International Journal of Game Theory (1997) 26: 403–408

Game Theory

Perfect Information Stochastic Games and Related Classes[1]

FRANK THUIJSMAN

Department of Mathematics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
email: frank@math.unimaas.nl

THIRUKKANNAMANGAI E.S. RAGHAVAN

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 South Morgan Room 517, Chicago, IL 60607-7045, USA
email: t.e.raghavan@uicvm.cc.uic.edu

Abstract: For n-person perfect information stochastic games and for n-person stochastic games with Additive Rewards and Additive Transitions (ARAT) we show the existence of pure limiting average equilibria. Using a similar approach we also derive the existence of limiting average ε-equilibria for two-person switching control stochastic games. The orderfield property holds for each of the classes mentioned, and algorithms to compute equilibria are pointed out.

1 Model

We deal with n-person stochastic games with finite state and action spaces. Only for some special classes of stochastic games limiting average ε-equilibria are known to exist, but generally their existence remains to be an open problem (see Thuijsman [1992] for a survey on equilibrium existence). For zerosum stochastic games Mertens & Neyman [1981] showed the existence of the limiting average value. Approaching this value generally involves the use of history dependent strategies. In this paper we show existence of limiting average equilibria (in 'almost stationary' behavior strategies) for perfect information stochastic games and for stochastic games with Additive Rewards and Additive Transitions (ARAT). For two-person stochastic games with switching control we show the existence of limiting average ε-equilibria. For none of these related classes (any perfect information game has ARAT as well as switching control structure) equilibrium existence was known before. Our method implies that the orderfield property holds for these classes, i.e. if payoffs and transitions are rational, then there are rational equilibrium strategies and rational equilibrium rewards as well. Algorithms to determine equilibria are also pointed out. Whenever we speak

[1] We gratefully acknowledge valuable remarks by J. Flesch and by anonymous referees on earlier versions.

about equilibria, optimal strategies, best replies etc., we shall always have limiting average rewards in mind.

Pure stationary optimal strategies exist for zerosum perfect information games (cf. Liggett & Lippman [1969]) as well as for zerosum ARAT games (cf. Raghavan et al. [1985]). For zerosum switching control stochastic games there are also stationary optimal strategies (cf. Filar [1981]), these however are not necessarily pure. An example at the end shows that in the non-zerosum case stationary solutions may fail to exist in any of the classes mentioned. Another example explains why for switching control games we have to restrict to the two-person case.

An n-person stochastic game is given by: (a) a set of players $N=\{1,2,\ldots,n\}$, (b) a set of states $S=\{1,2,\ldots,z\}$, (c) for each state s and each player i a finite set of actions A_s^i, (d) for each state s and each joint action $a\in A_s=\prod_{j\in N}A_s^j$ a payoff $r^i(s,a)\in\mathbb{R}$ to player i, and (e) for each state s and each joint action $a\in A_s$ a transition probability vector $p(s,a)=(p(1|s,a),(p(2|s,a),\ldots,p(z|s,a))\in[0,1]^z$. A perfect information stochastic game has the property:

$$\forall s\,\exists i\,\forall j\neq i:\quad \#A_s^j=1.$$

An ARAT stochastic game has additive rewards and additive transitions:

$$\forall s\,\forall i\,\forall a: r^i(s,a)=\sum_j r_j^i(s,a^j)\quad\text{and}\quad \forall s\,\forall t\,\forall a: p(t|s,a)=\sum_j p_j(t|s,a^j),$$

for some functions r_j^i and p_j. A switching control stochastic game has the property:

$$\forall s\,\exists i\,\forall a,b:\ \text{if } a^i=b^i, \text{ then } p(s,a)=p(s,b).$$

Play can start in any state of S and evolves by players $i\in N$ independently choosing actions $a_k^i\in A_{s_k}^i$, where s_k is the state visited at stage k. A strategy for player i is a rule to decide for any history $h_k=(s_1,a_1,s_2,a_2,\ldots,s_{k-1},a_{k-1},s_k)$, with $a_h\in A_{s_h}$, what mixed action (mixed over $A_{s_k}^i$) to use in state s_k at stage $k\in\mathbb{N}$. Generally, strategies are denoted by σ^i for player i. A joint strategy is denoted by σ and σ^{-i} denotes a joint strategy of the players in $N\setminus\{i\}$. For initial state s and joint strategy σ the reward to player i is given by $\gamma^i(s,\sigma)=E_{s\sigma}(\liminf_{T\to\infty}\frac{1}{T}\sum_{k=1}^T r^i(S(k),A(k)))$, where $S(k)$, $A(k)$ are random variables for state and action at stage k. Let $\gamma^i(\sigma)=(\gamma^i(1,\sigma),\gamma^i(2,\sigma),\ldots,\gamma^i(z,\sigma))$. Stationary strategies for player i are denoted by $x^i=(x_1^i,x_2^i,\ldots,x_z^i)$, where x_s^i is the mixed action to be used by player i in state s, whenever state s is being visited. A strategy is called pure if it never uses any randomization. We denote pure stationary strategies for player i by f^i.

2 Results

Theorem 1: Every n-person perfect information stochastic game has pure equilibria.

Proof: Given an n-person perfect information game Γ, consider the related zerosum game $\Gamma(i)$ where player i is maximizing his reward, while players $N \setminus \{i\}$ are jointly minimizing player i's reward. By Liggett & Lippman [1969]:

$$\exists v^i \, \exists f(i) \, \forall \sigma : \gamma^i(f^i(i), \sigma^{-i}) \geq v^i \geq \gamma^i(\sigma^i, f^{-i}(i)).$$

Then $\gamma^i(f^1(1), \ldots, f^n(n)) \geq v^i$ for all i. Since $f^i(i)$ is pure, any deviation of player i will be observed immediately and can always be punished by the players in $N \setminus \{i\}$ using $f^{-i}(i)$ because, for any state s and any action $a^i \neq f^i_s(i)$, the transition probability only depends on player i's action and, by the optimality of $f^i(i)$:

$$\sum_t p(t \mid s, a^i)\, v^i_t \leq v^i_s = \sum_t p(t \mid s, f^i_s(i)) v^i_t.$$

Now for each player i we define π^i by: play $f^i(i)$ as long as no player has deviated; if at some previous stage player $j \neq i$ has deviated, then play $f^i(j)$. It can be verified that $(\pi^1, \ldots, \pi^n)$ is a pure equilibrium in Γ. □

Theorem 2: Every n-person ARAT stochastic game has pure equilibria.

Proof: We now use the result of Raghavan et al. [1985] to take, for each i, pure strategies $f(i)$ as in the previous proof. Then, completely analogous to the above, we define $\pi^1, \ldots, \pi^n$ and, for the possibility of punishment, we observe that for any state s and any action $a^i \in A^i_s$:

$$\begin{aligned}
\sum_t p(t \mid s, a^i, \pi_s^{-i})\, v^i_t &= \sum_t p_i(t \mid s, a^i) v^i_t + \sum_t \sum_{j \neq i} p_j(t \mid s, \pi^j_s) v^i_t \\
&\leq \sum_t p_i(t \mid s, f^i_s(i)) v^i_t + \sum_t \sum_{j \neq i} p_j(t \mid s, \pi^j_s) v^i_t \\
&= \sum_t p(t \mid s, f^i_s(i), \pi_s^{-i}) v^i_t \leq \sum_t p(t \mid s, f^i_s(i), \pi_s^{-i}) \gamma^i(t, f^i(i), \pi^{-i}) \\
&= \gamma^i(s, f^i(i), \pi^{-i}).
\end{aligned}$$

This again yields that $(\pi^1, \ldots, \pi^n)$ is a pure equilibrium. □

Observe that, for the equilibria constructed in the previous proofs, the players play pure stationary while preventing any deviation by a constant threat of using a pure stationary punishment. The existence of pure stationary optimal strategies in the game $\Gamma(i)$ is of crucial importance in the above proofs, for this implies that the level player i can guarantee by playing $f^i(i)$ (i.e. the maxmin) is the same as the level at which his opponents can punish him by playing $f^{-i}(i)$ (i.e. the minmax). The following three-person game shows that maxmin and minmax need not be equal.

Example:

N

	L	R
T	0,0,0	1,1,1
B	1,1,1	1,1,1

F

1,1,1	1,1,1
1,1,1	0,0,0

Here there is only one state. Player 1 plays Top or Bottom, player 2 plays Left or Right, player 3 plays Near or Far. It can be seen that $\max_{x^1} \min_{x^2,x^3} \gamma^1(x^1, x^2, x^3) = 1/2 \leq 3/4 = \min_{x^2,x^3} \max_{x^1} \gamma^1(x^1, x^2, x^3)$. For perfect information games and for ARAT games a gap between maxmin and minmax cannot occur because the joint minmax strategy of the $n-1$ players is pure and stationary, and thus corresponds to $n-1$ pure stationary strategies for the individual players. The example shows that the above proof cannot be applied for 3-person switching control stochastic games.

Theorem 3: Every two-person switching control stochastic game has ε-equilibria.

Proof: By Filar [1981] we have:

$$\exists v^1, v^2\, \exists x, y\, \forall \sigma: \gamma^1(x^1, \sigma^2) \geq v^1 \geq \gamma^1(\sigma^1, x^2) \quad \text{and} \quad \gamma^2(\sigma^1, y^2) \geq v^2 \geq \gamma^2(y^1, \sigma^2).$$

Let g^2 be a pure stationary best reply for player 2 against x^1. Let R be the set of states that are recurrent w.r.t. (x^1, g^2). Let $\tilde{X}^1$ be the set of stationary strategies $\tilde{x}^1$ for player 1 with $\mathrm{Car}(\tilde{x}^1) \subseteq \mathrm{Car}(x^1)$ and for which in addition: a) $\tilde{x}^1_s = x^1_s$ for all $s \in R$, *b*) $|\mathrm{Car}(\tilde{x}^1_s)| = 1$ for all $s \notin R$ and *c*) all states in $S \setminus R$ are transient with respect to $(\tilde{x}^1, g^2)$. Let $\hat{x}^1 \in \tilde{X}^1$ be such that $\gamma^2(\hat{x}^1, g^2) \geq \gamma^2(\tilde{x}^1, g^2)$ for all $\tilde{x}^1 \in \tilde{X}^1$. Then also $\gamma^2(\hat{x}^1, g^2) \geq \gamma^2(x^1, g^2) \geq v^2$.

On the other hand, using the switching control structure, we also have $\gamma^1(\hat{x}^1, g^2) \geq v^1$. Moreover it can be verified that like in the previous proofs:

$$\forall s\, \forall a^1, a^2: \gamma^1(s, \hat{x}^1, g^2) \geq \sum_t p(t|s, a^1, g^2_s) v^1_t \quad \text{and}$$

$$\gamma^2(s, \hat{x}^1, g^2) \geq \sum_t p(t|s, \hat{x}^1_s, a^2) v^2_t,$$

which implies that the players can punish each other, whenever they observe any deviation. Since player 2 plays pure, his deviations can be observed immediately. Player 1, however, uses a mixed stationary strategy on R. There he could improve his reward by repeated deviations within the carrier of his strategy. For arbitrary $\varepsilon > 0$, such deviations can be observed by player 2 in the long run with probability at least $1 - \varepsilon$, since the action frequencies of player 1's actions should converge to $\hat{x}^1$ (cf. Thuijsman [1992]). Then, if we let π^1 consist of playing $\hat{x}^1$ and punishing by y^1 whenever necessary, and if we let π^2_ε consist of playing g^2 and punishing by x^2 whenever necessary, partly based on some ε-dependent testing, then $(\pi^1, \pi^2_\varepsilon)$ is an ε-equilibrium. □

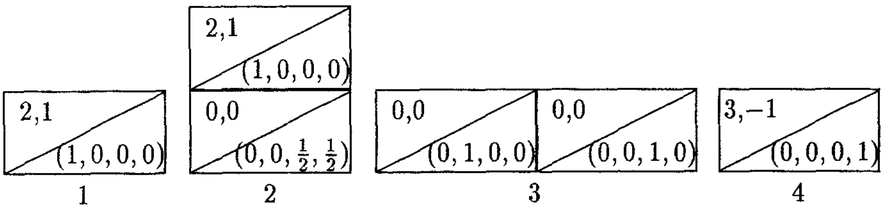
We remark that for these ε-equilibria both players are playing stationary while checking the opponent and threatening to use a stationary punishment. Notice that we would find a pure (0-)equilibrium in case $|\mathrm{Car}(x^1_s)| = 1$ for all $s \in R$.

Theorem 4: Perfect information-, ARAT- and two-person switching control games have the orderfield property.

Proof: The equilibria constructed for the perfect information case and for the ARAT case consist of pure strategies. Hence rational payoffs and transitions will give rational equilibrium rewards. For any two-person switching control game with rational data one can take rational stationary optimal strategies for the related zerosum games (cf. Filar [1981]). Since our ε-equilibria consist of these strategies combined with pure stationary strategies, the orderfield property holds. □

We remark that for zerosum ARAT stochastic games finite algorithms are known to compute pure stationary optimal strategies (cf. Raghavan et al. [1985]). The equilibria we constructed for the ARAT case (including perfect information), consist of zerosum optimal strategies. Hence, for these two cases the zerosum algorithms can straightforwardly be extended to yield finite algorithms for finding equilibria. For the two-person switching control case our approach first requires the determination of stationary optimal strategies, which can be done by the finite algorithm of Vrieze et al. [1983]. Next, this solution has to be combined with a solution to a perfect information game.

Example:



This example of a two-person perfect information game is taken from Flesch et al. [1996]. Player 1 plays rows, player 2 plays columns. Up left are the rewards for players 1 and 2 respectively; down right are the transition probabilities. Our construction gives the pure equilibrium: (π^1 with threat to play f^1, f^2), where $\pi^1 = (1, 0)$ and where $f^1 = f^2 = (0, 1)$. It is easy to verify that, for this example, there are no stationary ε-equilibria ($\varepsilon > 0$). However, for ARAT repeated games with absorbing states existence of stationary ε-equilibria has been shown by Evangelista et al. [1996].

References

Evangelista FS, Raghavan TES, Vrieze OJ (1994) Repeated ARAT games. In: Ferguson TS, Shapley LS (eds.) Statistics, probability and game theory, IMS Lecture Notes-Monograph Series, vol 30, pp. 13–28

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

Flesch J, Thuijsman F, Vrieze OJ (1996) Recursive repeated games with absorbing states. Math Oper Res 21: 1016–1022

Liggett TM, Lippman SA (1969) Stochastic games with perfect information and time average payoff. SIAM Review **11**: 604–607

Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory **10**: 53–66

Raghavan TES, Tijs SH, Vrieze OJ (1985) On stochastic games with additive reward and transition structure. J Opt Theory Appl **47**: 451–464

Thuijsman F (1992) Optimality and Equilibria in Stochastic Games. CWI-tract 82, Centre for Mathematics and Computer Science, Amsterdam

Vrieze OJ, Tijs SH, Raghavan TES, Filar JA (1983) A finite algorithm for the switching control stochastic game. OR Spektrum **5**: 15–24

Received September 1994
Revised version March 1995
Final version June 1996