NON-DISCOUNTED DENUMERABLE MARKOVIAN DECISION MODELS

by

Sheldon M. Ross

TECHNICAL REPORT NO. 94

May 2, 1967

Supported by the Army, Navy, Air Force and NASA under

Contract Nonr-225(53) (NR-042-002)

with the Office of Naval Research

Gerald J. Lieberman, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

NONTECHNICAL SUMMARY

A Markovian Decision Process is a process which is observed at distinct time points to be in some state. After observing the state of the system an action is chosen - corresponding to the action (and the present state) a cost is incurred and the transition probabilities for the next state are determined. A policy is any rule for choosing actions. Corresponding to each policy there is an expected long run average cost per unit time. This paper is concerned with finding an optimal policy - i.e. one whose associated average cost is as small as possible.

For example we might have a machine which deteriorates with time. The state of the system could be the condition of the machine, and the possible actions could be either to replace the machine or not. Associated with each state there would be an operating cost. Thus a policy is a rule for determining when to replace the machine and an optimal one is one which minimizes the long run average cost.

In this paper we let the state space be countable and present sufficient conditions for the existence of an optimal policy and for it to be of simple form. This form - called stationary deterministic - is of the form of a function from the state space to the action space. For example in the machine problem a stationary deterministic policy would replace whenever the machine is in a certain specified class of states.

In a special case the average cost criterion is shown to be equivalent to the discounted cost criterion. This latter criterion has been extensively studied. Under certain conditions the optimal discounted cost policies are shown to be almost optimal for the average cost criterion.

It is also shown that when a replacement action exists (as in the machine problem) then there always exists an optimal policy and the form of this policy is given. The final section gives a counterexample which shows that the optimal rule cannot always be taken to be of the stationary form where a stationary policy is one which at each state the action may be chosen according to some fixed randomization scheme. For instance in the machine problem a stationary policy is one which for each state gives a probability for replacing the machine.

NON-DISCOUNTED DENUMERABLE MARKOVIAN DECISION MODELS

Sheldon M. Ross

Introduction

We are concerned with a process which is observed at times $t = 0,1,2,\ldots$ 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) $P\{X_{t+1} = j \mid X_0,\Delta_0,\ldots X_t = i,\Delta_t = K\} = P(i,j:K)$

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 $\{D_K(X_0,\Delta_0, \ldots X_t)\}_{K=1}^{K_{X_t}}$ satisfying $0 \leq D_K(X_0,\Delta_0, \ldots X_t) \leq 1$ $K = 0,1 \ldots K_{X_t}$

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

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

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

We say that a rule R is stationary if $D_K(X_0,\Delta_0,\ldots X_t = i) = D_{i,K}$ independent of $X_0,\Delta_0,\ldots\Delta_{t-1}$ 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 [2] the process $\{(X_t,\Delta_t)\ t = 0,1,2,\ldots\}$ 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 $\beta\epsilon(0,1)$ and for 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 $X_0 = i$

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

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

Blackwell in [1] has shown for the discounted case that if $K_i < \infty$ and $\{C(i,K)\}$ bounded then there exists a stationary deterministic optimal rule. We shall be mainly concerned in this paper with the average cost criterion. The first results for the average cost criteria which did not assume a finite state space were given by Taylor [9] who worked with a replacement model. A replacement model is one in which there is a distinguished state 0 and action a_0 such that $X_0 = 0$ and $P(i,j:a_0) = \{\begin{smallmatrix}1 & j = 0 \\ 0 & \text{otherwise}\end{smallmatrix}$. Taylor showed that in the finite action replacement model if one can restrict attention to those rules whose expected time between replacements is uniformly bounded then there exists a stationary deterministic optimal rule and it is determined from a functional equation. Taylor's 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 (for each state) action general Markovian model. He treats the problem without using the results for the discounted problem and gives sufficient conditions 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 show that a sufficient condition for the above is that

(i) for each rule $R \epsilon C''$ 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 RϵC" where $M_{i0}(R)$ denotes the mean recurrence time from state i to state 0 when using rule R. [Note that for any rule RϵC" the resulting sequence of states forms a Markov chain].

In the first section of this paper, by following the approach of Taylor, we give a somewhat simpler proof of Derman's results. Also our sufficient conditions will be somewhat weaker: we won't require condition (i) and won't require that $M_{i0}(R) < T$ for all rules R. 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 recuced to the discounted cost case.

The third section deals with ϵ-optimal rules and a sufficient conditions is given for the opitmal 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.

The fifth section given an example of an optimal nonstationary rule which is better than any stationary rule.

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 $i\epsilon I$. Furthermore $\{\psi(i,\beta,R_\beta), i\epsilon I\}$ is the unique solution to (1) $\psi(i,\beta,R_\beta) = \min_K \{C(i,K) + \beta \sum_j P(i,j:K)\, \psi(j,\beta,R_\beta)\}$ $i\epsilon I$

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 $\beta\epsilon(0,1)$ $i,j\epsilon I$

(2) Let $f_\beta(i,j) = \psi(i,\beta,R_\beta) - \psi(j,\beta,R_\beta)$

One has by simple manipulations that

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

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 $|f_{\beta_r}(i,j)| < N$ for all $r = 1,2,\ldots$ all $i,j\epsilon I$.

Theorem 1.1: If Assumption(*)holds then there exists a bounded solution to the functional equation

$$(4)\quad g + f(i) = \min_K \{C(i,K) + \sum_j P(i,j:K) f(j)\} \quad i\epsilon I$$

Proof: Fix some state s. By Assumption (*) $f_{\beta_r}(i,s)$ is uniformly bounded for $r = 1,2,\ldots,$ and all $i \epsilon I$. 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 $\beta_{r'} \to 1$.

∴ 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 if i.

A special version of the next theorem was originally proven by Taylor [9]. Derman [4] proved it under the assumption that $K_i < \infty$ for all i; later in a joint paper with Lieberman [3] a proof not assuming this was given.

We shall give here a simple proof which follows a technique used in [9].

Theorem 1.2: If there exists a bounded solution to the functional equation

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

then there exists a stationary deterministic rule R* such that

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

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

Proof: for any rule R

$$E_R \{\sum_1^n [f(X_t) - E_R(f(X_t) \mid S_{t-1})]\} = 0$$

where X_t = state at time t, and

$S_{t-1} = (X_0,\Delta_0,\dots X_{t-1},\Delta_{t-1})$ = history up to time t.

$$\text{now } E_R[f(X_t) \mid S_{t-1}] = \sum_j P(X_{t-1},j:\Delta_{t-1})\, f(j)$$

$$= C(X_{t-1},\Delta_{t-1}) + \sum_j P(X_{t-1},j:\Delta_{t-1})f(j) - C(X_{t-1},\Delta_{t-1})$$

$$\geq \min_K \{C(X_{t-1},K) + \sum_j P(X_{t-1},j:K)f(j)\} - C(X_{t-1},\Delta_{t-1})$$

$$\geq g + f(X_{t-1}) - C(X_{t-1},\Delta_{t-1})$$

with equality for R* since R* is defined to take the minimizing action

$$\therefore\ 0 \leq E_R\{\sum_1^n f(X_t) - g - f(X_{t-1}) + C(X_{t-1},\Delta_{t-1})\}$$

with equality for R*.

$$\therefore\ g \leq \frac{E_R f(X_n)}{n} - \frac{E_R(f(X_0))}{n} + \frac{E_R \sum_1^n C(X_{t-1},\Delta_{t-1})}{n} \quad \text{with = for R*}$$

letting $n \to \infty$ and using fact that $f(X_n)$ is bounded, we have $g \leq \varphi(R,X_0)$ with equality for R* and for all possible values of X_0. QED.

Note: the above proof doesn't make use of the fact that $K_i < \infty$ or that $C(i,K)$ is bounded. Also it shows that $\lim_n E_{R*} \sum_0^n \frac{C(X_t,\Delta_t)}{n+1}$ exists and equals g, and that

$$g \leq \lim_n \inf E_R \sum_0^n \frac{C(X_t,\Delta_t)}{n+1} \quad \text{for every rule R.}$$

Thus, in this case, the fact that the average cost was defined by the lim sup as opposed to the lim inf is irrelevant.

For any rule $R\epsilon C''$

Let $i(R)$ = action R chooses when in state i - i.e. $D_{i,i(R)} = 1$

Def: For rules $R_n, R \epsilon C''$ 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 $n \geq N_i$.

Note that any countable sequence of rules $R_n \epsilon C''$ has a convergent subsequence.

The following theorem shows the relationship between R* and the β-discount optimal rules R_β.

Theorem 1.3: If Assumption(*) holds then

(i) for some subsequence $\{\beta_{r'}\}_{r'=1}^{\infty}$ of $\{\beta_r\}_{r=1}^{\infty}$

$R^* = \lim R_{\beta_{r'}}$.

(ii) if $R = \lim_{r'} R_{\beta_{r'}}$ where $\{\beta_{r'}\}$ is a subsequence of $\{\beta_r\}_{r=1}^{\infty}$ then R is optimal i.e. $\varphi(i,R) = g$ for all $i \epsilon I$

Proof: (i) Let $\{\beta_{r'}\}_{r'=1}^{\infty}$ be the subsequence for which

$$f_{\beta_{r'}}(i,s) \to f(i) \quad \text{and } g_{\beta_{r'}}(s) \to g \text{ as } \beta_{r'} \to 1$$

it is easily seen from the definition of $f_\beta(j)$ that $R_{\beta_{r'}}$ takes the actions which minimize $C(i,K) + \beta_{r'} \sum_j P(i,j:K) f_{\beta_{r'}}(j)$ but R* takes the actions which minimize $C(i,K) + \sum_j P(i,j:K) f(j)$. The result follows since $K_i < \infty$.

(ii) For any sequence $\{\beta_{r'}\}_{r'=1}^{\infty}$ we can get a subsequence $\{\beta_{r''}\}_{r''=1}^{\infty}$ for which $\lim f_{\beta_{r''}}(i,s)$ and $\lim g_{\beta_{r''}}$ exist. Denoting the limit by f(i) it follows from theorem 2 that any rule which minimizes $C(i,K) + \sum_j P(i,j:K) f(j)$ is optimal. But $R_{\beta_{r''}}$ minimizes

$C(i,K) + \beta_{r''} \sum_{j} P(i,j:K) f_{\beta_{r''}}(j)$ and $R_{\beta_{r''}} \to R$

$\therefore$ R is optimal. QED

Thus we see if $K_i < \infty$ for all $i \in I$, 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 $\{R_{\beta_r}\}$ is optimal.

The following theorem gives a sufficient condition 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 $i \in I$, $r = 1,2,\ldots$ 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: Consider the fixed rule R_{β_r}. Suppose the process starts at state i. Let t = time it takes to first get to state j.

$$\text{now} \quad \psi(i,\beta_r,R_{\beta_r}) = E_{R_{\beta_r}}(C_1) + E_{R_{\beta_r}}(C_2)$$

where C_1 = discounted costs incurred before one gets to state $j = \sum_{n=0}^{t-1} C(X_n,\Delta_n)\beta^n$

C_2 = discounted costs incurred after one gets to state $j = \sum_{n=t}^{\infty} C(X_n,\Delta_n)\beta^n$

$$\therefore \quad \psi(i,\beta_r,R_{\beta_r}) \leq M\,Et + \psi(j,\beta_r,R_{\beta_r})\, E_{R_{\beta_r}}(\beta_r^{\,t})$$

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

(recall that all costs are positive and bounded by M)

$$\therefore \quad \psi(i,\beta_r,R_{\beta_r}) - \psi(j,\beta_r,R_{\beta_r}) \leq MN \quad \text{for all } i,\ r = 1,2,\ldots$$

again $$\psi(i,\beta_r,R_{\beta_r}) = E_{R_{\beta_r}}(C_1) + E_{R_{\beta_r}}(C_2)$$

$$\geq E_{R_{\beta_r}}(C_2) = E_{R_{\beta_r}}(\beta_r^{\ t})\,\psi(j,\beta_r,R_{\beta_r})$$

$$\therefore \quad \psi(j,\beta_r,R_{\beta_r}) \leq \psi(i,\beta_r,R_{\beta_r}) + [1 - E_{R_{\beta_r}}(\beta_r^{\ t})]\,\psi(j,\beta_r,R_{\beta_r})$$

now $$\psi(j,\beta_r,R_{\beta_r}) \leq \frac{M}{1 - \beta_r}$$

and $E(\beta^t) \geq \beta^{Et} \geq \beta^N$ by Jensen's Inequality.

$$\therefore \quad [1 - E_{R_{\beta_r}}(\beta_r^{\ t})]\,\psi(j,\beta_r,R_{\beta_r}) \leq \frac{1 - \beta_r^{\ N}}{1 - \beta_r} M < NM$$

$$\therefore \quad |\psi(j,\beta_r,R_{\beta_r}) - \psi(i,\beta_r,R_{\beta_r})| \leq MN \quad \text{for all } r = 1,2,\ldots \ i\epsilon I$$

$$\therefore \quad |\psi(s,\beta_r,R_{\beta_r}) - \psi(i,\beta_r,R_{\beta_r})| \leq 2MN \ \text{for all } r = 1,2,\ldots \ i,s\epsilon I$$

$\therefore$ Assumption (*) holds. QED

<u>Lemma 1.5</u>: If for some $\alpha > 0$ and some state j

$P(i,j:K) \geq \alpha$ for all $i\epsilon I$, $K\epsilon K_i$

then Assumption (*) holds and there exists a

stationary deterministic optimal rule which is determined by the functional equation (5).

Proof: For any rule R, $M_{ij}(R) < 1/\alpha$ and so theorem 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: $\sup_{j \in I} \inf_{\substack{K \in K_i \\ i \in I}} P(i,j:K) > 0$

Note this is so if and only if there is a state j and $\alpha > 0$ such that $P(i,j:K) \geq \alpha$ for all $i \in I$, $K \in K_i$. For the sake of definiteness denote the state j for which the above holds by state 0. By Lemma 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 but with transition probabilities now given by

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

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 2 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 which 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 above is legitimate since $P(i,0:K) \geq \alpha$ for all i,K. Note also 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

$$\varphi(0,R) = E_R \sum_{t=0}^{T-1} C(X_t,\Delta_t)/ E_R T$$

$$= E_R (E_R \sum_{t=0}^{T-1} C(X_t,\Delta_t) \mid T) /E_R T$$

$$= \sum_{i=1}^{\infty} \alpha(1 - \alpha)^{i-1} \sum_{t=0}^{i-1} E_R[C(X_t,\Delta_t)\mid T = i] / 1/\alpha$$

now conditioning on T = i means that the transition probabilities used during times 0,1,... i - 2 were the 2nd stage transition probabilities, i.e. the transition probabilities of the prime problem.

$$\therefore \quad \text{for } t \leq i - 1, \; E_R[C(X_t,\Delta_t) \mid T = i] \; = \; E_R' C(X_t,\Delta_t)$$

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

$$\therefore \quad \varphi(0,R) \; = \alpha \sum_{i=1}^{\infty} \alpha(1-\alpha)^{i-1} \sum_{t=0}^{i-1} E_R' \, C(X_t,\Delta_t)$$

$$= \alpha \sum_{t=0}^{\infty} E_R' \, C(X_t,\Delta_t) \sum_{i=t+1}^{\infty} \alpha(1-\alpha)^{i-1} \quad \text{since everything is non-negative}$$

$$= \alpha \sum_{t=0}^{\infty} E_R' \, C(X_t,\Delta_t) \, (1-\alpha)^t$$

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

QED

Note that since $P(i,0:K) > \alpha$ for all i,K it follows that for any stationary rule R

$$\varphi(i,R) \; = \; \varphi(0,R)$$

and since a stationary deterministic optimal rule does exist it follows from the above theorem that this rule is precisely the optimal $1-\alpha$ discount rule with respect to the prime problem.

Letting $V(i) \; = \; \min\limits_{R} \varphi'(i,1-\alpha,R)$

we have that

$$V(i) \; = \; \min_K \{C(i,K) + (1-\alpha) \sum_j P'(i,j:K) \, V(j)\} \quad i \in I$$

or

$$V(i) \; = \; \min_K \{C(i,K) + \sum_j P(i,j:K) \, V(j) - \alpha V(0)\} \quad i \in I$$

and the optimal policy is the one which chooses the minimizing actions. Let $B(I)$ = Space of all bounded functions on I then defining the operator $T:B(I) \to B(I)$ by

$$(TU)_i = \min_K \{C(i,K) + \sum_j P(i,j:K)\, U(j) - \alpha U(0)\} \quad i \in I$$

Then under the supremum norm $\|U\| = \sup_{i \in I} |U(i)|$

We have that T is a contraction mapping with unique fixed point V, and V can be found by the simple to apply method of successive approximations i.e. for any $U \in B(I)$, $\lim_n T^n U = V$.

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 - can be applied.

Note that policy improvements for this $1-\alpha$ discount problem are by virtue of theorem 2.1 also policy improvements for the original average-cost problem.

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 $R \epsilon C''$ is ϵ-optimal for state i if

$$\varphi(i,R) \quad < \quad g(i) \quad + \quad \epsilon$$

$$\text{where} \qquad g(i) \quad = \quad \inf_{R} \varphi(i,R)$$

We say that $R\epsilon C''$ is ϵ-optimal if it is ϵ-optimal for every state i. One possible source of ϵ-optimal stationary deterministic rules are 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

$$\liminf_{\beta \to 1^-} \varphi(i,R_\beta) \quad = \quad g(i)$$

The following counter-example shows that this need not be the case.

$$\text{Let} \quad I \quad = \quad \begin{cases} (i,j) & j = 0,1,\ldots i, \quad i = 1,2,\ldots \\ & \end{cases}$$

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

The costs depend only on the state

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

$$C(\infty,\cdot) = \quad 2$$

the transition probabilities are as follows

$$P((i,0),(i + 1,0):1) \quad = \quad 1$$

$$P((i,0),(i,1):2) \quad = \quad 1$$

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

$$P((i,i),\infty:1) \quad = \quad 1$$

$$P(\infty,\infty:1) \quad = \quad 1$$

In words, when in state $(i,0)$ we can choose to go to state $(i + 1,0)$ at the cost of 1 unit for the next stage or we can elect to pay 0 dollars for the next i stages and 2 units for every stage after that.

Let $X_0 = (1,0)$

Let R_0 = rule which takes action 1 at all states, then it is easy to see that

$$\varphi((1,0),R_0) = 1$$

$$R \text{ stationary deterministic}, R \neq R_0 \Rightarrow \varphi((1,0),R) = 2$$

We now show that R_0 is not a β-optimal rule for any $\beta (0 < \beta < 1)$ and thus $\varphi((1,0),R_\beta) = 2$ for all $\beta \epsilon (0,1)$

$$\text{while} \quad \inf_R \varphi\,((1,0),R) = 1$$

now $\psi((1,0),\beta,R_0) = \frac{1}{1-\beta}$ since the cost at each stage is 1.

Let R_n = rule which takes action 2 at state $(n,0)$ and action 1 elsewhere.

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

$$= \frac{1 - \beta^n + 2\beta^{2n}}{1 - \beta}$$

now for n large $2\beta^{2n} - \beta^n < 0$

$\therefore$ for n large $\psi((1,0),\beta,R_n) < \frac{1}{1 - \beta}$

$\therefore$ for each β, $R_\beta \neq R_0$

$\therefore$ $\varphi((1,0), R_\beta) = 2$ for all β

and $\inf_R \varphi((1,0),R) = 1$ QED

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

<u>Theorem 3.1</u>: If for some sequence $\beta_r \to 1^-$, there exists an $N < \infty$, such that

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

then
$$\lim_{\beta_r \to 1} \varphi(i,R_{\beta_r}) = g(i) = \inf_{R} \varphi(i,R)$$

and so, for β_r large, R_{β_r} is ϵ-optimal for state i.

Proof: (i) Let $V_\beta(j) = \psi(j,\beta,R_\beta)$

$$\text{then } V_\beta(j) = \min_{K} \{C(j,K) + \beta\sum_{l} P(j,l:K)\, V_\beta(l)\}$$

and R_β takes the minimizing actions

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

$$\begin{aligned}
\text{and } E_{R_\beta}[V_\beta(X_t) \mid S_{t-1}] &= \sum_{j} P(X_{t-1},j:\Delta_{t-1})\, V_\beta(j) \\
&= \beta\sum_{j} P(X_{t-1},j:\Delta_{t-1})\, V_\beta(j) + C(X_{t-1},\Delta_{t-1}) \\
&\quad -C(X_{t-1},\Delta_{t-1}) + (1-\beta)\sum_{j} P(X_{t-1},j:\Delta_{t-1}) V_\beta(j) \\
&= V_\beta(X_{t-1}) - C(X_{t-1},\Delta_{t-1}) \\
&\quad + (1-\beta)\sum_{j} P(X_{t-1},j:\Delta_{t-1})\, V_\beta(j)
\end{aligned}$$

$$\therefore\ 0 = E_{R_\beta}[V_\beta(X_n) - V_\beta(X(0))] + E_{R_\beta}\sum_{1}^{n} C(X_{t-1},\Delta_{t-1}) - (1-\beta)\, E_{R_\beta}\sum_{1}^{n} V_\beta(X_t)$$

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

$$\therefore \quad \frac{1}{n} E_{R_{\beta_r}} \sum_1^n C(X_{t-1},\Delta_{t-1}) \leq (1-\beta_r)V_{\beta_r}(i) + (1-\beta_r)N - \frac{1}{n} E_{R_{\beta_r}} [V_{\beta_r}(X_n) - V_{\beta_r}(X_0)]$$

Letting $n \to \infty$ we have since $| V_{\beta_r}(X_n) - V_{\beta_r}(X_0) | < \frac{M}{1 - \beta_r}$

that $\varphi(X_0,R_{\beta_r}) \qquad - \beta_r) V_{\beta_r}(i) + (1 - \beta_r)N$ for any X_0.

now for any rule R

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

$$\therefore \quad \overline{\lim_{r\to\infty}}\, (1 - \beta_r)\, V_{\beta_r}(i) \leq \overline{\lim_{r\to\infty}}\, (1 - \beta_r)\ \psi(i,\beta_r,R) \leq \varphi(i,R)$$

where the second inequality follows from the Tauberian result that

$$\overline{\lim_{X\to 1^-}}\, (1 - X) \sum_{n=0}^{\infty} a_n X^n \leq \overline{\lim_{n}}\, \frac{1}{n} \sum_1^n a_i \text{ [see Titchmarsh - p. 227]}$$

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

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

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

Corollary 3.2: If for some sequence $\beta_r \to 1^-$, there exists $N_i < \infty$ for each $i \epsilon I$ 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

$$\overline{\lim_{r \to \infty}} \varphi(X_0, R_{\beta_r}) \leq g(i) \text{ for any } X_0\text{, any } i\text{, and also}$$

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

$$\therefore \ g(X_0) \leq g(i) \qquad \text{for any } i, X_0$$

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

(ii) The convergence is an immediate result of thecrem 3.1.

To show uniformity - fix some state i_0

The previous theorem yields that

$$\overline{\lim_{r \to \infty}} (1 - \beta_r) V_{\beta_r}(i_0) \leq g(i_0)$$

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

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

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

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

$\therefore$ $r > r_\epsilon \Rightarrow \varphi(j, R_{\beta_r}) \leq g(i_0) + \epsilon/2 + \epsilon/2 = g(i_0) + \epsilon$ 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 the condition that $\exists N < \infty$ such that $| \psi(j, \beta_r, R_{\beta_r}) - \psi(i, \beta_r, R_{\beta_r}) | < N$ for all r,i,j. This latter condition is Assumption (*).

From Theorem 1.4 we thus have

Theorem 3.3: If for some state j, there is $N < \infty$ such that $M_{ij}(R_{\beta_r}) < N$ for all $r = 1,2,\ldots$ all i then the condition for corollary 3.2 is satisfied and thus $g = \lim_{r \to \infty} \varphi(i, R_{\beta_r})$ uniformly in i.

Proof: see proof of Theorem 1.4.

Putting corollary 3.2 together with Theorems 1.1, 1.2 and 1.3 we have

Theorem 3.2: 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) $\zeta_0 = 0$

(ii) $P(i,j:a_0) = \begin{cases} 1 & j = 0 \\ 0 & \text{otherwise} \end{cases}$

Let $g = \inf_R \varphi(0,R)$ since $\chi_0 = 0$ we shall write $\varphi(R)$ for $\varphi(0,R)$

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

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

Proof: $\psi(i,\beta,R_\beta) = \min_K \{C(i,K) + \beta \sum_j P(i,j:K)\, \psi(j,\beta,R_\beta)\}$

$$\leq C(i,a_0) + \beta\psi(0,\beta,R_\beta)$$

$$\leq M + \psi(0,\beta,R_\beta) \quad \text{for all } i, \text{ all } \beta$$

and so the result follows from theorem 3.1. QED

[Recall that since C(i,K) is assumed bounded we can thus also assume without loss of generality that costs are non-negative and so $\psi(0,\beta,R_\beta) \geq 0$]

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).

Def: We say that rule R is a Markov rule if the action it chooses at time t only depends on the past history thru the state at time t, and t.

$$\text{i.e. } D_K(X_0,\Delta_0,\ldots X_t = i) = D_{i,K}(t)$$

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 $\varphi(R_n) < g + 1/2n$.

for each n, $\exists\, N_n$ such that $E_{R_n} \dfrac{\sum_0^{i-1} C(X_t,\Delta_t)}{i} \leq \varphi(R_n) + \dfrac{1}{2n}$ for all $i \geq N_n$

Let $\bar{N}_1$, be such that $\dfrac{E_{R_1} \sum_0^{\bar{N}_1-1} C(X_t,\Delta_t) + (N_2 + 1)M}{\bar{N}_1 + (N_2 + 1)} \leq \varphi(R_1) + 1$

where M is such that $C(i,K) < M$ for all i,K

Define $\bar{N}_i$ $i = 2,3,\ldots$ recursively by letting $\bar{N}_i$ be such that

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

Let R = non-random Markov be as follows

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

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

: :

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 $1/j < \epsilon$

We shall show that $n > \bar{N}_1 + \ldots \bar{N}_j + j \Rightarrow \dfrac{E_R \sum_1^n C(X_{t-1},\Delta_{t-1})}{n} < g + \epsilon$

now if $n > \bar{N}_1 + \dots + \bar{N}_j + j$ then for some $k \geq j$

$$\bar{N}_1 + \dots + \bar{N}_k + k \quad ' \quad \bar{N}_1 + \dots + \bar{N}_k + \bar{N}_{k+1} + k + 1$$

now $$\frac{E_R \sum_1^n C(X_{t-1},\Delta_{t-1})}{n}$$

$$\leq \frac{(\bar{N}_1 + \dots + \bar{N}_{k-1} + k)M + E_{R_k} \sum_1^{\bar{N}_k} C(X_{t-1}\Delta_{t-1}) + E_{R_{k+1}} \sum_1^{n-\sum_1^k \bar{N}_i - k} C(X_{t-1},\Delta_{t-1})}{n}$$

there are 2 cases

Case (i): $n - \sum_1^k \bar{N}_i - k < N_{k+1}$

Case (ii): $n - \sum_1^k \bar{N}_i - k \geq N_{k+1}$

If (i) then $$\frac{E_R \sum_1^n C(X_{t-1},\Delta_{t-1})}{n}$$

$$\leq \frac{(\bar{N}_1 + \dots + \bar{N}_{k-1} + k)M + E_{R_k} \sum_1^{\bar{N}_k} C(X_{t-1},\Delta_{t-1}) + N_{k+1} M}{\bar{N}_1 + \dots + \bar{N}_{k-1} + k + N_{k+1} + \bar{N}_k}$$

$$< \varphi(R_k) + \frac{1}{2k} < g + \frac{1}{2k} + \frac{1}{2k} = g + 1/k < g + \epsilon$$

If (ii) then $\dfrac{E_R \sum_{1}^{n} C(X_{t-1},\Delta_{t-1})}{n}$

$$\leq \frac{\left(\sum_{1}^{k-1} \bar{N}_i + k\right)M + E_{R_k} \sum_{1}^{\bar{N}_k} C(X_{t-1},\Delta_{t-1}) + \left(n - \sum_{1}^{k} \bar{N}_i - k\right)\left(\varphi(R_{k+1}) + \frac{1}{2(k+1)}\right)}{n}$$

$$\leq \frac{\left(\varphi(R_k) + \frac{1}{2k}\right)\left(\sum_{1}^{k-1} \bar{N}_i + k\right) + \left(n - \sum_{1}^{k} \bar{N}_i - k\right)\left(\varphi(R_{k+1}) + \frac{1}{2(k+1)}\right)}{n}$$

$$\leq \quad g + \frac{1}{k}$$

$$< \quad g + \epsilon$$

$\therefore \quad \varphi(R) = g$ 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 4.4: Let $I = \{0,1,2,\ldots\}$

$K_i = 3 \quad i = 0,1,\ldots$

The costs are as follows:

$C(i,0) = 1$ for all i

$C(i,1) = 1$ for all i

$C(i,2) = 1/i+1$ for all i

transition probabilities are as follows

$$P(i,0:0) = 1$$

$$P(i,i+1:1) = 1$$

$$P(i,i:2) = 1$$

In words when in state i we can choose to (1) remain in state i at the cost of 1/i+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].

Let R be any stationary deterministic rule

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

$$R_1 < \infty \Rightarrow \varphi(R) \geq \frac{1}{R_1} > 0$$

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

$$\therefore \quad R \text{ stationary deterministic} \Rightarrow \varphi(R) > 0$$

Let the non-random Markov rule R* be as follows:

When it first enters state i, i = 0,1,... 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** is also optimal. i.e. $\varphi(R^{**}) = 0$

where R** is the rule which when in state i selects action 2 with probability i/i+1 and action 1 with probability 1/i+1. QED

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

$$\text{where } \underline{\varphi}(R) = \liminf_n \frac{E_R \sum_0^{n-1} C(X_t,\Delta_t)}{n}$$

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

Theorem 4.4: For the replacement problem

$$\underline{g} = \bar{g} = g$$

Proof: Choose $\epsilon > 0$

Let R be such that $\underline{\varphi}(R) < \underline{g} + \epsilon/2$

$$\text{Choose N such that } \frac{E_R \sum_0^{N-1} C(X_t,\Delta_t) + M}{N+1} < \underline{\varphi}(R) + \epsilon/2$$

Define R' as follows

R' follows (takes the same actions as) R at times 0,... N-1

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

$$\bar{\varphi}(R') = \underline{\varphi}(R') = \frac{E_R \sum_{0}^{N-1} C(X_t,\Delta_t) + E_R C(X_N,a_0)}{N + 1}$$

$$\therefore \quad \bar{\varphi}(R') < \underline{\varphi}(R) + \epsilon/2 < \underline{g} + \varepsilon$$

$$\therefore \quad \bar{g} \leq \underline{g}$$

$$\therefore \quad \bar{g} = \underline{g} \text{ since by definition } \bar{g} \geq \underline{g} \qquad \text{QED}$$

Corollary 4.5: (i) 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) $\underline{\varphi}(R) \leq \bar{\varphi}(R)$ and so result follows from Corollary 4.2 and the above theorem.

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

5. Counterexample

In 4.4 an example is given of a process for which there is an optimal stationary rule but not an optimal stationary deterministic rule. This brings up the question first raised by Derman [4] of whether one need ever go outside the class of stationary rules. The following is an example for which there is an optimal nonstationary rule which is better than every stationary rule. The optimal rule may be taken to be a non-random Markov rule.

Example 5.1: Let $I = \{0,1,1',2,2',3,3',\ldots\}$

$$K_0 = 1, \; K_i = 2 \quad i = 1,2,\ldots, \; K_{i'} = 1 \quad i' = 1',2' \ldots$$

The cost depends only on the state and is zero except at state 0.

$$C(0,\cdot) = 1$$

$$C(i,\cdot) = 0 \qquad i = 1,2,\ldots$$

$$C(i',\cdot) = 0 \qquad i' = 1',2',\ldots$$

The transition probabilities are as follows:

$$P(i,i+1:1) = 1/2 \qquad i = 0,1,\ldots$$

$$P(i,0:1) = 1/2 \qquad i = 0,1,\ldots$$

$$P(i,i':2) = 1 - (1/2)^i \qquad i = 1,2,\ldots$$

$$P(i,0:2) = (1/2)^i \qquad i = 1,2,\ldots$$

$$P(i',i':1) = 1 - (1/2)^i \qquad i' = 1',2',\ldots$$

$$P(i',0:1) = (1/2)^i \qquad i' = 1',2',\ldots$$

Suppose that $X_0 = 0$

Let R_i be the stationary deterministic rule which takes action 2 when in state i and action 1 in all other states.

Define a cycle T as the time it takes to return to state 0. Then it is well known that

$$\varphi(0,R_i) = 1/E_{R_i}(T)$$

now
$$E_{R_i}(T) = \sum_{j=1}^{i} j(1/2)^j + (1/2)^i\,(i + 2^i)$$

$$= 3 - (1/2)^{i-1} < 3$$

$\therefore$
$$\varphi(0,R_i) = \frac{1}{3 - (1/2)^{i-1}} > \frac{1}{3} \qquad i = 1,2,\ldots$$

now let R be any stationary rule. Let P_i be the probability that R chooses action 2 when in state i,

now
$$\varphi(0,R) = 1/E_R(t)$$

and
$$E_R(T) = \sum_{i=1}^{\infty} \left(\prod_{j=1}^{i-1} (1 - P_j)\right) P_i\, E_{R_i}(T) + 2 \prod_{i=1}^{\infty} (1 - P_i)$$

$$< 3\left[\sum_{i=1}^{\infty} P_i \prod_{j=1}^{i-1} (1 - P_j) + \prod_{i=1}^{\infty} (1 - P_i)\right] = 3$$

$\therefore$
$$\varphi(0,R) > \frac{1}{3} \text{ for any stationary rule R.}$$

But if we define $\bar{N}_i$ $i = 1,2,\ldots$ as in theorem 4.3 we can let R* be the nonrandom Markov rule which uses R_1 for $t = 1, \ldots \bar{N}_1$

R_2 for $t = \bar{N}_1 + 1, \ldots \bar{N}_1 + \bar{N}_2$

$\vdots$

R_i for $t = \sum_{1}^{i-1} \bar{N}_j + 1, \ldots \sum_{1}^{i} \bar{N}_j$

$\vdots$

etc.

Then we can show as before that

$$\varphi(0,R^*) = \lim_{i} \varphi(0,R_i) = 1/3$$

$$\therefore \qquad \varphi(0,R^*) < \varphi(0,R) \text{ for all stationary } R.$$

It also follows by Theorem 3.1 that R* is optimal. QED.

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 Prcgramming, Annals of Mathematical Statistics, 36, 226-235.

[2] Derman, Cyrus (1965), Markovian Sequential Control Processes - Denumerable State Space, Journal of Mathematical Analysis and Applications, 10, 295-302.

[3] Derman, Cyrus and Lieberman, Gerald J. (1966), A Markovian Decision Model for a Joint Replacement and Stocking Problem, (To appear in Management Science).

[4] Derman, Cyrus (1966), Denumerable State Markovian Decision Processes - Average Cost Criterion, Annals of Mathematical Statistics, 37, 1545-1554.

[5] Derman, Cyrus and Veinott, Arthur, (1966), A Solution to a Countable System of Equations Arising in Markovian Decision Processes, Technical Report No. 89, Dept. of Statistics, Stanford University.

[6] Maitra, Ashok (1965, Dynamic Programming for Countable State Systems, Sankhya Ser A, 27, 259-266.

[7] Maitra, Ashok (1966), A Note on Undiscounted Dynamic Programming, Annals of Mathematical Statistics, 37, 1042-1044.

[8] Strauch, Ralph (1966), Negative Dynamic Programming, Annals of Mathematical Statistics, 37, 871-890.

[9] Taylor, Howard (1965), Markovian Sequential Replacement Processes, Annals of Mathematical Statistics, 36, 1677-1694.

[10] Titchmarsh, E.C. (1932), The Theory of Functions, Oxford University Press.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)

Stanford University
Department of Statistics
Stanford, California

2a REPORT SECURITY CLASSIFICATION

Unclassified

2b GROUP

3 REPORT TITLE

Non-discounted Denumerable Markovian Decision Models

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5 AUTHOR(S) (Last name, first name, initial)

Ross, Sheldon, M.

6 REPORT DATE

May 2, 1967

7a TOTAL NO OF PAGES

33

7b NO OF REFS

10

8a CONTRACT OR GRANT NO

Nonr-225(53)

b. PROJECT NO.

NR-042-002

c

d

9a ORIGINATOR'S REPORT NUMBER(S)

Technical Report No. 94

9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11 SUPPLEMENTARY NOTES

12 SPONSORING MILITARY ACTIVITY

Logistics and Mathematical Statistics Branch
Office of Naval Research
Washington, D.C. 20360

13 ABSTRACT

Countable state, finite action Markovian decision processes are studied under the average cost criterion. The problem is studied by using the known results for the discounted-cost problem. Sufficient conditions are given for the existence of an optimal rule which is of the stationary deterministic type. This rule is shown to be, in some sense, a limit point of the optimal discounted-cost rules. Sufficient conditions are also given for the optimal discounted-cost rules to be ϵ-optimal with respect to the average cost criterion. It is shown that if there is a replacement action then there exists an optimal rule but it may not be of the stationary deterministic type. It is also shown how, in a special case, the average cost criterion can be reduced to the discounted cost criterion. Lastly, an example is given of a process for which there exists an optimal nonstationary rule which is better than any stationary rule.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Markovian Decision Process						
Markovian Replacement Process						
Discrete Dynamic Programming						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_______________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_______________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_______________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.