The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 424–434

ON RECURRENT DENUMERABLE DECISION PROCESSES

BY LLOYD FISHER

University of Washington

1. Summary. This paper considers decision processes on a denumerable state space. At each state a finite number of decisions is allowed. The main assumption is that if one always chooses the same decision at each state the resulting Markov chain is ergodic (i.e. positive recurrent). Under this assumption it is shown that all possible decision procedures are (in an appropriate sense) uniformly ergodic.

2. Introduction. We follow the notation of Derman [2]. We are concerned with control of certain types of dynamic systems. The system is observed periodically and classified into one of a denumberable number of states. I denotes the state space. After each observation one of a finite number of possible decisions is made. Let K_i be the finite number of possible decisions when the system is in the state i. Let Y_t and Δ_t, $t = 0, 1, 2 \cdots$, denote the sequences of states and decisions. We assume that

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

where the $q_{ij}(k)$'s are known.

A rule or policy R for controlling the system is a set of functions $D_k(Y_0, \Delta_0, \cdots, Y_t)$ satisfying $0 \leqq D_k(Y_0, \cdots, Y_t) \leqq 1$, for every k and $\sum_{k=1}^{K_i} D_k(Y_0, \Delta_0, \cdots, Y_t = i) = 1$, for every history $Y_0, \Delta_0, \cdots, Y_t$ $(t = 0, 1, 2, \cdots)$. $D_k(Y_0, \Delta_0, \cdots, Y_t)$ is the instruction at time t to make decision k with probability $D_k(Y_0, \Delta_0, \cdots, Y_t)$ if the partial history $Y_0, \Delta_0, \cdots, Y_t$ has occurred.

Let C denote the class of all decision rules. Let C' denote the class of stationary Markovian rules, i.e. $D_k(Y_0, \Delta_0, \cdots, Y_t = i) \equiv D_{ik}$ independent of t and the past history except for the present state. Let C'' denote the sub-class of C' for which $D_{ik} = 0$ or 1. C'' is the class of nonrandomized stationary Markovian rules.

If $R \varepsilon C'$, Y_t is a denumerable Markov chain with transition probabilities

$$P_{ij} = \sum_{k=1}^{K_i} D_{ik} q_{ij}(k) \quad \text{for} \quad i, j \varepsilon I.$$

Let E' denote the set of transition matrices resulting from $R \varepsilon C'$. E'' bears a similar to C''.

We shall be working under the assumption that each $P \varepsilon E''$ is the transition matrix of an ergodic (i.e. positive recurrent) chain. For $P \varepsilon E''$ let $\alpha(P)$ be the unique nonnegative stationary vector of mass one associated with P. Let $\alpha(E'') = \inf_{P \varepsilon E''} \alpha(P)$. (This is the pointwise inf, component by component.)

We shall show later that each $P \varepsilon E'$ is ergodic. Thus, we similarly define $\alpha(P)$ for $P \varepsilon E'$ and $\alpha(E')$. Each rule $R \varepsilon C$ will also be shown to be ergodic in the

Received 9 June 1967; revised 6 November 1967.

following sense: For any state $0 \varepsilon I$ and decision rule R, let

$$\alpha(R)_0 = \liminf_{T\to\infty} E_R \text{ (number of visits to 0 up to time } T \mid \text{Start at } 0)/T.$$

Thus, we define $\alpha(C) = \inf_{R\varepsilon C} \alpha(R)$. Then $\alpha(C) > 0$.

3. Uniform ergodicity of E''. This section is devoted to showing that $\alpha(E'') > 0$ We will say that a sequence of transition matrices $P(n)$ converges to P if for each i $\exists$ $n(i)$ such that $P_{ij}(n) = P_{ij}$ for $n \geqq n(i)$ and all j. We shall write $P(n) \to P$. In other words, $P(n) \to P$ if and only if each row is fixed from some n on, where n may depend on the row. Since each $P \varepsilon E''$ has only a finite number of possible ith rows it is easy to see that any sequence $P(n)$, $P(n) \varepsilon E''$, has a subsequence converging to a member of E''.

To show that $\alpha(E'') > 0$ we need to show that for each $i \varepsilon I$ there is no sequence $P(n)$ of elements of E'' such that $\alpha(P(n))_i \to 0$. We will arrive at this result by finding a contradiction. Let 0 be some fixed element of I.

LEMMA 1. *Let $P(n) \to P$, $\{P(n)\} \subseteq E''$ and $\alpha(P(n))_0 \to 0$, then $\alpha(P(n))_j \to 0$ for each $j \varepsilon I$.*

PROOF. Since $P \varepsilon E''$ is ergodic, 0 and j commute in the chain P. Pick a sequence of states $j_1, \cdots, j_{m-1}$ such that $P_{0j_1}P_{j_1j_2} \cdots P_{j_{m-1}j} > 0$. For $l \geqq N$ let $P_{0\cdot}(l), P_{j_1\cdot}(l), \cdots, P_{j_{m-1}\cdot}(l)$ be fixed (and hence having the same values as P). Then for $l \geqq N$ we have $P_{0j}^{(n+m)}(l) \geqq P_{00}^{(n)}(l)P_{0j_1}P_{j_1j_2} \cdots P_{j_{m-1}j} = P_{00}^{(n)}(l)P^*$ where $P^* > 0$. Thus,

$$\alpha(P(l))_j = \lim_{M\to\infty} [\textstyle\sum_{t=0}^{M} P_{0j}^{(t)}(l)/M] = \lim_{M\to\infty} [\textstyle\sum_{t=m}^{M} P_{0j}^{(t)}(l)/M]$$

$$\leqq P^* \lim_{M\to\infty} [\textstyle\sum_{t=0}^{M-m} P_{00}^{n}(l)/(M-m)]\cdot(M-m)/M = P^*\alpha(P(l))_0.$$

By Lemma 1 we know that if $\alpha(P(n))_0 \to 0$ then $\alpha(P(n))_j \to 0$ for each j. In the next few lemmas the following situation will occur. Suppose that $P(n) \to P$ and $\alpha(P(n)) \to 0$. It will be desirable to replace the kth row of $P(n)$ by one fixed decision at state k. Call the new sequence $\bar{P}(n)$. We want to consider whether or not $\alpha(\bar{P}(n)) \to 0$.

LEMMA 2. *Let $P(n) \to P$ and $\alpha(P(n)) \to 0$. Form a new sequence $\bar{P}(n)$ by replacing the* kth *row by a row corresponding to a fixed decision at state k. If $\alpha(\bar{P}(n_m)) \not\to 0$ for any subsequence n_m then for all $j \neq k$, $M_{jk}(n)$ is bounded in n, where $M_{jk}(n)$ is the mean time to go from j to k in the Markov chain $P(n)$. Also $M_{kj}(n) \to +\infty$.*

PROOF. Let $M_{jj}(n)$ be the expected time of first return to j after starting at j. Since $\alpha(P(n))_j = 1/M_{jj}(n)$, $\alpha(P(n)) \to 0$ is the same as $M_{jj}(n) \to \infty$ for each j. The mean return time to j is less than or equal to the mean return time to j via $k \neq j$ or

(i) $M_{jk}(n) + M_{kj}(n) \geqq M_{jj}(n)$ for $k \neq j$.

Form the new sequence $\bar{P}(n) \to \bar{P}$ and suppose we cannot find a subsequence n_k with $\alpha(P(n_k)) \to 0$. Then $\bar{M}_{jj}(n)$ is bounded for each j where $\bar{M}_{jj}(n)$ is the M_{jj} for the chain $\bar{P}(n)$. Since $\bar{P}(n) \to \bar{P}$ we may use the trick of Lemma 1 to show that for $n \geqq N(j, k), j \neq k$,

$$\bar{M}_{kk}(n) \geqq P(j, k) M_{jk}(n) \quad \text{where} \quad P(j, k) > 0.$$

Since $\bar{M}_{kk}(n)$ is bounded, $M_{jk}(n)$ is bounded and by (i), $M_{kj}(n) \to +\infty$. (Note that $M_{jk}(n)$ depends only on $P(n)$ and not $\bar{P}(n)$.)

LEMMA 3. *Let $P(n) \to P$, $\alpha(P(n)) \to 0$. With the exception of at most one row, say the kth row, we may replace the ith row (giving $\bar{P}(n) \to P$) and have $\alpha(\bar{P}(n_K)) \to 0$ for some subsequence n_K.*

PROOF. Suppose that we may not replace the kth row. Then by Lemma 2, $M_{ik}(n)$ is bounded for $i \neq k$. Since $M_{ik}(n) + M_{ki}(n) \geqq M_{ii}(n)$ this implies that $M_{ki}(n) \to +\infty$. Thus it is not true that $\forall_j \neq i$, $M_{ji}(n)$ is bounded (i.e. let $j = k$). Thus, by Lemma 2, we may change the ith row.

Before proceeding to Lemma 4 we introduce some notation and terminology. Let $P(n) \to P$ and $K \subseteq I$. We shall say that K is a fixed set if $k \varepsilon K$ implies that $P_{k\cdot}(n) = P_{k\cdot}$.

For any Markov chain with transition matrix P and $B \subseteq I$ we define $M_{00}(P, B) = \sum_{t=1}^{\infty} tP$ (beginning at 0 that the first return is at time t and the path up to the return is entirely in $B \cup \{0\}$.) Clearly, $M_{00}(P, B) \leqq M_{00}(P)$.

LEMMA 4. *Suppose that $P(n) \to P$, $\{P(n)\} \subseteq E''$ with a finite fixed subset K, and that $\alpha(P(n)) \to 0$. Let $N > 0$. We can find a new sequence $\{\bar{P}(n)\} \subseteq E''$, $\bar{P}(n) \to \bar{P}$ with a finite fixed subset $L \supseteq K$ such that $\bar{P}_{k\cdot}(n) = P_{k\cdot}(n)$ for $k \varepsilon K$ and $M_{00}(\bar{P}, L) \geqq N$.*

PROOF. Since $M_{00}(n) \to +\infty$ choose m such that $M_{00}(m) > 2N$. By monotone convergence choose a finite set $L \subseteq I$ and $K \subseteq L$ such that $M_{00}(P(m), L) > N$.

It would be nice if we could form a new sequence $\tilde{P}(n) \to \tilde{P}$ where $\tilde{P}(n) = P(n)$ except that all rows of $\tilde{P}(n)$ corresponding to $j \varepsilon L$ are the same as the jth row of $P(m)$. If $\alpha(\tilde{P}(n)) \to 0$ we are done.

We try to effect $\tilde{P}(n)$ by taking our current sequence $P(n)$ and replacing one row corresponding to a $j \varepsilon L$ by the corresponding row of $P(m)$ and choosing a subsequence $\bar{P}(n_k) \to \bar{P}$ such that $\alpha(\bar{P}(n_k)) \to 0$. If L has more than two elements by Lemma 3 we may replace at least one element of L. We keep trying in this manner to replace more and more rows corresponding to $j \varepsilon L$ (and each time taking a subsequence of the current subsequence). As we replace a row the row which we are not allowed to replace may shift (or not exist at all), but eventually we have a new sequence $P'(n) \to P'$, $\alpha(P'(n)) \to 0$, where $P'(n)$ has the same jth rows as $P(m)$ for all $j \varepsilon L$ with possibly one exception. If all the rows for $j \varepsilon L$ are the same we are done.

Thus, suppose that we were not able to replace one row. Let this row be j_0. Since there are only a finite number of possibilities for the j_0th row we may choose a new subsequence where the j_0th row is constant. We may assume that $P'(n) \to P'$, $\alpha(P'(n)) \to 0$. L is a fixed set for $P'(n)$ and $P'_{j\cdot}(n) = P_{j\cdot}(m)$ for $j \varepsilon L, j \neq j_0$.

We now begin the whole procedure over again using L as the fixed set. We choose a k and a set $M \supseteq L$ such that $M_{00}(P'(k), M) > N$. We would like to replace all rows in $P'(n)$ corresponding to elements of M by the corresponding row of $P'(k)$ and still have $\alpha \to 0$. As above we replace one row at a time (taking a subsequence of the last subsequence). The exceptional row which can not be replaced may move around, but eventually we have replaced all rows except one. Suppose that the one which cannot be replaced is the j_0th row. Then we are done since $P'(k)$ has the given j_0th row (since that was the same for all $P'(n)$). If the exceptional row is not the j_0th row then replace the j_0th row by the j_0th row of $P(m)$, (the thing we originally desired to do). Then we are done by the fact that $M_{00}(P(m), L) > N$.

Finally, it is noted that in either eventuality K remains a fixed set.

The following theorem eliminates the need for assumption (F) in Derman [1] since Derman's condition (C) implies (F) which is used only in conjunction with (C).

THEOREM 1. $$\alpha(E'') > 0.$$

PROOF. Suppose not. Pick a subsequence $P(n) \subseteq E''$, $P(n) \to P$, $\alpha(P(n))_0 \to 0$. Using Lemma 4, choose a new subsequence $P^1(n) \to P^1$, $\{P^1(n)\} \subseteq E''$ with $L(1)$ fixed and $M_{00}(P^1, L(1)) \geqq 1$. Proceeding inductively choose $P^k(n) \to P^k$, $\{P^k(n)\} \subseteq E''$, $L(k) \supseteq L(k-1)$ fixed and $M_{00}(P^k, L(k)) \geqq k$.

Let $P \varepsilon E''$ and $P_{l\cdot} = P^k_{l\cdot}$ if $l \varepsilon L(k)$ for any k. Then

$$M_{00}(P) \geqq M_{00}(P^k, L(k)) \geqq k.$$

This implies $M_{00}(P) = \infty$ which contradicts the assumption that all $P \varepsilon E''$ are ergodic, i.e. $\alpha(P)_0 = 1/M_{00}(P) = 0$. Contradiction. Therefore one cannot find $\{P(n)\} \subseteq E''$, $\alpha(P(n))_0 \to 0$. Thus $\alpha(E'') > 0$.

COROLLARY 1. $\sup_{P \varepsilon E''} \alpha(P) < 1$ *and* $M_{jj}(P)$ *is uniformly bounded for* $P \varepsilon E''$, (*provided that* I *has at least two elements*).

4. Uniform ergodicity of all rules $R \varepsilon C$. We extend Theorem 1 to all rules $R \varepsilon C$ by first extending it to C' and then to C.

THEOREM 2. $$\alpha(E') = \alpha(E'').$$

PROOF. Since $E' \supseteq E''$ it is clear that $\alpha(E') \leqq \alpha(E'')$.

(A) We first show that if we take $P' \varepsilon E''$ and replace the policy at one state, say 0, by a randomized Markovian strategy the resulting Markov chain P has $\alpha(P) \geqq \alpha(E'')$. Thus, let $P' \varepsilon E''$ and $P = P'$ except for the 0th row where $P_{0\cdot} = \sum_{i=1}^{K_0} \lambda_i P_{0\cdot}(i)$ where $\lambda_i \geqq 0$, $\sum_{i=1}^{K_0} \lambda_i = 1$ and $P_{0\cdot}(i)$ corresponds to a stationary Markovian decision at state 0. Let $M_{jj} = \sup_{P \varepsilon E''} M_{jj}(P)$. To show that $\alpha(P) \geqq \alpha(E'')$ is equivalent to $M_{jj}(P) \leqq M_{jj}$ for each $j \varepsilon I$.

(i) $M_{00}(P) \leqq M_{00}$. It is easy to see that $M_{00}(P) = \sum_{i=1}^{K_0} \lambda_i M_{00}(P(i))$ where $P(i) = P'$ except for the 0th row where $P_{0\cdot}(i)$ is used. Thus, since $P(i) \varepsilon E''$ and $M_{00}(P(i)) \leqq M_{00}$, $M_{00}(P) \leqq M_{00}$.

(ii) $j \neq 0$, $M_{jj}(P) \leqq M_{jj}$.

We digress slightly before continuing the proof. Let $\tilde{P}$ be an ergodic Markov chain. Let t be the stopping time of the first visit to 0 or j after time 0. Let

$M(\tilde{P}) = E(t \mid \text{begin in state } j)$ and $M'(\tilde{P}) = E(t \mid \text{begin in state } 0)$. Let ${}^kH_{ij}(\tilde{P})$ be the probability starting in i of hitting state j before hitting state k (after time 0). Then

$$M_{jj}(\tilde{P}) = M(\tilde{P}) + {}^jH_{j0}(\tilde{P})M_{0j}(\tilde{P}) \qquad \text{and}$$
$$M_{0j}(\tilde{P}) = M'(\tilde{P}) + {}^jH_{00}M_{0j}(\tilde{P}) \qquad \text{or}$$
$$M_{0j}(\tilde{P}) = M'(\tilde{P})/(1 - {}^jH_{00}(\tilde{P})).$$

Thus, if $\tilde{P} \varepsilon E''$,

$$(1) \qquad M_{jj}(\tilde{P}) = M(\tilde{P}) + {}^jH_{j0}(\tilde{P})M'(\tilde{P})/(1 - {}^jH_{00}(\tilde{P})) \leqq M_{jj}.$$

Returning to our proof let $P(i)$ be the matrix P' with the 0th row given by $P_{0\cdot}(i)$. It is probabalistically clear that $M(P) = M(P') = M(P(i))$ and ${}^jH_{j0}(P) = {}^jH_{j0}(P') = {}^jH_{j0}(P(i))$ for $i = 1, 2, \cdots, K_0$ since the above quantities do not depend upon the transition probabilities at zero. Thus,

$$(2) \qquad M_{jj}(P) = M(P') + {}^jH_{j0}(P')M_{0j}(P)$$

where

$$M_{0j}(P) = \sum_{i=1}^{K_0} \lambda_i(M'(P(i)) + {}^jH_{00}(P(i))M_{0j}(P))$$

so that

$$M_{0j}(P) = \sum_{i=1}^{K_0} \lambda_i M'(P(i))/(1 - \sum_{i=1}^{K_0} \lambda_i\, {}^jH_{00}(P(i))).$$

Using $M(P(i)) = M(P')$, ${}^jH_{j0}(P(i)) = {}^jH_{j0}(P')$ and (1)

$${}^jH_{j0}(P')M'(P(i)) \leqq (M_{jj} - M(P'))(1 - {}^jH_{00}(P(i))),$$

$i = 1, 2, \cdots, K_0$, whence

$${}^jH_{j0}(P') \sum_{i=1}^{K_0} \lambda_i M'(P(i)) \leqq (M_{jj} - M(P'))(1 - \sum_{i=1}^{K_0} \lambda_i\, {}^jH_{00}(P(i)))$$

or ${}^jH_{j0}(P')M_{0j}(P) \leqq M_{jj} - M(P')$ which, combined with (2), gives $M_{jj}(P) \leqq M_{jj}$ the desired result.

(B) We now note that we may take any $P \varepsilon E''$ and randomize a finite number of rows (i.e. change them to correspond to a randomized policy at that state) and still have the new P' satisfying $\alpha(P') \geqq \alpha(E'')$. We have just proved this for one state. Suppose that this is true for any subset of n states. We may form a new E_N'' which consists of all of the old $P \varepsilon E''$ and also allowing the n randomized rows (or policies) desired. Then E_N'' satisfies all the conditions of E'' and $\alpha(E_N'') = \alpha(E'')$ by induction. By using part (A) we may now randomize an $n + 1$st row. By induction we may do this for any finite number of rows.

(C) Let $P \varepsilon E'$. Suppose that $\alpha(P)_0 < \alpha(E'')_0$, then we may choose a finite set $S \subseteq I$ such that $M_{00}(P, S) > M_{00}$. Form P' by $P'_{i\cdot} = P_{i\cdot}$ for $i \varepsilon S$ and $P'_{j\cdot}$ corresponds to a fixed policy at j if $j \not\varepsilon S$. Then $M_{00}(P') \geqq M_{00}(P, S) > M_{00}$ so that $\alpha(P') < \alpha(E'')_0$ contradicting (B). Thus, $\alpha(P) \geqq \alpha(E'')$ for each $P \varepsilon E'$, hence, $\alpha(E') \geqq \alpha(E'')$ finishing the proof.

For any $R \varepsilon C$ let $M_{00}(R)$ be the expected value of the first return to zero under policy R if the system begins in state 0. As before $M_{00} = \sup_{P \varepsilon E''} M_{00}(P)$.

A rule $R \varepsilon C$ is called memoryless if $D_k(Y_0, \Delta_0, \cdots, Y_t = i) = D_{ik}^{(t)}$. A memoryless rule depends only on t and the current position i but not on the whole past history. We modify an argument used by Derman and Strauch [2] to prove the following:

LEMMA 5. *For each $R \varepsilon C$ there is a memoryless rule $R_0 \varepsilon C$ such that*

$$M_{00}(R) = M_{00}(R_0).$$

PROOF. Given an $R \varepsilon C$ we define R_0 by

$$D_{ik}^{(t)} = P_R(\Delta_t = k \mid Y_0 = 0,\ Y_1 \neq 0,\ Y_2 \neq 0,\ Y_{t-1} \neq 0,\ Y_t = i)$$

if the conditioning event has positive probability. Otherwise we define $D_{ik}^{(t)} = \delta_{k1}$. $D_{0k}^{(0)} = D_k(Y_0 = 0)$ for R. We proceed to show by induction that

$$(3)\quad P_{R_0}(Y_0 = 0, Y_1 \neq 0, \cdots, Y_{t-1} \neq 0, Y_t = i) = P_R \quad \text{(Same Quantity)}$$

for each $i \varepsilon I$ and time t. For $t = 0$ and $t = 1$ we clearly have equality. Assume this is true for $t - 1$, then

$$\tilde{P} \equiv P_R(Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-1} \neq 0,\ Y_t = i)$$
$$= \textstyle\sum_{l\varepsilon I, l\neq 0} \sum_{k=1}^{K_l} \sum_{*} P_R(Y_t = i \mid \Delta_{t-1} = k,\ Y_{t-1} = l)$$
$$\cdot P_R(\Delta_{t-1} = k,\ Y_{t-1} = l,\ Y_0 = 0, \Delta_0 = k_0,\ Y_1 = i_1, \Delta_1 = k_1,\ \cdots,\ Y_{t-1} = l)$$

where $\sum_*$ is the sum over all $(k_0, i_1, k_1, i_2, \cdots, k_{t-2})$ allowed with $i_j \neq 0$, $j = 1, \cdots, t-2$. If we move the $\sum_*$ past $P_R(Y_t = i \mid \Delta_{t-1} = k,\ Y_{t-1} = l)$ since $\sum_*$ does not affect i, or k we have

$$(4)\quad P = \textstyle\sum_{l\varepsilon I, l\neq 0} \sum_{k=1}^{K_l} P_R(Y_t = i \mid \Delta_{t-1} = k,\ Y_{t-1} = l)$$
$$\cdot P_R(\Delta_{t-1} = k,\ Y_{t-1} = l,\ Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-2} \neq 0).$$

Now, $P_R(Y_t = i \mid \Delta_{t-1} = k,\ Y_{t-1} = l) = q_{li}(k) = P_{R_0}(Y_t = i \mid \Delta_{t-1} = k,\ Y_{t-1} = l)$. The second P_R term in (4) is

$$D_k^{(t-1)} \cdot P_R(Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-2} \neq 0,\ Y_{t-1} = l)$$
$$= D_k^{(t-1)} P_{R_0}(Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-2} \neq 0,\ Y_{t-1} = l)$$
$$= P_{R_0}(\Delta_{t-1} = k,\ Y_{t-1} = l,\ Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-2} \neq 0).$$

Thus,

$$\tilde{P} = \textstyle\sum_{l\varepsilon I, \neq 0} \sum_{k=1}^{K_l} P_{R_0}(Y_t = i \mid \Delta_{t-1} = k,\ Y_{t-1} = l)$$
$$\cdot P_{R_0}(\Delta_{t-1} = k,\ Y_{t-1} = l,\ Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-2} \neq 0)$$
$$= P_{R_0}(Y_0 = 0,\ Y_1 \neq 0,\ \cdots,\ Y_{t-2} \neq 0,\ Y_{t-1} = i).$$

Thus, we have

$$P_R \quad \text{(Return to } 0 \text{ for the first time at time } n)$$
$$= P_{R_0} \quad \text{(Same Quantity) which implies } M_{00}(R) = M_{00}(R_0).$$

LEMMA 6. *For each* $R \varepsilon C$, $M_{00}(R) \leqq M_{00}$.

PROOF. Suppose that $M_{00}(R) > M_{00}$. By Lemma 4 we may assume that R is a memoryless rule.

Let t_n be the stopping time given by the minimum of n and the first return to zero. By monotone convergence we may find a finite m such that $E_R(t_m) > M_{00}$. Form the new rule $\mathbf{R} \varepsilon C$ such that $\mathbf{R}$ follows the rule R until time m. After time m all transitions take place according to a fixed transition matrix $P \varepsilon E''$. $\mathbf{R}$ is still a memoryless rule. Since $P \varepsilon E''$ under the rule $\mathbf{R}$ the process returns to 0 with probability one. Consider the new rule R^* which behaves according to rule $\mathbf{R}$ until a return to origin, R^* then acts as if $\mathbf{R}$ were beginning again at zero.

Consider a state space S consisting of zero and pairs (i, t), $i \varepsilon I$, $i \neq 0$, $t = 1, 2, \cdots$. Consider the transition matrix P on S,

$$P_{0,(i,1)} = P_{R^*}(Y_1 = i \mid Y_0 = 0) \quad \text{for} \quad i \neq 0,$$

$$P_{0,(i,t)} = 0 \quad \text{for} \quad t > 1,$$

$$P_{0,0} = P_{R^*}(Y_1 = 0 \mid Y_0 = 0).$$

For $i \neq 0$, $P_{(i,t)(j,t+1)} = P_{R^*}(Y_{t+1} = j \mid Y_0 = 0, Y_1 \neq 0, Y_2 \neq 0, \cdots, Y_t = i)$,

$$P_{(i,t),0} = P_{R^*}(Y_{t+1} = 0 \mid Y_0 = 0, Y_1 \neq 0, \cdots, Y_{t-1} \neq 0, Y_t = i),$$

$$P_{(i,t),(j,t')} = 0, \qquad t' \neq t + 1.$$

Then one has a recurrent Markov chain on S and $M_{00}(R^*) = M_{00}(P)$. Let α be the unique positive stationary vector for P where possibly $\sum_{i \varepsilon S} \alpha_i = +\infty$. Since $\mathbf{R}$ was memoryless

$$P_{(i,t)(j,t+1)} = \sum_{k=1}^{K_i} \lambda_{i,t}(k) q_{ij}(k) \quad \text{for appropriate} \quad \lambda_{i,t}(k).$$

(i) for $i \neq 0$, $t > 1$,

$$\alpha(i, t) = \sum_{j \varepsilon I, j \neq 0} \alpha(j, t-1) \sum_{k=1}^{K_j} \lambda_{j,t-1}(k) q_{ji}(k).$$

(ii) for $i \neq 0$, $t = 1$,

$$\alpha(i, 1) = \alpha(0) \sum_{k=1}^{K_0} \lambda_0(k) q_{0i}(k).$$

(iii) for 0,

$$\alpha(0) = \sum_{i \varepsilon I, i \neq 0} \sum_{t=1}^{\infty} \alpha(i, t) \sum_{k=1}^{K_i} \lambda_{i,t}(k) q_{i0}(k) + \alpha(0) \sum_{k=1}^{K_0} \lambda_0(k) q_{00}(k).$$

For $i \neq 0$ we define

$$\alpha_i = \sum_{t=1}^{\infty} \alpha(i, t), \ \bar{\lambda}_i(k) = \sum_{t=1}^{\infty} \alpha(i, t) \lambda_{i,t}(k)/\alpha_i.$$

Then we see that from (i) and (ii),

$$\alpha_i = \alpha(0) \sum_{k=1}^{K_0} \lambda_0(k) q_{0i}(k) + \sum_{t=1}^{\infty} \sum_{j \varepsilon I, j \neq 0} \alpha(j, t) \sum_{k=1}^{K_j} \lambda_{j,t}(k) q_{ji}(k),$$

$$\alpha_i = \alpha(0) \sum_{k=1}^{K_0} \lambda_0(k) q_{0i}(k)$$
$$+ \sum_{j \varepsilon I, j \neq 0} \sum_{k=1}^{K_j} \alpha_j \Big(\sum_{t=1}^{\infty} \alpha(j, t) \lambda_{j,t}(k)/\alpha_j \Big) q_{ji}(k),$$

(5) $\alpha_i = \sum_{j\varepsilon I, j\neq 0} \alpha_j \sum_{k=1}^{K_j} \bar{\lambda}_j(k) q_{ji}(k) + \alpha(0) \sum_{k=1}^{K_0} \lambda_0(k) q_{0i}(k).$

Similarly,

(6) $\alpha(0) = \alpha(0) \sum_{k=1}^{K_0} \lambda_0(k) q_{00}(k) + \sum_{j\varepsilon I, j\neq 0} \alpha_j \sum_{k=1}^{K_j} \bar{\lambda}_j(k) q_{j0}(k).$

If we consider the matrix $P \varepsilon E'$ with $D_{ik} = \bar{\lambda}_i(k)$ for $i \neq 0$ and $D_{0k} = \lambda_0(k)$ we see that from (5) and (6), $\alpha P = \alpha$. Thus, normalizing α so that $\sum_{i\varepsilon I} \alpha_i = 1$ we have $\alpha(0) = 1/M_{00}(P) \geqq \alpha(E'')$ by Theorem 2. Since α has finite mass P is an ergodic Markov chain. Since $1/M_{00}(P) = \alpha(0)$ we have $M_{00}(R^*) = M_{00}(P) \leqq M_{00}$, which contradicts $M_{00} < E_{\mathbf{R}}(t_m) \leqq M_{00}(\mathbf{R}) = M_{00}(R^*)$.

THEOREM 3. *If $R \varepsilon C$, $\alpha(R) \geqq \alpha(E'')$.*

PROOF. In the previously mentioned paper of Derman and Strauch [2] during the proof of Theorem 2 it is shown that for each $R \varepsilon C$ there is a memoryless $R_0 \varepsilon C$ such that $P_R(Y_t = i \mid Y_0 = 0) = P_{R_0}(Y_t = i \mid Y_0 = 0)$ for each i and t. Although their proof was written under the assumption of a finite I the proof as written is valid for denumerable I. Thus,

$$\alpha(R)_0 = \liminf_{T\to\infty} E_R\left(\sum_{i=1}^{T} \delta_{0Y_i}\right)/T = \liminf_{T\to\infty} E_{R_0}\left(\sum_{i=1}^{T} \delta_{0Y_i}\right)/T = \alpha(R_0)_0$$

so that without loss of generality we may assume that we are working with a memoryless rule R.

Let X_n be the (random) time between the $n-1$st and nth return to zero. Since our rule is memoryless the distribution of X_n is completely specified if we know $S_{n-1} = \sum_{i=1}^{n-1} X_i$. By Lemma 5 we have $E(X_n \mid S_{n-1}) \leqq M_{00}$. Let $M_0 = 0$, $M_n = S_n - nM_{00}$ for $n \geqq 1$. Then

$$E(M_{n+1} \mid M_n) = M_n + E(X_{n+1} \mid S_n) - M_{00} \leqq M_n.$$

Thus M_n is a supermartingale.

Let N_n be the number of return visits to zero up to and including time n. Then $N_n + 1$ is a bounded stopping time and $E(|S_{N_n+1}|) \leqq E(S_{n+1}) < \infty$. By a standard martingale theorem (e.g. Kemeny, Snell and Knapp, [4] see remarks following Theorem 3–15) we have

$$E(M_{N_n+1}) \leqq E(M_0) = 0 \quad \text{or} \quad E(N_n + 1)M_{00} \geqq E(S_{N_n+1}).$$

Noting that $S_{N_n+1} \geqq n$ we have

$$E(N_n + 1)/n \geqq E(S_{N_n+1})/M_{00}n = \alpha(E'')_0 E(S_{N_n+1})/n \geqq \alpha(E'')_0.$$

Thus,

$$\alpha(R)_0 = \liminf_{n\to\infty} E(N_n)/n = \liminf_{n\to\infty} E(N_n + 1)/n \geqq \alpha(E'')_0.$$

The proof of the theorem is complete.

COMMENTS. To prove part (A) in the proof of Theorem 2 one cannot use the argument that if P and P^1 are matrices of ergodic chains and $0 \leqq \lambda \leqq 1$, then $P' = \lambda P + (1 - \lambda)P^1$ is ergodic, i.e. convex combinations preserve ergodicity. To see that this is false let us consider the state space of the nonnegative integers.

Let

$$P_{0,3} = 1,\ P_{2i,2i-2} = \tfrac{2}{3},\ P_{2i,2i+2} = \tfrac{1}{3},\ P_{2i+1,2i} = \tfrac{1}{10},\ P_{2i+1,2i+3} = \tfrac{9}{10}$$

$$\text{for } i = 1, 2, \cdots P_{10} = 1.$$

$$P^1_{0,2} = 1,\ P^1_{1,0} = 1,\ P^1_{2i,2i+2} = \tfrac{8}{9},\ P^1_{2i,2i-1} = \tfrac{1}{9},\ P^1_{2i+1,2i-1} = \tfrac{2}{3},\ P^1_{2i+1,2i+3} = \tfrac{1}{3}.$$

It is easy to check that P and P^1 are ergodic, but $P' = \frac{1}{2}P + \frac{1}{2}P^1$ is the transition matrix of a transient chain. In fact if one takes a convex combination of two transition matrices each of which may be either transient, null recurrent or ergodic the result may be any of the three types.

One may construct examples of denumerable decision processes where each $R \varepsilon C''$ leads to an ergodic chain by using the criteria given in a paper by Lamperti [5].

These results may also be used to provide an example of a decision process which has an optimal rule, but no optimal rule in C' (Fisher and Ross [3]).

REFERENCES

[1] DERMAN, C. (1966). Denumerable state Markovian decision processes—average cost criterion. *Ann. Math. Statist.* **37** 1545–1553.

[2] DERMAN, C. and STRAUCH, R. E. (1966). A note on memoryless rules for controlling sequential control processes. *Ann. Math. Statist.* **37** 276–278.

[3] FISHER, L. and ROSS, S. M. (1968). An example in denumerable decision processes. *Ann. Math. Statist.* **39** 674–675.

[4] KEMENY, J., SNELL, J. L. and KNAPP, A. (1965). *Denumerable Markov Chains.* Van Nostrand, Princeton.

[5] LAMPERTI, J. (1963). Criteria for stochastic processes: passage-time moments. *J. Math. Anal Appl.* **7** 127–145.