The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 412–423

NON-DISCOUNTED DENUMERABLE MARKOVIAN DECISION MODELS[1]

By Sheldon M. Ross

Stanford University

0. Introduction. We are concerned with a process which is observed at times $t = 0, 1, 2, \cdots$ to be in one of a possible number of states. We let I (assumed denumerable) denote the number of possible states. If at time t the system is observed in state i then one of K_i possible actions must be taken. Unless otherwise noted we shall assume throughout that $K_i < \infty$ for all i.

If the system is in state i at time t and action K is chosen then two things occur:

(i) We incur an expected cost $C(i, K)$ and

(ii) $P\{X_{t+1} = j \mid X_0, \Delta_0, \cdots, X_t = i, \Delta_t = K\} = P(i, j:K)$ where $\{X_r\}_{r=0}^{t+1}$ denotes the sequence of states and $\{\Delta_r\}_{r=0}^{t+1}$ the sequence of decisions up to time $t + 1$.

Thus both the costs and the transition probabilities are functions only of the last state and the subsequently made decision. It is assumed that both the expected costs $C(i, K)$ and the transition probabilities $P(i, j:K)$ are known. Furthermore it is assumed that the expected costs are bounded and we let M be such that $|C(i, K)| < M$ for all i, K.

A rule or policy R for controlling the system is a set of functions $\{D_K(X_0, \Delta_0, \cdots, X_t)\}_{K=1}^{K_{X_t}}$ satisfying

$$0 \leqq D_K(X_0, \Delta_0, \cdots, X_t) \leqq 1, K = 0, 1 \cdots, K_{X_t}$$

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

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

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

We say that a rule R is a stationary if $D_K(X_0, \Delta_0, \cdots, X_t = i) = D_{i,K}$ independent of $X_0, \Delta_0, \cdots, \Delta_{t-1}$ and t. We say that a rule R is stationary deterministic if it is stationary and also $D_{i,K} = 0$, or 1. Thus the stationary deterministic rules are those non-randomized rules whose actions at t just depend on the state at time t. We denote by C'' the class of stationary deterministic rules.

Following Derman [4] the process $\{(X_t, \Delta_t)t = 0, 1, 2, \cdots\}$ will be called a *Markovian decision process.*

Two possible measures of effectiveness of a rule governing a Markovian decision process are the expected total discounted cost and secondly the expected average cost per unit time. The first assumes a discount factor $\beta \varepsilon (0, 1)$ and for

Received 29 May 1967; revised 17 October 1967.

[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.