider games where equilibria $x \in X$ are characterized via variational inequalities (Sorin and Wang, 2016)

$$\langle \Phi(x), x - y \rangle = \sum_{i \in I} \langle \Phi^i(x), x^i - y^i \rangle \geq 0, \quad \forall y \in X. \tag{8}$$

where $X^i \subset \boldsymbol{R}^n$, compact, convex is the strategy set of player i, $X = \prod_i X^i$.
Φ^i is a map from X to X^i.
Examples are:
- finite games
- $\mathcal{C}^1$ concave games
- population games

Potential games are such that there exist P from X to $\boldsymbol{R}$ with:

$$\langle \Phi^i(x) - \nabla_i P(x), x - y \rangle = 0, \qquad y \in X$$

Dissipative games satisfy:

$$\langle \Phi(x) - \Phi(y), x - y \rangle \leq 0$$

Examples of dynamics expressed in terms of Φ. For the first three S^i is finite and $X^i = \Delta(S^i)$.
(1) *Replicator dynamics* (RD) (Taylor and Jonker [53])

$$\dot{x}^i_p = x^i_p[\Phi^i_p(x) - \overline{\Phi}^i(x)], \quad p \in S^i, i \in I,$$

where

$$\overline{\Phi}^i(x) = \langle x^i, \Phi^i(x) \rangle = \sum_{p \in S^i} x^i_p \Phi^i_p(x)$$

is the average evaluation for participant i.
(2) *Brown-von-Neumann-Nash dynamics* (BNN) (Brown and von Neumann [36], Smith [51], Hofbauer [43])

$$\dot{x}^i_p = \hat{\Phi}^i_p(x) - x^i_p \sum_{q \in S^i} \hat{\Phi}^i_q(x), \quad p \in S^i, i \in I,$$

where $\hat{\Phi}^i_q(x) = [\Phi^i_q(x) - \overline{\Phi}^i(x)]^+$ is called the "excess evaluation" of q.

(3) *Smith dynamics* (Smith) (Smith [50])

$$\dot{x}^i_p = \sum_{q \in S^i} x^i_q [\Phi^i_p(x) - \Phi^i_q(x)]^+ - x^i_p \sum_{q \in S^i} [\Phi^i_q(x) - \Phi^i_p(x)]^+, p \in S^i, i \in I,$$

where $[\Phi^i_p(x) - \Phi^i_q(x)]^+$ corresponds to pairwise comparison [48].

(4) *Local/direct projection dynamics* (LP) (Dupuis and Nagurney [38], Lahkar and Sandholm [45])

$$\dot{x}^i = \Pi_{T_{X^i}(x^i)}[\Phi^i(x)], \quad i \in I,$$

where $T_{X^i}(x^i)$ denotes the tangent cône to X^i at x^i.

(5) *Global/target projection dynamics* (GP) (Friesz et al. [39], Tsakas and Voorneveld [54])

$$\dot{x}^i = \Pi_{X^i}[x^i + \Phi^i(x)] - x^i, \quad i \in I.$$

(6) *Best reply dynamics* (BR) (Gilboa and Matsui [40])

$$\dot{x}^i \in BR^i(x) - x^i, \quad i \in I,$$

where

$$BR^i(x) = \{y^i \in X^i,\ \langle y^i - z^i, \Phi^i(x)\rangle \geq 0, \forall z^i \in X^i\}.$$

Properties

Properties expressed in terms of Φ.

Definition

Dynamics $\mathcal{B}_\Phi$ satisfies:

i) *positive correlation (PC)* (Sandholm [47]) if:

$$\langle \mathcal{B}^i_\Phi(x), \Phi^i(x)\rangle > 0, \quad \forall i \in I, \forall x \in X \text{ s.t. } \mathcal{B}^i_\Phi(x) \neq 0.$$

(This corresponds to MAD (myopic adjustment dynamics) (Swinkels [52]): given a configuration, any unilateral change should increase the evaluation);

ii) *Nash stationarity* if: for $x \in X$, $\mathcal{B}_\Phi(x) = 0$ if and only if x is an equilibrium of $\Gamma(\Phi)$.

Proposition

(RD), (BNN), (Smith), (LP), (GP) and (BR) satisfy (PC).
all except (RD) satisfy (NS).

Lyapounov functions

Consider a potential game and a dynamic satisfying (PC).
Then P is a Lyapounov function

Similar properties for dissipative games with ad hoc Lyapounov functions.

Further results:

elimination of dominated strategies
stability of pure strict equilibria
convergence to a profile from inside implies Nash
Lyapounov implies Nash

Hofbauer, Sandholm, Panayotis, Coucheney, Gaujal, Leslie, Laraki, Staugigl, Viossat, ...

0-sum games
Population games
Congestion games
Extension to composite games

Concluding comments

- connection : discrete/ continuous time
Stochastic approximation
- new concepts= attractors, ICT
pertubation of games and components of equilibra
- links
FP in terms of strategies or in term of payoffs (different interpretation)
perturbed best-reply / smooth FP
connection with no-regret procedures
RD and external consistency
Time average RD and Best reply dynamics
- no regret in learning, games and convex optimization
references on line learning/convex analysis
Bubeck
Hazan
Shalev Schwartz

Bibliography I

- AUER P., CESA-BIANCHI N., FREUND Y. AND R.E. SHAPIRE (2002) The nonstochastic multiarmed bandit problem, *SIAM J. Comput.*, **32**, 48-77.
- BENAIM M., (1999) *Dynamics of Stochastic Algorithms* , Séminaire de Probabilités, XXIII, Azéma J. and alii eds, LNM **1709**, Springer, 36-72.
- BENAIM M., J. HOFBAUER AND S. SORIN (2005) Stochastic approximations and differential inclusions, *SIAM J. Opt. and Control*, **44**, 328-348.
- BENAIM M., J. HOFBAUER AND S. SORIN (2006) Stochastic approximations and differential inclusions. Part II: applications, *Mathematics of Operations Research*, **31**, 673-695.

Bibliography II

- Blackwell D. (1956) An analog of the minimax theorem for vector payoffs, *Pacific Journal of Mathematics*, **6**, 1-8.
- Blum A. and Y. Mansour (2007) From external to internal regret, *Journal of Machine Learning Reserach*, **8**, 1307-1324.
- Brown G. W.(1951) Iterative solution of games by fictitious play, in Koopmans T.C. (ed.) *Activity Analysis of Production and Allocation* , Wiley, 374-376.
- Cesa-Bianchi N. and G. Lugosi (2003) Potential-based algorithms in on-line prediction and game theory, *Machine Learning*, **51**, 239-261.
- Cesa-Bianchi N. and G. Lugosi (2006) *Prediction, Learning and Games*, Cambridge University Press.

- FOSTER D. AND R. VOHRA (1999) Regret in the on-line decision problem, *Games and Economic Behavior*, **29**, 7-35.
- FREUND Y. AND R.E. SCHAPIRE (1999) Adaptive game playing using multiplicative weights, *Games and Economic Behavior*, **29**, 79-103.
- FUDENBERG D. AND D. K. LEVINE (1995) Consistency and cautious fictitious play, *Journal of Economic Dynamics and Control*, **19**, 1065-1089.
- FUDENBERG D. AND D. K. LEVINE (1998) *The Theory of Learning in Games*, MIT Press.
- FUDENBERG D. AND D. K. LEVINE (1999) Conditional universal consistency *Games and Economic Behavior*, **29**, 104-130.

Bibliography IV

- GAUNERSDORFER A. AND J. HOFBAUER (1995) Fictitious play, Shapley polygons and the replicator equation, *Games and Economic Behavior*, **11**, 279-303.
- GILBOA I. AND A. MATSUI (1991) Social stability and equilibrium, *Econometrica*, **59**, 859-867.
- HANNAN J. (1957) Approximation to Bayes risk in repeated plays, *Contributions to the Theory of Games, III*, Drescher M., A.W. Tucker and P. Wolfe eds., Princeton University Press, 97-139.
- HART S. (2005) Adaptive heuristics, *Econometrica*, **73**, 1401-1430.
- HART S. AND A. MAS-COLELL, (2000) A simple adaptive procedure leading to correlated equilibrium, *Econometrica*, **68**, 1127-1150.

- HART S. AND A. MAS-COLELL (2001) A general class of adaptive strategies, *Journal of Economic Theory*, **98**, 26-54.
- HART S. AND A. MAS-COLELL (2003) Regret-based continuous time dynamics, *Games and Economic Behavior*, **45**, 375-394.
- HOFBAUER J. AND W. H. SANDHOLM (2002) On the global convergence of stochastic fictitious play, *Econometrica*, **70**, 2265-2294.
- HOFBAUER J. AND K. SIGMUND, (1998) *Evolutionary Games and Population Dynamics*, Cambridge U.P.
- HOFBAUER J. AND K. SIGMUND, (2003) Evolutionary Games Dynamics, *Bulletin A.M.S.*, **40**, 479-519.

Bibliography VI

- HOFBAUER J. AND S. SORIN, (2006) Best response dynamics for continuous zero-sum games, *Discrete and Continuous Dynamical Systems -Series B*, **6**, 215-224.
- HOFBAUER J., S. SORIN AND Y. VIOSSAT (2009) Time average replicator and best reply dynamics, *Mathematics of Operations Research*, **34**, 263-269.
- LITTLESTONE N. AND M.K. WARMUTH (1994) The weighted majority algorithm, *Information and Computation*, **108**, 212-261.
- LUGOSI G., S. MANNOR AND G. STOLTZ (2008) Strategies for prediction under imperfect monitoring, *Mathematics of Operations Research*, **33**, 513-528.
- ROBINSON J., (1951) An iterative method of solving a game, *Annals of Mathematics*, **54**, 296-301.

Bibliography VII

- Shapley L.S. (1964) Some topics in two-person games, in *Advances in Game Theory*, M. Dresher, L.S. Shapley and A. W. Tucker (eds.), Annals of Mathematical Studies, **52**, Princeton University Press, 1-28.
- SORIN S. (2009) Exponential weight algorithm in continuous time, *Mathematical Programming*, Ser. B **116**, 513-528.
- Sorin S., *On some global and unilateral adaptive dynamics*, in Evolutionary Game Dynamics, K. Sigmund ed., Proceedings of Symposia in Applied Mathematics, **69**, A.M.S., 81-109, 2011.
- Sorin S. and C. Wang, *Finite composite games: equilibria and dynamics*, Journal of Dynamics and Games, **3**, 101-120, 2016.

- STOLTZ G. AND G. LUGOSI (2005) Internal Regret in On-Line Portfolio Selection, *Machine Learning*, **59**, 125-159.
- VIOSSAT Y. (2007) The replicator dynamics does not lead to correlated equilibria, *Games and Economic Behavior*, **59**, 397-407
- (MR0039220)
  G. W. Brown and J. von Neumann,
  Solutions of games by differential equations,
  in *Contibutions to the Theory of Games, I* (ed. H.W. Kuhn and A.W. Tucker), *Ann. Math. Studies*, **24** (1950), 73–79.

(MR639734) [10.1287/trsc.14.1.42]
S. C. Dafermos,
Traffic equilibrium and variational inequalities,
Transportation Sci., **14** (1980), 42–54.

(MR1246835) [10.1007/BF02073589]
P. Dupuis and A. Nagurney,
Dynamical systems and variational inequalities,
Ann. Oper. Res., **44** (1993), 9–42.

(MR1307808) [10.1287/opre.42.6.1120]
T. L. Friesz, D. Bernstein, N. J. Mehta, R. L. Tobin and S. Ganjalizadeh,
Day-to-day dynamic network disequilibria and idealized traveler information systems,
Oper. Res., **42** (1994), 1120–1136.

Bibliography X

(MR1106514) [10.2307/2938230]
I. Gilboa and A. Matsui,
Social stability and equilibrium,
Econometrica, **59** (1991), 859–867.

(MR929023) [10.1287/trsc.22.1.39]
P. T. Harker,
Multiple equilibrium behaviors on networks,
Transportation Sci., **22** (1988), 39–46.

S. Hart and A. Mas-Colell,
Uncoupled dynamics do not lead to Nash equilibrium,
Am. Econ. Rev., **93** (2003), 1830–1836.

[10.1556/Select.1.2000.1-3.8]
J. Hofbauer,
From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS,
Selection, **1** (2000), 81–88.

(MR2887006) [10.1016/j.jet.2009.01.007]
J. Hofbauer and W. H. Sandholm,
Stable games and their dynamics,
J. Econom. Theory, **144** (2009), 1665–1693.

(MR2479677) [10.1016/j.geb.2008.02.002]
R. Lahkar and W. H. Sandholm,
The projection dynamic and the geometry of population games,
Games Econom. Behav., **64** (2008), 565–590.

Bibliography XII

(MR1393599) [10.1006/game.1996.0044]
D. Monderer and L. S. Shapley,
Potential games,
Games Econom. Behav., **14** (1996), 124–143.

(MR1816280) [10.1006/jeth.2000.2696]
W. H. Sandholm,
Potential games with continuous player sets,
J. Econom. Theory, **97** (2001), 81–108.

W. H. Sandholm,
Population Games and Evolutionary Dynamics,
MIT Press, Cambridge, MA, 2011.

(MR551841) [10.1016/0191-2615(79)90022-5]
M. J. Smith,
The existence, uniqueness and stability of traffic equilibria,
Transportation Res. Part B, **13** (1979), 295–304.

(MR878966) [10.1287/trsc.18.3.245]
M. J. Smith,
The stability of a dynamic model of traffic assignment – an application of a method of Lyapunov,
Transportation Sci., **18** (1984), 245–252.

(MR774332) [10.1007/BF00935463]
M. J. Smith,
A descent algorithm for solving monotone variational inequalities and monotone complementarity problems,
J. Optim. Theory Appl., **44** (1984), 485–496.

(MR1227919) [10.1006/game.1993.1025]
J. M. Swinkels,
Adjustment dynamics and rational play in games,
Games Econom. Behav., **5** (1993), 455–484.

(MR0489983) [10.1016/0025-5564(78)90077-9]
P. D. Taylor and L. B. Jonker,
Evolutionary stable strategies and game dynamics,
Math. Biosci., **40** (1978), 145–156.

(MR2574130) [10.1016/j.geb.2009.01.003]
E. Tsakas and M. Voorneveld,
The target projection dynamic,
Games Econom. Behav., **67** (2009), 708–719.
Aumann (1976) Agreing to disagree, Annals of Stat
Blackwell and Dubins. Merging of opinions with increasing information, Ann Math. Stats, 1962.

Kalai and Lehrer. Rational learning leads to Nash equilibria, Ecta, 1993.

Geanakoplos and Polemarchakis. We can't disagree forever, JET, 1982.

Rosenberg, Solan and Vieille. Informational externalities and emergence of consensus, GEB, 2009.

Mossel E., A. Sly and O. Tamuz (2015) Strategic learning and the topology of social networks, Econometrica, 83, 1755-1794.

Peres Y., Racz M., Sly A. and I. Stuhl (2018) How fragile are information cascades? arcXiv:1711.04024v2