On semi-Markov games

By

Kensuke TANAKA and Kazuyoshi WAKUTA

(Received October 25, 1975)

1. Introduction

This paper is a continuation of our paper [4] and is concerned with semi-Markov games with some criterion. In the Markov games, time until the transition from a state to a next state occurs is a unit time, but it does not seem general enough. For this reason, in the paper, we shall consider the semi-Markov game which time until the transition occurs is a known random variable.

However, so far as we know, such the games have not been tried up to the present. Hence, at first, we shall give the formulation of semi-Markov game with the criterion of long-run average reward as the game proceeds over the infinite future. Then, we shall show that the game has a value and there exist the optimal stationary strategies for both players under this criterion and some assumptions. Moreover, we shall give a sufficient condition for some important assumption.

This paper consists of four sections. In Section 2, we shall give the formulation of the problem treated by us in this paper. In Section 3, we shall show the existence of optimal stationary strategies and, in Section 4, we shall give a sufficient condition.

2. The formulation of the problem

In this paper, we determine "*semi-Markov game*" by six objects (S, A, B, q, F, r). Here, S is a non-empty Borel subset of a Polish space, the set of states of a system; A is a non-empty Borel subset of a Polish space, the set of actions available to player I; B is a non-empty Borel subset of a Polish space, the set of actions available to player II; q is the law of motion of the system, it associates Borel measurably with each triple $(s, a, b) \in S \times A \times B$ a probability measure $q(\cdot \mid s, a, b)$ on the Borel measurable space $(S, \mathcal{B}(S))$, where $\mathcal{B}(S)$ is the σ-field generated by the metric on S; $F(\cdot \mid s, a, b, s')$ is a distribution of time until the transition from s to s' occurs, given that the next state is s'; r, the reward function, is a bounded Borel measurable function on $S \times A \times B \times R$, where R is a real line. At successive random times, player I and player II observe the current state s of the system and choose actions a and b, respectively, according to the full knowledge of the history of