Math Meth Oper Res
DOI 10.1007/s00186-007-0158-9

ORIGINAL ARTICLE

Fictitious play in stochastic games

G. Schoenmakers · J. Flesch · F. Thuijsman

Received: 18 January 2006 / Revised: 12 February 2007

Abstract In this paper we examine an extension of the fictitious play process for bimatrix games to stochastic games. We show that the fictitious play process does not necessarily converge, not even in the $2 \times 2 \times 2$ case with a unique equilibrium in stationary strategies. Here $2 \times 2 \times 2$ stands for 2 players, 2 states, 2 actions for each player in each state.

Keywords Non-cooperative games · Stochastic games · Fictitious play

1 Introduction

A bimatrix game is given by a set of players $I = \{1, 2\}$, a set $A = \{1, 2, \ldots, n_A\}$ of pure actions for player 1 and, similarly, a set $B = \{1, 2, \ldots, n_B\}$ of pure actions for player 2, and a payoff function $r : A \times B \to \mathbb{R}^2$. Independently of each other the players have to choose actions. They are allowed to randomize over their actions, so player 1 would generally choose a mixed action from F, which is the set of probability distributions over A, and similarly player 2 chooses actions from G, the set of probability distributions over B. We shall write f and g to denote elements of F and G. Clearly, pure actions can be seen as mixed actions as well. When the action pair $(a, b) \in A \times B$ is played, then player $i \in I$ receives a payoff $r_i(a, b)$. If the players play the mixed actions f and g, then the expected payoff to player i is

$$r_i(f, g) := \sum_{a \in A} \sum_{b \in B} f(a) g(b) r_i(a, b)$$

G. Schoenmakers · J. Flesch · F. Thuijsman (✉)
Mathematics Department, Maastricht University, PO Box 616, 6200MD Maastricht, The Netherlands
e-mail: frank@math.unimaas.nl

The *fictitious play process for bimatrix games* is based on the assumption that the bimatrix game is played repeatedly at stages $n \in \{1, 2, 3, \ldots\}$, where at stage n player i plays a best reply against the "observed behavior" of his opponent. This means that if player 2 has played $b(1), b(2), b(3), \ldots, b(n) \in G$ at stages $1, 2, 3, \ldots, n$, then player 1 has observed the action frequencies $g(n) = \frac{1}{n}\sum_{m=1}^{n} b(m)$. He may deduce that player 2 is playing according to the mixed action $g(n)$ and play a best reply $a(n+1)$ at stage $n+1$. Player 2 is assumed to respond similarly by playing at stage $n+1$ a best reply $b(n+1)$ against the action frequencies $f(n)$ based on the observed actions $a(1), a(2), a(3), \ldots, a(n)$ by player 1. The process is initiated by taking $a(1) = b(1) = 1$.

The fictitious play process is said to converge if $(f(n), g(n))_{n=1}^{\infty}$ converges. A game has the fictitious play property if every fictitious play process converges to an equilibrium.

Fictitious play processes were introduced by Brown (1951) and Robinson (1951), who proved the fictitious play property for two-player zero-sum games. Miyasawa (1961) proved the fictitious play property for generic 2×2-games. A geometric proof for this class of games is provided by Metrick and Polak (1994). Convergence was also shown by Monderer and Shapley (1996) for n-player games in which all players have the same number of actions and identical payoff functions. Shapley (1964), however, provided an example of a bimatrix game where each player has three actions and where the fictitious play process does not converge.

In this paper we use a discrete fictitious play process. For continuous fictitious play processes we refer to Krishna and Sjöström (1998) and Sela (2000).

The 2-player stochastic game model was introduced by Shapley (1953). It can be described as follows: let $I = \{1, 2\}$ be the set of players and let S be a finite set of states. For each state $s \in S$ there are finite sets of actions A_s and B_s for players 1 and 2, respectively. Play can start in any state s. If in state s player 1 chooses a_s and player 2 chooses b_s, then two things happen: (1) player i receives a payoff $r_{i,s}(a_s, b_s)$ and (2) with probability $p_s(t|a_s, b_s)$ play moves to state $t \in S$, where actions have to be chosen at the next stage. The number of stages is assumed to be infinite. Again, each player is allowed to randomize over his pure actions in each state. Both players are assumed to evaluate the infinite sequence of payoffs by means of a limiting average reward. We shall use the following notations. The set of joint pure actions of player 1 is denoted by $A = \times_{s\in S} A_s$. An element $a \in A$ is called a joint pure action of player 1. For player 2 the set B is defined analogously. Furthermore $f_s \in F_s$ shall denote a mixed action of player 1 in state s, where F_s is the set of all mixed actions of player 1 in state s. For player 2 we use g_s and G_s. We write $F = \times_{s\in S} F_s$ for the set of joint mixed actions f of player 1, and G for player 2. A strategy π of player 1 is an infinite sequence of joint actions: $\pi = (f(n))_{n=1}^{\infty}$, where for all n the joint action $f(n)$ may depend on the history up to stage $n-1$. Likewise $\sigma = (g(n))_{n=1}^{\infty}$ denotes a strategy for player 2. A stationary strategy is a strategy, which prescribes the same joint mixed action each stage: $x = (f)^{\infty}$ and $y = (g)^{\infty}$. The limiting average reward for player i is defined by $\gamma_i(s, \pi, \sigma) = \liminf_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} E_{s\pi\sigma}(R_i(n))$, where s is the starting state, $R_i(n)$ the random variable payoff to player i at stage n, and E is the expectation. A pair of strategies (π, σ) is called an equilibrium if $\gamma_1(s, \pi', \sigma) \le \gamma_1(s, \pi, \sigma)$ and

$\gamma_2(s, \pi, \sigma') \leq \gamma_2(s, \pi, \sigma)$ for all s, for all π' and for all σ', i.e. π and σ are best replies against each other for all initial states.

Generally, equilibria fail to exist in stochastic games, a famous example of which was provided by Gillette (1957). If, instead of considering best replies one considers ε-best replies ($\varepsilon > 0$), then ε-equilibria are known to exist for the 2-player case (cf. Vieille 2000a,b). These ε-equilibria generally require the use of history-dependent strategies. Equilibria, 0-equilibria that is, are known to exist only for classes of stochastic games that have some additional structure on the payoffs and/or transitions.

It is well-known that against a fixed stationary strategy the opponent always has a pure stationary strategy as a best reply (cf. Hordijk et al. 1983).

We now define a fictitious play process for stochastic games as a generalization of the fictitious play process described above: $f(1) = a(1)$ and $g(1) = b(1)$ are the joint actions for players 1 and 2, respectively, that consist of action 1 in each state. Let $a(2)$ be a joint action for player 1 with the property that $(a(2))^\infty$ is a best reply to $(g(1))^\infty$, and define $f(2) = \frac{1}{2}a(1) + \frac{1}{2}a(2)$; define $b(2)$ and $g(2)$ analogously. Continue recursively for $n \geq 3$ by letting $a(n)$ denote a joint action for player 1 with the property that $(a(n))^\infty$ is a best reply to $(g(n-1))^\infty$. Similarly $b(n)$ will be defined as a best reply to $(f(n-1))^\infty$. Next, $f(n) = \frac{1}{n}\sum_{m=1}^{n} a(m)$ and $g(n) = \frac{1}{n}\sum_{m=1}^{n} b(m)$ can be used to derive $a(n+1)$ and $b(n+1)$ analogously.

We would like to emphasize that this fictitious play process does not correspond to any play of the game itself. Nevertheless, for one-state stochastic games this extension coincides with the original fictitious play process for bimatrix games. We would also like to stress that this definition of fictitious play is different from the one introduced by Vrieze and Tijs (1982). In their paper a ficititious play process is defined for so-called β-discounted zero-sum stochastic games, in which the stage payoffs are evaluated by discounting. Their approach uses the fact that in any stochastic game stationary β-discounted optimal strategies exist and can be derived from related auxiliary matrix games. Such is not possible for limiting average reward stochastic games.

In this paper we examine a particular example of a $2 \times 2 \times 2$ stochastic game. Here $2 \times 2 \times 2$ stands for 2 players, 2 states, 2 actions for each player in each state. We show for this example that the fictitious play process does not converge, even though the game has a unique equilibrium in stationary strategies. Moreover the example is a so-called irreducible single-controller stochastic game with state independent transitions, i.e. with probability 1 both states will be visited infinitely often in any play, only one player's actions determine the transition probabilities and the transition probabilities are independent of the states. It is well known for irreducible stochastic games and for single controller stochastic games that stationary equilibria always exist (cf. Rogers 1969; Sobel 1971; Filar 1981).

2 The example

Consider the following $2 \times 2 \times 2$ stochastic game:

2, 1 (0.9; 0.1)	4, 0 (0.9; 0.1)
0, 0 (0.1; 0.9)	7, 1 (0.1; 0.9)

state 1

0, 1 (0.9; 0.1)	2, 0 (0.9; 0.1)
2, 0 (0.1; 0.9)	4, 1 (0.1; 0.9)

state 2

In each state the rows and the columns are the actions of player 1 and player 2, respectively. In each cell the numbers in the upper-left part are the payoffs to players 1 and 2, respectively, and the numbers in the lower-right part are the transition probabilities to states 1 and 2, respectively. Notice that the transition probabilities in this game depend only on the action of player 1 and they are independent of the state. Furthermore the game is irreducible, which means that irrespective of the players' strategies both states will be visited infinitely often with probability 1 and the limiting average rewards of the game do not depend on the starting state.

We now show that this game has a unique equilibrium in stationary strategies (f^*, g^*) where $f^* = ((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))^\infty$ and $g^* = ((\frac{1}{5}, \frac{4}{5}), (\frac{9}{20}, \frac{11}{20}))^\infty$.

Suppose player 1 plays stationary strategy $f = ((f_1, 1-f_1), (f_2, 1-f_2))^\infty$ and player 2 plays $g = ((g_1, 1-g_1), (g_2, 1-g_2))^\infty$. Given these strategies the invariant distribution over the states, i.e. the proportions of time that the states are being visited, is given by

$$\left(\frac{\frac{1}{10} + \frac{4}{5} f_2}{1 - \frac{4}{5} f_1 + \frac{4}{5} f_2}, \frac{\frac{9}{10} - \frac{4}{5} f_1}{1 - \frac{4}{5} f_1 + \frac{4}{5} f_2} \right)$$

and, using the expected payoffs in each of these states, it follows that

$$\begin{aligned} \gamma_1(f, g) = {} & \frac{\frac{1}{10} + \frac{4}{5} f_2}{1 - \frac{4}{5} f_1 + \frac{4}{5} f_2} (5 f_1 g_1 - 3 f_1 - 7 g_1 + 7) + \frac{\frac{9}{10} - \frac{4}{5} f_1}{1 - \frac{4}{5} f_1 + \frac{4}{5} f_2} \\ & \times (4 - 2 f_2 - 2 g_2) \\ \gamma_2(f, g) = {} & \frac{\frac{1}{10} + \frac{4}{5} f_2}{1 - \frac{4}{5} f_1 + \frac{4}{5} f_2} (1 - f_1 - g_1 + 2 f_1 g_1) + \frac{\frac{9}{10} - \frac{4}{5} f_1}{1 - \frac{4}{5} f_1 + \frac{4}{5} f_2} \\ & \times (1 - f_2 - g_2 + 2 f_2 g_2) \end{aligned}$$

It is straightforward to verify that there are no equilibria in which at least one player uses a pure stationary strategy.

To see that (f^*, g^*) is an equilibrium observe that, if $h_1 = ((1, 0), (1, 0))^\infty$, $h_2 = ((1, 0), (0, 1))^\infty$, $h_3 = ((0, 1), (1, 0))^\infty$ and $h_4 = ((0, 1), (0, 1))^\infty$, then

$$\gamma_1(h_1, g^*) = \gamma_1(h_2, g^*) = \gamma_1(h_3, g^*) = \gamma_1(h_4, g^*) = \gamma_1(f^*, g^*) = 3.35$$

$$\gamma_2(f^*, h_1) = \gamma_2(f^*, h_2) = \gamma_2(f^*, h_3) = \gamma_2(f^*, h_4) = \gamma_2(f^*, g^*) = 0.5$$

Because all pure stationary strategies of player 1 are best replies against g^*, we conclude that f^* is a best reply as well. Similarly for player 2. Therefore, (f^*, g^*) is an equilibrium. Uniqueness of this equilibrium follows straightforwardly from the best reply structure, which is examined in more detail in the next section.

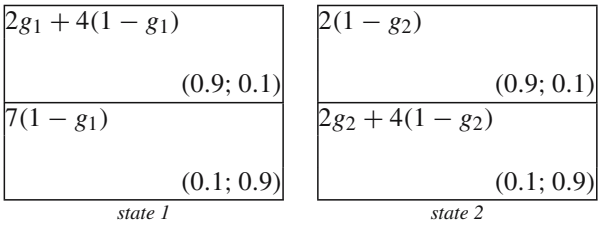
Now we state our main theorem.

Theorem *The fictitious play process for* $2 \times 2 \times 2$ *stochastic games does not need to converge.*

The proof of this theorem, which is based on an analysis of the best reply structure in the stationary strategy spaces, is given in the next section. The key of the proof is the observation of a cyclic pattern in the fictitious play process for the example presented.

3 The proof

We examine the best reply structure for stationary strategies in the example. We start with player 1. Take a fixed stationary strategy $g = ((g_1, 1 - g_1), (g_2, 1 - g_2))^\infty$ of player 2. Then player 1 faces the following Markov Decision Problem (MDP):



Let $v_{(a_1,a_2)}$ denote player 1's limiting average reward in the above MDP, when she plays the pure stationary strategy $(a_1, a_2)^\infty$. Notice that $(a_1, a_2)^\infty$ is a best reply to g if and only if $v_{(a_1,a_2)}$ is maximal.

We can calculate $v_{(1,1)}$ as follows. Suppose player 1 plays $(1, 1)^\infty$, then state 1 will, in expectation, be visited 9 stages out of 10 and

$$\begin{aligned} v_{(1,1)} &= 0.9(2g_1 + 4(1 - g_1)) + 0.1 \times 2(1 - g_2) \\ &= 3.8 - 1.8g_1 - 0.2g_2. \end{aligned}$$

The other values are:

$$\begin{aligned} v_{(1,0)} &= 4 - g_1 - g_2 \\ v_{(0,1)} &= 4.5 - 3.5g_1 - g_2 \\ v_{(0,0)} &= 4.3 - 0.7g_1 - 1.8g_2. \end{aligned}$$

Now we calculate the values of g_1 and g_2 for which player 1 is indifferent between some of her pure stationary strategies:

$$v_{(1,1)} = v_{(1,0)} \iff 3.8 - 1.8g_1 - 0.2g_2 = 4 - g_1 - g_2,$$

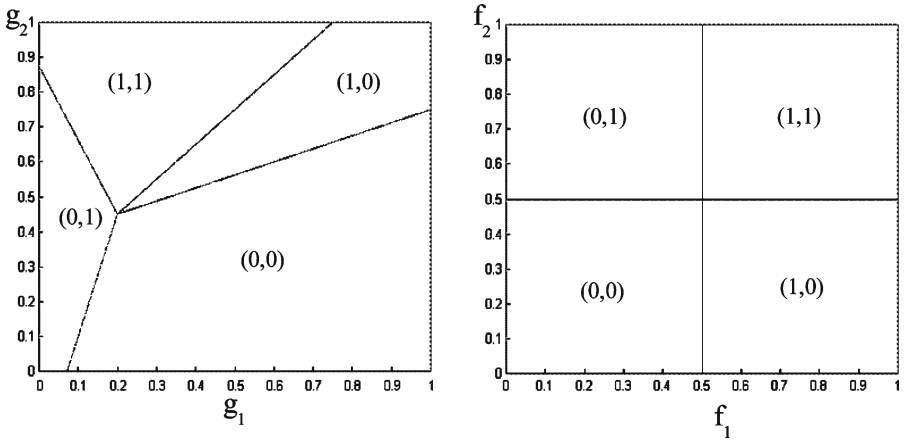


Fig. 1 Best reply structure for stationary strategies

hence

$$v_{(1,1)} = v_{(1,0)} \Longleftrightarrow g_2 = g_1 + \frac{1}{4}.$$

Analogously

$$\begin{aligned} v_{(1,1)} = v_{(0,1)} &\Longleftrightarrow g_2 = -\frac{17}{8}g_1 + \frac{7}{8} \\ v_{(1,0)} = v_{(0,0)} &\Longleftrightarrow g_2 = \frac{3}{8}g_1 + \frac{3}{8} \\ v_{(0,1)} = v_{(0,0)} &\Longleftrightarrow g_2 = \frac{7}{2}g_1 - \frac{1}{4}. \end{aligned}$$

From these equations we deduce the left part of Fig. 1, showing the best replies of player 1 against g. The lines in this figure correspond with the equations above. The lines divide the square into four regions. If (g_1, g_2) is in one of the regions, then the pure stationary strategy mentioned in the region is the pure best reply for player 1 against $g = ((g_1, 1 - g_1), (g_2, 1 - g_2))^\infty$. The common point of these regions corresponds to the equilibrium strategy g^*.

Since player 2 can only maximize her one-shot payoff we can easily deduce the right part of Fig. 1 showing the best replies of player 2 against an arbitrary stationary strategy $f = ((f_1, 1 - f_1), (f_2, 1 - f_2))^\infty$ of player 1. The two relevant indifference lines are $f_1 = \frac{1}{2}$ and $f_2 = \frac{1}{2}$.

Notice that Fig. 1 also indicates that $(1, 1)^\infty$ and $(0, 0)^\infty$ can only be best replies simultaneously at the equilibrium point. The same holds for $(1, 0)^\infty$ and $(0, 1)^\infty$. From Fig. 1 it is clear that for each player there is a unique stationary strategy against which all pure strategies of the opponent are best replies. This implies the uniqueness of the stationary equilibrium (f^*, g^*).

We will now derive some properties on how the fictitious play process evolves. This will be done in terms of so-called runs:

Definition 1 A run $[(a_1, a_2), (b_1, b_2)]$ is a part $((a(n_1), b(n_1)), (a(n_1 + 1), b(n_1 + 1)), \ldots, (a(n_2), b(n_2)))$ of the fictitious play process $(a(n), b(n))_{n=1}^{\infty}$, such that $(a(n), b(n)) = ((a_1, a_2), (b_1, b_2))$ for all $n \in \{n_1, \ldots, n_2\}$, whereas equality fails for $n = n_1 - 1$ and for $n = n_2 + 1$.

The next lemma shows how the different runs follow each other.

Lemma 1 *The following runs will succeed each other cyclically: first* $[(1, 1), (1, 1)]$*, then* $[(1, 0), (1, 1)]$*, then* $[(1, 0), (1, 0)]$*, then* $[(0, 0), (1, 0)]$*, then* $[(0, 0), (0, 0)]$*, then* $[(0, 1), (0, 0)]$*, then* $[(0, 1), (0, 1)]$*, then* $[(1, 1), (0, 1)]$ *and then we return to* $[(1, 1), (1, 1)]$ *and start a new cycle.*

Proof The proof is based on the fact that if we are in run $[(a_1, a_2), (b_1, b_2)]$ at stage n, then the action frequencies will change in the following way:

$$f(n) = \frac{n-1}{n} \cdot f(n-1) + \frac{1}{n}(a_1, a_2)$$
$$g(n) = \frac{n-1}{n} \cdot g(n-1) + \frac{1}{n}(b_1, b_2).$$

So, as n increases, $f(n)$ and $g(n)$ move along a straight line in the direction of the corner points (a_1, a_2) resp. (b_1, b_2).

Recall that the fictitious play process starts with a $[(1, 1), (1, 1)]$-run, hence both $f(1)$ and $g(1)$ are $(1, 1)$. So at stage 2 a $[(1, 0), (1, 1)]$-run is started, hence f moves in the direction of $(1, 0)$ and g stays at $(1, 1)$. At a certain stage in the right part of Fig. 1 the line $f_2 = \frac{1}{2}$ will be crossed and g starts moving towards $(1, 0)$, causing a $[(1, 0), (1, 0)]$-run to start. During this run both f and g move towards $(1, 0)$. But then at a certain stage in the left part of Fig. 1 the line between the $(1, 0)$-part and the $(0, 0)$-part will be crossed and the $(0, 0)$-part will be entered, which causes a $[(0, 0), (1, 0)]$-run to start. Analogous reasonings can be held to prove the occurrence of the other switches of run types. □

We will prove the nonconvergence of the fictitious play process by defining other processes on the left part of Fig. 1, called trajectories. We will show that these trajectories do not converge to the equilibrium point and that the fictitious play process follows lines that run even further away from the equilibrium point than the trajectories do.

Definition 2 Consider Fig. 2. A trajectory t is a set of four connected line segments that satisfies the following conditions:

(1) A trajectory starts and ends at line segment a, which corresponds to the equation $q_2 = \frac{3}{8}q_1 + \frac{3}{8}$, where $q_1 \in [\frac{1}{5}, 1]$. The starting point of a trajectory t is called $s(t)$ and the end point $e(t)$.
(2) In areas A, B, C and D the trajectory moves in the direction of the respective corner points $(0, 0)$, $(0, 1)$, $(1, 1)$ and $(1, 0)$.

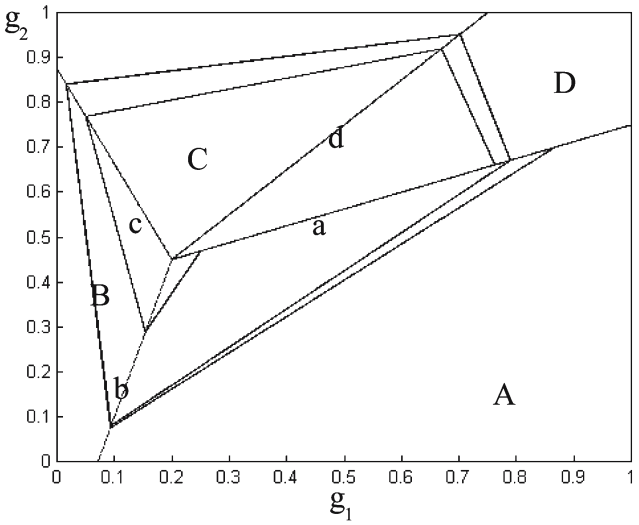


Fig. 2 Areas and trajectories

A trajectory t is called an orbit if $s(t) = e(t)$. An orbit $\bar{t}$, with $s(\bar{t}) = e(\bar{t}) = \psi$ is stable if for some small $\delta > 0$ the following contraction property holds: for all trajectories $t \neq \bar{t}$, if $||s(t) - \psi|| < \delta$, then $||e(t) - \psi|| < ||s(t) - \psi||$.

Lemma 2 *There are precisely 2 orbits, a stable one with starting point* $(\frac{15}{19}, \frac{51}{76})$ *and a non-stable one being the equilibrium point* $(\frac{1}{5}, \frac{9}{20})$.

Proof Finding orbits boils down to finding fixed points of a function h that assigns the finishing value $e(t)$ to the starting value $s(t)$ for each trajectory t.

For an arbitrary trajectory we have $s(t) = (\frac{1}{5} + \varepsilon, \frac{9}{20} + \frac{3}{8}\varepsilon)$ with $\varepsilon \in [0, \frac{4}{5}]$, which is at line segment a in Fig. 2, corresponding to the equation $q_2 = \frac{3}{8}q_1 + \frac{3}{8}$. The trajectory enters area A and moves in the direction of $(0, 0)$. As long as the trajectory is in area A it moves on the line $q_2 = \frac{\frac{9}{20} + \frac{3}{8}\varepsilon}{\frac{1}{5} + \varepsilon} q_1$. The trajectory leaves area A at line segment b, which corresponds to the equation $q_2 = \frac{7}{2}q_1 - \frac{1}{4}$. So at that moment we have

$$q_1 = \frac{\frac{1}{5} + \varepsilon}{1 + 12\frac{1}{2}\varepsilon} \quad \text{and} \quad q_2 = \frac{\frac{9}{20} + \frac{3}{8}\varepsilon}{1 + 12\frac{1}{2}\varepsilon}$$

and the trajectory enters area B. As long as the trajectory is in area B it moves on the line $1 - q_2 = \frac{\frac{11}{20} + 12\frac{1}{8}\varepsilon}{\frac{1}{5} + \varepsilon} q_1$. The trajectory leaves area B and enters area C at line segment c, corresponding to the equation $q_2 = -\frac{17}{8}q_1 + \frac{7}{8}$, so at that moment we have

$$q_1 = \frac{\frac{1}{5} + \varepsilon}{1 + 80\varepsilon} \quad \text{and} \quad q_2 = \frac{\frac{9}{20} + 67\frac{7}{8}\varepsilon}{1 + 80\varepsilon}.$$

As long as the trajectory is in area C it moves on the line $1-q_2 = \frac{\frac{11}{20}+12\frac{1}{8}\varepsilon}{\frac{4}{5}+79\varepsilon}(1-q_1)$. The trajectory leaves area C and enters area D at line segment d, which has $q_2 = q_1 + \frac{1}{4}$ as its equation, meaning that at that moment we have

$$q_1 = \frac{\frac{1}{5} + 188\frac{1}{2}\varepsilon}{1 + 267\frac{1}{2}\varepsilon} \quad \text{and} \quad q_2 = \frac{\frac{9}{20} + 255\frac{3}{8}\varepsilon}{1 + 267\frac{1}{2}\varepsilon}.$$

As long as the trajectory is in area D it moves on the line $q_2 = \frac{\frac{9}{20}+255\frac{3}{8}\varepsilon}{\frac{4}{5}+79\varepsilon}(1-q_1)$. At the end of the trajectory we are back on line segment a, so at that moment

$$q_1 = \frac{\frac{1}{5} + 301\varepsilon}{1 + 380\varepsilon} \quad \text{and} \quad q_2 = \frac{\frac{9}{20} + 255\frac{3}{8}\varepsilon}{1 + 380\varepsilon}.$$

Hence the function h is as follows:

$$h\left(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon\right) = \left(\frac{\frac{1}{5} + 301\varepsilon}{1 + 380\varepsilon}, \frac{\frac{9}{20} + 255\frac{3}{8}\varepsilon}{1 + 380\varepsilon}\right).$$

We have $h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) = (\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon)$ if and only if $\varepsilon = 0$ or $\varepsilon = \frac{56}{95}$. Therefore there are precisely 2 orbits with starting points $(\frac{1}{5}, \frac{9}{20})$, which is the equilibrium point, and $(\frac{15}{19}, \frac{51}{76})$.

For all $\varepsilon \in (0, \frac{56}{95})$ we have that $(\frac{15}{19}, \frac{51}{76}) > h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) > (\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon)$ and for all $\varepsilon \in (\frac{56}{95}, \frac{4}{5}]$ we have that $(\frac{15}{19}, \frac{51}{76}) < h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) < (\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon)$. Hence the orbit starting at the point $(\frac{15}{19}, \frac{51}{76})$ is stable and the equilibrium point by itself is a non-stable orbit. □

Now we need a few notations and definitions. Let e^* be the equilibrium point: $e^* = (\frac{1}{5}, \frac{9}{20})$. In view of Lemma 4 there is a stable orbit t^* with starting point $(\frac{15}{19}, \frac{51}{76})$. For each $x, y \in [0, 1]^2$ let $l[x, y]$ denote the line segment starting at x and finishing at y. Let $t^*(X)$ be the part of t^*, which is in part X in Fig. 2 for each $X \in \{A, B, C, D\}$ and let t^{*l} be the point where t^* and line segment $l \in \{a, b, c, d\}$ intersect in Fig. 2. Notice that $t^*(A) = l[t^{*a}, t^{*b}] \subset l[t^{*a}, (0, 0)]$; similarly for the regions B, C, D. We say that a point $x \in [0, 1]^2$ is *outside* t^* if $l[x, e^*] \cap t^* \neq \emptyset$.

Lemma 3 *$g(n)$ is outside t^* for each n.*

Proof Since $g(1) = (1, 1)$ is outside t^*, it is sufficient to show that if $g(n)$ is outside t^*, then $g(n+1)$ is outside t^*. Suppose $g(n)$ is outside t^*. Suppose also that $g(n) \in A$. For the other areas similar proofs can be given. Notice that by Lemma 2, if the fictitious play process is in area A, then the current run can only be $[(0, 0), (1, 0)]$ or $[(0, 0), (0, 0)]$. Consequently either

$$g(n+1) = \frac{n}{n+1}g(n) + \frac{1}{n+1}(0, 0) \quad \text{or} \quad g(n+1) = \frac{n}{n+1}g(n) + \frac{1}{n+1}(1, 0),$$

so g can only move towards $(0, 0)$ or $(1, 0)$. In the latter case $g(n+1)$ is clearly outside t^*, while in the former case $g(n+1) \in l[(0,0), g(n)]$.

Suppose first that $g(n+1) \in A$. Observe that $l[(0,0), g(n)]$ and $l[(0,0), t^{*a}]$ intersect in a single point, namely $(0,0)$, which is not in A. Also observe that $t^*(A) \subset l[(0,0), t^{*a}]$, while both $g(n)$ and t^{*a} are in A. Hence, we have that $g(n+1)$ is outside t^*. Secondly, if $g(n+1) \in B$, then notice that $(\frac{1}{14}, 0) = (l[(0,0), (1,0)] \cap b) \in B$ and $g(n+1) \in ConvHull\{(0,0), (\frac{1}{14}, 0), t^{*b}\}$. The latter set is completely in B and each of the extreme points is outside t^*. Hence $g(n+1)$ is outside t^*. □

Proof of main theorem According to Lemma 2 the different runs follow each other cyclically. This means that if the fictitious play process converges, then it must converge to the unique common point of the areas in Fig. 1, which is the equilibrium point. However, according to Lemma 5 the fictitious play process is always outside the stable orbit t^*. Therefore it cannot converge at all. □

4 Concluding remarks

In general the best reply structure in stochastic games is non-linear. In the model examined above it is the single-controller condition that guarantees the linearity. Moreover, we also had the additional structures of irreducibility and state independent transitions. Even so, the fictitious play process does not converge. An interesting question would be to find payoff and/or transition structure that do imply the fictitious play property. The question of convergence of the process for zero-sum stochastic games is open for future research.

References

Brown GW (1951) Iterative solution of games by fictitious play. In: Koopmans TC (ed) Activity analysis of production and allocation. Wiley, New York, pp 374–376

Filar JA (1981) Ordered field property for stochastic games when the player who controls transitions changes from state to state. J Opt Theory Appl 34:503–515

Gillette D (1957) Stochastic games with zero stop probabilities. In: Dresher M, Tucker AW, Wolfe P (eds) Contributions to the theory of games III, annals of mathematical studies, vol 39. Princeton University Press, Princeton, pp 179–187

Hordijk A, Vrieze OJ, Wanrooij GL (1983) Semi-markov strategies in stochastic games. Int J Game Theory 12:81–89

Krishna V, Sjöström T (1998) On the convergence of fictitious play. Math Oper Res 23:479–511

Metrick A, Polak B (1994) Fictitious play in 2×2 games: a geometric proof of convergence. Econ Theory 4:923–933

Miyasawa K (1961) On the convergence of the learning process in 2×2 non-zero-sum two-person games. Res Mem no 33, Economic Research Program. Princeton University, Princeton

Monderer D, Shapley LS (1996) Fictitious play property for games with identical interests. J Econ Theory 68:258–265

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

Rogers PD (1969) Non-zerosum stochastic games. PhD thesis, Report ORC 69–8, Operations Research Center, University of California, Berkeley

Sela A (2000) Fictitious play in 2×3-games. Games Econ Behav 31:152–162

Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39:1095–1100

Shapley LS (1964) Some topics in two-person games. In: Dresher M, Shapley LS, Tucker AW (eds) Advances in game theory. Princeton University Press, pp 1–28

Sobel MJ (1971) Noncooperative stochastic games. Ann Math Stat 42:1930–1935
Vieille N (2000a) Two-player stochastic games I: a reduction. Isr J Math 119:55–91
Vieille N (2000b) Two-player stochastic games II: the case of recursive games. Isr J Math 119:93–126
Vrieze OJ, Tijs SH (1982) Fictitious play applied to sequences of games and discounted stochastic games. Int J Game Theory 11:71–85