The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 2118–2122

ARBITRARY STATE MARKOVIAN DECISION PROCESSES[1]

BY SHELDON M. ROSS[2]

Stanford University

1. Introduction. We are concerned with a process which is observed at times $t = 0, 1, 2, \cdots$ and classified into one of a possible number of states. We let $\mathfrak{X}$ denote the state space of the process. $\mathfrak{X}$ is assumed to be a Borel subset of a complete separable metric space, and we let $\mathfrak{B}$ be the σ-algebra of Borel subsets of $\mathfrak{X}$. After each classification an action must be chosen and we let A, assumed finite, denote the set of all possible actions.

Let $\{X_t ; t = 0, 1, 2, \cdots\}$ and $\{\Delta_t ; t = 0, 1, 2, \cdots\}$ denote the sequence of states and actions; and let $S_{t-1} = (X_0, \Delta_0, \cdots, X_{t-1}, \Delta_{t-1})$. It is assumed that for every $x \varepsilon \mathfrak{X}$, $k \varepsilon A$ there is a known probability measure $P(\cdot \mid x, k)$ on $\mathfrak{B}$ such that, for some version, $P\{X_{t+1} \varepsilon B \mid X_t = x, \Delta_t = k, S_{t-1}\} = P(B \mid x, k)$ for every $B \varepsilon \mathfrak{B}$, and all histories S_{t-1}. It is also assumed that for every $k \varepsilon A$, $B \varepsilon \mathfrak{B}$, $P(B \mid \cdot, k)$ is a Baire function on $\mathfrak{X}$.

Whenever the process is in state x and action k is chosen then a bounded (expected) cost $C(x, k)$—assumed, for fixed k, to be a Baire function in x—is incurred.

A policy R is a set of Baire functions $\{D_k(S_{t-1}, x)\}_{k \varepsilon A}$ satisfying $D_k(S_{t-1}, x) \geqq 0$ for all $k \varepsilon A$, and $\sum_{k \varepsilon A} D_k(S_{t-1}, x) = 1$ for every (S_{t-1}, x). The interpretation being: if at time t the history S_{t-1} has been observed and $X_t = x$ then action k is chosen with probability $D_k(S_{t-1}, x)$. R is said to be stationary if $D_k(S_{t-1}, x) = D_k(x)$ for every S_{t-1}; R is said to be stationary deterministic if $D_k(x)$ equals 0 or 1 for all x, k.

For any policy R, let

$$\varphi(x, R) = \limsup_{n\to\infty} (n + 1)^{-1} \sum_{t=0}^{n} E_R[C(X_t, \Delta_t) \mid X_0 = x].$$

Thus $\varphi(x, R)$ is the expected average cost per unit time when the process starts in state x and policy R is used.

In [4], under the assumption that $\mathfrak{X}$ is denumerable, a number of results dealing with the average cost criterion were proven. The method employed was to treat the average cost problem as a limit (as the discount factor approaches unity) of the discounted cost problem. In this paper we generalize some of these results to arbitrary state spaces.

2. Stationary deterministic optimal policies. The following theorem was originally proven by Derman [2] for the special case that $\mathfrak{X}$ is denumerable. The following proof is new; it makes use of a technique used by Taylor [5].

THEOREM 1. *If there exists a bounded Baire function $f(x)$, $x \varepsilon \mathfrak{X}$ and a constant g,*

Received 27 November 1968.

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

[2] Now at the University of California, Berkeley.