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 **S** and the action space **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.