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 the system as it has evolved to the present. As a consequence of the actions chosen by the players and the duration time of s, player II pays player I $r(s, a, b, t)$ units of money, and the system moves to a new state s' according to the conditional distribution $q(\cdot|s, a, b)$ after some duration of state s. Then the whole process is repeated from the state s'. Here, our optimization problem is to maximize the limit of expected reward of player I gained during the first n transitions divided by the expected length of the first n transitions as the game proceeds over the infinite future and to minimize the limit of expected loss of player II incurred during the first n transitions divided by the expected length of the first n transitions.

A *strategy* π for player I is a sequence of $\pi_1, \pi_2, \ldots$, where π_n specifies the nth action to be chosen by player I by associating Borel measurably with each history $h_n=(s_1, a_1, b_1, t_1, \ldots, s_{n-1}, a_{n-1}, b_{n-1}, t_{n-1}, s_n)$ of the system a probability distribution $\pi_n(\cdot|h_n)$ on $(A, \mathcal{B}(A))$, where s_i, a_i, b_i, and t_i are the ith state, the ith action chosen by player I, the ith action chosen by player II and the ith duration time, respectively. A strategy π is, particularly, said to be *stationary* if there is a Borel measurable map f from S to P_A, where P_A is the set of all probability measures on $(A, \mathcal{B}(A))$, such that $\pi_n=f$ for all n and in this case, π is denoted by f^∞. Π denotes the class of all strategies for player I. Strategies and stationary strategies for player II are defined analogously. Γ denotes the class of all strategies for player II.

In order to ensure that the transitions do not take place too quickly, we shall need to assume the following:

Assumption 1. There exists $\delta>0$, $\varepsilon>0$ such that

$$\int_S F(\delta|s, a, b, s')\,dq(s'|s, a, b)<1-\varepsilon \tag{2. 1}$$

for all s, a and b.

Definition 1. A strategy π^* is optimal for player 1 if for each $\sigma\in\Gamma$ and $s_1\in S$,

$$\inf_{\sigma\in\Gamma}\sup_{\pi\in\Pi}\varlimsup_{n\to\infty}\frac{E_{\pi,\sigma}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi,\sigma}[\sum_{i=1}^{n} t_i|s_1]} \tag{2. 2}$$

$$\leq\varliminf_{n\to\infty}\frac{E_{\pi^*,\sigma}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi^*,\sigma}[\sum_{i=1}^{n} t_i|s_1]},$$

where $E_{\pi,\sigma}$ denotes the expectation by the pair (π, σ) of strategies for player I and player II.

A strategy σ^* is optimal for player II if for each $\pi\in\Pi$ and $s_1\in S$,

$$\sup_{\pi\in\Pi}\inf_{\sigma\in\Gamma}\underline{\lim}_{n\to\infty}\frac{E_{\pi,\sigma}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi,\sigma}[\sum_{i=1}^{n} t_i \mid s_1]} \tag{2. 3}$$

$$\geq \underline{\lim}_{n\to\infty}\frac{E_{\pi,\sigma^*}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi,\sigma^*}[\sum_{i=1}^{n} t_i \mid s_1]}$$

DEFINITION 2. Semi-Markov game has a value if for each $s_1 \in S$,

$$\inf_{\sigma\in\Gamma}\sup_{\pi\in\Pi}\overline{\lim}_{n\to\infty}\frac{E_{\pi,\sigma}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi,\sigma}[\sum_{i=1}^{n} t_i \mid s_1]} \tag{2. 4}$$

$$=\sup_{\pi\in\Pi}\inf_{\sigma\in\Gamma}\underline{\lim}_{n\to\infty}\frac{E_{\pi,\sigma}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi,\sigma}[\sum_{i=1}^{n} t_i \mid s_1]}.$$

In the case semi-Markov game has a *value*, the quantity

$$\inf_{\sigma\in\Gamma}\sup_{\pi\in\Pi}\overline{\lim}_{n\to\infty}\frac{E_{\pi,\sigma}[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)]}{E_{\pi,\sigma}[\sum_{i=1}^{n} t_i \mid s_1]},$$

as a function on S, is called the value function.

3. Existence of optimal strategies

In this section, we are concerned with the existence of optimal stationary strategies for our semi-Markov game. We have to impose some assumptions on S, A, B, q, F and r to ensure that there exist optimal strategies.

ASSUMPTION 2. (i) A, B and S are compact metric spaces, (ii) whenever $s_n \longrightarrow s_o$, $a_n \longrightarrow a_o$ and $b_n \longrightarrow b_o$, $q(\cdot \mid s_n, a_n, b_n)$ converges weakly to $q(\cdot \mid s_o, a_o, b_o)$.

ASSUMPTION 3. (i) $\int_0^\infty t dF(t \mid s, a, b, s') \equiv \tau(s, a, b, s')$ is a continuous function on $S \times A \times B \times S$, (ii)
$\int_0^\infty r(s, a, b, t) dF(t \mid s, a, b, s') \equiv r(s, a, b, s')$ is a continuous function on $S \times A \times B \times S$.
From these assumptions we can show the following lemma.

LEMMA 3. 1. *$\bar{\tau}(s, a, b)$ and $\bar{r}(s, a, b)$ are bounded continuous functions on $S \times A \times B$, where*

$$\bar{\tau}(s, a, b)=\int_S \tau(s, a, b, s')\, dq(s' \mid s, a, b)$$

and

$$\bar{r}(s, a, b)=\int_S r(s, a, b, s')\, dq(s' \mid s, a, b).$$

PROOF. It holds that

$$\begin{aligned}
&|\bar{\tau}(s_n, a_n, b_n)-\bar{\tau}(s_0, a_0, b_0)| \qquad (3.1)\\
&\leqq \int_S |\tau(s_n, a_n, b_n, s')-\tau(s_0, a_0, b_0, s')|\, dq(s' \mid s_n, a_n, b_n)\\
&+\Big|\int_S \tau(s_0, a_0, b_0, s')\, dq(s' \mid s_n, a_n, b_n)\\
&-\int_S \tau(s_0, a_0, b_0, s')\, dq(s' \mid s_0, a_0, b_0)\Big|.
\end{aligned}$$

Then, from Assumption 2 (i), (ii), Assumption 3 (i) and (3.1), we can prove that $\bar{\tau}(s, a, b)$ is a bounded continuous function on $S\times A\times B$. Similarly, $\bar{r}(s, a, b)$ is a bounded continuous function on $S\times A\times B$.

ASSUMPTION 4. There exist a continuous function $u(s)$ on S and a constant d such that for each $s\in S$,

$$u(s)=\sup_{\mu\in P_A} \inf_{\lambda\in P_B} \{ \bar{r}(s, \mu, \lambda)+\int_S u(s')dq(s' \mid s, \mu, \lambda)-d\bar{\tau}(s, \mu, \lambda)\}, \qquad (3.2)$$

where for each $\mu\in P_A$, $\lambda\in P_B$ and $E\in\mathcal{B}(S)$,

$$\bar{r}(s, \mu, \lambda) \equiv \int_A\int_B \bar{r}(s, a, b)\, d\lambda(b)\, d\mu(a),$$

$$\bar{r}(s, \mu, \lambda) \equiv \int_A\int_B \bar{\tau}(s, a, b)\, d\lambda(b)\, d\mu(a),$$

and

$$q(E \mid s, \mu, \lambda) \equiv \int_A\int_B q(E \mid s, a, b)\, d\lambda(b)\, d\mu(a).$$

Since A and B are compact metric spaces, P_A and P_B, endowded with weak topology, are compact metric spaces.

LEMMA 3.2. *Let $u(s, a, b)$ be a continuous, real-valued function on $S\times A\times B$. Then under Assumption* 2 (i),

$$u(s, \mu, \lambda) = \int_A\int_B u(s, a, b)\, d\lambda(b)\, d\mu(a),$$

$s\in S$, $\mu\in P_A$, $\lambda\in P_B$ is continuous function on $S\times P_A\times P_B$.

LEMMA 3. 3. *Let u be a bounded, continuous function on $X\times Y$, where X is a Borel subset of a polish space and Y is a compact metric space. Then, $u^*(x)=\max_{y\in Y} u(x, y)$ is continuous. Moreover, $u_*(x)=\min_{y\in Y} u(x, y)$ is also continuous.*

LEMMA 3. 4. *Let u be a bounded, continuous function on $X\times Y$, where X is a Borel subset of a Polish space and Y is a compact metric space. Then, there exist Borel measurable maps f and g from X into Y such that $u(x, f(x))=\max_{y\in Y} u(x, y)$ and $u(x, g(x))=\min_{y\in Y} u(x, y)$, $x\in X$.*

The proofs of these lemmas are given in Lemma 2. 1, Lemma 2. 2 and Lemma 2. 3 of [1].

Let $C(S)$ denote the family of all bounded, continuous functions on S. For $u\in C(S)$ we define $\|u\|=\sup_{s\in S}|u(s)|$. Then $(C(S), d)$ is a complete metric space, where $d(u, v)=\|u-v\|$ for each $u, v\in C(S)$.

Now, for $u\in C(S)$ and d in Assumption 4, we define

$$K(s, \mu, \lambda)\equiv \bar{r}(s, \mu, \lambda)+\int_S u(s')dq(s'|s, \mu, \lambda)-d\bar{\tau}(s, \mu, \lambda). \tag{3. 3}$$

Then, by virtue of Lemma 3. 2, $K(s, \mu, \lambda)$ is a continuous function on $S\times P_A\times P_B$. $K(s, \mu, \lambda)$, P_A and P_B satisfy the conditions of Sion's minimax theorem (Theorem 3. 4 of [3]) because of it's bilinearity in (μ, λ) and, consequently,

$$\sup_{\mu\in P_A}\inf_{\lambda\in P_B} K(s, \mu, \lambda)=\inf_{\lambda\in P_B}\sup_{\mu\in P_A} K(s, \mu, \lambda),\ s\in S. \tag{3. 4}$$

Moreover, since $K(s, \mu, \lambda)$ is continuous on $S\times P_A\times P_B$ and P_A, P_B are compact, sup and inf can be replaced by max and min, respectively. Thus, (3. 2) can be written as follows:

$$u(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B} K(s, \mu, \lambda)=\min_{\lambda\in P_B}\max_{\mu\in P_A} K(s, \mu, \lambda),\ s\in S. \tag{3. 5}$$

LEMMA 3. 5. *There exist Borel measurable maps μ_* and λ_* from S into P_A and P_B, respectively, such that*

$$\begin{aligned}\min_{\lambda\in P_B} K(s, \mu_*, \lambda)&=\max_{\mu\in P_A}\min_{\lambda\in P_B} K(s, \mu, \lambda)\\ &=\min_{\lambda\in P_B}\max_{\mu\in P_A} K(s, \mu, \lambda)\\ &=\max_{\mu\in P_A} K(s, \mu, \lambda_*),\ s\in S.\end{aligned} \tag{3. 6}$$

The proof is immediate from Lemma 3. 3 and Lemma 3. 4.

THEOREM 3. 1. *Under the Assumptions* 1, 2, 3, 4, *semi-Markov game has a value and both players have optimal stationary strategies.*

PROOF. By Assumption 4 and Lemma 3. 5, there exists a Borel measurable map μ_* from S into P_A such that

$$u(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\{\bar{r}(s,\mu,\lambda)+\int_S u(s')dq(s'|s,\mu,\lambda)-d\bar{\tau}(s,\mu,\lambda)\}$$

$$=\max_{\mu\in P_A}\min_{\lambda\in P_B}K(s,\mu,\lambda)$$

$$=\min_{\lambda\in P_B}K(s,\mu_*,\lambda).$$

For a stationary strategy $\mu_*^{\infty}=(\mu_*,\mu_*,\dots)$ for player I and any strategy σ for player II, we have

$$E_{\mu_*^{\infty},\sigma}\left[\sum_{t=2}^{n+1}\{u(s_t)-E_{\mu_*^{\infty},\sigma}[u(s_t)|h_{t-1}]\}\right]=0. \tag{3. 7}$$

But, for each t, it holds that

$$E_{\mu_*^{\infty},\sigma}[u(s_t)|h_{t-1}]=\int_S u(s')dq(s'|s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1}) \tag{3. 8}$$

$$=\{\bar{r}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})+\int_S u(s')dq(s'|s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})$$

$$-d\bar{\tau}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})\}-\{\bar{r}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})$$

$$-d\bar{\tau}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})\}\geqq\min_{\lambda\in P_B}\{r(s_{t-1},\mu_*(s_{t-1}),\lambda)$$

$$+\int_S u(s')dq(s'|s_{t-1},\mu_*(s_{t-1}),\lambda)-d\bar{\tau}(s_{t-1},\mu_*(s_{t-1}),\lambda)\}$$

$$-\{\bar{r}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})-d\bar{\tau}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})\}$$

$$=u(s_{t-1})-\{\bar{r}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})-d\bar{\tau}(s_{t-1},\mu_*(s_{t-1}),\lambda_{t-1})\},$$

where λ_{t-1} denotes a probability measure on $(B,\mathcal{B}(B))$ determined by $\sigma_{t-1}(\cdot|h_{t-1})$. Hence, from (3. 7) and (3. 8),

$$0\leqq E_{\mu_*^{\infty},\sigma}\left[\sum_{t=2}^{n+1}\{u(s_t)-(u(s_{t-1})-\bar{r}(s_{t-1},\mu_*,\lambda_{t-1})+d\bar{\tau}(s_{t-1},\mu_*,\lambda_{t-1}))\}\right] \tag{3. 9}$$

$$=E_{\mu_*^{\infty},\sigma}\left[u(s_{n+1})-u(s_1)+\sum_{t=2}^{n-1}(\bar{r}(s_{t-1},\mu_*,\lambda_{t-1})-d\bar{\tau}(s_{t-1},\mu_*,\lambda_{t-1}))\right],$$

or

$$d\leqq\frac{E_{\mu_*^{\infty},\sigma}\left[\sum_{t=2}^{n+1}\bar{r}(s_{t-1},\mu_*,\lambda_{t-1})\right]}{E_{\mu_*^{\infty},\sigma}\left[\sum_{t=2}^{n+1}\bar{\tau}(s_{t-1},\mu_*,\lambda_{t-1})\right]}+\frac{E_{\mu_*^{\infty},\sigma}[u(s_{n+1})-u(s_1)]}{E_{\mu_*^{\infty},\sigma}\left[\sum_{t=2}^{n+1}\bar{\tau}(s_{t-1},\mu_*,\lambda_{t-1})\right]} \tag{3. 10}$$

$$=\frac{E_{\mu_*^\infty,\sigma}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\mu_*^\infty,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]}+\frac{E_{\mu_*^\infty,\sigma}[u(s_{n+1})-u(s_1)]}{E_{\mu_*^\infty,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]}.$$

By Assumption 1, it is easy to see that

$$E_{\mu_*^\infty,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right] \geqq n\varepsilon\delta \longrightarrow \infty \quad \text{as} \quad n \longrightarrow \infty. \tag{3. 11}$$

Since u is bounded, from (3. 10) and (3. 11), we obtain

$$d \leqq \underline{\lim}_{n\to\infty} \frac{E_{\mu_*^\infty,\sigma}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\mu_*^\infty,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]} \qquad \text{for any} \quad \sigma \in \Gamma \tag{3. 12}$$

Thus, from (3. 12), it holds that for any $s_1 \in S$

$$d \leqq \sup_{\pi\in\Pi} \inf_{\sigma\in\Gamma} \underline{\lim}_{n\to\infty} \frac{E_{\pi,\sigma}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\pi,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]}. \tag{3. 13}$$

Similarly, from Assumption 4 and Lemma 3. 5, it holds that, for any $\pi \in \Pi$ and $s_1 \in S$

$$d \geqq \overline{\lim}_{n\to\infty} \frac{E_{\pi,\lambda_*^\infty}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\pi,\lambda_*^\infty}\left[\sum_{i=1}^{n} t_i \mid s_1\right]} \tag{3. 14}$$

$$\geqq \inf_{\sigma\in\Gamma} \sup_{\pi\in\Pi} \overline{\lim}_{n\to\infty} \frac{E_{\pi,\sigma}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\pi,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]}.$$

On the other hand, generally we have

$$\sup_{\pi\in\Pi} \inf_{\sigma\in\Gamma} \underline{\lim}_{n\to\infty} \frac{E_{\pi,\sigma}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\pi,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]} \tag{3. 15}$$

$$\leqq \inf_{\sigma\in\Gamma} \sup_{\pi\in\Pi} \overline{\lim}_{n\to\infty} \frac{E_{\pi,\sigma}\left[\sum_{i=1}^{n} r(s_i, a_i, b_i, t_i)\right]}{E_{\pi,\sigma}\left[\sum_{i=1}^{n} t_i \mid s_1\right]} \qquad \text{for any} \quad s_1 \in S.$$

By (3. 13), (3. 14) and (3. 15), our semi-Markov game has a value d and μ_*^∞ and λ_*^∞ are stationary optimal strategies for player I and player II, respectively. Thus the proof is complete.

4. Sufficient condition

In this section, we shall give the sufficient condition to ensure Assumption 4.

First we define an operator T_α: $C(S)\longrightarrow C(S)$ as follows: for $\alpha>0$

$$(T_\alpha u)(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}[\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\{\bar{r}(s,a,b)$$

$$+\int_S u(s')\,dq(s'\mid s,a,b)\}\,d\lambda(b)\,d\mu(a)]. \tag{4. 1}$$

LEMMA 4. 1. *The operator T_α is a contraction mapping on $C(S)$ for any $\alpha>0$.*

PROOF. It is easily proved that, for any $u, v\in C(S)$, $\|T_\alpha u-T_\alpha v\|$

$$\leqq\max_{\mu\in P_A}\min_{\lambda\in P_B}[\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\|u-v\|\,d\lambda(b)\,d\mu(a)] \tag{4. 2}$$

and, by Assumption 1,

$$\bar{\tau}(s,a,b)\geqq\delta\,\varepsilon. \tag{4. 3}$$

From (4. 2) and (4. 3), we have

$$\|T_\alpha u-T_\alpha v\|\leqq e^{-\alpha\delta\varepsilon}\|u-v\|.$$

This completes the proof.

Since $C(S)$ is a complete metric space, T_α has a unique fixed point in $C(S)$, by virtue of the Banach fixed point theorem. Let u_α^* be the unique fixed point of T_α. Then it holds that, for each $s\in S$,

$$u_\alpha^*(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}[\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\{\bar{r}(s,a,b)$$

$$+\int_S u_\alpha^*(s')\,dq(s'\mid s,a,b)\}\,d\lambda(b)\,d\mu(a)]. \tag{4. 4}$$

Now, fix some state 0 and let

$$f_\alpha(s)=u_\alpha^*(s)-u_\alpha^*(0). \tag{4. 5}$$

From (4. 4) and (4. 5), we have

$$u_\alpha^*(0)+f_\alpha(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}[\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\{\bar{r}(s,a,b)$$

$$+\int_S f_\alpha(s')\,dq(s'\mid s,a,b)\}\,d\lambda(b)\,d\mu(a)$$

$$+u_\alpha^*(0)\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\,d\lambda(b)\,d\mu(a)]. \tag{4. 6}$$

But

$$\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\,d\lambda(b)\,d\mu(a)=1-\alpha\bar{\tau}(s,\mu,\lambda)+\int_A\int_B 0(\alpha)\,d\lambda(b)\,d\mu(a). \tag{4. 7}$$

Hence, from (4. 6) and (4. 7),

$$f_\alpha(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}[\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\{\bar{r}(s,a,b)$$

$$+\int_S f_\alpha(s')\,dq(s'\,|\,s,a,b)\}\,d\lambda(b)\,d\mu(a)-\alpha u_\alpha^*(0)\bar{\tau}(s,\mu,\lambda)$$

$$+u_\alpha^*(0)\sum], \tag{4. 8}$$

where

$$\sum=\int_A\int_B 0(\alpha)\,d\lambda(b)\,d\mu(a).$$

THEOREM 4. 1. *If $\{f_\alpha(s),\ 0<\alpha<c\}$ is a uniformly bounded, equi-continuous family of functions on S for some $0<c<\infty$. Then, Assumption 4 holds.*

PROOF. By Ascoli-Arzela's theorem there exist a sequence $\alpha_\nu\longrightarrow 0$ and a continuous $u(s)$ such that $f_{\alpha_\nu}(s)$ converges uniformly to $u(s)$ on S.

Now we show that $\{\alpha u_\alpha^*(0),\ 0<\alpha<c\}$ is bounded. By virtue of Lemma 3. 2 and Lemma 3. 3, for a fixed point u_α^*, there exist Borel measurable maps μ_* and λ_* from S into P_A and P_B such that for each $s\in S$

$$u_\alpha^*(s)=\int_A\int_B e^{-\alpha\bar{\tau}(s,a,b)}\{\bar{r}(s,a,b)$$

$$+\int_S u_\alpha^*(s')\,dq(s'\,|\,s,a,b)\}\,d\lambda_*(b)\,d\mu_*(a). \tag{4. 9}$$

From (4. 9), it is easy to see that for each $s_1\in S$

$$u_\alpha^*(s_1)=E_{\mu_*^\infty,\lambda_*^\infty}\Big[\sum_{n=1}^{\infty}e^{-\sum_{k=1}^{n}\alpha\bar{\tau}(s_k,a_k,b_k)}\,\bar{r}(s_n,a_n,b_n)\Big]. \tag{4. 10}$$

Then, since $|r|\leqq M$ and $|\bar{\tau}|\geqq\varepsilon\,\delta$, from (4. 10) $|\alpha u_\alpha^*|$ is bounded. Hence, we can require that $\alpha_\nu u_{\alpha_\nu}^*(0)$ converges to d as $\alpha_\nu\longrightarrow 0$ and we can show that $u_\alpha^*(0)\Sigma$ converges uniformly to zero in μ and λ as $\alpha_\nu\longrightarrow 0$. Thus, from (4. 9), we get

$$u(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\{\bar{r}(s,\mu,\lambda)+\int_S u(s')\,dq(s'\,|\,s,\mu,\lambda)-d\bar{\tau}(s,\mu,\lambda)\}.$$

This completes the proof.

References

[1] A. MAITRA and T. PARTHASARATHY, *On stochastic games*, Journ. Opti. Theory and its Appl., 5 (1970), 289–300.

[2] M. ROSS, *Average cost semi-Markov decision processes*, J. appl. Prob., 7 (1970), 649–656.

[3] M. SION, *On general minimax theorems*, Pacific J. Math., 8 (1958), 171–176.

[4] K. TANAKA, S. IWASE and K. WAKUTA, *On Markov games with the expected average reward criterion, Sci. Rep.* Niigata Univ., ser. A. No. 13, (1976), 31–41.