ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Stochastic Games : recent results

Nicolas Vieille

Octobre 2001

Cahier n° 2002-003

LABORATOIRE D'ECONOMETRIE

Stochastic Games : recent results

Nicolas Vieille[1]

Octobre 2001

Cahier n° 2002-003

Résumé: Nous présentons des résultats récents sur les jeux stochastiques finis. Ce texte est à paraître dans le Handbook of Game Theory, vol 3., eds. R.J. Aumann et S. Hart

Abstract: This is a survey of recent results on stochastic games, prepared for the le Handbook of Game Theory, vol 3., eds. R.J. Aumann et S. Hart

Mots clés :   Jeux stochastiques

Key Words :   Stochastic games

Classification AMS: 91A06, 91A15

[1] Ecole Polytechnique and Département Finance et Economie, HEC, 78 351 Jouy-en-Josas, France.
e-mail: vieille@hec.fr

Stochastic Games: Recent Results

Nicolas Vieille

February 28, 2001

1 Introduction

The purpose of this chapter is to complement the previous chapter of Mertens (henceforth [M]), and to present selected recent developments of the theory that are not covered in it. We limit ourselves to some of the themes addressed there, without claiming to be exhaustive. In particular, we will not discuss topics such as algorithms or applications of stochastic games.

Our notations follow those of [M], and we refer to Section 1.1 there for a general presentation of stochastic games. Unless otherwise specified, the state space Ω and the action sets S^i, $i \in I$, are assumed finite.

In this finite setup, any undiscounted zero-sum stochastic game has a value (Mertens and Neyman (1981); see also Theorem 5.1 in [M]). In recent years, this fundamental result has been extended in several directions: to non-zero-sum games (do n-player stochastic games have an equilibrium payoff?) and to zero-sum games with imperfect monitoring and/or incomplete information (do such games have a $\max\min$ and a $\min\max$?).

In Section 1, we report on the result that any two-player stochastic game has an equilibrium payoff. Section 2 contains the known results for games with more than two players. These two sections relate to Section 7 in [M]. Section 3 deals with the extension of the theory of zero-sum games to games with imperfect monitoring in the zero-sum case. Finally, Section 4 contains new results on games with large state spaces (see Section 7 of [M]) and lists miscellaneous results.

To facilitate an independent reading of this chapter, we recall now the definition of equilibrium payoffs, and several other notions.

Definition 1 *Let Γ be an N-player game. The vector $d \in \mathsf{R}^N$ is an ε-equilibrium payoff of Γ if there exists a strategy vector σ, and $N_0 \in \mathsf{N}$, such that, for $n \geq N_0$:*

1. $\|\gamma_n(\sigma) - d\| \leq \varepsilon$;

2. σ *is an* ε*-equilibrium of* Γ_n.

In this definition, Γ_n is the n-stage version of Γ. We then say that σ is an ε-equilibrium; d is an equilibrium payoff if it is an ε-equilibrium payoff, for each $\varepsilon > 0$. Given $\varepsilon' > \varepsilon > 0$, an ε-equilibrium is also an ε'-equilibrium in all discounted games, provided the discount factor is close enough to zero. We refer to Sorin (1992) for a discussion of this concept.

A strategy is *stationary* (or stationary Markov) if the distribution used to select an action depends only on the current state of the game. Thus, a stationary strategy x^i of player i can be identified with an element $(x^i_\omega)_\omega$ of $\Delta(S^i)^\Omega$, where x^i_ω is the lottery used by player i in state ω. Stationary strategies will be denoted by Latin letters, and arbitrary strategies by Greek letters. Given a profile x of (stationary) strategies, the sequence $(\omega_n)_{n \in \mathbb{N}}$ of states is a Markov chain. The mixed extensions of the stage payoff g and of the transition probability P will also be denoted by g and P. A *perturbation* of a stationary strategy $x^i = (x^i_\omega)$ is a stationary strategy $\mathbf{e}^i = (\mathbf{e}^i_\omega)$ such that the support of x^i_ω is a subset of the support of $\mathbf{e}^i_\omega$, for each $\omega \in \Omega$.

2 Two-player non-zero-sum games

The purpose of this section is to present the main ideas of the proof of the next theorem, due to Vieille (2000a,2000b). Several tools of potential use for subsequent studies are introduced.

Theorem 2 *Every two-player stochastic game has an equilibrium payoff.*

An overview

W.l.o.g., we assume that the stage payoff function (g^1, g^2) of the game satisfies $g^1 < 0 < g^2$. The basic idea is to devise an ε-equilibrium profile that takes the form of a stationary-like strategy vector σ, supplemented by threats of indefinite punishment.

We give a heuristic description of σ. The profile σ essentially coincides with a stationary profile $\bar{x}$. For the Markov chain defined by $\bar{x}$, consider the partition of the state space Ω into recurrent sets and transient states. (This partition depends on $\bar{x}$, since the transitions depend on actions.) The recurrent sets are classified into *solvable* and *controlled* sets. The solvable sets

are those recurrent sets C for which the average payoff induced by $\bar{x}$ starting from C is high for both players; the controlled sets are the remaining sets. In each controlled set C, σ plays a perturbation of $\bar{x}$, designed so that the play leaves C in finite time. In the other states, σ coincides with $\bar{x}$. Given σ, the play eventually reaches some solvable set (and remains within it). Whenever the play is in a controlled or solvable set, each player monitors the behavior of the other player, using statistical tests.

This description is oversimplified and inaccurate in some fairly important respects, such as the fact that we use a generalized notion of recurrent set, called a *communicating set.*

The construction of σ consists of two independent steps: first, to construct the solvable sets and some controlled sets, and reduce the existence problem to a class of recursive games; second, to deal with the class of recursive games.[1]

Some terminology

Before proceeding with these steps, we shall provide a formal definition of the notions of communicating, solvable and controlled sets.

Definition 3 *A subset C is* communicating *given a profile x if, given any $\omega, \omega' \in C$, there exists a perturbation $\mathcal{e}$ of x for which, starting from ω, the probability of reaching ω' without leaving C equals one.*

In particular, C is closed given x: $P(C|\omega, x_\omega) = 1$, for each $\omega \in C$. Note that any recurrent set for x is communicating given x.

It is convenient to have the initial state of the game vary. Given an initial state ω, we denote by $(v^1(\omega), v^2(\omega))$ the threat point of the game.[2] If, facing a stationary strategy x^2, player 1 is to play s^1 in state ω, and to be punished immediately afterwards, his best future payoff is measured by the expectation $E\left[v^1|\omega, s^1, x^2_\omega\right]$ of v^1 under $P(\cdot|\omega, s^1, x^2_\omega)$. For $C \subseteq \Omega$, and given x^2, we set

$$H^1(x^2, C) = \max_{s^1 \in S^1} \max_{\omega \in C} E\left[v^1|\omega, s^1, x^2_\omega\right],$$

which somehow measures the threat point for player 1, against x^2, and given that the play visits all states of C. The definition of $H^2(x^1, C)$ is the symmetric one. It is easily seen that $H^1(x^2, C) \geq \max_C v^1(\omega)$.

---

[1] A game is recursive if the payoff function is identically zero outside absorbing states. Recursive games were introduced by Everett (1957).

[2] By definition, $v^i(\omega)$ is the value of the zero-sum stochastic game deduced from Γ, where the other player minimizes player i's payoff.

Let a profile x, and a recurrent set R for x be given. The (long-run) average payoff $E_{\omega,x}[\bar{g}_n]$ exists for each ω and is independent of $\omega \in R$. We denote it by $\gamma_R(x)$.

The definition builds upon a notion first introduced by Thuijsman and Vrieze (1991).

Definition 4 *A set $C \subseteq \Omega$ is* solvable *if, for some profile x, the following two conditions are fulfilled:*

1. *C is a communicating set given x.*

2. *There exists a point $\gamma = (\gamma^1, \gamma^2) \in \text{co}\,\{\gamma_R(x),\ R \text{ recurrent subset of } C\}$ such that*

$$(\gamma^1, \gamma^2) \geq (H^1(x^2, C), H^2(x^1, C)). \qquad (1)$$

This concept is motivated by the following observation. The communication requirement ensures that the players are able to visit the recurrent subsets of C cyclically by playing appropriate small perturbations of x. Given the interpretation of $(H^1(x^2, C), H^2(x^1, C))$ as a threat point, the inequality (1) may be interpreted as an individual rationality requirement. By a standard proof, one can show that γ is an equilibrium payoff of the game, provided the initial state belongs to C.

The set of equilibrium payoffs of the game does not increase when one replaces each state in a solvable set C by an absorbing state, with payoff the vector γ associated with C. Therefore, we assume throughout the chapter that all such sets coincide with absorbing states.

We now describe controlled sets. A pair $(\omega, s^i) \in C \times S^i$ is a *unilateral exit* of player i from $C \subseteq \Omega$ given a strategy x^{-i} if $P(C|\omega, s^i, x_\omega^{-i}) < 1$.

A triplet $(\omega, s^1, s^2) \in C \times S^1 \times S^2$ is a *joint exit* from C given x if $P(C|\omega, s^1, s^2) < 1$, and none of the pairs (ω, s^1) and (ω, s^2) is a unilateral exit.

Definition 5 *Let $C \subseteq \Omega$ be a communicating set given a profile x.*

The set C is controlled by player i *if there is a unilateral exit (ω, s^i) of player i (from C given x^{-i}) such that*

$$\left(E\left[v^1|\omega, s^i, x_\omega^{-i}\right], E\left[v^2|\omega, s^i, x_\omega^{-i}\right]\right) \geq (H^1(x^2, C), H^2(x^1, C)). \qquad (2)$$

The set C is jointly controlled *if there exists*

$$\gamma \in \text{co}\left\{E\left[v|\omega, s^1, s^2\right],\ (\omega, s^1, s^2) \text{ joint exit from } C \text{ given } x\right\}$$

such that

$$\gamma \geq (H^1(x^2, C), H^2(x^1, C)).$$

4

The rationale behind this definition runs as follows. Let $C \subseteq \Omega$ be a set controlled by player 1, and let x, $(\omega, s^1) \in C \times S^1$ be the associated profile and exit. Assume for simplicity that $P(C|\omega, s^1, x_\omega^2) = 0$.

Assume that we are given for each $\omega' \notin C$ an equilibrium payoff $\gamma(\omega')$ for the game starting at ω'. Then $E[\gamma(\cdot)|\omega, s^1, x_\omega^2]$ is an equilibrium payoff of the game, for every initial state in C.

We give few hints for this fact. By using appropriate perturbations of x, the players are able to come back repeatedly to ω without leaving C. If player 1 slightly perturbs x^1 by s^1 in each of these visits, the play leaves C in finite time and the exit state is distributed according to $P(\cdot|\omega, s^1, x_\omega^2)$. Given such a scenario, it takes many visits to ω before the plays leaves C. Hence player 1 may check the empirical choices of player 2 in these stages. Condition (2) implies that

- player 2 prefers playing x^2 in state ω to playing any other distribution and being punished; he prefers waiting for player 1 to use the exit (ω, s^1) to using any of his own exits, since

$$E\left[\gamma^2(\cdot)|\omega, s^1, x_\omega^2\right] \geq E\left[v^2(\cdot)|\omega, s^1, x_\omega^2\right] \geq H^2(x^1, C).$$

- player 1 prefers using the exit (ω, s^1) (and getting $E[\gamma^1(\cdot)|\omega, s^1, x_\omega^2]$) to using any other exit and being punished; he prefers using the exit (ω, s^1) to using no exit at all and being punished.

A similar property holds for jointly controlled sets.

A reduction to positive recursive games

To any controlled set C, we associate in a natural way a distribution μ_C of exit, i.e., a distribution such that $\mu_C(C) < 1$. If C is controlled by player i, let $\mu_C = P(\cdot|\omega, s^i, x_\omega^{-i})$ (with the notations of Definition 5). If C is jointly controlled, let μ_C be a convex combination of the distributions $P(\cdot|\omega, s^1, s^2)$, $((\omega, s^1, s^2)$ joint exit from C given $x)$ such that $E_{\mu_C}[v] = \gamma$.

Given a controlled set C, with its distribution μ_C of exit, define a *changed game* Γ_C by changing the transitions in each state of C to μ_C. For a collection $\mathcal{C}$ of disjoint controlled sets, the changed game $\Gamma_{\mathcal{C}}$ is obtained by applying this procedure to each element of $\mathcal{C}$.

In general, there is no inclusion between the equilibrium payoff sets of the original and the changed games Γ and $\Gamma_{\mathcal{C}}$. The goal of the next proposition, which is the main result in Vieille (2000a), is to exhibit a family $\mathcal{C}$ such that:

(i) such an inclusion holds and (ii) the changed game $\Gamma_{\mathcal{C}}$ has very specific properties.

Remember that, by assumption, the solvable sets of Γ coincide with the absorbing states of Γ.

Proposition 6 *There exists a family $\mathcal{C}$ of disjoint controlled sets with changed game $\Gamma_{\mathcal{C}}$ having the following property: for each strategy x^1 there exists a strategy x^2 such that (i) the play reaches an absorbing state in finite time; (ii) for each initial state ω_1, the expected termination payoff to player 2 is at least $v^2(\omega_1)$.*

Two remarks are in order. First, by (i), there must exist an absorbing state in Γ. The existence of solvable sets is therefore a corollary to the proposition. Next, the two games Γ and $\Gamma_{\mathcal{C}}$ need not have the same threat point v. The value $v^2(\omega_1)$ that appears in the statement is that of Γ.

Let $\mathcal{C}$ be given by this proposition. Let $\Gamma_{\mathcal{C}}^0$ be the game obtained from $\Gamma_{\mathcal{C}}$, after setting the payoff function to zero in each non-absorbing state.

Note that $\Gamma_{\mathcal{C}}^0$ is a recursive game such that:

P.1 all absorbing payoffs to player 2 are positive;

P.2 player 2 can force the play to reach an absorbing state in finite time: for any profile $x = (x^1, x^2)$ where x^2 is fully mixed, the play reaches an absorbing state in finite time, whatever the initial state.

Property **P.1** is a consequence of the assumption $g^2 > 0$; property **P.2** follows from Proposition 6 (i). Recursive games that satisfy both properties **P.1** and **P.2** are called positive recursive games.

It can be shown[3] that each equilibrium payoff of $\Gamma_{\mathcal{C}}^0$ is also an equilibrium payoff of the initial game Γ. The main consequence of Proposition 6 is thus that one is led to study positive recursive games.

Existence of equilibrium payoffs in positive recursive games

We now present some of the ideas in the proof of the result:

Proposition 7 *Every (two-player) recursive game which satisfies* P.1 *and* P.2 *has an equilibrium payoff.*

---

[3]This is where the assumption $g^1 < 0 < g^2$ comes into play.

In zero-sum recursive games, ε-optimal strategies do exist (Everett (1957)). In non-zero-sum positive recursive games, stationary ε-equilibria need not exist. For instance, in the game[4]

$$
\begin{array}{|cc|} \hline
\text{½} & \omega_1 \\
\frac{1}{5} & (-1,3)^* \\
\frac{4}{5} & \omega_2 \\ \hline
\end{array}
\;\; \omega_0
\qquad\qquad
\begin{array}{|c|c|} \hline
\omega_0 & 1,1^* \\ \hline
\end{array}
\;\; \omega_1
$$

$$
\begin{array}{|c|c|} \hline
\omega_0 & 2,2^* \\ \hline
\end{array}
\;\; \omega_2
$$

Example 1

one can check that no stationary profile x exists that would be an ε-equilibrium for every initial state.

Throughout this section, take Γ to be a fixed positive recursive game. The basic idea of the proof is to approximate the game by a sequence of constrained games. For each $\varepsilon > 0$, let Γ_ε be the game in which player 2 is constrained to use stationary strategies that put a weight of at least ε on each single action. Player 1 is unconstrained. A crucial feature of Γ_ε is that the average payoff function, defined for stationary profiles x by $\gamma(x) = \lim_{n\to\infty} \gamma_n(x)$, is continuous.

Next, one defines B_ε as an analog of the best-reply correspondence on the space of constrained stationary profiles. This correspondence is well-behaved so that: (i) it has a fixed point x_ε for each $\varepsilon > 0$, and (ii) the graph of fixed points (as a function of ε) is semialgebraic, hence there is a selection $\varepsilon \mapsto x_\varepsilon$ of fixed points such that $x_\varepsilon^{i,\omega}(s^i)$ has an expansion in Puiseux series in the neighborhood of zero (see [M], Section 4). This can be shown to imply that the limits $x_0 = \lim_{\varepsilon\to 0} x_\varepsilon$ and $\gamma = \lim_{\varepsilon\to 0} \gamma(x_\varepsilon)$ do exist. Finally, one proves that, for each ω, γ_ω is an equilibrium payoff for Γ starting in ω; an associated ε-equilibrium consists in playing a history-dependent perturbation of x_0, sustained by appropriate threats.

Solan (2000) proves that, by taking the usual best-reply map for B_ε, the program sketched in the previous paragraph works for games in which there are no more than two non-absorbing states, but not for more general games.

Before defining B_ε in greater detail, we assume we have an intuitive notion of what it is. Given a fixed point x_ε of B_ε, we begin by describing the asymptotic behavior of the play, as ε goes to zero. This discussion will point out some of the requirements that a satisfactory definition of B_ε should meet.

---

[4] In this example, each entry contains only the transitions. Transitions are deterministic except in state ω_0, when player 1 plays the Bottom row; the play then moves, with probability $\frac{4}{5}$, to the state ω_2, and to an absorbing state with payoff $(-1,3)$ otherwise.

For each $\varepsilon > 0$, given $x_\varepsilon = (x_\varepsilon^1, x_\varepsilon^2)$, the play reaches an absorbing state in finite time, since x_ε^2 is fully mixed and since Γ satisfies **P.2**. As ε goes to zero, the probability of some actions may vanish, and there may exist recurrent sets for x_0 that contain non-absorbing states.

Define a binary relation $\to$ on the non-absorbing states by $\omega \to \omega'$ if and only if the probability (starting from ω, computed for x_ε) that the play visits ω' converges to one as ε goes to zero. Define an equivalence relation by

$$\omega \sim \omega' \ \Leftrightarrow \ (\ \omega \to \omega' \text{ and } \omega' \to \omega).$$

The different equivalence classes define a partition of the set of non-absorbing states. Note that a transient state (given x_0) may be included in a larger equivalence class, or constitute an equivalence class by itself. One can check that each class is either a transient state, or a set that is communicating given $x_0 = \lim_{\varepsilon \to 0} x_\varepsilon$.

Consider an equivalence class C of the latter type, and let $\varepsilon > 0$ be fixed. Since the play reaches the set of absorbing states in finite time, C is transient under x_ε. Hence, given an initial state in C, the distribution Q_C^ε of the exit state[5] from C is well-defined. This distribution usually depends on the initial state in C. Since (x_ε) has a Puiseux expansion in the neighborhood of zero, it can be shown that the limit $Q_C = \lim_{\varepsilon \to 0} Q_C^\varepsilon$ exists. Moreover, it is *independent* of the initial state in C. Next, the distribution Q_C has a natural decomposition as a convex combination of the distributions

$$P(\cdot|\omega, s^i, x_0^{-i}), \text{ where } (\omega, s^{-i}) \text{ is a unilateral exit of } C \text{ given } x_0$$

and

$$P(\cdot|\omega, s^1, s^2), \text{ where } (\omega, s^1, s^2) \text{ is a joint exit from } C \text{ given } x_0.$$

It is straightforward to observe that the limit payoff vector $\gamma(\cdot) = \lim_{\varepsilon \to 0} \gamma(\cdot, x_\varepsilon)$ is such that, for $\omega \in C$, $\gamma(\omega)$ coincides with the expectation $E_{Q_C}[\gamma(\cdot)]$ of $\gamma(\cdot)$ under Q_C.

The main issue in designing the family $(B_\varepsilon)_\varepsilon$ of maps is to ensure that C is somehow controlled, in the following sense. Assuming that $\gamma(\omega')$ is an equilibrium payoff for the game starting from $\omega' \notin C$, it should be the case that $\gamma(\omega) = E_{Q_C}[\gamma(\cdot)]$ is an equilibrium payoff starting from $\omega \in C$. The main difficulty arises when the decomposition of Q_C involves two unilateral exits $(\bar{\omega}, \bar{s}^2)$, $(\boldsymbol{\omega}, \boldsymbol{s}^2)$ of player 2, such that $E\left[\gamma^2(\cdot)|\bar{\omega}, x_{0,\bar{\omega}}^2, \bar{s}^2\right] >$

---

[5] Which is defined to be the actual current state, at the first stage for which the current stage does not belong to C.

$E\left[\gamma^2(\cdot)|\underline{\omega}, x^2_{0,\underline{\omega}}, \underline{s}^2\right]$. Indeed, in such a case, player 2 is not indifferent between the two exits, and would favor using the exit $(\bar{\omega}, \bar{s}^2)$.

The approach in Vieille (2000b) is similar to proper ε-equilibrium. Given $x = (x^1, x^2)$, one measures for each pair $(\omega, s^2) \in \Omega \times B$ the *opportunity cost* of using s^2 in state ω by $\max_{S^2} E\left[\gamma^2(x)|\omega, x^1_\omega, \cdot\right] - E\left[\gamma^2(x)|\omega, x^2_\omega, s^2\right]$ (it thus compares the expected continuation payoff by playing s^2 with the maximum achievable). $B^2_\varepsilon(x)$ consists of those $\bar{x}^2$ such that whenever the pair (ω, s^2) has a higher opportunity cost than $(\bar{\omega}, \bar{s}^2)$, then the probability $\bar{x}^2_\omega(s^2)$ assigned by $\bar{x}^2$ to s^2 at state ω is quite small compared with $\bar{x}^2_{\bar{\omega}}(\bar{s}^2)$. One then sets $B_\varepsilon(x) = B^1_\varepsilon(x) \times B^2_\varepsilon(x)$, where B^1_ε is the best-reply map of player 1.

We conclude by giving a few stylized properties that show how to deal with the difficulties mentioned above. Since both exits $(\bar{\omega}, \bar{s}^2)$ and $(\underline{\omega}, \underline{s}^2)$ have a positive contribution to Q_C, it follows that $\underline{\omega}$ is visited (infinitely, as ε goes to zero) more often than $\bar{\omega}$, and also that, in some sense, facing x^1_0, player 2 can not reach $\bar{\omega}$ from $\underline{\omega}$, hence communication from $\underline{\omega}$ to $\bar{\omega}$ can be blocked by player 1. Thus player 1 is able to influence the relative frequency of visits in $\underline{\omega}$ and $\bar{\omega}$, hence the relative weight of the two exits $(\bar{\omega}, \bar{s}^2)$, $(\underline{\omega}, \underline{s}^2)$. It must therefore be the case that player 1 is indifferent between the two exits $(\underline{\omega}, \underline{s}^2)$ and $(\bar{\omega}, \bar{s}^2)$. The ε-equilibrium profile will involve a lottery performed by player 1, who chooses which of the two exits (if any) should be used to leave C.

Comments

1- The lack of symmetry between the two players may appear somewhat unnatural. However, it is not an artifact of the proof since symmetric stochastic games need not have a symmetric ε-equilibrium. For instance, the only equilibrium payoffs of the symmetric game

$0,0$	$2, 1^*$
$1, 2^*$	$1, 1^*$

are $(1, 2)$ and $(2, 1)$.

2- All the complexity of the ε-equilibrium profiles lies in the punishment phase.

3- The main characteristics of the ε-equilibrium profile (solvable sets, controlled sets, exit distributions, stationary profiles that serve as a basis for the perturbations) are independent of ε. The value of $\varepsilon > 0$ has an influence on the statistical tests used to detect potential deviations, the size of the perturbations used to travel within a communicating set, and the specification of the punishment strategies.

4- The above proof has many limitations. Neither of the two parts extends to games with more than two players. The ε-equilibrium profiles have no subgame perfection property. Finally, in zero-sum games, the value exists as soon as payoffs are observed (in addition to the current state). For non-zero-sum games, the tests check past choices. Whether an equilibrium exists when only the vector of current payoffs is publicly observed, is not known.

5- These ε-equilibrium profiles involve two phases: after a solvable set is reached, players accumulate payoffs (and check for deviations); before a solvable set is reached, they care only about transitions (about which solvable set will eventually be reached). This distinction is similar to the one which appears in the proof of existence of equilibrium payoffs for games with one-sided information (Simon et al. (1995)), where a phase of information revelation is followed by payoff accumulation. This (rather vague) similarity suggests that a complete characterization of equilibrium payoffs for stochastic games would intertwine the two aspects in a complex way, by analogy with the corresponding characterization for games with incomplete information (Hart (1985)).

6- In Example 1, the following holds: given an initial state ω, and $\varepsilon > 0$, the game starting at ω has a stationary ε-equilibrium. Whether this holds for any positive recursive game is not known.

3   Games with more than two players

It is as yet unknown whether n-player stochastic games always have an equilibrium payoff. We describe a partial result for three-player games, and explain what is specific to this number of players.

The first contribution is due to Flesch, Thuijsman and Vrieze (1997), who analyzed Example 2 below.

$0,0,0$	$0,1,3^*$			$3,0,1^*$	$1,1,0^*$
$1,3,0^*$	$1,0,1^*$			$0,1,1^*$	$0,0,0^*$

<div align="center">Example 2</div>

This example falls in the class of repeated games with absorbing states: there is a single non-absorbing state (in other words, the current state changes once at most during any play). We follow customary notations (see [M]). Players 1, 2 and 3 choose respectively a row, a column and a matrix. Starting from

<div align="center">10</div>

the non-absorbing state, the play moves immediately to an absorbing state, unless the move combination (Top, Left, Left) is played.

In this example, the set of equilibrium payoffs coincides with those convex combinations $(\gamma^1, \gamma^2, \gamma^3)$ of the three payoffs $(1, 3, 0), (0, 1, 3), (3, 0, 1)$ such that $(\gamma^1, \gamma^2, \gamma^3) \geq (1, 1, 1)$, and $\gamma^i = 1$ for at least one player i. Corresponding ε-equilibrium profiles involve cyclic perturbations of the profile of stationary (pure) strategies (Top, Left, Left). Rather than describe this example in greater detail, we discuss a class of games below that includes it.

This example gave the impetus for the study of three-player games with absorbing states (see Zamir (1992), Section 5 for some motivation concerning this class of games). The next result is due to Solan (1999).

Theorem 8 *Every three-player repeated game with absorbing states has an equilibrium payoff.*

Sketch of the Proof: Solan defines an auxiliary stochastic game in which the current payoff $\underline{g}(x)$ is defined to be the (coordinatewise) minimum of the current vector payoff $g(x)$ and of the threat point.[6] He then uses Vrieze and Thuijsman's (1989) idea of analyzing the asymptotic behavior (as $\lambda \to 0$) of a family $(x_\lambda)_{\lambda>0}$ of stationary equilibria of the auxiliary λ-discounted game.

The limits $\lim_{\lambda \to 0} x_\lambda$ and $\lim_{\lambda \to 0} \gamma_\lambda(x_\lambda)$ do exist, up to a subsequence. If it happens that $\lim_{\lambda \to 0} \gamma_\lambda(x_\lambda) = \gamma(\lim_{\lambda \to 0} x_\lambda)$, then $x = \lim_{\lambda \to 0} x_\lambda$ is a stationary equilibrium of the game. Otherwise, it must be the case that the nature of the Markov chain defined by x_λ changes at the limit: for $\lambda > 0$ close enough to zero, the non-absorbing state is transient for x_λ, whereas it is recurrent for x.

In this case, the limit payoff $\lim_{\lambda \to 0} \gamma_\lambda(x_\lambda)$ can be written as a convex combination of the non-absorbing payoff $\underline{g}(x)$ (which by construction is dominated by the threat point) and of payoffs received in absorbing states reached when perturbing x. By using combinatorial arguments, Solan constructs an ε-equilibrium profile that coincides with cyclic perturbations of x, sustained by appropriate threats.

In order to illustrate Example 2 above and Solan's proof, we focus on the following games, called *quitting games*. Each player has two actions: *quit* and *continue:* $S^i = \{c^i, q^i\}$. The game ends as soon as at least one player chooses to quit (if no player ever quits, the payoff is zero). For simplicity, we assume that a player receives 1 if he is the only one to quit.

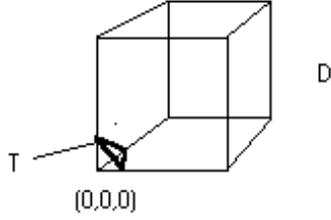
[6]In particular, the current payoff is not multilinear.

Figure 1

A stationary strategy is characterized by the probability of quitting, i.e., by a point in $[0,1]$. Hence the space of stationary profiles is the unit cube $D = [0,1]^3$, with $(0,0,0)$ being the unique non-absorbing profile.

Assume first that, for some player, say player 1, the payoff vector $\gamma(q^1, c^2, c^3)$ is of the form $(1, +, +)$, where the $+$ sign stands for "a number higher than or equal to one". Then the following stationary profile is an ε-equilibrium, provided α is small enough: player 1 quits with probability α, player 2 and 3 continue with probability 1.

We now rule out such configurations. For $\varepsilon > 0$ small, consider the constrained game where the players are restricted to stationary profiles x that satisfy $\sum_{i=1}^{3} x^i \geq \varepsilon$, i.e., the points below the triangle $T = \{x \in D, x^1 + x^2 + x^3 = \varepsilon\}$ are chopped off D (see Figure 1).

If it happens that at every point $x \in T$, one has $\gamma^i(x) < 1$ for some[7] i, then any stationary equilibrium of the constrained game (which exists by standard fixed-point arguments) is a stationary equilibrium of the true game.

It therefore remains to discuss the case where $\gamma(x_0) = (+, +, +)$ for some $x_0 \in T$. Given $x \in T$, the probability that two players quit simultaneously is of order ε^2, hence γ is close on T to the linear function

$$x^1 \gamma(q^1, c^{-1}) + x^2 \gamma(q^2, c^{-2}) + x^3 \gamma(q^3, c^{-3}).$$

Since $\gamma^1(q^1, c^{-1}) = 1$, and $\gamma^1(x_0) \geq 1$, it must be that $\gamma^1(q^2, c^{-2}) \geq 1$ or $\gamma^1(q^3, c^{-3}) \geq 1$. Similar observations hold for the other two players.

If $\gamma(q^1, c^{-1})$ were of the form $(1, -, -)$, one would have $\gamma(q^2, c^{-2}) = (+, 1, +)$ or $\gamma(q^3, c^{-3}) = (+, +, 1)$, which has been ruled out. Up to a permutation of players 2 and 3, one can assume $\gamma(q^1, c^{-1}) = (1, +, -)$. The signs of $\gamma(q^2, c^{-2})$ and $\gamma(q^3, c^{-3})$ are then given by $(-, 1, +)$ and $(+, -, 1)$.

---

[7]Player i would then rather quit than let x be played. In geometric terms, the best-reply map points inwards on T.

(-,1,+)

B

(+,+,+)

C

A
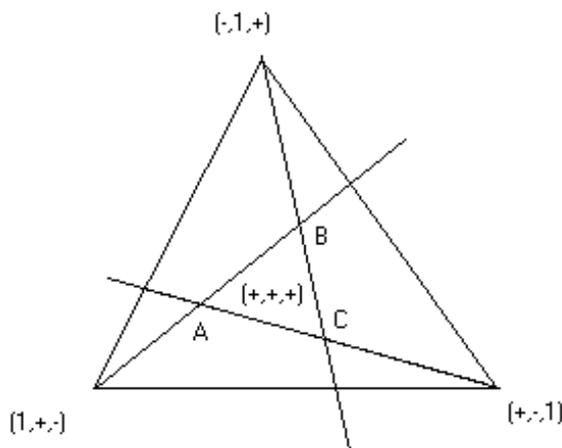
(1,+,-)   (+,-,1)

Figure 2

Draw the triangle T together with the straight lines $\{x, \gamma^i(x) = 1\}$, for $i = 1, 2, 3$.

The set of $x \in T$ for which $\gamma(x) = (+, +, +)$ is the interior of the triangle (ABC) delineated by these straight lines. We now argue that for each x on the edges of (ABC), $\gamma(x)$ is an equilibrium payoff. Consider for instance $\gamma(A)$ and let σ be the strategy profile that plays cyclically: according to the stationary profile $(\eta, 0, 0)$ during N_1 stages, then according to $(0, \eta, 0)$ and $(0, 0, \eta)$ during N_2 and N_3 stages successively. Provided N_1, N_2, N_3 are properly chosen, the payoff induced by σ coincides with $\gamma(A)$. Provided η is small enough, in the first N_1 stages (*resp.* next N_2, next N_3 stages), the continuation payoff[8] moves along the segment joining $\gamma(A)$ to $\gamma(B)$ (*resp.*, $\gamma(B)$ to $\gamma(C)$, $\gamma(C)$ to $\gamma(A)$). Therefore, σ is an ε-equilibrium profile associated with $\gamma(A)$.

Clearly, this approach relies heavily upon the geometry of the three-dimensional space. Note that, for such games, there is a stationary ε-equilibrium or an equilibrium payoff in the convex hull of $\{\gamma(q^i, c^{-i}), i = 1, 2, 3\}$. Solan and Vieille (2000b) devised a four-player quitting game for which this property does not hold. Whether or not n-player quitting games do have equilibrium payoffs remains an intriguing open problem.[9]

---

[8]I.e., the undiscounted payoff obtained in the subgame starting at that stage.

[9]A partial existence result is given in Solan and Vieille (2000b).

13

An important trend in the literature is to identify classes of stochastic games for which there exist ε-equilibrium profiles (see for instance Thuijsman and Raghavan (1997)) that exhibit a simple structure (stationary, Markovian, etc.).

To conclude, we mention that the existence of (extensive-form) correlated equilibrium payoffs is known (Solan and Vieille (2000a)).

Theorem 9 *Every stochastic game has an (autonomous) extensive-form correlated equilibrium payoff.*

The statement of the result refers to correlation devices that send (private) signals to the players at each stage. The distribution of the signals sent in stage n depends on the signal sent in stage $n-1,$ and is independent of any other information.

Idea of the proof: The first step is to construct a "good" strategy profile, meaning a profile that yields all players a high payoff, and by which no player can profit by a unilateral deviation that is followed by an indefinite punishment. One then constructs a correlation device that imitates this profile: the device chooses for each player a recommended action according to the probability distribution given by the profile. It also reveals to all players what its recommendations were in the previous stage. In this way, a deviation is detected immediately.

4   Zero-sum games with imperfect monitoring

These are games where, at any stage, each player receives a private signal which depends, possibly randomly, on the choices of the players (see Sorin (1992), Section 5.2 for the model). In constrast to (complete information) repeated games, dropping the perfect monitoring assumption already has important implications in the zero-sum case.

It is instructive to consider first the following striking example (Coulomb (1999)):

	L	M	R
T	100	1*	0*
B	100	0	1

Example 3

When player 1 plays B, he receives the signal a if either the L or M column was chosen by player 2, and the signal b otherwise. The signals to player 2, and to player 1 when he plays the T row, are irrelevant for what follows.

Note that the right-hand side

1^*	0^*
0	1

of the game coincides (up to an affine transformation on payoffs) with the Big Match (see [M], Section 2), which was shown to have the value $\frac{1}{2}$. We now show that the addition of the L column, which is apparently dominated, has the effect of bringing the $\max\min$ down to zero.

Indeed, let σ be any strategy of player 1, and let y be the stationary strategy of player 2 that plays L and R with probabilities $1 - \varepsilon$ and ε, respectively. Denote by θ the absorbing stage, i.e. the first stage in which one of the two move profiles (T, M) or (T, R) is played.

If $P_{\sigma,y}(\theta < +\infty) = 1$, then $\gamma_n(\sigma, y) \to \varepsilon$ as n goes to infinity. Otherwise, choose an integer N large enough so that $P_{\sigma,y}(N \leq \theta < +\infty) < \varepsilon^2$. In particular,

$$P_{\sigma,y}(\theta \geq N, \text{ player 1 ever plays } T \text{ after stage } N) \leq \varepsilon.$$

Let y' be the stationary strategy of player 2 that plays M and R with probabilities $1 - \varepsilon$ and ε, respectively, and call τ the strategy that coincides with y up to stage N, and with y' afterwards. Since (B, L) and (B, M) yield the same signal to player 1, the distributions induced by (σ, y) and by (σ, τ) on sequences of signals to player 1 coincide up to the first stage after stage N, in which player 1 plays T.

Therefore, $P_{\sigma,\tau}$-almost surely,

$$\bar{g}_n \to 0 \text{ if } \theta < N$$
$$\bar{g}_n \to 1 - \varepsilon \text{ if } N \leq \theta < +\infty$$
$$\bar{g}_n \to \varepsilon \text{ if } \theta = +\infty.$$

Since $P_{\sigma,y}(N \leq \theta < +\infty) \leq \varepsilon$, $\lim_{n \to +\infty} E_{\sigma,\tau}[\bar{g}_n] \leq 2\varepsilon$. Thus, player 2 can defend zero.

Since player 1 clearly guarantees 0 and can defend $\frac{1}{2}$, the game has no value and the $\max\min$ is equal to zero.

The following theorem is due to Coulomb (1995,1999).

Theorem 10 *Every zero-sum repeated game with absorbing states and partial monitoring has a* $\max\min$.

Sketch of the proof: Following the steps of Kohlberg (1974), the max min is first shown to exist for so-called generalized Big Match games, then for all games with absorbing states.

The class of generalized Big Match games includes Example 3. Player 1 has only two actions, T and B, while the action set S^2 of player 2 is partitioned into $\bar{S}^2$ and $\mathcal{S}^2$. For $s^2 \in \bar{S}^2$, transitions and payoffs are independent of $s^1 \in \{T, B\}$. For $s^2 \in \mathcal{S}^2$, the probability of reaching an absorbing state is positive given (T, s^2) and equals zero given (B, s^2). Coulomb (1999) characterizes the max min for such games; as in Example 3, it depends only on the signal structure to player 1, given the action B.

As might be expected, the max min is quite sensitive to the signalling structure. For instance, consider again Example 3. Assume that the signal a associated with the entry (B, M) is replaced by a random device that sends the signal a with probability $1 - \eta$, and the signal a' otherwise. If $a' = b$, the max min is still close to zero for η small (the M column is indistinguishable from a convex combination of the L and R columns). If $a' \neq b$, the max min is equal to $\frac{1}{2}$, whatever the value of $\eta > 0$.

Let Γ be any game with absorbing states. To any pair $x^1, \mathbf{x}^1$ of distributions over S^1 is associated a (fictitious) generalized Big Match $\Gamma_{x^1, \mathbf{x}^1}$, in which the B and T rows correspond, respectively, to the mixed moves x^1 and x^1, slightly perturbed by $\mathbf{x}^1$. It is shown by Coulomb (1995) that the max min of Γ is equal to the supremum over $(x^1, \mathbf{x}^1)$ of the max min of the auxiliary game $\Gamma_{x^1, \mathbf{x}^1}$. The difficult part is to show that player 2 can defend such a quantity.

5 Stochastic games with large state space

Consider first a stochastic game with countable state space Ω and finite actions sets S^1 and S^2. Maitra and Sudderth (1992) prove that, with $\limsup_{n \to +\infty} g_n$ as the payoff function for the infinite game,[10] the game has a value. This result was considerably extended by Martin (1998). Let Ω, S^1 and S^2 be endowed with the discrete topology, and the set H_∞ of plays be given the product topology. Let the payoff function of the infinite game be *any* Borel function f on H_∞. (Martin does not deal with stochastic games, but, as argued in Maitra and Sudderth (2000), the extension to stochastic games is immediate.)

Theorem 11 *The game with payoff function f has a value.*

---

[10]This payoff function includes many cases of interest, including discounted stochastic games.

See Martin (1998) for the proof. The proof relies on another theorem of Martin (1975) for games of perfect information. We also refer to Maitra and Sudderth (2000) for an introduction to the proof.

We conclude by citing miscellaneous results. In stochastic games with incomplete information on one side, a lottery chooses at stage 0 the stochastic game to be played, and only player 1 is informed. Such games may be analyzed through an auxiliary stochastic game in which the current posterior held by player 2 on the true game being played is part of the state variable. It is conjectured that the max min exists and coincides with $\lim_{n\to\infty} v_n$ and $\lim_{\lambda\to 0} v_\lambda$. The basic intuition is that the max min should coincide with the value of the auxiliary game, which is not known to exist (see Mertens (1987)). Only scattered results are available so far. This has been verified by Sorin (1984) for games of the Big Match type, and by Rosenberg and Vieille (2000) when the possible stochastic games are recursive. For games with absorbing states, it is known that $\lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$ (Rosenberg (2000)).

References

[1] J.M. Coulomb. Repeated games with absorbing states and signaling structure. *Mathematics of Operations Research*, forthcoming, 1995.

[2] J.M. Coulomb. Generalized big-match. *Mathematics of Operations Research*, 24:795–816, 1999.

[3] H. Everett. Recursive games. In M. Dresher, A.W. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games, Vol. III*, pages 47–78. Princeton University Press, 1957.

[4] S. Hart. Non-zero-sum two-person repeated games with incomplete information. *Mathematics of Operations Research*, 10:117–153, 1985.

[5] J. Flesch, F. Thuijsman and O.J. Vrieze. Cyclic Markov equilibrium in stochastic games. *International Journal of Game Theory*, 26:303–314, 1997.

[6] E. Kohlberg. Repeated games with absorbing states. *Annals of Statistics*, 2:724–738, 1974.

[7] A. Maitra and W. Sudderth. An operator solution for stochastic games. *Israel Journal of Mathematics*, 78:33–49, 1992.

[8] A. Maitra and W. Sudderth. Stochastic games with Borel payoffs. In A. Neyman and S. Sorin, editors, *To Appear in Proceedings of the NATO ASI on Stochastic Games, Stony Brook 1999*, forthcoming, 2000.

[9] D.A. Martin. Borel determinacy. *Annals of Mathematics*, 102:363–371, 1975.

[10] D.A. Martin. The determinacy of Blackwell games. *The Journal of Symbolic Logic*, 63:1565–1581, 1998.

[11] J.F. Mertens. Repeated games. In *Proceedings of the International Congress of Mathematics (Berkeley, 86)*, pages 1528–1577. Americal Mathematical Society, 1987.

[12] J.F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10(2):53–56, 1981.

[13] D. Rosenberg. Zero-sum absorbing games with incomplete information on one side: Asymptotic analysis. *SIAM Journal of Control and Optimization*, 39:208–225, 2000.

[14] D. Rosenberg and N. Vieille. The MaxMin of recursive games with incomplete information on one side. *Mathematics of Operations Research*, 25:23–35, 2000.

[15] R.S. Simon, S. Spiez and H. Torunczyk. The existence of equilibria in certain games, separation for families of convex functions and a theorem of Borsuk-Ulam type. *Israel Journal of Mathematics*, 92:1–21, 1995.

[16] E. Solan. Three-player absorbing games. *Mathematics of Operations Research*, 24:669–698, 1999.

[17] E. Solan. Stochastic games with two non-absorbing states. *Israel Journal of Mathematics*, 119:29–54, 2000.

[18] E. Solan and N. Vieille. Correlated equilibrium in stochastic games. *Games and Economic Behavior*, forthcoming, 2000a.

[19] E. Solan and N. Vieille. Quitting games. *Mathematics of Operations Research*, forthcoming, 2000b.

[20] S. Sorin. Big match with lack of information on one side (part 1). *International Journal of Game Theory*, 13:201–255, 1984.

[21] S. Sorin. Repeated games with complete information. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications, Vol. I*, chapter 4, pages 71–104. North-Holland, 1992.

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

[23] F. Thuijsman and O.J. Vrieze. Easy initial states in stochastic games. In T.E.S. Raghavan and Al., editors, *Stochastic Games and Related Topics*, pages 85–100. Kluwer Academic Publishers, 1991.

[24] N. Vieille. Two-player stochastic games I: A reduction. *Israel Journal of Mathematics*, 119:55–92, 2000a.

[25] N. Vieille. Two-player stochastic games II: The case of recursive games. *Israel Journal of Mathematics*, 119:93–126, 2000b.

[26] N. Vieille. Small perturbations and stochastic games. *Israel Journal of Mathematics*, 119:127–142, 2000c.

[27] O.J. Vrieze and F. Thuijsman. On equilibria in stochastic games with absorbing states. *International Journal of Game Theory*, 18:293–310, 1989.

[28] S. Zamir. Repeated games of incomplete information: zero-sum. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications, Vol. I*, chapter 5, pages 109–154. North-Holland, 1992.