The Annals of Statistics
1976, Vol. 4, No. 5, 936-953

CONDITIONS FOR THE EQUIVALENCE OF OPTIMALITY CRITERIA IN DYNAMIC PROGRAMMING

BY JAMES FLYNN

University of Chicago

This paper examines the relationships between optimality criteria which are commonly used for undiscounted, discrete-time, countable state Markovian decision models. One approach, due to Blackwell, is to maximize the expected discounted total return as the discount factor approaches 1. Another, due to Veinott, is to maximize the Cesaro means of the finite horizon expected returns as the horizon tends to infinity. Derman's is to maximize the long-run average gain. Denardo, Miller and Lippman showed that Blackwell's and Veinott's approaches are equivalent for finite state and action spaces. As shown here, that equivalence breaks down when the state space is countable. Also, policies optimal according to Blackwell's or Veinott's approach need not be optimal according to Derman's. On the positive side, fairly weak conditions are given under which Blackwell's and Veinott's criteria imply Derman's, and somewhat stronger conditions under which Blackwell's and Veinott's criteria are equivalent.

1. Introduction. Our formulation of the Markovian decision model follows Blackwell's (1965). Consider a system with a countable *state space* S and a finite *action space* A. Each day the current state $s \in S$ is observed and an action $a \in A$ is selected. This results in (1) an *immediate income* $i(s, a)$ and (2) a transition to a new state s' with probability $q(s' \mid s, a)$. The incomes are assumed bounded. The problem is to control the system in the most effective manner over an infinite future. A rule or *policy* π for controlling the system specifies for each $n \geqq 1$ what act to choose on the nth day as a function of the system's current history $h = (s_1, a_1, \cdots, s_n)$ or, more generally, π specifies for each h a probability distribution on A. A (nonrandomized) *stationary* policy is a policy which is specified by a single function f mapping S into A: under it, you select act $f(s)$ whenever the system is in state s. Given an initial state s and a policy π, let $r_j(s, \pi)$ denote the *expected return* on the jth day ($j = 1, 2, \cdots$); then the expected *N-stage return* is $V^N(s, \pi) = \sum_{j=1}^{N} r_j(s, \pi)$, the *average gain* in the first N periods is $V^N(s, \pi)/N$, and the expected *discounted total return* is $V_\beta(s, \pi) = \sum_{j=1}^{\infty} \beta^{j-1} r_j(s, \pi)$ where $0 \leqq \beta < 1$ is the discount factor.

For the discounted problem, there is only one reasonable criterion: a policy π_* is said to be *β-optimal* if

$$V_\beta(s, \pi_*) = V_\beta^*(s) \equiv \sup_\pi V_\beta(s, \pi), \qquad s \in S. \tag{1}$$

Received June 1974; revised February 1976.

AMS 1970 *subject classifications.* Primary 49C15, 62L99, 90C40, 93C55; Secondary 60J10, 60J20.

Key words and phrases. Dynamic programming, Markovian decision process, optimality criteria, average overtaking criteria, average gain, discounting, small interest rates.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

Blackwell (1965) and Maitra (1965) established the existence of a β-optimal stationary policy for each $0 \leqq \beta < 1$.

There are, however, a number of competing criteria for the undiscounted problem. Derman (1964), (1966), and Ross (1968) used criteria which depend on the average gain:

DEFINITION 1. A policy π_* is *lim sup average optimal* if

$$\limsup_n V^n(s, \pi_*)/n \geqq \limsup_n V^n(s, \pi)/n\,, \qquad s \in S \tag{2}$$

for every policy π; it is *lim inf average optimal* if

$$\liminf_n V^n(s, \pi_*)/n \geqq \liminf_n V^n(s, \pi)/n\,, \qquad s \in S \tag{3}$$

for every π; and it is *average optimal* if

$$\liminf_n (V^n(s, \pi_*) - V^n(s, \pi))/n \geqq 0\,, \qquad s \in S \tag{4}$$

for every π. Given any $\varepsilon > 0$, the policy π_* is *ε-average optimal* if

$$\liminf_n (V^n(s, \pi_*) - V^n(s, \pi))/n \geqq -\varepsilon\,, \qquad s \in S \tag{5}$$

for every π.

Unfortunately, those criteria are underselective since they depend only on the tail of the returns and not on the returns during the first millenium. One can, of course, get around this with the criterion of selecting a π_* such that $\liminf_N (V^N(s, \pi_*) - V^N(s, \pi)) \geqq 0$ for all $s \in S$ and all π. However, the latter is overselective since even when S is finite, there need not exist any policy satisfying it (see Denardo and Miller (1968)). One can circumvent both problems by using the following criterion, which is due to Veinott (1966):

DEFINITION 2. A policy π_* is *average overtaking optimal* if

$$\liminf_N \sum_{n=1}^{N} (V^n(s, \pi_*) - V^n(s, \pi))/N \geqq 0\,, \qquad s \in S \tag{6}$$

for every policy π.

Another way of approaching the undiscounted problem is to study the discounted problem for the case of *small interest rates* (i.e., values of β close to 1). The following criteria are due to Blackwell (1962) (see Remark 2 at the end of this section).

DEFINITION 3. A policy π_* is *optimal* if there exists a $\beta_0 \in (0, 1)$ such that

$$V_\beta(s, \pi_*) \geqq V_\beta(s, \pi)\,, \qquad s \in S,\ \beta \in (\beta_0, 1) \tag{7}$$

for every policy π.

DEFINITION 4. A policy π_* is *1-optimal* if

$$\liminf_{\beta \to 1-} (V_\beta(s, \pi_*) - V_\beta(s, \pi)) \geqq 0\,, \qquad s \in S \tag{8}$$

for every policy π.

Optimality certainly implies 1-optimality. Using an Abelian argument (see

Hobson (1926)), one can establish the following general result:

$$\text{(9)} \qquad \begin{matrix} \text{optimality} & \searrow & \\ & & \text{1-optimality.} \\ \text{average overtaking optimality} & \nearrow & \end{matrix}$$

It is also clear that

$$\text{(10)} \qquad \text{average optimality} \begin{matrix} \nearrow \text{ lim sup average optimality} \\ \searrow \text{ lim inf average optimality.} \end{matrix}$$

One is tempted to link (9) and (10) by the statement: "1-optimality implies average optimality." Surprisingly, this is not the case. In Flynn (1974), we give an example where an optimal policy is not lim inf average optimal. (We also show that this "pathology" cannot occur when S is finite.) Here we give examples of policies which are both optimal and average overtaking optimal but are not lim inf average optimal and lim sup average optimal, respectively (see Section 5, Examples 3 and 4). We also investigate:

QUESTION 1. When does

$$\text{(11)} \qquad \text{1-optimality} \Rightarrow \text{average optimality?}$$

Denardo and Miller (1968) and Lippman (1969) established the following:

THEOREM 1. *If S is finite, then average overtaking optimality and 1-optimality are equivalent.*

This paper contains an example which shows that optimality need not imply average overtaking optimality (see Section 5, Example 1). Hence, Theorem 1 does not extend to countable S. Also, Blackwell (1962) gave an example (with S finite) which indicates that neither 1-optimality nor average overtaking optimality imply optimality. It follows that (9) cannot be strengthened without imposing some conditions.

QUESTION 2. When is it true that

$$\text{(12)} \qquad \text{average overtaking optimality} \Leftrightarrow \text{1-optimality?}$$

The following related—but different (see Ross (1971))—questions are of interest:

QUESTION 1′. When does 1-optimality among stationary policies imply average optimality among stationary policies?

QUESTION 2′. When is average overtaking optimality among stationary policies equivalent to 1-optimality among stationary policies?

To give this second group of questions proper meaning, we require:

DEFINITION 5. Let Ω be a class of policies and let $\varepsilon > 0$ be a real number. A policy π_* is *average overtaking optimal*, *optimal*, *1-optimal*, *ε-average optimal*, or *average optimal among π in Ω* if (6), (7), (8), (5) or (4), respectively, holds for

all $\pi \in \Omega$. When Ω denotes the class of stationary policies, replace the phrase "among π in Ω" by "among stationary policies."

REMARK 1. One can easily show (9) holds if we replace each type of optimality by the corresponding type of optimality "among $\pi \in \Omega$" where Ω is arbitrary. Also, Theorem 1 remains valid if we replace each type by the corresponding type "among stationary policies."

REMARK 2. Our definition of "1-optimality" (see Definition 4) is equivalent to Veinott's (1969) definition of "0-discount optimality" but differs slightly from his definition of "1-optimality." Veinott (1966) calls π_* "1-optimal" if it satisfies

$$\lim_{\beta \to 1-} (V_\beta(s, \pi_*) - V_\beta^*(s)) = 0, \qquad s \in S. \tag{13}$$

(This is the same as Blackwell's (1962) definition of "nearly optimal.") Our definitions are equivalent when S is finite since there always exists an optimal policy in that case (see Blackwell (1962)). However, for general S, average overtaking optimality would not imply 1-optimality if 1-optimality were defined via (13) (see Section 5, Example 7).

2. Outline and discussion. Our results on Questions 1 and 1′ appear in Section 3, while those on Questions 2 and 2′ appear in Section 4. All of our examples are in Section 5. The proofs of some of our technical results are relegated to Section 6. In this section, we focus on some of our more striking results.

Our answer to Question 1 is fairly complete. In Corollary 2 we show that the following condition is sufficient for (11): For every $\varepsilon > 0$, there exists a policy which is ε-average optimal. (This result is fairly useful: see Remark 3 below.) Surprisingly, the analogous result does not hold for Question 2. In Example 2, we construct a model in which there exists an average overtaking optimal policy and a 1-optimal policy which is not average overtaking optimal. Apparently, the relationship between 1-optimality and average overtaking optimality is not all that strong. It would be interesting to answer:

QUESTION 3. Is the existence of an average overtaking optimal policy sufficient for optimality to imply average overtaking optimality?

REMARK 3. By Corollary 2, the existence of an average optimal policy suffices for 1-optimality to imply average optimality. Now, many authors (e.g., Derman (1964), (1966), Derman and Veinott (1967), and Ross (1968), (1971)) have investigated the question of the existence of a stationary lim inf average optimal policy; for the most part, their results apply directly to the problem of the existence of an average optimal policy (see Theorem 2 below). That condition (iv) of Theorem 2 would be enough to eliminate the "pathology" described in Flynn (1974) was originally conjectured by Bennett Fox.

THEOREM 2. *There exists a stationary policy which is average optimal in each of the following cases*:

(i) *S is finite.*

(ii) *There exists a bounded set of numbers* $\{g, f(s)\}$, $s \in S$, *satisfying*

$$g + f(s) = \min_{a \in A} \{i(a, s) + \sum_{s' \in S} q(s' \mid s, a) f(s')\} \qquad s \in S . \tag{14}$$

In this case, any stationary policy π_* *which, for each* s, *selects the action which minimizes the* RHS *of* (14) *is average optimal.*

(iii) $V_\beta^*(s') - V_\beta^*(s'')$ *is bounded uniformly in* β, s' *and* s''.

(iv) *There exists a family* $\{\pi^\beta, 0 \leqq \beta < 1\}$ *of* β*-optimal stationary policies and a state* s_0 *with the property: under the Markov chain associated with any* π^β, *the system eventually reaches* s_0 *with probability* 1; *moreover, the mean recurrence time from* $s \in S$ *to* s_0 *under* π_β *is bounded uniformly in* β *and* s.

PROOF. The sufficiency of (i) follows from Theorem 4.2 of Brown (1965); that of (ii) follows from Remark 1 of Ross (1968), who also showed that both (iii) and (iv) imply (ii).

In our investigation of Questions 1′ and 2′, we determine conditions under which the following are true (see Remark 4 below):

STATEMENT 1. Any stationary policy π_* which is 1-optimal among stationary policies is average optimal among stationary policies.

STATEMENT 2. Any stationary policy π_* which is 1-optimal among stationary policies is average overtaking optimal among stationary policies.

In Theorem 4, we show that the following condition implies Statement 1: (EPR) *Under the Markov chain associated with each stationary policy the system eventually reaches a positive recurrent state with probability* 1. Note that (EPR) holds when S is finite (Chung (1967)).

In Theorem 5, we find conditions which imply Statement 2. In order to give an example where these conditions apply, we introduce the notion of *uniform recurrence*. Let C be an indecomposable class of states in a Markov chain with n-step transition matrix $P^{(n)} = (P_{ij}^{(n)})$, $(i, j, n = 1, 2, \cdots)$. We say that C is *uniformly recurrent* if C is closed and if for some $j \in C$ there exists a $\delta > 0$ and a positive integer n such that $\sum_{k=1}^{n} P_{ij}^{(k)} \geqq \delta$ for all $i \in C$. One can show that a uniformly recurrent class of states consists of a nonempty set of positive recurrent states and a (possibly empty) set of transient states. The expected time spent in the transient states is bounded. (Uniform recurrence is equivalent to the notion of uniform ϕ-recurrence which appears in Orey (1971) when ϕ is the measure which assigns unit mass to j and zero mass to every other state.) Theorem 6 states that the following condition implies Statement 2: (UR) *Under the Markov chain associated with each stationary policy, the state space can be expressed as a union of uniformly recurrent classes.* (Observe that the classes need not be disjoint.) Note that (UR) does not include the case where S is finite. (Statement 1 is, of course, always true in that case: see Remark 1.).

REMARK 4. If a (possibly nonstationary) policy π_* is optimal among stationary policies, then under (EPR) it is average optimal among stationary policies

while under (UR) it is average overtaking optimal among stationary policies (see Corollaries 3 and 4). It would be interesting to answer:

QUESTION 4. Under (EPR), need a (possibly nonstationary) 1-optimal policy be average optimal among stationary policies?

Note that the answer to the analogous question for average overtaking optimal policies is "no" (see Example 8).

REMARK 5. None of the positive results in Sections 3 and 4 require that A be finite; however, some of them fail if we relax the condition that income be bounded (see Examples 5 and 6).

3. Questions 1 and 1′. This section deals with the relationship between 1-optimality and average optimality. Corollary 2 summarizes our main result on Question 1, while Theorem 4 and Corollary 3 summarize our main results on Question 1′. The next result is our key.

THEOREM 3. *Let Ω be any class of policies such that for each positive ε, Ω contains a policy which is ε-average optimal among π in Ω. Then any $\pi_* \in \Omega$ which is 1-optimal among π in Ω must also be average optimal among π in Ω.*

The following corollaries are immediate consequences of Theorem 3.

COROLLARY 1. *If $\lim_n V^n(s, \pi)/n$ exists for all $s \in S$ and $\pi \in \Omega$, then any $\pi_* \in \Omega$ which is 1-optimal among π in Ω must also be average optimal among π in Ω.*

COROLLARY 2. *If for every $\varepsilon > 0$ there exists a policy which is ε-average optimal, then 1-optimality implies average optimality.*

REMARK 6. By Corollary 1, the following convergence condition implies Statement 1: (C) $\lim_n V^n(s, \pi)/n$ *exists for all* $s \in S$ *and all stationary* π. Note that (C) does not imply the existence of stationary ε-average optimal policies (see Ross (1971)). In Lemma 2 below we show that (EPR) implies (C), hence the following:

THEOREM 4. *Condition* (EPR) *implies Statement* 1.

We shall show that Theorem 4 implies:

COROLLARY 3. *Under* (EPR), *if π_* is optimal among stationary policies, then it is average optimal among stationary policies.*

We do not know whether the analogue of Corollary 3 holds for 1-optimal π_* (see Remark 4 above). Note that both Theorem 4 and Corollary 3 fail if we relax the boundedness condition on $i(\cdot, \cdot)$ (see Example 5).

Theorem 3 requires the following lemma, the proof of which is relegated to Section 6.

LEMMA 1. *For every $M > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\liminf_{\beta \to 1-} (1 - \beta) \sum_{j=1}^{\infty} \beta^{j-1} a_j - \liminf_n \sum_{j=1}^{n} a_j/n < \varepsilon \tag{15}$$

for any sequence $\{a_j\}_{j=1}^{\infty}$ *satisfying*

$$|a_j| \leqq M, \qquad j = 1, 2, \cdots, \tag{16}$$

and

$$\limsup_n \sum_{j=1}^{n} a_j/n - \liminf_{\beta \to 1-} (1-\beta) \sum_{j=1}^{\infty} \beta^{j-1} a_j < \delta . \tag{17}$$

PROOF OF THEOREM 3. We can assume that there exists a policy π_* which is 1-optimal among π in Ω. By hypothesis, there exist policies $\pi_1, \pi_2, \cdots$ satisfying

$$\liminf_n \sum_{j=1}^{n} [r_j(s, \pi_k) - r_j(s, \pi)]/n \geqq -1/k \tag{18}$$

for every $\pi \in \Omega$, $(k = 1, 2, \cdots)$. In particular,

$$\limsup_n \sum_{j=1}^{n} [r_j(s, \pi_*) - r_j(s, \pi_k)]/n \leqq 1/k , \qquad k = 1, 2, \cdots . \tag{19}$$

The fact that π_* is 1-optimal among π in Ω implies

$$\liminf_{\beta \to 1-} (1-\beta) \sum_{j=1}^{\infty} \beta^{j-1} (r_j(s, \pi_*) - r_j(s, \pi)) \geqq 0 \tag{20}$$

for every $\pi \in \Omega$. By (19) and (20), the fact that $i(\cdot, \cdot)$ is bounded, and Lemma 1, there exists a subsequence $\{\pi_{m_k}\}_{k=1}^{\infty}$ of policies satisfying

$$\liminf_n \sum_{j=1}^{n} (r_j(s, \pi_*) - r_j(s, \pi_{m_k}))/n \geqq -1/k , \qquad k = 1, 2, \cdots . \tag{21}$$

Thus, (18) and (21) imply that

$$\begin{aligned} \liminf_n \sum_{j=1}^{n} (r_j(s, \pi_*) - r_j(s, \pi))/n &\geqq \liminf_n \sum_{j=1}^{n} (r_j(s, \pi_*) - r_j(s, \pi_{m_k}))/n \\ &\quad + \liminf_n \sum_{j=1}^{n} (r_j(s, \pi_{m_k}) - r_j(s, \pi))/n \\ &\geqq -1/k - 1/m_k \end{aligned}$$

for every policy $\pi \in \Omega$, $(k = 1, 2, \cdots)$. Since k is arbitrary, (4) holds. The theorem follows.

Theorem 4 requires the following lemma (see Remark 6).

LEMMA 2. *Condition* (EPR) *implies condition* (C).

PROOF. Let the random variables $X_1, X_2, \cdots$ represent the Markov chain associated with a given stationary policy π and initial state $s_0 \in S$. Define the function f on S by $f(s) = i(s, \pi(s))$, $(s \in S)$. Observe that f is bounded. This and condition (EPR) imply that $(f(X_1) + \cdots + f(X_n))/n$ converges almost everywhere to a bounded random variable R (see Chung (1967): 1.15, Theorem 2). By the bounded convergence theorem, $(Ef(X_1) + \cdots + Ef(X_n))/n$ converges to ER. To finish the proof of the lemma, note that $Ef(X_n) = r_n(s, \pi)$, $(n = 1, 2, \cdots)$.

PROOF OF COROLLARY 3. Let π_* be optimal among stationary policies. Since there exist β-optimal stationary policies for $0 \leqq \beta < 1$ (see Blackwell (1965)), π_* must be optimal. Using this together with the arguments in the proof of Theorem 8.3 of Strauch (1966), one can show there exists an optimal stationary policy π_{**}; moreover,

$$r_n(s, \pi_*) = r_n(s, \pi_{**}) , \qquad s \in S;\ n = 1, 2, \cdots . \tag{22}$$

Corollary 3 follows easily from (22) and Theorem 4.

4. Question 2 and 2′. This section deals with the relationship between 1-optimality and average overtaking optimality. Our positive results on Questions 2 and 2′ are less impressive than those on Questions 1 and 1′; in particular, the analogues of Theorem 3 and Corollary 2 are false: the existence of an average overtaking optimal policy does not imply any equivalence between 1-optimality and average overtaking optimality (see Example 2). However, we do have the following analogue of Corollary 1.

THEOREM 5. *Let Ω be a class of policies such that $g(s, \pi) \equiv \lim_n V^n(s, \pi)/n$ and $b(s, \pi) \equiv \lim \sum_{j=1}^N (V^j(s, \pi) - jg(s, \pi))/N$ exist and are finite for all $s \in S$ and all $\pi \in \Omega$. If $\pi_* \in \Omega$ and π_* is 1-optimal among $\pi \in \Omega$, then π_* is average overtaking optimal among $\pi \in \Omega$.*

REMARK 7. In Section 3, we showed that (C) implies Statement 1 (see Remark 6). Unfortunately, (C) does not imply Statement 2 (see Example 1). By Theorem 5, Statement 2 holds under the following convergence condition which is stronger: (CS) *$g(s, \pi) \equiv \lim_n V^n(s, \pi)/n$ and $b(s, \pi) \equiv \lim_N \sum_{j=1}^N (V^j(s, \pi) - jg(s, \pi))/N$ exist and are finite for all $s \in S$ and all stationary π.* This condition holds when S is finite (see Doob (1953)). Note that (EPR) does not imply (CS) (counterexamples are easy to construct). In Lemma 3 below, we show that (UR) implies (CS), hence the following:

THEOREM 6. *Condition* (UR) *implies Statement* 2.

Arguing as in the proof of Corollary 3, one can establish:

COROLLARY 4. *Under* (UR), *if π_* is optimal among stationary policies, then it is average overtaking optimal among stationary policies.*

The analogue of Corollary 4 does not hold for 1-optimal π_*: under (UR), a nonstationary policy can be 1-optimal without being average overtaking optimal (see Example 8). Note that both Theorem 6 and Corollary 4 fail if we relax the boundedness condition on $i(\cdot, \cdot)$ (see Example 6).

PROOF OF THEOREM 5. Let $\pi_* \in \Omega$ be 1-optimal among policies in Ω and let π be an arbitrary policy in Ω. Fix $s \in S$. By Corollary 1, $g(s, \pi_*) \geqq g(s, \pi)$. One can easily show that $g(s, \pi_*) > g(s, \pi)$ implies $\lim_N \sum_{n=1}^N (V^n(s, \pi_*) - V^n(s, \pi)) = \infty$, which is stronger than (6). Thus we need only consider the case $g(s, \pi_*) = g(s, \pi)$. In this case, $\liminf_{\beta \to 1-} (V_\beta(s, \pi_*) - V_\beta(s, \pi)) = \liminf_{\beta \to 1-} \{\sum_{j=1}^\infty \beta^{j-1}(r_j(s, \pi_*) - g(s, \pi_*)) - \sum_{j=1}^\infty \beta^{j-1}(r_j(s, \pi) - g(s, \pi))\}$. By condition (CS) and an Abelian argument (see Hobson (1926)), the RHS equals

$$\lim_N \sum_{n=1}^N \sum_{j=1}^n (r_j(s, \pi_*) - g(s, \pi_*))/N - \lim_N \sum_{n=1}^N \sum_{j=1}^n (r_j(s, \pi) - g(s, \pi))/N.$$

Since $g(s, \pi_*) = g(s, \pi)$, the latter equals $\lim_N \sum_{n=1}^N \sum_{j=1}^n (r_j(s, \pi_*) - r_j(s, \pi))/N$. This and the 1-optimality of π_* give us (6).

Theorem 6 requires only the following lemma (see Remark 7).

LEMMA 3. *Condition* (UR) *implies condition* (CS).

PROOF. Let π be a stationary policy. Fix $s \in S$. Since (UR) implies (EPR), the existence of $\lim_n V^n(s, \pi)/n$ follows from Lemma 2. Using Theorem 7.1 of Orey (1971) and the interpretation of uniform ϕ-recurrence given in Section 2, one can show that $\lim_N \sum_{j=1}^N (V^j(s, \pi) - j \lim_n V^n(s, \pi)/n)/N$ exists and is finite. We are done.

5. Counterexamples. This section contains all of our counterexamples. In Example 1, there exists a stationary optimal policy which is not average overtaking optimal. In the second example, there exists a stationary average overtaking optimal policy, and a stationary 1-optimal policy which is not average overtaking optimal. (Hence, the existence of an average overtaking optimal policy is not sufficient for (12).) In the third and fourth, there exist stationary policies which are both optimal and average overtaking optimal but are not lim inf average optimal and lim sup average optimal, respectively. The fifth example shows that Theorem 4 and Corollary 3 require the boundedness condition on $i(\cdot, \cdot)$, while the sixth shows that the same is true for Theorem 6 and Corollary 4. Example 7 shows that (9) fails if we define 1-optimality differently (see Remark 2). Finally, Example 8 shows that even under (UR), a nonstationary 1-optimal policy need not be average overtaking among stationary policies (see Remark 4).

The first two examples depend on Lemma 5 (see below), which establishes the existence of a bounded sequence of real numbers $\{a_j\}_{j=1}^\infty$ satisfying

$$\begin{aligned} &\liminf_N \sum_{n=1}^N \sum_{j=1}^n a_j/N = -1\,, \\ &\lim_{\beta \to 1-} \sum \beta^{j-1} a_j = \limsup_N \sum_{n=1}^N \sum_{j=1}^n a_j/N = 0\,. \end{aligned} \tag{23}$$

EXAMPLE 1. Let the state space consist of 0, 1, 2, $\cdots$, ∞. To each state there correspond two actions, 0 and 1. The functions of $q(\cdot \mid \cdot, \cdot)$ and $i(\cdot, \cdot)$ satisfy:

$$\begin{gathered} q(j+1 \mid j, 0) = q(j+1 \mid j, 1) = 1\,, \qquad j = 1, 2, \cdots, \\ q(1 \mid 0, 1) = 1\,, \qquad q(0 \mid 0, 0) = q(\infty \mid 0, 0) = \tfrac{1}{2}\,, \\ q(\infty \mid \infty, 0) = q(\infty \mid \infty, 1) = 1\,; \\ i(j, 0) = i(j, 1) = a_{j+1}\,, \qquad j = 1, 2, \cdots, \\ i(0, 1) = a_1\,, \qquad i(0, 0) = -\tfrac{1}{4}\,, \qquad i(\infty, 0) = i(\infty, 1) = 0\,. \end{gathered}$$

Let π_j denote the policy which always selects action j ($j = 0, 1$). One can easily show that π_1 is optimal. That π_1 is not average overtaking optimal follows from

$$\liminf_N \sum_{n=1}^N \sum_{j=1}^n (V^j(0, \pi_1) - V^j(0, \pi_0))/N = \liminf_N (\sum_{n=1}^N \sum_{j=1}^n a_j)/N + \tfrac{1}{2}$$

and (23).

EXAMPLE 2. Let the state space consist of 0, 1, 2, $\cdots$. To each state, there correspond two actions, 0 and 1. The functions $q(\cdot \mid \cdot, \cdot)$ and $i(\cdot, \cdot)$ satisfy:

$$\begin{gathered} q(1 \mid 0, 1) = q(0 \mid 0, 0) = q(j+1 \mid j, 0) = q(j+1 \mid j, 1) = 1\,, \qquad j = 1, 2, \cdots, \\ i(0, 1) = a_i\,, \qquad i(0, 0) = 0\,, \qquad i(j, 0) = i(j, 1) = a_{j+1}\,, \qquad j = 1, 2, \cdots. \end{gathered}$$

Define π_j $(j = 0, 1)$ as in Example 1. Using (23), one can show that π_0 is both 1-optimal and average overtaking optimal while π_1 is 1-optimal but not average overtaking optimal.

Our next two examples depend on the following lemma, the proof of which appears in Section 6.

LEMMA 4. *There exists a bounded sequence of real numbers* $\{v_j\}_{j=1}^{\infty}$ *which satisfy*

$$\lim_N \sum_{n=1}^{N} \sum_{j=1}^{n} v_j/N = \infty\,, \qquad \liminf_n \sum_{j=1}^{n} v_j/n < 0\,. \tag{24}$$

EXAMPLE 3. Let the state and action spaces be as in Example 2. Transitions are deterministic:

$$q(0\,|\,0, 0) = q(1\,|\,0, 1) = q(j+1\,|\,j, 0) = q(j+1\,|\,j, 1) = 1\,, \qquad j = 1, 2, \cdots.$$

The immediate income depends only on the state:

$$i(0, 0) = i(0, 1) = 0\,, \qquad i(j, 0) = i(j, 1) = v_j\,, \qquad j = 1, 2, \cdots.$$

Define π_j $(j = 0, 1)$ as in Example 1. Clearly π_1 is average overtaking optimal but is not lim inf average optimal. That π_1 is optimal follows from the fact (see Hobson (1926)) that for any policy π and any $s \in S$,

$$\begin{aligned} \liminf_N \sum_{n=1}^{N} V^n(s, \pi)/N &\leqq \liminf_{\beta \to 1-} V_\beta(s, \pi) \leqq \limsup_{\beta \to 1-} V_\beta(s, \pi) \\ &\leqq \limsup_N \sum_{n=1}^{N} V^n(s, \pi)/N\,. \end{aligned} \tag{25}$$

EXAMPLE 4. Let the state space, action space and law of transition be as in Example 3. The only difference is that we define $i(\cdot, \cdot)$ by $i(0, 0) = i(1, 0) = 0$ and $i(j, 0) = i(j, 1) = w_j \equiv -v_j$ for $j = 1, 2, \cdots$. Define π_j $(j = 0, 1)$ as in Example 1. Clearly, π_0 is average overtaking optimal, but is not lim sup average optimal, since by (24) we have $\lim_N \sum_{n=1}^{N} \sum_{j=1}^{n} w_j/N = -\infty$ and $\limsup \sum_{j=1}^{n} w_j/n > 0$. That π_0 is optimal follows from (25).

EXAMPLE 5. Let the state and action space be as in Example 1. Transitions are described as follows:

$$\begin{aligned} q(\infty\,|\,\infty, 0) = q(\infty\,|\,\infty, 1) = q(0\,|\,0, 0) = 1\,, \qquad q(1\,|\,0, 1) = \tfrac{1}{2} = q(\infty\,|\,0, 1)\,, \\ q(j+1\,|\,j, 0) = q(j+1\,|\,j, 1) = q(\infty\,|\,j, 1) = q(\infty\,|\,j, 0) = \tfrac{1}{2}\,, \\ j = 1, 2, \cdots. \end{aligned}$$

The immediate income depends only on the state:

$$i(0, 0) = i(0, 1) = 0 = i(\infty, 0) = i(\infty, 1)\,, \qquad i(j, 0) = i(j, 1) = 2^j w_j\,,$$

where w_j is defined as in Example 4 for $j = 1, 2, \cdots$. Define π_j $(j = 0, 1)$ as in Example 1. Observe that under any stationary policy, the system eventually reaches an absorbing state with probability 1 (condition (EPR)). However, using the arguments of Example 4, one can show that π_0 is optimal but not lim sup average optimal.

EXAMPLE 6. Let the state and action space be as in Example 1. Transitions

are described as follows:

$$q(\infty \mid \infty, 0) = q(\infty \mid \infty, 1) = 1\,, \qquad q(1 \mid 0, 1) = q(\infty \mid 0, 1) = \tfrac{1}{2}\,,$$
$$q(0 \mid 0, 0) = q(\infty \mid 0, 0) = \tfrac{1}{2}\,,$$
$$q(j+1 \mid j, 0) = q(j+1 \mid j, 1) = q(\infty \mid j, 0) = q(\infty \mid j, 1) = \tfrac{1}{2}\,, \qquad j = 1, 2, \cdots .$$

The immediate income depends only on the state:

$$i(j, 0) = i(j, 1) = 2^{j+1}a_{j+1}\,, \qquad j = 1, 2, \cdots,$$
$$i(0, 1) = 2a_1\,, \qquad i(0, 0) = -\tfrac{1}{4}\,, \qquad i(\infty, 0) = i(\infty, 1) = 0\,,$$

where a_j, $(j = 1, 2, \cdots)$, satisfies (23). Define π_j $(j = 0, 1)$ as in Example 1. Using the arguments of Example 1, one can show that π_1 is optimal, but not average overtaking optimal. Condition (UR) follows from the fact that the probability of moving to state ∞ in any transition is always $\geqq \frac{1}{2}$.

EXAMPLE 7. Let the S consist of 0, 1, 2, $\cdots$ together with $(k, 1)$, $(k, 2)$, $\cdots$, $(k, 2k)$ for $k = 1, 2, \cdots$. To each state, there correspond two actions, 0 and 1. Transitions are deterministic. When in state k $(k = 1, 2, \cdots)$, we can either (Act 0) move to $k + 1$ or (Act 1) move to $(k, 1)$. Once we reach state $(k, 1)$, we proceed along the path connecting (k, j) to $(k, j + 1)$ until we hit $(k, 2k)$; then we move directly to state 0. State 0 is an absorbing state. Formally,

$$q(0 \mid 0, i) = 1 = q(0 \mid (k, 2k), i) = q((k, j+1) \mid (k, j), i)\,,$$
$$q(k+1 \mid k, 0) = 1 = q((k, 1) \mid k, 1)\,,$$
$$i = 0, 1;\ j = 1, \cdots, 2k-1;\ k = 1, 2, \cdots .$$

The immediate income depends only on the state:

$$i(k, i) = 0\,, \qquad i = 0, 1;\ k = 0, 1, 2, \cdots,$$

and

$$i((k, j), i) = 1\,, \qquad 1 \leqq j \leqq k\,,$$
$$= -2\,, \quad k < j \leqq 2k\,, \qquad i = 0, 1;\ k = 1, 2, \cdots .$$

Let π_* denote the policy which always chooses action 0. Also, for $N = 1, 2, \cdots$, let π_N be the stationary policy which selects action 1 in state N and action 0 in all other states. One can easily show that $\lim_{\beta \to 1-} \max_N V_\beta(1, \pi_N) > 0$. Since $V_\beta(1, \pi_*) = 0$ for all β, we have $\liminf_{\beta \to 1-} V_\beta(1, \pi_*) - V_\beta^*(1) < 0$. Thus π_* is not 1-optimal if 1-optimality is defined by (13), yet π_* satisfies (6).

The next example depends on Lemma 5 below, the proof of which appears in Section 6.

LEMMA 5. (a) *There exists a sequence of integers* $N_0 = 0, N_1, N_2, \cdots$ *satisfying*

$$0 < N_k^2 + 4N_k < N_{k+1}^2\,, \qquad k = 1, 2, \cdots, \tag{26}$$

$$N_k^2 \beta^{N_k^2}(1-\beta)(1-\beta^{4N_k}) \leqq 2^{-k}\,, \qquad 0 < \beta < 1;\ k = 1, 2, \cdots, \tag{27}$$

and

$$(1 - (\tfrac{1}{2})^{k-1})^{I_k} \leqq \tfrac{1}{2}, \qquad k = 1, 2, \cdots, \tag{28}$$

where I_k denotes the largest integer $\leqq (N_k^2 - N_{k-1}^2 - 4N_{k-1} - 1)/(k-1)$.

(b) *Given* $N_0 = 0, N_1, N_2, \cdots$ *satisfying* (26) *and* (27), *define* $\{a_j\}_{j=1}^{\infty}$ *by*

$$\begin{aligned} a_j &= 0 \quad \text{if} \quad N_{k-1}^2 + 4N_{k-1} < j \leqq N_k^2, \\ &= -1 \quad \text{if} \quad N_k^2 < j \leqq N_k^2 + N_k \quad \text{or} \quad N_k^2 + 3N_k < j \leqq N_k^2 + 4N_k, \\ &= 1 \quad \text{if} \quad N_k^2 + N_k < j \leqq N_k^2 + 3N_k, \quad \text{for} \quad k = 0, 1, \cdots. \end{aligned} \tag{29}$$

Then $\{a_j\}_{j=1}^{\infty}$ *satisfies* (23).

EXAMPLE 8. Let $N_0, N_1, N_2, \cdots$ and $a_1, a_2, \cdots$ satisfy the conditions of Lemma 5, and let $\alpha_1, \alpha_2, \cdots$ satisfy

$$\alpha_k^{N_k^2} \geqq \tfrac{1}{2}, \qquad 0 < \alpha_k < 1, \qquad k = 1, 2, \cdots. \tag{30}$$

Let the state space consist of $1, 2, \cdots$ together with $(k, 1), \cdots, (k, 4N_k)$ for $(k = 1, 2, \cdots)$. To each state, there correspond four actions: 0, 1, 2 and 3. When in state k, we can either (Act 0) move directly to state $(k, 1)$, (Act 1) move directly to state 1, (Act 2) move to states $k + 1$ and 1 with probability $\frac{1}{2}$ each by tossing a fair coin, or (Act 3) remain in state k with probability α_k and move to state 1 with probability $1 - \alpha_k$ by tossing a biased coin. Specifically,

$$\begin{aligned} q((k, 1) \mid k, 0) = 1 = q(1 \mid k, 1), \qquad q(k + 1 \mid k, 2) = \tfrac{1}{2} = q(1 \mid k, 2), \\ q(k \mid k, 3) = \alpha_k = 1 - q(1 \mid k, 3), \qquad k = 1, 2, \cdots. \end{aligned}$$

Once we reach state $(k, 1)$, we proceed along the path connecting (k, j) to $(k, j + 1)$ until we hit $(k, 4N_k)$; then we move directly to state 1, i.e.,

$$\begin{aligned} q((k, j + 1) \mid (k, j), i) = q(1 \mid (k, 4N_k), i) = 1, \\ i = 0, \cdots, 3; \; j = 1, \cdots, 4N_k - 1; \; k = 1, 2, \cdots. \end{aligned}$$

All coin tosses are assumed independent. The immediate income depends only on the state:

$$\begin{aligned} i(k, i) &= 0, \qquad i = 0, 1, 2; \; k = 1, 2, \cdots, \\ i((k, j), i) &= -1, \qquad 1 \leqq j \leqq N_k \quad \text{or} \quad 3N_k < j \leqq 4N_k \\ &= 1, \qquad N_k < j \leqq 3N_k, \qquad i = 0, 1; \; k = 1, 2, \cdots. \end{aligned}$$

One can show that given any (nonrandomized) stationary policy, the state space of the associated Markov chain can be expressed as a union of uniformly recurrent classes (see Section 2). These classes are, of course, not disjoint. In fact, for each such class, the set of positive recurrent states consists of the states accessible from state 1. The role played by j in the definition of uniform recurrence (see Section 2) is assumed by state 1. Hence condition (UR) holds.

Let π_1 denote the policy which always selects action 1. One can easily show that π_1 is both optimal and average overtaking optimal. We shall construct a

nonstationary policy π_* satisfying

$$\sum_{j=1}^{\infty} \beta^{j-1} a_j \leqq \sum_{j=1}^{\infty} \beta^{j-1} r_j(1, \pi_*) \leqq 0\,, \qquad 0 \leqq \beta < 1\,, \tag{31}$$

$$\sum_{n=1}^{N} V^n(1, \pi_*) \leqq (\tfrac{1}{4}) \sum_{n=1}^{N} (\sum_{j=1}^{n} a_j)\,, \qquad N = 1, 2, \cdots, \tag{32}$$

and

$$r_j(s, \pi_*) = r_j(s, \pi_1)\,, \qquad s \neq 1;\ j = 1, 2, \cdots. \tag{33}$$

Now, (23), (31) and (32) imply

$$\liminf_N \sum_{n=1}^{N} V^n(1, \pi_*)/N \leqq -\tfrac{1}{4} \quad \textit{and} \quad \lim_{\beta \to 1-} \sum_{j=1}^{n} \beta^{j-1} r_j(1, \pi_*) = 0\,. \tag{34}$$

Using (33), (34), and the fact that π_1 is both optimal and average overtaking optimal, one can show that any such π_* is 1-optimal but not average overtaking optimal. We leave the details to the reader.

Now we construct π_*. Given any initial state $s \neq 1$, always choose *act* 1. (Clearly, (33) holds.) For the case in which the initial state is 1, follow the *program*:

I. *Select act* 3 *from time* 1 *until time* N_1^2. *Select act* 0 *from time* N_1^2 *until time* $N_1^2 + 4N_1 + 1$. *Then set* $K = 1$.

II. *Select act* 2 *from time* $N_K^2 + 4N_K + 1$ *until time* N_{K+1} *or the first time that the system reaches state* $K + 1$, *whichever occurs first.*

III. *If the system reaches state* $K + 1$ *before time* N_{K+1}^2, *then choose act* 3 *from the time the system reaches* $K + 1$ *until time* N_{K+1}^2.

IV. *If the system occupies state* $K + 1$ *at time* N_{K+1}^2, *then choose act* 0 *from time* N_{K+1}^2 *until time* $N_{K+1}^2 + 4N_{K+1} + 1$; *otherwise, choose act* 1 *from time* N_{K+1}^2 *until time* $N_{K+1}^2 + 4N_{K+1} + 1$.

V. *Set* $K = K + 1$. *Then go to* II.

We need only establish (31) and (32). Assume π_* is used and the initial state is 1. For any k ($k = 1, 2, \cdots$), one can easily show that if the system occupies state k at time N_k^2, then the income for period j will equal a_j for values of $j \in [N_{k-1}^2 + 4N_{k-1} + 1, N_k^2 + 4N_k]$; while if the system does not occupy state k at time N_k^2, then the income will be zero for all such values of j. Using this and (29), one can establish (31). As the reader can verify, (32) will also follow once we establish

$$P[\text{system } \textit{occupies} \text{ state } k \text{ at time } N_k^2] \geqq \tfrac{1}{4}\,, \qquad k = 1, 2, \cdots. \tag{35}$$

To prove (35), we need

$$P[\text{system } \textit{reaches} \text{ state } k \text{ by time } N_k^2] \geqq \tfrac{1}{2}\,, \qquad k = 1, 2, \cdots. \tag{36}$$

Toward establishing (36), observe that each time act 2 is chosen a fair coin is tossed. Call a toss "favorable" if it does not direct the system to state 1. Now, one can show that the system will always be in state 1 at time $N_{k-1}^2 + 4N_{k-1} + 1$, ($k = 1, 2, \cdots$). Hence one way in which the system can reach state k by time

N_k^2 is for one of the first I_k groups of successive coin tosses to constitute a favorable run of length $k-1$ (I_k is defined in Lemma 5(a)). The probability of none of those runs being favorable equals $(1-(\frac{1}{2})^{k-1})^{I_k}$ by the independence of coin tosses. Applying (28), we get (36). Now, given that the system reaches state k at time $\tau < N_k^2$, the system remains in state k between time τ and time N_k^2 with probability ${\alpha_k}^{N_k^2-\tau} > {\alpha_k}^{N_k^2}$, which is $\geqq \frac{1}{2}$ by (28). This and (36) give us (35).

6. Proofs. The only things left are the proofs of Lemmas 1, 4 and 5.

PROOF OF LEMMA 1. Let $\{a_j\}_{j=1}^{\infty}$ be any sequence satisfying (16). It is convenient to work with the sequence $\{x_j\}_{j=1}^{\infty}$ defined by

$$x_j = a_j - \liminf_{\beta\to 1-}(1-\beta)\sum_{j=1}^{\infty}\beta^{i-1}a_i\,, \qquad j=1,2,\cdots. \tag{37}$$

Clearly,

$$\liminf_{\beta\to 1-}(1-\beta)\sum_{j=1}^{\infty}\beta^{j-1}x_j = 0\,, \qquad |x_j|\leqq 2M\,, \qquad j=1,2,\cdots. \tag{38}$$

By (37) and (38), (15) is equivalent to

$$\liminf \sum_{j=1}^{n} x_j/n > -\varepsilon\,, \tag{39}$$

while (17) is equivalent to

$$\limsup \sum_{j=1}^{n} x_j/n < \delta\,. \tag{40}$$

These equivalences allow us to establish the lemma by showing that

$$\begin{aligned}&\liminf_n \textstyle\sum_{j=1}^{n} x_j/n \leqq -\varepsilon\,,\\ &\limsup_n \textstyle\sum_{j=1}^{n} x_j/n < \delta(\varepsilon, M) \equiv \varepsilon(1-e^{-\varepsilon/3M})/6e\end{aligned} \tag{41}$$

is impossible, since it will then follow that $\delta = \delta(\varepsilon, M)$ meets the required conditions. (The reader should verify this.)

Assume (41). Let $S_n = \sum_{j=1}^{n} x_j$ $(n=1,2,\cdots)$ and let $\rho = \limsup S_n/n$. Clearly, there exists an integer N such that

$$S_n \leqq 2\rho n\,, \qquad n \geqq N\,. \tag{42}$$

By (41), there exist integers $N_1, N_2, \cdots$ satisfying

$$N < N_j < N_{j+1}\,, \qquad \lim_j N_j = \infty\,, \qquad S_{N_j} \leqq -(\tfrac{2}{3})\varepsilon N_j\,, \qquad j=1,2,\cdots. \tag{43}$$

Now, (43) and the fact that the x_j's are bounded in absolute value by $2M$ (see (38)) imply that

$$S_n \leqq -(\tfrac{1}{3})\varepsilon N_j\,, \qquad L_j \leqq n \leqq U_j\,, \tag{44}$$

where L_j is the smallest integer $\geqq N_j(1-\varepsilon/6M)$ and U_j is the largest integer $\leqq N_j(1+\varepsilon/6M)$, $(j=1,2,\cdots)$. It is easy to show that

$$(1-\beta)\sum_{j=1}^{\infty}\beta^{j-1}x_j = (1-\beta)^2\sum_{j=1}^{\infty}\beta^{j-1}S_j\,.$$

Clearly, the RHS equals

$$(1-\beta)^2\Big(\sum_{j=1}^{N}\beta^{j-1}S_j + \sum_{j=N+1}^{L_k}\beta^{j-1}S_j + \sum_{j=L_k+1}^{U_k}\beta^{j-1}S_j + \sum_{j=U_k+1}^{\infty}\beta^{j-1}S_j\Big)\,,$$

which according to (38), (42) and (44) is bounded above by

$$2(1-\beta)^2N^2M + (1-\beta)^2\sum_{j=1}^{\infty}\beta^{j-1}(2\rho j) - (1-\beta)^2\sum_{j=L_k+1}^{U_k}\beta^{j-1}(\tfrac{1}{3})\varepsilon N_k$$

for $k = 1, 2, \cdots$. Consequently,

$$(45)\qquad (1-\beta)\sum_{j=1}^{\infty}\beta^{j-1}x_j \leqq 2(1-\beta)^2N^2M + 2\rho - (\tfrac{1}{3})\varepsilon N_k(1-\beta)\beta^{L_k}(1-\beta^{U_k-L_k}),$$

for $k = 1, 2, \cdots$. Let $\beta_k = 1 - 1/N_k$, $(k = 1, 2, \cdots)$. Evidently, $N_k(1-\beta_k) = 1$ and $\lim_k \beta_k{}^{N_k} = e^{-1}$. Now, one can easily show that

$$\lim_k N_k(1-\beta_k)\beta_k{}^{L_k}(1-\beta_k{}^{U_k-L_k}) = \lim_k \beta_k{}^{L_k} - \beta_k{}^{U_k} = e^{-(1-\varepsilon/6M)} - e^{-(1+\varepsilon/6M)} > (1-e^{-\varepsilon/3M})/e .$$

Thus, the fact that $\lim_k (1-\beta_k)^2N^2M = 0$, and (45) imply

$$\limsup_k (1-\beta_k)\sum_{j=1}^{\infty}(\beta_k)^{j-1}x_j < 2\rho - \varepsilon(1-e^{-\varepsilon/3M})/3e .$$

Consequently, $\liminf_{\beta\to 1-}(1-\beta)\sum_{j=1}^{\infty}\beta^{j-1}x_j < 0$ whenever $\rho < \delta(\varepsilon, M)$ (see (41)), contradicting (38). Lemma 1 follows.

PROOF OF LEMMA 4. Our proof is constructive. For each k $(k = 0, 1, \cdots)$, define r_j by

$$r_j = 1, \quad 2(2^k-1) < j \leqq 2(2^k-1) + 2^k$$
$$= -1, \quad 2(2^k-1) + 2^k < j \leqq 2(2^{k+1}-1) .$$

Let $v_j(\delta) = r_j - \delta$ for all real δ. Clearly, $\liminf_n \sum_{j=1}^{n} v_j(\delta)/n = -\delta$. Hence, all we need is a $\delta > 0$ satisfying

$$(46)\qquad \liminf_N \sum_{n=1}^{N}\sum_{j=1}^{n} v_j(\delta)/N = \infty .$$

We will show that $\delta = 2^{-11}$ works. Let $S_n = \sum_{j=1}^{n} r_j$ $(n = 1, 2, \cdots)$ and $m(k) = 2(2^k-1)$, $(k = 0, 1, \cdots)$. For any $\delta > 0$,

$$(47)\qquad \sum_{j=1}^{N}(S_j - j\delta)/N \geqq \sum_{j=1}^{m(k)} S_j/m(k+1) - \sum_{j=1}^{m(k+1)} j\delta/m(k), \quad m(k) \leqq N \leqq m(k+1) .$$

Now, the RHS of (47) equals

$$(\sum_{j=1}^{m(k)} S_j/m(k))(m(k)/m(k+1)) - \sum_{j=1}^{m(k)} j\delta/m(k)(\sum_{j=1}^{m(k+1)} j/\sum_{j=1}^{m(k)} j) .$$

Using the well-known formula for the sum of consecutive positive integers, one gets $\sum_{j=1}^{m(k)} j = (2^k-1)(2^{k+1}-1)$, $(k = 1, 2, 3, \cdots)$. Also, $2^{k-1} \leqq 2^k - 1$, $(k = 1, 2, \cdots)$. Hence, $\sum_{j=1}^{m(k+1)} j/\sum_{j=1}^{m(k)} j \leqq 8$ and $m(k)/m(k+1) \geqq \frac{1}{4}$, $(k = 1, 2, \cdots)$. This implies that the RHS of (47) is $\geqq (\frac{1}{4})\sum_{j=1}^{m(k)}(S_j - 32j\delta)/m(k)$, $(k = 1, 2, \cdots)$. Consequently, if $\delta > 0$ satisfies

$$(48)\qquad \lim_k \sum_{j=1}^{m(k)}(S_j - j\delta)/m(k) = \infty ,$$

then $2^{-5}\delta$ satisfies (46). We will show that $\delta = 2^{-6}$ satisfies (48).

As the reader can easily verify,

$$S_{m(j)+h} = h, \quad \text{if} \quad h = 1, \cdots, 2^j ,$$
$$= 2^{j+1} - h, \quad \text{if} \quad h = 2^j + 1, \cdots, 2^{j+1} ,$$

$(j = 0, 1, \cdots)$. This gives us

$$\sum_{i=m(k-1)+1}^{m(k)} S_i = 2^{2(k-1)}, \qquad k = 1, 2, \cdots .$$

Using the fact that $\sum_{j=1}^{m(k)} j \leqq 2^{2k+1}$, we have

$$\sum_{j=1}^{m(k)} (S_j - j\delta) \geqq 2^{2k-4} - 2^{2k+1}\delta \tag{49}$$

whenever $\delta > 0$. Observe that for $\delta = 2^{-6}$, the RHS of (49) $\geqq 2^{2k-5}$. Consequently, for $\delta = 2^{-6}$,

$$\liminf_k \sum_{j=1}^{m(k)} (S_j - j\delta)/m(k) \geqq \liminf_k (2^{2k-5})2^{-k-1} = \infty .$$

Lemma 5 requires the following result:

LEMMA 6. *For any $\varepsilon > 0$, there exists a positive integer N_0 such that $N \geqq N_0$ implies*

$$N^2\beta^{N^2}(1 - \beta)(1 - \beta^{4N}) \leqq \varepsilon , \qquad 0 < \beta < 1 .$$

PROOF. Let $\rho = -\ln \beta$. We need only show that

$$F(\rho, x) = x^2e^{-\rho x^2}(1 - e^{-\rho})(1 - e^{-4\rho x})$$

converges to 0 uniformly in ρ, $(0 \leqq \rho < \infty)$, as x approaches ∞. We will do this by proving that the existence of a $\delta > 0$ and sequences $\{x_j\}_{j=1}^{\infty}$ and $\{\rho_j\}_{j=1}^{\infty}$ satisfying

$$\lim_j x_j = \infty , \qquad x_j < x_{j+1} , \qquad F(x_j, \rho_j) \geqq \delta , \qquad j = 1, 2, \cdots \tag{50}$$

leads to a contradiction. To begin with, whenever $\{x_j\}_{j=1}^{\infty}$ and $\{\rho_j\}_{j=1}^{\infty}$ satisfy (50), we have

$$\lim_j \rho_j x_j = 0 , \qquad \lim_j \rho_j = 0 . \tag{51}$$

Otherwise, there exist a $\lambda > 0$ and subsequences $\{x_{n_j}\}_{j=1}^{\infty}$ and $\{\rho_{n_j}\}_{j=1}^{\infty}$ such that $x_{n_j}\rho_{n_j} \geqq \lambda$, $(j = 1, 2, \cdots)$. But then $F(\rho_{n_j}, x_{n_j}) \leqq x_{n_j}^2 e^{-\lambda x_{n_j}}$, $(j = 1, 2, \cdots)$, which is impossible by (50). Now (51) and the fact that $1 - e^{-z} \leqq ze^z$ for all positive z imply that for j sufficiently large,

$$F(\rho_j, x_j) \leqq (5\rho_j x_j) \cdot (\rho_j x_j^2 e^{-\rho_j x_j^2}) .$$

But the second factor in the RHS is bounded while the first converges to 0 by (51). This contradicts (50), finishing the proof.

PROOF OF LEMMA 5. First, part (a). Evidently there exist integers $I(1)$, $I(2)$, $\cdots$ such that

$$(1 - (\tfrac{1}{2})^{k-1})^I \leqq \tfrac{1}{2} , \qquad I \geqq I(k), \; k = 2, 3, \cdots . \tag{52}$$

By Lemma 6, there exist integers $N_1, N_2, \cdots$ satisfying (27). Lemma 6 also allows us to choose this sequence so that $N_k^2 \geqq (k - 1)I(k) + N_{k-1}^2 + 4N_{k-1} + 1$ for $k = 1, 2, \cdots$, where $I(k)$ satisfies (52) and $N_0 = 0$. One can easily show that any such sequence must satisfy (26) and (28). Part (a) follows.

For part (b), let $N_0 = 0, N_1, N_2, \cdots$ satisfy (26) and (27), and let $J_k = \{j \mid N_k^2 + 1 \leqq j \leqq N_k^2 + 4N_k\}$ for $k = 1, 2, \cdots$. Define $s_i = \sum_{j=1}^{i} a_j$ and $T_i = \sum_{j=1}^{i} s_j$, $(i = 1, 2, \cdots)$. Evidently, if $i \in J_k$ for some k, then s_i satisfies

$$
\begin{aligned}
s_{N_k^2+j} &= -j && \text{if } 0 \leqq j \leqq N_k \\
&= -(2N_k - j) && \text{if } N_k \leqq j \leqq 3N_k \\
&= 4N_k - j && \text{if } 3N_k \leqq j \leqq 4N_k ;
\end{aligned} \tag{53}
$$

while if $i \notin \bigcup_k J_k$, then $s_i = 0$. Using (53) and the well-known formula for the sums of consecutive positive integers, one can show that $T_{N_k^2+2N_k} = -N_k^2$ and $-N_k^2 \leqq T_j \leqq 0$, $(j \in J_k)$, for $k = 1, 2, \cdots$. Also, $T_j = 0$, $j \notin \bigcup_k J_k$. One can easily show that these facts imply $\limsup_N T_N/N = 0$ and $\liminf_N T_N/N = \lim_k (T_{N_k^2+2N_k})/(N_k^2 + 2N_k) = -1$. The only thing left to establish is

$$
\lim_{\beta \to 1-} \sum_{j=1}^{\infty} \beta^{j-1} a_j = 0 . \tag{54}
$$

We will make use of the identity

$$
(1 - \beta)^2 \sum_{j=1}^{\infty} \beta^{j-1} T_j = \sum_{j=1}^{\infty} \beta^{j-1} a_j , \qquad 0 < \beta < 1 . \tag{55}
$$

Since $T_j \geqq -N_k^2$ for $j \in J_k$, we have $\sum_{j \in J_k} \beta^{j-1} T_j \geqq -N_k^2 \beta^{N_k^2} (\sum_{j=1}^{4N_k} \beta^{j-1})$. Hence,

$$
(1 - \beta)^2 \sum_{j \in J_k} \beta^{j-1} T_j \geqq -N_k^2 \beta^{N_k^2} (1 - \beta)(1 - \beta^{4N_k}) . \tag{56}
$$

Let $\delta > 0$ be arbitrary. Condition (54) will follow if we can select a $\beta_0 < 1$ so that

$$
\sum_{j=1}^{\infty} \beta^{j-1} a_j \geqq -\delta , \qquad \beta_0 \leqq \beta < 1 . \tag{57}
$$

Select k_0 and $\beta_0 < 1$ so that $\sum_{k=k_0+1}^{\infty} 2^{-k} \leqq \delta/2$ and $(1 - \beta) N_{k_0}^2 (1 - \beta^{N_{k_0}^2}) \leqq \delta/2$ for $\beta_0 \leqq \beta < 1$. The identity (55) implies that

$$
\sum_{j=1}^{\infty} \beta^{j-1} a_j = (1 - \beta)^2 \sum_{j=1}^{N_{k_0}^2} \beta^{j-1} T_j + (1 - \beta)^2 \sum_{j=N_{k_0}^2+1}^{\infty} \beta^{j-1} T_j . \tag{58}
$$

By (56), the fact that $T_j = 0$ for $j \notin \bigcup_k J_k$, and the fact that $t_j \geqq -N_k^2$ for $j \in J_k$ $(k = 1, 2, \cdots)$, the RHS of (58) is $\geqq (1 - \beta)^2 \sum_{j=1}^{N_{k_0}^2} N_{k_0}^2 \beta^{j-1} - \sum_{k=k_0+1}^{\infty} N_k^2 \beta^{N_k^2} (1 - \beta)(1 - \beta^{4N_k})$. But the latter is $\geqq -(1 - \beta) N_{k_0}^2 (1 - \beta^{N_{k_0}^2}) - \delta/2$ by (27) and our choice of k_0. This and our method of selecting β_0 give us (57). We are done.

Acknowledgments. We wish to thank Ben Fox for his helpful comments.

REFERENCES

BLACKWELL, D. (1962). Discrete dynamic programming. *Ann. Math. Statist.* **33** 719–726.

BLACKWELL, D. (1965). Discounted dynamic programming. *Ann. Math. Statist.* **36** 226–235.

BROWN, B. (1965). On the iterative method of dynamic programming on a finite space discrete time Markov process. *Ann. Math. Statist.* **36** 1279–1285.

CHUNG, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, Berlin.

DENARDO, E. and MILLER, B. (1968). An optimality condition for discrete dynamic programming with no discounting. *Ann. Math. Statist.* **39** 1220–1227.

DERMAN, C. (1964). On sequential control processes. *Ann. Math. Statist.* **35** 341–349.

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

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

DOOB, J. (1953). *Stochastic Processes.* Wiley, New York.

FLYNN, J. (1974). Averaging versus discounting in dynamic programming: a counterexample. *Ann. Statist.* **2** 411–413.

HOBSON, E. (1926). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series.* Cambridge Univ. Press.

LIGGETT, T. and LIPPMAN, S. (1969). Stochastic games with perfect information and time average payoff. *SIAM Rev.* **11** 604–607.

LIPPMAN, S. (1969). Criterion equivalence in discrete dynamic programming. *Operations Res.* **17** 920–923.

MAITA, A. (1965). Dynamic programming for countable state systems. *Sankhyā Ser. A* **27** 259–266.

OREY, S. (1971). *Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand, Princeton.

ROSS, S. (1968). Non-discounted denumerable Markovian decision models. *Ann. Math. Statist.* **39** 412–423.

ROSS, S. (1971). On the nonexistence of ε-optimal randomized stationary policies in average cost Markov decision models. *Ann. Math. Statist.* **42** 1567–1568.

STRAUCH, R. (1966). Negative dynamic programming. *Ann. Math. Statist.* **37** 871–890.

VEINOTT, A. (1966). On finding optimal policies in discrete dynamic programming with no discounting. *Ann. Math. Statist.* **37** 1284–1294.

VEINOTT, A. (1969). Discrete dynamic programming with sensitive discount optimality criteria. *Ann. Math. Statist.* **40** 1635–1660.

SCHOOL OF BUSINESS ADMINISTRATION
WAYNE STATE UNIVERSITY
DETROIT, MICHIGAN 48202