SIAM REVIEW
Vol. 11, No. 4, October 1969

STOCHASTIC GAMES WITH PERFECT INFORMATION AND TIME AVERAGE PAYOFF*

THOMAS M. LIGGETT AND STEVEN A. LIPPMAN†

There has been a score of papers in the game theory and dynamic programming literature which have used Tauberian theorems to obtain results about average cost type criteria from results pertaining to discounted costs. The purpose of this paper is to point out that an extension of the Hardy–Littlewood theorem that has been suggested is false and to provide a new proof for the theorem in stochastic games where this false extension was first used.

The Hardy–Littlewood theorem states that if $\langle a_k \rangle$ is a nonnegative sequence of constants and if $\lim_{\beta \uparrow 1} A(\beta)$ exists, then

$$\lim_{N \to \infty} C_N = \lim_{\beta \uparrow 1} A(\beta),$$

where $C_N = N^{-1} \sum_{k=1}^{N} a_k$ and $A(\beta) = (1 - \beta) \sum_{k=1}^{\infty} a_k \beta^k$.

In a stimulating paper, Gillette [4] stated[1] that if $\langle a_k \rangle$ is a nonnegative sequence of constants, then

$$\underline{C} \equiv \liminf_{N \to \infty} C_N \geqq \liminf_{\beta \uparrow 1} A(\beta) \equiv \underline{A}.$$

In the example below, we show that his assertion is false even if the sequence of constants is bounded.

Example 1. For the moment, let $\langle q_n \rangle$ be any set of positive integers. Define $\langle p_n \rangle$ by $p_1 = 0$ and $p_{n+1} = p_n + q_n$. Let

$$a_k = \begin{cases} 1, & \text{if } 2p_n < k \leqq 2p_n + q_n \text{ for some } n, \\ 0, & \text{otherwise.} \end{cases}$$

As defined, $a_k = 1$ for $k = 2p_n + 1, \cdots, 2p_n + q_n$ and $a_k = 0$ for $k = (2p_n + q_n) + 1, \cdots, (2p_n + q_n) + q_n$. Hence, $\underline{C} = \frac{1}{2}$. Also,

$$A(\beta) = (1 - \beta) \sum_{k=1}^{\infty} a_k \beta^k = (1 - \beta) \sum_{n=1}^{\infty} \beta^{2p_n + 1}(1 - \beta^{q_n})/(1 - \beta),$$

* Received by the editors November 5, 1968.

† Western Management Science Institute, University of California, Los Angeles, California 90024.

[1] He suggested a proof along the lines of Karamata's ingenious proof of the Hardy–Littlewood theorem [6, p. 227]. However, Karamata's method cannot be used since the analogue of

$$\lim_{\beta \uparrow 1} (1 - \beta) \sum_{n=1}^{\infty} a_n \beta^n P(\beta^n) = \lim_{\beta \uparrow 1} A(\beta) \int_0^1 P(t)\, dt$$

does not hold for limit inferior if the polynomial P has negative coefficients.

the last by summing a partial geometric series. Fix β temporarily and let $\alpha_n = \beta^{p_n} - \beta^{p_{n+1}}$, then $\beta^{p_n} = \sum_{k=n}^{\infty} \alpha_k$. Rearranging $A(\beta)$, we have

$$\begin{aligned} A(\beta) &= \beta \sum_{n=1}^{\infty} \alpha_n \beta^{p_n} = \beta \sum_{n=1}^{\infty} \alpha_n \sum_{k=n}^{\infty} \alpha_k \\ &= \frac{\beta}{2}\left[\sum_{n=1}^{\infty} \alpha_n^2 + \left(\sum_{n=1}^{\infty} \alpha_n\right)^2\right] \qquad (1) \\ &= \frac{\beta}{2}\left[\sum_{n=1}^{\infty} \alpha_n^2 + 1\right]. \end{aligned}$$

Next, fix ε in $(0, \frac{1}{4})$ and define $c = \log \varepsilon / \log(1 - \sqrt{\varepsilon})$, which assures $c > 2$. Now, impose the condition on $\langle q_r \rangle$ that it increases fast enough so that $q_n > 2p_n \div (c - 2)$ for each n. Define x_n and y_n by

$$x_n = \frac{|\log(1 - \sqrt{\varepsilon})|}{q_n}, \qquad y_n = \frac{|\log \varepsilon|}{2p_n}. \tag{2}$$

Clearly, $y_{n+1} < y_n$ and $y_n \to 0$, while $cq_n > 2p_n + 2q_n = 2p_{n+1}$, which implies $x_n < y_{n+1}$. Hence, $x_n < y_{n+1} < y_n$, and the union of the intervals (x_n, y_n) is $(0, \infty)$. Consequently, for every $\beta < 1$, there exists an n satisfying $x_n \leqq |\log \beta| < y_n$. Substituting in (2), $|\log(1 - \sqrt{\varepsilon})| \leqq |\log \beta^{q_n}|$ and $|\log \varepsilon| > |\log \beta^{2p_n}|$. Equivalently, $1 - \sqrt{\varepsilon} \geqq \beta^{q_n}$ and $\varepsilon < \beta^{2p_n}$. Thus, $\alpha_n^2 = \beta^{2p_n}(1 - \beta^{q_n})^2 > \varepsilon^2$, and substituting in (1) yields $A(\beta) \geqq \beta[\varepsilon^2 + 1]/2$. Hence, $\underline{A} \geqq [\varepsilon^2 + 1]/2 > \frac{1}{2} = \underline{C}$.

Gillette asserted [4, Lemma 2] that for a stochastic game of perfect information with M positions,[2] there are pure stationary strategies[3] (x^*, y^*) such that for all strategies x and y,

$$\underline{L}^i(x, y^*) \leqq \underline{L}^i(x^*, y^*) \leqq \underline{L}^i(x^*, y), \qquad 1 \leqq i \leqq M, \tag{3}$$

where

$$\underline{L}^i(x, y) = \liminf_{n \to \infty} n^{-1} \sum_{j=1}^{n} a_j^i(x, y),$$

$$L^i(x, y) = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} a_j^i(x, y),$$

and $a_j^i(x, y)$ is the payoff to player I at time j when play initially starts from position i and player I (II) uses strategy x (y).

His proof, however, rests upon the extension of the Hardy–Littlewood theorem which was shown to be false in Example 1. We rectify this by establishing a slightly stronger theorem.

[2] In a stochastic game, perfect information means that at each of the M positions one or the other of the players is restricted to a single decision.

[3] A pure stationary strategy is one which prescribes for a player the same decision each time the same position is reached.

THEOREM 1. *For a stochastic game of perfect information with M positions, there are pure stationary strategies* (x^*, y^*) *such that for all strategies x and y,*

$$\bar{L}^i(x, y^*) \leqq L^i(x^*, y^*) \leqq \underline{L}^i(x^*, y), \qquad 1 \leqq i \leqq M, \tag{4}$$

where

$$\bar{L}^i(x, y) = \limsup_{n\to\infty} n^{-1} \sum_{j=1}^{n} a_j^i(x, y).$$

Proof. For $\beta \in [0, 1)$, let $D_\beta^i(x, y) = \sum_{n=1}^{\infty} \beta^{n-1} a_n^i(x, y)$, let X (Y) be the set of pure stationary strategies available to player I (II), and define $D_\beta(x, y)$, $\bar{L}(x, y)$, $\underline{L}(x, y)$ and $L(x, y)$ to be M-vectors with the obvious components. Observe that if we fix player I's (II's) strategy, then our stochastic game is precisely the standard [1] finite state and action space Markovian decision process. Hence, we can conclude from Blackwell's Theorem 5 [1] that for each $x \in X$ $(y \in Y)$ there is a $y_x \in Y$ $(x_y \in X)$ and a number β_x (β_y) such that for all $y \in Y$ $(x \in X)$ and all $\beta \geqq \beta_x(\beta_y)$ we have

$$D_\beta(x, y_x) \leqq D_\beta(x, y), \qquad (D_\beta(x, y) \leqq D_\beta(x_y, y)). \tag{5}$$

Shapley has shown [5, Theorem 2] that for each $\beta \in [0, 1)$ there is a pair $(x_\beta, y_\beta) \in X \times Y$ such that for any pair (x, y) of strategies we have

$$D_\beta(x, y_\beta) \leqq D_\beta(x_\beta, y_\beta) \leqq D_\beta(x_\beta, y). \tag{6}$$

Then letting $\beta^* = \max\{\max\{\beta_x : x \in X\};\ \max\{\beta_y : y \in Y\}\}$ and defining $(x^*, y^*) = (x_{\beta^*}, y_{\beta^*})$, we can conclude from (5) and (6) that

$$D_\beta(x, y^*) \leqq D_\beta(x^*, y^*) \leqq D_\beta(x^*, y) \tag{7}$$

for any pair (x, y) of strategies and all $\beta \geqq \beta^*$.

Consequently, we can employ Brown's result [2, Theorem 4.2] that

$$H^n(x^*, y^*) - \sup_x H^n(x, y^*) \text{ is bounded uniformly in } n, \tag{8}$$

where $H^n(x, y) = \sum_{i=1}^{n} a_i(x, y)$. Hence, it follows that for each strategy x

$$\begin{aligned} L(x^*, y^*) &= \lim_{n\to\infty} n^{-1} \sup_x H^n(x, y^*) \\ &\geqq \limsup_{n\to\infty} n^{-1} H^n(x, y^*) = \bar{L}(x, y^*). \end{aligned}$$

Similarly, we can show that $L(x^*, y^*) \leqq \underline{L}(x^*, y)$ since $H^n(x^*, y^*) - \inf_y H^n(x^*, y)$ is bounded uniformly in n. This completes the proof.

As noted in footnote 2, perfect information in a stochastic game means that at each of the M positions one or the other of the players is restricted to a single decision. We can extend this definition to encompass the following two cases: (i) There is a subset S of the M positions such that if at any time the players are in position $i \in S$, then player I makes his decision after having observed player II's decision. Similarly, player I decides first if $i \notin S$. (ii) Player I (II) makes his decision after having observed player II's (I's) decision at times $1, 3, 5, \cdots$ $(2, 4, 6, \cdots)$. The first extension is established as follows. First, we can show that if B and C are two matrices of the same size and if val $[B]$ and val $[C]$ denote the values of the two games with payoff matrices B and C, where one player chooses after having observed the choice of the other player, then $|\text{val}[B] - \text{val}[C]| \leqq \max_{i,j} |b_{ij} - c_{ij}|$. Second, note that optimal strategies for the above games are pure. Now, for each β in $[0, 1)$ we can use Shapley's techniques [5] to obtain the existence of pure stationary strategies which satisfy (6). The proof now continues as in Theorem 1. The above argument also suffices to establish the second extension after the state space has been enlarged in the obvious manner to $2M$ positions.

Finally, we note that as a consequence of (7) we can establish (see [3]) a stronger form of Theorem 1 by replacing (4) with

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \tilde{H}^j(x, y^*) &\leqq \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \tilde{H}^j(x^*, y^*) \\ &\leqq \liminf_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \tilde{H}^j(x^*, y), \end{aligned} \tag{9}$$

where $\tilde{H}^j(x, y) = H^j(x, y) - jL(x^*, y^*)$.

REFERENCES

[1] D. BLACKWELL, *Discrete dynamic programming*, Ann. Math. Statist., 33 (1962), pp. 719–726.

[2] B. W. BROWN, *On the iterative method of dynamic programming on a finite space discrete time Markov process*, Ibid., 36 (1965), pp. 1279–1285.

[3] E. V. DENARDO AND B. L. MILLER, *An optimality condition for discrete dynamic programming with no discounting*, Ibid., 39 (1968), pp. 1220–1228.

[4] D. GILLETTE, *Stochastic games with zero stop probabilities*, Contributions to the Theory of Games, vol. III, M. Dresher, A. W. Tucker, P. Wolfe, eds., Princeton University Press, Princeton, 1957, pp. 179–187.

[5] L. S. SHAPLEY, *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A., 39 (1953), pp. 1095–1100.

[6] E. C. TITCHMARCH, *The Theory of Functions*, 2nd ed., Oxford University Press, London, 1939.