An adjusted payoff-based procedure

AlgoGT, Grenoble
20-21 Juin 2011.

Mario Bravo

UPMC
Institut de Mathématiques de Jussieu
Équipe Combinatoire et Optimisation

Outline

Motivation

We dispose of a N-player normal form game. For instance:

	L	m	R
T	$(0,0)$	$(1,0)$	$(0,1)$
M	$(0,1)$	$(0,0)$	$(1,0)$
B	$(1,0)$	$(0,1)$	$(0,0)$

Motivation

Information at each time.
Matrix unknown
Actions played:

- After each time, Player 1 is informed of its own payoff. The opponent's moves are unknown.

	L	m	R
T	?	?	?
M	?	?	?
B	?	?	?

Motivation

Information at each time.
Matrix unknown
Actions played: (M, R)

- After each time, Player 1 is informed of its own payoff. The opponent's moves are unknown.

	L	m	R
T	?	?	?
M	?	?	$(1, ?)$
B	?	?	?

Motivation

Information at each time.
Matrix unknown
Actions played: (M, R)-(T, L)

- After each time, Player 1 is informed of its own payoff. The opponent's moves are unknown.

	L	m	R
T	$(0, ?)$	$?$	$?$
M	$?$	$?$	$(1, ?)$
B	$?$	$?$	$?$

Motivation

Information at each time.
Matrix unknown
Actions played: (M, R)-(T, L)-(B, L)

- After each time, Player 1 is informed of its own payoff. The opponent's moves are unknown.

	L	m	R
T	$(0, ?)$	$?$	$?$
M	$?$	$?$	$(1, ?)$
B	$(1, ?)$	$?$	$?$

Motivation

Information at each time.
Matrix unknown
Actions played: (M, R)-(T, L)-(B, L)

- After each time, Player 1 is informed of its own payoff. The opponent's moves are unknown. Eventually:

	L	m	R	$R2$
T	$(0, ?)$	?	?	?
M	?	?	$(1, ?)$	?
B	$(1, ?)$	?	?	?

The model

- Setting
  - $A = \{1, \ldots, N\}$ is the set of players.
  - $S^i = \{1, \ldots, N_i\}$ is the finite set of strategies for player i.
  - $\Delta^i = \{z \in \mathbb{R}^{|S^i|}, z^s \geq 0, \sum_{s \in S^i} z^s = 1\}$ is the mixed strategy set for player i.
  - Payoff to each player i, $G^i : S = S^i \times S^{-i} \to \mathbb{R}$, where $S^{-i} = \prod_{j \neq i} S^j$. We keep the notation for its multilinear extension to $\Delta = \prod_j \Delta_j$.
- This game is repeated infinitely and we denote by $n \in \mathbb{N}$ the time.
  - $\sigma_n^i \in \Delta_i$ is the mixed strategy for player i at time n .
  - s_n^i is the move of Player i at stage n and $s_n = (s_n^1, \ldots, s_n^N) \in S$ is the profile of moves at stage n.
  - g_n^i is the payoff received by Player i at stage n, i.e., $g_n^i = G^i(s_n)$.
  - $\mathcal{F}_n$ is the σ-algebra generated by the history $\{s_1, s_2, \ldots, s_n\}$ up to time n.

- Then the discrete dynamics goes like
  - At stage n, Player i

$$x_{n-1}^i \in R^{N_i}$$
$$\downarrow$$
$$\sigma_n^i = \sigma^i(x_{n-1}^i) \qquad \text{Choice Rule}$$
$$\downarrow$$
$$s_n^i$$
$$\downarrow$$
$$g_n^i = G(s_n^i, s_n^{-i})$$
$$\downarrow$$
$$x_n^i = f(x_{n-1}^i, g_n^i, s_n^i, n) \qquad \text{Updating Rule}$$

where x_n^i is a state variable (*the perception vector*) for each player and we assume that $\sigma^i(\cdot)$ is strictly positive on each component for all $n \in \mathbb{N}$.

- Let $X \subseteq \mathbb{R}^{N \cdot \sum_i |S^i|}$ be the state space for the perception vectors $x = (x^1, \ldots, x^N)$.
- We assume the following on the decision rule $\sigma : X \to \prod_i \Delta^i$:

$$\text{the function } \sigma^i : \mathbb{R}^{|S^i|} \to \Delta^i \text{ is continuous, and}$$

$$\text{for all } s \in S^i \text{ and } x^i \in \mathbb{R}^{|S^i|}, \; \sigma^{is}(x^i) > 0.$$

The tool

- We want to study the asymptotic behavior of a process of the form

$$z_{n+1} - z_n = \gamma_{n+1}\big(H(z_n) + V_{n+1}\big) \tag{1}$$

where V_n is a (deterministic or random) perturbation.

- The idea is to connect the asymptotic behavior of (1) with the asymptotic behavior of the continuous dynamics

$$\dot{z} = H(z), \tag{2}$$

by means of the following general theorem:

The tool

- **Theorem (Benaim, 99).**
In the discrete process (1), if H is a globally Lipschitz function and the following holds
  (a) $\gamma_n \geq 0$, $\sum_n \gamma_n = +\infty$ and $\gamma_n \to 0$,
  (b) For all $T > 0$
$$\lim_{n\to+\infty} \sup_k \left(\left\| \sum_{i=n}^{k-1} \gamma_{i+1} V_{i+1} \right\| : k = n+1, \ldots, m(\sum_{j=1}^{n} \gamma_j + T) \right) = 0,$$
where $m(t)$ is the largest integer l such that $t \geq \sum_{j=1}^{l} \gamma_j$, and
  (c) $\sup_{n\in\mathbb{N}} \|z_n\| < +\infty$,
then the set of limits points of the sequence $(z_n)_{n\geq 1}$ is an ICT set of the dynamics (12).

Back to the model

In this work we consider the following updating rule

Updating rule

$$x^{is}_{n+1} = \begin{cases} (1 - \frac{1}{\theta^{is}_n})x^{is}_n + \frac{1}{\theta^{is}_n} g^i_{n+1} & \text{if } s = s^i_{n+1}, \\ x^{is}_n & \text{otherwise,} \end{cases}$$

where θ^{is}_n is the number of times that strategy s has been used by player $i \in A$ up to time n. Set λ^{is}_n as the frequency of action s for player i up to time n, i.e., $\lambda^{is}_n = \frac{\theta^{is}_n}{n+1}$.
The variation of λ_n can be computed as

$$\lambda^{is}_{n+1} - \lambda^{is}_n = \frac{1}{n+1}\left(\mathbb{1}_{\{s=s^i_{n+1}\}} - \lambda^{is}_n + O\Big(\frac{1}{n}\Big)\right).$$

Then we can write the previous process in the following manner

$$\begin{cases} x_{n+1}^{is} - x_n^{is} = \frac{1}{n+1}\left[\frac{\sigma^{is}(x_n^i)}{\lambda_n^{is}}(G^i(s,\sigma^{-i}(x_n)) - x_n^{is}) + U_{n+1}^{is}\right], \\ \\ \lambda_{n+1}^{is} - \lambda_n^{is} = \frac{1}{n+1}\left[\sigma^{is}(x_n^i) - \lambda_n^{is} + M_{n+1}^{is}\right], \end{cases} \tag{APD}$$

where the noise terms are explicitly

$$U_{n+1}^{is} = \frac{1}{\lambda_n^{is}}(g_{n+1}^i - x_n^{is})\mathbb{1}_{\{s=s_{n+1}^i\}} - \left[\frac{\sigma_n^{is}}{\lambda_n^{is}}(G^i(s,\sigma^{-i}(x_n)) - x_n^{is})\right],$$

$$M_{n+1}^{is} = \mathbb{1}_{\{s=s_{n+1}^i\}} - \sigma_n^{is} + O\left(\frac{1}{n}\right).$$

Our process is related with the following continuous dynamics.

Continuous dynamics

$$\begin{aligned} \dot{x}_t^{is} &= \frac{\sigma^{is}(x_t^i)}{\lambda_t^{is}}\left(G^i(s, \sigma^{-i}(x_t)) - x_t^{is}\right) \\ \dot{\lambda}_t^{is} &= \sigma^{is}(x_t^i) - \lambda_t^{is}, \end{aligned} \tag{D}$$

CMS procedure (Cominetti et al, 2010)

Our procedure turns out to be a variation of the following process studied in [CMS10].

$$x_{n+1}^{is} = \begin{cases} (1 - \frac{1}{n+1})x_n^{is} + \frac{1}{n+1} g_{n+1}^i & \text{if } s = s_{n+1}^i \\ x_n^{is} & \text{otherwise,} \end{cases}$$

For convenience, we write the process as

$$x_{n+1}^{is} - x_n^{is} = \frac{1}{n+1} \sigma^{is}(x_n^i)\big(G^i(s, \sigma^{-i}(x_n)) - x_n^{is} + \tilde{U}_{n+1}^i\big) \tag{3}$$

where $\mathbb{E}(\tilde{U}_{n+1}^i | \mathcal{F}_n) = 0$.
The authors link the asymptotic behavior of (3) with the asymptotic behavior of the continuous dynamics

$$\dot{x}_t^{is} = \sigma^{is}(x_t^i)\big(G^i(s, \sigma^{-i}(x_t)) - x_t^{is}\big) \tag{D'}$$

Asymptotic analysis

Observe that, since $\sigma_n^{is} \geq \overline{\sigma} > 0$, then $\liminf \lambda_n^{is} \geq \overline{\sigma}$ almost surely for all $n \in \mathbb{N}$, $i \in \{1, \ldots, N\}$ and $s \in S^i$.

Proposition

The process (APD) converges almost surely to an *ICT* set for the continuous dynamics (D).

Let $F : X \to X$ defined by $F^{is}(x) = G^i(s, \sigma^{-i}(x))$.

Proposition

Asumme that, for every $i \in A$, the function σ^i is Lipschitz for the infinity norm. Assume also that F is contracting for the infinity norm. Then there exists a unique rest point $(x_*, \sigma(x_*)) \in X \times \Delta$ of (D). Furthermore, the set $\{(x_*, \sigma(x_*))\}$ is a global attractor for (D) and the process (APD) converges almost surely to $(x_*, \sigma(x_*))$.

Logit rule

From now on we assume that the players use the *Logit rule*, i.e., for each i and all $s \in S^i$

$$\sigma^{is}(x^i) = \frac{\exp(\beta_i x^{is})}{\sum_{r \in S^i} \exp(\beta_i x^{ir})} \tag{4}$$

Lemma (CMS10)

If $(x, \lambda) \in X \times \Delta$ is a rest point of the , then $\lambda = \sigma(x)$ is a Nash equilibrium of a game where the strategy set for the each player i is Δ^i and her payoff $\overline{G}^i : \Delta \to \mathbb{R}$ is given by

$$\overline{G}^i(\pi) = \sum_{s \in S^i} \pi^{is} G^i(s, \pi^{-i}) - \frac{1}{\beta_i} \sum_{s \in S^i} \pi^{is} \big(\ln(\pi^{is}) - 1 \big). \tag{5}$$

Summary

Almost sure convergence

Proposition

If $2\eta\alpha < 1$, the discrete process (APD) converges almost surely to the unique rest point $(x_*, \sigma(x_*))$ of the dynamics (D).

Here η is the maximal unilateral deviation payoff that a single player can face, i.e.,

$$\eta = \max_{\substack{i \in A, s \in S^i \\ r_1, r_2 \in \tilde{S}^{-i}}} |G^i(s, r_1) - G^i(s, r_2)|, \tag{6}$$

where $\tilde{S}^{-i} = \{(r_1, r_2) \in S^{-i} \times S^{-i}; r_1^k \neq r_2^k \text{ for exactly one } k\}$, and

$$\alpha = \max_{i \in A} \sum_{j \neq i} \beta_j$$

- $\rho(H) = \max\{Re(\mu_j)$ with μ_j $, j \in \{1, \ldots, k\}$ eigenvalues of $H\}$.

We redefine, for simplicity, the two dynamics involved as

$$\underbrace{\dot{x} = \Phi(x)}_{CMS} \qquad \underbrace{(\dot{x}, \dot{\lambda}) = \Psi(x, \lambda)}_{Adjusted}.$$

Lemma

Assume that $2\eta\alpha < 1$. Let (x_*, λ_*) and x_* be the unique rest points of the dynamics (D) and (D') respectively. Then

$$-1 \leq \rho(\nabla\Psi(x_*, \lambda_*)) < -\frac{1}{2} \leq -\frac{N}{\sum_{k\in A} |S^k|} \leq \rho(\nabla\Phi(x_*)) < 0. \qquad (7)$$

Proposition

Under the assumptions of the previous Lemma, the following estimates hold,

(*i*) for almost all trajectories of (3)

$$n^{\delta}(x_n - x_*) \longrightarrow 0, \quad \text{as} \;\; n \longrightarrow +\infty,$$

for every $\delta \in (0, |\rho(\nabla\Phi(x_*))|)$,

(*ii*) for almost all trajectories of (APD)

$$n^{\delta}\big((x_n, \lambda_n) - (x_*, \lambda_*)\big) \longrightarrow 0, \quad \text{as} \;\; n \longrightarrow +\infty,$$

for every $\delta \in (0, 1/2)$.

Example

	L	m	R
T	$(0,0)$	$(1,0)$	$(0,1)$
M	$(0,1)$	$(0,0)$	$(1,0)$
B	$(1,0)$	$(0,1)$	$(0,0)$

Rest Point$= (x_*, \lambda_*) = (x_*, \sigma(x_*))$, where $x_* = (1/3, 1/3, 1/3)$.

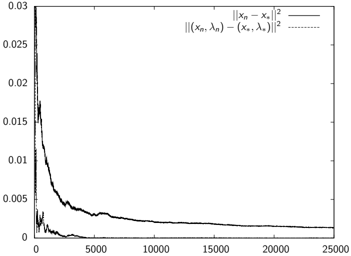


Figure: $\|(x_n, \lambda_n) - (x_*, \lambda_*)\|^2$ v/s $\|x_n - x_*\|^2$.

Summary

Covergence with positive probability

The crucial concept is the following

Definition

Let $(z_n)_n$ be a discrete stochastic process with state space Z. A point $z \in Z$ is attainable by $(z_n)_n$ if for each $m \in \mathbb{N}$ and every open neighborhood U of z, $\mathbb{P}(\exists n \geq m,\ z_n \in U) > 0$.

Lemma

Fix $\lambda = (\lambda^1, \ldots, \lambda^N) \in \Delta$. Set $x^i \in \mathbb{R}^{|S^i|}$ such that $x^{is} = G^i(s, \lambda^{-i})$ for all $s \in S^i$ and put $x = (x^1, \ldots, x^N) \in X$. Then, the point $(x, \lambda) \in X \times \Delta$ is attainable by the process $(x_n, \lambda_n)_n$. In particular, any rest point of the dynamics (D) is attainable.

Let $\mathcal{Y}$ be the set of rest points of the dynamics (D). Recall that $\mathcal{L}(z_n)$ is the limit set of the sequence $(z_n)_n$.

Lemma

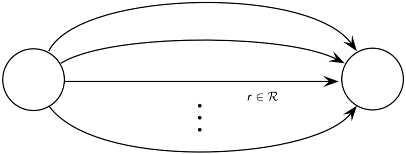
Assume that $1 \leq 2\eta\alpha < 2$. Then $\mathcal{Y}$ reduces to one point (x_*, λ_*) which is an attractor for (D).

Let $B(\mathcal{A})$ the basin of attraction corresponding to the attractor $\mathcal{A}$.

Proposition

If an attractor $\mathcal{A}$ for the dynamics (D) satisfies that $B(\mathcal{A}) \cap \mathcal{Y} \neq \emptyset$, then $\mathbb{P}(\mathcal{L}(x_n, \lambda_n) \subseteq \mathcal{A}) > 0$. In particular, under the Logit decision rule decision, if $1 \leq 2\eta\alpha < 2$, then $\mathcal{Y}$ reduces to one point (x_*, λ_*) and $\mathbb{P}((x_n, \lambda_n) \to (x_*, \lambda_*)) > 0$.

A traffic game



Each route $r \in \mathcal{R}$ in the network is characterized by an increasing sequence of values $c_1^r \leq \cdots \leq c_N^r$ where c_u^r represents the average travel time when r carries a load of u users.

This traffic game is shown to be a potential game in the sense that there exists a function $\Lambda : [0,1]^{N \times |\mathcal{R}|} \longrightarrow \mathbb{R}$ such that

$$\frac{\partial \Lambda}{\partial \lambda^{is}}(\lambda) = G^i(s, \lambda^{-i}),$$

and

$$-c_u^{\mathbf{r}^i} = G^i(\mathbf{r}).$$

We assume that $\beta_i = \beta$ for all i.

A traffic game

Observe that the value η translates to

$$\eta = \max\{\eta^r_u\,;\ r \in \mathcal{R},\ 2 \leq u \leq N\} = \max\{c^r_u - c^r_{u-1}\,;\ r \in \mathcal{R},\ 2 \leq u \leq N\}.$$

Proposition

If $\eta\beta < 1$, (D) has a unique rest point $(x_*, \lambda_*) \in X \times \Delta$ which is symmetric in the sense that $x_* = (\hat{x}, \ldots, \hat{x})$ and $\lambda_* = (\hat{\lambda}, \ldots, \hat{\lambda}) = \sigma(x_*)$. Furthermore, $\{(x_*, \lambda_*)\}$ is an attractor for (D) and $\mathbb{P}((x_n, \lambda_n) \to (x_*, \lambda_*)) > 0$.

Summary

Nonconvergence

If the smoothing parameters grow then the asymptotic behavior can change completely.
We consider the following "class" of games: 2 players, Logit rule with identical smoothing parameter $\beta > 0$ and payoffs given by

$$\begin{pmatrix} 0 & a & -b \\ -b & 0 & a \\ a & -b & 0 \end{pmatrix},$$

where $a > b > 0$ (the *good* Rock-Scissors-Paper game).

Lemma

If the parameter $\beta > 0$ is sufficiently large then there exists an equilibrium $(\underline{x}, \underline{\lambda})$ which is linearly unstable, i.e., there exists an eigenvalue μ of the matrix $\nabla\Psi(\underline{x}, \underline{\lambda})$ such that $\mathrm{Re}(\mu) > 0$.

Then he have the following

Lemma

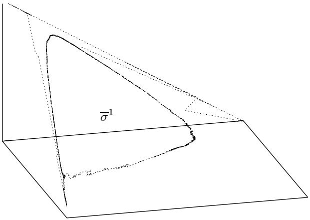
Under the hypothesis of the previous Lemma, there exists a parameter $\beta > 0$ sufficiently large and at least one equilibrium $(\underline{x}, \underline{\lambda}) \in X \times \Delta$ of (D) such that, for the discrete procedure (APD),

$$\mathbb{P}\big(\lim_{n \to +\infty} (x_n, \lambda_n) = (\underline{x}, \underline{\lambda})\big) = 0.$$

Observe that for the game above, the equilibrium is independent of β (in fact, is the centroid of the product of simplexes) and we had: for β small, $(x_n, \lambda_n) \to (\underline{x}, \underline{\lambda})$ almost surely.

Come back to the example

It seems to be a cycle that attracts the trajectories (σ_n^1, $\beta = 4$).



However, consider the simple zero-sum game where $G_i = G$

$$G = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then the (unique) equilibrium $\{(x_*, \sigma(x_*))\}$ where $x_*^1 = x_*^2 = (-e^{\beta}/(1+e^{\beta}), 1/(1+e^{\beta}))^T$ is an attractor and then $\mathbb{P}((x_n, \lambda_n) \to (x_*, \sigma(x_*))) > 0$ for all $\beta > 0$.

Random Environment

We allow the Nature to control the matrices that the players receive at each time $n \in \mathbb{N}$.

- The matrix payoff are now $G_i : S \times W \to \mathbb{R}$ where W is a finite set
- The player i scores $\overline{g}_n^i = G^i(s_n, w_n)$ at time n, where $(w_n)_n$ is a controlled (by the frequencies of play) Markov Chain, that is, there exists a family of transition matrices $(P(\lambda))_\lambda$ where
- $P_{(w_n,w)}(\lambda_n) = \mathbb{P}(w_{n+1} = w \mid \mathcal{F}_n)$, where now $\mathcal{F}_n$ denote the σ-algebra generated by the history $(s_1, w_1, \ldots s_n, w_n)$ up to time n

We assume that

(A1) For a fixed $\lambda \in \Delta$ the Markov Chain with probability transition $P(\lambda)$ has a unique invariant probability $\tau(\lambda) \in \Delta(W)$ (the set of probabilities over W).

(A2) The function $\mathbf{P} : \Delta \to M$ (the set of stochastic matrices of dimension of $|W| \times |W|$), $\mathbf{P}(\lambda) = P_{(\cdot,\cdot)}(\lambda)$ is of class $\mathcal{C}^1$.

Then the new (analogous) scheme is

Updating rule

$$x_{n+1}^{is} = \begin{cases} (1 - \frac{1}{\theta_n^{is}})x_n^{is} + \frac{1}{\theta_n^{is}}\overline{g}_{n+1}^{i} & \text{if } s = s_{n+1}^i, \\ x_n^{is} & \text{otherwise,} \end{cases}$$

And the coupled procedure is

$$
\begin{aligned}
x^{is}_{n+1} - x^{is}_n &= \frac{1}{n+1}\Big[\frac{\sigma^{is}(x^i_n)G_i(s,\sigma^{-i}(x_n),\tau(\lambda_n))}{\lambda^{is}_n} + \frac{\overline{U}^{is}_{n+1}}{\lambda^{is}_n}\Big] \\
\lambda^{is}_{n+1} - \lambda^{is}_n &= \frac{1}{n+1}\big[\sigma^{is}_n - \lambda^{is}_n + M^{is}_{n+1}\big],
\end{aligned}
\qquad \text{(APDp)}
$$

where

$$
\overline{U}^{is}_{n+1} = (\overline{g}^{i}_{n+1} - x^{is}_n)\mathbb{1}_{\{s=s^i_{n+1}\}} - (\sigma^{is}(x^i_n)G_i(s,\sigma^{-i}(x_n),\tau(\lambda_n)) - x^{is}_n)
$$

$$
M^{is}_{n+1} = \mathbb{1}_{\{s=s^i_{n+1}\}} - \sigma^{is}_n + O\Big(\frac{1}{n}\Big).
$$

Observe that the process $(\overline{U}_n)_n$ is not a martingale difference sequence while the $(\overline{M}_n)_n$ sequence (up to a vanishing term) still is.

The associated continuous dynamics in this case is

$$\begin{cases} \dot{x}_t^{is} = \frac{\sigma^{is}(x_t^i)}{\lambda_t^{is}} \left[G^i(s, \sigma^{-i}(x_t), \tau(\lambda)) - x_t^{is} \right], \\ \\ \dot{\lambda}_t^{is} = \sigma^{is}(x_t^i) - \lambda_t^{is}. \end{cases} \tag{D''}$$

Proposition

Under the Logit decision rule, let η_w be the maximal unilateral deviation that a player can face for a fixed $w \in W$ (see (6)). Set $\overline{\eta} = \max_w \eta_w$ and $\tilde{\eta} = \max_{i,\mathbf{s},w,w'} |G^i(\mathbf{s}, w) - G^i(\mathbf{s}, w')|$. If

$$\begin{cases} 2\overline{\eta}\alpha + \tilde{\eta}\sqrt{|W|}k < 1, \\ 2\max_i \beta_i < 1, \end{cases}$$

where k is the $\|\cdot\|_\infty$-Lipschitz constant for the function τ, then the discrete process (APDp) converges almost surely to the global attractor $\{(x_*, \lambda_*)\}$ for the dynamics (D").

The constant case

Assume that $\mathbf{P}(\lambda) = P$ for every $\lambda \in \Delta$.
The condition to ensure almost sure convergence reduces to $2\overline{\eta}\alpha < 1$.
Let $\mathcal{Y}$ be the set of rest points of (D").
Also, we have the following

Proposition

If an attractor $\mathcal{A}$ for the dynamics (D") verifies that $\mathcal{A} \subseteq \overline{\mathcal{Y}}$, then $\mathbb{P}(\mathcal{L}(x_n, \lambda_n) \subseteq \mathcal{A}) > 0$. In particular, under the Logit decision rule decision, if $1 \leq 2\overline{\eta}\alpha < 2$, then $\overline{\mathcal{Y}}$ reduces to one point (x_*, λ_*) and $P((x_n, \lambda_n) \to (x_*, \lambda_*)) > 0$.

Merci!