# NOTES

## A NOTE ON MEMORYLESS RULES FOR CONTROLLING SEQUENTIAL CONTROL PROCESSES[1]

BY CYRUS DERMAN AND RALPH E. STRAUCH[2]

*Columbia University and University of California, Berkeley*

**1. Introduction and summary.** We are concerned with a dynamic system which at times $t = 0, 1, \cdots$ is observed to be in one of a finite number of states. We shall denote the space of possible states by $I$. After each observation the system is "controlled" by making one of a finite number of possible decisions. We shall denote by $K_i$ the set of possible decisions when the system is in state $i$, $i \varepsilon I$. Let $\{Y_t\}$, $t = 0, 1, \cdots$, denote the sequence of observed states and $\{\Delta_t\}$, $t = 0, 1, \cdots$, the sequence of observed decisions. The fundamental assumption regarding $\{Y_t, \Delta_t\}$, $t = 0, 1, \cdots$, is

(A) $P(Y_{t+1} = j \mid Y_0, \Delta_0, \ldots, {}_tY_t = i, \Delta_t = k) = q_{ij}(k), \quad t = 0, 1, \cdots; j \varepsilon I; k \varepsilon K_i$

where the $q_{ij}(k)$'s are non-negative numbers satisfying $\sum_j q_{ij}(k) = 1$, $k \varepsilon K_i$; $i \varepsilon I$.

A rule for making successive decisions can be summarized in the form of a collection of non-negative functions

$$D_k(Y_0, \Delta_0, \cdots, \Delta_{t-1}, Y_t), \qquad t = 0, 1, \cdots; k \varepsilon K_{Y_t},$$

where in every case $\sum_k D_k(\cdot) = 1$. We set

$$P(\Delta_t = k \mid Y_0, \Delta_0, \cdots, \Delta_{t-1}, Y_t) = D_k(Y_0, D_0, \cdots, \Delta_{t-1}, Y_t)$$

for $t = 0, 1, \cdots$. Thus, given $Y_0 = i$ and any rule $R$ for making successive decisions, the sequence $\{Y_t, \Delta_t\}$, $t = 0, 1, \cdots$, is a stochastic process with its probability measure dependent upon the rule $R$. We refer to such a process as a *sequential control process*.

Let $C$ denote the class of *all* possible rules; $C'$ denote the class of all rules such that

$$D_k(Y_0, \Delta_0, \cdots, \Delta_{t-1}, Y_t = i) = D_{ik}, \qquad t = 0, 1, \cdots; k \varepsilon K_i; i \varepsilon I.$$

That is, $C'$ is the class of all rules such that the mechanism for making a decision at any time $t$ is dependent only on the state of the system at time $t$. A rule $R \varepsilon C'$ has a stationary Markovian character and, indeed, when $R \varepsilon C'$ is used,

Received 3 March 1965; revised 1 August 1965.

[1] Work supported in part by the Office of Naval Research and a National Science Foundation Graduate Fellowship. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at The RAND Corporation.

the resulting process $\{Y_t\}$, $t = 0, 1, \cdots$, is a Markov chain with stationary transition probabilities. We let $C''$ denote the subclass of $C'$ where the $D_{ik}$'s are zero or one. Rules in $C'$ allow for randomization; the rules in $C''$ are non-randomized.

For a given $R \varepsilon C$ and initial state $Y_0 = i$, let

$$X_{T,j,k,R}(i) = (T + 1)^{-1}\textstyle\sum_{t=0}^{T} P_R(Y_t = j, \Delta_t = k \mid Y_0 = i)$$

and let $X_{T,R}(i)$ denote the vector of components $X_{T,j,k,R}(i)$ for all $k \varepsilon K_j$ and $j \varepsilon I$. Denote by $H_R(i)$ the set of limit points of $X_{T,R}(i)$ as $T \to \infty$. Let

$$H(i) = \textstyle\bigcup_{R\varepsilon C} H_R(i), \qquad H'(i) = \bigcup_{R\varepsilon C'} H_R(i), \qquad H''(i) = \bigcup_{R\varepsilon C''} H_R(i);$$

and let $\bar{H}'(i)$ and $\bar{H}''(i)$ denote the convex hulls of $H'(i)$ and $H''(i)$, respectively. In [5] was proved

THEOREM 1. (a) $\bar{H}'(i) = \bar{H}''(i) \supset H(i)$.

(b) *If the Markov chain corresponding to $R$ is irreducible for every $R \varepsilon C''$, then* $\bar{H}''(i) = H'(i) = H(i) = \bigcup_{i\varepsilon I} H(i)$.

Examples were given in [4] and [5] showing that $H(i)$ can be larger than $H'(i)$. In (b) the irreducibility assumption can be weakened to the condition that for each $R \varepsilon C''$ the corresponding Markov chain has, at most, one ergodic class.

Blackwell [1], [2], and Maitra [6] have considered *memoryless* rules. By a memoryless rule we mean a rule $R$ such that

$$D_t(Y_0, \Delta_0, \cdots, \Delta_{t-1}, Y_t = i) = D_{ik}^{(t)} \qquad t = 0, 1, \cdots, k \varepsilon K_i, i \varepsilon I.$$

That is, with a memoryless rule the mechanism for making a decision is a function of the time $t$ and the state $i$ at time $t$. The memoryless rules of Blackwell and Maitra are non-randomized; i.e., $D_{ik}^{(t)} = 0$ or 1. We shall let $C^M$ denote the class of memoryless rules (both randomized and non-randomized). Thus $C \supset C^M \supset C' \supset C''$. If $R \varepsilon C^M - C'$, then $\{Y_t\}$, $t = 0, 1, \cdots$, is a finite state Markov chain with non-stationary transition probabilities.

We remark that it is the memoryless rules (non-randomized) that are considered the rules of interest in the usual finite horizon dynamic programming problems. See Blackwell [2] for interesting remarks along these lines. We are concerned with optimization problems where the criterion to be optimized are functions of the points in $H(i)$. In [5] it was shown that one can construct problems where the optimal rule is in $C - C'$. This can occur, e.g., if the criterion to be optimized is a linear functional over the points in $H(i)$ but where a solution must also satisfy one or more linear constraints in $H(i)$. It is for the purpose of treating optimization problems of this kind that we are interested in the limit points of $X_{T,R}(i)$ for rules belonging to the various important sub-classes of $C$.

Let $H^M(i) = \bigcup_{R\varepsilon C^M} H_R(i)$. The result of this note (which is similar to Theorem 4.1 of [7]) is

THEOREM 2.

$$H(i) = H^M(i).$$

*In fact, for any $R \varepsilon C$ there exists an $R_0 \varepsilon C^M$ such that $X_{tR_0}(i) = X_{tR}(i)$ for all $t$.*

PROOF. Define $R_0$ by

$$D_{jk}(t) = P_{R_0}(\Delta_t = k \mid Y_t = j) = P_R(\Delta_t = k \mid Y_t = j, Y_0 = i).$$

It is enough to show that for all $t$, $k \varepsilon K_j$ and $j \varepsilon I$,

$$(*) \quad P_{R_0}(Y_t = j, \Delta_t = k \mid Y_0 = i) = P_R(Y_t = j, \Delta_t = k \mid Y_0 = i).$$

The relation $(*)$ holds for $t = 0$, since

$$P_R(Y_0 = j, \Delta_0 = k \mid Y_0 = i) = 0 = P_{R_0}(Y_0 = j, \Delta_0 = k \mid Y_0 = i)$$

if $j \neq i$ and

$$\begin{aligned} P_R(Y_0 = i, \Delta_0 = k \mid Y_0 = i) &= P_R(\Delta_0 = k \mid Y_0 = i) \\ &= P_{R_0}(\Delta_0 = k \mid Y_0 = i) \end{aligned}$$

by definition. Now assume $(*)$ holds for $t = 0, \cdots, T-1$. Then

$$P_R(Y_T = j, \Delta_T = k \mid Y_0 = i) = P_R(Y_T = j \mid Y_0 = i) P_R(\Delta_T = k \mid Y_T = j, Y_0 = i)$$

but

$$P_R(\Delta_T = k \mid Y_T = j, Y_0 = i) = D_{jk}(t)$$

by definition, and

$$\begin{aligned} P_R(Y_T = j \mid Y_0 = i) &= \textstyle\sum_{l \varepsilon I} \sum_{k \varepsilon K_l} P_R(Y_{T-1} = l, \Delta_{T-1} = k \mid Y_0 = i) q_{lj}(k) \\ &= \textstyle\sum_{l \varepsilon I} \sum_{k \varepsilon K_l} P_{R_0}(Y_{T-1} = l, \Delta_{T-1} = k \mid Y_0 = i) q_{lj}(k) \\ &= P_{R_0}(Y_T = j \mid Y_0 = i) \end{aligned}$$

by the induction hypothesis. Thus

$$\begin{aligned} P_R(Y_T = j, \Delta_T = k \mid Y_0 = i) &= P_{R_0}(Y_T = j \mid Y_0 = i) D_{jk}(t) \\ &= P_{R_0}(Y_T = j, \Delta_T = k \mid Y_0 = i). \end{aligned}$$

This completes the proof.

## REFERENCES

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

[2] BLACKWELL, DAVID (1964). Memoryless strategies in finite-stage dynamic programming. *Ann. Math. Statist.* **35** 863–865.

[3] DERMAN, CYRUS (1962). On sequential decisions and Markov chains. *Management Sci.* **9** 16–24.

[4] DERMAN, CYRUS (1963). Stable sequential rules and Markov chains. *J. Math. Anal. Appl.* **6** 257–265.

[5] DERMAN, CYRUS (1964). On sequential control processes. *Ann. Math. Statist.* **35** 341–349.

[6] MAITRA, ASHOK (1963). Dynamic programming for countable state systems. Doctoral dissertation, Univ. of California, Berkeley.

[7] STRAUCH, RALPH EUGENE (1965). Negative dynamic programming. Doctoral dissertation, Univ. of California, Berkeley.