ON SEQUENTIAL CONTROL PROCESSES[1]

By Cyrus Derman

Columbia University

1. Introduction. Consider a dynamic system which at times $t = 0, 1, \cdots$ is observed to be in one of $L + 1$ states $0, \cdots, L$. After each observation the system is "controlled" by making one of K decisions $d_1, \cdots, d_K$. Let $\{Y_t\}$, $t = 0, 1, \cdots$, denote the sequence of observed states and $\{\Delta_t\}$, $t = 0, 1, \cdots$, the sequence of decisions. We shall assume that

$$P(Y_{t+1} = j \mid Y_0, \Delta_0, \cdots, Y_t = i, \Delta_t = d_k) = q_{ij}(k),$$
$$t = 0, 1, \cdots; \quad j = 0, \cdots, L; \quad k = 1, \cdots, K$$

where the $q_{ij}(k)$'s are non-negative numbers satisfying

$$\sum_{j=0}^{L} q_{ij}(k) = 1, \qquad i = 0, \cdots, L; \quad k = 1, \cdots, K.$$

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

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

where in every case $\sum_{k=1}^{K} D_k = 1$. We set

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

for $t = 0, 1 \cdots; k = 1, \cdots, K$. Thus, given any rule R for making the successive decisions, the sequence $\{Y_t\}$, $t = 0, 1, \cdots$, is a stochastic process possessing a finite state space $0, \cdots, L$ with its probability measure dependent upon the way the rule brings into play the conditional probabilities $q_{ij}(k)$. In particular, if the rule R is of the form

$$D_k(Y_0, \Delta_0, \cdots, \Delta_{t-1}, Y_t = i) = D_{ik}, \qquad t = 0, 1, \cdots; \quad i = 0, \cdots, L,$$

where $\sum_{k=1}^{K} D_{ik} = 1$, $i = 0, \cdots, L$, then $\{Y_t\}$ is a finite state Markov chain with stationary transition probabilities

$$p_{ij} = \sum_{k=1}^{K} D_{ik} q_{ij}(k), \qquad i = 0, \cdots, L; \quad j = 0, \cdots, L. \tag{1.1}$$

Let C denote the class of *all* possible rules, C', the above class of randomized stationary type rules, and C'', the finite sub-class of C' for which the D_{ik}'s are either 0 or 1.

Received 5 July 1963.

[1] Work sponsored by the Office of Naval Research.

For every $R \varepsilon C$ and any initial state $Y_0 = i$, we consider the $K(L + 1)$ component vectors

$$\Phi_T^R(i) = \{x_{T01}, \cdots, x_{TLK}\}, \qquad T = 1, 2, \cdots,$$

where

$$x_{Tjk} = (T + 1)^{-1}\sum_{t=0}^{T} P(Y_t = j, \Delta_t = d_k \mid Y_0 = i),$$

$$j = 0, \cdots, L; k = 1, \cdots, K.$$

Let $\Phi^R(i) = \lim_{T\to\infty} \Phi_T^R(i)$ whenever the limit exists. This will be the case when $R \varepsilon C'$ (See Chung [3], p. 32.). At any rate, denote by $H_R(i)$ the set of limit points of $\{\Phi_T^R(i)\}$ as $T \to \infty$. Let

$$H(i) = \bigcup_{R\varepsilon C} H_R(i), \quad H'(i) = \bigcup_{R\varepsilon C'} H_R(i), \quad \text{and} \quad H''(i) = \bigcup_{R\varepsilon C''} H_R(i);$$

and let $\bar{H}'(i)$ denote the closure of the convex hull of $H'(i)$, and $\bar{H}''(i)$ the convex hull of $H''(i)$. One of the main results of this paper is

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* $\bar{H}''(i) = H'(i) = H(i) = \bigcup_{i=0}^{L} H(i)$.

This result is an extension of a theorem proved by Derman [7]. It is of interest when dealing with the determination of optimal rules. For if the criterion to be optimized is a function of the limit points of $\{\Phi_T^R(i)\}$, then the theorem is useful in reducing the problem to a consideration of rules in class C', C'', or an initial randomized selection of such rules. When consideration is limited to rules in C' or C'', methods of functional equations, linear programming, and Markov chains (see e.g. Bellman [1], [2], Blackwell [3], Derman [5], [6], Freimar [8], Howard [9], Klein [11], Manne [13]) are applicable.

Our other result is a strong law. Suppose $Y_0 = 1$ with probability 1. Define

$$Z_{tjk} = 1, \qquad \text{if } Y_t = j, \quad \Delta_t = d_k$$
$$= 0, \qquad \text{otherwise.}$$

Let $\tilde{\Phi}_T^R(i)$ be the vectors with the components

$$\tilde{x}_{Tjk} = (T + 1)^{-1}\sum_{t=0}^{T} Z_{tjk}, \qquad j = 0, \cdots, L; \quad k = 1, \cdots, K.$$

For a fixed $R \varepsilon C$ denote by ω a sample sequence of the joint process $\{Y_t, \Delta_t\} t = 0, 1, \cdots$. Let $U^R(i, \omega)$ be the set of limit points of $\tilde{\Phi}_T^R(i)$ as $T \to \infty$.

THEOREM 2. *For each $R \varepsilon C$, $P(U^R(i, \omega) \subset \bar{H}'') = 1$, where $\bar{H}''$ is the convex hull of $\bigcup_{j=0}^{L} H''(j)$.*

This theorem is also of some use in finding optimal rules when the criterion is in terms of sample frequencies rather than expected frequencies as in Theorem 1.

2. Proof of Theorem 1. The proof of part (b) follows that in [7] and will not be given here.

Let

$$f(\Phi_T^R(i)) = \sum_{j=0}^{L} \sum_{k=1}^{K} w_{jk} x_{Tjk},$$

where $\{w_{jk}\}, j = 0, \cdots, L; k = 1, \cdots, K$, are any given numbers. The following was proven in [5].

LEMMA 1. *There exists an* $R^* \varepsilon C''$ *such that*

$$f(\Phi^{R^*}(i)) = \min_{R \varepsilon C} \limsup_{T \to \infty} f(\Phi_T^R(i)).$$

We shall say a state i is recurrent with respect to the class C' if for every j there is a rule $R(j) \varepsilon C'$ such that $P(Y_t = i \text{ for some } t > 0 \mid Y_0 = j) = 1$ when rule $R(j)$ is used. We shall prove

LEMMA 2. *If* i *is recurrent with respect to* C', *then* $\bar{H}(i) \subset \bar{H}'(i)$.

PROOF. We prove this in two steps. Let C^* be the class of rules R for which $\lim_{T\to\infty} \Phi_T^R(i)$ exists; let $H^*(i) = \bigcup_{R \varepsilon C^*} H_R(i)$. We shall first show that $H^*(i) \subset \bar{H}'(i)$. Suppose, to the contrary, that there exists a rule $R \varepsilon C^*$ such that

$$\lim_{T\to\infty} \Phi_T^R(i) = \Phi^R(i) = z = (z_{01}, \cdots, z_{LK}) \not\varepsilon \bar{H}'(i).$$

It is a well-known property of convex sets (see Karlin [10] p. 397) that there exist numbers $w_{01}, \cdots, w_{LK}$ such that

$$\sum_{j=0}^{L} \sum_{k=1}^{K} w_{jk} z_{jk} < \sum_{j=0}^{L} \sum_{k=1}^{K} w_{jk} x_{jk} \tag{2.1}$$

for all $x = (x_{01}, \cdots, x_{LK}) \varepsilon \bar{H}'(i)$. However, by Lemma 1, there exists a rule $R^* \varepsilon C''$ such that $f(\Phi^{R^*}(i)) \leqq f(\Phi^R(i))$, a contradiction of (2.1), proving the assertion.

Thus having shown that $H^*(i) \subset \bar{H}'(i)$, we now show that for any $R \varepsilon C$ with $z = (z_{01}, \cdots, z_{LK}) \varepsilon H(i)$ a limit point of $\{\Phi_T^R\{(i)\}$, we can construct a sequence of rules $R_v \varepsilon C^*$, $v = 1, \cdots$, such that $\lim_{v\to\infty} \Phi^{R_v}(i) = z$. Since $\bar{H}'(i)$ is closed and $H^*(i) \subset \bar{H}'(i)$ this implies that $z \varepsilon \bar{H}'(i)$. Let $T_1, T_2, \cdots$ be such that $\lim_{v\to\infty} \Phi_{T_v}^R(i) = z$. We construct R_v as follows. Use rule R until $t = T_v$; then use, as if starting from $t = 0$, rule $R(j)$ (if $Y_{T_v} = j, j = 0, \cdots, L$) until i is reached for the first time after $t = t_v$. (We can assume that under $R(j)$, $P(Y_t = i \text{ for some } t > 0 \mid Y_0 = j) = 1$ because of the hypothesis on the lemma.) Then revert, as if starting from $t = 0$, to the use of rule R for T_v units of time, $\cdots$, etc. Let η_{ji} denote the mean first passage time from j to i using $R(j)$ and let $M = \max_j \{\eta_{ji}\}$. By well-known results from finite state Markov chain theory, $M < \infty$. Since R_v generates a renewal type process it is easily seen by standard methods that $R_v \varepsilon C^*$ and

$$\begin{aligned} \frac{1}{T_v + M} \sum_{t=1}^{T_v} P(Y_t = j, \Delta_t = d_k \mid Y_0 = i) &\leqq x_{jk}^{(v)} \\ &\leqq \frac{1}{T_v} \left\{ \sum_{t=1}^{T_v} P(Y_t = j, \Delta_t = d_k \mid Y_0 = i) + M \right\} \end{aligned} \tag{2.2}$$

for $j = 0, \cdots L$; $k = 1, \cdots K$ where $x_{jk}^{(v)}$ is the (j, k)th coordinate of $\Phi^{R_v}(i)$. On letting $v \to \infty$, it follows from (2.2) that $\lim_{v\to\infty} \Phi^{R_v}(i) = z$, which was to be proved.

LEMMA 3. *There exists an* $R^{**} \varepsilon C''$ *such that*

$$f(\Phi^{R^{**}}(i)) = \max_{R\varepsilon C} \limsup_{T\to\infty} f(\Phi_T^R(i)).$$

PROOF. First suppose that i is recurrent with respect to C'. Since f is a linear function there exists an extreme point $x_0 \varepsilon \bar{H}'(i)$ for which $\max_{x\varepsilon \bar{H}'(i)} f(x) = f(x_0)$. Since x_0 is an extreme point of $\bar{H}'(i)$ there exists a rule $R_0 \varepsilon C'$ such that $\Phi^{R_0}(i) = x_0$. Let R be an arbitrary rule in C, and let $T_1, T_2, \cdots$ be such that $\lim_{v\to\infty} f(\Phi_{T_v}^R(i)) = \limsup_{T\to\infty} f(\Phi_T^R(i))$ and $\lim_{v\to\infty} \Phi_{T_v}^R(i) = x$. By Lemma 2, $x \varepsilon \bar{H}'(i)$. Therefore, by the continuity of f, we have

$$\limsup_{T\to\infty} f(\Phi_T^R(i)) = \lim_{v\to\infty} f(\Phi_{T_v}^R(i)) = f(x) \leqq f(x_0) = f(\Phi^{R_0}(i)).$$

Now suppose i is not recurrent with respect to C'. Let us then "artificially" adjoin the decision d_{K+1} to the possible decisions such that $q_{ji}(K + 1) = 1$, $j = 0, \cdots, L$, and $\{-w_{jK+1}\}$, $j = 0, \cdots L$, are arbitrarily large. With the adjoined decision, i is recurrent with respect to C'; however, with the w_{jK+1}'s so chosen, the $R_0 \varepsilon C'$ as given above cannot have the property that $\Delta_t = d_{K+1}$ for any state at any time t. Thus R_0 maximizes $\limsup_{T\to\infty} f(\Phi_T^R(i))$ over the original decision possibilities $d_1, \cdots, d_K$; that is, we have shown there is a rule $R_0 \varepsilon C'$ such that

$$f(\Phi^{R_0}(i)) = \max_{R\varepsilon C} \limsup_{T\to\infty} (\Phi_T^R(i)).$$

From Lemma 1 there exists a rule $R^{**} \varepsilon C''$ such that

$$f(\Phi^{R^{**}}(i)) = \max_{R\varepsilon C} \liminf_{T\to\infty} f(\Phi_T^R(i))$$
$$\geqq \liminf_{T\to\infty} f(\Phi_T^{R_0}(i)) = \lim_{T\to\infty} f(\Phi_T^{R_0}(i)) = f(\Phi^{R_0}(i));$$

hence $f(\Phi^{R^{**}}(i)) = f(\Phi^{R_0}(i))$ and Lemma 3 is proved.

Lemmas 1 and 3 imply that for every $z \varepsilon H(i)$ and any linear function f there exists a rule $R_1 \varepsilon C''$ such that

$$f(z) \leqq f(\Phi^{R_1}(i)). \tag{2.3}$$

However, this is enough to imply part (a) of Theorem 1. For if $z \varepsilon H(i)$ but $z \not\varepsilon \bar{H}''(i)$, then there exists a linear function f such that $f(z) > f(x)$ for all $x \varepsilon \bar{H}''(i)$. This is a contradiction of (2.3). Thus $H(i) \subset \bar{H}''(i)$. The same argument can be made to show that if $z \varepsilon \bar{H}'(i)$ then $z \varepsilon \bar{H}''(i)$ since any vector in $\bar{H}'(i)$ can be expressed as a linear combination of vectors in $H'(i)$. Thus $\bar{H}'(i) \subset \bar{H}''(i)$. Trivially $\bar{H}'(i) \supset \bar{H}''(i)$, since $H'(i) \supset H''(i)$. Therefore, $\bar{H}'(i) = \bar{H}''(i)$.

3. Corollaries of Theorem 1. With respect to finding optimal rules the results of Section 2 have some immediate implications.

Corollary 1. *There exist rules $R^* \varepsilon C''$ and $R^{**} \varepsilon C''$ such that*

$$f(\Phi^{R^*}(i)) = \min_{R \varepsilon C} \limsup_{T\to\infty} f(\Phi_T^R)i)), \qquad i = 0, \cdots, L,$$

and

$$f(\Phi^{R^{**}}(i)) = \max_{R \varepsilon C} \limsup_{T\to\infty} f(\Phi_T^R(i)), \qquad i = 0, \cdots, L.$$

This corollary asserts that the same rule holds for all initial states $Y_0 = i$. The proof of this follows from Lemmas 1 and 3 together with

Lemma 4. *Suppose $R_i \varepsilon C'$, $i = 0, \cdots, L$, satisfy*

$$f(\Phi^{R_i}(i)) = \min_{R \varepsilon C'} (\max) f(\Phi^R(i)), \qquad i = 0, 1, \cdots, L.$$

Denote by D_i the ith *row of the matrix characterizing R_i, $i = 0, \cdots, L$; then $R^* \varepsilon C'$, defined as the rule characterized by the matrix D^* having D_i for its* ith *row, $i = 0, \cdots, L$, satisfies $f(\Phi^{R^*}(i)) = f(\Phi^{R_i}(i))$, $i = 0, \cdots, L$.*

Proof. Let p_{ij} be given by (1.1) for the rule R^*. Note that

$$\begin{aligned} f(\Phi^{R_i}(i) &= f\left(\sum_{j=0}^{L} p_{ij}\Phi^{R_i}(j)\right) \\ &= \sum_{j=0}^{L} p_{ij} f(\Phi^{R_i}(j)) = \sum_{j=0}^{L} p_{ij} f(\Phi^{R_j}(j)). \end{aligned}$$

On iterating we get

$$f(\Phi^{R_i}(i)) = \sum_{j=0}^{L} p_{ij}^{(t)} f(\Phi^{R_j}(j)), \qquad t \geqq 1,$$

where $p_{ij}^{(t)}$ is the t-step transition probability under R^*. However, the right hand side is $f(\Phi^{R^*}(i))$, which proves the lemma.

Corollary 2. *If g is a concave (convex) function on $K(L + 1)$-Euclidean space, then there exists a rule $R_0 \varepsilon C''$ such that*

$$g(\Phi^{R_0}(i) = \min_{R \varepsilon C} (\max) \limsup_{T\to\infty} g(\Phi_T^R(i)).$$

Using part (a), the proof of Corollary 2 is a matter of details keeping in mind that the minimum (maximum) of a concave (convex) function over a closed bounded convex set is obtained at an extreme point.

Corollary 3. *If the Markov chain corresponding to R is irreducible for every $R \varepsilon C''$, and if g is any continuous function on $K(L + 1)$-Euclidean space, then there exists a rule $R_0 \varepsilon C'$ such that*

$$g(\Phi^{R_0}(i) = \min_{R \varepsilon C} (\max) \limsup_{T\to\infty} g(\Phi_T^R(i)).$$

The corollary is a consequence of part (b) of Theorem 1.

4. A Counter-example. The question arises whether we can assert that $H(i) = H'(i)$ under weaker conditions than the irreducibility condition of Theorem 1. An example was given in [7] to show that in general it is not true.

However, one may conjecture that it is true under the assumption that i is recurrent with respect to C'. The following example proves the contrary.

Suppose we have states 0, 1, 2 and decisions d_1 and d_2 such that

$$q_{00}(1) = 0,\quad q_{01}(1) = 1,\quad q_{02}(1) = 0,\quad q_{00}(2) = 0,\quad q_{01}(2) = 1,\quad q_{02}(2) = 0;$$

$$q_{10}(1) = 0,\quad q_{11}(1) = 1,\quad q_{12}(1) = 0,\quad q_{10}(2) = 0,\quad q_{11}(2) = 0,\quad q_{12}(2) = 1;$$

$$q_{20}(1) = 0,\quad q_{21}(1) = 0,\quad q_{22}(1) = 1,\quad q_{20}(2) = 1,\quad q_{21}(2) = 0,\quad q_{22}(2) = 0.$$

In words, whenever the system is in state 0 a transition will take place to state 1 irrespective of which decision is made; when the system is in state 1, it remains there if decision d_1 is made and it changes to state 2 if decision d_2 is made; whenever the system is in state 2 it remains in state 2 or proceeds to state 0 according to whether decision d_1 or d_2 is made. Here state 0 is recurrent with respect to C'. Let

$$\pi_{T0j} = (T + 1)^{-1}\sum_{t=0}^{T} P(Y_t = j \mid Y_0 = 0), \qquad j = 0, 1, 2.$$

When $R \varepsilon C'$, $\lim_{T\to\infty} \pi_{T0j}$ exists. Let $\pi(0) = \lim_{T\to\infty} (\pi_{T00}, \pi_{T01}, \pi_{T02})$, actually $\pi(0) = \sum_{k=1}^{2} \lim_{T\to\infty} \Phi_T^R(0) = \lim_{T\to\infty} \sum_{k=1}^{2} \Phi_T^R(0)$. However, it follows from well-known results in Markov chain theory that under any rule $R \varepsilon C'$, the only possible vectors $\pi(0)$ are of the form $\pi(0) = (0, 1, 0)$, $(0, 0, 1)$, or (a, b, c) where $a > 0$, $b > 0$, $c > 0$ and $a + b + c = 1$. The vector of the first form will occur if $D_{11} = 1$; the second, if $D_{11} < 1$ and $D_{21} = 1$; the third, if $D_{11} < 1$, $D_{21} < 1$.

However, consider the rule $R \not\varepsilon C'$ as follows. Given that $Y_0 = 0$, let $\Delta_0 = 1$ or 2. Then, of course, $Y_1 = 1$ with probability 1. Let P $(\Delta_t = 1 \; Y_0 = 0, Y_1 = 1, \cdots, Y_t = 1) = e^{-(\frac{1}{2})^t}$; $P(\Delta_t = 1 \mid Y_t = 2) = 1$. Then $P(Y_t = 0 \mid Y_0 = 0) = 0$, $t > 0$, and $P(Y_t = 1 \mid Y_0 = 0) = \exp[-\frac{1}{2} \sum_{v=0}^{t-2} (\frac{1}{2})^v] \to e^{-1}$ as $t \to \infty$. Hence, using R, $\pi(0) = (0, e^{-1}, 1 - e^{-1})$ which is not possible under any rule in C'. It follows, that $\Phi^R(0)$ is not a member of $H'(0)$.

5. Proof of Theorem 2. Before proceeding to the proof of Theorem 2 we shall need to prove a preliminary result.

LEMMA 5. *If $R^{**} \varepsilon C''$ is as in Corollary 1 to Theorem 1, then*

$$\limsup_{T\to\infty} \sup_{j, R\varepsilon C} f(\Phi_T^R(j)) \leqq \max_i \{f(\Phi^{R^{**}}(i))\}.$$

PROOF. We prove this under the assumption that every state is recurrent with respect to the class C'. The device of adjoining additional "artificial" decisions with arbitrarily large values of $-w_{jk}$'s associated with the adjoined decisions will apply as in the proof of Lemma 3. Now suppose the lemma is not true. Then there exist subsequences $\{T_v\}$, $\{R_v\}$, $\{i_v\}$, $v = 1, 2, \cdots$, such that

$$\lim_{v\to\infty} f(\Phi_{T_v}^{R_v}(i_v)) = \theta > \max_i \{f(\Phi^{R^{**}}(i))\}.$$

Construct, as in the proof of Lemma 2, the sequence of rules $\{R_v'\}$ as follows. Use rule R_v until $t = T_v$; then making use of the recurrency assumption use the appropriate rule in C' until $Y_t = i_v$ for the first time after T_v ; then, as if starting from $t = 0$, revert to the use of R_v ; $\cdots$, etc. From renewal type arguments, it follows that $\lim_{T\to\infty} \Phi_T^{R_v'}(i_v) = \Phi^{R_v'}(i_v)$ exists and that $\lim_{v\to\infty} f(\Phi^{R_v'}(i_v)) = \theta$. However, this means that

$$f(\Phi^{R_v'}(i_v)) > \max_i \{f(\Phi^{R^{**}}(i))\}$$

for some v, which is a contradiction of Corollary 1.

To prove Theorem 2 let R be an arbitrary rule in C. Suppose that the theorem does not hold. Then there exists a sphere S with positive radius such that $S \cap \bar{H}'' = \phi$, the null set, and $P(U^R(i, \omega) \cap S \neq \phi) > 0$. This is so since $\bar{H}_c''$, the complement of $\bar{H}''$, can be covered by a denumerable number of such spheres S_v , $v = 1, \cdots$, and

$$P(U^R(i, \omega) \cap \bar{H}_c'' \neq \phi) \leqq \sum_{v=1}^{\infty} P(U^R(i, \omega) \cap S_v \neq \phi).$$

$\bar{H}''$ is a closed and bounded convex set. S is convex. Since they are disjoint the two sets can be separated by a hyper-plane; i.e. there exists a set of numbers $w_{01}, \cdots, w_{LK}$ such that

$$f(r) < f(s) \tag{5.1}$$

for all $r = (r_{01}, \cdots, r_{LK}) \varepsilon \bar{H}''$ and $s = (s_{01}, \cdots, s_{LK}) \varepsilon S$. Define $g_t = w_{jk}$, if $Y_t = j$, $\Delta_t = d_k$, $j = 0, \cdots, L$; $k = 1, \cdots, K$; $t = 0, 1, \cdots$. Note that

$$(T + 1)^{-1} \sum_{t=0}^{T} g_t = f(\tilde{\Phi}_T^R(i)). \tag{5.2}$$

For a fixed N let

$$W_v = \sum_{t=(v-1)N+1}^{vN} g_t, \qquad v = 1, \cdots, [T/N]$$

and

$$W' = \sum_{t=[T/N]N+1}^{T} g_t, \qquad \text{if } [T/N] < T/N$$

$$= 0, \qquad \text{if } [T/N] = T/N.$$

Clearly, $|W_v|$, $v = 1, \cdots, [T/N]$, and $|W'|$ are bounded and

$$\begin{aligned} E(W_v \mid W_1, \cdots, W_{v-1}) & \\ &\leqq \max_i \sum_{j=0}^{L} \sum_{k=1}^{K} w_{jk} \sum_{t=(v-1)N+1}^{vN} P(Y_t = j, \Delta_t = d_k \mid Y_{(v-1)N} = i) \\ &\leqq \sup_{i, R\varepsilon C} \sum_{j=0}^{L} \sum_{k=1}^{K} w_{jk} \sum_{t=0}^{N} P(Y_t = j, \Delta_t = d_k \mid Y_0 = i) \\ &= \sup_{i, R\varepsilon C} N f(\Phi_N^R(i)). \end{aligned} \tag{5.3}$$

By Lemma 5, given any $\epsilon > 0$ there exists an N such that

$$\sup_{i,R\epsilon C} Nf(\Phi_N^R(i)) \leqq N\{f(\Phi^{R^{**}}(i_0)) + \epsilon\} \tag{5.4}$$

where i_0 is the state at which $\max_i f(\Phi^{R^{**}}(i))$ is obtained. However, by a strong law of large numbers for dependent random variables (Loève [12], p. 387) we have

$$\lim_{T\to\infty} [T/N]^{-1} \sum_{v=1}^{[T/N]} (W_v - E(W_v \mid W_1, \cdots, W_{v-1})) = 0$$

with probability 1. Thus, from (5.3) and (5.4) we have

$$\limsup\, [T/N]^{-1} \sum_{v=1}^{[T/N]} W_v \leqq N\{f(\Phi^{R^{**}}(i_0)) + \epsilon\} \tag{5.5}$$

with probability 1.

But then,

$$\begin{aligned} \limsup_{T\to\infty} T^{-1} \sum_{t=1}^{T} g_t &= \limsup_{T\to\infty} \left\{ \sum_{v=1}^{[T/N]} W_v + W' \right\} \\ &\leqq \limsup_{T\to\infty} T^{-1} \sum_{v=1}^{[T/N]} W_v + \limsup_{T\to\infty} W'/T \\ &= \limsup_{T\to\infty} T^{-1} \sum_{v=1}^{[T/N]} W_v \\ &\leqq N^{-1} \limsup_{T\to\infty} [T/N]^{-1} \sum_{v=1}^{[T/N]} W_v \\ &\leqq f(\Phi^{R^{**}}(i_0)) + \epsilon \end{aligned} \tag{5.6}$$

with probability 1. Since ϵ can be arbitrarily small, we have from (5.2) and (5.6)

$$\limsup_{T\to\infty} f(\tilde{\Phi}_T^R(i)) \leqq \{f(\Phi^{R^{**}}(i_0))\} \tag{5.7}$$

with probability 1. However, in order that, for any ω, a limit point $\tilde{x}$ of $\{\tilde{\Phi}_T^R(i)\}$ be a member of S we must have from (5.1)

$$f(\tilde{x}) > f(\Phi^{R^{**}}(i_0)),$$

since $\Phi^{R^{**}}(i_0)\ \epsilon\ \bar{H}''$. Thus, by (5.7)

$$P(S \cap U^R(i, \omega) \neq \phi) \leqq P\{\limsup_{T\to\infty} f(\tilde{\Phi}_T^R(i)) > f(\Phi^{R^{**}}(i_0))\} = 0,$$

a contradiction proving the theorem.

REFERENCES

[1] Bellman, Richard (1957). *Dynamic Programming*. Princeton Univ. Press.

[2] Bellman, Richard (1957). A Markovian decision process. *J. Math. Mech.* **6** 679–684.

[3] Blackwell, David (1962). Discrete dynamic programming. *Ann. Math. Statist.* **33** 719–726.

[4] CHUNG, KAI LAI (1960). *Markov Chains With Stationary Transition Probabilities.* Springer, Berlin.

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

[6] DERMAN, CYRUS (1963). Optimal replacement and maintenance under Markovian deterioration with probability bounds on failure. *Management Sci.* **9** 478–481.

[7] DERMAN, CYRUS (1963). Stable sequential rules and Markov chains. *J. Math. Analysis and Applications.* **6** 257–265.

[8] FREIMAR, M. (1961). On solving a Markovian decision problem by linear programming. Unpublished Technical Report, Institute for Defense Analysis, Cambridge.

[9] HOWARD, RONALD A. (1960). *Dynamic Programming and Markov Processes.* Technology Press of M.I.T. and Wiley, New York.

[10] KARLIN, SAMUEL (1959). *Mathematical Methods and Theory in Games, Programming and Economics.* **1** Addison-Wesley, Massachusetts.

[11] KLEIN, MORTON (1962). Inspection-maintenance-replacement under Markovian deterioration. *Management Sci.* **9** 25–32.

[12] LOÈVE, M. (1955). *Probability Theory,* (1st ed.). Van Nostrand, New York.

[13] MANNE, A. (1960). Linear programming and sequential decisions, *Management Sci.* **6** 259–267.