stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 14/71 SEPTEMBER

A. HORDIJK

A SUFFICIENT CONDITION FOR THE EXISTENCE OF AN OPTIMAL POLICY WITH RESPECT TO THE AVERAGE COST CRITERION IN MARKOVIAN DECISION PROCESSES

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

A sufficient condition for the existence of an optimal policy with respect to the average cost criterion in Markovian decision processes.

A. Hordijk

1. INTRODUCTION.

We are concerned with a dynamic system which at times $t = 0,1, \ldots$ is observed to be in one of a possible number of states. Let I denote the space of all possible states. We assume I to be countable. If at time t the system is observed in state i then a decision k must be chosen from a given finite set K(i). Let $\{y_t\}$ and $\{\Delta_t\}$, $t = 0,1, \ldots$, denote the sequences of states and decisions.

If the system is in state i at time t and decision k is chosen, then two things occur:

(i) A known cost w_{ik} is incurred; assume that this cost function is bounded uniformly in i and k.

(ii) $P\{y_{t+1} = j \mid y_0, \Delta_0, \ldots, y_t = i, \Delta_t = k\} = q_{ij}(k)$ i.e., the transition probabilities from one state to another are functions only of the last observed state and the subsequent decision. It is assumed that the $q_{ij}(k)$'s are known.

A rule or policy R for controlling the system is a set of funtions $\{D_k(y_0,\Delta_0, \ldots, y_t)\}$ satisfying for every history $y_0, \Delta_0, \ldots, y_t$ $(t = 0,1, \ldots)$, $0 \leq D_k(y_0,\Delta_0,\ldots,y_t) \leq 1$, for every k, and $\sum_{k\in K(i)} D_k(y_0,\Delta_0, \ldots, y_t = i) = 1$. $D_k(y_0,\Delta_0, \ldots, y_t)$ is the instruction at time t to make decision k with probability $D_k(y_0,\Delta_0, \ldots, y_t)$ if the particular history $y_0,\Delta_0, \ldots, y_t$ has occurred.

The process $\{(y_t,\Delta_t)\ t = 0,1, \ldots\}$ is called a Markovian decision process.

Let C denote the class of all possible policies. Let C^M

denote the class of all memoryless rules, i.e. $D_k(y_0,\Delta_0, \ldots, y_t = i) = D^t_{ik}$ independent of the past history except for the present state. A stationary rule is a memoryless rule for which $D^t_{ik} = D_{ik}$ independent of t. Let C^S denote the class of nonrandomized stationary rules, i.e. $D_{ik} = 0$, or 1.

For any rule $R \in C$ and state $i \in I$, let

$$\phi(i,R) = \lim_{T\to\infty} \sup\ (T+1)^{-1} \sum_{t=0}^{T} \sum_{j,k} P_R(y_t = j,\ \Delta_t = k | y_0 = i) w_{jk}$$

where $P_R(y_t = j,\ \Delta_t = k \mid y_0 = i)$ denotes the probability of being at time t in state j and then making decision k when starting in i and using policy R. The quantity $\phi(i,R)$ represents the expected average cost per unit time when the initial state is i and rule R is used.

We say that a rule $R^* \in C$ is optimal with respect to the average cost criterion if $\phi(i,R^*) \leq \phi(i,R)$ for all $R \in C$ and all $i \in I$.

For any rule $R \in C$ and state $i \in I$ and $0 < \alpha < 1$, let

$$\psi(i,\alpha, R) = \sum_{t=0}^{\infty} \alpha^t \sum_{j,k} P_R(y_t = j,\ \Delta_t = k | y_0 = i) w_{jk}.$$

$\psi(i,\alpha,R)$ represents the expected total discounted cost with discountfactor α when the initial state is i and rule R is used.

We say that a rule $R^* \in C$ is optimal with respect to the discounted cost criterion with discountfactor α if $\psi(i,\alpha,R^*) \leq \psi(i,\alpha,R)$ for all $R \in C$ and all $i \in I$.

When I is finite for each of the criteria there always exists an optimal nonramdomized stationary policy. For proofs and references we refer to the book Finite State Markovian Decision Processes by C. Derman [7]. When I is denumerable Blackwell [1] proved under the assumption of bounded cost function that there always exists an optimal nonrandomized stationary rule with respect to the discounted cost criterion

for each discountfactor $0 < \alpha < 1$. If the boundedness condition on w_{ik} is weakened an optimal rule may not exist [10]. An optimal rule with respect to the average cost criterion does not always exists when I is denumerable. To our knowledge the first counterexample is due to Maitra; it can be found in [5]. A striking counterexample was given by Fisher and Ross [9]. In this example the resulting Markov chain $\{y_t\}$ is positive recurrent for each $R \in C^S$, i.e. all states belong to one communicating class and are positive recurrent (see [3]). Also there is an element of C^M which is an optimal policy, but an optimal stationary rule does not exist.

Several authors have stated sufficient conditions for the existence of an optimal nonrandomized stationary policy with respect to the average cost criterion in denumerable state Markovian desicion processes. Derman [5] proved that a sufficient condition is the existence of a bounded solution $\{g,v_j\}$, $j \in I$ of the functional equation

$$(1) \qquad g+v_i = \min_{k\in K(i)} \{w_{ik} + \sum_{j\in I} q_{ij}(k)v_j\}, \; i \in I.$$

Derman's paper [5] in conjunction with a later joint paper [8] of Derman and Veinott show that the following two conditions together ensure the existence of a bounded solution of (1):

(i) For each $R \in C^S$ the resulting Markov chain is positive recurrent

(ii) There exists some state (say 0) and a constant $T < \infty$ such that $M_{i0}(R) < T$ for all i and all $R \in C^S$ where $M_{i0}(R)$ denotes mean recurrence time from state i to state 0 when using rule R.

Ross [13,14] proved that the following weaker condition is also sufficient for the existence of a bounded solution of (1): There exists a sequence $\{\alpha_r\}_{r=1}^{\infty}$ of discountfactors with $\alpha_r \to 1^-$ as $r \to \infty$ and a constant $N < \infty$ such that $|\psi(i,\alpha_r) - \psi(j,\alpha_r)| < N$ for all $r = 1,2, \ldots$, and all $i,j \in I$ where $\psi(i,\alpha_r)$ denotes the minimal expected discounted cost with discountfactor α_r.

In [6] Derman noted that in all likelihood, a better approach to the existence question would avoid equation (1). It is the purpose of this paper to make a first step in this direction. In section 2 we propose conditions A and B and under this conditions it is proved that each limitpoint of discounted-optimal policies (see definition 2) is optimal with respect to the average cost. Since condition A is not easy to verify, we state a more easily verifiable set of conditions (C and D) implying the original conditions A and B. In section 3 we discuss a simple infinite period inventory model to show the applicability of the conditions C and D.

As to the question of the generalization of the results of section 2 it can be said that the conditions C and D can be generalized to arbitrary state spaces directly. It seems that in order to generalize the theorems of section 2 we have to impose continuity conditions on the cost function and the transition probabilities. We have not as yet investigated this further.

2. SUFFICIENT CONDITIONS FOR OPTIMALITY.

DEFINITION 1. Following Derman [4] we say that for policies $R_n, R \in C^S$, $R_n \to R$ as $n \to \infty$ if $D_{ik}(R_n) \to D_{ik}(R)$ as $n \to \infty$ for all i, k.

NOTATION 1. For $R \in C^S$ let $P_{ij}(R)$ be $\sum_{k \in K(i)} q_{ij}(k) D_{ik}(R)$. $P_{ij}(R)$ denotes the matrix of transition probabilities of the resulting Markov chain when nonrandomized stationary policy R is followed. Let $\pi_{ij}(R)$ be $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} P_{ij}^t(R)$. This limit always exists (see [3]).

LEMMA 1. If $\lim_{n\to\infty} R_n = R$ then $\lim_{n\to\infty} P_{ij}(R_n) = P_{ij}(R)$ for all i,j. If I is finite then moreover $\lim_{n\to\infty} \pi_{ij}(R_n) = \pi_{ij}(R)$ for all i,j.

PROOF. Because $P_{ij}(R) = \sum_{k \in K(i)} q_{ij}(k) D_{ik}(R)$ for all $R \in C^S$ the first statement is an immediate consequence of $\lim_{n\to\infty} D_{ik}(R_n) = D_{ik}(R)$. For $R \in C^S$ it holds that $D_{ik}(R) = 0$, or 1. So if $\lim_{n\to\infty} D_{ik}(R_n) = D_{ik}(R)$ then there exists integer n(i) such that $D_{ik}(R_n) = D_{ik}(R)$ as soon as $n > n(i)$. If I is finite then $\sup_{i \in I} n(i)$ is finite and it follows that as soon as $n > \sup_{i \in I} n(i)$, $P_{ij}(R_n) = P_{ij}(R)$ for all i,j. Consequently $\pi_{ij}(R_n) = \pi_{ij}(R)$ for all i,j as soon as $n > \sup_{i \in I} n(i)$. □

CONDITION A. If $\lim_{n\to\infty} R_n = R$ then $\lim_{n\to\infty} \pi_{ij}(R_n) = \pi_{ij}(R)$ for all i,j.

It follows from lemma 1 that condition A holds when I is finite. When I is denumerable condition A is not always satisfied even if we add the condition that for all $R \in C^S$ the resulting Markov chain is positive recurrent. To show this we give the following counterexample which is constructed by simplifying a counterexample of Fisher and Ross [9].

COUNTEREXAMPLE. Let the state space be the nonnegative integers 0,1, ... , and suppose there are two decisions 1 and 2. The transition probabilities are given as follows:

$q_{0i}(1) = q_{0i}(2) = 3.4^{-i}$ for i = 1,2, ...

$q_{i0}(1) = q_{i(i+1)}(1) = 2^{-1}$ for i = 1,2, ...

$q_{i0}(2) = 1 - q_{ii}(2) = 2^{-i}$ for i = 1,2, ...

$M_{i0}(R)$ with $R \in C^S$ denotes the mean recurrence time from state i to state 0 when using rule R. If $R \in C^S$ is such that $D_{i2}=1$ then $M_{i0}(R)=\sum_{t=1}^{\infty} t\, 2^{-i}(1-2^{-i})^{t-1} = 2^i$.

Let R_n denote the rule with $D_{i1} = 1$ for $i \leq n-1$ and $D_{i2} = 1$ for $i \geq n$. And let R_∞ denote the rule with $D_{i1} = 1$ for all i. Then $\lim_{n\to\infty} R_n = R_\infty$. For $i \leq n-1$ it follows that

$$M_{i0}(R_n) = \sum_{t=1}^{n-i} t\, 2^{-t} + 2^{-(n-i)} \{(n-i) + M_{n0}(R_n)\}$$
$$= 2 - 2.2^{-(n-i)} + 2^i.$$

Therefore,

$$M_{00}(R_n) = 1 + \sum_{i=1}^{\infty} 3.4^{-i} M_{i0}(R_n)$$
$$= 1 + \sum_{i=1}^{n-1} 3.4^{-i} \{2-2.2^{-(n-i)} + 2^i\} + \sum_{i=n}^{\infty} 3.4^{-i} . 2^i$$
$$= 1 + \sum_{i=1}^{n-1} 6.4^{-i} - 2^{-n} \sum_{i=1}^{n-1} 6.2^{-i} + \sum_{i=1}^{\infty} 3.2^{-i}$$
$$= 4 + \sum_{i=1}^{n-1} 6.4^{-i} - 2^{-n} \sum_{i=1}^{n-1} 6.2^{-i} \text{ and it follows that}$$

$$\lim_{n\to\infty} M_{00}(R_n) = 6.$$

However,

$$M_{00}(R_\infty) = 1 + \sum_{i=1}^{\infty} 3.4^{-i} M_{i0}(R_\infty)$$
$$= 1 + \sum_{i=1}^{\infty} 3.4^{-i}.2 = 3. \text{ As } \pi_{00}(R) = (M_{00}(R))^{-1} \text{ for}$$

all $R \in C^S$ (see [3]), it follows that $\lim_{n\to\infty} \pi_{00}(R_n) \neq \pi_{00}(R_\infty)$.

CONDITION B. For all $R \in C^S$, $\sum_{j\in I} \pi_{ij}(R) = 1$ for all $i \in I$.

We shall prove that the conditions A and B together are sufficient for the existence of an optimal nonrandomized stationary rule with respect to the average cost criterion. To do this we need some further notation and some lemmas.

NOTATION 2. Let $w_i(R)$ denote the cost incurred in state i when using rule $R \in C^S$, i.e. $w_i(R) = \sum_{k\in K(i)} D_{ik}(R) w_{ik}$.

LEMMA 2. If condition B holds then $\phi(i,R) = \sum_{j\in I} \pi_{ij}(R) w_j(R)$ and $\phi(i,R) = (1-\alpha) \sum_{j\in I} \pi_{ij}(R) \psi(j,\alpha,R)$ for all i, all $R \in C^S$, and all $0 < \alpha < 1$.

PROOF. Because $\sum_{j\in I} \pi_{ij}(R) = 1$ for $R \in C^S$, it follows from a theorem of Scheffé [15] that $\lim_{T\to\infty} (T+1)^{-1} \sum_{t=0}^{T} \sum_{j\in E} P_{ij}^t(R) = \sum_{j\in E} \pi_{ij}(R)$ uniformly for all subsets $E \subset I$. Since w_{ik} is bounded implies $w_i(R)$ is bounded, it follows that

$$\phi(i,R) = \lim_{T\to\infty} (T+1)^{-1} \sum_{t=0}^{T} \sum_{j\in I} P_R(y_t = j | y_0 = i) w_j(R)$$

$$= \sum_{j\in I} \pi_{ij}(R) w_j(R).$$

Note that ϕ defined as a lim sup can be written as a limit.

Since $\sum_{j\in I} \pi_{ij}(R)\, P_{jm}^t(R) = \pi_{im}(R)$ for all $t = 1,2, \ldots$ (see [3]), we find by using the dominated convergence theorem:

$$(1-\alpha) \sum_{j\in I} \pi_{ij}(R)\psi(j,\alpha,R) = (1-\alpha) \sum_{j\in I} \pi_{ij}(R) \sum_{t=0}^{\infty} \alpha^t \sum_{m\in I} P_{jm}^t(R) w_m(R)$$

$$= (1-\alpha) \sum_{t=0}^{\infty} \alpha^t \sum_{m\in I} \sum_{j\in I} \pi_{ij}(R) P_{jm}^t(R) w_m(R)$$

$$= (1-\alpha) \sum_{j=0}^{\infty} \alpha^t \phi(i,R)$$

$$= \phi(i,R). \quad \square$$

LEMMA 3. For all i, all $R \in C$ it holds that

$$\limsup_{\alpha\to 1^-} (1-\alpha)\psi(i,\alpha,R) \le \phi(i,R).$$

PROOF. The theorem is essentially a Tauberian theorem. For fixed i, let $w_t(R)$ denote $\sum_{j\in I} \sum_{k\in K(j)} P_R(y_t=j, \Delta_t=k \mid y_0=i) w_{jk}$ and let $W_t(R)$ denote $w_0(R) + \ldots + w_t(R)$. $w_t(R)$ is the expected cost at time t and $W_t(R)$ is the cumulative cost until time t when using rule $R \in C$.

Then $\psi(i,\alpha,R) = \sum_{t=0}^{\infty} w_t(R)\alpha^t$ and therefore $(1-\alpha)^{-1}\psi(i,\alpha,R) = \sum_{t=0}^{\infty} w_t(R)\alpha^t \cdot \sum_{t=0}^{\infty} \alpha^t = \sum_{t=0}^{\infty} W_t(R)\alpha^t$.

Choosing $\varepsilon > 0$ arbitrarily small we have since $\limsup_{t\to\infty} (t+1)^{-1} W_t(R) = \phi(i,R)$ that there exists an integer T such that $(t+1)^{-1} W_t(R) \le \phi(i,R) + \varepsilon/2$ for $t > T$.

Since $(1-\alpha)^{-2} = \sum_{t=0}^{\infty} (t+1)\alpha^t$, it follows that

$$(1-\alpha)^{-1}\psi(i,\alpha,R) - (1-\alpha)^{-2}\phi(i,R) = \sum_{t=0}^{T} \{W_t(R) - (t+1)\phi(i,R)\}\alpha^t +$$

$$+ \sum_{t=T+1}^{\infty} \{(t+1)^{-1} W_t(R) - \phi(i,R)\}(t+1)\alpha^t \le$$

$$\leq \max_{t=0,1,\ldots,T} \{W_t(R)-(t+1)\phi(i,R)\}(1-\alpha)^{-1}(1-\alpha^{T+1}) +$$

$$+ \varepsilon/2\,(1-\alpha)^{-2}.$$

$$\Longrightarrow (1-\alpha)\psi(i,\alpha,R)-\phi(i,R) \leq$$

$$\leq \max_{t=0,1,\ldots,T} \{W_t(R)-(t+1)\phi(i,R)\}(1-\alpha)(1-\alpha^{T+1}) +$$

$+ \varepsilon/2 \leq \varepsilon$ for α sufficiently near 1. Hence

$\limsup_{\alpha\to 1^-} (1-\alpha)\psi(i,\alpha,R) \leq \phi(i,R)$. □

NOTATION 3. For all i,all $0<\alpha<1$ let $\psi(i,\alpha)$ denote $\inf_{R\in C} \psi(i,\alpha,R)$.

LEMMA 4. If for a constant g there exists a sequence $\{\alpha_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} \alpha_n = 1^-$ and $\lim_{n\to\infty} (1-\alpha_n)\psi(i,\alpha_n) = g$ for some $i\in I$, then $g \leq \inf_{R\in C} \phi(i,R)$.

PROOF. Since $\psi(i,\alpha_n) \leq \psi(i,\alpha_n,R)$ for all $R \in C$, we have

$$g = \lim_{n\to\infty} (1-\alpha_n)\psi(i,\alpha_n) \leq \limsup_{n\to\infty} (1-\alpha_n)\Psi(i,\alpha_n,R) \leq$$

$\leq \limsup_{\alpha\to 1} (1-\alpha)\psi(i,\alpha,R)$ for all $R \in C$. From lemma 3 it follows then $g \leq \phi(i,R)$ for all $R \in C$. □

DEFINITION 2. We say that R is a limitpoint of discounted-optimal rules if there exist sequences $\{\alpha_n\}_{n=1}^{\infty}$ and $\{R_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \alpha_n = 1^-$ and $R_n \in C^S$ is an optimal policy with respect to the expected discounted cost with discountfactor α_n and moreover it holds that $\lim_{n\to\infty} R_n = R$.

THEOREM 1. If conditions A and B hold then each limitpoint of discounted-optimal rules is an optimal policy with respect to the average cost criterion.

PROOF. Suppose R^* is a limitpoint of discounted-optimal rules, so that there exist sequences $\{\alpha_n\}_{n=1}^{\infty}$ and $\{R_n\}_{n=1}^{\infty}$ as in definition 2.

Since $\psi(i,\alpha_n,R_n) = \sum_{t=0}^{\infty} \alpha_n^t \sum_{j\in I} P_{ij}^t(R_n)w_j(R_n)$, it follows that

$$(2) \quad \psi(i,\alpha_n,R_n) = w_i(R_n) + \alpha_n \sum_{j\in I} P_{ij}(R_n)\psi(j,\alpha_n,R_n).$$

Recalling that w_{ik} is bounded we let M denote an upperbound of $|w_{ik}|$. Then also $|(1-\alpha)\psi(i,\alpha,R)| \leq M$ for all i, all $R \in C$. From this it follows that there exists a subsequence $\{n_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} (1-\alpha_{n_k})\psi(i,\alpha_{n_k},R_{n_k})$ exists for all i. Let us denote the limit by g(i). Then by lemma 4 we have that $g(i) \leq \inf_{R\in C} \phi(i,R)$. We shall prove in the following that actually $g(i) = \phi(i,R^*)$ for all i. This in turn implies that R^* is optimal with respect to the average cost criterion.

From (2) it is easily seen that

$$(1-\alpha_{n_k})\psi(i,\alpha_{n_k},R_{n_k}) = (1-\alpha_{n_k})w_i(R_{n_k}) +$$

$$+ \alpha_{n_k} \sum_{j\in I} P_{ij}(R_{n_k})(1-\alpha_{n_k})\psi(j,\alpha_{n_k},R_{n_k}).$$

By taking limits and using a convergence theorem of Scheffé [15] we find with lemma 1 : $g(i) = \sum_{j\in I} P_{ij}(R^*)g(j)$. Iterating this equality and using condition B we obtain:

(3) $\quad g(i) = \sum_{j\in I} \pi_{ij}(R^*)\, g(j)$ for all i.

Lemma 2 gives $\phi(i,R_{n_k}) = \sum_{j\in I} \pi_{ij}(R_{n_k})\, w_j(R_{n_k})$ for all k. By the uniform boundedness of $w_i(R_{n_k})$ and condition A it follows that

(4) $\quad \lim_{k\to\infty} \phi(i,R_{n_k}) = \phi(i,R^*)$ for all i.

Lemma 2 also gives $\phi(i,R_{n_k}) = \sum_{j\in I} \pi_{ij}(R_{n_k})(1-\alpha_{n_k})\psi(j,\alpha_{n_k},R_{n_k})$.

By using the same arguments we find

(5) $\quad \lim_{k\to\infty} \phi(i,R_{n_k}) = \sum_{j\in I} \pi_{ij}(R^*)\, g(j)$ for all i.

Combining the equations (3), (4) and (5) we find that $\phi(i,R^*) = g(i)$ for all i. □

REMARK 1. Theorem 1 remains true if we weaken conditions A and B as follows: For $0 < \beta < 1$ let C_β be the subset of C^S consisting of those rules which are optimal with respect to the expected discounted cost with discountfactor α, $\beta < \alpha < 1$.

Then the weaker condition is: There exists a constant β_0, $0 < \beta_0 < 1$, such that $\sum_{j\in I} \pi_{ij}(R) = 1$ for $R \in C_{\beta_0}$ and such that if R_n, $R \in C_{\beta_0}$ and $\lim_{n\to\infty} R_n = R$ then $\lim_{n\to\infty} \pi_{ij}(R_n) = \pi_{ij}(R)$ for all i,j.

THEOREM 2. If conditions A and B hold then there exists an optimal nonrandomized stationary policy with respect to the average cost criterion.

PROOF. We only have to prove that there exists a limitpoint of discounted-optimal rules. We therefore need the following result given by Blackwell [1]: If $K(i) < \infty$ and if $|w_{ik}| < M$ for all i,k, then under the α-discounted criterion with $0 < \alpha < 1$ there exists a nonrandomized stationary rule R_α such that $\psi(i,\alpha,R_\alpha) = \inf_{R\in C} \psi(i,\alpha,R)$ for all i.
Now since I is denumerable and $K(i) < \infty$, for an arbitrary sequence $\{\alpha_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} \alpha_n = 1^-$ there exists a subsequence $\{\alpha_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} D_{ik}(R_{\alpha_{n_k}})$ exists for all i,k. □

Conditions A and B though sufficient for the existence of an optimal policy are not easy to verify. Therefore we will give a new set of conditions implying the conditions A and B.

CONDITION C. For each $R \in C^S$ the resulting Markov chain does not have two disjoint closed sets.

Before we state the next condition, we give the following definition.

DEFINITION 3. A collection of probability measures $\mathcal{P}$ on a metric space S is called tight if there exists for any $\varepsilon > 0$ a compact subset $A \subset S$ with the property that $P(A) \geq 1-\varepsilon$ for all $P \in \mathcal{P}$ (see [2]).

In our case the state space I is denumerable so the topology on I will be a discrete one and a collection of probability measures $\mathcal{P}$ on I will be called tight if there exists for any $\varepsilon > 0$ a FINITE subset $A \subset I$ with $P(A) \geq 1-\varepsilon$ for all $P \in \mathcal{P}$.

CONDITION D. The collection of probability measures on I

$$\{q_{i.}(k) \mid i \in I,\ k \in K(i)\} \text{ is tight.}$$

THEOREM 3. The conditions C and D imply the conditions A and B.

PROOF. Condition D states that for any $\varepsilon > 0$ there exists a finite subset $A_\varepsilon \subset I$ with $\sum_{j \in A_\varepsilon} q_{ij}(k) \geq 1-\varepsilon$ for all i, all $k \in K(i)$. Consequently we have $\sum_{j \in A_\varepsilon} P_{ij}(R) \geq 1-\varepsilon$ for all i, all $R \in C^S$. By induction on t it then easily follows that for any t, $\sum_{j \in A_\varepsilon} P^t_{ij}(R) \geq 1-\varepsilon$ for all i, all $R \in C^S$. Because A_ε is a finite set we have that $\sum_{j \in A_\varepsilon} \pi_{ij}(R) = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{j \in A_\varepsilon} P^t_{ij}(R) \geq 1-\varepsilon$ for all i, all $R \in C^S$. We state two conclusions:

i) for any $R \in C^S$, $\sum_j \pi_{ij}(R) = 1$ for all $i \in I$, and so condition B is satisfied.

ii) the collection of probability measures on I

$$\{\pi_{i.}(R) \mid i \in I,\ R \in C^S\} \text{ is tight.}$$

It follows from conditions B and C that $\pi_{ij}(R) = \pi_{jj}(R)$ for all $i,j \in I$, all $R \in C^S$ (see [3]). If $\pi_j(R)$ denote $\pi_{jj}(R)$ then we have from ii):

6) the collection of probability measures $\{\pi_.(R) \mid R \in C^S\}$ is tight.

Let us suppose that $\lim_{n \to \infty} R_n = R^*$. We call $\{\pi_i\}_{i \in I}$ a limit-point if there exists a subsequence of the natural numbers say $\{n_k\}_{k=1}^{\infty}$ with $\lim_{k \to \infty} \pi_i(R_{n_k}) = \pi_i$ for all i. To prove that also condition A is satisfied, we shall show that any limit-point $\{\pi_i\}_{i \in I}$ satisfies $\pi_i = \pi_i(R^*)$ for all i.

Suppose $\{\pi_i\}_{i \in I}$ is limitpoint and

$$(7) \quad \lim_{k \to \infty} \pi_i(R_{n_k}) = \pi_i \text{ for all i.}$$

Tightness implies $\sum_{i \in I} \pi_i = 1$ which can be deduced from a general theorem of Prohorov (see [2]) or alternatively this can

be deduced directly. Consequently, (see [15] and lemma 1)

$$(8)\quad \lim_{k\to\infty} \sum_{j\in I} \pi_j(R_{n_k})P_{ji}(R_{n_k}) = \sum_{j\in I} \pi_j P_{ji}(R^*).$$

The conditions B and C together (see [3]) imply that for $R \in C^S$, $\{\pi_i(R)\}_{i\in I}$ is the unique solution of the equations in $\{u_i\}_{i\in I}$

$$(9)\quad \begin{cases} u_i = \sum_{j\in I} u_j P_{ji}(R) \\ 1 = \sum_{j\in I} u_j \end{cases}$$

Combining (7), (8) and (9) we find that $\pi_i = \sum_{j\in I} \pi_j P_{ji}(R^*)$. Since we already found that $\sum_{j\in I} \pi_j = 1$, it follows from (9) that $\pi_i = \pi_i(R^*)$ for all i. □

3. AN INFINITE PERIOD STATIONARY INVENTORY MODEL WITH BACKLOGGING.

Let y_t denote the level of inventory at time t and let Δ_t be the amount ordered after observing y_t. Assume delivery of the Δ_t units is instantaneous so that at the moment of ordering, the inventory level is $y_t + \Delta_t$. Suppose the sequence of demands $\{D_t\}$ for the product during each of the periods is a sequence of independent and identically distributed random variables, say $P\{D_t = j\} = p_j$, $j = 0,1,\ldots$ and $\sum_{j=0}^{\infty} p_j = 1$. We allow negative inventory, that is, backlogging of demand, and suppose a denumerable state space. Then:

$$q_{ij}(k) = P\{y_{t+1} = j \mid y_t = i, \Delta_t = k\}$$

$$= P\{\text{Demand} = i+k-j\}$$

$$= p_{i+k-j} \quad \text{for } i+k \geq j, \text{ 0 otherwise.}$$

CONDITION E. There exist integers C_1 and C_2, such that the set of ordering decisions in state $i \leq C_2$ is given by

$$K(i) = \{k \mid C_1 \leq i+k \leq C_2\}.$$

As a consequence of theorems 1 and 3 we have the following theorem.

THEOREM 4. If $p_j > 0$ for $j = 0,1,\ldots$ and condition E is satisfied it follows that each limitpoint of discounted-optimal rules is an optimal policy with respect to the average cost criterion.

PROOF. Since $q_{iC_1}(k) = p_{i+k-C_1}$ for all $i \leq C_2$, all $k \epsilon K(i)$, we see that state C_1 can be reached from each state and under each policy. So it follows that for any $R \in C^S$ there do not exist two disjoint closed sets and so condition C is satisfied.

To show that condition D holds we choose $\varepsilon > 0$ arbitrarily. Let $K < \infty$ be such that $\sum_{j=0}^{K+C_1} p_j \geq 1-\varepsilon$, it then follows that $\sum_{j=-K}^{C_2} q_{ij}(k) = \sum_{j=0}^{i+k+K} p_j \geq 1-\varepsilon$ for all $i \leq C_2$, all $k \epsilon K(i)$. □

A nonrandomized stationary rule which prescribes no ordering in state i when $i \geq s$ and prescribes an order of $S - i$ units when $i < s$ is called an (s,S) policy. Under certain conditions on the cost function it can be proved that there exist optimal (s,S) policies with respect to the expected discounted cost (see for instance [11], [12] and [16]).
As a consequence of theorem 4 we state the following:

COROLLARY. If the cost function is such that there exist optimal (s,S) policies with respect to the expected discounted cost criteria, then there exists an optimal (s,S) policy with respect to the average cost criterion when it is assumed that $p_j > 0$ for $j = 0,1,\ldots$ and it is assumed that condition E holds.

REFERENCES.

[1] D. Blackwell: Discounted dynamic programming. Ann. Math. Statist. 36(1965), 226-235.

[2] P. Billingsley: Convergence of probability measures. Wiley, New York 1968.

[3] K.L. Chung: Markov chains with stationary transition probabilities. Springer, Berlin 1960.

[4] C. Derman: Markovian sequential control processes-denumerable state space. J. Math. Anal. Appl. 10(1965), 295-302.

[5] C. Derman: Denumerable state Markovian decision processes-average cost criterion. Ann.Math. Statist. 37 (1966), 1545-1554.

[6] C. Derman: Markovian decision processes-average cost criterion. In: Mathematics of the decision sciences, 1968 American Mathematical Society, part 2, 139-148.

[7] C. Derman: Finite state Markovian decision processes. Academic Press, New York 1970.

[8] C. Derman and A. Veinott: A solution to a countable system of equations arising in Markovian decision processes. Ann. Math. Statist. 38 (1967), 582-585.

[9] L. Fisher and S. Ross: An example in denumerable decision processes. Ann. Math. Statist. 39(1968), 674-675.

[10] A. Hordijk and H.C. Tijms: A counterexample in discounted dynamic programming. Report BW 7/71(1971), Mathematisch Centrum, Amsterdam (to appear in J. Math. Anal. Appl.).

[11] D.L. Iglehart: Optimality of (s,S) inventory policies in the infinite horizon dynamic inventory problem. Management Sci. 9 (1963), 259-267.

[12] E.L. Johnson: On (s,S) policies, Management Sci. 15 (1968), 80-101.

[13] S. Ross: Non-discounted denumerable Markovian decision models. Ann. Math. Statist. 39 (1968), 412-423.

[14] S. Ross: Applied probability models with optimization applications. Holden-Day, San Francisco 1970.

[15] H. Scheffé: A useful convergence theorem for probability distributions. Ann. Math. Statist. 18 (1947), 434-438.

[16] H.C. Tijms: The optimality of (s,S) inventory policies in the infinite period model. Statistica Neerlandica 25 (1971), 29-43.