*STOCHASTIC GAMES**

BY L. S. SHAPLEY

PRINCETON UNIVERSITY

Communicated by J. von Neumann, July 17, 1953

Introduction.—In a stochastic game the play proceeds by steps from position to position, according to transition probabilities controlled jointly by the two players. We shall assume a finite number, N, of positions, and finite numbers m_k, n_k of choices at each position; nevertheless, the game may not be bounded in length. If, when at position k, the players choose their ith and jth alternatives, respectively, then with probability $s_{ij}^k > 0$ the game stops, while with probability p_{ij}^{kl} the game moves to position l. 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 first player takes a_{ij}^k from the second whenever the pair i, j is chosen at position k. If we define the bound M:

$$M = \max_{k, i, j} |a_{ij}^k|,$$

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

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

The process therefore depends on $N^2 + N$ matrices

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

$$A^k = (a_{ij}^k | i = 1, 2, \ldots, m_k; j = 1, 2, \ldots, n_k),$$

with $k, l = 1, 2, \ldots, N$, with elements satisfying

$$p_{ij}^{kl} \geq 0, \ |a_{ij}^k| \leq M, \ \sum_{l=1}^{N} p_{ij}^{kl} = 1 - s_{ij}^k \leq 1 - s < 1.$$

By specifying a starting position we obtain a particular game Γ^k. The term "stochastic game" will refer to the collection $\Gamma = \{\Gamma^k | k = 1, 2, \ldots, N\}$.

The full sets of pure and mixed strategies in these games are rather cumbersome, since they take account of much information that turns out to be irrelevant. However, we shall have to introduce a notation only for certain behavior strategies,[1] namely those which prescribe for a player the same probabilities for his choices every time the same position is reached, by whatever route. Such *stationary strategies*, as we shall call them, can be represented by N-tuples of probability distributions, thus:

$$\vec{x} = (x^1, x^2, \ldots, x^N), \qquad \text{each } x^k = (x_1^k, x_2^k, \ldots, x_{m_k}^k),$$

for the first player, and similarly for the second player. This notation applies without change in all of the games belonging to Γ.

Note that a stationary strategy is not in general a mixture of *pure* stationary strategies (all x_i^k zero or one), since the probabilities in a behavior strategy must be uncorrelated.

Existence of a Solution.—Given a matrix game B, let val$[B]$ denote its minimax value to the first player, and $X[B]$, $Y[B]$ the sets of optimal mixed strategies for the first and second players, respectively.[2] If B and C are two matrices of the same size, then it is easily shown that

$$\left|\text{val}[B] - \text{val}[C]\right| \leq \max_{i,j} \left|b_{ij} - c_{ij}\right|. \tag{2}$$

Returning to the stochastic game Γ, define $A^k(\vec{\alpha})$ to be the matrix of elements

$$a_{ij}^k + \textstyle\sum_l p_{ij}^{kl} \alpha^l,$$

$i = 1, 2, \ldots, m_k;\ j = 1, 2, \ldots, n_k$, where $\vec{\alpha}$ is any N-vector with numerical components. Pick $\vec{\alpha}_{(0)}$ arbitrarily, and define $\vec{\alpha}_{(t)}$ by the recursion:

$$\alpha_{(t)}^k = \text{val}[A^k(\vec{\alpha}_{(t-1)})], \qquad t = 1, 2, \ldots.$$

(If we had chosen $\alpha_{(0)}^k$ to be the value of A^k, for each k, then $\alpha_{(t)}^k$ would be the value of the truncated game $\Gamma_{(t)}^k$ which starts at position k, and which is cut off after t steps if it lasts that long.) We shall show that the limit of $\vec{\alpha}_{(t)}$ as $t \to \infty$ exists and is independent of $\vec{\alpha}_{(0)}$, and that its components are the values of the infinite games Γ^k.

Consider the transformation T:

$$T\vec{\alpha} = \vec{\beta}, \qquad \text{where } \beta^k = \text{val}[A^k(\vec{\alpha})].$$

Define the norm of $\vec{\alpha}$ to be

$$\|\vec{\alpha}\| = \max_k \left|\alpha^k\right|.$$

Then we have

$$\begin{aligned} \|T\vec{\beta} - T\vec{\alpha}\| &= \max_{k} \left| \mathrm{val}[A^k(\vec{\beta})] - \mathrm{val}[A^k(\vec{\alpha})] \right| \\ &\leq \max_{k,i,j} \left| \textstyle\sum_l p_{ij}^{kl}\beta^l - \sum_l p_{ij}^{kl}\alpha^l \right| \\ &\leq \max_{k,i,j} \left| \textstyle\sum_l p_{ij}^{kl} \right| \max_{l} \left| \beta^l - \alpha^l \right| \\ &= (1 - s)\|\vec{\beta} - \vec{\alpha}\|, \end{aligned} \tag{3}$$

using (2). In particular, $\|T^2\vec{\alpha} - T\vec{\alpha}\| \leq (1 - s)\|T\vec{\alpha} - \vec{\alpha}\|$. Hence the sequence $\vec{\alpha}_{(0)}, T\vec{\alpha}_{(0)}, T^2\vec{\alpha}_{(0)}, \ldots$ is convergent. The limit vector $\vec{\phi}$ has the property $\vec{\phi} = T\vec{\phi}$. But there is only one such vector, for $\vec{\psi} = T\vec{\psi}$ implies

$$\|\vec{\psi} - \vec{\phi}\| = \|T\vec{\psi} - T\vec{\phi}\| \leq (1 - s)\|\vec{\psi} - \vec{\phi}\|,$$

by (3), whence $\|\vec{\psi} - \vec{\phi}\| = 0$. Hence $\vec{\phi}$ is the unique fixed point of T and is independent of $\vec{\alpha}_{(0)}$.

To show that ϕ^k is the value of the game Γ^k, we observe that by following an optimal strategy of the finite game $\Gamma_{(t)}^k$ for the first t steps and playing arbitrarily thereafter, the first player can assure himself an amount within $\epsilon_t = (1 - s)^t M/s$ of the value of $\Gamma_{(t)}^k$; likewise for the other player. Since $\epsilon_t \to 0$ and the value of $\Gamma_{(t)}^k$ converges to ϕ^k, we conclude that ϕ^k is indeed the value of Γ^k. Summing up:

THEOREM 1. *The value of the stochastic game* Γ *is the unique solution* $\vec{\phi}$ *of the system*

$$\phi^k = \mathrm{val}[A^k(\vec{\phi})], \qquad k = 1, 2, \ldots, N.$$

Our next objective is to prove the existence of optimal strategies.

THEOREM 2. *The stationary strategies* $\vec{x}^*$, $\vec{y}^*$, *where* $x^l \in X[A^l(\vec{\phi})]$, $y^l \in Y[A^l(\vec{\phi})]$, $l = 1, 2, \ldots, N$, *are optimal for the first and second players respectively in every game* Γ^k *belonging to* Γ.

Proof: Let a finite version of Γ^k be defined by agreeing that on the tth step the play shall stop, with the first player receiving the amount $a_{ij}^h + \sum_l p_{ij}^{hl}\phi^l$ instead of just a_{ij}^h. Clearly, the stationary strategy $\vec{x}^*$ assures the first player the amount ϕ^k in this finite version. In the original game Γ^k, if the first player uses $\vec{x}^*$, his expected winnings after t steps will be at least

$$\phi^k - (1 - s)^{t-1} \max_{h,i,j} \textstyle\sum_l p_{ij}^{hl}\phi^l,$$

and hence at least

$$\phi^k - (1 - s)^t \max_l \phi^l.$$

His total expected winnings are therefore at least

$$\phi^k - (1 - s)^t \max_l \phi^l - (1 - s)^t M/s.$$

Since this is true for arbitrarily large values of t, it follows that $\vec{x}^*$ is optimal in Γ^k for the first player. Similarly, $\vec{y}^*$ is optimal for the second player.

Reduction to a Finite-Dimensional Game.—The non-linearity of the "val" operator often makes it difficult to obtain exact solutions by means of Theorems 1 and 2. It therefore becomes desirable to express the payoff directly in terms of stationary strategies. Let $\bar{\Gamma} = \{\bar{\Gamma}^k\}$ denote the collection of games whose *pure* strategies are the stationary strategies of Γ. Their payoff functions $\mathfrak{H}^k(\vec{x}, \vec{y})$ must satisfy

$$\mathfrak{H}^k(\vec{x}, \vec{y}) = x^k A^k y^k + \sum_l x^k P^{kl} y^k \mathfrak{H}^l(\vec{x}, \vec{y}),$$

for $k = 1, 2, \ldots, N$. This system has a unique solution; indeed, for the linear transformation $T_{\vec{x}\vec{y}}$:

$$T_{\vec{x}\vec{y}}\,\vec{\alpha} = \vec{\beta}, \text{ where } \beta^k = x^k A^k y^k + \sum_l x^k P^{kl} y^k \alpha^l$$

we have at once

$$\|T_{\vec{x}\vec{y}}\,\vec{\beta} - T_{\vec{x}\vec{y}}\,\vec{\alpha}\| = \max_k \left|\sum_l x^k P^{kl} y^k (\beta^l - \alpha^l)\right| \leq (1 - s)\|\vec{\beta} - \vec{\alpha}\|,$$

corresponding to (3) above. Hence, by Cramer's rule,

$$\mathfrak{H}^k(\vec{x}, \vec{y}) = \frac{\begin{vmatrix} x^1P^{11}y^1 - 1 & x^1P^{12}y^1 & \ldots & -x^1A^1y^1 & \ldots & x^1P^{1N}y^1 \\ x^2P^{21}y^2 & x^2P^{22}y^2 - 1 & & & & \\ \ldots & & & \ldots & & \ldots \\ x^NP^{N1}y^N & \ldots & \ldots & -x^NA^Ny^N & \ldots & x^NP^{NN}y^N - 1 \end{vmatrix}}{\begin{vmatrix} x^1P^{11}y^1 - 1 & x^1P^{12}y^1 & \ldots & x^1P^{1k}y^1 & \ldots & x^1P^{1N}y^1 \\ x^2P^{21}y^2 & x^2P^{22}y^2 - 1 & & \ldots & & \\ \ldots & & & x^kP^{kk}y^k - 1 & & \ldots \\ & & & \ldots & & \\ x^NP^{N1}y^N & \ldots & \ldots & x^NP^{Nk}y^N & \ldots & x^NP^{NN}y^N - 1 \end{vmatrix}}.$$

THEOREM 3. *The games* $\bar{\Gamma}^k$ *possess saddle points:*

$$\min_{\vec{y}} \max_{\vec{x}} \mathfrak{H}^k(\vec{x}, \vec{y}) = \max_{\vec{x}} \min_{\vec{y}} \mathfrak{H}^k(\vec{x}, \vec{y}), \tag{4}$$

for $k = 1, 2, \ldots, N$. *Any stationary strategy which is optimal for all* $\Gamma^k \in \Gamma$ *is an optimal pure strategy for all* $\bar{\Gamma}^k \in \bar{\Gamma}$, *and conversely. The value vectors of* Γ *and* $\bar{\Gamma}$ *are the same.*

The proof is a simple argument based on Theorem 2. It should be pointed out that a strategy $\vec{x}$ may be optimal for one game Γ^k (or $\bar{\Gamma}^k$) and not optimal for other games belonging to Γ (or $\bar{\Gamma}$). This is due to the possibility that Γ might be "disconnected"; however if none of the p_{ij}^{kl} are zero this possibility does not arise.

It can be shown that the sets of optimal stationary strategies for Γ are closed, convex polyhedra. A stochastic game with rational coefficients does not necessarily have a rational value. Thus, unlike the minimax theorem for bilinear forms, the equation (4) is not valid in an arbitrary ordered field.

Examples and Applications.—1. When $N = 1$, Γ may be described as a simple matrix game A which is to be replayed according to probabilities that depend on the players' choices. The payoff function of Γ is

$$\mathfrak{H}(x, y) = \frac{xAy}{xSy},$$

where S is the matrix of (non-zero) stop probabilities. The minimax theorem (4) for rational forms of this sort was established by von Neumann;[3] an elementary proof was subsequently given by Loomis.[4]

2. By setting all the stop probabilities s_{ij}^k equal to $s > 0$, we obtain a model of an indefinitely continuing game in which future payments are discounted by a factor $(1 - s)^t$. In this interpretation the actual transition probabilities are $q_{ij}^{kl} = p_{ij}^{kl}/(1 - s)$. By holding the q_{ij}^{kl} fixed and varying s, we can study the influence of interest rate on the optimal strategies.

3. A stochastic game does not have perfect information, but is rather a "simultaneous game," in the sense of Kuhn and Thompson.[1] However, perfect information can be simulated within our framework by putting either m_k or n_k equal to 1, for all values of k. Such a stochastic game of perfect information will of course have a solution in stationary *pure* strategies.

4. If we set $n_k = 1$ for all k, effectively eliminating the second player, the result is a "dynamic programming" model.[5] Its solution is given by any set of integers $\vec{i} = \{i_1, i_1, \ldots, i_N | 1 \leq i_k \leq m_k\}$ which maximizes the expression

$$\mathfrak{K}^k(\vec{i}) = \frac{\begin{vmatrix} p_{i_1}^{11} - 1 & p_{i_1}^{12} & \dots & -a_{i_1}^{1} & \dots & p_{i_1}^{1N} \\ p_{i_2}^{21} & p_{i_2}^{22} - 1 & & & & \\ \dots & & & \dots & & \dots \\ p_{i_N}^{N1} & & \dots & -a_{i_N}^{N} & \dots & p_{i_N}^{NN} - 1 \end{vmatrix}}{\begin{vmatrix} p_{i_1}^{11} - 1 & p_{i_1}^{12} & \dots & p_{i_1}^{1k} & \dots & p_{i_1}^{1N} \\ p_{i_2}^{21} & p_{i_2}^{22} - 1 & & \dots & & \\ \dots & & & p_{i_k}^{kk} - 1 & & \dots \\ & & & \dots & & \\ p_{i_N}^{N1} & & \dots & p_{i_N}^{Nk} & \dots & p_{i_N}^{NN} - 1 \end{vmatrix}}.$$

For example (taking $N = 1$), let there be alternative procedures $i = 1, \dots, m$ costing $c_i = -a_i$ to apply and having probability s_i of success. The above then gives us the rule: adopt that procedure i^* which maximizes the ratio a_{i*}/s_{i*}, or equivalently, the ratio s_{i*}/c_{i*}.

5. Generalizations of the foregoing theory to infinite sets of alternatives, or to an infinite number of states, readily suggest themselves (see for example ref. 6). We shall discuss them in another place.

* The preparation of this paper was sponsored (in part) by the Office of Naval Research.

[1] Kuhn, H. W., *Contributions to the Theory of Games II*, Annals of Mathematics Studies No. 28, Princeton, 1953, pp. 209–210.

[2] von Neumann, J., and Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton, 1944 and 1947, p. 158.

[3] von Neumann, J., *Ergebnisse eines Math. Kolloquiums*, **8**, 73–83 (1937).

[4] Loomis, L. H., these PROCEEDINGS, **32**, 213–215 (1946).

[5] Bellman, R., these PROCEEDINGS, **38**, 716–719 (1952).

[6] Isbell, J. R., *Bull. A. M. S.*, **59**, 234–235 (1953).

EXISTENCE OF GENERALIZED LOCAL CLASS FIELDS

BY G. WHAPLES

MATHEMATICS DEPARTMENT, UNIVERSITY OF SOUTHERN CALIFORNIA*

Communicated by Saunders MacLane, July 27, 1953

1. *Introduction.*—By generalized local class field theory[1] we mean the theory of abelian extensions of a field k which is complete under a non-Archimedean rank one valuation and has a residue class field $\bar{k}$ which satisfies the two axioms: (1) $\bar{k}$ has no inseparable extensions, and (2) for each positive integer n, the algebraic closure of $\bar{k}$ contains exactly one subfield of degree n over $\bar{k}$. Two of the three main theorems—namely the existence and limitation theorems—have been proved in an entirely