Working Paper

Cycling with a Generalized Urn Scheme and a Learning Algorithm for 2×2 Games

Martin Posch

WP-94-76
August, 1994

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 71521 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Cycling with a Generalized Urn Scheme and a Learning Algorithm for 2×2 Games

Martin Posch

WP-94-76
August, 1994

International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 71521 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Cycling with a Generalized Urn Scheme and a Learning Algorithm for 2x2 Games*

Martin Posch

August 19, 1994

Abstract

In this paper we explore a learning algorithm for 2x2 games. We assume that the players neither know the payoff matrix of their opponent nor their own and can only observe their own actions and their own payoffs. We prove that the learning process, which is modelled by a generalized urn scheme, converges to a pure strategy profile if the game has at least one strict Nash equilibrium. In case there is no strict Nash equilibrium, the learning algorithm exhibits oscillations. We derive sufficient conditions that cycling occurs in a generalized urn scheme. *Journal of Economic Literature Classification Number: C73*

*I wish to thank Yuri Kaniovski and Karl Sigmund for stimulating suggestions and helpful comments.

1 Introduction

The idea of Nash equilibrium is probably the most important concept in game theory. There are essentially two interpretations of Nash equilibrium. One belongs to the static approach of traditional game theory, whereas the other is a dynamic interpretation based on an evolutionary viewpoint.

Traditional game theory assumes that the players are rational and therefore can determine Nash equilibria or some refinement thereof by deduction. However, this view has been strongly challenged on the ground that the players would have to know all the possible actions and the preferences of the other players to be able to evaluate a Nash equilibrium. This is a very strong assumption, since information about the preferences may not be public and the evaluation of the Nash equilibrium may cost much effort. Furthermore if the game has multiple Nash equilibria it is often necessary to assume a preplay bargaining process to guarantee that the players agree on the same Nash equilibrium.

In the dynamic interpretation a Nash equilibrium is understood as the result of a learning or evolutionary process. The game is played repeatedly, and after each round the players update their strategies. Thus the Nash equilibrium need not be evaluated by supernaturally intelligent agents but can be found iteratively by following more or less simple rules.

One class of learning processes is based on the idea of fictitious play. One assumes that the players can observe the actions of the others and can compute a best response, provided they know the strategies of the other players. Thus, based on prior beliefs and the history of the play, the players can make hypotheses about the strategies of their opponents and then play a best reply

to the expected behavior of their opponents. After each round the players update their beliefs. These so called Bayesian learning processes have been studied recently in several game theoretic contexts by e.g. Eichberger *et al.* (1991), Jordan (1991) and Milgrom and Roberts (1991). The main topic of these papers is to study convergence of Bayesian learning to Nash equilibria.

A more naive approach will be explored in this paper. We assume that players can only observe their own actions and their own payoffs. Thus, they are not able to evaluate a best response, but can only naively learn by trial and error.

Kraines and Kraines (1993) studied what they call *Pavlovian* learning algorithm for the Prisoners Dilemma. They assume that the players consider payoffs below a certain limit as non satisfactory. Their learning algorithm works as follows: The players start with a mixed strategy (that is a probability distribution on the actions) and choose an action at random. If the payoff is satisfactory they increase the probability of repeating this action, otherwise they decrease it. Kraines and Kraines show that in the case of the Prisoners Dilemma both players will end up cooperating, hence by not playing Nash equilibrium for the one round game.

In contrary to the Pavlovian approach, which formalizes negative and positive conditioning, we consider a self-reinforcing learning model, which turns out to be a stochastic version of the replicator dynamics for the game. Again the player starts with some mixed strategy. He or she chooses an action at random and increases the probability of playing that action (and as a consequence decreases the probability of playing one of the other actions). The amount of the increment depends on the payoff he receives. Additionally one assumes that the longer the game evolves, the smaller the changes of probabilities at each round become. This seems reasonable, since individuals

are usually less ready to change their behaviour if they already have lots of experience.

Since both players are learning simultaneously a basic result is that the learning algorithm need not converge, but a cycling of strategy profiles may occur.

Cycling in learning models is often assumed to be an unrealistic feature, since players should be able to detect cycles that emerge (see e.g. Mailath, 1992). Nevertheless, cyclic behavioral patterns have been observed in many human conflicts (see e.g. the pig cycle in Rosenmüller (1972)). Thus, the bounded rationality assumption may not be as implausible as it seems. In any case, realistic or not, cycling is an interesting property of learning mechanisms.

In this paper we will study the oscillating dynamics of a learning algorithm in the context of a two person normal form game. To model the learning algorithm we use a generalized Polya urn scheme, which will be described in the next section. In section 3 we describe the game and the learning algorithm. In sections 4 and 5 we give results on convergence and cycling for urn processes which we use in section 6 to classify the learning dynamics. The proofs are left for the appendix.

2 Urn Models and a First Learning Algorithm

To give an insight into the development of urn models, we first describe the urn model formulated by Polya and Eggenberger (1923). Consider an urn of infinite capacity that contains one black and one white ball. Now balls are

iteratively added to the urn according to the following rule: Draw a ball from the urn at random, replace it and add one additional ball of the same color to the urn. Will the frequency of black (resp. white) balls oscillate randomly between $[0,1]$ or will it converge to a limiting frequency X? Polya (1931) proved that the frequencies converge indeed, and that the limit frequencies are uniformly distributed on [0,1].

This model has been extensively generalized. In the Polya urn model the probability that a black (resp. white) ball is added to the urn is equal to the current frequency of black (resp. white) balls. Hill *et al.* (1980) introduced a model where the probability to add a black ball is given by an arbitrary function of the frequencies, called the *urn function.* Brian Arthur, Yuri Ermoliev and Yuri Kaniovski (1984; 1987; 1988;) generalized this model further by considering urns with more than two types of balls, and urns where more than one ball at a time may be added. Finally, Dosi and Kaniovski (1994) considered models with several urns, where the urn function depends on the frequencies of the balls in all urns.

Before we study the learning algorithm for the normal form game, we introduce a learning algorithm for a decision problem that can be realized by a generalized urn scheme. It is a simplified version of a learning algorithm studied in Arthur (1993), where it was also mentioned that it can be applied to normal form games.

Consider an agent that can choose between two actions I and II. Action I leads in 10% of the cases to a payoff of 100 units and in 90% to a payoff of only 10 units. Action II leads in 90% to a payoff of 50 Units and in 10% to a payoff of 10 units. Obviously in the long run action II is the optimal choice, but since the agent has no prior information about the probability distribution of the payoff he has to learn.

The learning algorithm now works as follows: The agent has an urn with infinite capacity containing an arbitrary but positive number of balls of type I and II. To determine the next action he draws a ball from the urn at random and replaces it. Then he triggers the according action and observes his payoff. Now he adds as many balls of the type he has drawn to the urn as he received units of payoff.

Thus the frequency of balls of type I and II gives the probability that he chooses the first resp. second action. The initial urn composition can be interpreted as his prior belief. Brian Arthur (1993) asserts for a qualitatively equal model that the frequencies of balls converge a.s. such that in the limit the optimal strategy is chosen with probability one.

This learning model has some very plausible properties. On the one hand it is self-reinforcing, so that if one has chosen a certain action, the probability that the same action will be chosen next time increases. This is realistic, since changing the behaviour usually is attended with expenditure and effort.

Another important property is that the learning process becomes more and more stable as time evolves. In the beginning the frequencies of balls will fluctuate strongly due to stochastic events. But later the stochastic fluctuations have only little impact on the frequencies of balls since the total number of balls is growing very fast.

Brian Arthur compared this algorithm to the learning behaviour of humans and found that humans are much faster in exploiting the gained knowledge than this algorithm, so that they may get locked in a non optimal action. If they e.g. got by chance the first ten times a payoff of 100 units per round with action I but only 10 units with action II, they would stick with action I and no longer test action II. The learning algorithm in the contrary keeps

exploring alternative strategies and thus converges to the optimal action. Thus. the urn scheme is a sort of "zero hypothesis" for a learning algorithm, the most simple learning rule one can think of, which has to be modified to fit to actually observed learning mechanisms, since it is learning much slower.

One can imagine a realization of this learning rule in an organism if one thinks of cells containing two substances SI and SII instead of urns and balls. The probabilities of choosing the actions I and II are given by the concentrations $SI/(SI + SII)$ and $SII/(SI + SII)$, respectively. The cells could be neurones whose firing rates are proportional to the concentration of some substance and the choice of action could be determined by which neurone fires first (cf. Maynard Smith (1982)).

3 The Learning Algorithm for the 2x2 Normal Form Game

We consider two agents (A and B) playing a repeated normal form game. Each agent has two possible actions, I and II at his disposal. The payoffs they receive after every round depend on the payoff matrices $\mathcal{A} = (a_{jk})$ for player A and $\mathcal{B} = (b_{jk})$ for player B, where $j, k = 1, 2$. Thus, if player A chooses action j and his opponent action k he gets a_{jk} units and the other player b_{kj} units of payoff. We assume that a_{jk} and b_{jk} are positive, where $j, k = 1.2$.

Again the strategy of each player is a probability distribution on the two actions. which can be represented by the frequencies of balls in an urn.

Assume that every player has an urn with balls of type I and II. Before every

round of the game he draws one ball at random. He triggers this action and observes his payoff. Now he adds to the urn as many balls of the type drawn as he has received units of payoff. Since we did not require that the payoffs are integers, the numbers of balls need not be whole numbers.

Finally, without changing the relative frequencies of balls, he renormalizes the number of balls in the urn, such that at round n there are n balls in the urn. Thus the total number of balls in each urn is increasing linearly. The last step of the algorithm is for technical reasons only. Since it guarantees that at every time instant n the total number of balls in both urns is equal, it simplifies the analysis essentially.

This algorithm is a special case of the learning algorithm introduced by Brian Arthur (1993). In his model the total number of balls (which he calls strength) is renormalized to $C \cdot n^{\nu}$, where C and ν are positive numbers. Hence, we consider the case $C = \nu = 1$.

A similar learning algorithm was studied by A. Ianni (1993), but since she puts the emphasis on convergence results, she does not need the normalization of the number of balls in the urn.

To study the dynamics of the strategies (that is the frequencies of balls) we introduce some notation. First we note that the frequency of balls in each urn is well defined by the relative frequency of balls of type I (resp. II). Thus, it suffices to analyze the dynamics of the relative frequency of the type I balls.

Denote by S_n^A (resp. S_n^B) the total number of type I balls in the urn of player A (resp. B) at time n. Since we assumed that at time n there are n balls in

each urn, the relative frequencies of type I balls (denoted as x_n^A resp. x_n^B) are

$$x_n^i := \frac{S_n^i}{n}, \quad i = A, B.$$

and (x^A, x^B) lies in the square $Q := [0,1] \times [0,1]$.

Now let σ_n^A denote the random variable describing the number of type I balls that are added to the urn of player A at time n. Thus we have

$$\sigma_n^A := \begin{cases} \textit{Payoff of Player } A & \text{if player A chose action I;} \\ 0 & \text{if he chose action II.} \end{cases}$$

The distribution of σ_n^A is given by

$$\begin{aligned} P(\sigma_n^A = a_{11}) &= x_n^A \cdot x_n^B; \\ P(\sigma_n^A = a_{12}) &= x_n^A \cdot (1 - x_n^B); \\ P(\sigma_n^A = 0) &= 1 - x_n^A. \end{aligned}$$

By analogy we define σ_n^B to be the increment of type I balls in the urn of player B.

Let P_n^A (resp. P_n^B) denote the random variable describing the payoff of player A (resp. B) in round n.

Thus for the dynamics of the total numbers of type I balls we get

$$S_{n+1}^i = (S_n^i + \sigma_n^i) \cdot \frac{n+1}{n + P_n^i}, \quad i = A, B.$$

where the factor on the right comes from the normalization.

Hence for the relative frequencies we obtain

$$
\begin{aligned}
x^i_{n+1} &= \frac{S^i_{n+1}}{n+1} && (1)\\
&= x^i_n + \frac{1}{n+P^i_n}(\sigma^i_n - x^i_n\, P^i_n) && (2)\\
&= x^i_n + \frac{1}{n}(\sigma^i_n - x^i_n\, P^i_n) + \epsilon^i_n(x_n), && (3)
\end{aligned}
$$

where $i = A, B$ and $x_n := (x^A_n, x^B_n)$.

Since $\frac{a}{n+b} = \frac{a}{n} - \frac{ab}{n^2+nb}$ and since P^i_n as well as $(\sigma^i_n - x^i_n\, P^i_n)$ are bounded on Q we get $\epsilon^i_n(x) = O(\frac{1}{n^2})$.

Let $\mathbf{F_n}$ denote the σ-algebra generated by $\{x^A_1, x^B_1, x^A_2, \ldots, x^B_n\}$. We set

$$f^i(x_n) := E(\sigma^i_n - x^i_n P^i_n \,|\, \mathbf{F_n}), \quad i = A, B,$$

so that $n\, f^i(x_n)$, $i = A, B$ is the expected increment of x^i_n up to the ϵ-term, given the history of the game till time n.

We rewrite the difference equation (3) to

$$x^i_{n+1} = x^i_n + \frac{1}{n} f^i(x_n) + \frac{1}{n}\mu^i(x_n) + \epsilon^i_n(x_n), \quad x_n \in Q, \tag{4}$$

where $\mu^i(x_n) := \sigma^i_n - x^i_n\, P^i_n - f^i(x_n)$ and $i = A, B$.

Equation (4) consists of a deterministic "driving" part, a stochastic perturbational part (the μ-term in (4)) and an error term of order $O(1/n^2)$. Since $E(\mu^i(x_n) \,|\, \mathbf{F_n}) = 0$, the expected motion of the process (x^i_n) is given by the "driving" part of (4) up to an $O(1/n^2)$. Thus on the average the motion is directed by the term $f^i(x)$.

4 Convergence Results

Brian Arthur, Yuri Ermoliev and Yuri Kaniovski (1984; 1987; 1988;) studied very general urn processes. Using stochastic approximation results in Nevelson and Hasminskii (1973) they gave convergence results and a classification of the fixed points of the system into attainable and unattainable points which lead in our context to the following theorems.

To get a simpler notation we set $f := (f^A, f^B)$, $\mu := (\mu^A, \mu^B)$ and so on.

Since the function f gives the expected motion of the process, it is intuitively clear that if the process converges with positive probability to a point $\theta \in Q$ then $f(\theta) = 0$. These points are called the *fixed points* of the process (x_n).

However, the system need not converge at all: Even in the purely deterministic system $x^i_{n+1} := x^i_n + f^i(x_n)$, depending on the function f, cycles may emerge. (See e.g. the deterministic discrete game dynamics of Hofbauer (1994)). A sufficient condition for convergence is given by a Ljapunov function.

Theorem 1 *Let (x_n) be an urn processes as defined by (4) such that f is continuous. Let $B = \{ x \mid f(x) = 0\}$ be the set of fixed points of the deterministic system, and assume that B has only finitely many connected components. If there exists a C^2-Ljapunov function $v : Q \to \mathbf{R}$ such that*

1. $v(x) \geq 0, \quad \forall x \in Q;$
2. $\langle \nabla v(x), f(x) \rangle < 0, \quad \forall x \in Q - B.$

then $\lim_{n\to\infty} d(x_n, B) = 0$ a.s., where $d(x, B)$ denotes the distance of the point x to the set B.

Proof: The theorem is a consequence of Theorem 7.3 in Nevelson and Hasminskii (1973).

Hence, if all connected components of B are singletons, then the process (x_n) converges a.s. to a random vector $\bar{x}$ with $\bar{x} \in B$.

However not all the fixed points of f are attained in the limit with positive probability. There are fixed points such that the expected motion f points towards them, and fixed points where f points away. Hence we say that θ is

- a *sink* if the Jacobian $Df(\theta)$ has only eigenvalues with strictly negative real part.
- a *source* (resp. a *saddle*), if all (resp. at least one) eigenvalues have strictly positive real part.

Theorem 2 *Let $\theta \in Q$ be a sink of the process (x_n) defined by (4). Then*

$$P(\lim_{n\to\infty} x_n = \theta) > 0.$$

Proof: This is a direct generalization of Theorem 2 in Arthur *et al.* (1988).

Theorem 3 *Let $\theta \in intQ$ be a source or a saddle of the process (x_n) defined by (4). Then*

$$P(\lim_{n\to\infty} x_n = \theta) = 0.$$

Proof: For the proof we apply Theorem 5 in Arthur *et. al.* (1988).

Unfortunately there is still no result on the attainability of sources and saddles on the boundary of Q. Since at the boundary the variance of the process is vanishing it is much harder to get a corresponding result. Nevertheless it is conjectured that the theorem also holds for sources and saddles on the boundary.

Thus, to prove convergence of the learning algorithm we have to find appropriate Ljapunov functions. In section 6 we will give a classification of the 2x2 games and provide Ljapunov functions where they exist.

5 Cycling

If no strict Ljapunov function for the learning process exists, but instead an invariant of motion, the process exhibits cycling with positive probability. We derive sufficient conditions for cycling for a generalized urn scheme, which also covers the learning process (4). We give here the results for the learning process and leave the proof for the appendix.

Assume that the stochastic difference equation (4) has exactly one interior fixed point $\theta \in intQ$, i.e. $f(\theta) = 0$. Let H be an invariant of motion such that

a. $H \in C^2(intQ)$ and the second derivatives are bounded;

b. $\langle \nabla H, f \rangle = 0, \quad \forall x \in intQ$;

c. $H(x) \geq 0, \quad \forall x \in intQ$;

d. $H(x) = 0, \quad \forall x \in bdQ$;

e. θ is a global strict maximum of H and the only critical point.

Interpreting the function H as a mountain over the square Q, the conditions (c)-(e) imply that it has a unique peak at θ and level zero at the boundary of Q. Hence for every $c \in \mathbf{Im}\, H(Q)$ the set $H^{-1}(c)$ is a closed curve around the fixed point or the fixed point itself.

First we show that the process converges a.s. to these closed curves or to the fixed point. To this end we prove that the invariant of motion H applied to x_n converges a.s. for $n \to \infty$.

Proposition 1 *The limit* $\lim_{n\to\infty} H(x_n)$ *exists almost surely.*

Hence the process $H(x_n)$ converges to a random variable $\bar{H}$, which can take values in $\mathbf{Im}\, H(Q)$. Next we show that for every open interval I in $\mathbf{Im}\, H(Q)$ the probability that $\bar{H}$ is in I is positive.

Proposition 2 *For all* $c \in \mathbf{Im}\, H(Q)$ *and* $\epsilon > 0$ *we have*

$$P(\bar{H} \in]c-\epsilon,\, c+\epsilon[) > 0.$$

Since the sets $H^{-1}(]c-\epsilon,\, c+\epsilon[)$ are rings around the fixed point θ, we can deduce in particular that the process does not converge a.s. to the boundary of Q or the interior fixed point.

To prove that the learning process spins around the fixed point with positive probability we make a change of coordinates by moving the fixed point to the center $(0,0)$ and denote the new coordinates for simplicity again by x_n.

The angle between two points x_n, x_{n+1} is given by

$$\Delta\phi_n := \arctan\left(\frac{x_n^A x_{n+1}^B - x_{n+1}^A x_n^B}{x_n^A x_{n+1}^A + x_n^B x_{n+1}^B}\right). \tag{5}$$

Let

$$\phi_n := \sum_{n=N_0}^{n} \Delta\phi_n.$$

The process x_n spins around the fixed point θ if $|\phi_n| \to \infty$ for $n \to \infty$.

Theorem 4 *The process a.s. either*

- *converges to the boundary of Q or the interior fixed point;*

or

- *there is an N_0, such that for all $n > N_0$ the angles $\Delta\phi_n$ are well defined and we have*

$$|\phi_n| \to \infty \text{ and } \Delta\phi_n \to 0 \text{ for } n \to \infty.$$

The second dynamics emerges with positive probability.

For the proof we show that if the process does not converge to the fixed point or the boundary (which is by Proposition 2 with positive probability the case) it "follows" a.s. a solution of the differential equation $\dot{x} = f(x)$. Since H is an invariant of motion for this differential equation all its solutions in the interior of Q up to the fixed point θ are periodic. Thus, the learning process follows the periodic solutions of the differential equation with positive probability and hence cycles around the fixed point.

6 Classification of the Dynamics

Since the dynamics of the stochastic process (4) depends on the deterministic part, we first evaluate the expected motion f.

$$f^A(x^A, x^B) = x^A(1 - x^A)\big(\alpha_1 - x^B(\alpha_1 + \alpha_2)\big); \quad (6)$$

$$f^B(x^A, x^B) = x^B(1 - x^B)\big(\beta_1 - x^A(\beta_1 + \beta_2)\big); \quad (7)$$

where

$$\alpha_1 := a_{12} - a_{22}, \quad \alpha_2 := a_{21} - a_{11},$$
$$\beta_1 := b_{12} - b_{22}, \quad \beta_2 := b_{21} - b_{11}.$$

Hofbauer and Sigmund (1988) discussed the dynamics of the differential equation $\dot{x} = f(x)$ with f defined as in (6,7), which is the replicator dynamics for asymmetric games. They gave a classification of the dynamics, which will also be appropriate for the dynamics of the stochastic difference equation (4). We give here only the results of the analysis supplemented with the according Ljapunov functions.

To avoid degenerate cases we assume that both $\alpha_1 \cdot \alpha_2 \neq 0$ and $\beta_1 \cdot \beta_2 \neq 0$. Note that independent of the payoff matrices the four vertices of Q are zeros of f and thus fixed points of the process.

If $\alpha_1 \cdot \alpha_2 < 0$ then $f^A(x)$ does not change its sign in Q. If additionally $\beta_1 \cdot \beta_2 < 0$ then the same holds for $f^B(x)$ and the sum of the coordinates with appropriately chosen signs gives a Ljapunov function: $v(x^A, x^B) := \pm x^A + \pm x^B$. If $\beta_1 \cdot \beta_2 > 0$ then $f^B(x)$ changes its sign at $x^A = \frac{\beta_1}{\beta_1 + \beta_2}$. Hence

choosing the proper signs

$$v(x^A, x^B) := \pm x^A + \pm x^B(x^A - \frac{\beta_1}{\beta_1 + \beta_2})$$

is a Ljapunov function. By analogy we get a Ljapunov function if $\alpha_1 \cdot \alpha_2 > 0$ and $\beta_1 \cdot \beta_2 < 0$.

In the above cases, thus if $\alpha_1 \cdot \alpha_2 < 0$ or $\beta_1 \cdot \beta_2 < 0$, the game has only one Nash equilibrium. It is strict and coincides with the only sink of f. Since there is no fixed point in $intQ$, by Theorem 1 the process converges a.s. to a random vector $\bar{x}$ which can take values in the set of the fixed points on the vertices. Since at present there are no results on the attainability of saddles and sources on the boundary, we cannot prove that in the limit the players play the Nash equilibrium with probability one, although this seems to be the case. However, since the strict Nash equilibrium is a sink, we can deduce from Theorem 2 that it is attained in the limit with positive probability.

It remains to consider the case when $\alpha_1 \cdot \alpha_2 > 0$ and $\beta_1 \cdot \beta_2 > 0$. In this case there is a unique interior fixed point in intQ, given by

$$\theta = \left(\frac{\beta_1}{\beta_1 + \beta_2}, \frac{\alpha_1}{\alpha_1 + \alpha_2}\right).$$

θ is a Nash equilibrium but not strict. We have to distinguish two cases:

If $\alpha_1 \cdot \beta_1 > 0$ then θ is a saddle and there are two strict Nash equilibria on the vertices (see Fig. 1). Again we can find a Ljapunov function

$$v(x^A, x^B) := \pm \left(x^A x^B - x^A \frac{\alpha_1}{\alpha_1 + \alpha_2} - x^B \frac{\beta_1}{\beta_1 + \beta_2}\right).$$

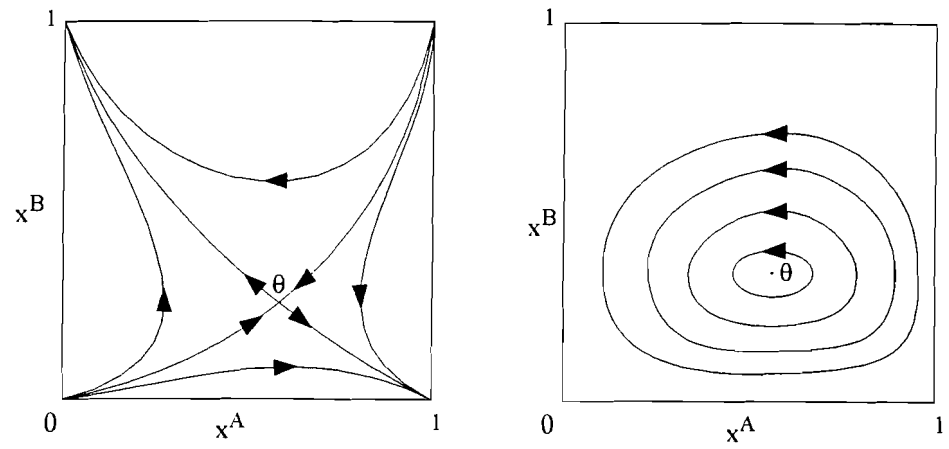
Thus, by Theorem 1 the process converges a.s. to a random vector $\bar{x}$ which can take values in the set of the fixed points on the vertices and the interior fixed point. Since the interior fixed point is a saddle, by Theorem 3 it is attained in the limit with probability 0. Two of the fixed points on the boundary are sinks (the two strict Nash equilibria) and the other sources. Hence, by Theorem 2 both strict Nash equilibria are attained in the limit with positive probability. Again we cannot prove that the process will converge to one of the sinks with probability one.

Finally, if $\alpha_1 \cdot \beta_1 < 0$ the interior fixed point is a center and there is no strict Nash equilibrium. For this case Hofbauer and Sigmund derived an invariant of motion

$$H(x^A, x^B) := (x^B)^{\alpha_1} (1 - x^B)^{\alpha_2} (x^A)^{-\beta_1} (1 - x^A)^{-\beta_2}. \tag{8}$$

Since the only critical point of H is the fixed point θ and since $H(x) = 0$ on the boundary of Q, all solutions of the differential equation $\dot{x} = f(x)$ in the interior of Q generate periodic orbits around the fixed point (see Fig. 2). Furthermore the time average of the strategies $\frac{1}{t} \int_0^t x(t)\, dt$ converges to the interior fixed point.

The invariant of motion (8) satisfies the conditions (a)-(e) in section 5. Hence we can apply Theorem 4 and deduce that with positive propability also the stochastic learning algorithm exhibits an oscillating behavior. However, since the step size of the learning process is of order $1/n$, the period of the cycles is growing exponentially. Thus, one cannot expect that the time average $\frac{1}{n} \sum_{k=1}^{n} x_k$ will converge.



Figures 1, 2. The phase portrait of $\dot{x} = f(x)$ in the two cases where there is an interior fixed point. The flow of the differential equation corresponds to the expected motion of the stochastic process.

On the tacit understanding that we exclude the degenerate cases where $\alpha_1 \cdot \alpha_2 = 0$ or $\beta_1 \cdot \beta_2 = 0$ we summarize the classification of the 2x2 games in the following theorems:

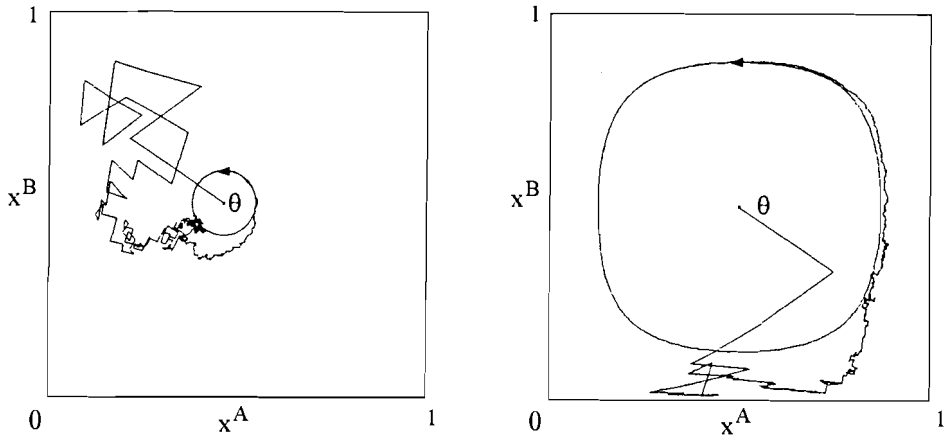
Theorem 5 *If there is at least one strict Nash equilibrium, then the learning algorithm a.s. converges to a pure strategy profile. With positive probability all strict Nash equilibria are attained in the limit .*

We conjecture that the process converges a.s. to the strict Nash equilibria.

Our main result is:

Theorem 6 *If there is no strict Nash equilibrium, then the process exhibits cycling with positive probability. If the process does not cycle, it a.s. converges either to the interior fixed point or to the boundary of Q.*

Figures 3 and 4 show two runs of the cycling learning process. For this plot we used the payoff matrices $\mathcal{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ and $\mathcal{B} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Here the interior fixed point (the Nash equilibrium) is $\theta = (\frac{1}{2}, \frac{1}{2})$, from where we started the process.



Figures 3, 4. Two runs of the learning process. For technical reasons the plots were calculated with the stochastic difference equation only until $n = 10^7$ and then were continued with the solution of the corresponding differential equation.

7 Appendix

Consider a stochastic process $x_n \in intQ := [0,1] \times [0,1]$, such that

$$x_{n+1} = x_n + \frac{1}{n} f(x_n) + \frac{1}{n} \mu(x_n) + \epsilon_n(x_n). \tag{9}$$

Let the following conditions hold

i. $f := (f^1, f^2)$ is Lipschitz-continuous;

ii. $E(\mu(x) \mid \mathbf{F_n}) = 0$ and $\mu(x)$ is bounded;

iii. $\epsilon_n(x)$ is a random variable such that $\epsilon_n(x) = O(1/n^2)$;

iv. The function f has exactly one interior zero θ;

v. For every $x_1 \in intQ$, every open set $U \subseteq Q$ and every N there is an $N_0 > N$ such that $P(x_{N_0} \in U) > 0$.

The learning process (4) obviously satisfies conditions (i)-(iv). Condition (v) says that every open set in Q can be reached with positive probability. In Lemma 6 we show that the learning algorithm satisfies this condition.

Let H be an invariant of motion satisfying the conditions (a)-(e) in section 5.

Proof of Proposition 1: An application of Taylor's theorem gives

$$H(x_{n+1}) = H\left(x_n + \frac{1}{n} f(x_n) + \frac{1}{n}\mu(x_n) + \epsilon_n(x_n)\right)$$

$$
\begin{aligned}
&= H(x_n) + \frac{1}{n} \underbrace{\langle \nabla H(x_n), f(x_n) \rangle}_{=0} \\
&\quad + \frac{1}{n} \langle \nabla H(x_n), \mu(x_n) \rangle + k_n(x_{n+1}) \left(\frac{1}{n}\right)^2 \\
&\leq H(x_n) + \frac{1}{n} \langle \nabla H(x_n), \mu(x_n) \rangle + K \left(\frac{1}{n}\right)^2 .
\end{aligned}
$$

For the estimation of $k_n(x)$ by a $K \in \mathbf{R}^+$ which is independent of n and x, we used the fact that the second derivatives of H are bounded on $intQ$, and that $\epsilon_n(x)$ is an $O(1/n^2)$.

Hence for the expectations we have

$$
\begin{aligned}
&E\big(H(x_{n+1}) \,|\, \mathbf{F_n}\big) \\
&\quad \leq E\left(H(x_n) + \frac{1}{n} \langle \nabla H(x_n), \mu(x_n) \rangle + K \left(\frac{1}{n}\right)^2 \,\middle|\, \mathbf{F_n} \right) \\
&\quad = E\big(H(x_n) \,|\, \mathbf{F_n}\big) + \frac{1}{n} \underbrace{E\big(\langle \nabla H(x_n), \mu(x_n) \rangle \,|\, \mathbf{F_n} \big)}_{=0} + K \left(\frac{1}{n}\right)^2 \\
&\quad = H(x_n) + K \left(\frac{1}{n}\right)^2 . \qquad (10)
\end{aligned}
$$

Thus we obtain for all n

$$E\big(H(x_{n+1}) - H(x_n) \,|\, \mathbf{F_n}\big) \leq K \left(\frac{1}{n}\right)^2. \tag{11}$$

We now define the random variable

$$G(x_n) = H(x_n) + K \sum_{j \geq n} \left(\frac{1}{j}\right)^2$$

and get

$$\begin{aligned} &E\big(G(x_{n+1}) - G(x_n) \,|\, \mathbf{F_n}\big) \\ &\quad = \underbrace{E\big(H(x_{n+1}) - H(x_n) \,|\, \mathbf{F_n}\big)}_{\leq K\left(\frac{1}{n}\right)^2} - K \left(\frac{1}{n}\right)^2 \leq 0. \end{aligned}$$

We see that $G(x_n)$ is a nonnegative supermartingale, and by the Martingale Convergence Theorem (see e.g. Williams, 1991) converges pointwise with probability 1. Since $G(x_n)$ converges pointwise to $H(x_n)$ for $n \to \infty$, $H(x_n)$ converges too. □

Proof of Proposition 2: Fix a $c \in \mathbf{Im}\, H(Q)$ and an $\epsilon > 0$. Let $v(x) := (H(x) - c)^2$. Since v is a function of the invariant of motion H, it is itself an invariant of motion and satisfies the conditions of Proposition 1. Hence $v(x_n)$ converges a.s. for $n \to \infty$ and by (11) there is a K such that

$$E\big(v(x_{n+1}) - v(x_n) \,|\, \mathbf{F_n}\big) \leq K \frac{1}{n^2}.$$

We choose an N such that

$$K \sum_{n=N}^{\infty} \frac{1}{n^2} < \frac{\epsilon^2}{4}.$$

Let $U_\epsilon(c) := H^{-1}\big(]c - \epsilon, c + \epsilon[\big)$. By condition (v) there is an $N_0 > N$, such that $P(x_{N_0} \in U_{\epsilon/2}(c)) > 0$.

Thus, setting E=$\{\, x_{N_0} \in U_{\epsilon/2}(c) \,\}$ we get $P(E) > 0$. On E we have $v(x_{N_0}) \leq \frac{\epsilon^2}{4}$.

Since $v(x_{N_0})$ is $\mathbf{F}_{N_0}$-measurable we have for all $n > N_0$ on E

$$\begin{aligned} E\big(v(x_n) \,|\, \mathbf{F}_{N_0}\big) &\leq E\big(v(x_{N_0}) \,|\, \mathbf{F}_{N_0}\big) + K \sum_{j=N_0}^{n} \frac{1}{j^2} \\ &\leq \underbrace{v(x_{N_0})}_{\leq \frac{\epsilon^2}{4}} + \frac{\epsilon^2}{4} \leq \frac{\epsilon^2}{2}. \end{aligned}$$

Let $F = \{\, \lim_{n\to\infty} v(x_n) > \epsilon^2 \,\} \cap E$ be the event that x_{N_0} is in $U_{\epsilon/2}(c)$ and the process does not enter $U_\epsilon(c)$ from a given time onward. Obviously $F \subseteq E$.

Assume $F = E$ a.s.. Then by the Lemma of Fatou we get on E

$$\epsilon^2 < E(\underbrace{\lim_{n\to\infty} v(x_n)}_{>\epsilon^2 \text{ on } F=E} \,|\, \mathbf{F}_{N_0}) \leq \lim_{n\to\infty} E(v(x_n) \,|\, \mathbf{F}_{N_0}) \leq \frac{\epsilon^2}{2}$$

which is a contradiction.

Since $P(E) > 0$ we obtain $P(E - F) > 0$ and get $P(\lim_{n\to\infty} v(x_n) < \epsilon^2) > 0$.
□

As in Nevelson (1973) we prove that the sum of the stochastic perturbations converges.

Lemma 1 *The stochastic process* $Y_n := \sum_{k=1}^{n} \frac{1}{k}\mu(x_k)$ *is an* L^2-*martingale.*

Hence we can apply the Martingale Convergence Theorem (see e.g. Williams (1991)) for L^2-martingales, and conclude that the pointwise limit $\lim_{n\to\infty} Y_n(\omega) = Y_\infty(\omega)$ exists a.s..

Proof: Since $E(Y_{n+1} - Y_n \,|\, \mathbf{F_n}) = \frac{1}{n+1} E(\mu(x_{n+1}) \,|\, \mathbf{F_n}) = 0$ we have that Y_n is a martingale.

Let $Y_0 = 0$. Since the martingale differences are orthogonal in L^2, we deduce

$$\|Y_n\|_{L^2}^2 = \sum_{k=1}^{n} \|Y_n - Y_{n-1}\|_{L^2}^2.$$

Hence

$$\|Y_n\|_{L^2}^2 = \sum_{k=1}^{n} \|\frac{1}{k}\mu(x_k)\|_{L^2}^2 \leq K \sum_{k=1}^{n} \frac{1}{k^2} \leq \infty.$$

□

Proof of Theorem 4: Let Ω be the event that $H(x_n)$ and $Y(x_n)$ converge. By Proposition 1 and Lemma 1 we have $P(\Omega) = 1$. Let $(x_n) := (x_n)(\omega)$, $\omega \in \Omega$ be a path which neither converges to the fixed point nor to the boundary. According to Proposition 2 this occurs with positive probability.

We will prove that this path (x_n) spins around the fixed point. We rewrite the difference equation (9) such that

$$x_{n+1} = x_n + \frac{1}{n} f(x_n) + \epsilon'_n,$$

where $\lim_{n\to\infty} H(x_n)$ exists and $\sum_{n=1}^{\infty} \epsilon'_n < \infty$.

Let $x(t, a, t_0)$ denote the solution of the differential equation (DE)

$$\dot{x} = f(x), \quad x(t_0) = a. \tag{12}$$

Since H is an invariant of motion for the DE up to the fixed point θ all solutions in $intQ$ are periodic. In the following we will prove, that (x_n) "follows" a solution of the DE (12).

Let $c := \lim_{n\to\infty} H(x_n)$ and $\gamma := H^{-1}(c)$. Since H is an invariant of motion for the DE (12), γ is the orbit of a solution in $intQ$. Since $\lim_{n\to\infty} H(x_n) = c$ we deduce that (x_n) converges to the set γ.

First we show that the angles $\Delta\phi_n$ are well defined:

From a certain time on (x_n) is very close to γ. Since in the new coordinates the global maximum of H is the origin and $x_n \not\to (0,0)$ we have $(0,0) \notin \gamma$. Hence there is an $\epsilon > 0$ s.t. $\exists N_0$ with $\|x_n\| > \epsilon \; \forall n > N_0$. Since the step size of the process converges to zero, the denominator is bounded from below. In addition, the nominator is bounded from above by an $O(1/n)$.

Hence we can choose an N_0 such that for all $n > N_0$, (5) is well defined.

The proof that the path (x_n) follows a solution $x(t, a, t_0)$ of (12) will be given in several steps. In Lemma 2 and 3 we prove that (x_n) stays *for some time* close to a solution of the DE, if it started close enough. In Lemma 5 we prove that for an adapted time scale s_n, with arbitrarily small steps, the solutions of the ODE (12) (with proper initial conditions) stay close to the path (x_n) *forever.*

To approximate the path (x_n) by a solution of the ODE we introduce a time scale t_n, s.t.

$$t_n := \sum_{k=1}^{n} \frac{1}{k} \cdot \tag{13}$$

In the first two lemmata we adapt a discrete version of Gronwalls Lemma, which has been proved by Benveniste et.al. (1990).

Lemma 2 *If* $\nu_r \leq r_1 \sum_{i=1}^{r} \gamma_i \nu_{i-1} + r_2$ *for* $r = 0, 1, ..., n$ *with* r_1, r_2, γ_i *positive, then*

$$\nu_n \leq (r_2 + \gamma_1 \nu_0) \exp(r_1 \sum_{i=1}^{n} \gamma_i).$$

Proof: We may suppose that $r_1 = 1$.

It is easily proved by induction that

$$1 + \sum_{i=1}^{r} \gamma_i \exp(\sum_{j=1}^{i-1} \gamma_j) \leq \exp(\sum_{i=1}^{r} \gamma_i)$$

holds for all $r \geq 1$.

Let P(r) denote the property: $\nu_r \leq (r_2 + \gamma_1 \nu_0) \exp(\sum_{i=1}^{r} \gamma_i)$.

P(1) reduces to $\nu_1 \leq \gamma_1 \nu_0 + r_2$ which is clearly true.

Suppose P(r) is true, then

$$\begin{aligned}
\nu_{r+1} &\leq \sum_{i=1}^{r+1} \gamma_i \nu_{i-1} + r_2 \\
&\leq \sum_{i=1}^{r+1} \gamma_i \left(r_2 \exp(\sum_{j=1}^{i-1} \gamma_j) + \gamma_1 \nu_0 \exp(\sum_{j=1}^{i-1} \gamma_j) \right) + r_2 \\
&\leq r_2 \left(1 + \sum_{i=1}^{r+1} \gamma_i \exp(\sum_{j=1}^{i-1} \gamma_j) \right) + \gamma_1 \nu_0 \sum_{i=1}^{r+1} \gamma_i \exp(\sum_{j=1}^{i-1} \gamma_j) \\
&\leq r_2 \exp(\sum_{i=1}^{r+1} \gamma_i) + \gamma_1 \nu_0 \exp(\sum_{i=1}^{r+1} \gamma_i).
\end{aligned}$$

Hence we proved $P(r+1)$. □

For $t > 0$ we denote the largest natural number n such that $\sum_{k=1}^{n} \frac{1}{k} < t$ by $M(t)$.

Lemma 3 *Let $\Delta T > 0$, and $a_0 \in Q$. Then for $N \leq n \leq M(t_N + \Delta T)$ we have*

$$\|x_n - x(t_n, a_0, t_N)\| \leq U(N) \exp(L \cdot \Delta T),$$

where L is the Lipschitz constant of f and $U(N) - \|x_N - a_0\| \to 0$ for $N \to \infty$.

Proof: For simplicity let $x(t) := x(t, a_0, t_N)$. Since L is the Lipschitz constant of f we have $\|f(x) - f(x')\| \leq L\|x - x'\|$ for all $x, x' \in Q$

Then for t_n defined in (13) we have

$$\begin{aligned} x(t_{n+1}) - x(t_n) &= \int_{t_n}^{t_{n+1}} f(x(s))\, ds \\ &= \frac{1}{n} f(x(t_n)) + \alpha_n, \end{aligned}$$

where $\|\alpha_n\| \leq L \left(\frac{1}{n}\right)^2$.

We wish to compare x_n and $x(t_n)$ for $n = N, \ldots, M(t_N + \Delta T)$.

Since

$$x_n - x(t_n) = x_{n-1} - x(t_{n-1}) + \frac{1}{n}[f(x_{n-1}) - f(x(t_{n-1}))] + \epsilon'_n + \alpha_n$$

we have

$$
\begin{aligned}
x_n - x(t_n) &= (x_N - a_0) + \sum_{k=N}^{n-1} \frac{1}{k+1}[f(x_k) - f(x(t_k))] + \\
&\quad \sum_{k=N}^{n-1} \epsilon'_k + \sum_{k=N}^{n-1} \alpha_k \\
\|x_n - x(t_n)\| &\leq \|x_N - a_0\| + L \sum_{k=N}^{n-1} \frac{1}{k+1} \|x_k - x(t_k)\| + \\
&\quad \sum_{k=N}^{n-1} \epsilon'_k + L \sum_{k=N}^{n-1} \left(\frac{1}{k+1}\right)^2 \\
&\leq \|x_N - a_0\| + L \sum_{k=N}^{n-1} \frac{1}{k+1} \|x_k - x(t_k)\| \\
&\quad + U_1(N) + U_2(N),
\end{aligned}
$$

where $U_1(N), U_2(N) \to 0$ for $N \to \infty$ since the sum $\sum_{k=N}^{n-1} \epsilon'_k$ converges for $n \to \infty$ by our assumptions.

Applying Lemma 2, we have for $N < n < M(t_N + \Delta T)$

$$
\|x_n - x(t_n)\| \leq \underbrace{\left(U_1(N) + U_2(N) + (1 + \frac{L}{N})\|x_N - a_0\|\right)}_{=:U(N)} \exp(L \cdot \Delta T)
$$

and obviously $U(N) - \|x_N - a_0\| \to 0$ for $N \to \infty$. □

Lemma 4 *Let $\Delta T > 0$. For every sufficiently small $\epsilon > 0$*

$$B_\epsilon(a) \cap \gamma \subseteq x(a,]-\Delta T, \Delta T[)$$

holds for all $a \in \gamma$, where $B_\epsilon(a)$ denotes an open ϵ-ball around the point a and

$$x(a,]-\Delta T, \Delta T[) := \{x(t,a,0) \mid t \in]-\Delta T, \Delta T[\}.$$

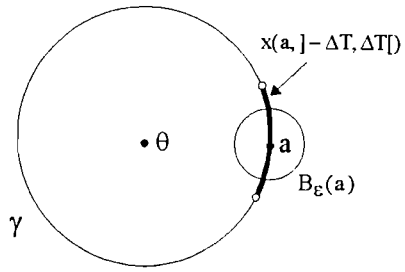


Figure 5. Illustration to Lemma 4.

Proof: For $\Delta T \geq \tau/2$, where τ is the period of the periodic orbit in γ, there is nothing to prove. Thus assume $\Delta T < \tau/2$.

Let $\epsilon_0 > 0$. We claim that for every $a \in \gamma$ we can find an $\epsilon_a < \epsilon_0$ s.t.

$$B_{\epsilon_a}(a) \cap (\gamma - x(a,]-\Delta T, \Delta T[)) = \emptyset. \tag{14}$$

Indeed, assume that this is not the case: Then there is a sequence (t_n) with $-\tau/2 < t_n \leq \tau/2$ such that $x(t_n, a, 0) \to a$ for $n \to \infty$ and $t_n \notin (-\Delta T, \Delta T)$. Thus we can find a converging subsequence (t_{n_k}) such that $t := \lim_{n\to\infty} t_{n_k}$. Obviously $t \neq z \cdot \tau$, $\forall z \in \mathbf{Z}$.

Since $x(t_{n_k}, a, 0) \to x(t, a, 0)$ we have $x(t, a, 0) = a$ and get a contradiction to the uniqueness of the solution of the ODE.

Since the balls $B_{\epsilon_a}(a)$ are open we can choose for every point a a maximal $\epsilon_a \leq \epsilon_0$ satisfying condition (14).

We still have to prove that we can choose the ϵ independently of a. To this end we show that the ϵ_a are bounded from below by a positive number.

Assume that this is not the case: Then there is a sequence (a_n) such that $\lim_{n\to\infty} \epsilon_{a_n} = 0$. Since γ is compact we can find a converging subsequence (a_{n_k}) such that $\lim_{k\to\infty} a_{n_k} =: a$. For this point a we choose an $\bar{\epsilon}_a > 0$ such that

$$B_{\bar{\epsilon}_a}(a) \cap \gamma \subseteq x(a,]-\Delta T/2, \Delta T/2[).$$

For k large enough we have $B_{\bar{\epsilon}_a/2}(a_{n_k}) \subset B_{\bar{\epsilon}_a}(a)$ and since there are no fixed points on γ we deduce for large k

$$\gamma(a, \Delta T/2) \subset x(a_{n_k},]-\Delta T, \Delta T[)$$

and get

$$B_{\bar{\epsilon}_a/2}(a_{n_k}) \subset x(a_{n_k},]-\Delta T, \Delta T[).$$

Since the $\epsilon_{a_{n_k}}$ were chosen to be maximal we have $\epsilon_{a_{n_k}} \geq \bar{\epsilon}_a/2$ for all large k. Hence the limit of the $\epsilon_{a_{n_k}}$ cannot be 0 and we obtain a contradiction. □

Lemma 5 *Let $\epsilon_0 > 0$. There is an N such that for all $n > N$ there are $s_n \in \mathbf{R}$ such that*

$$\|x_n - x(s_n)\| < \epsilon_0, \quad |s_{n+1} - s_n| \leq \epsilon_0 \text{ and } s_n \to \infty \text{ for } n \to \infty. \tag{15}$$

Proof:

In the following steps we choose a proper δ-neighbourhood of γ which we denote by $B_\delta(\gamma)$:

1. Choose ΔT_0 such that $\epsilon_0/2 > \Delta T_0 > 0$. By Lemma 4 we can choose an $\epsilon \leq \epsilon_0$ such that for all $a \in \gamma$ we have $B_\epsilon(a) \cap \gamma \subset \gamma(a, \Delta T_0)$.

   Set $\Delta T := \Delta T_0 + 2$.

2. By Lemma 3 we can choose $N_1 > 0$ and $\delta \leq \epsilon/2$, s.t. for all $n \geq N_1$ and $a \in Q$ with $\|a - x_n\| < \delta$ the following holds:

   For all k such that $n \leq k \leq M(t_n + \Delta T)$ we have:

   $$\|x_k - x(t_k, a, t_n)\| \leq \frac{\epsilon}{2}.$$

   Hence if the stochastic process and the solution of the DE are closer than δ at a time $n > N_1$, then for the time span ΔT their distance will not exceed $\epsilon/2$.

3. Choose $N \geq \max\{N_0, N_1, 1/(2\epsilon_0)\}$, such that $x_n \in B_\delta(\gamma)$ for $n \geq N$.

We will construct the times s_j, $j \geq N$ iteratively in blocks: First we choose an initial time s_{k_0} such that the distance of $x(s_{k_0})$ to $x_{s_{k_0}}$ is smaller than δ.

The iteration step: For $j = k_i + 1, \ldots, k_{i+1} - 1$ we use the original time scale. For these j the distance of the two processes is smaller than $\epsilon/2$. Then we choose a time $s_{k_{i+1}}$ such that the distance of $x_{s_{k_{i+1}}}$ to the solution of the differential equation $x(s_{k_{i+1}})$ is smaller then δ.

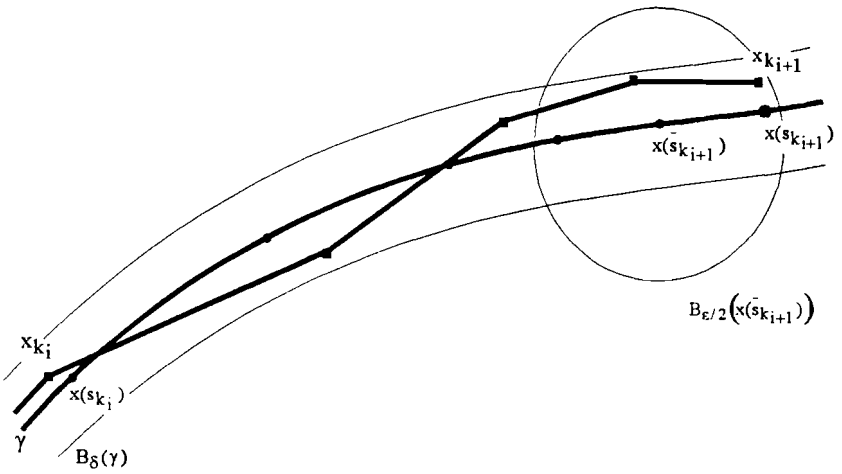


Figure 6. At round k_i the distance from x_{k_i} to $x(s_{k_i})$ is less than δ. By Lemma 3 we deduce that until round $k_i + 1$ the distance from x_j to $x(s_{k_i} + \sum_{l=k_i+1}^{j} \frac{1}{l})$ is less than $\epsilon/2$. Since we know that the process x_n is in the $\delta-$neighbourhood $B_\delta(\gamma)$, we can find a time $s_{k_{i+1}}$ close to $\bar{s}_{k_{i+1}} := s_{k_i} + \sum_{l=k_i+1}^{k_{i+1}} \frac{1}{l}$ such that the distance from $x(s_{k_{i+1}})$ to $x_{k_{i+1}}$ is less than δ.

Since $x_N \in B_\delta(\gamma)$ we can choose an $a_0 \in \gamma$ such that $\|a_0 - x_N\| < \delta$. Set $k_0 := N$ and $s_{k_0} := t_N$.

Let $x(s) := x(s, a_0, s_{k_0})$ denote the solution of the differential equation (12) starting at a_0.

Assume that we have constructed the sequence (s_j) until $j = k_i$, such that $\|x_{k_i} - x(s_{k_i})\| < \delta$.

We claim that there are s_j, $j = k_i + 1, \ldots, k_{i+1}$ such that we have

$$\|x_j - x(s_j)\| \leq \frac{\epsilon}{2}, \quad j = k_i + 1, \ldots, k_{i+1} \tag{16}$$

and

$$\|x_{k_{i+1}} - x(s_{k_{i+1}})\| \leq \delta, \tag{17}$$

where $k_{i+1} := M(t_{k_i} + \Delta T)$.

Additionaly we have

1. $s_{k_i} < s_j$ for $j = k_i + 1, \ldots, k_{i+1}$ and $s_{k_{i+1}} - s_{k_i} > 1$;
2. $|s_j - s_{j-1}| \leq \epsilon_0$ for $k_i + 1 \leq j \leq k_{i+1}$.

Proof of the claim:

Let $s_j := s_{k_i} + \sum_{l=k_i+1}^{j} \frac{1}{l}$ for $j = k_i + 1, \ldots, k_{i+1} - 1$.

Since the chosen N and δ satisfy the conditions of Lemma 3, we have for $j = k_i + 1, \ldots, k_{i+1} - 1$

$$\|x_j - \underbrace{x\big((s_j - s_{k_i}), x(s_{k_i}), s_{k_i}\big)}_{=x(s_j, a_0, s_{k_0})}\| < \frac{\epsilon}{2}$$

and obtain

$$\|x_j - x(s_j)\| < \frac{\epsilon}{2}.$$

Let $\bar{s}_{k_{i+1}} := s_{k_i} + \sum_{l=k_i+1}^{k_{i+1}} \frac{1}{l}$. By Lemma 3 we also have

$$\|x_{k_{i+1}} - x(\bar{s}_{k_{i+1}})\| < \frac{\epsilon}{2}. \tag{18}$$

Since $x_{k_{i+1}} \in B_\delta(\gamma)$ we can choose a time $s_{k_{i+1}}$, such that

$$\|x_{k_{i+1}} - x(s_{k_{i+1}})\| \leq \delta \tag{19}$$

and $|s_{k_{i+1}} - \bar{s}_{k_{i+1}}| \leq \tau/2$, where τ is the period of γ.

We still have to prove that $s_{k_{i+1}} - s_{k_i} > 1$.

To this end we deduce from $\delta < \epsilon/2$ and the inequalities (18),(19)

$$\|x(\bar{s}_{k_{i+1}}) - x(s_{k_{i+1}})\| \leq \epsilon.$$

Since we have chosen ϵ according to Lemma 4 we obtain

$$B_\epsilon(x(\bar{s}_{k_{i+1}})) \cap \gamma \subseteq x(x(\bar{s}_{k_{i+1}}),]-\Delta T_0, \Delta T_0[).$$

Thus we can deduce $\bar{s}_{k_{i+1}} - \Delta T_0 < s_{k_{i+1}} < \bar{s}_{k_{i+1}} + \Delta T_0$.

Hence

$$|\bar{s}_{k_{i+1}} - s_{k_{i+1}}| < \Delta T_0. \tag{20}$$

Since $\Delta T = \Delta T_0 + 2$ we have for the discrete times $\bar{s}_{k_{i+1}} - s_{k_i} > \Delta T_0 + 1$. (By switching to the discrete timescale we make at most an error of $1/N < 1$.) Using the triangle inequality we get with (20)

$$s_{k_{i+1}} - s_{k_i} > 1.$$

The step size $|s_j - s_{j-1}|$ is bounded by $1/N \leq \epsilon_0/2$ for $j = k_i + 1, \ldots, k_{i+1} - 1$ and for $j = k_{i+1}$ (according to (20)) by $\Delta T_0 + \epsilon_0/2 \leq \epsilon_0$. Thus the claim is proven.

Hence we can construct iteratively the sequence (s_n) with the properties stated in the lemma. □

Using Lemma 5 we finally prove the theorem.

Let $\dot{r}, \dot{o}$ denote the differential equation (12) expressed in polar coordinates, such that the fixed point is moved to the origin.

Since the polar coordinates depend continuously on the cartesian coordinates, according to Lemma 5 we can find for every $\epsilon_1 > 0$ an N and a solution $x(t)$ of the ODE (12) such that for all $n > N$ we have

$$|\sum_{k=N}^{n} \Delta\phi_k - \phi(s_n)| < \epsilon_1,$$

where

$$\phi(t) := \int_{t_N}^{t} \dot{\phi}(x(s))\, ds, \quad t > t_N.$$

Since we know that the solution of the ODE spins around the fixed point we have $|\phi(s_n)| \to \infty$ for $n \to \infty$ and hence $|\phi_n| \to \infty$ for $n \to \infty$.

Finally, since the step size of the process is an $O(1/n)$ the same holds for the angles and we get $\Delta\phi_n \to 0$ for $n \to \infty$. □

Lemma 6 *Let x_n be the learning process defined by (4) such that $x_1 \in intQ$. Then for every open set $U \subseteq Q$ and every N_0 there is an $N > N_0$ such that*

$$P(x_N \in U) > 0.$$

Proof: We examine the process (4) in its original shape (2):

$$x^i_{n+1} = x^i_n + \frac{1}{n + G^i_n}(\alpha^i_n - x^i_n G^i_n), \quad i = A, B.$$

Hence depending on the chosen actions the increments of x_n up to a factor of order $O(1/n)$ are given by

$$\text{I,I}: \ (R^A - x^A R^A,\ R^B - x^B R^B) \qquad \text{I,II}: \ (S^A - x^A S^A,\ -x^B T^B)$$

$$\text{II,I}: \ (-x^A T^A,\ S^B - x^B S^B) \qquad \text{II,II}: \ (P^A - x^A P^A,\ P^B - x^B P^B).$$

As long as $x_n \in intQ$ all actions are chosen with positive probability.

Let $\bar{x}$ be a point in U. We iteratively construct a path that converges to $\bar{x}$. Choose the actions according to the following rule:

1. If $x_n^A \leq \bar{x}^A$ and $x_n^B \leq \bar{x}^B$ choose the actions (I,I) until this inequality no longer holds.

   Note that by construction the path x_n does not come arbitrarily close to the bdQ. Hence the increments of x_n are bounded from below by an $O(1/n)$ and since $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ after finitely many steps the above inequality no longer holds.

2. If $x_n^A \leq \bar{x}^A$ and $x_n^B \geq \bar{x}^B$ choose action pair (I,II) until this inequality no longer holds, which is again the case in finite time.

In the other cases we choose the action pairs by analogy.

Since the increments are an $O(1/n)$, the path x_n will converge to the point $\bar{x}$. Hence there is a time N_1 after which the path does not leave the open set U. Thus for $N := \max(N_0, N_1)$ we have $x_N \in U$.

Since every step has positive probability, the path from x_1 to x_N has positive probability, too. □

References

[1] Arthur, W.B. (1993). "On Designing Economic Agents that Behave Like Human Agents," *J. Evol. Econ.* **3**, 1-22.

[2] Arthur, W.B., Ermoliev, Y.M., and Kaniovski, Y.M. (1984). "Strong Laws for a Class of Path-dependent Stochastic Processes with Applications," in *Proc. Conf. on Stochastic Optimization, Kiev 1984* (Arkin, Shiryayev, and Wets, Eds.), pp. 287-300. Berlin: Springer.

[3] Arthur, W.B., Ermoliev, Y.M., and Kaniovski, Y.M. (1987). "Nonlinear Urn Processes. Asymptotic Behavior and Applications," WP-87-85, International Institute for Applied Systems Analysis, Laxenburg Austria.

[4] Arthur, W.B., Ermoliev, Y.M, and Kaniovski, Y.M. (1988). "Nonlinear Adaptive Processes of Growth with General Increments. Attainable and Unattainable Components of Terminal Set," WP-88-86, International Institute for Applied Systems Analysis, Laxenburg Austria.

[5] Dosi, G., and Kaniovski, Y.M. (1994). "On 'Badly Behaved' Dynamics. Some Applications of Generalized Urn Schemes to Technological and Economic Change," Mimeo, International Institute for Applied Systems Analysis, Laxenburg Austria.

[6] Benveniste, A., Metivier and M., Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximation.* Berlin: Springer.

[7] Eichberger, J., Haller, H., and Milne, F. (1991). "Naive Bayesian Learning in 2x2 Matrix Games," Mimeo, University of Melbourne.

[8] Hill, B.M., Lane, D., and Sudderth, W. (1980). "A Strong Law for some Generalized Urn Processes," Ann. Prob. **8**, 214-226.

[9] Hofbauer, J. (1994). "Discrete Time Dynamics for Bimatrix Games," Mimeo, Univ. of Vienna.

[10] Hofbauer, J. and Sigmund, K. (1988). *The Theory of Evolution and Dynamical Systems.* Cambridge: Cambridge Univ. Press.

[11] Ianni, A. (1993). "On the Application of Generalized Urn Schemes to Evolutionary Models," Mimeo, Univ. College London.

[12] Jordan, J.S. (1991). "Bayesian Learning in Normal Form Games," *Games Econ. Behav.* **3**, 60-81.

[13] Kraines, D., and Kraines, V. (1993). "Learning to Cooperate with Pavlov. An Adaptive Stragey for the Iterated Prisoners Dilemma with Noise," *Theory and Decision* **35**, 107-150.

[14] Mailath, G.J. (1992). "Introduction. Symposium on Evolutionary Game Theory," *J. Econ. Theory* **57**, 259-277.

[15] Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge Univ. Press.

[16] Milgrom, P., and Roberts J. (1991). "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Econ. Behav.* **3**, 82-100.

[17] Nevelson, M.B. and Has'minskii, R.Z. (1973). *Stochastic Approximation and Recursive Estimation.* Amer. Math. Society Translations of Math. Monographs *47*, Providence.

[18] Pemantle, R. (1990). "Nonconvergence to Unstable Points in Urn Models and Stochastic Approximations," *The Ann. of Prob.* **18**, 698-712.

[19] Polya, G. and Eggenberger, F. (1923). "Über die Statistik verketteter Vorgänge," *Zeit. Angew. Math. Mech.* **3**, 279-289.

[20] Polya, G. (1931). "Sur quelques Points de la Théorie des Probabilités," *Ann. Inst. H. Poincaré* **1**, 117-161.

[21] Rosenmüller, J. (1972). "Konjunkturschwankungen," in *Selecta Mathematica IV* (Jacobs, K., ed.), pp. 143-173. Berlin: Springer.

[22] Williams, D. (1991). *Probability with Martingales.* Cambridge: Cambridge Univ. Press.