The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 316–318

A NOTE ON POSITIVE DYNAMIC PROGRAMMING

BY ASHOK MAITRA

Indian Statistical Institute, Calcutta

1. Introduction. A positive dynamic programming problem is determined by four objects, S, A, q and r. S and A are non-empty Borel subsets of Polish spaces, q is a transition function on S given $S \times A$ and r is a bounded, non-negative, Borel measurable function on $S \times A$. We interpret S as the state space of some system and A as the set of actions available at each state. When the system is in state s and we take action a, the system moves to a new state s' according to the distribution $q(\cdot/s, a)$ and we receive an immediate return $r(s, a)$. The process is then repeated from the new state s', and we wish to maximise the total expected return over the infinite future.

A plan π is a sequence $\pi_1, \pi_2, \cdots$, where π_n tells you how to choose an action on the nth day, as a function of the previous history $h = (s_1, a_1, \cdots, a_{n-1}, s_n)$, by associating with each h (Borel measurably) a probability distribution $\pi_n(\cdot/h)$ on the Borel subsets of A. Certain types of plans are of special interest. A *semi-Markov plan* is a sequence $f_1, f_2, \cdots$, where each f_n is a Borel measurable map from $S \times S$ into A, and $f_n(s_1, s_n)$ is the action we take on the nth day if we start in state s_1 and the state on the nth day is s_n. A *Markov plan* is a sequence $f_1, f_2, \cdots$ where each f_n is a Borel measurable map from S into A and $f_n(s)$ is the action we choose on the nth day if the nth state is s. A *stationary* plan is a Markov plan in which $f_n = f$ for some Borel measurable map f from S to A and all n.

A plan π associates (Borel measurably) with each initial states a total expected return $I(\pi)(s)$. We shall assume that the structure of the problem is such that the optimal return $v^* = \sup_\pi I(\pi)$ is a finite function on S. [Note that we are not assuming that v^* is bounded].

This problem has been studied by Blackwell [1], Strauch [6] and Barbosa Dantas [2]. An example due to Blackwell shows that ϵ-optimal plans need not exist (see Example 4.1 in [6]) and moreover, that the optimal return need not be Borel measurable. The purpose of this note is to impose certain topological conditions on A, q and r and show that under these assumptions there will exist ϵ-optimal plans and that the optimal return will be Borel measurable. Specifically, we shall prove the

THEOREM. *Let S be a Borel subset of a Polish space, A a compact metric space and r a bounded, non-negative, upper semi-continuous* (*abbreviated, hereafter, by* usc) *function on $S \times A$. Assume, furthermore, that $(s_n, a_n) \rightarrow (s_0, a_0)$ implies $q(\cdot/s_n, a_n)$ converges weakly to $q(\cdot/s_0, a_0)$. Then, for any $\epsilon > 0$, there exists an ϵ-optimal semi-Markov plan π* (*that is, $I(\pi) \geqq v^* - \epsilon$*)) *and, moreover, the optimal return v^* is a Baire function of the second class.*

Note that if S is countable and A finite, the conditions of the above theorem are fulfilled.

Received 7 February 1968.

2. Proof of theorem. Throughout this section, the conditions imposed on S, A, q and r in the theorem stated above will remain operative.

The proof of the theorem rests on a selection theorem due to Dubins and Savage [3]. We state it here in a form somewhat different from that in which Dubins and Savage have stated it but which is immediately applicable to our problem.

SELECTION THEOREM. *Let u be a bounded* usc *function on $S \times A$. Define u^*: $S \to (-\infty, \infty)$ by: $u^*(s) = \max_{a \varepsilon A} u(s, a)$. Then u^* is* usc *and there exists a Borel measurable function f from S to A such that $u^*(s) = u(s, f(s))$ for all $s \varepsilon S$.*

The proof may be found in [3], page 38 or in [5].

We shall also require the following:

LEMMA. *Let v be a bounded* usc *function on S. Then w: $S \times A \to (-\infty, \infty)$ defined by: $w(s, a) = \int v(\cdot)\, dq(\cdot/s, a)$ is* usc.

PROOF. First, note that if v' is any bounded continuous function on S, then it follows from the condition imposed on q that the function $(s, a) \to \int v'(\cdot)\, dq\cdot(\cdot/s, a)$ is continuous. Next, as v is a bounded usc function, there exists a sequence $\{v_n\}$ of bounded continuous functions on S such that $v_n \downarrow v$ (by Theorem 3.3.8 in [4]). Hence, the functions w_n on $S \times A$ defined by $w_n(s, a) = \int v_n(\cdot)\, dq(\cdot/s, a)$ are continuous, and, by the dominated convergence theorem, $w_n \downarrow w$. Consequently, w is usc ,which completes the proof of the lemma.

PROOF OF THEOREM. In the dynamic programming problem, denote, for each $n \geqq 1$, by u_n^* the optimal return over n days of play. Each u_n^* is then a bounded, non-negative function on S, and, moreover, $u_n^* \uparrow u^*$ (say). We shall show by induction that each u_n^* is usc on S. Note that

$$(1) \qquad u_1^*(s) = \max_{a \varepsilon A} r(s, a) \quad \text{for all} \quad s \varepsilon S,$$

so that it follows from the Selection Theorem that u_1^* is usc. Suppose for $n = m$, u_m^* is usc. Then it is easy to see that

$$(2) \quad u_{m+1}^*(s) = \max_{a \varepsilon A} [r(s, a) + \int u_m^*(\cdot)\, dq(\cdot/s, a)] \quad \text{for all} \quad s \varepsilon S.$$

The lemma above together with the inductive hypothesis ensures that the second term inside square brackets on the right-hand side of (2) is usc on $S \times A$, so that the entire expression within square brackets is usc on $S \times A$. Thus, the 'max' is justified in (2). Consequently, it follows once again from the Selection Theorem that u_{m+1}^* is usc on S. As u^* is a point-wise limit of the usc functions u_n^*, it is a Baire function of the second class. From (2), we get

$$(3) \qquad u_{m+1}^*(s) \geqq [r(s, a) + \int u_m^*(\cdot)\, dq(\cdot/s, a)] \qquad \text{for all } s, a \text{ and } m.$$

Keeping s and a fixed, let $m \to \infty$ in (3). By the monotone convergence theorem, we have:

$$(4) \qquad u^*(s) \geqq [r(s, a) + \int u^*(\cdot)\, dq(\cdot/s, a)] \qquad \text{for all } s \text{ and } a.$$

Theorem 2 in [1] now implies that the optimal return (over the infinite future) $v^* \leqq u^*$.

Again from the Selection Theorem and (1) and (2), we get the existence of

Borel measurable maps, f_n, $n \geqq 1$, from S to A such that

$$u_1^*(s) = r(s, f_1(s)) \quad \text{for all} \quad s \varepsilon S \tag{5}$$

and

$$u_{n+1}^*(s) = r(s, f_{n+1}(s)) + \int u_n^*(\cdot)\, dq(\cdot/s, f_{n+1}(s)) \qquad \text{for all } s \text{ and } n. \tag{6}$$

Now we can construct an ϵ-optimal semi-Markov plan as follows: Let $\epsilon > 0$ and let g be a fixed (but otherwise arbitrary) Borel measurable map from S to A. Define

$$S_1 = \{s: u_1^*(s) \geqq u^*(s) - \epsilon\} \qquad \text{and, for} \quad n \geqq 2,$$

$$S_n = \{s: u_{n-1}^*(s) < u^*(s) - \epsilon,\ u_n^*(s) \geqq u^*(s) - \epsilon\}.$$

The sets S_n are Borel, disjoint and $\bigcup_{n=1}^\infty S_n = S$. Define $g_1 = f_n$ on S_n, $n \geqq 1$, and for $m \geqq 2$, define $g_m(s, s') = g(s')$ if $s \ \varepsilon \ S_1 \cup S_2 \cup \cdots \cup S_{m-1}$, and $g_m(s, s') = f_{n-m+1}(s')$, if $s \varepsilon S_n$, $n \geqq m$. Then $\pi = \{g_1, g_2, \cdots\}$ is our required semi-Markov plan. For, it is easy to see, using (5) and (6), that if $s \varepsilon S_n$, $I(\pi)(s) \geqq u_n^*(s) \geqq u^*(s) - \epsilon$. Consequently, $I(\pi) \geqq u^* - \epsilon$, which proves that (as ϵ is arbitrary) $v^* = u^*$ and π is ϵ-optimal. Moreover, the optimal return is a Baire function of the second class. This completes the proof of the theorem.

Remark 1. Our theorem is the dynamic programming analogue of Theorem 2.16.1 in [3].

Remark 2. Blackwell has given an example in [1], which satisfies the conditions of our theorem, but for which an optimal plan does not exist. The same example shows that ϵ-optimal stationary plans need not exist. Whether or not, under our conditions, ϵ-optimal Markov plans exist, we have not been able to determine.

REFERENCES

[1] Blackwell, D. (1965). Positive dynamic programming. *Fifth Berkeley Symp. Math. Statist. Prob.* **1** 415–418. Univ. of California Press.

[2] Barbosa Dantas, C. A. (1966). On the existence of stationary optimal plans. Doctoral dissertation. Univ. of California, Berkeley.

[3] Dubins, L. E. and Savage, L. J. (1965). *How to Gamble If You Must.* McGraw-Hill, New York.

[4] McShane, E. J. and Botts, T. A. (1959). *Real Analysis.* Van Nostrand, Princeton.

[5] Maitra, A. (1968). Discounted dynamic programming on compact metric spaces. *Sankhyā Ser. A* **30** 211–216.

[6] Strauch, R. E. (1966). Negative dynamic programming. *Ann. Math. Statist.* **37** 871–890.