STOCHASTIC GAMES WITH ZERO STOP PROBABILITIES

Dean Gillette

§1. INTRODUCTION

A stochastic game has been defined by Shapley [1], to be a game consisting of a finite collection of positions among which two players pass according to jointly-controlled transition probabilities. Thus, if at the $i^{\underline{th}}$ position, the players choose pure strategies k and ℓ; the probability of passing to position j is given by $p^{ij}_{k\ell}$ and the (finite) payoff to the first player is $a^{i}_{k\ell}$. Payments are to accumulate throughout the game.

Shapley has, in addition, required that, at each position, there is a positive probability $s^{i}_{k\ell}$ of stopping the game. We may include the stop probability within our framework by requiring the existence of a position j with $a^{j}_{k\ell} \equiv 0$ and with $p^{ij}_{k\ell} \geq \delta > 0$ and $p^{ji}_{k\ell} = 0$ for all i, k, ℓ.

In this paper, we shall investigate the consequences of relaxing the requirement that the game shall certainly end, i.e., we will allow the stop probability to be zero at some or all positions. It follows, of course, that the accumulated payoff may not be well defined. We introduce effective payoffs, defined over the entire game, and attempt to find optimum strategies for the players relative to these effective payoffs.

§2. STRATEGIES AND EFFECTIVE PAYOFFS

A strategy is taken to be a selection of a probability distribution on the alternatives at each position of the infinite game. It is not in general required that a given distribution be maintained on successive returns to a given position; a player is allowed to change his mind. These strategies are more properly labeled behavior strategies, but since we deal with no others, we will omit the modifier.

Of particular interest are stationary strategies, in which for each of the finite number of distinct positions, a probability distribution is specified for use when that position is reached, independent of where it might occur in the infinite game. A stationary strategy for the first player, on a game of M positions, may be represented by an M-tuple of probability distributions:

$$x = (x^1, x^2, \ldots, x^M), \quad \text{each} \quad x^i = \left(x^i_1, x^i_2, \ldots, x^i_{k_i} \right)$$

and similarly for the second player. We shall also use a pair (x, y) to denote non-stationary strategies for the two players, but shall specify the type of strategy where necessary.

The probability of passing from position i to position j and the payoff to player one at position i, given stationary strategies (x, y), are respectively:

$$P^1_{ij}(x, y) = \sum_{k,\ell} p^{ij}_{k\ell}\, x^i_k\, y^i_\ell$$

$$A_i(x, y) = \sum_{k,\ell} a^i_{k\ell}\, x^i_k\, y^i_\ell \quad .$$

Let P^0_{ij} denote the identity matrix, and for $n \geq 1$, let:

$$P^{n+1}_{ij}(x, y) = \sum_{m=1}^{M} P^n_{im}(x, y)\, P^1_{mj}(x, y) \quad .$$

The payoff that player one has accumulated after traversing $N + 1$ positions, $N \geq 0$, starting with position i, given stationary strategies (x, y), is:

$$H^N_i(x, y) = \sum_{n=0}^{N} \sum_{j=1}^{M} \left[P^n_{ij}(x, y) \right] \left[A_j(x, y) \right] \quad .$$

For non-stationary strategies, this formulation of $H^N_i(x, y)$ is, of course, not applicable; however, we shall retain the notation for the accumulated payment, independent of the type of strategy.

An <u>effective payoff</u> is some function of the payoffs accumulated position by position during the course of the game.

If the payoff after $n + 1$ positions have been played is discounted by $(1 - s)^n$, $0 < s < 1$, we write the effective <u>s-discounted payoff</u> for strategies (x, y) as:

$$D_i^s(x, y) = \sum_{n=0}^{\infty} (1 - s)^n \left(H_i^n(x, y) - H_i^{n-1}(x, y) \right)$$

where $H_i^{-1}(x, y) \equiv 0$.

We also define a limiting average payoff, $L_i(x, y)$, by:

$$L_i(x, y) = \liminf_{N \to \infty} \frac{H_i^N(x,y)}{N} .$$

The lim inf is taken to insure existence. However any convex combination of lim inf and lim sup may be taken without destroying the character of later proofs.

Shapley's results [1], may be interpreted to demonstrate the validity of the following theorem.

THEOREM 1. For the s-discounted effective payoff on a stochastic game there are stationary strategies (x^*, y^*) such that for all strategies (x, y) and all $i = 1, \ldots, M$:

$$D_i^s(x, y^*) \leq D_i^s(x^*, y^*) \leq D_i^s(x^*, y) .$$

Moreover, if the game is of perfect information, a solution exists in stationary pure strategies.

(Note that this theorem concerns stochastic games, possibly with stop probabilities, and with a discount. However, with the formulation of a stop position as in the Introduction above, Shapley's Theorem 2 and Application 2 are applicable.)

In order to obtain an analogous result for the limiting average payoff, some lemmas are required.

§3. PRELIMINARY RESULTS

LEMMA 1. Suppose f_n is a sequence of functions converging uniformly on $X \times Y$ to a function f. If for each n there exist $(x_n, y_n) \in X \times Y$ such that for all $(x, y) \in X \times Y$:

$$\text{(A)} \qquad f_n(x, y_n) \leq f_n(x_n, y_n) \leq f_n(x_n, y)$$

then:

(i) $$\inf_{y\epsilon Y}\sup_{x\epsilon X} f(x, y) = \sup_{x\epsilon X}\inf_{y\epsilon Y} f(x, y) \quad .$$

Moreover, for each $\epsilon > 0$ there is an n such that if (x_n, y_n) satisfies (A) then for all $m \geq n$:

(ii) $$\begin{aligned} f_m(x_n, y_n) - \epsilon &< f_m(x_n, y) \\ f_m(x_n, y_n) + \epsilon &> f_m(x, y_n) \quad . \end{aligned}$$

Choosing n so that for all $m \geq n$ and for all $x, y,$ $|f_m(x, y) - f(x, y)| < \frac{\epsilon}{4}$, one may easily verify (ii) from which (i) follows.

LEMMA 2. If a_n is a sequence of non-negative numbers, then:

$$\limsup_{N \longrightarrow \infty} \frac{1}{N} \sum_{n=1}^{N} a_n \leq \limsup_{s \longrightarrow 0^+} s \sum_{n=1}^{\infty} (1 - s)^{n-1} a_n$$

$$\liminf_{N \longrightarrow \infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geq \liminf_{s \longrightarrow 0^+} s \sum_{n=1}^{\infty} (1 - s)^{n-1} a_n \quad .$$

A proof of this generalization of a theorem of Hardy and Littlewood may be constructed along the same lines as the proof of that theorem as given by Titchmarsh ([3], p. 227).

LEMMA 3. If $P^1 = \{p_{ij}^1\}$ is an M-dimensional stochastic matrix, with P^n the n^{th} power of P, then there is a stochastic matrix, $Q = \{q_{ij}\}$, such that:

$$\lim_{N \longrightarrow \infty} \frac{1}{N} \sum_{n=1}^{N} P^n = Q \quad .$$

Moreover, if there is an integer m and a number $\delta > 0$ such that

$$\min_{ij} p_{ij}^m \geq \delta$$

then the numbers q_{ij} are independent of i and

$$\left| \frac{1}{N} \sum_{n=0}^{N} (p_{ij}^n - q_{ij}) \right| \leq \frac{1}{N} \quad \frac{1}{(1-M\delta)(1-(1-M\delta)^{1/m})} \quad .$$

These results, from the theory of Markov processes, may be obtained by a slight extension of the results given in Doob ([2], pp. 173-175).

§4. THE LIMITING AVERAGE EFFECTIVE PAYOFF

We impose the condition that the payoffs to player one be non-negative. From the definition of the limiting average effective payoff, we see that this results in no loss of generality.

We first concentrate on stationary strategies. From the definitions and Lemma 3, we may write:

$$L_i(x, y) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{M} \left[P_{ij}^n(x, y) \right] \left[A_j(x, y) \right] \quad .$$

It follows by Lemma 2 that:

$$L_i(x, y) = \lim_{s \to 0^+} s \, D_i^s(x, y) \quad . \tag{B}$$

There is always a solution for the s-discounted payoff, consequently, from Lemma 1 there are at least ϵ-best strategies in the limiting average effective payoff provided the passage to the limit is uniform over a set of (x, y) simultaneously containing solutions of an infinite sequence of s-discounted payoffs with the discount approaching zero. We consider several cases.

CASE 1. Perfect Information. The convergence in (B) is uniform over the (finite) set of all pure stationary strategies. Since this set also contains solutions for the s-discounted games for all s, we find that there are pure stationary strategies x^*, y^* so that for all pure stationary strategies (x, y) for all i:

$$L_i(x^*, y) \geq L_i(x^*, y^*) \geq L_i(x, y^*) \quad .$$

Now, let (x, y) be any pair of strategies, stationary or not. From the definitions of the effective payoffs and with the aid of Lemma 2, we have:

$$L_i(x, y^*) \leq \limsup_{N \to \infty} \left[\frac{1}{N} H_i^N(x, y^*) \right] \leq$$

$$\limsup_{s \to 0^+} s \sum_{n=1}^{\infty} (1 - s)^{n-1} \left(H_i^n(x, y^*) - H_i^{n-1}(x, y^*) \right) =$$

$$\limsup_{s \to 0^+} \left[s\, D_i^s(x, y^*) \right] \leq \limsup_{s \to 0^+} \left[s\, D_i^s(x^*, y^*) \right] = L_i(x^*, y^*) =$$

$$\liminf_{s \to 0^+} \left[s\, D_i^s(x^*, y^*) \right] \leq \liminf_{s \to 0^+} \left[s\, D_i^s(x^*, y) \right] =$$

$$\liminf_{s \to 0^+} s \sum_{n=1}^{\infty} (1 - s)^{n-1} \left(H_i^n(x^*, y) - H_i^{n-1}(x^*, y) \right) \leq$$

$$\liminf_{N \to \infty} \frac{1}{N} H_i^N(x^*, y) = L_i(x^*, y) \quad .$$

Thus, we have:

> THEOREM 2. For the limiting average effective pay-off on a stochastic game of perfect information there are pure stationary strategies (x^*, y^*) such that for all strategies x and y, all $i = 1, \ldots, M$:
>
> $$L_i(x^*, y) \geq L_i(x^*, y^*) \geq L_i(x, y^*) \quad .$$

CASE 2. Cyclic Stochastic Games. A cyclic stochastic game is one in which there exists a $\delta > 0$ and $N > 1$ such that:

$$\min_{\substack{i,\, j,\, k_m,\, l_m \\ 1 \leq m \leq N}} \left\{ \sum_{j_1, \ldots, j_{N-1}} \left[p_{k_1 \ell_1}^{i j_1} \, p_{k_2 \ell_2}^{j_1 j_2} \cdots p_{k_N \ell_N}^{j_{N-1} j} \right] \right\} = \delta \quad .$$

Here, independent of the choice of strategies, there is a positive probability of passing from position i to position j in exactly N moves. In particular, there can be no "absorbing subsets" of the positions of the stochastic game nor is there a stop position.

It is clear that the transition probabilities $P_{ij}^1(x, y)$, for a cyclic game satisfy the conditions of the second part of Lemma 3, simultaneously in all stationary strategies (x, y). As a result of this, it may be shown that, for stationary strategies, the passage to the limit (B) is uniform over all stationary strategies. Consequently, from Lemma 1 and Theorem 1, for stationary strategies (x, y) and all i:

$$\text{(C)} \qquad \inf_y \sup_x L_i(x, y) = \sup_x \inf_y L_i(x, y) \quad .$$

Using the norm

$$\|x - \bar{x}\| = \sum_{i,k} |x_k^i - \bar{x}_k^i| \quad ,$$

it may be shown that the functions $s\, D_i^S(x, y)$ are continuous in x and y. From the uniform convergence, it follows that $L_i(x, y)$ is continuous so that the inf and sup of (C) may be replaced by min and max. Using this fact and using Lemma 2 as in Case 1 we have:

THEOREM 3. For the limiting average effective payoff on cyclic stochastic games, there are stationary strategies x^* and y^* such that for all i:

$$L_i(x, y^*) \leq L_i\ (x^*, y^*) \leq L_i(x^*, y) \quad .$$

<u>A Counter example for the General Case</u>. It is not in general the case that the unrestricted stochastic game will have a solution in stationary strategies for the limiting average effective payoff. Consider the following probabilities and payoffs for a 3-position game.

$$p_{k\ell}^{11} = \begin{vmatrix} 1 & 0 \\ 1 & 0 \end{vmatrix}; \quad p_{k\ell}^{12} = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}; \quad p_{k\ell}^{13} = \begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix}$$

$$p_{k\ell}^{21} = |0| \quad ; \quad p_{k\ell}^{22} = |1| \quad ; \quad p_{k\ell}^{23} = |0|$$

$$p_{k\ell}^{31} = |0| \quad ; \quad p_{k\ell}^{32} = |0| \quad ; \quad p_{k\ell}^{33} = |1|$$

$$a_{k\ell}^{1} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}; \quad a_{k\ell}^{2} = |0| \quad ; \quad a_{k\ell}^{3} = |1| \quad .$$

It may be verified that for stationary strategies:

$$\min_y \max_x L_1(x, y) = 1$$

$$\max_x \min_y L_1(x, y) = \tfrac{1}{2}$$

Here, we have a non-cyclic game (position 2 may be considered a stop position or a repeating position with zero payoff). Convergence in (B) is non-uniform and the function $L_i(x, y)$ is discontinuous in y at $y^1 = (1, 0)$.

§5. REMARKS

1. The fact that a solution exists in stationary strategies is dependent upon the effective payoff. Consider the following two-position game:

$$p_{k\ell}^{11} = |0\ 0| \qquad p_{k\ell}^{12} = |1\ 1|$$

$$p_{k\ell}^{21} = \left|\begin{smallmatrix}1\\1\end{smallmatrix}\right| \qquad p_{k\ell}^{22} = \left|\begin{smallmatrix}0\\0\end{smallmatrix}\right|$$

$$a_{k\ell}^{1} = |1\ -1| \qquad a_{k\ell}^{2} = \left|\begin{smallmatrix}-1\\1\end{smallmatrix}\right| \quad .$$

If we let the effective payoff be

$$E_i(x, y) = \limsup_{N \longrightarrow \infty} - |H_i^N(x, y)|$$

it may be verified that for any stationary x,

$$\min_y E_i(x, y) = -\infty, \qquad i = 1, 2,$$

but that there exists a non-stationary strategy $\bar{x}$, that gives:

$$\min_y E_1(\bar{x}, y) = 0, \qquad \min_y E_2(\bar{x}, y) = -1 \quad .$$

(The maximizing player should choose the same alternative number as did his opponent at the previous move.)

We note also that here the addition of a constant to the payoff may seriously affect the results of the game.

2. In general, the <u>value</u> of a stochastic game depends on the starting position. However, as may be deduced from Lemma 3, with the limiting average effective payoff for a cyclic game, the value is independent of the starting position.

3. It is easy to show that if there are finite sets X and Y such that if for all $s \leq s_0$, optimal strategies of the s-discounted payoff lie in (X, Y), then there is an $s_1 \leq s_0$ and pair (x, y) in (X, Y) that is uniformly the solution for the s-discounted payoff, $s \leq s_1$, and for the limiting average payoff.

In particular, games of perfect information fall into this category.

One may ask if there is an analogous uniformity of solution in cyclic games. The following example (due to Shapley) answers the question in the negative:

$$A^1 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \qquad A^2 = |0|$$

$$p^{11} = \begin{vmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{vmatrix} \quad p^{21} = |1/2|$$

$$p^{12} = \begin{vmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{vmatrix} \quad p^{22} = |1/2| \quad .$$

Here

$$D_1^s(\bar{x}, \bar{y}) = \frac{1 + s}{4s}, \; D_2^s(\bar{x}, \bar{y}) = \frac{1 - s}{4s}, \; \bar{x}^1 = \left(\frac{7 - s}{12}, \frac{5 + s}{12}\right), \; \bar{y}^1 = (1/2, 1/2),$$

where $\bar{x}$, $\bar{y}$ are the unique optimal strategies.

BIBLIOGRAPHY

[1] SHAPLEY, L. S., "Stochastic games," Proc. Nat. Acad. Sci., 39, (1953), pp. 1095-1100.

[2] DOOB, J. C., Stochastic Processes, Wiley, New York, 1953.

[3] TITCHMARSH, E. C., The Theory of Functions, 2nd Ed., Oxford, 1939.

Dean Gillette

Bell Telephone Laboratories, Inc.