Average-discounted equilibria in stochastic games

ELSEVIER

European Journal of Operational Research 112 (1999) 187–195

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Average-discounted equilibria in stochastic games

J. Flesch *, F. Thuijsman, O.J. Vrieze

Department of Mathematics, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Received 14 May 1997; accepted 13 October 1997

Abstract

In stochastic games with finite state and action spaces, we examine existence of equilibria where player 1 uses the limiting average reward and player 2 a discounted reward for the evaluations of the respective payoff sequences. By the nature of these rewards, the far future determines player 1's reward, while player 2 is rather interested in the near future. This gives rise to a natural cooperation between the players along the course of the play. First we show the existence of stationary ε-equilibria, for all $\varepsilon > 0$, in these games. However, besides these stationary ε-equilibria, there also exist ε-equilibria, in terms of only slightly more complex ultimately stationary strategies, which are rather in the spirit of these games because, after a large stage when the discounted game is not interesting any longer, the players cooperate to guarantee the highest feasible reward to player 1. Moreover, we analyze an interesting example demonstrating that 0-equilibria do not necessarily exist in these games, not even in terms of history dependent strategies. Finally, we examine special classes of stochastic games with specific conditions on the transition and payoff structures. Several examples are given to clarify all these issues.

Keywords: Game theory; Stochastic games; Equilibria; Discounted reward; Limiting average reward

1. Introduction

We deal with stochastic games with finite state and action spaces. Such games can be seen as Markov decision processes with more controllers, called players. Formally, such a game Γ can be given by a tuple $\langle S, \{I_s\colon s \in S\}, \{J_s\colon s \in S\}, r^1, r^2, p\rangle$, where S is a non-empty finite set of states, I_s and J_s are respective non-empty finite sets of (pure) actions for players 1 and 2 in state s, r^1 and r^2 are respective payoff functions that assign payoffs $r^1(s, i_s, j_s)$, $r^2(s, i_s, j_s)$ to any action pair (i_s, j_s) in any state s, and p is the transition probability map assigning a probability vector $(p(t \mid s, i_s, j_s))_{t \in S}$ to any action pair (i_s, j_s) in any state s.

The game is to be played at stages in $\mathbb{N}$ in the following way. The play starts at stage 1 in an initial state, where, simultaneously and independently, both players are to choose an action. These choices induce immediate payoffs to both players given by the respective payoff functions, and next, the play moves to a new state according to the corresponding transition probability vector. In the new state, at stage 2, new actions are to be chosen

* Corresponding author. Tel.: +31 43 388 3494; fax: +31 43 321 1889.

0377-2217/99/$ – see front matter
PII S0377-2217(97)00384-6

by the players. Then the players receive the corresponding payoffs given by the payoff functions, and afterwards the play moves to some new state according to the corresponding transition probability vector again, and so on. The players are assumed to have complete information and perfect recall.

A mixed action for a player in state s is a probability distribution on the set of his actions in state s. Mixed actions in state s will be denoted by x_s for player 1 and by y_s for player 2, and the sets of mixed actions in state s by X_s and Y_s, respectively. A strategy is a decision rule that prescribes a mixed action in the current state for any past history of the play. Such general strategies, so-called history dependent strategies, will be denoted by π for player 1 and by σ for player 2. If for all histories, the mixed actions prescribed by a strategy only depend on the current stage and state then the strategy is called Markov, while if they only depend on the current state then the strategy is called stationary. Thus the respective stationary strategy spaces are $X := \times_{s\in S} X_s$ and $Y := \times_{s\in S} Y_s$; while the respective Markov strategy spaces are $F := \times_{n\in\mathbb{N}} X$ and $G := \times_{n\in\mathbb{N}} Y$. We will use the notations x and y for stationary strategies and f and g for Markov strategies of the respective players. A stationary strategy is called pure, if it prescribes one pure action to be used for each state. Thus the respective spaces of pure stationary strategies are simply $I := \times_{s\in S} I_s$, $J := \times_{s\in S} J_s$. Pure stationary strategies will be denoted by i and j.

A strategy pair (π, σ) together with an initial state s determines a stochastic process on the payoffs. The sequences of payoffs need to be evaluated in some manner. The limiting average reward evaluates them by the long-term average payoffs, given for player $k \in \{1, 2\}$ by

$$\gamma^k(s, \pi, \sigma) := \liminf_{N\to\infty} \mathbb{E}_{s\pi\sigma}\left(\frac{1}{N}\sum_{n=1}^{N} r_n^k\right),$$

where r_n^k denotes the random variable for the payoff of player k at stage n. Hence the limiting average reward places its emphasis on the far future payoffs. Another widely used evaluation is the β-discounted reward, $\beta \in (0, 1)$, which is given for player $k \in \{1, 2\}$ by

$$\gamma_\beta^k(s, \pi, \sigma) := \mathbb{E}_{s\pi\sigma}\left((1-\beta)\sum_{n=1}^{\infty} \beta^{n-1} r_n^k\right).$$

In contrast with the limiting average reward, the β-discounted reward is obviously rather determined by the near future payoffs.

A strategy pair (π, σ) is called an ε-equilibrium, $\varepsilon \geqslant 0$, with respect to (ψ^1, ψ^2), where both ψ^1 and ψ^2 are one of the above rewards, if for all $s \in S$, for all $\bar\pi$ and $\bar\sigma$

$$\psi^1(s, \bar\pi, \sigma) \leqslant \psi^1(s, \pi, \sigma) + \varepsilon \quad \text{and}$$
$$\psi^2(s, \pi, \bar\sigma) \leqslant \psi^2(s, \pi, \sigma) + \varepsilon,$$

which means that for every initial state $s \in S$, neither player can gain more than ε with respect to his own reward function by a unilateral deviation. If both players use the limiting average reward then we speak of limiting average equilibria, while when the β-discounted rewards are used then we speak of β-discounted equilibria.

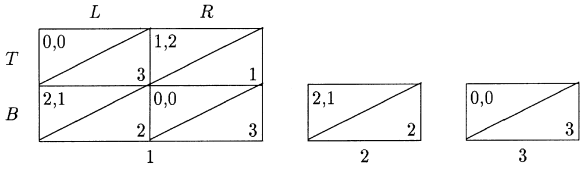
Fink (1964) and Takahashi (1964) showed that β-discounted 0-equilibria always exist in terms of stationary strategies. The structure of limiting average equilibria is, however, substantially more complex and the question of existence of limiting average ε-equilibria has not yet been answered. The famous zerosum game introduced by Gillette (1957), the Big Match, which was examined by Blackwell and Ferguson (1968), and the non-zerosum game in Sorin (1986) demonstrate that, in general, limiting average 0-equilibria do not necessarily exist and history dependent strategies are indispensable for establishing limiting average ε-equilibria.

In this paper we investigate existence of ε-equilibria in games where the players use different evaluations. We assume that player 1 uses the limiting average reward, while player 2 is interested in his β -discounted reward. We will call these games average-discounted games. By the nature of these rewards, as discussed above, the players are interested in different time periods of the play, which may lead to a natural cooperation between the players. First we show the existence of stationary ε-equilibria, for all $\varepsilon > 0$, in these games. So stationary strategies are not only sufficient for establishing equilibria in classical dis-

counted games but also in these average-discounted games. The existence of equilibria in terms of stationary strategies is appealing, since stationary strategies are rather simple strategies. On the other hand, however, these stationary equilibria have the drawback that they do not make use of the special nature of these games, they do not use that different time periods interest the players. Therefore we also prove the existence of ε-equilibria, where, after a large stage when the discounted game is not interesting any longer, the players cooperate to guarantee the highest feasible reward to player 1. These ε-equilibria are formed by only slightly more complex Markov strategies, which we call "ultimately stationary" (after finitely many stages stationary strategies are played forever). Next, we analyze an interesting example demonstrating that 0-equilibria do not always exist in these average-discounted games, not even in terms of history dependent strategies. Finally, we examine special classes of stochastic games, where specific conditions are imposed on the transition and payoff structures.

We now briefly discuss the following game to clarify the issues.

Example 1.1.



Here matrices represent the states of the game. The actions of player 1 are rows and the actions of player 2 are columns. In each entry, the corresponding payoffs are placed in the up-left corner, while the transition is placed in the bottom-right corner. In this game each transition is represented by the number of the state to which transition should occur with probability 1. Notice that the only interesting state is state 1, since states 2 and 3 are absorbing, i.e., once the play visits either of these states it stays there forever. Hence we assume the initial state to be state 1. Obviously, strategies only need to be defined for state 1.

Take an arbitrary discount factor $\beta \in (0,1)$. There are two really simple stationary equilibria with respect to $(\gamma^1, \gamma^2_\beta)$. One of them is playing entry (B,L) at stage 1, yielding absorption in state 2 and reward $(2,1)$, and the other one is to play entry (T,R) at each stage, which gives reward $(1,2)$. These stationary equilibria, however, are not really in the spirit of the game. The players could also decide to play entry (T,R) sufficiently long so that player 2's reward, which is rather determined by the near future payoffs, becomes almost 2, and then, when the rest of the play does not really interest player 2 any longer, to play entry (B,L) so as to give player 1 the highest feasible payoff 2 at each further stage. This plan, yielding a reward close to $(2,2)$, can be realized by ultimately stationary strategies (after finitely many stages stationary strategies are played forever). Note that rewards close to $(2,2)$ cannot be guaranteed by stationary ε-equilibria, with small $\varepsilon \geqslant 0$.

2. Stationary ε-equilibria

This section is devoted to the analysis of the existence of stationary ε-equilibria, $\varepsilon > 0$, in these average-discounted games. First we introduce a restricted strategy space for player 2. Let

$$\bar{\delta} := \min_{s \in S} \frac{1}{|J_s|}.$$

For $\delta \in [0, \bar{\delta}]$ let

$$Y(\delta) := \{y \in Y \,|\, y_s(j_s) \geqslant \delta \quad \forall s \in S,\ \forall j_s \in J_s\}$$

in words, $Y(\delta)$ is the set of stationary strategies of player 2 which use each action in each state with probability at least δ. Obviously, $Y(\delta)$ is a polytope, and by the choice of $\bar{\delta}$ it is non-empty. The following lemma states some well known properties of the rewards and sets of best reply strategies.

Lemma 2.1. *(i) The function $\gamma^1(s,x,\cdot)$ is continuous on $Y(\delta)$ for any $s \in S, x \in X, \delta \in (0,\bar{\delta}]$. (ii) Let $\bar{y} \in Y$. Then the set*

$$\mathscr{B}^1(\bar{y}) := \{x \in X \,|\, \gamma^1(s,x,\bar{y}) \geqslant \gamma^1(s,\hat{x},\bar{y}) \quad \forall s \in S,\ \forall \hat{x} \in X\}$$

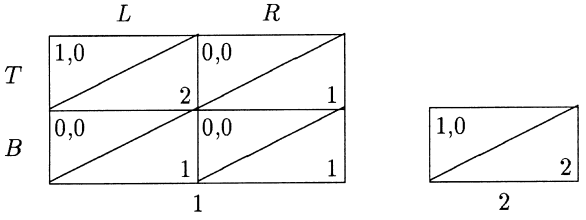
is non-empty and convex. (iii) The function $\gamma_\beta^2(s,\cdot,\cdot)$ *is continuous on* $X \times Y$ *for any* $s \in S$, $\beta \in (0,1)$. *(iv) Let* $\beta \in (0,1)$, $\bar{x} \in X$, $\delta \in [0,\bar{\delta}]$. *Then the set*

$$\mathscr{B}_\beta^2(\delta,\bar{x}) := \{y \in Y(\delta) \mid \gamma_\beta^2(s,\bar{x},y) \geqslant \gamma_\beta^2(s,\bar{x},\hat{y}) \quad \forall s \in S,\ \forall \hat{y} \in Y(\delta)\}$$

is non-empty, convex, and closed.

Note that (ii) and (iv) also show that against a fixed stationary strategy there exist stationary best replies for the other player. Notice that the above properties are weaker for the limiting average reward, which is clarified by the following example.

Example 2.2.



The notation is the same as in example 1.1. Since state 2 is trivial, stationary strategies for the players are fully determined by the mixed actions in state 1. Suppose that the initial state is state 1. Let $y^n := (1/n,(n-1)/n) \in Y$, and let $y := \lim_{n\to\infty} y^n = (0,1)$. For $x = (1,0)$ we have $\gamma^1(1,x,y^n) = 1$ while $\gamma^1(1,x,y) = 0$, so $\gamma^1(1,x,\cdot)$ is not continuous on Y; nevertheless on $Y(\delta)$, with $\delta > 0$, it is continuous (cf. (i) of Lemma 2.1) due to the fact that the ergodic structure of the Markov chain with respect to (x,y) is the same for any $y \in Y(\delta)$. One can readily verify that, for $\bar{y} = (\frac{1}{2},\frac{1}{2})$, we have $\mathscr{B}^1(\bar{y}) = \{(\lambda,1-\lambda) \mid \lambda \in (0,1]\}$, so $\mathscr{B}^1(\bar{y})$ is not closed (cf. (ii) of Lemma 2.1).

The main result of this section is the following theorem.

Theorem 2.3. *In any stochastic game, for any* $\varepsilon > 0$, *there exists a stationary* ε*-equilibrium with respect to* $(\gamma^1,\gamma_\beta^2)$, *where* $\beta \in (0,1)$.

Proof. Take arbitrary $\varepsilon > 0$ and $\beta \in (0,1)$. For a strategy $y \in Y$ let $\overline{\mathscr{B}}^1(y)$ denote the closure of $\mathscr{B}^1(y)$. By (iii) in Lemma 2.1, the function $\gamma_\beta^2(s,\cdot,\cdot)$ is continuous on the compact space $X \times Y$, for any $s \in S$, hence it is uniformly continuous as well. Therefore there exists a $\delta \in (0,\bar{\delta}]$ such that for all $s \in S$ we have

$$\sup_{x\in X}\left[\sup_{y\in Y}\gamma_\beta^2(s,x,y) - \sup_{y\in Y(\delta)}\gamma_\beta^2(s,x,y)\right] \leqslant \frac{\varepsilon}{2}. \tag{1}$$

Now consider the following set-valued map:

$$\Psi\colon (x,y) \in X \times Y(\delta) \mapsto \overline{\mathscr{B}}^1(y) \times \mathscr{B}_\beta^2(\delta,x) \subset X \times Y(\delta).$$

The set $X \times Y(\delta)$ is convex and compact, and, in view of Lemma 2.1, this correspondence Ψ is non-empty, convex, compact valued and upper semi-continuous. Hence the conditions of Kakutani's fixed point theorem (cf. Kakutani, 1941) are satisfied. Therefore Ψ has a fixed point, i.e., there exists a pair $(x,y) \in X \times Y(\delta)$ such that $(x,y) \in (\overline{\mathscr{B}}^1(y),\mathscr{B}_\beta^2(\delta,x))$.

Using this fixed point (x,y) we construct a stationary ε-equilibrium in the game. Since $x \in \overline{\mathscr{B}}^1(y)$ and $y \in \mathscr{B}_\beta^2(\delta,x)$, by the uniform continuity of $\gamma_\beta^2(s,\cdot,\cdot)$ on $X \times Y$ for all $s \in S$, there exists an $x' \in \mathscr{B}^1(y)$ such that for all $s \in S$ all the following inequalities (and equality) hold:

$$\gamma_\beta^2(s,x',y) + \frac{1}{2}\varepsilon \geqslant \gamma_\beta^2(s,x,y) + \frac{1}{4}\varepsilon = \sup_{\bar{y}\in Y(\delta)}\gamma_\beta^2(s,x,\bar{y}) + \frac{1}{4}\varepsilon \geqslant \sup_{\bar{y}\in Y(\delta)}\gamma_\beta^2(s,x',\bar{y}). \tag{2}$$

We show that (x',y) is an ε-equilibrium. Recall that, as discussed above, against a stationary strategy there always exist best replies in stationary strategies. Hence using $x' \in \mathscr{B}^1(y)$ we have for all $s \in S$ that

$$\gamma^1(s,\pi,y) \leqslant \sup_{\bar{x}\in X}\gamma^1(s,\bar{x},y) = \gamma^1(s,x',y) \quad \forall \pi.$$

For player 2, applying Eqs. (1) and (2), we obtain for all $s \in S$ that

$$\gamma_\beta^2(s,x',\sigma) \leqslant \sup_{\bar{y}\in Y}\gamma_\beta^2(s,x',\bar{y})$$

$$\leqslant \sup_{\bar{y}\in Y(\delta)}\gamma_\beta^2(s,x',\bar{y}) + \frac{1}{2}\varepsilon \leqslant \gamma_\beta^2(s,x',y) + \varepsilon \quad \forall\sigma.$$

Therefore (x',y) is an ε-equilibrium indeed. □

3. Ultimately stationary ε-equilibria

In the previous section we showed the existence of stationary ε-equilibria for all $\varepsilon > 0$, with respect to $(\gamma^1,\gamma_\beta^2)$. These stationary ε-equilibria are appealing, because simple strategies are used. In this section, however, we also prove the existence of ε-equilibria in terms of ultimately stationary strategies (after finitely many stages stationary strategies are played forever), where the players naturally cooperate, based on the different nature of their rewards. The idea is that after a large stage N player 2 becomes uninterested in the game due to the large powers of the discount factor β, so after stage N the players can cooperate to guarantee the highest feasible reward for player 1 in the future. During the first N stages, obviously, player 1 has to be careful not to decrease his future perspectives after stage N.

Theorem 3.1. *In any stochastic game, for any $\varepsilon > 0$, there exist ε-equilibria with respect to $(\gamma^1,\gamma_\beta^2)$, where $\beta \in (0,1)$, such that up to some stage N the players play Markov strategies, from stage $N+1$ they play stationary strategies, and if the play is at stage $N+1$ in state s, then player 1 receives $\rho_s := \sup_{\pi,\sigma}\gamma^1(s,\pi,\sigma)$.*

Proof. Consider a stochastic game Γ. Take arbitrary $\varepsilon > 0$ and $\beta \in (0,1)$. Let $N \in \mathbb{N}$ be so large that

$$\beta^{N+1}\left[\max_{s,i_s,j_s} r^2(s,i_s,j_s) - \min_{s,i_s,j_s} r^2(s,i_s,j_s)\right] \leqslant \varepsilon,$$

so after stage $N+1$ player 2 can only improve his β-discounted reward by at most ε. In the theory of Markov decision problems it is known that there exist pure stationary strategies $i^* \in I$ and $j^* \in J$ such that for ρ_s as defined in the theorem it holds that

$$\rho_s = \gamma^1(s,i^*,j^*) \quad \forall s \in S.$$

Consider the game Γ^N which is played up to stage N and in which player 1 maximizes the value of ρ in the final state and player 2 maximizes his N-stage β-discounted reward, given for N-stage strategies π^N,σ^N and initial state s by

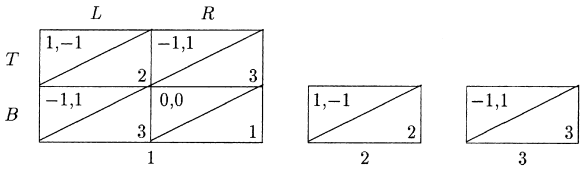
$$\mathbb{E}_{s\pi^N\sigma^N}\left((1-\beta)\sum_{n=1}^{N}\beta^{n-1}r_n^2\right).$$

Using backwards induction, one can construct an N-stage Markov 0-equilibrium (f^N,g^N) in the game Γ^N. Let f denote the Markov strategy which coincides with f^N for the first N stages and which prescribes the pure stationary strategy i^* afterwards. The definition of g is analogous. Thus by their definitions, f and g satisfy the requirements of the theorem. We only have to show that (f,g) is an ε-equilibrium. Observe that player 1's limiting average reward γ^1 is completely determined by the value of ρ in the state at stage $N+1$, which is exactly what he maximizes during the first N stages, so player 1 cannot improve at all. Player 2 can only improve his reward by ε after stage N, because of the choice of N; while during the first N stages, by his reward function in the game Γ^N, he cannot improve it at all. So (f,g) is an ε-equilibrium indeed. □

4. A game without average-discounted 0-equilibria

In Sections 2 and 3 we showed the existence of ε-equilibria, for all $\varepsilon > 0$, in terms of stationary and ultimately stationary strategies. The interesting example 3 will demonstrate that, in these average-discounted games, 0-equilibria do not always exist, not even in history dependent strategies. So as it might be expected, the solutions of average-discounted games are on the one hand more complex than that of discounted games, where stationary 0-equilibria always exist, but on the other hand simpler than that of limiting average games, where stationary ε-equilibria do not generally exist for small $\varepsilon \geqslant 0$.

Example 4.1 (*game Γ_3*).



The notation is the same as in example 1.1. We show that 0-equilibria do not exist for initial state 1.

Theorem 4.2. *In the game Γ_3, there exist no 0-equilibria for initial state 1 with respect to $(\gamma^1, \gamma^2_\beta)$ for any $\beta \in (0,1)$.*

Proof. As states 2 and 3 are trivial, strategies only need to be defined for histories where the initial state is 1 and no absorption has occurred. Notice that the only information carried by these histories is the current stage. Therefore all history dependent strategies are simply Markov strategies. Since any mixed action in state 1 can be represented by the probability assigned to the first action, any Markov strategy for any player is an element of the set $\times_{n=1}^{\infty}[0,1]$.

Suppose by way of contradiction that $(f,g) = (f(n), g(n))_{n=1}^{\infty}$ is a Markov 0-equilibrium with respect to $(\gamma^1, \gamma^2_\beta)$, where $\beta \in (0,1)$; here $f(n)$ and $g(n)$ denote the probabilities of playing action T and L, respectively, at stage n. Let $f^k := (f(n))_{n=k}^{\infty}$ and $g^k := (g(n))_{n=k}^{\infty}$ for any $k \in \mathbb{N}$, so f^k and g^k are the Markov strategies f and g starting from stage k. Let ξ^k denote player 1's limiting average reward when using (f^k, g^k) for initial state 1.

Based on the assumption that (f,g) is a 0-equilibrium, we subsequently derive that we should have:

(1) $\xi^1 > -1$;
(2) $0 < f(1) < 1$ and $0 < g(1) < 1$;
(3) (f^n, g^n) is a 0-equilibrium, $0 < f(n) < 1$, and $0 < g(n) < 1$ for all $n \in \mathbb{N}$;
(4) $\xi^n < \xi^{n+1}$ and $g(n) < g(n+1)$ for all $n \in \mathbb{N}$.

Next we show that these properties lead to a contradiction.

Proof of (1). Since (f,g) is a 0-equilibrium, it suffices to define a strategy $\bar{f}$ for player 1 which guarantees a reward larger than -1 when playing against g. For $n \in \mathbb{N}$ let

$$\bar{f}(n) := \begin{cases} 1 & \text{if } g(n) > 0 \\ 0 & \text{if } g(n) = 0. \end{cases}$$

Now with respect to $(\bar{f}, g)$, whenever the play is in state 1, either the cell (B,R) is played with probability 1 or the cell (T,L) is played with a positive probability, hence $\gamma^1(1, \bar{f}, g) > -1$.

Proof of (2). If $f(1) = 1$ then $g(1) = 0$, since it yields absorption in entry (T,R) giving the highest possible reward 1 for player 2. However, this contradicts $\xi^1 > -1$ (cf. (1)), hence $f(1) < 1$ must hold. If $f(1) = 0$ then $g(1) = 1$, which also contradicts $\xi^1 > -1$; hence $f(1) > 0$.

If $g(1) = 1$ then $f(1) = 1$ has to hold because f is a best reply against g, which contradicts $0 < f(1) < 1$. Hence $g(1) < 1$. Now suppose that $g(1) = 0$. Using (1) we have

$$-1 < \xi^1 = f(1)(-1) + (1 - f(1))\xi^2,$$

thus by $f(1) > 0$ we obtain $\xi^2 > \xi^1$, which means that player 1 would be better off by playing action B at stage 1 and playing f^2 from stage 2 on assuring reward ξ^2. This is in contradiction with the fact that $f(1) > 0$. Hence $g(1) > 0$ must hold.

Proof of (3). By (2), the probability of no absorption at stage 1 has a positive probability, therefore, clearly, (f^2, g^2) must be a 0-equilibrium as well. Using that (f^2, g^2) is a 0-equilibrium, one can show similarly that $0 < f(2) < 1$ and $0 < g(2) < 1$. Now repeating this argument yields the statement.

Proof of (4). The strategy f^1 is a best reply against g^1 and player 1 plays action B with a positive probability at stage 1 (cf. (3)), hence

$$\xi^1 = g(1)(-1) + (1 - g(1))\xi^2.$$

Now using (1) and $g(1) > 0$ (cf. (3)), we have $\xi^1 < \xi^2$ indeed. Repeating this argument leads to $\xi^n < \xi^{n+1}$ for all $n \in \mathbb{N}$.

At stages n and $n+1$, in view of (3), player 1 plays action T with positive probabilities, thus

$$\xi^n = g(n)\cdot 1 + (1-g(n))(-1),$$
$$\xi^{n+1} = g(n+1)\cdot 1 + (1-g(n+1))(-1).$$

Now from $\xi^n < \xi^{n+1}$ it follows that $g(n) < g(n+1)$.

Deriving a contradiction. Consider the strategy $\bar{f}_K$, $K \geqslant 2$, which prescribes action B up to stage $K-1$ and the strategy f^K from stage K on. Then with respect to $(\bar{f}_K, g)$, player 1's reward is -1 if absorption occurs during the first $K-1$ stages and equals ξ^K otherwise. Thus we have for all $K \geqslant 2$ that

$$\begin{aligned}
\gamma^1(\bar{f}_K, g) &= \left[1 - \prod_{n=1}^{K-1}(1-g(n))\right](-1) \\
&\quad + \left[\prod_{n=1}^{K-1}(1-g(n))\right]\xi^K \\
&= \left[1 - \prod_{n=1}^{K-2}(1-g(n))\right](-1) \\
&\quad + \left[\prod_{n=1}^{K-2}(1-g(n))\right]\big[g(K-1)(-1) \\
&\qquad (1-g(K-1))\xi^K\big] \\
&= \left[1 - \prod_{n=1}^{K-2}(1-g(n))\right](-1) \\
&\quad + \left[\prod_{n=1}^{K-2}(1-g(n))\right]\xi^{K-1} \\
&= \cdots \\
&= g(1)(-1) + (1-g(1))\xi^2 \\
&= \xi^1.
\end{aligned}$$

However, by properties 2 and 4 we have that $g(n) > g(1) > 0$ for all $n \in \mathbb{N}$. Therefore, if player 1 uses $\bar{f}_K$ with a large K then absorption occurs in entry (B,L) during the first $K-1$ stages with probability almost 1. Formally,

$$\lim_{K\to\infty}\left[1 - \prod_{n=1}^{K-1}(1-g(n))\right] = 1,$$

thus

$$\xi^1 = \lim_{K\to\infty}\gamma^1(\bar{f}_K, g) = -1,$$

which contradicts (1). Hence the basic assumption that (f,g) is a 0-equilibrium is false. □

5. Special classes of stochastic games

This section is devoted to the study of average-discounted equilibria in special classes of games. We briefly treat several classes of games in which (ε-)equilibria can be achieved by using other techniques.

Unichain games. A stochastic game is called unichain if, for any stationary strategy pair, there is just one ergodic set of states. This condition assures that the limiting average reward $\gamma^1(s,\cdot,\cdot)$ is also continuous on $X \times Y$ for all $s \in S$ and that the best reply sets $\mathscr{B}^1(\bar{y})$, $\bar{y} \in Y$, are closed (cf. (i) and (ii) in Lemma 2.1). In these games one can establish stationary 0-equilibria with respect to $(\gamma^1, \gamma^2_\beta)$ by simply applying Kakutani's fixed point theorem on $X \times Y$.

Perfect information, switching control and ARAT games. A stochastic game has perfect information if, in each state, one of the players has only one action available. A stochastic game with switching control is a stochastic game with the property that, in each state s, the transition probabilities only depend on the actions of one of the players, i.e., either $p(s,i_s,j_s) = p(s,i_s)$ for all i_s, j_s or $p(s,i_s,j_s) = p(s,j_s)$ for all i_s, j_s. Finally, a stochastic game is called an ARAT game (additive reward and additive transition structure) if it has the following property: for each pair of actions (i_s,j_s) in each state s, the payoffs $r^k(s,i_s,j_s)$, $k=1,2$, can be decomposed as $r^k(s,i_s,j_s) = r^k_1(s,i_s) + r^k_2(s,j_s)$ while, similarly, the transition probability $p(s,i_s,j_s)$ can be decomposed as $p(s,i_s,j_s) = p_1(s,i_s) + p_2(s,j_s)$. So by the definitions, perfect information stochastic games have switching control as well as ARAT structure.

In perfect information games and ARAT games one can establish average-discounted 0-equilibria, almost analogously as in the proof for the existence of limiting average 0-equilibria for these games in Thuijsman and Raghavan (1997). The idea is that player 1 has to play a pure stationary limiting average optimal strategy i, i.e.

$\inf_\sigma \gamma^1(s,i,\sigma) = \sup_\pi \inf_\sigma \gamma^1(s,\pi,\sigma)$ for all $s \in S$, (pure stationary limiting average optimal strategies always exist in these games, cf. Liggett and Lippman, 1969; Raghavan et al., 1985), and player 2 has to play a stationary β-discounted best reply y against the strategy i. This already implies that player 2 does not have a profitable deviation against i. Notice that, since the strategy i prescribes one pure action for each state, player 2 can immediately detect any deviation of player 1. Now in order to eliminate the profitability of deviations of player 1, if player 2 detects a deviation from i then he has to punish player 1 by switching to a strategy σ satisfying $\gamma^1(s,\pi,\sigma) \leqslant \gamma^1(s,i,y) + \delta$ for all s and π, where $\delta > 0$ is sufficiently small. Note that these punishments are effective due to the transition structure of these games.

In switching control stochastic games the proof is somewhat more complicated, because player 1 does not need to have pure stationary limiting average optimal strategies. Nevertheless, Filar (1981) showed that there exist stationary limiting average optimal strategies for player 1. Now the main difference is that player 2 cannot immediately detect deviations of player 1, but, as shown in Thuijsman and Raghavan (1997), player 2 can conduct statistical tests on the action frequencies of player 1, and by doing so he can detect deviations in the long run with probability almost 1. This way we obtain average-discounted ε-equilibria, for all $\varepsilon > 0$, for switching control games as well.

Repeated games with absorbing states. These are stochastic games where all the states but one are absorbing. Here one can establish average-discounted ε-equilibria, for all $\varepsilon > 0$, as follows. For any $\alpha \in (0,1)$, there exists a stationary equilibrium $(x^{\alpha\beta}, y^{\alpha\beta})$ with respect to $(\gamma^1_\alpha, \gamma^2_\beta)$ (cf. Fink, 1964; Takahashi, 1964); this game is in fact a discounted game with two different discount factors. Using techniques as in Vrieze and Thuijsman (1989) one can show that either $(x^{1\beta}, y^{1\beta})$ or $(x^{1\beta}, y^{\alpha\beta})$ with a large α can be supplemented with history dependent "punishment" strategies to establish an ε-equilibrium with respect to $(\gamma^1, \gamma^2_\beta)$; here $(x^{1\beta}, y^{1\beta})$ is the limit strategy pair of some sequence $(x^{\alpha_n\beta}, y^{\alpha_n\beta})$, $n \in \mathbb{N}$, with $\alpha_n \uparrow 1$.

6. Concluding remarks

In this article we have studied average-discounted games. Here in contrast with classical discounted or limiting average games, the players use different evaluations for their payoff sequences. These games and their solutions turn out to be more complex than discounted games, however, they still have a substantially simpler structure than limiting average games.

We also wish to remark that, in the literature of stochastic games and Markov decision processes, games have already been studied where, instead of using the discounted or the limiting average evaluation, the players (or the player) use convex combinations of several discounted rewards with different discount factors and the limiting average reward (cf. for example Filar and Vrieze, 1992; Feinberg and Shwartz, 1994, 1995). Although the ideas have something in common, the arising problems require a different analysis.

References

Blackwell, D., Ferguson, T.S., 1968. The big match. Annals of Mathematical Statistics 33, 159–163.

Feinberg, E.A., Shwartz, A., 1994. Markov decision models with weighted discounted criteria. Mathematics of Operations Research 19, 152–168.

Feinberg, E.A., Shwartz, A., 1995. Constrained Markov decision models with weighted discounted rewards. Mathematics of Operations Research 20, 302–320.

Filar, J.A., 1981. Ordered field property for stochastic games when the player who controls transitions changes from state to state. Journal of Optimization Theory and Applications 34, 503–515.

Filar, J.A., Vrieze, O.J., 1992. Weighted reward criteria in competitive Markov decision processes. Zeitschrift für Operations Research – Methods and models of operations research 36, 343–358.

Fink, A.M., 1964. Equilibrium in a stochastic *n*-person game. Journal of Science of Hiroshima University Series A-I 28, 89–93.

Gillette, D., 1957. Stochastic games with zero stop probabilities. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games III, Annals of Mathematical Studies, vol. 39. Princeton University Press, pp. 179–187.

Kakutani, S., 1941. A generalization of Brouwer's fixed point theorem. Duke Mathematical Journal 8, 161–167.

Liggett, T.M., Lippman, S.A., 1969. Stochastic games with perfect information and time average payoff. SIAM Review 11, 604–607.

Raghavan, T.E.S., Tijs, S.H., Vrieze, O.J., 1985. On stochastic games with additive reward and transition structure. Journal of Optimization Theory and Applications 47, 451–464.

Sorin, S., 1986. Asymptotic properties of a non-zerosum game. International Journal of Game Theory 15, 101–107.

Takahashi, 1964. Equilibrium points of stochastic noncooperative *n*-person games. Journal of Science of Hiroshima University, Series A-I 28, 95–99.

Thuijsman, F., Raghavan, T.E.S., 1997. Perfect information stochastic games and related classes. International Journal of Game Theory 26, 403–408.

Vrieze, O.J., Thuijsman, F., 1989. On equilibria in repeated games with absorbing states. International Journal of Game Theory 18, 293–310.