Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2145–2146, March 1982
Mathematics

Stochastic games have a value

(repeated games/stochastic processes/dynamic programming)

JEAN-FRANCOIS MERTENS* AND ABRAHAM NEYMAN†‡

*Center for Operations Research and Econometrics and Department of Mathematics, Université Catholique de Louvain, Louvain-la Neuve, Belgium; and †Department of Mathematics, University of California, Berkeley, California 94720

Communicated by Lloyd S. Shapley, December 21, 1981

ABSTRACT **Undiscounted nonterminating stochastic games in which the state and action spaces are finite have a value.**

A stochastic game is played in stages. At each stage t, the game is in one of finitely many states, and each of the two players observes the current state z_t and chooses one of finitely many actions. The pair of actions at stage t, together with z_t, determines the payoff x_t to be made by player II to player I at stage t and the probability used by the referee to select the next state. All of the referee's choices are made independently of the past. A player's strategy is a specification of a probability distribution over his actions at each stage, conditional on the current state and the history of the game up to that stage. Any pair of strategies, σ of player I and τ of player II, induces together with z_0 a probability distribution on the stream $(x_0, x_1, \ldots)$ of payoffs.

The definition of a value for the stochastic game depends on how the players evaluate a distribution over streams of payoffs. Shapley (1) proved that the λ-discounted game, the game with "evaluation"

$$E\left\{\sum_{t=0}^{\infty} \lambda(1-\lambda)^t x_t\right\} \qquad 0 < \lambda < 1,$$

has a value and that both players have optimal stationary strategies. Let v_λ^z denote the value of the λ-discounted game with initial state z, and let σ_λ denote a stationary optimal strategy of player I in the λ-discounted game. Using Tarski's principle for real closed fields, Bewley and Kohlberg (2) proved that both v_λ^z and σ_λ have a convergent expansion in fractional powers of λ and that the limit v_∞^z of v_λ^z as $\lambda \to 0$ exists. The question as to whether or not the undiscounted stochastic games—i.e., the games with "evaluation"

$$E\{\liminf_{n\to\infty} \bar{x}_n\},$$

where $\bar{x}_n = (1/n)\Sigma_{t<n} x_t$—always have a value was open for many years. The existence of a value has been proved only in special cases. Gillette (3) and Liggett and Lippman (4) proved the existence of the value when the undiscounted stochastic game has perfect information. Gillette (3) and Hoffman and Karp (5) proved that irreducible (cyclic) undiscounted stochastic games have a value. Blackwell and Ferguson (6) found in a particular example ("The Big Match") a strategy that would prove to be basic for further generalizations. Kohlberg (7) proved that all "games with absorbing states" have a value.

Our main result is that undiscounted stochastic games always have a value. We have the following *Theorem*.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MAIN THEOREM. *For every stochastic game and for every $\varepsilon > 0$, there exist strategies σ_ε of player I and τ_ε of player II and a number $N > 0$ such that, for all strategies τ of player II and σ of player I and for every initial state z,*

$$\varepsilon + E_{\sigma_\varepsilon,\tau}\{\liminf_{n\to\infty} \bar{x}_n\} \geq v_\infty^z \geq -\varepsilon + E_{\sigma,\tau_\varepsilon}\{\limsup_{n\to\infty} \bar{x}_n\},$$

and, for every $n > N$,

$$\varepsilon + E_{\sigma_\varepsilon,\tau}(\bar{x}_n) \geq v_\infty^z \geq -\varepsilon + E_{\sigma,\tau_\varepsilon}(\bar{x}_n).$$

Independently of ourselves, Monash (8) announced a weaker version of the present result: it is not claimed that the strategy σ_ε of player I is ε-optimal either in the infinite game or in sufficiently long finite games, but only that, for every strategy τ of player II, there exists N such that, for all $n \geq N$, $E_{\sigma_\varepsilon,\tau}(\bar{x}_n) \geq v_\infty^z - \varepsilon$.

The ε-optimal strategies

Let $0 \leq r < 1$, $B > 0$ be such that, for $0 < \lambda < 1$ and for every (initial) state z, $|v_\lambda^z - v_\infty^z| \leq B\lambda^{1-r}$. The existence of such r and B follows from the basic result of Bewley and Kohlberg (2). Let $\delta > 0$, $\ell \geq 1$ and $\ell \geq 2B/\delta$, and define $L(\lambda) = \inf\{n | n \geq \ell \lambda^{-r}\}$. Choose $\beta > 1$ such that $\beta r < 1$, and let $\alpha = \beta r$, $\gamma = \beta - \alpha$. Our strategy will depend on an additional constant M, sufficiently large to satisfy further requirements that will be specified later.

We define inductively:

$$s_0 = s \text{ arbitrary} \geq M$$

$$\lambda_i = s_i^{-\beta},\ L_i = L(\lambda_i)$$

$$B_0 = 0,\ B_{i+1} = B_i + L_i$$

$$\ell_i = w_{B_i}, \text{ where } w_j \text{ is } v_\infty^{z_j}$$

$$s_{i+1} = \text{Max}\left[M,\ s_i + \sum_{B_i \leq j < B_{i+1}} (x_j - \ell_{i+1} + 2\delta)\right].$$

The (M, s, δ, r, ℓ)-strategy ($\sigma_{M,\delta}$ for short) is to play at stage j the optimal strategy σ_{λ_i} of the λ_i-discounted game when $B_i \leq j < B_{i+1}$.

We prove that, for every $\varepsilon > 0$, there is a pair M, δ such that $\sigma_{M,\delta}$ is ε-optimal—i.e., satisfies the requirements of σ_ε in the *Main Theorem*. We actually prove that, for every $\varepsilon > 0$, there is $\delta_0 > 0$ such that for every $0 < \delta \leq \delta_0$ there is $M_0 > 0$ such that for $M > M_0$ $\sigma_{M,\delta}$ is ε-optimal.

‡ Present address: Economics Dept., Stanford Univ., Stanford, CA 94305.

Sketch of the proof

Using the definition of s_k, we show that for M sufficiently large,

$$\sum_{i<B_n} x_i \geq s_n - s_0 + \sum_{k=1}^{n} (B_k - B_{k-1})(\ell_k - 2\delta) - \sum_{k=1}^{n} I(s_k = M)2M^{\alpha}A\ell, \quad [1]$$

where A is twice the largest absolute value of payoffs appearing in the game matrices and I is the indicator function. This leads to the study of the $(s_k, \ell_k)_{k=0}^{\infty}$ process. For that we first show that there is $\lambda_\delta > 0$ such that, for every $0 < \lambda \leq \lambda_\delta$ and for every stopping time T,

$$E\left\{w_{T+L} - w_T + \lambda \sum_{i<L} (x_{T+i} - w_{T+L} + \delta) \middle| \mathcal{F}_T\right\} \geq 0$$

whenever player I uses constantly σ_λ between T and $T + L \equiv T + L(\lambda)$, where $\mathcal{F}_T$ denotes the σ-algebra of all events up to the stopping time T. Denoting $\mathcal{G}_k = \mathcal{F}_{B_k}$, we can rewrite this as

$$E\{s_{k+1} - s_k + s_k^{\beta}(\ell_{k+1} - \ell_k) | \mathcal{G}_k\} \geq \delta\ell s_k^{\alpha}, \text{ for } M \geq \lambda_\delta^{-1/\beta}. \quad [2]$$

Making a linear positive change of variables, we are led to the study of the following class of stochastic processes.

Let $(Q, \alpha, \beta, \delta)$ be a fixed "4-tuple", where Q is a fixed finite set of real numbers, $0 \leq \alpha < 1$, $1 < \beta$, and $\delta > 0$. For any given $M > 0$, we consider the class $\mathcal{A}(M)$ of all stochastic processes $[\Omega, \mathcal{F}, (\mathcal{F}_k)_{k=0}^{\infty}, s_k, \ell_k]$ where (s_k, ℓ_k) are $\mathcal{F}_k$-measurable $[M, \infty) \times Q$-valued random variables obeying:

$$|s_{k+1} - s_k| \leq s_k^{\alpha} \text{ and } E\{s_{k+1} - s_k + s_k^{\beta}(\ell_{k+1} - \ell_k)|\mathcal{F}_k\} \geq \delta s_k^{\alpha}.$$

Our formulas **1** and **2** imply that the result will be proved by the following *Proposition*.

PROPOSITION. *For every* $\varepsilon > 0$, *there is* $M_0 > 0$ *such that for every* $M \geq M_0$ *there is* $N > 0$ *such that for any stochastic process* $(s_k, \ell_k)_{k=0}^{\infty}$ *in* $\mathcal{A}(M)$:

(i) ℓ_k *converges a.e., say to* ℓ_∞ *and* $E(\ell_\infty) > \ell_0 - \varepsilon$;
(ii) *for any stopping time* T, $E(\ell_T) \geq \ell_0 - \varepsilon$;
(iii) *lim sup* $(1/n)\sum_{k=1}^{n} I(s_k = M) \leq \varepsilon$;
(iv) $E\{(1/n)\sum_{k=1}^{n} I(s_k = M)\} \leq \varepsilon$ *whenever* $n \geq N$.

In order to prove parts (*i*) and (*ii*) of the *Proposition*, we consider the function

$$\varphi(s) = A\left[1 - \prod_{i=0}^{\infty}(1 - \rho s_i^{-\gamma})\right],$$

where $\gamma = \beta - \alpha$, $0 < \rho = (1 - \varepsilon)/A$, and s_k are defined inductively by $s_0 = s$, $s_{k+1} = s_k + s_k^{\alpha}$. Using the theory of completely monotonic functions, we derive the convexity of φ for sufficiently large s. This, together with additional properties of φ, allows us to derive the following *Lemma*.

LEMMA. $E\{\ell_k - \ell_T|\mathcal{F}_k\} \leq \varphi(s_k) \leq \varphi(M) \underset{M\to\infty}{\longrightarrow} 0$, *where* $T = \inf\{j | j > k, \ell_k \neq \ell_j\}$ *and* ℓ_T *is defined as* ℓ_k *if* $T = +\infty$.

Together with the finiteness of Q (the range of the ℓ-process) this *Lemma* implies that for every given $\varepsilon > 0$, for sufficiently large M $E\{\ell_k - \ell_T|\mathcal{F}_k\} \leq \varepsilon$ for any stopping time $T \geq k$. On the one hand, this implies part (*ii*) of the *Proposition*. On the other hand, it implies, taking $\varepsilon < \min |x - y|$, $x, y \in Q$, $x \neq y$, that there is $\sigma > 1$ such that, for any stopping time T,

$$P\{\exists i, i \geq T, \ell_i < \ell_T|\mathcal{F}_T\} \leq \sigma < 1,$$

and this implies part (*i*) of the *Proposition* and the existence of a constant V for which

$$E\left\{\sum_{k=0}^{\infty} |\ell_{k+1} - \ell_k|\right\} \leq V.$$

To prove parts (*iii*) and (*iv*), we consider the functions $\psi(V,N)$, which are defined by

$$\psi(V,N) = \sup E\left\{\sum_{k=0}^{N} I(s_k = M)\right\},$$

where the sup is taken over all stochastic processes $(s_k, \ell_k)_{k=0}^{\infty}$ in $\mathcal{A}(M)$ with $E\{\sum_{i=0}^{\infty} |\ell_{k+1} - \ell_k|\} \leq V$. The function ψ satisfies the following properties:

(*i*) $\psi_M(V,N)$ is concave in V and $\psi_M(V, Nk) \leq k\psi_M(V/k, N)$;
(*ii*) $\psi_M(V, N) \to \psi_M(0, N)$ as $V \to 0$;

and

(*iii*) $\psi_M(0, N) \leq N\varepsilon/2$ for M, N sufficiently large.

These properties allow us to derive parts (*iii*) and (*iv*) of the *Proposition*.

This work was supported by National Science Foundation Grant MCS-79-06634 at the University of California, Berkeley. The work was initiated during an informal workshop at Center for Operation Research & Econometrics, Université Catholique de Louvain, Belgium, on repeated games in the winter of 1978–1979. Helpful comments and stimulus from the participants, especially S. Sorin and S. Zamir, is gratefully acknowledged. The work was continued in the summer of 1979, first in Stanford during the Summer Workshop of the Institute for Mathematical Studies in the Social Sciences and next at the Institute for Advanced Studies at the Hebrew University of Jerusalem. A first draft was finally put together during a visit of Neyman to the Mathematics Department at the Université Catholique de Louvain. It appeared as Center for Operation Research & Econometrics D.P. 8001 and contains a complete proof of the result presented in this report.

1. Shapley, L. (1953) *Proc. Natl. Acad. Sci. USA* **39**, 1095–1100.
2. Bewley, T. & Kohlberg, E. (1976) *Math. Oper. Res.* **1**, 197–208.
3. Gillette, D. (1957) *Contributions to the Theory of Games*, Annals of Mathematics Study 39, eds. Dresher, M., Tucker, A. W. & Wolfe, P. (Princeton Univ. Press, Princeton, NJ), Vol. 3, pp. 179–187.
4. Liggett, T. & Lippman, S. (1970) *SIAM Rev.* **11**, 604–607.
5. Hoffman, A. J. & Karp, R. M. (1966) *Manage. Sci.* **12**, 359–370.
6. Blackwell, D. & Ferguson, T. S. (1968) *Ann. Math. Stat.* **39**, 159–168.
7. Kohlberg, E. (1974) *Ann. Stat.* **2**, 724–738.
8. Monash, C. (1979) Dissertation (Harvard Univ., Cambridge, MA).