The Annals of Mathematical Statistics
1971, Vol. 42, No. 3, 919-926

CONTINUOUSLY DISCOUNTED MARKOV DECISION MODEL WITH COUNTABLE STATE AND ACTION SPACE[1]

By Prasadarao Kakumanu
New York University

1. Introduction. We are concerned with a continuous time Markov decision process in which both the state space $\mathbf{S}$ and the action space $\mathbf{A}$ are countable. The process is continuously observed and found in one of a possible state $i \in \mathbf{S}$, then an action $a \in \mathbf{A}$ is taken. As a result a return $r(i, a)$ is obtained and the process moves to a new state $j \in \mathbf{S}$, which is governed by the transition probability rates $q(j \mid i, a)$. Let $r(a)$ be the return vector whose ith element is $r(i, a)$, $i \in \mathbf{S}$. And let $Q(a)$ be the transition probability rate matrix whose (i, j)th element is $q_{ij}(a) = q(j \mid i, a)$; $i, j \in S$.

A deterministic memoryless policy π is a mapping from $\mathbf{S}X(0, \infty)$ into $\mathbf{A}$. At any epoch t, if the current state is $S_t = i$, our action is $A_t = \pi(i, t)$. We consider only deterministic memoryless policies. In addition, we assume that for every $i \in \mathbf{S}$, $\pi(i, \cdot)$ is Lebesgue measurable. Such a Lebesgue measurable, memoryless, deterministic policy we call a *Markov policy*. A Markov policy is called stationary if $\pi(i, t) = \pi(i)$, that is the action taken depends only on the current state $S_t = i$, and not on time t. Let $q_{ij}(t, \pi) = q(j \mid i, \pi(i, t))$; $i, j \in \mathbf{S}$ be the transition probability rates from the state i to the state j when the policy π is used. And let $Q(t, \pi) = \{q_{ij}(t, \pi); i, j \in \mathbf{S}\}$ be the transition probability rate matrix which we call the infinitesimal generator of the Markov decision process, when the policy π is used. When π is stationary we write $Q(\pi)$ instead of $Q(t, \pi)$. Throughout the paper we assume that for all $i \in \mathbf{S}$, $t \in [0, \infty)$ and for any given π:

Assumption 1. $q_{ij}(t, \pi) \geqq 0 \quad i \neq j, \sum_j q_{ij}(t, \pi) = 0$, and

Assumption 2. $|q_{ii}(t, \pi)| \leqq M$, for some positive number $M < \infty$.

Under these assumptions the author in [7], has shown the existence of a unique stochastic transition probability matrix function $F(s, t, \pi) = \{f_{ij}(s, t, \pi); i, j \in \mathbf{S}\}$, for any given Markov policy π, and that it, satisfies the Kolmogorov forward differential equations:

$$\frac{\partial F(s, t, \pi)}{\partial t} = F(s, t, \pi)\, Q(t, \pi) \qquad \text{with } F(s, s, \pi) = I \tag{1.1}$$

for almost all $t \geqq s \geqq 0$.

For any two vectors X_1 and X_2, we write $X_1 \geqq X_2$ if the inequality holds for all corresponding coordinates. We call any vector X is bounded if $\|X\| = \sup_i |x_i|$ is bounded. Let e be the infinite column vector with all coordinates unity.

Received February 2, 1970; revised October 30, 1970.

[1] This research is part of a doctoral dissertation at Cornell University and was supported by the National Science Foundation under Grant GK 856.

Let $r(i, t, \pi) = r(i, \pi(i, t))$, be the return when the state of the system is $i \in \mathbf{S}$ and the policy π is used. And let $r(t, \pi)$ be the return vector. If π is stationary we write $r(\pi)$ instead of $r(t, \pi)$. The return vector $r(\cdot, \cdot)$, which we assume bounded is a rate and the total return over a time interval $[s, t)$ is $\int_s^t r(u, \pi)\, du$. The ith element of the total discounted return vector $\Psi(\pi)$ to the system with the discount factor $\alpha > 0$ is defined to be:

$$\Psi(i, \pi) = \int_0^\infty e^{-\alpha t} \sum_j f_{ij}(0, t, \pi)\, r(j, t, \pi)\, dt, \qquad i \in \mathbf{S}. \tag{1.2}$$

For any Markov policy π and $\alpha > 0$, $\|\Psi(\pi)\|$ is bounded. For any $\varepsilon > 0$, π^* is called *ε-optimal* if for any measurable policy π, $\Psi(\pi^*) \geqq \Psi(\pi) - e\varepsilon$, and will be called *optimal* if it is ε-optimal for every $\varepsilon > 0$ or equivalently, if $\Psi(\pi^*) \geqq \Psi(\pi)$.

The problem is to find an optimal stationary policy π^* among the class of Markov policies, and how to obtain such a policy.

Howard [6], Martin-Löf [8], Miller [9], Rykov [11] and Veinott [13] and G. de Leve [3], [4] have studied the continuous time Markov sequential decision process under various conditions.

In this paper we show that $\sup_\pi \Psi(\pi)$ is the unique bounded solution g to the "dynamic programming optimality equation"

$$\alpha g = \sup_\pi \{r(\pi) + Q(\pi) g\}. \tag{1.3}$$

From this we then show the existence of ε-optimal stationary policies for $\varepsilon > 0$ when $\mathbf{A}$ is countable, and optimal stationary policies when $\mathbf{A}$ is finite. We also give a procedure which will yield an optimal stationary policy π^* and the corresponding optimal return $\Psi(\pi^*)$.

Throughout our discussion the Markov decision process starts from the origin. In view of this we write $F(t, \pi)$ instead of $F(0, t, \pi)$. If π is stationary the corresponding $F(t, \pi)$ is a time-homogeneous transition probability matrix function.

2. Existence of stationary optimal policies. We begin by considering equations similar to (1.3) and relate their solution to the expected discounted return $\Psi(\pi)$ as defined in (1.2).

THEOREM 2.1. *Let $\varepsilon \geqq 0$ and g a bounded vector be given. For a given stationary policy π,*

(a) *if $\alpha g \geqq r(\pi) + Q(\pi)\, g - e\varepsilon$ then $g \geqq \Psi(\pi) - \alpha^{-1} e\varepsilon$, and*

(b) *if $\alpha g \leqq r(\pi) + Q(\pi)\, g + e\varepsilon$ then $g \leqq \Psi(\pi) + \alpha^{-1} e\varepsilon$.*

PROOF. Let π be a given stationary policy, since π is a stationary policy the corresponding transition matrix $F(t, \pi)$ is time-homogeneous. In post multiplying the forward Kolmogorov equations (1.1) corresponding to given π by g we obtain after rearranging:

$$\frac{\partial}{\partial t} F(t, \pi) g = F(t, \pi)[Q(\pi) g]. \tag{2.1}$$

We have from (a)

$$Q(\pi)g \leqq \alpha g - r(\pi) + e\varepsilon. \tag{2.2}$$

Substituting the value of $Q(\pi)g$ in (2.1) after some simplification we arrive at

$$-\frac{\partial}{\partial t}[e^{-\alpha t}F(t, \pi)]g \geqq e^{-\alpha t}F(t, \pi)r(t, \pi) - e^{-\alpha t}e\varepsilon.$$

By integrating on both sides with respect to $t \in [0, \infty)$ we obtain $g \geqq \Psi(\pi) - \alpha^{-1}e\varepsilon$, which proves (a). (b) may be proved similarly. □

The hypothesis in the above theorem is stated for a given stationary π and the conclusion is about that π alone. The hypothesis in the following theorem is assumed for all actions $a \in \mathbf{A}$, and the conclusions concern all Markov π.

THEOREM. 2.2 *If g is a bounded vector on $\mathbf{S}$ which, for some $\varepsilon \geqq 0$ satisfies*

$$\alpha g \geqq r(a) + Q(a)g - e\varepsilon \qquad a \in A, \tag{2.3}$$

then $g \geqq \Psi(\pi) - \alpha^{-1}e\varepsilon$ for all Markov π.

PROOF. Since (2.3) is true for all $a \in \mathbf{A}$, in particular it is true for $a = \pi(i, t)$, the ith component of π, we have

$$\alpha g \geqq r(t, \pi) + Q(t, \pi)g - e\varepsilon. \tag{2.4}$$

Substituting for $Q(t, \pi)g$ in the forward Kolmogorov equations corresponding to Markov π. As in Theorem 2.1 after some simplification we arrive at $g \geqq \Psi(\pi) - \alpha^{-1}e\varepsilon$. Since π is arbitrary, the theorem is proved. □

Using Theorem 2.2, it is easy to show that, if

$$\alpha g = \sup_a\{r(a) + Q(a)g\}, \tag{2.5}$$

then $g \geqq \Psi(\pi)$ for all Markov π. This leads to the simple but often useful result which we state as:

COROLLARY 2.3. *If the expected return $\Psi(\pi^*)$ of a Markov policy π^* satisfies $\alpha\Psi(\pi^*) \geqq r(a) + Q(a)\Psi(\pi^*)$ for all $a \in \mathbf{A}$, then π^* is optimal.*

THEOREM 2.4. *Let π be a stationary policy. Then $g = \Psi(\pi)$ is the unique bounded solution to*

$$\alpha g = r(\pi) + Q(\pi)g. \tag{2.6}$$

PROOF. Using Theorem 2.1 it is easy to show that any solution g to (2.5) must be equal to $\Psi(\pi)$. We will now show that $\Psi(\pi)$ is a solution to (2.6). For any stationary policy π and $t > 0$

$$\Psi(\pi) = \int_0^t e^{-\alpha s}F(s, \pi)r(\pi)\,ds + e^{-\alpha t}F(t, \pi)\Psi(\pi).$$

Differentiating with respect to t and taking limits as $t \to 0^+$, in open form we have:

$$\begin{aligned} &\lim_{t\to 0^+} \textstyle\sum_j f_{ij}(t, \pi) r(j, \pi) \\ &+\lim_{t\to 0^+} \textstyle\sum_j f_{ij}(t, \pi)\Psi(j, \pi) \\ &+\lim_{t\to 0^+} \textstyle\sum_j (\partial/\partial t) f_{ij}(t, \pi)\Psi(j, \pi) = 0. \end{aligned} \tag{2.7}$$

From the definition of $F(t, \pi)$ we have for all $i, j \in \mathbf{S}$, $\lim_{t\to 0^+} f_{ij}(t, \pi) = \delta_{ij}$ and $\lim_{t\to 0^+} (\partial/\partial t) f_{ij}(t, \pi) = q_{ij}(\pi)$, and since $r(\pi)$ and $\Psi(\pi)$ are bounded vectors we obtain from (2.7):

$$r(i, \pi) - \alpha\Psi(i, \pi) + \textstyle\sum_j q_{ij}(\pi)\Psi(j, \pi) = 0, \qquad i \in \mathbf{S}.$$

which proves the theorem. □

The preceding theorems serve to bound and evaluate expected returns in terms of solutions of functional equations. The following theorem in some sense summarizes these results and more closely focuses our attention on the particular functional (2.5), whose solution has not yet been shown to exist.

THEOREM 2.5. *If there exists a bounded vector g which satisfies* (2.5), $\alpha g = \sup_a [r(a) + Q(a)g]$ *then* $g = \sup_\pi \Psi(\pi)$; *and for every* $\varepsilon > 0$ *there exists* ε*-optimal policies which are stationary. If* $\mathbf{A}$ *is finite. there exist optimal stationary policies.*

PROOF. By Corollary 2.3 $g \geqq \Psi(\pi)$ for every policy π. Now let $\varepsilon > 0$ be given. We define the stationary policy π_ε by taking for the action at state i, any action $\pi_\varepsilon(i)$ for which

$$\alpha g_i - \varepsilon \leqq r(i, \pi_\varepsilon(i)) + \textstyle\sum_j q_{ij}(\pi_\varepsilon) g_j, \qquad i \in \mathbf{S}.$$

By Theorem 2.1 (b) it follows

$$\Psi(\pi_\varepsilon) \geqq g - e\varepsilon \geqq \sup_\pi \Psi(\pi) - e\varepsilon.$$

Thus ε-optimal stationary policies exist and by letting $\varepsilon \to 0^+$, we have $g = \sup_\pi \Psi(\pi)$. To prove the last statement we note that when $\mathbf{A}$ is finite we may take $\varepsilon = 0$. □

Thus we see, that a key to the existence of optimal stationary policies is the existence of a solution g to (2.5), $\alpha g = \sup_a \{r(a) + Q(a)g\}$. We first show that this solution exists when $\mathbf{A}$ is finite and then extend to the case when $\mathbf{A}$ is countable.

3. Policy improvement and convergence. In this section we give a "policy space iterative" procedure which yields a sequence of stationary policies $\pi^0, \pi^1, \pi^2, \cdots$, such that $\Psi(\pi^n)$ converges to some bounded vector g.

Suppose $\mathbf{A}$ is finite. Let π^0 be any stationary policy with return $\Psi(\pi^0)$. For each state $i \in \mathbf{S}$ we define $\pi^1(i)$ to be any action for which

$$\begin{aligned} r(i, \pi^1(i)) + \textstyle\sum_j q_{ij}(\pi^1(i))\Psi(i, \pi^0(i)) \\ = \max_a [r(i, a) + \textstyle\sum_j q_{ij}(a)\Psi(i, \pi^0(i))]. \end{aligned}$$

Then

$$r(\pi^1)+Q(\pi^1)\Psi(\pi^0) \geqq r(\pi^0)+Q(\pi^0)\Psi(\pi^0).$$

But for any stationary policy π we have from Theorem 2.4

$$\alpha\Psi(\pi) = r(\pi)+Q(\pi)\Psi(\pi).$$

Using this relation for π^0 and π^1 the above equation may be written as

$$[\alpha I - Q(\pi^1)](\Psi(\pi^1)-\Psi(\pi^0)) \geqq 0.$$

Taking $g = \Psi(\pi^1)-\Psi(\pi^0)$ it follows from Corollary 2.3 that $\Psi(\pi^1) \geqq \Psi(\pi^0)$. We continue this process, defining a sequence of stationary policies $\pi^0, \pi^1, \pi^2, \cdots$, for which $\Psi(\pi^{n+1}) \geqq \Psi(\pi^n), n = 0, 1, 2, \cdots$. Let, $g = \lim_n \Psi(\pi^n) \leqq \alpha^{-1}\|r\| < \infty$.

LEMMA 3.1. *When* **A** *is finite there exists a bounded solution* g *to*

$$\alpha g = \max_a [r(a)+Q(a)g].$$

PROOF. Let $\{\pi^n\}$ be a sequence of policies obtained by "policy space iterative" procedure described above. Then we have at each stage:

$$\alpha\Psi(\pi^{n+1}) \geqq \max_a [r(a)+Q(a)\Psi(\pi^n)]-\varepsilon(\pi^n),$$

where

$$\varepsilon(\pi^n) = r(\pi^{n+1})+Q(\pi^{n+1})\Psi(\pi^n)-r(\pi^n)-Q(\pi^n)\Psi(\pi^n). \tag{3.1}$$

If $\varepsilon(\pi^n) \to 0$, (the proof of which appears in the following lemma) then using the monotone convergence theorem we get,

$$\alpha g \geqq \max_a[r(a)+Q(a)g]. \tag{3.2}$$

We also have for each n from Theorem 2.4 $\alpha\Psi(\pi^n) = r(\pi^n)+Q(\pi^n)\Psi(\pi^n)$. Since the set of stationary policies is a compact space of all functions from **S** to **A**, there exists a subsequence $\{\pi^{n'}\}$ of $\{\pi^n\}$ such that $\pi^{n'}$ converges to some stationary policy π^*. Since **A** is a finite set, $\pi^{n'} \to \pi^*$ means that for every pair (i, j); $q_{ij}(\pi^{n'}) = q_{ij}(\pi^*)$ for sufficiently large values of n'. Hence we have by taking limits as $n' \to \infty$, $\alpha g = r(\pi^*)+Q(\pi^*)g$. This may be written as,

$$\alpha g \leqq \max_a [r(a)+Q(a)g]. \tag{3.3}$$

The theorem follows from (3.2) and (3.3). □

LEMMA 3.2. *When* **A** *is finite* $\{\pi^n\}$ *is a sequence of policies obtained by the "policy space iterative" procedure then,* $\lim_{n\to\infty} \varepsilon(\pi^n) = 0$.

PROOF. We have $\alpha\{\Psi(\pi^{n+1})-\Psi(\pi^n)\} = r(\pi^{n+1})+Q(\pi^{n+1})\Psi(\pi^{n+1})-r(\pi^n)-Q(\pi^n)\Psi(\pi^n)$. Adding and subtracting $Q(\pi^{n+1})\Psi(\pi^n)$ on the right-hand side and rearranging term, we obtain

$$\alpha\{\Psi(\pi^{n+1})-\Psi(\pi^n)\} = \varepsilon(\pi^n)+Q(\pi^{n+1})\{\Psi(\pi^{n+1})-\Psi(\pi^n)\},$$

where $\varepsilon(\pi^n) \geqq 0$ and given by (3.1). Now applying Theorem 2.4 with g in that theorem being the vector $\Psi(\pi^{n+1}) - \Psi(\pi^n)$ and $r(\pi)$ in that theorem being the vector $\varepsilon(\pi^n)$ we can conclude,

$$\Psi(\pi^{n+1}) - \Psi(\pi^n) = \int_0^\infty e^{-\alpha t} F(t, \pi^{n+1}) \varepsilon(\pi^n)\, dt.$$

But $\Psi(\pi^n)$ converges and $f_{ii}(t, \pi^{n+1}) \geqq e^{-Mt}$ and $f_{ij}(t, \pi^{n+1}) \geqq 0$ for all $i, j, \in \mathbf{S}$. Hence $\varepsilon(\pi^n) \to 0$. □

We are now prepared to state and prove our main theorem.

THEOREM 3.3. (i) $g = \sup_\pi \Psi(\pi)$ *is the unique bounded solution to* (2.5). $\alpha g = \sup_a \{r(a) + Q(a)g\}$; (ii) *If* $\varepsilon > 0$ *there exist* ε*-optimal policies which are stationary and if* $\mathbf{A}$ *is finite there exist stationary optimal policies.*

PROOF. Under the hypothesis that g is a bounded solution to (2.5). Theorem 2.5 states both that $g = \sup_\pi \Psi(\pi)$, hence is unique, and part (ii) of the present theorem. Thus we need only show the existence of a bounded g satisfying $\alpha g = \sup_a \{r(a) + Q(a)g\}$. Let $\mathbf{A}_n$ be the set of actions $\{1, 2, \cdots, n\}$. Let $g(n)$ be a solution to $\alpha g(n) = \max_{a \varepsilon \mathbf{A}_n} \{r(a) + Q(a)g(n)\}$. We have for $n \geqq m$, $\alpha g(n) \geqq \max_{a \varepsilon \mathbf{A}_m} \{r(a) + Q(a)g(n)\}$. Clearly $g(n) \leqq g(n+1) \leqq \alpha^{-1} \|r\|$, and letting $g = \lim_n g(n)$, by the dominated convergence theorem, we obtain

$$\alpha g \geqq \max_{a \varepsilon \mathbf{A}_m} [r(a) + Q(a)g]$$

for $m = 1, 2, \cdots$. Letting $m \to \infty$ we get

$$\alpha g \geqq \sup_a [r(a) + Q(a)g]. \tag{3.4}$$

From Lemma 3.1 there exists $a_n \in \mathbf{A}_n$ such that

$$g_i(n) = r(i, a_n) + \sum_j q_{ij}(a_n) g_j(n), \qquad i \in S. \tag{3.5}$$

Since r and q_{ij} are bounded and the argument is countable there exists a subsequence $\{a_{n'}\}$ of $\{a_n\}$ such that $r(i, a_{n'}) \to r^*(i)$ and $q_{ij}(a_{n'}) \to q_{ij}^*$ for all $i, j \varepsilon S$. Hence from (3.5) we obtain by taking the limit on both sides as $n' \to \infty$

$$g_i = r^*(i) + \sum_j q_{ij}^* g_j, \tag{3.6}$$

for all $i \in \mathbf{S}$. For any given $\varepsilon > 0$, we can find n^* such that

$$|r(i, a_{n^*}) - r^*(i)| \leqq \varepsilon/2$$

and

$$|q_{ij}(a_{n^*}) - q_{ij}^*| \leqq \alpha\varepsilon/(2\|r\| \, \|q_{ii}(\cdot)\|).$$

Using this n^* from (3.6) we obtain,

$$g_i \leqq r(i, a_{n^*}) + \sum_j q_{ij}(a_{n^*}) g_j + \varepsilon \qquad \text{for all } i \varepsilon \mathbf{S}.$$

Since $\varepsilon > 0$ is arbitrary we have

$$g \leqq \sup_a [r(a) + Q(a)g]. \tag{3.7}$$

From (3.4) and (3.7) the theorem follows. □

Combining Theorem 3.1, Theorem 2.2 and Corollary 2.3 we obtain, if a stationary policy π^* is ε-optimal among stationary policies then it is ε-optimal over all Markov policies. When **A** is finite and π^* is optimal among stationary policies then it is optimal over all Markov policies.

4. Some further results. In this section we state some simple but important results concerning the total discounted return $\Psi(\pi)$.

LEMMA 4.1. *If* $\Psi^T(\pi) = \int_0^T e^{-\alpha t} F(t, \pi) r(t, \pi)\, dt$ *then* $\|\Psi(\pi) - \Psi^T(\pi)\| \to 0$ *as* $T \to \infty$ *for any given Markov policy* π.

LEMMA 4.2. *If* $\pi(t)$ *and* $\pi'(t)$ *are such that* $\pi(t) = \pi'(t)$ *for* $t \leqq T$, *then*

$$\|\Psi(\pi) - \Psi(\pi')\| \leqq 2\|r\|\alpha^{-1} e^{-\alpha T}.$$

Let π and π' be two policies, such that for t sufficiently large, it is better to use π up to time t and then switch to π' than to use π' from the beginning. Then it is better to use π forever than to use π' forever. Let $\Psi(\pi^T \pi')$ be the return obtained by using the policy π up to time T and then using π'. Then we have the following:

THEOREM 4.3. *If* π *and* π' *are policies for which there exists a* T *such that*

$$\Psi(\pi^T\, \pi') \geqq \Psi(\pi') \text{ for } t \geqq T, \text{ then } \Psi(\pi) \geqq \Psi(\pi').$$

PROOF. We have from Lemma 4.2

$$\|\Psi(\pi) - \Psi(\pi^t, \pi')\| \leqq 2\|r\|\alpha^{-1} e^{-\alpha t}.$$

By using the hypothesis this may be written as,

$$\Psi(\pi) \geqq \Psi(\pi') - 2\|r\|\alpha^{-1} e^{-\alpha t}, \qquad \text{for } t \geqq T.$$

Now taking the limit as $t \to \infty$ we obtain the required result. □

Acknowledgment. The author acknowledges a great debt of gratitude to his thesis advisor Professor Howard M. Taylor III for his guidance and encouragement during the preparation of the thesis and this paper which is based upon the thesis. He is also indebted to the referee for his helpful comments and suggestions.

REFERENCES

[1] BLACKWELL, D. (1965). Discounted dynamic programming. *Ann. Math. Statist.* **36** 220–235.
[2] CHUNG, K. L. (1967). *Markov Chains With Stationary Transition Probabilities*, 2nd ed. Springer-Verlag, Berlin.
[3] DE LEVE, G. (1964a). *Generalized Markovian Decision Process I. Model and Method.* Mathematical Centre Tracts 3.
[4] DE LEVE, G. (1964b). *Generalized Markovian Decision Process II. Probabilistic Background.* Mathematical Centre Tracts 4.
[5] DYNKIN, E. B. (1961). *Theory of Markov Process*, tr. D. E. Brown from Russian. Prentice-Hall, Englewood Cliffs.

[6] HOWARD, R. A. (1960). *Dynamic Programming and Markov Processes.* Wiley, New York.
[7] KAKUMANU, P. V. (1969). Continuous time Markov decision models with applications to optimization problems. Technical Report 63, Dept. of Operations Research, Cornell Univ.
[8] MARTIN-LÖF. (1967). Optimal control of a continuous-time Markov chain with periodic transition probabilities. *Operations Research* **15** 872–881.
[9] MILLER, B. L. (1968). Finite state continuous-time Markov decision process with applications to a class of optimization problems in queuing theory. *SIAM J. Control* **6** 226–280.
[10] REUTER, G. E. H. and LEDERMANN, W. (1953). On the differential equations for the transition probabilities of Markov processes with enumerably many states. *Proc. Cambridge Philos. Soc.* **49** 247–262.
[11] ROYKOV, V. V. (1966). Markov decision process with finite state and decision space. *Theory Probability Appl.* **11** 302–311.
[12] TAYLOR, H. M. Markov sequential decision process. (Mimeographed.) Cornell Univ.
[13] VEINOTT, A. F., Jr. (1969). Discrete dynamic programming with sensitive discount optimality criteria. *Ann. Math. Statist.* **40** 1635–1660.