The Annals of Statistics
1975, Vol. 3, No. 2, 554-558

LOWER SEMICONTINUOUS STOCHASTIC GAMES WITH IMPERFECT INFORMATION

BY SAILES K. SENGUPTA

University of Missouri—Kansas City

Shapley's stochastic game is considered in a more general setting, with the accumulated payoff being regarded as a function on the space of infinite trajectories, and the set of states of the system taken as a compact metric space. It has been shown that any game with a lower semicontinuous payoff has a value and one of the players has an optimal strategy. As a consequence, in Shapley's game both players have optimal strategies.

1. Introduction. Infinite move games with perfect information have been considered by several authors [2], [3], [4], [5], [6] with successive improvements in their results, although the question as to exactly when such games have a value remains open. Blackwell [1] has considered infinite games with a simple kind of information lag, namely, at each state the *simultaneous* choices of the two players determine the play. He has proved that if the payoff is the indicator function of a G_δ set on the space of trajectories, then the game is determined. We consider here a variation suggested by Blackwell where the simultaneous choices of the players determine the transition probabilities to the state space according to a given law and prove that any game with a payoff which is lower semicontinuous on the trajectory space is determined. This result, as a special case, implies the determinedness of Shapley's stochastic games in [6].

2. Statement of the theorem. Let I, J be two finite sets and S a compact metric space. Let $Z = I \times J \times S$ and H is the space of all infinite sequences $h = (z_1, z_2, \cdots)$ and $z_n \in Z$ and φ is a bounded Baire function on H. To each φ and initial state $s_0 \in S$ corresponds a 0-sum two person game $\Gamma_\varphi(s_0)$, to be played as follows: starting from state s_0 (known to both) players A and B respectively choose $i_1 \in I$ and $j_1 \in J$ simultaneously. Depending on their choices a referee moves the state of the system to $s_1 \in S$ according to the transition law $q(\cdot \mid s_0, i_1, j_1)$ known to both, and then announces the triple (i_1, j_1, s_1). The process is repeated with the simultaneous choices of $i_2 \in I$ by A and $j_2 \in J$ by B followed by the referee's move to $s_2 \in S$ according to the given transition law, etc. The result, an infinite sequence of triples preceded by $s_0 \in S$ is an element in $S \times H$. Let Π be the set of all partial histories $p = (s_0, z_1, \cdots, z_n)$, $n = 0, 1, 2, \cdots$; then a strategy $\alpha(\beta)$ for A(B) associates with each partial history a regular conditional probability distribution on $I(J)$. A pair (α, β) together with the transition law q determines a probability distribution $P_{\alpha\beta}$ on H ($P_{\alpha\beta}$ depends on s_0, a fact which will be understood) and A's expected income is $E_{\alpha\beta}(\varphi)$, $E_{\alpha\beta}$ denoting the expectation under $P_{\alpha\beta}$.

Received November 1973; revised April 1974.

The lower and upper values of the game Γ_φ starting at s_0 will be denoted by

$$L(\varphi) = \sup_\alpha \inf_\beta E_{\alpha\beta}(\varphi)$$

and

$$U(\varphi) = \inf_\beta \sup_\alpha E_{\alpha\beta}(\varphi)$$

respectively. If $L(\varphi) = U(\varphi)$, the common value is called the value of the game Γ_φ.

THEOREM. *If φ is lower semicontinuous on H then Γ_ϕ has a value and* B *has an optimal strategy.*

PROOF. For any partial history p and for a fixed k we define a game $\Gamma^k(p)$, *that starts from* p and has a payoff equal to

$$\varphi_k(h) = \inf_{h'} \varphi(h')$$

the infimum being taken over all h' that agrees with h up to the kth triplet (denote this by $h' \overset{k}{=} h$). We first prove the

PROPOSITION. (a) *The game $\Gamma^k(p)$ has a value $v^k(p)$ and*

(b) *The value of $\Gamma^k(p)$ is equal to the value of the matrix game whose (i, j)th coordinate is*

$$\int v_k(p, (i, j), u)\, dq(u \mid s, i, j)\ .$$

We prove the existence in (a) by induction on k. The case $k = 0$ is trivial since we have a game with constant payoff. The induction hypothesis is: The game $\Gamma^m(p)$ has a value and both players have optimal strategies for all p and $m \leqq k - 1$ and *all bounded measurable function* φ. We then prove the same for $m = k$. Without loss of generality we may assume p has less than k triplets, for otherwise we have again a constant payoff situation.

We set

$$\gamma(z_1, \cdots, z_{k-1}, (i, j)) = \int \varphi_k(p, \cdots, z_{k-1})(i, j, u)\, dq(u)\ .$$

In the matrix game with payoff $\gamma(z_1, \cdots, z_{k-1}, (i, j))$, there exists a value say γ^* and optimal strategies λ^*, μ^* where λ^*, μ^* are probability distributions on I and J respectively (given $z_1, \cdots, z_{k-1}$). By the induction hypothesis, there exists a value of the game with payoff $\gamma^*(z_1, \cdots, z_{k-1})$ (considered as an infinite game) and A and B have both optimal strategies say $\alpha_1{}^*$ and $\beta_1{}^*$ respectively. Let us define α^* for A as the strategy which follows α^* up to the $(k - 1)$th stage and then takes λ^* followed by arbitrary moves afterwards and likewise β^* for B is $(\beta_1{}^*, \mu^*)$ followed by arbitrary moves thereafter. We claim that α^*, β^* are respectively the optimal strategies for A and B in $\Gamma^k(p)$ and hence the game has a value, namely $E_{\alpha^*\beta^*}\varphi_k(z_1, \cdots, z_k)$. We first note that by construction of γ, λ^*, μ^*,

$$\begin{aligned} E_{\lambda^*\mu} \int \varphi_k(z_1, \cdots, z_{k-1}, i, j, u)\, dq(u) &\geqq E_{\lambda^*\mu^*} \int \varphi_k(z_1, \cdots, z_{k-1}, i, j, u)\, dq(u) \\ &\geqq E_{\lambda\mu^*} \int \varphi_k(z_1, \cdots, z_{k-1}, i, j, u)\, dq(u)\ , \end{aligned}$$

where $\mu(\lambda)$ are the conditional distributions on the kth coordinate space $J(I)$

given $z_1, \cdots, z_{k-1}$ under an arbitrary strategy $\beta(\alpha)$. Thus for this β we have

$$\begin{aligned} E_{\alpha_1{}^*\beta} E_{\lambda^*\mu} \int \varphi_k(z_1, \cdots, z_{k-1})(i, j, u)\, dq(u) \\ &\geqq E_{\alpha_1{}^*\beta} E_{\lambda^*\mu^*} \int \varphi_k(z_1, \cdots, z_{k-1})(i, j, u)\, dq(u) \\ &= E_{\alpha_1{}^*\beta}\gamma^* \geqq E_{\alpha_1{}^*\beta_1{}^*}\gamma^* = v_k \quad \text{say.} \end{aligned}$$

The last inequality follows by optimality of $\beta_1{}^*$. Thus for all β for B we have

$$E_{\alpha^*\beta}(\varphi_k) \geqq v_k\,.$$

Similarly for arbitrary strategy α of A we get

$$E_{\alpha\beta^*}(\varphi_k) \leqq v_k\,.$$

Thus v_k is the value and α^*, β^* are respectively optimal strategies for A and B.
We now prove the second assertion in the proposition:

$$v_k(p) = \text{Val}\, ||\int v_k(p, (i, j, u))\, dq(u \mid i, j)||\,,$$

where $\text{Val}\, ||A_{ij}||$ for any matrix (A_{ij}) stands for the value of the corresponding matrix game. By the previous result, for any $z \in Z$ and any partial history p there exist optimal strategies $\alpha^*(z)$, $\beta^*(z)$ satisfying

$$v_k(pz) = E_{\alpha^*(z), \beta^*(z)}[\varphi_k(p'h)]\,,$$

where p' is the partial history p followed by $z = (i, j, u)$.

Now consider the matrix game with payoff

$$u_k(p, (i, j)) = \int v_k(pz)\, dq(u)\,.$$

This game has optimal strategies λ^*, μ^* say.

Consider the strategy $\alpha^* = (\lambda^*, \alpha^*(z))$ that is λ^* followed by $\alpha^*(z)$. Then for any strategy β (which can be considered as a probability μ on J followed by $\beta(z)$) of B one has

$$\begin{aligned} E_{\alpha^*\beta}[\varphi_k(ph)] &= \int dP_{\lambda^*\mu}(i, j) \int dq(u) \int \varphi_k(p'h)\, dP_{\alpha^*(z)\beta(z)} \\ &\geqq \int dP_{\lambda^*\mu}(i, j) \int v_k(p, z)\, dq(u \mid i, j) \\ &= E_{\lambda^*\mu}[u_k(p, (i, j))] \\ &\geqq E_{\lambda^*\mu^*}[u_k(p, (i, j))]\,. \end{aligned}$$

Similarly for any strategy α of A,

$$E_{\alpha\beta^*}[\varphi_k(ph)] \leqq E_{\lambda^*\mu^*}[u_k(p, (i, j))]\,.$$

Thus $E_{\lambda^*\mu^*}[u_k(p, (i, j))]$, the value of the matrix game $||u_k(p, (i, j))||$ is also the value of $\Gamma_k(p)$.

We now come back to the proof of the theorem. First note that $v_k(p)$ increases to $W(p)$ say, as $k \to \infty$. The value of a matrix game being a continuous function of its elements it follows that

$$W(p) = \text{Val}\, ||\int W(p, i, j, u)\, dq(u)||\,. \tag{1}$$

Taking $p = (s_0)$, since $v_k(s_0) \uparrow W(s_0)$ and A can guarantee at least $v_k(s_0)$ in $\Gamma_k(s_0)$, and so certainly in Γ_φ (remembering $\varphi_k \leqq \varphi$), the lower value $L(\varphi)$ in Γ_φ satisfies $L(\varphi) \geqq W(s_0)$.

We now exhibit a strategy for B which would guarantee the expected payoff within $W(s_0)$ and this would imply that B has an optimal strategy and the game has a value. We give here a construction analogous to Blackwell's. Given p, let B play optimally as in the matrix game

$$\|\textstyle\int W(p, (i, j, u))\, dq(u \mid (i, j))\| .$$

Then under arbitrary strategy of A if the resulting play is $z_1, z_2, \cdots$, then

$$W(s_0),\ W(s_0, z_1),\ W(s_0, z_1, z_2),\ \cdots$$

is an expectation decreasing martingale. This follows from (1) and the construction of the optimal strategy, since

$$\begin{aligned} E(W(s_0, z_1, \cdots, z_{n+1}) \mid W(s_0), \cdots, W(s_0, z_1, \cdots, z_n)) \\ \leqq \mathrm{Val}\, \|\textstyle\int W(p, i, j, u)\, dq\, (u \mid i, j)\| \\ = W(p) \qquad \text{where} \quad p = (s_0, z_1, \cdots, z_n) . \end{aligned}$$

Hence

$$W(s_0) \geqq E(W(s_0. z_1)) \geqq \cdots \geqq E(W(s_0, z_1, \cdots, z_n)) \geqq E(v_n(s_0, z_1, \cdots, z_n)) .$$

But $\varphi_n = v_n(z_1, \cdots, z_n)$. Hence $E(\varphi_n) \leqq W(s_0)$. But by lower semicontinuity of φ, $\varphi_n \uparrow \varphi$. Hence $E(\varphi) \leqq W(s_0)$, proving that B can guarantee an expected payoff within $W(s_0)$ and the corresponding strategy of B is optimal.

COROLLARY. *If φ is continuous, then the game is determined and both players have optimal strategies.*

PROOF. Interchange the roles of the players and consider the payoff $- \varphi$.

REMARK 1. In Shapley's ([6]) stochastic game the payoff is accumulated over stages, the nth stage of the game producing a payoff $a(u_{n-1}, z_n) = A_n$ say, with a discounting factor β^n ($\beta < 1$), so that the accumulated payoff equals

$$\lim_{n\to\infty} \left(\textstyle\sum_{k=1}^{n} A_k \beta^k\right) . \tag{2}$$

With S finite, the fact that finite dimensional functions on the history space are continuous and the fact that the A_n's assume only a finite number of values and so are bounded, it follows that the expression (2) is the uniform limit of a sequence of continuous functions on the history space. So (2) is continuous itself, the game is determined and both players have optimal strategies.

REMARK 2. If one could prove that for a φ satisfying the property that the set $[h\colon \varphi(h) \geqq c]$, is a G_δ set for any real c, the game Γ_φ has a value, then he could conclude that Shapley's game with *expected average cost criterion* is determined too.

REMARK 3. The function $\varphi_k(h) = \inf_{h' \overset{k}{=} h} \varphi(h')$ has been assumed to be measurable in the preceding. In fact it is lower semicontinuous. Here is a proof. It is enough to show that for all c, if

$$(3) \qquad \varphi_k(h^0) > c \qquad \text{and} \qquad h^n \to h^0\,, \qquad \text{then} \quad \liminf \varphi_k(h^n) \geqq c\,.$$

Suppose not; then there exists a subsequence $\{n'\}$ of $\{n\}$ such that for all n', $\varphi_k(h^{n'}) < c$. This implies that there exists $\bar{h}^{n'} \stackrel{k}{=} h^{n'}$ such that

$$(4) \qquad \varphi(\bar{h}^{n'}) < c \qquad \text{for all} \quad n'\,.$$

The space Ω being compact, there exists a convergent subsequence $\{\bar{h}^{n''}\}$ of $\{\bar{h}^{n'}\}$ converging to $\bar{h}^0$ say. Thus from (4) $\liminf \varphi(\bar{h}^{n''}) \leqq c$. But $\liminf \varphi(\bar{h}^{n''}) \geqq \varphi(\bar{h}^0)$ by l.s.c. of φ. Hence $c \geqq \varphi(\bar{h}^0) \geqq \varphi_k(h^0)$, contradicting (3).

REMARK 4. It will be interesting to know if B has a *stationary* optimal strategy.

Acknowledgments. I am indebted to Professor D. Blackwell who introduced me to the problem and offered valuable suggestions. I am also thankful to Professor R. N. Bhattacharya for several stimulating discussions. Thanks also to the referee for his comments that led to some improvements of the results.

REFERENCES

[1] BLACKWELL, DAVID (1969). Infinite G_δ games with imperfect information. *Zastos. Mat.* **10** 99–101.
[2] DAVIS, MORTON (1964). Infinite games of perfect information. *Advances in Game Theory.* Princeton Univ. Press.
[3] GALE, DAVID and STEWART, F. M. (1953). Infinite games with perfect information. *Contribution to The Theory of Games II.* Princeton Univ. Press.
[4] MYCIELSKI, JAN and STEINHAUS, HUGO (1962). A mathematical axiom contradicting the axiom of choice. *Bull. Acad. Polon. Sci. Sér. Math. Astron. Phys.* **10** 1–3.
[5] OXTOBY, J. C. (1957). The Banach-Mazur game and Banach category theorem. *Contributions to The Theory of Games III.* Princeton Univ. Press.
[6] SHAPLEY, L. S. (1953). Stochastic games. *Proc. Nat. Acad. Sci.* **39** 1095–1100.
[7] WOLFE, PHILIP (1955). The strict determinateness of certain infinite games. *Pacific J. Math.* **5** 891–897.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MISSOURI
KANSAS CITY, MISSOURI 64110