Int J Game Theory (2003) 32:3–16
DOI 10.1007/s001820300135

International Journal of Game Theory

A two armed bandit type problem*

Michel Benaïm[1], **Gerard Ben Arous**[2]

[1] Department of Mathematics, Université Cergy-Pontoise, 2 Avenue A. Chauvin, 95302, Cergy-Pontoise, FRANCE (Email: benaim@math.u-cergy.fr)
[2] Courant Institute, 251 Mercer Street, New York, NY 10012, USA
(Email: benarous@cims.nyu.edu)

Received: February 2000/Revised: March 2003

Let f_0 and f_1 be two real valued continuous functions defined on $[0,1]$. Given $X = (X_1, X_2, \ldots) \in \{0,1\}^{\mathbf{N}_*}$ define

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}, \quad n \geq 1$$

and

$$Q(X) = \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f_{X_{k+1}}(\bar{X}_k).$$

Suppose that a "player" has to choose an infinite sequence X of zero and one and is rewarded at time $n+1$ by $f_{X_{n+1}}(\bar{X}_n)$. Then $Q(X)$ is the *cumulative payoff* corresponding to *strategy* X.

The purpose of this note is to address the following question:

What are the good strategies?

The answer is given by the next theorem.

Theorem 1. *Let*

$$Q^* = \max_{0 \leq x \leq 1} x f_1(x) + (1-x) f_0(x)$$

and let $0 \leq x^* \leq 1$ *be such that* $Q^* = x^* f_1(x^*) + (1-x^*) f_0(x^*)$.

* We are very grateful to Josef Hofbauer – who showed us how to improve significantly our proof of the main result- and Sylvain Sorin, as well as to three anonymous referees and an associate editor for their useful comments.

(a) *Suppose X is a sequence such that* $\lim_{n\to\infty} \overline{X}_n = x^*$. *Then,* $Q(X) = Q^*$.
(b) *For all* $X \in \{0,1\}^{\mathbf{N}_*}$, $Q(X) \leq Q^*$.

In words, our player cannot do better in the long run than playing a sequence of IID random Bernoulli variables with parameter x^*.

Apart from its intrinsic interest[1] this result is motivated by certain questions arising in the *theory of learning in games* and *stochastic fictitious play* (for a recent overview see (Fudenberg and Levine, 1998)). In particular, it implies that in a 2×2 repeated game, continuous "fictitious play type strategies" are closed under best-response; that is, that an equilibrium using such strategies remains an equilibrium if arbitrary strategies are allowed.

Section 1 will make precise this last sentence. Section 2 is devoted to the proof of a theorem generalizing Theorem 1 to an arbitrary number of choices, provided the associated functions $f_0, f_1, f_2, \ldots$ satisfy a certain condition (which is always satisfied if the player has only two actions at hand). Section 3 provides a class of counterexamples showing that this condition is necessary.

1. Fictitious-play type strategies

This section is based on some (published and unpublished) work by Benaïm and Hirsch.

Let Γ be a 2×2 game in strategic form. Players are labeled $i = 1, 2$ and the opponent to player i is conventionally denoted $-i$ (that is $-i = 3 - i$). The set of *pure strategies* of player i is $A^i = \{0,1\}$ and the set of *mixed strategies* is $\Delta^i = [0,1]$ where $x^i \in [0,1]$ is the mixed strategy which assigns weight x^i to 1 and weight $(1 - x^i)$ to 0. The set of *action profiles* is $A = A^1 \times A^2$ and the set of *strategy profiles* is $\Delta = \Delta^1 \times \Delta^2$. The *payoff function* to player i is denoted $u^i : A \to \mathbf{R}$. It extends to a function $\Delta \to \mathbf{R}$ (still denoted u^i) in the usual way.

Consider the infinitely-repeated game (with complete information) where Γ is played over and over at times $k = 1, 2 \ldots$ At the beginning of stage k player i chooses an action $a_k^i \in A^i$ independently of the other player. As a result of these choices she gets the payoff $u^i(a_k)$ where $a_k = (a_k^1, a_k^2)$ denotes the *action profile* at time k.

At the end of each stage both players are informed of the actions chosen in that stage. For $n \geq 1$, let $A^{(n)} = A \times \cdots \times A$ be the Cartesian product of n copies of A.

A *behavior strategy* for player i is a sequence $h^i = \{h_n^i\}_{n \in \mathbf{N}}$ defined by

(a) $h_0^i \in [0,1]$ is the probability with which player i chooses action 1 at the beginning of the first game.
(b) For all $n \geq 1$, h_n^i is a function $h_n^i : A^{(n)} \to [0,1]$ where $h_n^i(a_1, \ldots, a_n)$ is the probability that player i plays action 1 at time $n+1$ when the actions $a_1, \ldots, a_n$ have been played at times $k = 1, \ldots, n$.

Let $h = (h^1, h^2)$ be a pair of strategies. The *(lower) payoff* to player i corresponding to h is the random variable

[1] We find this problem quite amusing and we spent some good time trying to solve it and challenging several of our colleagues.

$$\underline{P}^i(h) = \liminf_{n\to\infty} \frac{1}{n}(u^i(a_1) + \cdots + u^i(a_n))$$

where $a_1, \ldots, a_n$ is the process generated by the pair of strategies $h = (h^1, h^2)$. The upper payoff $\bar{P}^i(h)$ is defined similarly with lim sup instead of lim inf.

As usual, a *Nash equilibrium* is a pair (h^1, h^2) of strategies such that for each $i = 1, 2$:

$$\underline{P}^i(h) \geq \underline{P}^i(\tilde{h})$$

for all $\tilde{h}$ with $\tilde{h}^{-i} = h^{-i}$.

When h is an equilibrium, the random variable $(\underline{P}^1(h), \underline{P}^2(h))$ is called a *(lower) equilibrium payoff.*

Let Ψ^i be some function from $[0, 1]$ into itself. A *Fictitious-Play Type Strategy* (FTS) for player *i induced by* Ψ^i is a strategy h^i such that

$$h^i_n(a_1, \ldots, a_n) = \Psi^i(x_n^{-i})$$

for all $n \geq 1$, where

$$x_n^i = \frac{1}{n}(a_1^i + \cdots + a_n^i), \quad i = 1, 2.$$

If Ψ^i is continuous (Lipschitz, smooth) we call h^i a continuous (Lipschitz, smooth) FTS.

Given a function $\Psi : [0, 1] \to [0, 1]$ we denote

$$Gr(\Psi) = \{(x, \Psi(x)) : x \in [0, 1]\}$$

and

$$Gr'(\Psi) = \{(\Psi(x), x); x \in [0, 1]\}.$$

The following theorem describes the long term behavior of a repeated game where *both* players play a Lipschitz FTS.

Theorem 2 (Benaïm and Hirsch, 1999). *Assume that players play Lipschitz FTS h^1, h^2 induced by the functions Ψ^1 and Ψ^2. Let $\{x_n\} = \{(x_n^1, x_n^2)\}$ denote the sequence of empirical strategies, and $L(\{x_n\})$ the* limit set *of $\{x_n\}$. Then, with probability one:*

(i) *$L(\{x_n\})$ is a compact connected subset of $Gr'(\Psi^1) \cap Gr(\Psi^2)$.*

(ii)

$$\inf\{u^i(y) : y \in L(\{x_n\})\} \leq \underline{P}^i(h) \leq \bar{P}^i(h) \leq \max\{u^i(y) : y \in L(\{x_n\})\}$$

In particular, if $Gr'(\Psi^1) \cap Gr(\Psi^2)$ is finite, then $\{x_n\}$ converges almost surely to a point $x_* \in Gr'(\Psi^1) \cap Gr(\Psi^2)$ and $\underline{P}^i(h) = \bar{P}^i(h) = u^i(x_*)$.

Proof: Follows from Theorems 2.2 and 4.1 of Benaïm and Hirsch (1999). **QED**

When Ψ^1 and Ψ^2 are certain smooth approximations to the best response functions, part (i) of Theorem 2 has been proved by (Fudenberg and Kreps, 1993) and (Kaniovski and Young, 1995) extending convergence results of the standard fictitious play proved by Robinson (1951) and Miyasawa (1961).

The next theorem is the key result of this section. It asserts that if a player plays a continuous FTS then it is rational for her opponent to play a continuous FTS.

Theorem 3. *Suppose player 2 plays a continuous FTS h^2 induced by Ψ^2. Let*

$$v^1 = \max_{x^1 \in \Delta^1} u^1(x^1, \Psi^2(x^1)).$$

Then

(i) *Player 1 cannot get more than v^1. That is, for any strategy h^1*

$$\underline{P}^1(h) \leq v^1.$$

(ii) *Suppose furthermore that Ψ^2 is Lipschitz. Let x^1_* be a point realizing the equality $v^1 = u^1(x^1_*, \Psi^2(x^1_*))$, $x^2_* = \Psi^2(x^1_*)$ and h^1 a Lipschitz FTS induced by any function Ψ^1 such that $Gr'(\Psi^1) \cap Gr(\Psi^2) = \{(x^1_*, x^2_*)\}$ (for example the function $\Psi^1 : x^2 \to x^1_*$). Then*

$$\underline{P}^1(h) = \bar{P}^1(h) = v^1.$$

Proof: For notational convenience set $X_k = a^1_k$ (the action played by 1 at time k). Let $\mathscr{F}_n$ be the sigma field generated by the variables $X_1, \ldots, X_n$. Let

$$f_1(x) = u^1(1,1)\Psi^2(x) + u^1(1,0)(1 - \Psi^2(x))$$

and

$$f_0(x) = u^1(0,1)\Psi^2(x) + u^1(0,0)(1 - \Psi^2(x)).$$

Then

$$E(u^1(a_{n+1}) \,|\, \mathscr{F}_n) = E(X_{n+1}|\mathscr{F}_n) f_1(\bar{X}_n) + (1 - E(X_{n+1}|\mathscr{F}_n)) f_0(\bar{X}_n).$$

Hence

$$E(u^1(a_{n+1}) \,|\, \mathscr{F}_n) = E(f_{X_{n+1}}(\bar{X}_n) \,|\, \mathscr{F}_n). \quad (1)$$

The law of large number for L^2 martingales implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (u^1(a_{k+1}) - E(u^1(a_{k+1}) \,|\, \mathscr{F}_k)) = 0$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(f_{X_{k+1}}(\bar{X}_k)-E(f_{X_{k+1}}(\bar{X}_k)\mid\mathscr{F}_k))=0$$

almost surely. It then follows from (1) that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(u^1(a_{k+1})-f_{X_{k+1}}(\bar{X}_k))=0 \tag{2}$$

almost surely. This last equality shows that $\underline{P}^1(h)=Q(X)$ where $\underline{P}^1(h)$ is the (lower) payoff obtained by player 1 and $Q(X)$ is as in Theorem 1. Then, by Theorem 1

$$\underline{P}^1(h)\le\max_{0\le x\le 1}xf_1(x)+(1-x)f_0(x)=\max_{0\le x\le 1}u^1(x,\Psi^2(x)).$$

This proves part (i) of Theorem 3. Part (ii) of Theorem 3 follows from Theorem 2. **QED**

A vector $v=(v^1,v^2)\in\mathbf{R}^2$ is said *feasible* if there exists a mixed strategy profile $x=(x^1,x^2)\in\Delta$ such that $v^i=u^i(x)$. If $v=(v^1,v^2)$ is a feasible vector we denote

$$D^i(v^i)=\{x\in\Delta:u^i(x)\le v^i\}.$$

For $i=1,2$, player's *i reservation utility* is

$$\underline{v}^i=\min_{x^{-i}\in\Delta^{-i}}\left(\max_{x^i\in\Delta^i}u^i(x)\right).$$

By combining Theorems 2 and 3 we obtain the following version of the *folk theorem*

Corollary 1. *Let $v=(v^1,v^2)$ be a feasible payoff vector with $v^i>\underline{v}^i$. Then:*

(i) *there exist Lipschitz functions $\Psi^i:\Delta^{-i}\to\Delta^i$ such that*
  (a) *$Gr'(\Psi^1)\subset D^2(v^2)$,*
  (b) *$Gr(\Psi^2)\subset D^1(v^1)$,*
  (c) *For all $x\in Gr'(\Psi^1)\cap Gr(\Psi^2)$ and $i\in\{1,2\}$ $u^i(x)=v^i$.*

(ii) *For any such pair of functions the corresponding fictitious play type strategies $h=(h^1,h^2)$ constitute a Nash equilibrium whose equilibrium payoff is v. Furthermore, $\underline{P}^i(h)=\bar{P}^i(h)=v^i$.*

Proof: Choose $x_*\in\Delta$ such that $u^i(x_*)=v^i$, $i=1,2$. Let $x^1\in[0,1]$. If $u^1(x^1,x_*^2)\le v^1$ set $\Psi^2(x^1)=x_*^2$. If $u^1(x^1,x_*^2)>v^1$ set $\Psi^2(x^1)=y$ where y is the solution to $u^1(x^1,y)=v^1$. Such a solution always exists (for otherwise we would have $\underline{v}^1=\min_y\max_{x^1}u^1(x^1,y)>v^1$) and can be written as $y=\dfrac{v^1-cx-d}{ax+b}$ where a,b,c,d are obtained by rewriting u^1 in the form

$$u^1(x^1, x^2) = x^2(ax^1 + b) + cx^1 + d.$$

By exchanging the labels of the players we define the function Ψ^1. Functions (Ψ^1, Ψ^2) satisfy assertions (a), (b), (c) and are piecewise C^1, hence Lipschitz continuous.

Let h^1, h^2 be the pair of FTS induced by Ψ^1 and Ψ^2. Theorem 1 implies that $P^i(h) = v^i$ almost surely and part (i) of Theorem 3 implies that (h^1, h^2) is a Nash equilibrium. **QED**

2. Proof of Theorem 1

In this section we state and prove a theorem generalizing Theorem 1.

Let $S = \{e^0, \dots, e^d\}$ be a finite set of cardinal $d + 1$ and let

$$\mathscr{P}(S) = \left\{ x = (x_0, \dots, x_d) \in \mathbf{R}^{d+1} : x_i \geq 0, \sum_i x_i = 1 \right\}$$

denote the d-dimensional unit simplex. The extremal points of $\mathscr{P}(S)$ are identified with the elements of S, so that

$$e^i_j = \delta_{i,j}$$

for $i, j \in \{0, \dots, d\}$.

Let $f_0, \dots, f_d$ be $d + 1$ real valued continuous functions defined on $\mathscr{P}(S)$. Given $X = (X_1, X_2, \dots) \in S^{\mathbf{N}^*}$ define

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \in \mathscr{P}(S), \quad n \geq 1$$

and

$$Q(X) = \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f_{X_{k+1}}(\bar{X}_k) \in \mathbf{R},$$

where here (and throughout) f_{e^i} is identified with f_i. Let

$$U = \left\{ z \in \mathbf{R}^d : z_i > 0, \sum_{i=1}^{d} z_i < 1 \right\}$$

and

$$g : \bar{U} \to \mathbf{R}^d,$$

$$z \to g(z) = (g_1(z), \dots, g_d(z))$$

be the function defined by

$$g_i(z_1,\dots,z_d) = f_0\left(1-\sum_{i=1}^{d} z_i, z_1,\dots,z_d\right) - f_i\left(1-\sum_{i=1}^{d} z_i, z_1,\dots,z_d\right).$$

We say that $(f_0,\dots,f_d)$ has a *gradient structure* if there exists a smooth real valued function V defined on some neighborhood of $\bar{U}$ such that

$$g(z) = \nabla V(z)$$

for all $z \in \bar{U}$.

Remark that this condition is always satisfied when $d = 1$. When $d \geq 1$ and g is C^2 this is equivalent to the condition that $\left(\frac{\partial g_i}{\partial y_j}\right)_{i,j}$ is symmetric.

Theorem 4. *Let*

$$Q^* = \max_{x\in\mathscr{P}(S)}\left(\sum_{i=0}^{d} x_i f_i(x)\right)$$

and let $x^* \in \mathscr{P}(S)$ *be such that* $Q^* = \sum_{i=0}^{d} x_i^* f_i(x_i^*)$.

(a) *Suppose* X *is a sequence such that* $\lim_{n\to\infty}\bar{X}_n = \bar{x} \in \mathscr{P}(S)$ *Then* $Q(X) = \sum_{i=0}^{d}\bar{x}_i f_i(\bar{x})$. *In particular,* $Q(X) = Q^*$ *if* $\lim_{n\to\infty}\bar{X}_n = x^*$.

(b) *Suppose* $(f_0,\dots,f_d)$ *has a gradient structure. Then* $Q(X) \leq Q^*$ *for all* $X \in S^{\mathbf{N}^*}$.

Proof of part (a)

Let μ_n be the probability measure defined on $S\times\mathscr{P}(S)$ by

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n}\delta_{X_{k+1},\bar{X}_k}.$$

We claim that $\{\mu_n\}$ converges weakly (i.e. for the weak* topology) toward the product measure of the measure $\sum_i \bar{x}_i\delta_{e^i}$ and the Dirac measure $\delta_{\bar{x}}$.

By compactness of $S\times\mathscr{P}(S)$ the sequence $\{\mu_n\}$ is relatively compact for the topology of weak* convergence. Let $\mu = \lim_{n_i\to\infty}\mu_{n_i}$ be a limit point of $\{\mu_n\}$. For any Borel set $A\subset\mathscr{P}(S)$ and $\theta\in S$ set $\mu_\theta(A) = \mu(\{\theta\}\times A)$. Then for any continuous function $g : S\times\mathscr{P}(S)\to\mathbf{R}$

$$\lim_{n_i\to\infty}\frac{1}{n_i}\sum_{k=1}^{n_i} g(X_{k+1},\bar{X}_k) = \sum_{\theta\in S}\int_{\mathscr{P}(S)} g(\theta,x)\mu_\theta(dx). \tag{3}$$

Applying equality (3) with the function $g(\theta,x) = h(x)$ gives

$$h(\bar{x}) = \int_{\mathscr{P}(S)} h(x)\sum_{\theta\in S}\mu_\theta(dx).$$

Since this is true for any continuous function $h : \mathscr{P}(S) \to \mathbf{R}$ we have $\sum_{\theta \in S} \mu_\theta = \delta_{\bar{x}}$. Now, applying (3) with $g(\theta, x) = \theta$ gives $\bar{x}_i = \int_{\mathscr{P}(S)} \mu_i(dx)$. Thus $\mu_i = \bar{x}_i \delta_{\bar{x}}$. This proves the claim.

The proof of part (a) is now obvious, since by weak* convergence we get

$$Q(X) = \lim_{n \to \infty} \int_{S \times \mathscr{P}(S)} f_\theta(x) \mu_n(d\theta \times dx) = \sum_{\theta \in S} x^*_\theta f_\theta(x^*) = Q^*$$

Proof of part (b)

First, observe that it suffices to prove the result for smooth functions. Indeed, if $f_0, \ldots, f_d$ are only continuous, let $\psi(z) = (2\pi)^{-d/2} \exp(-\|z\|^2/2)$ and $V_\epsilon(z) = \int_{\mathbf{R}^d} V(z + \epsilon u) \psi(u)\, du$. The function V_ϵ, hence $g_\epsilon = \nabla V_\epsilon$, is C^∞. Furthermore, for any $\delta > 0$ there exists $\epsilon > 0$ such that

$$\|g_\epsilon - g\|_\infty = \max_{z \in \bar{U}} \|g(z) - g_\epsilon(z)\| \le \delta.$$

Set $\tilde{f}_0 = f_0$ and $\tilde{f}_i = f_0 - g_{\epsilon, i}$ for $i = 1, \ldots, d$. Then

$$Q(X) \le \tilde{Q}(X) + \delta$$

where $\tilde{Q}(X)$ is defined like $Q(X)$ with $\tilde{f}_i$ instead of f_i.

Therefore (assuming that the result holds for smooth functions) we obtain

$$Q(X) \le \tilde{Q}^* + \epsilon \le Q^* + 2\delta.$$

Since δ is arbitrary this proves that $Q(X) \le Q^*$.

From now on, we then assume that $f_0, \ldots, f_d$ are C^1.

For $\theta \in S$, let F_θ be the vector field on $\mathscr{P}(S) \times \mathbf{R}$ defined by

$$F_\theta(x, y) = (-x + \theta, -y + f_\theta(x)).$$

Let $\mathscr{H}$ denote the set of functions $\Theta : \mathbf{R}_+ \to S$ which are right continuous and have at most finitely many discontinuities on each finite interval of time.

For each $\Theta \in \mathscr{H}$ the ordinary differential equation

$$(\dot{x}, \dot{y}) = F_{\Theta(t)}(x, y) \tag{4}$$

is well defined and induces a non-autonomous semi-flow

$$\xi_\Theta : \{(s, t) \in \mathbf{R}^2 \colon t \ge s \ge 0\} \times (\mathscr{P}(S) \times \mathbf{R}) \to (\mathscr{P}(S) \times \mathbf{R})$$

where for all $t \ge s \ge 0$, $p \in \mathscr{P}(S) \times \mathbf{R}$ $t \to \xi_\Theta(s, t, p)$ is the solution to (4) with initial condition $\xi_\Theta(s, s, p) = p$.

Set $\tau_0 = 0$ and $\tau_n = 1 + \dfrac{1}{2} + \cdots + \dfrac{1}{n}$ for $n \in \mathbf{N}^*$. Given $X \in S^{\mathbf{N}_*}$ let $\mathscr{X} \in \mathscr{H}$ be defined by $\mathscr{X}(t) = X_{n+1}$ on $[\tau_n, \tau_{n+1}[$ and let $P : \mathbf{R}_+ \to \mathscr{P}(S) \times \mathbf{R}$ be the function which is affine on each interval $[\tau_n, \tau_{n+1}]$ and satisfies

$$P(\tau_n) = \left(\overline{X}_n, \frac{1}{n}\sum_{k=1}^{n} f_{X_k}(\overline{X}_{k-1})\right)$$

for all $n \in \mathbf{N}$, where $\overline{X}_0$ is arbitrary chosen in $\mathscr{P}(S)$.

The idea of the proof is as follow. First, we will show (Lemma 1) that for s large enough, $t \to \xi_{\mathscr{X}}(s, s+t, P(s))$ shadows $t \to P(t+s)$ over every finite interval of time. Then, we will prove[2] (Lemma 2), that $s \to \xi_{\Theta}(s, s+t, p)$ eventually enters the subset of $\mathscr{P}(S) \times \mathbf{R}$ defined by the inequality:

$$y \leq \sum_{i=0}^{d} x_i f_i(x).$$

Lemma 1. *For all* $T > 0$

$$\lim_{s\to\infty}\left[\max_{0\leq t\leq T} dist(\xi_{\mathscr{X}}(s, s+t, P(s)), P(s+t))\right] = 0$$

Proof: By definition of P and $\mathscr{X}$

$$P(\tau_{n+1}) - P(\tau_n) = (\tau_{n+1} - \tau_n)F_{\mathscr{X}(\tau_n)}(P(\tau_n)).$$

Thus

$$P(s+t) - P(s) = \int_s^{s+t} [F_{\mathscr{X}(u)}(P(u)) + \epsilon(u)]\, du$$

where $\epsilon(u) = F_{\mathscr{X}(u)}(P(\tau_n)) - F_{\mathscr{X}(u)}(P(u))$ for $u \in [\tau_n, \tau_{n+1}[$.

Let $\varepsilon(s, T) = \max_{s\leq u\leq s+T}\|\epsilon(u)\|$ and let $L > 0$ be a Lipschitz constant for F_i, $i = 0, \ldots, d$. Then for all $0 \leq t \leq T$

$$\|\xi_{\mathscr{X}}(s, s+t, P(s)) - P(s+t)\|$$
$$\leq \int_s^{s+t} L\|\xi_{\mathscr{X}}(s, s+u, P(s)) - P(u)\|\, du + T\varepsilon(s, T).$$

Thus by Gronwall's inequality

$$\|\xi_{\mathscr{X}}(s, s+t, P(s)) - P(s+t)\| \leq e^{LT} T\varepsilon(s, T).$$

For all $u \in [\tau_n, \tau_{n+1}[$

$$\|\epsilon(u)\| \leq L\|P(u) - P(\tau_n)\| \leq \frac{1}{n+1} L \max_{p\in K, \theta=0,1} \|F_\theta(p)\|.$$

Hence $\varepsilon(s, T) \leq Ce^{-s}$ for some constant $C > 0$. This proves the lemma. **QED**

[2] In the first version of the paper we proved this result for $d = 1$. Josef Hofbauer has shown us how to extend the result to $d \geq 1$ provided $(f_0, \ldots, f_d)$ has a gradient structure.

Choose $R > 0$ large enough so that $|f_\theta(x)| < R$ for all $x \in \mathscr{P}(S)$, $\theta \in S$ and set

$$K = \mathscr{P}(S) \times [-R, R].$$

It is easy to see that if $p \in K$ then $\xi_\Theta(s, t, p) \in K$ for all $t \geq s \geq 0$.

Let

$$\mathscr{U} = \left\{ p = (x, y) \in K : y > \sum_{i=0}^{d} x_i f_i(x) \right\}.$$

Lemma 2. *Let $A \subset \mathscr{U}$ be a compact set. There exists $T_A > 0$ such that for all $\Theta \in \mathscr{H}$, $s \geq 0$ and $p \in A$*

$$\{\xi_\Theta(s, s+t, p)\ 0 \leq t \leq T_A\} \cap (K \backslash A) \neq \varnothing$$

Proof [With help of Josef Hofbauer]. To shorten notation set $f(x) = (f_0(x), \dots, f_d(x))$. Then (4) rewrites

$$\dot{x} = -x + \Theta(t), \quad \dot{y} = -y + \langle f(x), \Theta(t) \rangle$$

where $\langle\, , \rangle$ is the usual inner product on $\mathbf{R}^{d+1}$. Thus

$$\dot{y} = -y + \langle f(x), x + \dot{x} \rangle.$$

Hence

$$\frac{y(s+t) - y(s)}{t} = \frac{1}{t} \int_s^{s+t} [\langle f(x(u)), x(u) \rangle - y(u)]\, du + \frac{1}{t} \int_s^{s+t} \langle f(x(u)), \dot{x}(u) \rangle\, du.$$

Now, the normalization constraint $\sum_{i=0}^{d} x_i = 1$ yields

$$\langle f(x(u)), \dot{x}(u) \rangle = \sum_{i=0}^{d} f_i(x(u)) \dot{x}_i(u) = -\sum_{i=1}^{d} g_i(x(u)) \dot{x}_i(u) = -\frac{d}{du} V(x(u))$$

where we write $V(x(u))$ for $V(x_1(u), \dots, x_d(u))$. Therefore

$$\frac{y(s+t) - y(s)}{t} = \frac{1}{t} \int_s^{s+t} [\langle f(x(u)), x(u) \rangle - y(u)]\, du - \frac{1}{t} [V(x(s+t)) - V(x(s))].$$

Letting $t \to \infty$ leads to

$$\lim_{t\to\infty}\frac{1}{t}\int_s^{s+t}[\langle f(x(u)),x(u)\rangle - y(u)]\,du = 0.$$

This proves the result. **QED**

We now conclude the proof of part (b). Suppose $Q(X) > Q_*$. Choose real numbers a and b such that $Q_* < a < b < Q(X)$. Let

$$A = \{p = (x,y) \in K : y \geq a\}, \quad B = \{p = (x,y) \in K : y \geq b\}$$

and T_A be as in Lemma 2. For all s large enough $P(s) \in B$. Thus by Lemma 1 there exists $s_0 \geq 0$ such that $\xi_{\mathscr{X}}(s, s+t, P(s)) \in A$ for all $s \geq s_0$ and $0 \leq t \leq T_A$. This is contradictory with Lemma 2.

Remarks: • The proof of Lemma 2 shows that the conclusion to Lemma 2 remains true if $\mathscr{U}$ is replaced by

$$\mathscr{V} = \left\{p = (x,y) \in K : y < \sum_{i=0}^{d} x_i f_i(x)\right\}.$$

Using this last remark it is not hard to see that one cannot expect a result similar to Theorem 1 with a payoff defined with a "lim sup" (or "lim") instead of the "lim inf."

• Theorem 4 can be generalized as follows. Let (γ_n) be a sequence of nonnegative numbers such that $\gamma_n \to 0$ and $\sum_n \gamma_n = \infty$. Let x_n and q_n be defined by

$$x_{n+1} - x_n = \gamma_{n+1}(-x_n + X_{n+1}),$$

$$q_{n+1} - q_n = \gamma_{n+1}(-q_n + f_{X_{n+1}}(x_n));$$

and let $Q(X) = \liminf_{n\to\infty} q_n$. Then the conclusions of Theorem 4 continue to hold with x_n in place of $\overline{X}_n$.

3. More than two strategies

The proof given in the preceding section shows that the condition that $f = (f_0, \ldots, f_d)$ has a gradient structure is sufficient to Theorem 4. We give here a class of examples (for $d > 1$) for which both this condition and the theorem fail.

Let

$$E_\alpha^d = \left\{x = (x_0, \ldots, x_d) \in \mathbf{R}^{d+1} : \sum_{i=0}^{d} x_i = \alpha\right\}$$

and let

$$G : E_1^d \to E_0^d,$$

$$x \to (G_0(x), \ldots, G_d(x))$$

be a smooth bounded vector field. The flow induced by G is the family $u = \{u_t\}$ of diffeomorphims defined by $u_0(x) = x$ and

$$\frac{du_t(x)}{dt} = G(u_t(x)).$$

A set $\Gamma \subset E_1^d$ is said to be *invariant* under G if for every $x \in \Gamma$ $u_t(x) \in \Gamma$ for all $t \in \mathbf{R}$.

An *attractor* for G is a nonempty compact invariant set $\Gamma \subset E_1^d$ having an open neighborhood U such that

$$\lim_{t\to\infty} \mathsf{dist}(u_t(x), \Gamma) = 0$$

uniformly in $x \in U$. Suppose that

(a) G points inward $\mathscr{P}(S)$ at the boundary of $\mathscr{P}(S)$. That is $G_i(x) > 0$ for $x_i = 0$.
(b) There exists an attractor $\Gamma \subset \mathscr{P}(S)$ for G that contains no equilibrium. That is

$$\Gamma \cap G^{-1}(0) = \emptyset.$$

(c) The map $p = Id + G$ maps $\mathscr{P}(S)$ into its interior.

It is easy to construct a vector field verifying these properties for $d > 1$. For (a) and (b) it suffices, for example, to choose a vector field G on E_1^d pointing inward $\mathscr{P}(S)$ at $\partial\mathscr{P}(S)$ and having an attracting periodic orbit $\Gamma \subset \mathscr{P}(S)$. Then replace G by ϵG. For $\epsilon > 0$ small enough this ensures property (c) without affecting properties (a) and (b).

We now construct our counterexample. Set

$$f_0(x) = -\sum_{i=1}^{d} x_i G_i(x)$$

and

$$f_i = f_0 + G_i$$

for $i = 1, \ldots, d$. Clearly f cannot have a gradient structure; for otherwise G would be a gradient vector field, contradicting assumption (b) above.

It follows from the definition of f that

$$\sum_i x_i f_i(x) = 0. \tag{5}$$

Let

$$\delta = \inf_{x \in \Gamma} \left(\sum_{i=1}^{d} G_i^2(x) \right) > 0.$$

We claim that there exists a strategy X ensuring a payoff

$$Q(X) \geq \delta > Q^* = 0.$$

Set $x_n = \overline{X}_n$ and $y_n = \frac{1}{n} \sum_{k=0}^{n-1} f_{X_{k+1}}(x_n)$. The pair (x_n, y_n) evolves according to the dynamics

$$x_{n+1} - x_n = \frac{1}{n+1}(-x_n + X_{n+1}),$$

$$y_{n+1} - y_n = \frac{1}{n+1}(-y_n + f_{X_{n+1}}(x_n)).$$

Now, suppose that at time $n+1$ the player chooses to play action e_i (i.e. $X_{n+1} = e_i$) with probability $p_i(x_n)$. By a classical method in the *theory of stochastic approximations* the long term behavior of this dynamics can be described in terms of the following system of differential equations

$$\frac{dx}{dt} = -x + E(X_{n+1} \mid x_n = x) = -x + p(x),$$

$$\frac{dy}{dt} = -y + E(f_{X_{n+1}}(x_n) \mid x_n = x) = -y + \langle f(x), p(x) \rangle. \tag{6}$$

Using (5) we get

$$\langle f(x), p(x) \rangle = \langle f(x), x + G(x) \rangle = \langle f(x), G(x) \rangle = \sum_{i=1}^{d} G_i^2(x).$$

Therefore the system (6) rewrites

$$\frac{dx}{dt} = G(x),$$

$$\frac{dy}{dt} = -y + \sum_{i=1}^{d} G_i^2(x). \tag{7}$$

Let $L(\{x_n\})$ and $L(\{(x_n, y_n)\})$ respectively denote the limit sets of $\{x_n\}$ and $\{(x_n, y_n)\}$. It follows from stochastic approximation theory that

(i) The event $\mathscr{E} = \{L(\{x_n\}) \subset \Gamma\}$ has positive probability (see Theorem 7.3 in (Benaïm, 1999))
(ii) $L(\{(x_n, y_n)\})$ is a compact set invariant by (7) (see e.g. section 5 in (Benaïm, 1999))

On $\mathscr{E}$, $L(\{(x_n, y_n)\})$ is therefore a compact invariant set of (7) restricted to $\Gamma \times \mathbf{R}$. But, for $(x, y) \in \Gamma \times \mathbf{R}$ $\frac{dy}{dt} \geq -y + \delta$. Hence $L(\{(x_n, y_n)\}) \subset \Gamma \times [\delta, \infty[$ on $\mathscr{E}$. Since $\mathscr{E}$ has positive probability, we have proved the existence of a strategy $X \in S^{\mathbf{N}^*}$ ensuring a payoff $Q(X) \geq \delta$.

References

Benaïm M (1999) Dynamics of stochastic algorithms. In J. Azema et al., eds, *Séminaire de Probabilités XXXIII*, L.N. in Math. 1708, 1–68 Springer Verlag, New York

Benaïm M and Hirsch MW (1999) Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games. *Game and Economic Behavior* **29**, 36–72

Fudenberg D and Levine DK (1998) *The Theory of Learning in Games*. MIT Press, Cambridge, Massachusetts, London, England

Fudenberg D and Kreps K (1993) Learning mixed equilibria, *Games and Economic Behavior* **5**, 320–367

Kaniovski Y and Young H (1995) Learning dynamics in games with stochastic perturbations. *Games and Economic Behavior* **11**, 330–363

Miyasawa K (1961) On the convergence of the learning process in a 2×2 non-zero-sum two-person game, Res. Mem. No. 33, Econ. Res. Program, Princeton Univ., Princeton, NJ

Robinson J (1951) An iterative method of solving a game, *Ann. Math.* **54**, 296–301