The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 412–423

NON-DISCOUNTED DENUMERABLE MARKOVIAN DECISION MODELS[1]

BY SHELDON M. ROSS

Stanford University

0. Introduction. We are concerned with a process which is observed at times $t = 0, 1, 2, \cdots$ to be in one of a possible number of states. We let I (assumed denumerable) denote the number of possible states. If at time t the system is observed in state i then one of K_i possible actions must be taken. Unless otherwise noted we shall assume throughout that $K_i < \infty$ for all i.

If the system is in state i at time t and action K is chosen then two things occur:

(i) We incur an expected cost $C(i, K)$ and

(ii) $P\{X_{t+1} = j \mid X_0, \Delta_0, \cdots, X_t = i, \Delta_t = K\} = P(i, j:K)$ where $\{X_r\}_{r=0}^{t+1}$ denotes the sequence of states and $\{\Delta_r\}_{r=0}^{t+1}$ the sequence of decisions up to time $t + 1$.

Thus both the costs and the transition probabilities are functions only of the last state and the subsequently made decision. It is assumed that both the expected costs $C(i, K)$ and the transition probabilities $P(i, j:K)$ are known. Furthermore it is assumed that the expected costs are bounded and we let M be such that $|C(i, K)| < M$ for all i, K.

A rule or policy R for controlling the system is a set of functions $\{D_K(X_0, \Delta_0, \cdots, X_t)\}_{K=1}^{K_{X_t}}$ satisfying

$$0 \leqq D_K(X_0, \Delta_0, \cdots, X_t) \leqq 1, K = 0, 1 \cdots, K_{X_t}$$

$$\text{and} \quad \sum_{K=1}^{K_{X_t}} D_K(X_0, \Delta_0, \cdots, X_t) = 1$$

for every history $X_0, \Delta_0, \cdots, X_t, t = 0, 1, \cdots$.

The interpretation being: if at time t we have observed the history $X_0, \Delta_0, \cdots, X_t$ then action K is chosen with probability $D_K(X_0, \cdots, X_t)$.

We say that a rule R is a stationary if $D_K(X_0, \Delta_0, \cdots, X_t = i) = D_{i,K}$ independent of $X_0, \Delta_0, \cdots, \Delta_{t-1}$ and t. We say that a rule R is stationary deterministic if it is stationary and also $D_{i,K} = 0$, or 1. Thus the stationary deterministic rules are those non-randomized rules whose actions at t just depend on the state at time t. We denote by C'' the class of stationary deterministic rules.

Following Derman [4] the process $\{(X_t, \Delta_t) t = 0, 1, 2, \cdots\}$ will be called a *Markovian decision process.*

Two possible measures of effectiveness of a rule governing a Markovian decision process are the expected total discounted cost and secondly the expected average cost per unit time. The first assumes a discount factor $\beta \varepsilon (0, 1)$ and for

Received 29 May 1967; revised 17 October 1967.

[1] This work was supported in part by the Army, Navy, Air Force and NASA under contract Nonr 225(53) (NR-042-002) with the Office of Naval Research.

a starting state $X_0 = i$ the objective is to minimize

$$\psi(i, \beta, R) = E_R \sum_{t=0}^{\infty} C(X_t, \Delta_t)\beta^t.$$

The second criteria tries to minimize for a given $X_0 = i$

$$\varphi(i, R) = \limsup_{n\to\infty} E_R \sum_{t=0}^{n} C(X_t, \Delta_t)(n+1)^{-1}.$$

Since costs are bounded and adding a constant to all the costs $C(i, K)$ will affect all rules identically in both criteria we may without loss of generality assume that costs are non-negative.

We shall be concerned in this paper with the average cost criterion. The first results for the average cost criterion which did not assume a finite state space were given by Taylor [9]. Taylor worked with a replacement model (see Section 4) and gave sufficient conditions for the existence of a stationary deterministic optimal rule. His method was to treat the average cost problem via the known results of the discounted cost problem.

Derman [4] has recently dealt with the countable state, finite action general Markovian model and has given a sufficient condition for the existence of a stationary deterministic optimal rule. Unfortunately, this condition—the existence of a bounded solution of the functional equation $g + f(i) = \min_K \{C(i, K) + \sum_j P(i, j:K)f(j)\}$—cannot be checked directly. Derman's paper [4], however, in conjunction with a later joint paper [5] of Derman and Veinott shows that a sufficient condition for the above is that: (i) for each rule $R \varepsilon C''$ the resulting Markov chain is positive recurrent, and (ii) there exists some state (say 0) and a constant $T < \infty$ such that $M_{i0}(R) < T$ for all i, and all $R \varepsilon C''$ where $M_{i0}(R)$ denotes the mean recurrence time from state i to state 0 when using rule R.

In the first section of this paper, by following the approach of Taylor, weaker sufficient conditions than those given by Derman are determined. We also show the connection between the average cost optimal rule and the optimal discounted cost rules—speaking loosely, the former is a limit point of the latter rules.

The second section shows how, in a special case, the average cost case can be reduced to the discounted cost case.

The third section deals with ϵ-optimal rules and a sufficient condition is given for the optimal discounted rules to be ϵ-optimal.

The fourth section deals with the replacement problem; and it is shown that an optimal rule always exists, but it may not be of the stationary deterministic type.

1. On the existence of a stationary deterministic optimal rule. We shall need the following result given by Blackwell [1]: If $K_i < \infty$ and $C(i, K) < M$ for all i, K, then under the β-discounted criteria with $0 < \beta < 1$ there exists a stationary deterministic rule R_β such that $\psi(i, \beta, R_\beta) = \min_R \psi(i, \beta, R)$ for all $i \varepsilon I$. Furthermore, $\{\psi(i, \beta, R_\beta), i \varepsilon I\}$ is the unique solution to

$$(1) \quad \psi(i, \beta, R_\beta) = \min_K \{C(i, K) + \beta \sum_j P(i, j:K)\psi(j, \beta, R_\beta)\}, \qquad i \varepsilon I,$$

and any stationary deterministic rule which when in state i selects an action which minimizes the right side of (1) is optimal.

Following Taylor, for any $\beta \varepsilon (0, 1) i, j \varepsilon I$, let

$$f_\beta(i, j) = \psi(i, \beta, R_\beta) - \psi(j, \beta, R_\beta). \tag{2}$$

One has by simple manipulations that

$$g_\beta(j) + f_\beta(i, j) = \min_K \{C(i, K) + \beta \sum_s P(i, s{:}K) f_\beta(s, j)\} \tag{3}$$

where $g_\beta(j) = (1 - \beta)\psi(j, \beta, R_\beta)$. Note that $|g_\beta(j)| < M$ for all β, j. We need the following assumption:

ASSUMPTION (*). For some sequence $\beta_r \to 1^-$ there exists a constant $N < \infty$ such that

$$|f_{\beta_r}(i, j)| < N \qquad \text{for all} \quad r = 1, 2, \cdots, \qquad \text{all} \quad i, j \varepsilon I.$$

THEOREM 1.1. *If Assumption (*) holds, then there exists a bounded solution to the functional equation*

$$g + f(i) = \min_K \{C(i, K) + \sum_j P(i, j{:}K) f(j)\}, \qquad i \varepsilon I. \tag{4}$$

PROOF. Fix some state s. By Assumption (*) $f_{\beta_r}(i, s)$ is uniformly bounded for $r = 1, 2, \cdots$, and all $i \varepsilon I$. Since I is denumerable we can get a subsequence $\{\beta_r'\}_{r=1}^\infty$ such that $f_{\beta_r'}(i, s) \to f(i)$ for all i. Since $g_\beta(s)$ is bounded for all β we can also require that $g_{\beta_r'}(s) \to g$ as $r \to \infty$. Therefore, by (3) and the bounded convergence theorem we have that

$$g + f(i) = \min_K \{C(i, K) + \sum_j P(i, j{:}K) f(j)\}. \qquad \text{QED}$$

REMARK. If $\psi(i, \beta, R_\beta)$ is an increasing function of i for each β, then $f(i)$ is an increasing function of i.

THEOREM 1.2. *If there exists a bounded set of numbers $\{g, f(i)\}$, $i \varepsilon I$, such that*

$$g + f(i) = \min_K \{C(i, K) + \sum_j P(i, j{:}K) f(j)\}, \qquad i \varepsilon I, \tag{5}$$

then there exists a stationary deterministic rule R^* such that

$$g = \varphi(i, R^*) = \min_R \varphi(i, R) \qquad \text{for all} \quad i$$

and R^* is any rule which, for each i, prescribes an action which minimizes the right side of (5).

PROOF. See Derman [4] or Derman and Lieberman [3].

REMARK 1. It also follows from [3] that $g = \lim_n \sum_{t=0}^n E_{R^*}[C(X_t, \Delta_t) | X_0 = i]/n$ for all i; and that $g \leqq \liminf_n \sum_{t=0}^n E_R[C(X_t, \Delta_t) | X_0 = i]/n$ for all rules R, all i.

REMARK 2. For any subsequence of $\{\beta_r\}_{r=1}^\infty$ there is a sub-subsequence $\{\beta_r''\}_{r=1}^\infty$ such that $\lim_{r\to\infty} g_{\beta_r''}(s)$ exists. By Theorem 1.2 this limit must be g. Therefore, $g = \lim_r g_{\beta_r}(s) = \lim_r (1 - \beta_r)\psi(s, \beta_r, R_{\beta_r})$, for all states s. For $R \varepsilon C''$ let $i(R)$ be the action R chooses when in state i.

DEFINITION. For rules R_n, $R \varepsilon C''$, we say that R_n converges to R ($R_n \to R$, or $\lim_n R_n = R$) if for each i there exists N_i such that $i(R_n) = i(R)$ for all $n \geqq N_i$ Note that any countable sequence of rules $R_n \varepsilon C''$ has a convergent subsequence.

THEOREM 1.3. *If Assumption* (*) *holds, then*:

(i) *for some subsequence* $\{\beta_r'\}_{r=1}^{\infty}$ *of* $\{\beta_r\}_{r=1}^{\infty}$ *and some* R^*, $R^* = \lim_r R_{\beta_r'}$,

(ii) *if* $R = \lim_r R_{\beta_r'}$, *where* $\{\beta_r'\}_{r=1}^{\infty}$ *is a subsequence of* $\{\beta_r\}_{r=1}^{\infty}$,

then R *is optimal*, i.e., $\varphi(i, R) = g$ *for all* $i \varepsilon I$.

PROOF OF (i). Let $\{\beta_r'\}_{r=1}^{\infty}$ be the subsequence for which $f_{\beta_r'}(i, s) \to f(i)$ and $g_{\beta_r'}(s) \to g$ as $r \to \infty$. Now it is easily seen from the definition of $f_\beta(j, s)$ that, when in state i, $R_{\beta_r'}$ selects the action which minimizes $C(i, K) + \beta_r' \sum_j P(i, j:K) f_{\beta_r'}(j, s)$. But R^* selects the action which minimizes $C(i, K) + \sum_j P(i, j:K) f(j)$. The result follows since $K_i < \infty$.

PROOF OF (ii): Fix s and let $\{\beta_r''\}_{r=1}^{\infty}$ be a subsequence of $\{\beta_r'\}_{r=1}^{\infty}$ for which $\lim_r g_{\beta_r''}(s)$ and $\lim_r f_{\beta_r''}(i, s)$ exist for all i. Denoting these limits by g and $f(i)$ it follows from Theorem 1.2 that any rule which, when in state i, selects the action which minimizes $C(i, K) + \sum_j P(i, j:K) f(j)$ is optimal. But $R_{\beta_r''}$ minimizes $C(i, K) + \beta_r'' \sum_j P(i, j:K) f_{\beta_r''}(j, s)$ and $R_{\beta_r''} \to R$. Therefore, R is optimal. QED

Thus we see if $K_i < \infty$ for all $i \varepsilon I$, and Assumption (*) holds, then there exists an optimal stationary deterministic rule which is a limit point of $\{R_\beta : 0 < \beta < 1\}$; and any rule which is a limit point of $\{R_{\beta_r}\}_{r=1}^{\infty}$ is optimal. The following theorem gives a sufficient conditon for Assumption (*) to hold.

THEOREM 1.4. *If for some state* j *and sequence* $\beta_r \to 1$ *there is a constant* $N < \infty$ *such that* $M_{ij}(R_{\beta_r}) < N$ *for all* $i \varepsilon I$, $r = 1, 2, \cdots$, *then Assumption* (*) *holds*; *where* $M_{ij}(R_{\beta_r})$ *is the mean recurrence time to go from state* i *to state* j *when using the* β_r*-optimal discount rule* R_{β_r}.

PROOF. Let

$$T = \min\{t : X_t = j\},$$

$$\psi(i, \beta_r, R_{\beta_r}) = E_{R_{\beta_r}} \sum_{n=0}^{T-1} C(X_n, \Delta_n)\beta^n + E_{R_{\beta_r}} \sum_{n=T}^{\infty} C(X_n, \Delta_n)\beta^n,$$

where all expectations are understood to be conditioned on $X_0 = i$. Therefore,

$$\psi(i, \beta_r, R_{\beta_r}) \leqq M E_{R_{\beta r}} T + \psi(j, \beta_r, R_{\beta_r}) E_{R_{\beta r}}(\beta_r^{T})$$

$$\leqq MN + \psi(j, \beta_r, R_{\beta_r})$$

(recall that all costs are positive and bounded by M). Now by the above we have that

$$\psi(i, \beta_r, R_{\beta_r}) \geqq \psi(j, \beta_r, R_{\beta_r}) E_{R_{\beta r}}(\beta_r^{T})$$

$$\therefore \psi(j, \beta_r, R_{\beta_r}) \leqq \psi(i, \beta_r, R_{\beta_r}) + [1 - E_{R_{\beta r}}(\beta_r^{T})]\psi(j, \beta_r, R_{\beta_r});$$

now $\psi(j, \beta_r, R_{\beta_r}) \leqq M(1 - \beta_r)^{-1}$ and $E(\beta^T) \geqq \beta^{ET} \geqq \beta^N$, by Jensen's inequality,

$$\therefore [1 - E_{R_{\beta r}}(\beta_r^{T})]\psi(j, \beta_r, R_{\beta_r}) \leqq (1 - \beta_r^{N})(1 - \beta_r)^{-1} M < NM,$$

$$\therefore |\psi(j, \beta_r, R_{\beta_r}) - \psi(i, \beta_r, R_{\beta_r})| \leqq MN \quad \text{for all} \quad r, i. \qquad \text{QED}$$

LEMMA 1.5. *If for some state j and $\alpha > 0$, $P(i, j:K) \geqq \alpha$ for all $K \varepsilon K_i$, $i \varepsilon I$, then Assumption (*) holds.*

PROOF. For any $R \varepsilon C''$, $M_{ij}(R) \leqq 1/\alpha$, and so Theorem 1.4 applies. QED

2. Determination of optimal policy by reduction of average-cost case to discounted-cost case. We shall need the following assumption.

ASSUMPTION. $\sup_{j \varepsilon I} \inf_{K \varepsilon K_i, i \varepsilon I} P(i, j:K) > 0$. Note this is so if and only if there is a state j and $\alpha > 0$ such that $P(i, j:K) \geqq \alpha$ for all $i \varepsilon I$, $K \varepsilon K_i$. For the sake of definiteness denote the state j for which the above holds by state 0. By Lemma 1.5 there exists a stationary deterministic optimal rule for this process.

Consider now a new process (the prime process) with identical state and action spaces and with the same cost structure but with transition probabilities now given by

$$\begin{aligned} P'(i, j:K) &= P(i, j:K)(1 - \alpha)^{-1}, && j \neq 0, \\ &= (P(i, 0:K) - \alpha)(1 - \alpha)^{-1}, && j = 0. \end{aligned}$$

Denote by $\psi'(i, \beta, R)$ the total expected β-discounted costs when using rule R with respect to the new (prime) process. Note that any rule for the prime process can also be considered as a rule for the original process and vice versa. The fundamental theorem in the reduction is the following:

THEOREM 2.1. *For any stationary rule R,*

$$\varphi(0, R) = \alpha\psi'(0, 1 - \alpha, R).$$

PROOF. In the original problem we shall think of the transitions as taking place in two stages. During stage 1 a coin with probability α of coming up heads is flipped. If heads comes up, then the process goes to state 0; if not, then the process moves to the next state according to the second stage transition probabilities, which are the transition probabilities that are necessary in order to make the total transition probability what it should be, i.e., if action K is chosen, then the total probabilities must be $P(i, j:K)$. Note that the desired second stage transition probabilities are exactly the transition probabilities of the prime problem. Define a cycle as the time between successive occurrences of heads. Let T = time of cycle. Then it is well known (follows from the strong law of large numbers and the bounded convergence theorem) that for any stationary R,

$$\begin{aligned} \varphi(0, R) &= E_R \textstyle\sum_{t=0}^{T-1} C(X_t, \Delta_t)/E_R T \\ &= E_R\{E_R \textstyle\sum_{t=0}^{T-1} C(X_t, \Delta_t) \mid T\}/E_R T \\ &= \textstyle\sum_{i=1}^{\infty} \alpha(1 - \alpha)^{i-1} \sum_{t=0}^{i-1} E_R[C(X_t, \Delta_t) \mid T = i]/1/\alpha. \end{aligned}$$

Now conditioning on $T = i$ means that the transition probabilities used during times $0, 1, \cdots, i - 2$ were the second stage transition probabilities. Therefore, for $t \leqq i - 1$,

$$E_R[C(X_t, \Delta_t) \mid T = i] = E_R'C(X_t, \Delta_t),$$

where E_R' denotes the expected cost with respect to the prime problem. Thus,

$$\begin{aligned}\varphi(0, R) &= \alpha \sum_{i=1}^{\infty} \alpha(1-\alpha)^{i-1} \sum_{t=0}^{i-1} E_R' C(X_t, \Delta_t) \\ &= \alpha \sum_{t=0}^{\infty} E_R' C(X_t, \Delta_t) \sum_{i=t+1}^{\infty} \alpha(1-\alpha)^{i-1} \\ &= \alpha \sum_{t=0}^{\infty} E_R' C(X_t, \Delta_t)(1-\alpha)^t \\ &= \alpha \psi'(0, 1-\alpha, R). \qquad \text{QED}\end{aligned}$$

Since $P(i, 0{:}K) \geqq \alpha$ for all i, K it follows that for $R \varepsilon C''$, $\varphi(i, R) = \varphi(0, R)$; and since an optimal rule does exist in C'' it follows that it is precisely the optimal $1 - \alpha$ discount rule with respect to the prime problem.

Thus if our assumption holds then we have reduced the average-cost problem to a discounted-cost problem, and any of the well-known methods of successive approximations or policy improvements (see [1] for details) may be applied.

3. On ϵ-optimal rules. It is known (see [4]) that even under the conditions that $K_i < \infty$ for all i and $C(i, K)$ uniformly bounded that there need not exist an optimal rule in the average cost sense. Also there may exist an optimal rule, but there may be no stationary deterministic rule which is optimal.

This brings up the question whether there always exist ϵ-optimal stationary deterministic rules. We say that $R \varepsilon C''$ is ϵ-optimal for state i if $\varphi(i, R) < g(i) + \epsilon$, where $g(i) = \inf_R \varphi(i, R)$. We say that $R \varepsilon C''$ is ϵ-optimal if it is ϵ-optimal for every state i. One possible source of ϵ-optimal stationary deterministic rules is the optimal β-discount rules $\{R_\beta{:}0 < \beta < 1\}$. One might conjecture, that for any state i, that these rules are ϵ-optimal for state i in the sense that $\lim \inf_{\beta \to 1-} \varphi(i, R_\beta) = g(i)$. The following counter-example shows that this need not be the case.

EXAMPLE.

$$I = (i, j), \quad 0 \leqq j \leqq i, \quad i \geqq 1, \quad K_{(i,j)} = \begin{cases} 2, & j = 0, \\ 1, & j \neq 0. \end{cases} \quad K_\infty = 1, \quad = \infty.$$

The costs depend only on the state

$$C(i, j), \cdot) = \begin{cases} 1, & j = 0, \\ 0, & j \neq 0. \end{cases} \qquad C(\infty, \cdot) = 2,$$

The transition probabilities are as follows:

$$P((i, 0), (i+1, 0){:}1) = 1 = P((i, 0), (i, 1){:}2),$$
$$P((i, j), (i, j+1){:}1) = 1 \quad \text{for} \quad 0 < j < i,$$
$$P((i, i), \infty{:}1) = 1 = P(\infty, \infty{:}1).$$

In words, when in state $(i, 0)$ we can choose to go to state $(i + 1, 0)$ at the cost of one unit for the next stage or we can elect to pay 0 dollars for the next i stages and two units for every stage after that. Let R_0 be the rule which takes action

1 at all states, then it is easy to see that

$$\varphi((1, 0), R_0) = 1, \qquad R \varepsilon C'', \quad R \neq R_0 \Rightarrow \varphi((1, 0), R) = 2.$$

Let R_n be the rule which takes action 2 at state $(n, 0)$ and action 1 elsewhere:

$$\psi((1, 0), \beta, R_n) = \sum_{i=0}^{n-1} \beta^i + \sum_{i=2n}^{\infty} 2\beta^i = (1 - \beta^n + 2\beta^{2n})(1 - \beta)^{-1}$$

$$\therefore \text{ for } n \text{ large } \psi((1,0), \beta, R_n) < (1 - \beta)^{-1} = \psi((1,0), \beta, R_0),$$

$$\therefore \text{ for each } \beta \varepsilon (0, 1), R_\beta \neq R_0,$$

$$\therefore \varphi((1, 0), R_\beta) = 2 \text{ for all } \beta \text{ and } \inf_R \varphi((1, 0), R) = 1.$$

Thus it is not necessarily true that the β-optimal discount rules are ϵ-optimal, or even ϵ-optimal for a specific state. We shall now give sufficient conditions for R_β to be:

(i) ϵ-optimal for a particular state (for β near 1),

(ii) ϵ-optimal for all β near 1.

THEOREM 3.1. *If for some sequence* $\beta_r \to 1^-$ *there exists an* $N < \infty$ *such that*

$$\psi(j, \beta_r, R_{\beta_r}) - \psi(i, \beta_r, R_{\beta_r}) < N \qquad \text{for all } j \varepsilon I, r = 1, 2, \cdots,$$

then $\lim_{r\to\infty} \varphi(i, R_{\beta_r}) = g(i) = \inf_R \varphi(i, R)$ *and so, for* r *large,* R_{β_r} *is* ϵ*-optimal for state* i.

PROOF. Let

$$V_\beta(j) = \psi(j, \beta, R_\beta),$$

then

$$V_\beta(j) = \min_K \{C(j, K) + \beta \sum_l P(j, l{:}K) V_\beta(l)\}$$

and R_β takes the minimizing actions. Now

$$E_{R_\beta} \sum_{t=1}^{n} [V_\beta(X_t) - E_{R_\beta}[V_\beta(X_t) \mid S_{t-1}]] = 0$$

and

$$\begin{aligned} E_{R_\beta}[V_\beta(X_t) \mid S_{t-1}] &= \sum_j P(X_{t-1}, j{:}\Delta_{t-1}) V_\beta(j) \\ &= \beta \sum_j P(X_{t-1}, j{:}\Delta_{t-1}) V_\beta(j) + C(X_{t-1}, \Delta_{t-1}) \\ &\quad - C(X_{t-1}, \Delta_{t-1}) + (1 - \beta) \sum_j P(X_{t-1}, j{:}\Delta_{t-1}) V_\beta(j) \\ &= V_\beta(X_{t-1}) - C(X_{t-1}, \Delta_{t-1}) \\ &\quad + (1 - \beta) \sum_j P(X_{t-1}, j{:}\Delta_{t-1}) V_\beta(j) \\ \therefore 0 &= E_{R_\beta}[V_\beta(X_n) - V_\beta(X_0)] + E_{R_\beta} \sum_1^n C(X_{t-1}, \Delta_{t-1}) \\ &\quad - (1 - \beta) E_{R_\beta} \sum_1^n V_\beta(X_t). \end{aligned}$$

Using our condition we have that $V_{\beta_r}(X_t) < V_{\beta_r}(i) + N$ for all t

$$\begin{aligned} \therefore n^{-1} E_{R_{\beta_r}} \sum_1^n C(X_{t-1}, \Delta_{t-1}) &\leqq (1 - \beta_r) V_{\beta_r}(i) \\ &\quad + (1 - \beta_r) N - n^{-1} E_{R_{\beta_r}}[V_{\beta_r}(X_n) - V_{\beta_r}(X_0)]. \end{aligned}$$

Letting $n \to \infty$ we have, since $|V_{\beta_r}(X_n) - V_{\beta_r}(X_0)| < M/(1 - \beta_r)$, that $\varphi(X_0, R_{\beta_r}) \leqq (1 - \beta_r)V_{\beta_r}(i) + (1 - \beta_r)N$ for any X_0. Now for any rule R,

$$(1 - \beta)V_\beta(i) \leqq (1 - \beta)\psi(i, \beta, R)$$

$$\therefore \limsup_{r\to\infty} (1 - \beta_r)V_{\beta_r}(i) \leqq \limsup_{r\to\infty} (1 - \beta_r)\psi(i, \beta_r, R) \leqq \varphi(i, R),$$

where the second inequality follows from the Tauberian result [see Titchmarsh, p. 227] that $\limsup_{x\to 1^-} (1 - x) \sum_{n=0}^{\infty} a_n x^n \leqq \limsup_n n^{-1} \sum_1^n a_i$,

$$\therefore \limsup_{r\to\infty} \varphi(X_0, R_{\beta_r}) \leqq \varphi(i, R) \quad \text{for any } R, \text{ any } X_0,$$

$$\therefore \limsup_{r\to\infty} \varphi(X_0, R_{\beta_r}) \leqq g(i) \quad \text{for any } X_0,$$

$$\therefore \lim_{r\to\infty} \varphi(i, R_{\beta_r}) = g(i). \qquad \text{QED}$$

COROLLARY 3.2. *If for some sequence $\beta_r \to 1^-$ there exists $N_i < \infty$ for each $i \varepsilon I$ such that $\psi(j, \beta_r, R_{\beta_r}) - \psi(i, \beta_r, R_{\beta_r}) < N_i$ for all r, j, then:*

(i) $\lim_{r\to\infty} \varphi(i, R_{\beta_r}) = g(i)$ *for all i, and the convergence is uniform in i, and thus R_{β_r} is ϵ-optimal for r large;*

(ii) $g(i) = g(j) = g$ *for all i, j.*

PROOF. We first prove (ii). From the proof of the previous theorem we have that

$$\limsup_{r\to\infty} \varphi(X_0, R_{\beta_r}) \leqq g(i) \quad \text{for any } X_0, \text{ any } i$$

and also that

$$g(X_0) = \lim_{r\to\infty} \varphi(X_0, R_{\beta_r}) \quad \text{for any } X_0$$

$$\therefore g(X_0) \leqq g(i) \quad \text{for any } i, X_0$$

$$\therefore g(i) = g(j) = g \quad \text{for all } i, j.$$

Now we prove (i). The convergence is an immediate result of Theorem 3.1. To show uniformity—fix some state i_0. The previous theorem yields that

$$\limsup_{r\to\infty} (1 - \beta_r)V_{\beta_r}(i_0) \leqq g(i_0)$$

and

$$\varphi(j, R_{\beta_r}) \leqq (1 - \beta_r)V_{\beta_r}(i_0) + N_{i_0}(1 - \beta_r) \quad \text{for any state } j.$$

For $\epsilon > 0$, let r_ϵ be such that $r > r_\epsilon$ implies:

(i) $(1 - \beta_r)V_{\beta_r}(i_0) < g(i_0) + \epsilon/2$,

(ii) $(1 - \beta_r)N_{i_0} < \epsilon/2$.

Therefore, $r > r_\epsilon \Rightarrow \varphi(j, R_{\beta_r}) \leqq g(i_0) + \epsilon/2 + \epsilon/2 = g(i_0) + \epsilon$ for any j, but $g(i_0) = g$ and so convergence is uniform. QED

Note that the condition in the above corollary is weaker than Assumption (*). Thus putting Corollary 3.2 together with Theorems 1.1, 1.2 and 1.3 we have

THEOREM 3.3. *If Assumption (*) holds, then there exists a stationary deterministic optimal rule which is a limit point of the optimal β_r-discount rules, and for any ϵ the β_r-discount rules are ϵ-optimal for r large.*

4. Replacement process.

DEFINITION. A Markovian replacement process is a Markovian decision process with a distinguished state—call it state 0—and a distinguished action —call it a_0—such that:

(i) $X_0 = 0$,

(ii) $P(i, j:a_0) = 1, \quad j = 0,$

$= 0, \quad$ otherwise.

Let $g = \inf_R \varphi(0, R)$. Since $X_0 = 0$ we shall write $\varphi(R)$ for $\varphi(0, R)$, and we shall say that R is optimal (ϵ-optimal) if it is optimal (ϵ-optimal) for state 0.

Let R_β be the β-optimal discount rule. As an immediate consequence of Theorem 3.1 we have

THEOREM 4.1. *In the replacement process*

$$\lim_{\beta \to 1-} \varphi(R_\beta) = g.$$

PROOF.

$$\begin{aligned} \psi(i, \beta, R_\beta) &= \min_K \{C(i, K) + \beta \textstyle\sum_j P(i, j:K)\psi(j, \beta, R_\beta)\} \\ &\leqq C(i, a_0) + \beta\psi(0, \beta, R_\beta) \\ &\leqq M + \psi(0, \beta, R_\beta) \qquad \text{for all } i, \text{ all } \beta, \end{aligned}$$

and so the result follows from Theorem 3.1. QED

The following corollary is immediate.

COROLLARY 4.2. (i) *There exist ϵ-optimal stationary deterministic rules for the replacement problem.*

(ii) *If R is optimal among the stationary deterministic rules, then R is optimal (for the replacement problem).*

DEFINITION. We say that rule R is a Markov rule if the action it chooses at time t only depends on the past history through the state at time t, and t, i.e., $D_R(X_0, \Delta_0, \cdots, X_t = i) = D_{i,R}(t)$.

We say that R is non-random Markov if it is Markov and non-random.

THEOREM 4.3. *For the replacement model there exists a non-random Markov rule which is optimal.*

PROOF. For each n, let R_n be a stationary deterministic rule such that $\varphi(R_n) < g + \frac{1}{2}n$. For each n, $\exists$ N_n such that

$$E_{R_n} \textstyle\sum_0^{i-1} C(X_t, \Delta_t) i^{-1} \leqq \varphi(R_n) + (2n)^{-1},$$

for all $i \geqq N_n$. Let $\bar{N}_1$ be such that

$$[E_{R_1} \textstyle\sum_0^{\bar{N}_1 - 1} C(X_t, \Delta_t) + (N_2 + 1)M][\bar{N}_1 + (N_2 + 1)]^{-1} \leqq \varphi(R_1) + 1,$$

where M is such that $C(i, K) < M$ for all i, K. Define $\bar{N}_i$, $i = 2, 3, \cdots$, recursively by letting $\bar{N}_i$ be such that

$$[E_{R_i} \textstyle\sum_0^{\bar{N}_i - 1} C(X_t, \Delta_t) + M(\sum_{j=1}^{i-1} \bar{N}_j + i + N_{i+1})] \cdot [\bar{N}_i + (\sum_{j=1}^{i-1} \bar{N}_j + i + N_{i+1})]^{-1} < \varphi(R_i) + (2i)^{-1}.$$

Let R be the non-random Markov rule which is defined as follows:

use R_1 for $t = 1, \cdots, \bar{N}_1$, then take action a_0,

use R_2 for the next $\bar{N}_2$ stages, then take a_0,

$\vdots$

use R_i for the next $\bar{N}_i$ stages, then take a_0,

etc.

CLAIM. $\varphi(R) = g$. For any $\epsilon > 0$, let j be such that $j^{-1} < \epsilon$. This claim is then verified by showing that

$$n > \bar{N}_1 + \cdots + \bar{N}_j + j \Rightarrow E_R \sum_1^n C(X_{t-1}, \Delta_{t-1}) n^{-1} < g + \epsilon. \qquad \text{QED}$$

Thus in the replacement problem there always exists an optimal non-randomized rule. That this rule cannot always be taken to be stationary is shown by the following example.

EXAMPLE.

$$I = \{0, 1, 2, \cdots\} K_i = 3 \qquad \text{for all } i,$$

$$C(i, 0) = C(i, 1) = 1, \quad C(i, 2) = 1/i + 1 \qquad \text{for all } i,$$

$$P(i, 0{:}0) = P(i, i + 1{:}1) = P(i, i{:}2) = 1 \qquad \text{for all } i.$$

In words, when in state i we can choose to: (1) remain in state i at the cost of $(i + 1)^{-1}$ units, or (2) go to state $i + 1$ at the cost of 1 unit, or (3) return to state 0 at the cost of 1 unit. (Actually the replacement action is superfluous in the sense that action 1 is always a better action).

For any stationary deterministic rule R let $i(R)$ be the action R chooses when in state i. Let

$$R_1 = \min \{i{:}i(R) \neq 1\},$$

then it is easy to see that

$$R_1 < \infty \Rightarrow \varphi(R) \geqq R_1^{-1} > 0,$$

$$R_1 = \infty \Rightarrow \varphi(R) = 1 > 0.$$

Therefore, R stationary deterministic $\Rightarrow \varphi(R) > 0$.

Now let the non-random Markov rule R^* be as follows: when it first enters state i, $i \geqq 0$, R^* chooses action 2 i times and then it chooses action 1. It is easy to see that $\varphi(R^*) = 0$.

It is also interesting to note that the stationary (but non-deterministic) rule R^{**} which when in state i selects action 2 with probability $i(i + 1)^{-1}$ and action 1 with probability $(i + 1)^{-1}$ is also optimal, i.e., $\varphi(R^{**}) = 0$.

We defined, for the replacement problem, the average cost in terms of the lim sup as opposed to the lim inf. The question arises whether or not this is a meaningful difference. We show that it is not, and both criteria are in a sense alike.

Let

$$\underline{g} = \inf_R \underline{\varphi}(R) \qquad \text{and} \qquad \bar{g} = \inf_R \bar{\varphi}(R)$$

where

$$\underline{\varphi}(R) = \liminf_n E_R \sum_0^{n-1} C(X_t, \Delta_t) n^{-1},$$
$$\bar{\varphi}(R) = \limsup_n E_R \sum_0^{n-1} C(X_t, \Delta_t) n^{-1}.$$

THEOREM 4.4. *For the replacement problem,* $\underline{g} = \bar{g} = g$.

PROOF. Choose $\epsilon > 0$ and let R be such that $\underline{\varphi}(R) < \underline{g} + \epsilon/2$. Choose N such that

$$[E_R \sum_0^{N-1} C(X_t, \Delta_t) + M](N+1)^{-1} < \underline{\varphi}(R) + \epsilon/2.$$

Define R' as follows: R' follows (takes the same actions as) R at times 0, 1, $\cdots$, $N-1$ and then R' takes action a_0 at time N. Thus the process is now in state 0, and we consider it as starting all over again, i.e., we forget that the history up to this time has ever taken place. R' now follows R for the next N stages, then takes a_0, then follows R (pretending the previous history never took place) for the next N stages, then takes a_0, etc. Then it is easy to see that

$$\bar{\varphi}(R') = \underline{\varphi}(R') = [E_R \sum_0^{N-1} C(X_t, \Delta_t) + E_R C(X_N, a_0)](N+1)^{-1}$$
$$\therefore \bar{\varphi}(R') < \underline{\varphi}(R) + \epsilon/2 < \underline{g} + \epsilon,$$
$$\therefore \quad \bar{g} \leqq \underline{g},$$
$$\therefore \quad \bar{g} = \underline{g} \text{ since by definition } \bar{g} \geqq \underline{g}. \qquad \text{QED}$$

COROLLARY 4.5. (i) *For the replacement problem there exist* ϵ*-optimal stationary deterministic rules with respect to the* lim inf *criteria.*

(ii) *The* lim sup *optimal non-randomized Markov rule* R *of Theorem* 4.3 *has* $\underline{\varphi}(R) = g$.

PROOF. (i) $\underline{\varphi}(R) \leqq \bar{\varphi}(R)$ and so the result follows from Corollary 4.2 and the above theorem.

(ii) $\underline{g} \leqq \underline{\varphi}(R) \leqq \bar{\varphi}(R) = g$. QED

5. Acknowledgement. The author would like to express his deep appreciation to Professor Gerald J. Lieberman whose guidance and inspiration helped make this paper possible.

REFERENCES

[1] BLACKWELL, DAVID (1965). Discounted dynamic programming. *Ann. Math. Statist.* **36** 226–235.

[2] DERMAN, CYRUS (1965). Markovian sequential control processes—denumerable state space. *J. Math. Anal. Appl.* **10** 295–302.

[3] DERMAN, CYRUS and LIEBERMAN, GERALD J. (1966). A Markovian decision model for a joint replacement and stocking problem. *Management Sci.* **13** 609–617.

[4] DERMAN, CYRUS (1966). Denumerable state Markovian decision processes—average cost criterion. *Ann. Math. Statist.* **37** 1545–1554.

[5] DERMAN, CYRUS and VEINOTT, ARTHUR (1967). A solution to a countable system of equations arising in Markovian decision processes. *Ann. Math. Statist.* **38** 582–585.

[6] MAITRA, ASHOK (1965). Dynamic programming for countable state systems. *Sankhyā Ser. A* **27** 259–266.

[7] MAITRA, ASHOK (1966). A note on undiscounted dynamic programming. *Ann. Math. Statist.* **37** 1042–1044.

[8] STRAUCH, RALPH (1966). Negative dynamic programming. *Ann. Math. Statist.* **37** 871–890.

[9] TAYLOR, HOWARD (1965). Markovian sequential replacement processes. *Ann. Math. Statist.* **36** 1677–1694.

10] TITCHMARSH, E. C. (1932). *The Theory of Functions*. Oxford Univ. Press.