A RELATIONSHIP BETWEEN SINGLE CONTROLLER STOCHASTIC GAMES AND CERTAIN MATRIX GAMES

Kumari Julee

Research Scholar,
Univ. Dept. of Math., T.M. Bhagalpur University, Bhagalpur, Bihar.

Abstract : A stochastic game on finitely many stages with limiting average payoff. Here we assume that the law of motion depends on the actions of one player only. It is shown that the value of the stochastic game for each state is the same as the value of the matrix game whose rows and columns are the pure stationary strategies of the players. In this paper we established a relationship between single controller stochastic games and certain matrix games.

Key words: stochastic game, matrix game, players, probability, Markovian decision process.

INTRODUCTION

A stochastic game on finitely many stages with limiting average payoff. Here we assume that the law of motion depends on the actions of one player only. It is shown that the value of the stochastic game for each state is the same as the value of the matrix game whose rows and columns are the pure stationary strategies of the players. A relationship between single controller stochastic games and certain matrix games is also established.

We shall now define the stochastic game more precisely. In a stochastic game the play proceeds by steps from position to position according to transition probabilities (changing from one step to another) controlled jointly by the two players.

We shall assume a finite number N of positions and finite number m_k, n_k of choices of each position. If, when at position k, the players choose their i^{th} and j^{th} alternatives, respectively, then with probability $s_{ij}^k > 0$ the game stops, while with probability p_{ij}^{kl} the game moves to position 1. We define

$$s = \min_{k,i,j} s_{ij}^k$$

Since s is positive, the game ends with probability 1 after a finite number of steps, because for any number t, the probability, that it has not stopped after t steps is not more than $(1 - s)^t$.

Payments accumulate throughout the course of play : the 1st player takes a_{ij}^k from the IInd whenever the pair i, j is chosen at position k.

If we define the bound M:

$$M = \min_{k,i,j} \left| a_{ij}^k \right|$$

then we see that the expected total gain or loss is bounded by

$M + (1 - s) M + (1 - s)^2 M + \ldots = M/s$

The process therefore depends on the following matrices

$$P^{kl} = (p_{ij}^{kl} \mid i = 1, 2, ..., m_k;\ j = 1, 2, ..., n_k)$$

$$A^{kl} = (a_{ij}^{kl} \mid i = 1, 2, ..., m_k;\ j = 1, 2, ..., n_k)$$

with $k = 1, 2, ..., N$, with elements satisfying

$$p_{ij}^{kl} \geq 0, \left| a_{ij}^{kl} \right| \leq M$$

$$\sum_{l=1}^{n} p_{ij}^{kl} = 1 - \left| a_{ij}^{kl} \right| \leq 1 - s < 1$$

A stochastic game is in a state s on a given day if players I and II play the matrix game

$$A_s = \left(a_{ij}^s \right)_{i,j=1}^{m,n}$$

on that day.

We shall assume that there are S such states. The probability that the game moves from state s to state s' tomorrow is given by $q(s'/s, i, j)$ where i, j are the row and column of A, chosen by players I and II, respectively. In general, player's strategies will depend on the complete past histories. However, we will be concerned primarily with the special class of strategies called stationary strategies. A stationary strategy, f, for player I consists of S probability vectors

$$f(s) = (f_1(s), f_2(s), \ldots, f_m(s)),$$

where $s = 1, S$, and $f_1(s)$ denotes the probability that player I will choose the i^{th} row of A, whenever the game is in state s. Player II's stationary strategies are similarly defined.

Let F and G be the sets of all stationary strategies for players I and II. A stationary strategy which is degenerated at each s is called pure stationary strategy.

Any such pure stationary strategy a for player I can be denoted by an S-tuple $\sigma = (i(1), i(2), \ldots, i(S))$ with the understanding that player I chooses row $i(s)$ whenever state s is reached.

Similarly we denote by v = (j(1), j(2), ..., j(S)) any pure stationary strategy for player II. There are $t=\prod_{s-1}^{s} m_{s'}$, such strategies for I and $p=\prod_{s-1}^{s} n_s$ such strategies for Π.

Once we specify the initial state and a strategy pair (f, g) for the two players, we have a probability distribution over all sequences of states and actions which can occur during the game and consequently over all the sequences of payoffs to player I.

Let $\prod_n (f,g,s)$ = expected income to player I on the nth day when the players use the strategy pair (f, g) and the game begins at state s.

The undiscounted pay-off by player II to player I is defined by

$$\phi(f,g)(s)=\liminf_{n\to\infty}\left(\frac{1}{N}\right)\sum_{n=1}^{N}\prod_n (f,g,s) \quad (1)$$

We shall denote by $v(s)$ the value of the undiscounted stochastic game starting in state s. We impose the following restriction on the law of motion: Player II alone controls the transition probabilities. That is

$q(s'/s, i, j)\ q(s'/s, j)$ for all i, j, s, s' (2)

Unless stated otherwise, the above restriction is implicitly assumed in this section.

In an undiscounted stochastic game restricted by assumption equation (2) the value and a pair of optimal stationary strategies exist.

It should be clear that a choice of a stationary strategy, g, by player II determines an $S \times S$ matrix

$$Q(g) = (q(s'/s, g)),$$

where $q(s'/s,g)=\sum_{j=1}^{n_s} q(s'/s,j)gj(s)$ for all s, s'.

Further, it is known that there exists a matrix

$$Q^*(g)=\lim_{n\to\infty}(1/(N+1))\sum_{n=0}^{N} Q^n(g) \quad (3)$$

where $Q^0(g) = I$, the identity matrix.

Now, if $(f, g) \in F \times G$ we define a current pay-off vector associated with this pair of strategies by

$$r(f,g)=(r(f,g,1),r(f,g,2),...,r(f,g,S))$$

where for each S

$$r(f,g)(s)=\sum_{i=1}^{m_s}\sum_{j=1}^{n_s} a_{ij}^s f_i(s) g_j(s) \quad (4)$$

To simplify the already complex notation we shall not differentiate between row and column vectors.

Lemma 1. $\phi(f,g)=Q^*(g)r(f,g)$ for all $(f, g) \in F \times G$

It should be noted that an undiscounted stochastic game with a stationary strategy for one of the players fixed is equivalent to an average reward Markovian decision process (AMD process) as far as the other player is concerned. We shall have the following well-known theorem in Markovian decision theory.

Theorem 2. In an AMD process there exists a pure stationary optimal policy.

In this section we shall show that $\phi(f, g)$ is linear in f or each stationary g. The following conventions will be adopted.

We recall that a pure stationary strategy, σ, for player I is an S-tuple $i(1), i(2), \ldots, i(S)$, where $i(S)$ corresponds to the row chosen by I in the s^{th} state whenever this state is visited. Since there are t such S-tuples, these pure stationary strategies for I can be labelled $\sigma_1, \sigma_2, \ldots, \sigma_t$ according to lexicographical ordering. That is

σ_1 = (1, 1, ..., 1),

σ_2 = (1, 1, ...1, 2), etc.

Now given a probability vector $\xi = (\xi_1, \xi_2, \ldots, \xi_t)$ we can construct a random vector $X = (x_1, x_2, ..., x_s)$ which takes values k with probabilities ξ_t. That is

$Pr\{X = \sigma_k\} = \xi_k$, k = 1, 2, ..., t (5)

The marginal distribution of X induces a unique stationary strategy, f, for player I defined by

$$f_i(s)=\sum_{\{k:\sigma_k(s)=i\}} \xi_k \quad (6)$$

The summation above is over all k's which correspond to these pure stationary strategies that select i^{th} row in s^{th} state. Conversely, given a stationary strategy, f, for I we can construct a random S-vector X taking values $\sigma_1, \sigma_2, \ldots, \sigma_t$ according to the joint distribution.

$$\Pr\{X=\sigma_k\}=\prod_{S=1}^{S} f_{i(s)}(s) \quad (7)$$

where $\sigma_k = (i(1), i(2), \ldots, i(S))$.

Lemma 3. Fix any stationary strategy g for player II. Then

(a) $\phi(f,g)=Q^*(g)r(f,g)$ for every $f \in F$, where $\xi = \{\xi_1, \xi_2, \xi_3, ..., \xi_t\}$ is constructed via (7).

(b) If $\xi = \{\xi_1, \xi_2, \xi_3, ..., \xi_t\}$ is any probability vector and $f \in F$ is constructed as in equation (6) then the equation in (a) bolds again.

Proof. Lemma 1 and the fact that $Q^*(g)$ is independent of f imply that it is sufficient to show that under the conditions of either (a) or (b)

$$\sum_{i=1}^{t}\xi_k r(\sigma_k,g)(s)=r(f,g)(s) \text{ for every } s = 1, 2, ..., S. \quad (8)$$

It is easy to check the validity of equation (8) in the case where S = 2 and A_2 are 2 × 2 matrices. The argument for the general case follows the same lines.

MATRIX GAMES

It is easy to check the validity of equation (8) in the case where $S = 2$ and A_2 are 2×2 matrices. The argument for the general case follows the same lines.

Let $v_1, v_2, v_3, \ldots, v_p$ be the enumeration of all pure stationary strategies for player II. We define the following set of *txp* matrices:

$$A(s) = \phi(\sigma, v_j)(s)_{i,j=1}^{t,p} \qquad s = 1, 2, \ldots, S. \qquad (9)$$

It turns out that these matrix games contain some relevant information about the original stochastic game as seen from the following theorems.

Theorem 4. If $v(s)$ is the value of the stochastic game starting at state s, then

$v(s) = \text{val}\,[A(s)]$ for each $s = 1, 2, .., S$.

Further, there exists a common optimal strategy for player I for all the matrix games $A(s)$.

Proof. There exists $(f^0, g^0) \in F \times G$ optimal for players I and II in the stochastic game.

By Lemma 3 part (a) for every s

$$\phi(f^0, g)(s) = \sum_{i=1}^{t} \xi_1^0 \phi(\sigma_i, g)(s) \geq v(s) \quad \text{for all } g \in G \qquad (10)$$

Thus val $[A(s)] \geq v(s)$ for all (s).

Let if possible $\text{val}[A(s)] \geq v(\bar{s})$ for some state $\bar{s}$. Then any optimal mixed strategy $\bar{\xi}$ for player I in the matrix game $A(\bar{s})$ induces a stationary strategy $\bar{f}$ as in equation (6) such that by lemma 3 (part(b))

$$\phi(\bar{f}, v_j)(\bar{s}) = \sum_{i=1}^{t} \bar{\xi}_i \phi(\sigma_i, v)(\bar{s}) \geq v(s) \quad \text{for all } v_j \qquad (11)$$

By theorem 2.2 $\min_{(g)} \phi(\bar{f}, g)$ is achieved at a pure stationary strategy, say v_{j^0}. Thus equation (11) implies that

$$\phi(\bar{f}, g^0)(\bar{s}) \geq \phi(\bar{f}, v_j)(\bar{s}) > v(\bar{s}).$$

However this, contradicts the optimality of g^0 in the stochastic game. The second part of the theorem follows from equation (2) since the mixed strategy $(\xi_1^0, \xi_2^0, \xi_3^0, \ldots, \xi_1^0)$ induced by f^0 by equation (7) is now optimal for player I in all the matrix games $A(s)$.

It is now clear that with any ξ^* which is a common optimal for player I in the matrix games $A(s)$ we can associate a stationary strategy f^* by equation (6) which is optimal for player I in the stochastic game.

The above theorem is not true in general as is easily seen from the following example:

$$S = 1 \quad \begin{bmatrix} 2 & 3 \\ 6 & 0 \end{bmatrix} \begin{matrix} \to 1 \\ \to 2 \end{matrix} \qquad\qquad S = 2 \quad \begin{matrix} \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \\ \downarrow \;\; \downarrow \\ 1 \;\; 2 \end{matrix}$$

The arrows indicate the state s reached when player I(II) chooses a given row (column). For instance, if I chooses row 2 in state 1, the game moves to state 2.

A pair of optimal stationary strategies for this game is (f^0, g^0),

where $f^0 = ((6/7, 1/7), (13/33, 20/33))$ and

$g^0 = ((19/33, 14/33), (4/5, 1/5))$.

The value of the stochastic game is 80/33 in both states. However the value of

$$A(1) = \begin{bmatrix} 2 & 2 & 3 & 3 \\ 2 & 2 & 3 & 3 \\ 4 & 0 & 1 & 0 \\ 7/2 & 4 & 1/2 & 4 \end{bmatrix}$$

is 43/18 with

$x^0 = (0, 27/36, 1/36, 8/36)$ and

$y^0 = (9/18, 2/18, 7/18, 0)$

being a pair of optimal strategies in the matrix game $A(1)$. Theorem 4 does hold a certain other special classes of stochastic games.

In a matrix game player II(I) can decide whether a given action can be eliminated without any loss to him by simply suppressing the corresponding column (row) and testing whether the value of the new matrix game has increased (decreased). In view of theorem 4 an analogous procedure could be applied to single controller stochastic games. For instance, the elimination of the k^{th} column of as corresponds to the elimination of a whole "block" of columns of $A(s)$ for each s.

Clearly, if the values of the so truncated games $A(s)$, $s = 1, 2, s$ have not changed then the above column can be eliminated without loss to player II. The next theorem implies the less obvious fact that it is sufficient to compare the values of $A(s^*)$ and $A(s^*)$ only.

We prove this theorem for a class of games larger than the single controller games.

Theorem 5. Let $\bar{\Gamma}$ be the stochastic subgame obtained from the original game Γ by deletion of a column (row) in state S^*. Further, assume that in both $\bar{\Gamma}$ and Γ the players possess optimal stationary strategies.

If $\bar{v}(s)$ denotes the value of the game $\bar{\Gamma}$ starting at t, then $\bar{v}(s^*) = v(s^*)$ implies that $\bar{v}(s) = v(s)$ for every s.

Proof. Let (f^0, g^0) and (f^*, g^*) be two pairs of optimal stationary strategies in Γ and $\bar{\Gamma}$, respectively.

Further, let $\hat{g}$ be the following nonstationary strategy for player II in Γ : he plays g^0 as long as s^* is not reached, as soon as s^* is reached he switches over permanently to the strategy g^*.

Correspondingly, let $\hat{f}$ be that strategy for player I in Γ which uses f^* as long as s^* is not reached and which switches to f^0 as soon as s^* is reached.

The strategy pair (f^*, g^0) partitions the states into irreducible chains $C_1, C_2, ..., C_k$ and a set of transient states H. Suppose that there exists some $\bar{s}$ for which $\bar{v}(\bar{s}) > v(\bar{s})$ (note that $\bar{v}(\bar{s}) > v(s)$ for every s since player II has one action less in $\bar{\Gamma}$), then

$$\phi(f^*, g)(\bar{s}) > v(\bar{s}) \text{ for all } g, \tag{12}$$

where $\bar{\phi}$ denotes the undiscounted pay-off in F we shall show that equation (12) is impossible.

Case (i)

The state s^* is non transient under (f^*, g^0) without loss of generality, $(s^* \in C_k)$ and assume that the starting state is $\bar{s}$.

Let T_i be the random variable which denotes the first time the game enters the state

$S^i \in C_i$, $i = 1, 2, ..., K$.

There is no loss of generality in the assumption, $s^k = s^*$.

Let $P_{f^*g^0}(T_i < \infty) = \Pr\{T_i < \infty |$ the pair (f^*, g^0) is used) then

$$\phi(f^*, \hat{g})(\bar{s}) = \sum_{i=1}^{k-1} \phi(f^*, g^0)(s^i) P_{f^*g^0}(T_i < \infty) + \phi f(f^*, g^*)(s^*) P_{f^*g^0}(T_k < \infty) \tag{13}$$

$$\phi(\hat{f}, g^0)(\bar{s}) = \sum_{i=1}^{k-1} \phi(f^*, g^0)(s^i) P_{f^*g^0}(T_i < \infty) + \phi f(f^0, g^0)(s^*) P_{f^*g^0}(T_k < \infty) \tag{14}$$

Since we are using average payoffs, equations (13) and (14) are intuitively obvious; they can be proved rigorously, by arguments. From the above and the hypothesis $\bar{v}(s^*) > v(s^*)$ together with the definitions of $(\hat{f}, \hat{g})(f^*, g^*)$ and (f^0, g^0) we have

$$\begin{aligned} &\phi(f^*, \hat{g})(\bar{s}) = \phi(f^*, \hat{g})(\bar{s}) \\ &\phi(\hat{f}, g^0)(\bar{s}) \le v(\bar{s}) \end{aligned} \tag{15}$$

which contradicts equation (2.20)

Case (ii)

The state s^* is transient under (f^*, g^0) and $\bar{s}$ is the initial state. Let U_i be the first time the game enters $s^i \in C_i$, without passing through

S^*, $i = 1, 2, k$.

Let U_{k+1} be the first time the game reaches S^*. Since s^* is transient under (f^*, g^0).

$$\begin{aligned} 1 &= \sum_{i=1}^{k+1} P_{f^*,\hat{g}}(U_i < \infty) \\ &= \sum_{i=1}^{k+1} P_{\hat{f},g^0}(U_i < \infty) \\ &= \sum_{i=1}^{k+1} P_{f^*,g^0}(U_i < \infty) \end{aligned} \tag{16}$$

A proof analogus to the previous case can now be given using U_i's in place of T_i's.

REFERENCES

[1] Dresher, M.; Shapley, L.S. and Tucker, A.W.Ed. : Advances in game theory Ann. of Math Studies No. 52, Princeton Univ. Press, Princeton (1964).

[2] Kuhn, H.W. and Tucker : Contributions of theory of games V-1 Ann. A.W. Ed. of Math Studies, 24 Princeton Univ. Press, Princeton, 1950.

[3] Mickinsey, J.C.C. : Introduction to the theory of Game, Mac-Graw Hill Book Co. New York, 1952.

[4] Owein, G. : Multilinear Extension of Games, Management Science 18 (5) Part-2, (1872).