International Journal of Game Theory, Vol. 15, Issue 2, page 101–107

Asymptotic Properties of a Non-Zero Sum Stochastic Game

By S. Sorin[1]

Summary: An example of a non-zero sum stochastic game is given where: i) the set of Nash Equilibrium Payoffs in the finitely repeated game and in the game with discount factor is reduced to the threat point; ii) the corresponding set for the infinitely repeated game is disjoint from this point and equals the set of feasible, individually rational and Pareto optimal payoffs.

I Introduction

The Folk Theorem states that the set of Nash Equilibrium Payoffs (NEP for short) in an infinitely repeated game coincides with the set D of individually rational payoffs that are feasible.

Moreover D contains the set of NEP in the game with discount factor λ and the set of NEP in the finitely repeated game. (For these results see e.g. Aumann.) On the other hand, for zero-sum stochastic games the value of the infinitely repeated game exists (Mertens/Neyman) and equals the limit of the values of both, discounted and finitely repeated games. Finally it can be shown (Mertens) that the set of NEP in non-zero sum λ discounted stochastic games has at least one limit point as λ goes to 0. It was thus a natural conjecture to ask whether this point was a NEP in the infinitely repeated game (also proving the non-emptiness of this set).

In fact this attempt fails and in the following example we show that neither the finitely repeated game nor the discounted game is a good approximation of the infinitely repeated game, as far as the equilibrium concept is concerned.

II The Example

Recall that a (two person) stochastic game is played in stages. At stage m, the players observe the current state z_m and choose some actions (i_m, j_m). The triple (z_m, i_m, j_m) determines the current (vector) payoff x_m at that stage and the probability according to which the new state is choosen. Hence a (behavioral) strategy is the random choice

[1] Sylvain Sorin, Université Louis Pasteur, Département de Mathématique, 7 rue René Descartes, 67084 Strasbourg, France.

0020–7276/86/020101–107 $2.50

of an action at each stage m, conditionally to the history $(z_1, i_1, j_1, \ldots, i_{m-1}, j_{m-1}, z_m)$ up to that stage. We denote by S and T the strategy sets of both players. We shall consider the stochastic game described by the following payoff matrix:

$$\begin{bmatrix} (1.0)^* & (0,2)^* \\ (0,1) & (1,0) \end{bmatrix}$$

where a star denotes an absorbing payoff (i.e. this entry, once reached, determines the payoff for all remaining stages).

Given (σ, τ) in $S \times T$, they induce a (vector) payoff $X_n(\sigma, \tau)$ in the n-stage game $G(n)$ where:

$$X_n(\sigma, \tau) = E_{\sigma,\tau}\left(\frac{1}{n} \sum_{m=1}^{n} x_m\right)$$

We denote by $E(n)$ the set of NEP in $G(n)$.

Similarly (σ, τ) induce a payoff $X_\lambda(\sigma, \tau)$ in the λ-discounted game $G(\lambda)$ with:

$$X_\lambda(\sigma, \tau) = E_{\sigma,\tau}\left(\lambda \sum_{m=1}^{\infty} (1-\lambda)^{m-1} x_m\right), \quad \text{for } \lambda \in (0, 1]$$

and $E(\lambda)$ is the set of NEP in $G(\lambda)$.

Note that Nash's theorem implies that $E(n)$ and $E(\lambda)$ are non-empty closed sets.

Now $E(\infty)$, set of NEP in the undiscounted infinitely repeated game $G(\infty)$ is defined as the set of payoffs $y = (y^1, y^2)$ satisfying:

$\forall \epsilon > 0$, $\exists\, \sigma_\epsilon \in S$, $\exists\, \tau_\epsilon \in T$, $\exists\, N \in \mathrm{N}$ such that, for all $\sigma \in S$, $\tau \in T$ and all $n \geqslant N$:

$$\left.\begin{aligned} X_n^1(\sigma, \tau_\epsilon) - \epsilon \leqslant y^1 \leqslant X_n^1(\sigma_\epsilon, \tau_\epsilon) + \epsilon \\ X_n^2(\sigma_\epsilon, \tau) - \epsilon \leqslant y^2 \leqslant X_n^2(\sigma_\epsilon, \tau_\epsilon) + \epsilon \end{aligned}\right\} \qquad (1)$$

(Recall that it is hopeless to look for "exact" Nash equilibrium since already in the zero-sum case only ϵ-optimal strategies may exist).

Let us now remark that the sets of feasible payoffs, namely:

$$C(n) = \{y \mid \exists\, (\sigma,,\tau) \text{ in } S \times T \text{ with } X_n(\sigma, \tau) = y\}$$

$$C(\lambda) = \{y \mid \exists\, (\sigma, \tau) \text{ in } S \times T \text{ with } X_\lambda(\sigma, \tau) = y\}$$

$$C(\infty) = \{y \mid \exists\, (\sigma, \tau) \text{ in } S \times T \text{ with } L(X_n(\sigma, \tau)) = y\}$$

(where L is a Banach limit) are equal to the following set:

$$C = \mathrm{Co}\,\{(1, 0), (0, 1), (0, 2)\}.$$

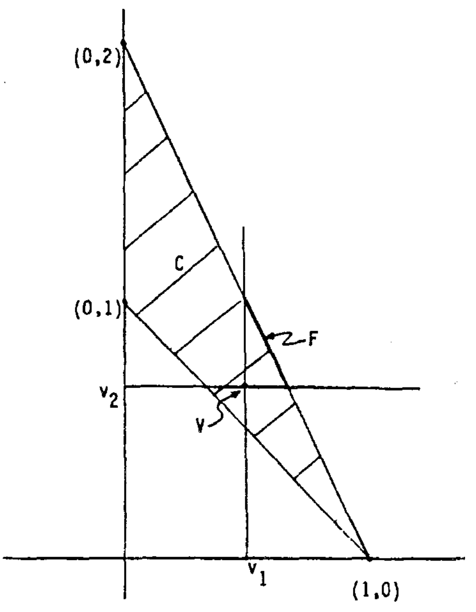
Finally the minimax, ν_1, for player 1 is given by the value of the following zero-sum game:

$$\begin{bmatrix} 1^* & 0^* \\ 0 & 1 \end{bmatrix}$$

which is exactly the "Big Match" of Blackwell/Ferguson. We recall that the value of this game is 1/2. An optimal strategy for player 2 is to play (1/2, 1/2) i.i.d. Player 1 has no optimal strategy but an ϵ optimal strategy can be constructed as follows: Let $M \geqslant 1/\epsilon$ and g_n be the sum of the (non-absorbing) payoff up to stage n; then player 1 plays Top at stage $n+1$ with probability $(M+g_n-n/2)^{-2}$. Intuitively player 1 will always play Top with a small probability; but this probability will decrease very fast if player 2 is using a frequency of Left less than 1/2 and increase otherwise. Similarly $\nu_2 = 2/3$ (the corresponding strategies being 2/3, 1/3) i.i.d. for player 2 and Top with probability $(M-g_n+2n/3)^{-2}$ for player 1) and we shall denote by $V=(\nu_1, \nu_2)$ the threat point.

It follows that the set F of feasible individually rational and Pareto optimal payoffs is given by

$$F = \{(a; 2(1-a)); 1/2 \leqslant a \leqslant 2/3\}$$



We can now state our results.

Theorem 1: $E(n) = E(\lambda) = \{V\}, \forall n \geqslant 1, \forall \lambda \in (0, 1]$.

Note that we obviously have $E(1) = \{V\}$ but this does not a priori even imply: $\{V\}$ included in $E(n)$ or in $E(\lambda)$, as in repeated games with only one state.

Theorem 2: $E(\infty) = F$.

III Proof of Theorem 1

Given (σ, τ) pair of equilibrium strategies in $G(n)$ or $G(\lambda)$ we denote by (s, t) the corresponding mixed actions of both players at the first stage, where s (resp. t) stands for the probability of Top (resp. Left).

It follows immediatly that $s = 1$ is impossible (P II would play $t = 0$) and similarly for $t = 1$ (P I would play $s = 1$).

We now consider the two classes of games.

1 Study of $E(\lambda)$

For a given λ, let w be the maximal equilibrium payoff for player 2 and let this payoff be obtained by some equilibrium pair (σ, τ).

Define w' (resp. w'') to be the normalized payoffs induced by (σ, τ) from stage 2 on, given the history (Bottom, Left) (resp. (Bottom, Right)) at stage 1.

Assume first $s = 0$. Since $t < 1, w''$ is a NEP for P II and the equality $w = (1 - \lambda)w''$ contradicts the definition of w.

On the other hand $t = 0$ would imply $s = 0$.

Hence we are left with the case where s and t belong to the open interval $(0, 1)$. The equilibrium conditions are thus:

$$w = (1 - s)(\lambda + (1 - \lambda)w') = 2s + (1 - s)(1 - \lambda)w''.$$

By definition of w and using the fact that now both w' and w'' are NEP we obtain:

$$(1 - s)(\lambda + (1 - \lambda)w) \geqslant w$$

$$(1 - s)(2 - (1 - \lambda)w) \leqslant 2 - w.$$

These inequalities give:

$$(2 - w)(\lambda + (1 - \lambda)w) \geqslant w(2 - (1 - \lambda)w)$$

hence:

$$2\lambda \geqslant 3\lambda w$$

Since $w \geqslant v_2$ we obtain $w = v_2$ and $E(\lambda)$ is included in the line $y_2 = v_2$.

Let now (σ, τ) be specified in order that player 1 achieves his maximum equilibrium payoff u. u' and u'' being defined like w' and w'' above we obtain:

$$u = t = t(1 - \lambda)u' + (1 - t)(\lambda + (1 - \lambda)u'')$$

and since u' and u'' are NEP for player 1 it follows that:

$$u \leqslant t(1 - \lambda)u + (1 - t)(\lambda + (1 - \lambda)u)$$

but $u = t$, hence:

$$u \leqslant u^2(1-\lambda) + (1-u)(\lambda + (1-\lambda)u)$$

which gives

$$2u\lambda \leqslant \lambda.$$

Thus we finally get $u \leqslant \frac{1}{2}$, hence $u = v_1$ and $E(\lambda) = \{V\}$ for all $\lambda \in (0, 1]$.

The equilibrium strategies in $G(\lambda)$ are given by: $s = \lambda/(\lambda + 2)$ and $t = 1/2$. (Note that these strategies are stationary.)

2 Study of $E(n)$

The analysis is roughly similar, except the fact that after one stage both players play in a different game, namely $G(n-1)$.

So let us define m such that $E(n)$ is included in $\{y_2 = v_2\}$ for all $n \leqslant m$. As above we introduce (σ, τ) corresponding to some NEP w of player 2 in $G(m+1)$, and (s, t) being the random actions at stage 1.

Here again, assuming $s = 0$ gives $(m+1)\,w = mv_2$ hence a contradiction, and similarly if $t = 0$.

The equilibrium conditions for s and t in $(0, 1)$ are now:

$$(m+1)w = (1-s)(1+mv_2) = 2s(m+1) + (1-s)mv_2.$$

$$v_2 = 2/3 \text{ gives } s = 1/(2m+3) \text{ and } w = 2/3.$$

An analogous computation for player 1 leads to:

$$(m+1)u = (m+1)t = (1-t)1 + mv_1$$

hence $v_1 = 1/2$ implies $t = 1/2$ and $u = 1/2$.

This shows by induction that for all $n \geqslant 1$ $E(n) = \{V\}$.

The equilibrium strategies for $G(n)$ are given by: $s = 1/(2n+1)$ and $t = 1/2$ at stage one, and then equilibrium in $G(n-1)$ if Bottom.

This ends the proof of Theorem 1.

IV Proof of Theorem 2

We shall split the proof into 2 parts.

Step 1: $E(\infty) \subset F$.

Obviously any NEP lies in C and Pareto dominates V.

It remains thus to show that it belongs to the line $[(1, 0), (0, 2)]$.

The idea of the proof is very simple: if the probability of getting an absorbing payoff on the equilibrium path, is less than one, then after some stage P I is essentially playing bottom; the corresponding feasible payoffs, from this stage on, are not individually rational, hence a contradiction.

In fact given $\bar{y}$ in $E(\infty)$ and $\epsilon > 0$, let $(\sigma_\epsilon, \tau_\epsilon)$ satisfy (1).

Define $\underset{\sim}{m}$ to be the stopping time corresponding to the first action Top of player 1 and note that on the event $\{\underset{\sim}{m} = +\infty\}$ the average payoff lies on the line $y_1 + y_2 = 1$, and in particular it is not individually rational.

Assume now that $p = \text{Prob}_{\sigma_\epsilon,\tau_\epsilon}(\{\underset{\sim}{m} = +\infty\})$ is strictly positive.

For any $\delta > 0$, we can introduce some N such that:

$$\text{Prob}(\{\underset{\sim}{m} = +\infty\} \mid \{\underset{\sim}{m} \geqslant N\}) \geqslant 1 - \delta.$$

Hence by deviating from stage N on, one player will gain at least

$$(v_1 + v_2 - 1)\frac{1}{2}(1 - \delta) - 2\delta, \quad \text{conditionally on } \{\underset{\sim}{m} \geqslant N\}.$$

For δ small enough this amount is greater than $1/13$; it follows now from (1) that p has to be less than some 14ϵ.

Hence $\bar{y}$ belongs to any neighbourhood of the Pareto boundary, corresponding to $\{\underset{\sim}{m} < +\infty\}$ and this achieves the proof.

Step 2: $F \subset E(\infty)$.

It remains now to prove that the points in F can actually be achieved as NEP (showing in particular that $E(\infty)$ is not empty).

Let $y = (a, 2(1-a))$ in F with $1/2 \leqslant a \leqslant 2/3$.

We describe the strategies as follows:

For player 2, τ is defined by playing left with probability a, i.i.d. at each stage. For player 1, σ is a δ-optimal strategy in the following game:

$$\begin{bmatrix} (1-a)^* & -a^* \\ -(1-a) & a \end{bmatrix}$$

as constructed in Blackwell/Ferguson or more generally in Mertens/Neyman.

Recall that such a strategy can be defined as follows: play optimally at stage n in the game with discount factor λ_n where λ_n is a function of the history defined by $\lambda_n = (M + g_n)^{-2}$.

This implies in particular that, if $\bar{t}_n$ denotes the frequency of Left up to stage n, there exists some N such that σ satisfies:

$$\text{Prob}_{\sigma,\tau}(\underset{\sim}{m} \leqslant N) \geqslant 1 - \delta \tag{2}$$

and for all $n \geqslant N$ and all τ' in T:

$$\text{Prob}_{\sigma,\tau'}(\underset{\sim}{m} \leqslant n) \leqslant 1 - \delta \Rightarrow E_{\sigma,\tau'}(\bar{t}_n \mid \underset{\sim}{m} > n) \leqslant a + \delta \tag{3}$$

$$\mathrm{Prob}_{\sigma,\tau'}(\underset{\sim}{m} \leqslant n)(\mathrm{Prob}\,(j_{\underset{\sim}{m}} = \text{Left} \mid \underset{\sim}{m} \leqslant n) - a) \geqslant -\delta^2. \tag{4}$$

Since a is greater than 1/2 it follows from (2) that σ is a best reply to τ up to 2δ.

Now given σ, either (3) holds hence the non absorbing average payoff is at most $a + \delta$ thus less then $2(1-a) + \delta$ (a being less than 2/3) or the probability of such a payoff is less than δ.

As for the absorbing part, either it has a weight greater than δ but by (4) the corresponding payoff is at most $2(1-a+\delta)$, or the absorbing probability is less than δ.

It follows easily that, given $\epsilon > 0$, by taking δ small enough the above (σ, τ) form an equilibrium pair associated to y.

V Concluding Remarks

The main feature of this example is the fact that $E(n)$ and $E(\lambda)$ are constant and disjoint from $E(\infty)$.

This implies that the difference between the infinite game and the two approximations cannot be reduced by taking a stronger concept of Equilibrium.

It is worthwhile to remark moreover that the NEP in $G(\infty)$ are precisely the "good" outcomes while $E(n)$ and $E(\lambda)$ are reduced to the threat point.

This last property exhibits another phenomena, already noticed by Aumann/Maschler: in $G(n)$ and $G(\lambda)$ both players are requested to play at equilibrium strategies which induce a payoff V, without guaranteeing it, while strategies guaranteeing it do exist for both (Typically player 2 has to play (1/2, 1/2) i.i.d. while (2/3, 1/3) i.i.d. is his minmax strategy.)

References

Aumann RJ (1981) Survey of repeated games. In: Essays in game theory and mathematical economical in honor of Oskar Morgenstern. Bibliographisches Institut, Mannheim Wien Zürich

Aumann RJ, Maschler M (1972) Some thoughts on the minimax principle. Management Science 18/5, part II:54–63

Blackwell D, Ferguson TS (1968) The big match. Annals of Mathematical Statistics 39:159–163

Mertens JF (1982) Repeated games: an overview of the zero-sum case. In: Hildenbrand W (ed) Advances in economic theory. Cambridge UP

Mertens JF, Neyman A (1981) Stochastic games. International Journal of Game Theory 10:53–66

Received June 1984
(Revised version February 1985)