

# On the robustness of learning in games with stochastically perturbed payoff observations

Mario Bravo

Universidad de Santiago de Chile  
mario.bravo.g@usach.cl

Joint work with Panayotis Mertikopoulos (CNRS, Grenoble)

Stochastic Methods in Game Theory: Workshop on Learning  
18 November 2015

# Outline

- 1 The model
- 2 Unilateral
- 3 Dominated strategies
- 4 Stability and convergence
- 5 Empirical frequencies for 2 players

## Notation

A finite game in normal form is a tuple  $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{A}, u)$  consisting of

- A finite set of players  $\mathcal{N} = \{1, \dots, N\}$ ;
- A finite set  $\mathcal{A}_k$  of actions per player  $k \in \mathcal{N}$ ;
- The players' payoff functions  $u_k: \mathcal{A} \rightarrow \mathbb{R}$ , where  $\mathcal{A} \equiv \prod_k \mathcal{A}_k$  denotes the set of all joint action profiles  $(\alpha_1, \dots, \alpha_N)$ .
- The set of mixed actions of  $k$  will be  $\mathcal{X}_k \equiv \Delta(\mathcal{A}_k)$  and we denote  $\mathcal{X} \equiv \prod_k \mathcal{X}_k$ . Also, we define  $v_{k\alpha}(x) \equiv u_k(\alpha; x_{-k})$

## The model – Deterministic case

We consider the following general model

$$y_k(t) = \int_0^t v_k(x(s)) ds,$$
$$x_k(t) = Q_k(\eta_k(t)y_k(t)),$$

where :

- $y_k(t) = (y_{k\alpha}(t))_{\alpha \in \mathcal{A}_k}$  is a score vector for player  $k$ , for every  $\alpha \in \mathcal{A}_k$ , based on its cumulative payoff  $y_{k\alpha}(t) = \int_0^t v_{k\alpha}(x(s)) ds$  up to time  $t$ .
- $Q_k(y_k)$  is a regularized best response map the player uses to select a mixed strategy  $x_k(t) \in \mathcal{X}_k$  and receives the whole stream of payoff.
- $\eta_k(t) > 0$  is a learning parameter which can be tuned by the player.

$$Q_k(y_k) = \operatorname{argmax}_{x_k \in \mathcal{X}_k} \{\langle y_k, x_k \rangle - h_k(x_k)\},$$

where the *penalty function*  $h_k: \mathcal{X}_k \rightarrow \mathbb{R}$  satisfies the following properties :

- a)  $h_k$  is continuous on  $\mathcal{X}_k$ .
- b)  $h_k$  is smooth on the relative interior of every face of  $\mathcal{X}_k$ .
- c)  $h_k$  is strongly convex on  $\mathcal{X}_k$  : there exists some  $K > 0$  such that

$$h_k(tx_k + (1-t)x'_k) \leq th_k(x_k) + (1-t)h_k(x'_k) - \frac{1}{2}Kt(1-t)x'_k - x_k^2,$$

for all  $x_k, x'_k \in \mathcal{X}_k$  and for all  $t \in [0, 1]$ .

### Assumption

$\eta_k(t)$  is  $C^1$ -smooth, nonincreasing and  $\lim_{t \rightarrow \infty} t\eta_k(t) = +\infty$ .

**Remark** We will need the parameter  $\eta$  for some of the results.

## Examples

- $h(x) = \sum_{\alpha} x_{\alpha} \log x_{\alpha}$ .

The induced regularized best response is then given by the so-called logit map :

$$G_{\alpha}(y) = \frac{\exp(y_{\alpha})}{\sum_{\beta} \exp(y_{\beta})}.$$

For constant  $\eta = 1$ , the logit map leads to the continuous-time exponential weight algorithm [Sorin 09].

Differentiation yields

$$\dot{x}_{k\alpha} = \frac{e^{y_{k\alpha}} \dot{y}_{k\alpha}}{\sum_{\beta \in \mathcal{A}_k} e^{y_{k\beta}}} - \frac{e^{y_{k\alpha}} \sum_{\beta \in \mathcal{A}_k} e^{y_{k\beta}} \dot{y}_{k\beta}}{\left(\sum_{\beta \in \mathcal{A}_k} e^{y_{k\beta}}\right)^2} = x_{k\alpha} \left[ v_{k\alpha}(x) - \sum_{\beta \in \mathcal{A}_k} x_{k\beta} v_{k\beta}(x) \right],$$

which is the (multi-population) replicator equation [Taylor and Jonker '78] for population evolution under natural selection.

## Examples

- $h(x) = \frac{1}{2} \sum_{\alpha} x_{\alpha}^2$ .

This penalty function leads to the projected best response map

$$\Pi(y) = \operatorname{argmin}_{x \in \Delta} \left\{ \langle y, x \rangle - \frac{1}{2} \|x\|^2 \right\} = \operatorname{argmin}_{x \in \Delta} \|y - x\|^2,$$

The induced trajectories  $x(t) = \Pi(y(t))$  satisfy the so-called projection dynamics

$$\dot{x}_{k\alpha} = \begin{cases} v_{k\alpha}(x) - |\operatorname{supp}(x_k)|^{-1} \sum_{\beta \in \operatorname{supp}(x_k)} v_{k\beta}(x) & \text{if } x_{k\alpha} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

This dynamic was introduced in game theory by [Friedman '91] as a geometric model of the evolution of play in population games;

## Stochastic Perturbations

- Measurements are constantly subject to stochastic fluctuations which introduce noise to the input of any learning algorithm.
- We consider the stochastically perturbed process

$$\begin{aligned} dY_{k\alpha} &= v_{k\alpha}(X) dt + \sigma_{k\alpha}(X) dW_{k\alpha}, \\ X_k &= Q_k(\eta_k Y_k), \end{aligned}$$

where

- $W_{k\alpha}$  is a family of Wiener processes
- The diffusion coefficients  $\sigma_{k\alpha} : \mathcal{X} \rightarrow \mathbb{R}$  (assumed Lipschitz) measure the strength of the players' payoff observation noise

For simplicity, we only present here the special case where each player's penalty function is of the decomposable :

$$h_k(x_k) = \sum_{\alpha \in \mathcal{A}_k} \theta_k(x_{k\alpha})$$

**Remark :** Correlations can be considered (complicated to write) but every result we present here continues to hold. Also non-decomposable penalty function can be analyzed.



## Stochastic perturbations

## Proposition

Let  $X(t)$  be an orbit of the process in  $\mathcal{X}$  and let  $I$  be an open interval over which the support of  $X(t)$  remains constant. Then, the evolution of  $X(t)$  over  $I$  is governed by the SDE

$$\begin{aligned} dX_{k\alpha} = & \frac{\eta_k}{\theta''_{k\alpha}} \left[ v_{k\alpha} - \Theta''_k \sum_{\beta} v_{k\beta} / \theta''_{k\beta} \right] dt \\ & + \frac{\eta_k}{\theta''_{k\alpha}} \left[ \sigma_{k\alpha} dW_{k\alpha} - \Theta''_k \sum_{\beta} \sigma_{k\beta} / \theta''_{k\beta} dW_{k\beta} \right] \\ & + \frac{\dot{\eta}_k}{\eta_k} \frac{1}{\theta''_{k\alpha}} \left[ \theta'_{k\alpha} - \Theta''_k \sum_{\beta} \theta'_{k\beta} / \theta''_{k\beta} \right] dt \\ & - \frac{1}{2} \frac{1}{\theta''_{k\alpha}} \left[ \theta'''_{k\alpha} U_{k\alpha}^2 - \Theta''_k \sum_{\beta} \theta'''_{k\beta} / \theta''_{k\beta} U_{k\beta}^2 \right] dt, \end{aligned}$$

where all summations are taken over  $\beta \in \text{supp}(X_k)$  and :

- $\theta'_{k\alpha} = \theta'_k(X_{k\alpha})$ ,  $\theta''_{k\alpha} = \theta''_k(X_{k\alpha})$ ,  $\theta'''_{k\alpha} = \theta'''_k(X_{k\alpha})$ ,
- $\Theta''_k = \left( \sum_{\beta} 1 / \theta''_{k\beta} \right)^{-1}$ ,
- $U_{k\alpha}^2 = \left( \frac{\eta_k}{\theta''_{k\alpha}} \right)^2 \left[ \sigma_{k\alpha}^2 (1 - \Theta''_k / \theta''_{k\alpha})^2 + \sum_{\beta \neq \alpha} (\Theta''_k / \theta''_{k\beta})^2 \sigma_{k\beta}^2 \right]$ .

## Examples

- $h(x) = \sum_{\alpha} x_{\alpha} \log x_{\alpha}$  ( $\theta(x) = x \log x$ )

$$\begin{aligned}
 dX_{k\alpha} = & \eta_k X_{k\alpha} \left[ v_{k\alpha} - \sum_{\beta \in \mathcal{A}_k} X_{k\beta} v_{k\beta} \right] dt \\
 & + \eta_k X_{k\alpha} \left[ \sigma_{k\alpha} dW_{k\alpha} - \sum_{\beta \in \mathcal{A}_k} \sigma_{k\beta} X_{k\beta} dW_{k\beta} \right] \\
 & + \frac{\dot{\eta}_k}{\eta_k} X_{k\alpha} \left[ \log X_{k\alpha} - \sum_{\beta \in \mathcal{A}_k} X_{k\beta} \log X_{k\beta} \right] dt \\
 & + \frac{1}{2} X_{k\alpha} \left[ \sigma_{k\alpha}^2 (1 - 2X_{k\alpha}) - \sum_{\beta \in \mathcal{A}_k} \sigma_{k\beta}^2 X_{k\beta} (1 - 2X_{k\beta}) \right] dt.
 \end{aligned}$$

**Remark :** For constant  $\eta$ , this is the version of the stochastic replicator dynamics of exponential learning studied by [Mertikopoulos and Moustakas '10].

## Examples

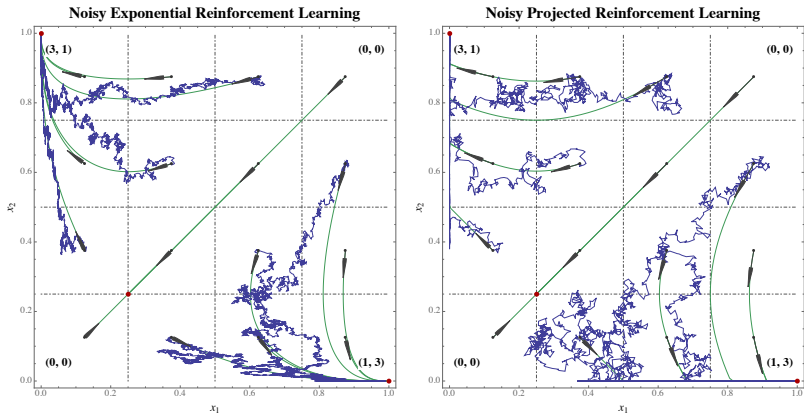
- $h(x) = \sum_{\alpha} x_{\alpha}^2$  ( $\theta(x) = x^2$ )

leads to the stochastic projection dynamics :

$$\begin{aligned}
 dX_{k\alpha} = & \left[ v_{k\alpha} - |\text{supp}(X_k)|^{-1} \sum_{\beta \in \text{supp}(X_k)} v_{k\beta} \right] dt \\
 & + \left[ \sigma_{k\alpha} dW_{k\alpha} - |\text{supp}(X_k)|^{-1} \sum_{\beta \in \text{supp}(X_k)} \sigma_{k\beta} dW_{k\beta} \right] \\
 & + \frac{\dot{\eta}_k}{\eta_k} [X_{k\alpha} - |\text{supp}(X_k)|^{-1}] dt.
 \end{aligned}$$

**Remark :** Valid only on intervals over which the support of  $X(t)$  remains constant...

## Examples



**Figure :** Evolution of play logit (left) and projected best responses (right) for a  $2 \times 2$  game ( $\sigma_{k\alpha} \equiv 1$ )

# Outline

- 1 The model
- 2 Unilateral**
- 3 Dominated strategies
- 4 Stability and convergence
- 5 Empirical frequencies for 2 players

- We take the point of view of player 1
- The stream of payoff  $v(t) \in \mathbb{R}^{|\mathcal{A}_1|}$  is (for instance) supposed to be measurable and bounded.
- We focus on the unilateral process :

$$\begin{aligned}dY_\alpha(t) &= v_\alpha(t) dt + \sigma_\alpha(t) dW_\alpha(t), \\X(t) &= Q(\eta(t) Y(t)),\end{aligned}$$

We consider the following notion of external regret :

Regret

$$\text{Reg}(t) = \max_{\alpha \in \mathcal{A}} \int_0^t v_\alpha(s) ds - \int_0^t \langle v(s), x(s) \rangle ds,$$

- We take the point of view of player 1
- The stream of payoff  $v(t) \in \mathbb{R}^{|\mathcal{A}_1|}$  is (for instance) supposed to be measurable and bounded.
- We focus on the unilateral process :

$$\begin{aligned}dY_\alpha(t) &= v_\alpha(t) dt + \sigma_\alpha(t) dW_\alpha(t), \\X(t) &= Q(\eta(t) Y(t)),\end{aligned}$$

We consider the following notion of external regret :

## Regret

$$\text{Reg}(t) = \max_{\alpha \in \mathcal{A}} \int_0^t v_\alpha(s) ds - \int_0^t \langle v(s), x(s) \rangle ds,$$

## Theorem

Assume  $\lim_{t \rightarrow \infty} \eta(t) = 0$  and that  $Y(0) = 0$ . Then, almost surely,

$$\text{Reg}(t) \leq \frac{\Omega}{\eta(t)} + \sigma_{\max}^2 \frac{|\mathcal{A}|}{2K} \int_0^t \eta(s) ds + \mathcal{O}(\sigma_{\max} \sqrt{t \log \log t}),$$

where  $\Omega = \max\{h(x) - h(x') : x, x' \in \mathcal{X}\}$ ,  $\sigma_{\max} = \sup_t \max_{\beta} \sigma_{\beta}(t)$  and  $K$  is the strong convexity constant of the player's penalty function  $h$ .

### Remark :

- The role of  $\eta$  is important here.
- When  $\sigma = 0$ , known result [Sorin 09, Hofbauer et al 09].
- Unfortunately, we do not have an example when  $\eta$  is constant and the bound given above does not hold.



## Idea of the proof

- Proof obtained by analyzing

$$\frac{1}{\eta} F(p, \eta Y) = \frac{1}{\eta} \cdot [h(p) + h^*(\eta Y) - \langle \eta Y, p \rangle],$$

where  $h^*(y) = \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\}$  is the Fenchel conjugate of  $h$ .

- No need to use the evolution equation for  $X$ .

By choosing  $\eta(t) = t^{-\gamma}$  for some  $\gamma \in (0, 1)$ , we have :

### Corollary

*Assume that  $\eta(t) \sim t^{-\gamma}$  for some  $\gamma \in (0, 1)$ . Then :*

$$\text{Reg}(t) = \begin{cases} (t^{1-\gamma}) & \text{if } 0 < \gamma < \frac{1}{2}, \\ (\sqrt{t \log \log t}) & \text{if } \gamma = \frac{1}{2}, \\ (t^\gamma) & \text{if } \frac{1}{2} < \gamma < 1. \end{cases}$$

## Outline

- 1 The model
- 2 Unilateral
- 3 Dominated strategies**
- 4 Stability and convergence
- 5 Empirical frequencies for 2 players

## Game framework

Given a finite game  $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{A}, u)$ , we say that  $p_k \in \mathcal{X}_k$  is dominated by  $p'_k \in \mathcal{X}_k$  (and we write  $p_k \prec p'_k$ ) if

$$\langle v_k(x), p_k \rangle < \langle v_k(x), p'_k \rangle \quad \text{for all } x \in \mathcal{X}.$$

- A pure strategy  $\alpha \in \mathcal{A}_k$  becomes extinct along  $x(t)$  if  $x_{k\alpha}(t) \rightarrow 0$  as  $t \rightarrow \infty$ .
- More generally, we will say that the mixed strategy  $p_k \in \mathcal{X}_k$  becomes extinct along  $x(t)$  if  $\min\{x_{k\alpha}(t) : \alpha \in \text{supp}(p_k)\} \rightarrow 0$ ;

## Game framework

## Theorem

*If  $p_k \in \mathcal{X}_k$  is dominated (even iteratively), then it becomes extinct along  $X(t)$  almost surely.*

- we have to show that

$$F_k(p_k, \eta_k Y_k) \geq h_k(p_k) - h_k(p'_k) + \eta_k \cdot [c_k + m_k t + \xi_k(t)],$$

- and use the fact that if  $F(p, y_n) \rightarrow +\infty$  for some sequence  $y_n$ , the sequence  $x_n = Q(y_n)$  converges to the union of faces of  $\Delta$  that do not contain  $p$ . Therefore :  $\liminf_{n \rightarrow \infty} \{x_{n,\alpha} : \alpha \in \text{supp}(p)\} = 0$ .

## Bounds

## Proposition

Let  $\alpha \in \mathcal{A}_k$  be dominated by  $\beta \in \mathcal{A}_k$  and assume that  $\lim_{x \rightarrow 0^+} \theta'_k(x) = -\infty$ . Then, for all  $\delta, \varepsilon > 0$  and for all large enough  $t$ , we have :

$$X_{k\alpha}(t) \leq \phi_k \left[ C_k - \eta_k(t) \left( m_k t - 2(1 + \varepsilon) \sigma_{\alpha\beta} \sqrt{t \log \log t} \right) \right] \quad (\text{a.s.})$$

and

$$\mathbf{P}(X_{k\alpha}(t) > \delta) \leq \frac{1}{2} \operatorname{erfc} \left[ \frac{1}{2\sigma_{\alpha\beta}} \left( m_k \sqrt{t} - \frac{C_k - \theta'_k(\delta)}{\eta_k(t) \sqrt{t}} \right) \right]$$

where :

- $\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt$  is the complementary error function.
- $\phi_k = (\theta'_k)^{-1}$  (note that  $\lim_{z \rightarrow -\infty} \phi_k(z) = 0$  by assumption).
- $m_k = \min_{x \in \mathcal{X}} \{v_{k\beta}(x) - v_{k\alpha}(x)\} > 0$  is the minimum payoff difference between  $\alpha$  and  $\beta$ .
- $\sigma_{\alpha\beta}^2 = \frac{1}{2} \max_{x \in \mathcal{X}} \{\sigma_{k\alpha}^2(x) + \sigma_{k\beta}^2(x)\} > 0$ .
- $C_k$  is a constant that depends only on the initial conditions.

### Proposition

Assume  $\eta_k$  is constant. If  $\tau_\delta = \inf\{t > 0 : X_{k\alpha}(t) \leq \delta\}$ , then :

$$\mathbb{E}[\tau_\delta] \leq \frac{[C_k - \theta'_k(\delta)]_+}{\eta_k m_k}.$$

## Comments

Different type (by perturbing fitness in populations) of deduction gives :

$$\begin{aligned}dX_{k\alpha} &= X_{k\alpha} \left[ v_{k\alpha} - \sum_{\beta} X_{k\beta} v_{k\beta} \right] dt \\ &+ X_{k\alpha} \left[ \sigma_{k\alpha} dW_{k\alpha} - \sum_{\beta} \sigma_{k\beta} X_{k\beta} dW_{k\beta} \right] \\ &- X_{k\alpha} \left[ \sigma_{k\alpha}^2 X_{k\alpha} - \sum_{\beta} \sigma_{k\beta}^2 X_{k\beta}^2 \right] dt,\end{aligned}$$



## Comments

Different type (by perturbing fitness in populations) of deduction gives :

$$\begin{aligned} dX_{k\alpha} = & X_{k\alpha} \left[ v_{k\alpha} - \sum_{\beta} X_{k\beta} v_{k\beta} \right] dt \\ & + X_{k\alpha} \left[ \sigma_{k\alpha} dW_{k\alpha} - \sum_{\beta} \sigma_{k\beta} X_{k\beta} dW_{k\beta} \right] \\ & - X_{k\alpha} \left[ \sigma_{k\alpha}^2 X_{k\alpha} - \sum_{\beta} \sigma_{k\beta}^2 X_{k\beta}^2 \right] dt, \end{aligned}$$

$$\begin{aligned} dX_{k\alpha} = & X_{k\alpha} \left[ v_{k\alpha} - \sum_{\beta} X_{k\beta} v_{k\beta} \right] dt \\ & + X_{k\alpha} \left[ \sigma_{k\alpha} dW_{k\alpha} - \sum_{\beta} \sigma_{k\beta} X_{k\beta} dW_{k\beta} \right] \\ & + \frac{1}{2} X_{k\alpha} \left[ \sigma_{k\alpha}^2 (1 - 2X_{k\alpha}) - \sum_{\beta} \sigma_{k\beta}^2 X_{k\beta} (1 - 2X_{k\beta}) \right] dt. \end{aligned}$$

**Remark** Quite different properties [Imhof '05, Hofbauer and Imhof '09].

## Outline

- 1 The model
- 2 Unilateral
- 3 Dominated strategies
- 4 Stability and convergence**
- 5 Empirical frequencies for 2 players

## Stability

- As we discussed, in the case  $\sigma = 0$  with constant learning rates ( $\eta = 1$ ) our process leads to the Replicator Dynamics (RD).
- In that case, we have that
  - a) If a solution orbit converges to  $x^*$ , then  $x^*$  is Nash.
  - b) If  $x^* \in \mathcal{X}$  is (Lyapunov) stable, then it is also Nash.
  - c) Strict Nash equilibria are asymptotically stable.
- Our idea in this part is to explore how far can we go in this direction using this stochastic approach.

## Definition

Let  $x^* \in \mathcal{X}$ . We will say that :

- 1  $x^*$  is stochastically (Lyapunov) stable if, for every  $\varepsilon > 0$  and for every neighborhood  $U_0$  of  $x^*$  in  $\mathcal{X}$ , there exists a neighborhood  $U \subseteq U_0$  of  $x^*$  such that

$$\mathbb{P}(X(t) \in U_0 \text{ for all } t \geq 0) \geq 1 - \varepsilon,$$

whenever  $X(0) \in U$ .

- 2  $x^*$  is stochastically asymptotically stable if, for every  $\varepsilon > 0$  and for every neighborhood  $U_0$  of  $x^*$  in  $\mathcal{X}$ , there exists a neighborhood  $U \subseteq U_0$  of  $x^*$  such that

$$\mathbb{P}\left(X(t) \in U_0 \text{ for all } t \geq 0 \text{ and } \lim_{t \rightarrow \infty} X(t) = x^*\right) \geq 1 - \varepsilon,$$

whenever  $X(0) \in U$ .

## Theorem

Let  $x^* \in \mathcal{X}$ . Then :

- (1) If  $\mathbb{P}(\lim_{t \rightarrow \infty} X(t) = x^*) > 0$ ,  $x^*$  is a Nash equilibrium of  $\mathcal{G}$ .
- (2) If  $x^*$  is stochastically (Lyapunov) stable, it is also Nash.
- (3) If  $x^*$  is a strict Nash equilibrium of  $\mathcal{G}$ , it is stochastically asymptotically stable.

## Outline

- 1 The model
- 2 Unilateral
- 3 Dominated strategies
- 4 Stability and convergence
- 5 Empirical frequencies for 2 players**

## Time Averages

- We want to study the asymptotic behavior of the players' empirical distribution of play

$$\bar{X}(t) = \frac{1}{t} \int_0^t X(s) ds,$$

- Our analysis is motivated by the original deterministic results showing that  $\bar{X}(t)$  converges to Nash equilibrium under the replicator dynamics, in some cases.

## Averages

## Theorem

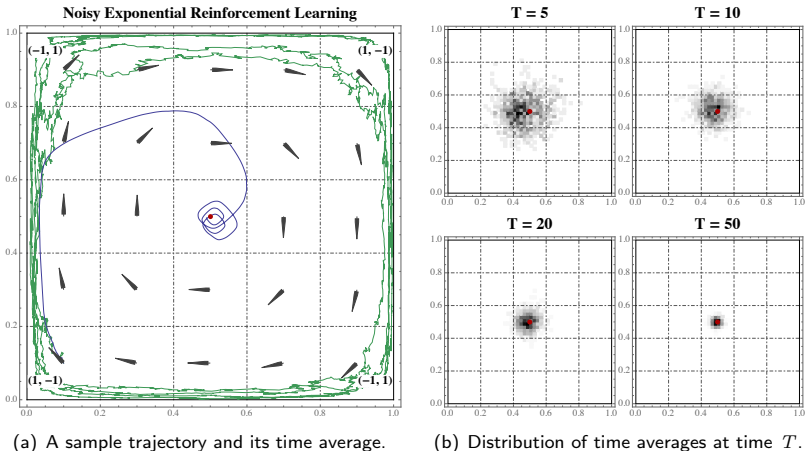
For a 2-player game  $\mathcal{G}$ , the  $\omega$ -limit set of the empirical distribution of play  $\bar{X}(t)$  is an ICT set for the deterministic best response dynamics [Gliboa and Matsui '91] :

$$\dot{x}_k \in BR_k(x) - x_k,$$

where  $BR_k(x) \equiv \operatorname{argmax}_{x'_k \in \mathcal{X}_k} \langle v_k(x), x'_k \rangle$ .

- If the empirical distribution of play converges, its limit is a Nash equilibrium.
- In zero-sum games, the empirical distribution of play converges almost surely to the set of Nash equilibria of  $\mathcal{G}$ .
- The only ICT sets of the best response dynamics in potential games consist of components of Nash equilibria. Then, the empirical measure converges, almost surely, to one of such components.





**Figure :** Time averages with logit best responses in a game of Matching Pennies (as in Fig. 1, Nash equilibria are depicted in red and the game's payoff's are displayed inline; for benchmarking purposes, we also took  $\sigma_{k\alpha} = 1$  for all  $\alpha \in \mathcal{A}_k$ ,  $k = 1, 2$ ). Fig. 2(a) shows the evolution of a sample trajectory and its time average; in Fig. 2(b), we show a density plot of the distribution of  $10^4$  time-averaged trajectories for different values of the integration horizon  $T$ .

Thanks !