Learning Behavior of Variable-Structure Stochastic Automata in a Three Person Zero-Sum Game

By Kenshiro OKAMURA*, Taiho KANAOKA**, Toshihiko OKADA**
and Shingo TOMITA**

(Received July 1, 1983)

Abstract

This paper investigates the learning behavior of variable-structure stochastic automata in a three person zero-sum game. The game has three variable-structure stochastic automata and a random environment. In the game the players do not possess prior information concerning the payoff matrix and at the end of every play all the players update their own strategies on the basis of the response from the random environment. Under such situations if a payoff matrix satisfies some conditions, it can be shown that the learning behavior of the automata converges to the optimal strategies.

1. Introduction

The learning behavior of variable-structure stochastic automata operating in a random environment has been studied extensively by many authors (1–8). These automata have the capability of learning the desired state with updating their probabilities of actions. Since Chandrasekan and Shen (2) studied the behavior of variable-structure stochastic automata in two person zero-sum games, various papers of automata games have been published.

Lakshmivarahan and Narendra (6) show that the learning behavior of variable-structure stochastic automata converges to the optimal pure strategies when the game matrix has a saddle point.

However, most of the work in competitive games has been limited to two person zero-sum games.

This paper investigates the learning behavior of variable-structure stochastic automata taking part in a three person zero-sum game as the players. In two person zero-sum games, a gain of one player corresponds to a loss of another player, however, in three person zero-sum game this relation is not satisfied and the payoffs of the players affect each other.

In the game the players don't possess prior information concerning the payoff matrix and the available strategies, and during the course of the game all the players update their strategies using their reinforcement on the basis of the response from the environment. At every play the environment responds to the automaton's action by producing a response. Under such situations when a payoff matrix satisfies some con-

* Graduate Student, Electronics Engineering
** Department of Electronics Engineering

ditions given in Section 4, the learning behavior of the automata converges to the optimal strategies.

After a brief introduction to the variable-structure stochastic automaton, the outline of automata game is stated in Section 3. Further, some solutions as the set of optimal strategies in three person zero-sum game are defined and the collective behavior of stochastic automata in the game having the solution is studied in Section 4. Finally, as illustrative examples, some games are simulated on a computer in Section 5.

2. Formulation of Learning Automaton

The definitions associated with a variable-structure stochastic automaton in a random environment and a reinforcement are presented here.

The variable-structure stochastic automaton (*VSSA*) A is defined by the sixtuple

$$A=\{X,\ Y,\ W,\ g,\ \boldsymbol{P}(t),\ T\} \qquad \text{(referring to Fig. 1)} \tag{1}$$

where $X=\{0, 1\}$ is the set of two inputs (0: reward, 1: penalty), $Y=\{y_1, y_2,\ldots, y_r\}$ $(r\geqq 2)$ is the set of r outputs, $W=\{w_1, w_2,\ldots, w_r\}$ is the set of r states, g is the output function $f=g(w_i)$ $(1\leqq i\leqq r)$ which is a one-to-one mapping from the state set to the output set. In this paper the states of the automaton are regarded as identical with the outputs. The vector $\boldsymbol{P}(t)=(p_1(t), p_2(t),\ldots, p_r(t))$ $(\sum_{i=1}^{r} p_i(t)=1)$ is the state probability vector at instant t, where $p_i(t)$ denotes the probability of the choice of the i th state w_i. T defines the reinforcement scheme which generates $\boldsymbol{P}(t+1)$ from $\boldsymbol{P}(t)$. T can be written formally

$$P(t+1)=T[P(t),\ X(t),\ W(t)] \tag{2}$$

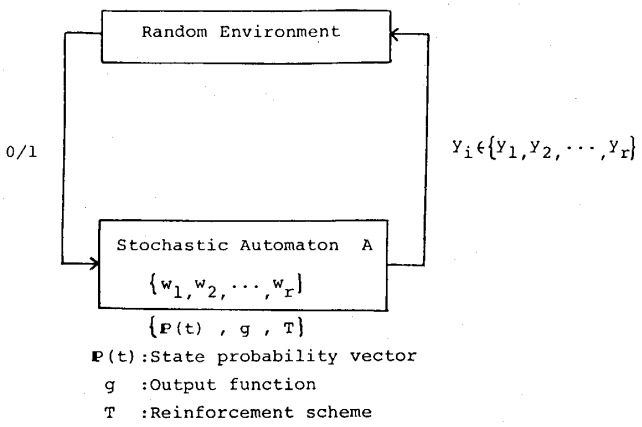

Fig. 1 Variable structure stochastic automaton.

where $X(t)$ and $W(t)$ denote the input and the state of the automaton at instant t, respectively.

A number of reinforcement schemes such as L_{R-I}, N_{R-I}, N_{R-p} have been reported (5), in this paper the linear reward-inaction scheme L_{R-I} which is described in the

following is used.

Let the automaton choose the state w at instant t. If the environment responds with

1) reward (0), then set

$$p_i(t+1)=p_i(t)+\beta(1-p_i(t)), \qquad 0<\beta<1$$
$$p_j(t+1)=(1-\beta)p_j(t), \qquad 1\leqq j\leqq r,\ j\neq i \tag{3}$$

2) penalty (1), then set

$$p_j(t+1)=p_j(t), \qquad 1\leqq j\leqq r \tag{4}$$

where β is a parameter affecting the rate of learning.

The basic idea behind L_{R-I} scheme is following. If A chooses the state w_i at instant t and the environment outputs a reward, the state probability $p_i(t)$ is increased, and the other components of $\boldsymbol{P}(t)$ is decreased so that $\boldsymbol{P}(t)$ is stochastic. For a penalty, $\boldsymbol{P}(t)$ is not changed. Thus, A updates $\boldsymbol{P}(t)$ with L_{R-I} scheme on the basis of the output of A and the response from the environment as much as possible to receive the reward.

In the sense of game theory, we can see the outputs of A as its strategies, the inputs of A as its payoffs and the environment as a referee of a game. With these meaning, we describe the 3-automaton game in the next section.

3. Three Person Zero-sum Game of Automata

In this section a three person zero-sum game of automata is stated. Fig. 2 describes schematically the three-automaton game. The game has 3 *VSSA* as its players and a environment as its referee.

Let $A_l(1\leqq l\leqq 3)$ a player in the game. A_l has r_l outputs (strategies)

$$Y_l=\{y_i^l \mid 1\leqq i\leqq r\} \tag{5}$$

If at an instant each player chooses y_i^1, y_j^2 and y_k^3, respectively. $\{y_1^i, y_j^2, y_k^3\}$ is a play at this time. During the game, such plays are repeated continuously.

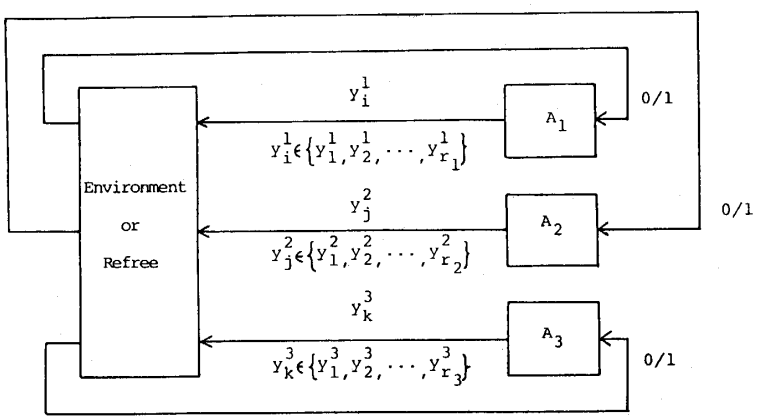

Fig. 2 Automata game.

Next, let the output probability vector (mixed strategy) of player A_l

$$\boldsymbol{P}_l(t)=(p_1^l(t),\ p_2^l(t),\ldots,\ p_{r_l}^l(t)),\ \sum_{i=1}^{r_l} p_i^l(t)=1 \tag{6}$$

where $t(t>0)$ denotes the number of plays, and $p_i^l(t)$ denotes the probability with which A_l chooses the ith output $y_i^{l+)}$.

The environment (referee) determines the payoffs $M^l(i, j, k)$ $(1\leqq l\leqq 3)$ to three players depending on the play $\{y_i^1, y_j^2, y_k^3\}$ and the payoff matrix M. M specifies the payoffs to three players corresponding to $r(r=r_1r_2r_3)$ kinds of plays and is the $r\times 3$ matrix. Then, the environment gives the penalty out to the player A_l in a random manner with the probability

$$C_{i,j,k}^l=\frac{1}{2}(1-M(i, j, k)), \tag{7}$$

where it is also said that the environment gives the reward out with the probability $1-C_{i,j,k}^l$. Each player changes its output probability vector $\boldsymbol{P}^l(t)$ to $\boldsymbol{P}^l(t+1)$ using the reinforcement scheme as much as possible to receive the reward on the basis of the response from the environment. The tth play is done in this way, and this play is made sequentially.

In this three person zero-sum game, all the players possess no prior information concerning the game, and each player chooses its output without knowing the other players' outputs. Therefore, these things mean that the players do not know the payoff matrix and the number of players participating in the game. And the only available information concerning the game for each player is the response from the environment.

4. Collective Behavior of Automata

In this section some solutions of three person zero-sum game and the collective behavior of variable-structure stochastic automata in the game having the solution are stated.

EQUILIBRIUM POINT

We define equilibrium points E_I, E_{II} and E_{III} as the sets of optimal strategies. For these solutions, there must exist a dominant strategy d_I for at least one player. Note that this strategy implies that, for any fixed pure strategies of other players, the payoff of this strategy is greater than the corresponding payoff.

(Definition 1) Equilibrium point E_I
$y_{i_I}^1$ is said to be the dominant strategy d_I of A_1 if

$$^\forall j,\ ^\forall k\ (1\leqq j\leqq r_2,\ 1\leqq k\leqq r_3)$$

$$M(i_I, j, k)\geqq M(i, j, k),\qquad 1\leqq i\leqq r,\ i\neq i_I \tag{8}$$

+) In eq. (1) g is one-ot-one mapping from the set of states to the set of outputs, so states and outputs are regarded synonymous, thus we don't describe the state of A_l, especially.

This definition can be given in the same way for other players. The pure strategies $X_I=(y^1_{i_I}, y^2_{j_I}, y^3_{k_I})$ is said the equilibrium point E_I of the game.

(Definition 2) Equilibrium point E_{II}
When there exists a dominant strategy d_I of A_1 $y^1_{i_I}$, y^2_j is said to be the dominant strategy d_{II} of A_2 if

$$\forall k(1 \leqq k \leqq r_3)$$

$$M(i_I, j_{II}, k) \geqq M(i_I, j, k), \qquad 1 \leqq j \leqq r_2,\ j \neq j_{II} \tag{9}$$

This definition can be given in the same way for other players. The element of the set $X_{II}=\{(y^1_{i_I}, y^2_{j_{II}}, y^3_{k_{II}}), (y^1_{i_{II}}, y^2_{j_I}, y^3_{k_{II}}), (y^1_{i_{II}}, y^2_{j_{II}}, y^3_{k_I})\}$ is said the equilibrium point E_{II} of the game.

(Definition 3) Equilibrium point E_{III}
When there exists a dominant strategy d_I of A_1 $y^1_{i_I}$ and a dominant strategy d_{II} of A_2 $y^2_{j_{II}}$, $y^3_{k_{III}}$ is said to be the dominant strategy d_{III} of A_3 if

$$M(i_I, j_{II}, k_{III}) \geqq M(i_I, j_{II}, k), \qquad 1 \leqq k \leqq r_3,\ k \neq k_{III} \tag{10}$$

This definition can be given in the same way for other players. The element of the set $X_{III}=\{(y^1_{i_I}, y^2_{j_{II}}, y^3_{k_{III}}), (y^1_{i_I}, y^2_{j_{III}}, y^3_{k_{II}}), (y^1_{i_{II}}, y^2_{j_I}, y^3_{k_{III}}), (y^1_{i_{III}}, y^2_{j_I}, y^3_{k_{II}}), (y^1_{i_{II}}, y^2_{j_{III}}, y^3_{k_I}), (y^1_{i_{III}}, y^2_{i_{II}}, y^3_{k_I})\}$ is said the equilibrium point E_{III} of the game.

For these equilibrium points we have the following proposition.

(Proposition 1) Let M_I, M_{II} and M_{III} be the set of the payoff matrix having E_I, E_{II} and E_{III}, respectively. Then

$$M_I \subsetneqq M_{II} \subsetneqq M_{III}.$$

proof. It is clear from Definition 1, 2 and 3. Q.E.D.

In the game having the equilibrium point stated in Definition 1, 2 and 3, when all the players update their own output probability vectors using L_{R-I} schemes on the basis of the response from the environment, we have the following theroems.

(Theorem 1) In a game having the equilibrium point E_I, if all the players update their own output probability vectors using L_{R-I} schemes, the collective behavior of the players converges to the equilibrium point E_I with positive probability.

This implies that $p^l_{i_I}(t)$ in (6) satisfies

$$\forall \varepsilon(\varepsilon > 0)$$

$$\lim_{t \to \infty} p^l_{i_I}(t) > 1-\varepsilon, \qquad 1 \leqq l \leqq 3, \tag{11}$$

where y is a dominant strategy d_I of A_l.

proof. Let $(y^1_{i_I}, y^2_{j_I}, y^3_{k_I})$ be an equilibrium point E_I. The conditional expection of the random variable $p^1_{i_I}$ with respect to $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ is given by

$$E\{p^1_{i_I}(t+1)|\boldsymbol{P}(t)\} = p^1_{i_I}[\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p^2_j p^3_k C^1_{i_I,j,k} p^1_{i_I} + (1-C^1_{i_I,j,k})\{p^1_{i_I}+(1-\beta)p^1_{i_I}]$$

$$+\sum_{\substack{i=1\\ i\neq i_I}}^{r_1}\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p^1_i p^2_j p^3_k\{C^1_{i,j,k}p^1_{i_I}+(1-C_{i,j,k})(1-\beta)p^1_{i_I}\} \quad (12)$$

where for simplicity we put $p^l_i = p^l_i(t)$, and $\beta(0<\beta<1)$ is the parameter of L_{R-I} scheme used by A. In (12) $C^1_{i,j,k}$ denotes the probability that A_1 receives a penalty when each play chooses the output y^1_i, y^2_j and y^3_k, respectively, and from (7) $C^1_{i,j,k}$ is given by

$$C^1_{i,j,k} = \frac{1}{2}(1-M'(i,j,k)). \quad (13)$$

Substituting (13) into (12) yields

$$E\{p^1_{i_I}(t+1)|\boldsymbol{P}(t)\} = \frac{1}{2}\beta p^1_{i_I}\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p^1_i p^2_j p^3_k (M'(i_I,j,k)-M'(i,j,k)) + p^1_{i_I} \quad (14)$$

Since $y^1_{i_I}$ is the dominant strategy d_I of A_1, (8) must be satisfied. Thus, it follows from (8) and (14) that

$\forall t(t>0)$

$$E\{p(t+1)\,|\,\boldsymbol{P}(t)\} - p^1_{i_I}(t)\} \geqq 0, \quad (15)$$

where the equality sign holds only at $p^1_{i_I}(t)=0$ or 1. Noticing that

$\forall t(t>0)$

$$0 \leqq p^1_{i_I}(t) \leqq 1, \qquad 1 \leqq i \leqq r_1$$

from the submartingale theorem (10), it is easily seen that

$\forall \varepsilon(\varepsilon>0)$

$$\lim_{t\to\infty} p^1_{i_I}(t) > 1-\varepsilon \quad (16)$$

For other players it can be shown in the same way. Hence the collective behavior of the players converges to the equilibrium point E_I. Q. E. D.

(Theorem 2) In a game where there exists the equilibrium point E_{II}, if all the players update their own output probability vectors using L_{R-I} schemes, the collective behavior of the players converges to the equilibrium point E_{II} with positive probability.

proof. Let $(y^1_{i_I}, y^2_{j_{II}}, y^3_{k_{II}})$ be an equilibrium point E_{II}. Since there exists the dominant strategy d_I for A_1, $p^1_{i_I}(t)$ converges to unity with positive probability proved in (16). And in this time the conditional expection of the random variable $P^2_{j_{II}}$ is given by

$$E\{p^2_{j_{II}}(t+1)|\boldsymbol{P}(t) = \frac{1}{2}\beta p^2_{j_{II}}\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p^2_j p^3_k (M^2(i_I,j_{II},k)-M^2(i_I,j,k)) + p^2_{j_{II}} \quad (17)$$

Note that $y^2_{j_{II}}$ is the dominant strategy d_{II} of A_2. From (9) and (17)

$^{\forall}t(t>0)$

$$E\{p^2_{j_{II}}(t+1)\,|\,\boldsymbol{P}(t)\}-p^2_{j_{II}}(t)\geqq 0 \tag{18}$$

Hence, from Theorem 1 it can be seen that

$^{\forall}\varepsilon(\varepsilon>0)$

$$\lim_{t\to\infty} p^2_{j_{II}}(t)>1-\varepsilon$$

For A_3 it can be shown in the same way. Q. E. D.

In this case the convergence of A_2 and A_3 is conditioned on the behavior of A_1. This means that A_2 and A_3 behave the conditioned convergence.

(Theorem 3) In a game where there exists the equilibrium point E_{III}, if all the players update their own output probability vectors using L_{R-I} schemes, the collective behavior of the players converges to the equilibrium point E_{III} with positive probability.

proof. The proof is similar to Theorem 2. Q. E. D.

The solutions in Definition 1, 2 and 3 have the dominant strategy d_I and the every payoff of this strategy is greater than the corresponding payoff of all the other strategies. There rarely exists such a strategy in the game situation, so the equilibrium points E_I, E_{II} and E_{III} are the particular solutions.

SADDLE POINT

Adding to the solutions stated in Definition 1, 2 and 3, we define the saddle point in three person zero-sum game as a solution.

(Definition 4) Saddle point

Let

$$\left.\begin{aligned} M(i_1, j_1, k_1)&=\max_{i}.\min_{j,k}.\{M(i, j, k)\}\\ M(i_2, j_2, k_2)&=\max_{j}.\min_{i,k}.\{M(i, j, k)\}\\ M(i_3, j_3, k_3)&=\max_{k}.\min_{j,j}.\{M(i, j, k)\}\end{aligned}\right\} \tag{19}$$

Then, if the relations $i_1=i_2=i_3$, $j_1=j_2=j_3$ and $k_1=k_2=k_3$ are satisfied, we define the pure strategies $S=(y^1_{i_1}, y^2_{j_1}, y^3_{k_1})$ as the saddle point of three person zero-sum game.

The saddle point S is more practical than the equilibrium points E_I, E_{II} and E_{III} in the meaning that all the players behave so as to minimize the payoffs of other two players and maximize the own payoff.

In a game having the saddle point S, when all the players update their own output probability vectors using L_{R-I} schemes, we then have the following theorem.

(Theorem 4) In a game having the saddle point S, if all the players update their own output probability vectors using L_{R-I} schemes with a proper parameter, the collective

behavior of the players converges to the saddle point S with positive probability.

proof. Let $S=(y_a^1, y_b^2, y_c^3)$ be a saddle point. When all the players update their own output probability vectors using L_{R-I} schemes with the same parameter, the conditional expection of the random variable $p_a^1 p_b^2 p_c^3$ is given by

$$E\{p_a^1(t+1)p_b^2(t+1)p_c^3(t+1)\,|\,\boldsymbol{P}(t)\}=\beta p_a^1 p_b^2 p_c^3(H_1+\beta H_2+\beta^2 H_3)+p_a^1 p_b^2 p_c^3, \tag{20}$$

where $H_1=\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p_j^2 p_k^3 \overline{C_{a,j,k}^1}+\sum_{i=1}^{r_1}\sum_{k=1}^{r_3} p_i^1 p_j^2 \overline{C_{i,b,k}^2}+\sum_{i=1}^{r_1}\sum_{j=1}^{r_2} p_i^1 p_j^2 \overline{C_{i,j,c}^3}-\frac{3}{2}$, (21)

and for simplicity we put $\overline{C_{j,j,k}^1}\,\overline{C^2}\,\overline{C^3}=\overline{C_{i,j,k}^1 C_{i,j,k}^2 C_{i,j,k}^3}$ so on.
The derivation of (21) is given in the Appendix.

In (21),

$$\overline{C_{i,j,k}^l}=1-C_{i,j,k}^l=\frac{1}{2}(1+M^l(i, j, k)), \tag{22}$$

where $\overline{C_{i,j,k}^l}$ denotes the probability that A_l receives a reward from the environment when A_1, A_2 and A_3 choose ith output y_i^1, jth output y_j^2 and kth output y_k^3, respectively.

Substituting (22) into (21),

$$H_1=\frac{1}{2}\left\{\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p_j^2 p_k^3 M^1(a, j, k)+\sum_{i=1}^{r_1}\sum_{k=1}^{r_3} p_i^1 p_k^3 M^2(i, b, k)+\sum_{i=1}^{r_1}\sum_{j=1}^{r_2} p_i^1 p_j^2 M^3(i, j, c)\right\} \tag{23}$$

From the assumption that (y_a^1, y_b^2, y_c^3) is the saddle point, it can be seen that

$$\left.\begin{aligned} M^1(a, b, c)&=\min_{j,k}\{M^1(a, j, k)\}\\ M^2(a, b, c)&=\min_{i,k}\{M^2(i, b, k)\}\\ M^3(a, b, c)&=\min_{i,j}\{M^3(i, j, c)\}\end{aligned}\right\} \tag{24}$$

From (23) and (24)

$$\begin{aligned} H_1&\geqq\frac{1}{2}\{M^1(a, b, c)\sum_{j=1}^{r_2}\sum_{k=1}^{r_3} p_j^2 p_k^3+M^2(a, b, c)\sum_{i=1}^{r_1}\sum_{k=1}^{r_3} p_i^1 p_k^3+M^3(a, b, c)\sum_{i=1}^{r_1}\sum_{j=1}^{r_2} p_i^1 p_j^2\}\\ &=\frac{1}{2}\{M(a, b, c)+M(a, b, c)+M(a, b, c)\}\end{aligned} \tag{25}$$

Recall that

$$\forall i, \forall j, \forall k(1\leqq i\leqq r_1, 1\leqq j\leqq r_2, 1\leqq k\leqq r_3)$$

$$\sum_{l=1}^{3} M^l(i, j, k)=0 \tag{26}$$

It follows from (25) and (26) that

$$H_1\geqq 0, \tag{27}$$

where the equality sign holds only at $p_a^1=p_b^2=p_c^3=1$. And it is clearly that H_2 and H_3 are bounded. Thus, being independent of signs of H_2 and H_3, there exists $\beta(0<\beta<1)$ such that

$$H_1+\beta H_2+\beta^2 H_3>0 \tag{28}$$

From (20) and (28), for a proper parameter it is clearly that

$$E\{p_a^1(t+1)p_b^2(t+1)p_c^3(t+1)|\boldsymbol{P}(t)\}-p_a^1(t)p_b^2(t)p_c^3(t)\geqq 0, \tag{29}$$

where the equality sign holds only at $p_a^1=p_b^2=p_c^3=0$ or 1. This completes the proof of the theorem. Q. E. D.

As the well known there is a Nash play defined as a solution of N person zero-sum game. The solutions defined in Definition 1, 2, 3 and 4 are all the Nash plays. Let M_N, M_I, M_{II}, M_{III} and M_s be the sets of the payoff matrices having the Nash play, the equilibrium points E_I, E_{II}, E_{III} and the saddle point, respectively. Then the relation given in Fig. 3 is satisfied. And it is clearly that the solutions stated in Definition 1, 2, 3 and 4 are the optimal plays.

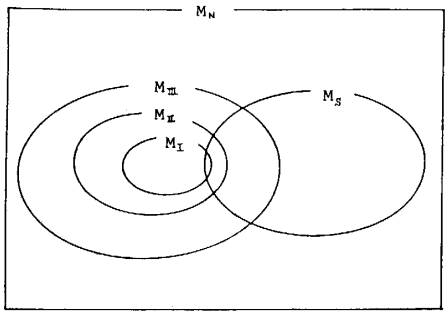

Fig. 3 Relation between solutions.

5. Simulation

To illustrate the collective behavior of the automata in the last section, computer simulations are carried out. In all examples reported below, it is assumed that each player has the set of two outputs {1, 2} and they use L_{R-I} schemes with same parameter β. Each game is played 10 times and the averaged probabilities with which each player chooses the optimal strategy are shown in Fig. 4—Fig. 8.

The payoff matrices used in the simulation are given in Table 1.—Table 4. And it is assumed that each player's payoff is in $[-1, 1]$. The payoff matrix given in Table. 1. has the equilibrium point $E_I(1, 1, 1)$. The behavior of the automata in this game is shown in Fig. 4. The probabilities $p_1^l(t)$ $(1\leqq l\leqq 3)$ with which each automaton chooses the output 1 converge to unity in all 10 experiments.

The payoff matrix Table 2. has the equilibrium point $E_{II}(1, 1, 1)$, and A_1, A_2 and A_3 have the dominant strategy d_I, d_{II} and d_{II}, respectively. The payoff matrix given in Table 3. has the equilibrium point $E_{III}(1, 1, 1)$, and A_1, A_2 and A_3 have the dominant

strategy d_I, d_{II} and d_{III}, respectively. The behavior of the automata in the games having the payoff matrices Table 2. and Table 3. are shown in Fig. 5 and Fig. 6, respectively. In both two games given in Fig. 5 and Fig. 6, the probability p_1^1 with which A_1 chooses the dominant strategy d_I converges to unity fastest and smoothly. And in Fig. 6 the probability p_1^2 with which A_2 chooses the dominant strategy d_{II} converges faster than p_1^3 with which A_3 chooses the dominant strategy d_{III}. As mentioned before dominant strategy d_I always makes the maximum payoff for every fixed outputs of other players, and the existence of the dominant strategy d_{II} needs d_I and the existence of d_{III} needs d_I and d_{II}. From these relations it might be said that the less the influence of other players' behavior is, the faster the convergence is.

The payoff matrix given in Table 4. has the saddle point S (1, 1, 1). The behaviors of the automata using L_{R-I} schemes with the parameter $\beta=0.04$ and 0.08 in the game having Table 4. are shown in Fig. 7 and Fig. 8, respectively. In the case $\beta=0.04$ the behavior of the automata converges to the saddle point S in all 10 experiments. In the case $\beta=0.08$ the automata fail to learn in some experiments. For example, p_1^1 converges to 0.7 in Fig. 8 and this shows that A_1 fails to learn three times. These behaviors coincide with the results described in the last section.

Table 1. Payoff matrix with equilibrium point E_{I}.

y_i^1	y_j^2	y_k^3	M_1	M_2	M_3
1	1	1	0.5	−0.1	−0.4
1	1	2	0.9	0.0	−0.9
1	2	1	0.2	−0.5	0.3
1	2	2	0.8	−0.6	−0.2
2	1	1	0.0	0.1	−0.1
2	1	2	0.1	0.2	−0.3
2	2	1	−0.1	−0.8	0.9
2	2	2	0.0	−0.4	0.4

Table 2. Payoff matrix with equilibrium point E_{II}.

y_i^1	y_j^2	y_k^3	M_1	M_2	M_3
1	1	1	0.4	−0.3	−0.1
1	1	2	0.4	0.0	−0.4
1	2	1	0.3	−0.5	0.2
1	2	2	0.4	−0.4	0.0
2	1	1	−0.3	−0.2	0.5
2	1	2	−0.5	−0.2	0.7
2	2	1	−0.3	−0.2	0.5
2	2	2	−0.1	−0.5	0.6

Table 3. Payoff matrix with equilibrium point E_{III}.

y_i^1	y_j^2	y_k^3	M_1	M_2	M_3
1	1	1	0.2	−0.1	−0.1
1	1	2	0.2	0.2	−0.4
1	2	1	0.5	−0.5	0.0
1	2	2	0.4	−0.6	0.2
2	1	1	−0.2	−0.2	0.4
2	1	2	−0.5	−0.1	0.6
2	2	1	−0.3	0.0	0.3
2	2	2	−0.8	0.0	0.8

Table 4. Payoff matrix with saddle point S.

y_i^1	y_j^2	y_k^3	M_1	M_2	M_3
1	1	1	0.0	0.0	0.0
1	1	2	0.2	0.1	−0.3
1	2	1	0.1	−0.2	0.1
1	2	2	0.1	−0.2	0.1
2	1	1	−0.2	0.1	0.1
2	1	2	−0.5	0.2	0.3
2	2	1	0.3	−0.6	0.3
2	2	2	−0.9	0.4	0.5

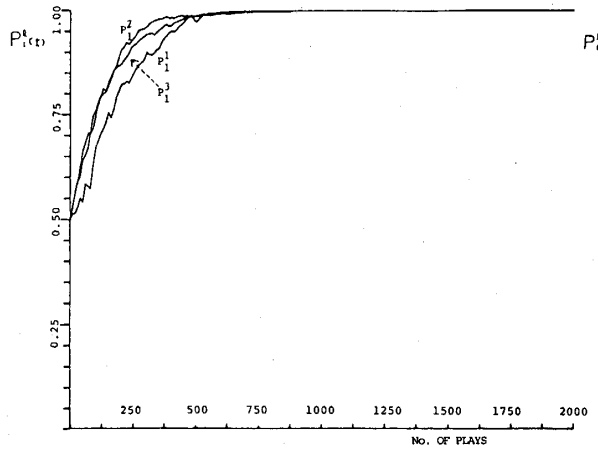

Fig. 4 Behavior of P_1^1, P_1^2 and P_1^3 in the game given by Table. 1 (Average of 10 exepts. with $\beta=0.04$).

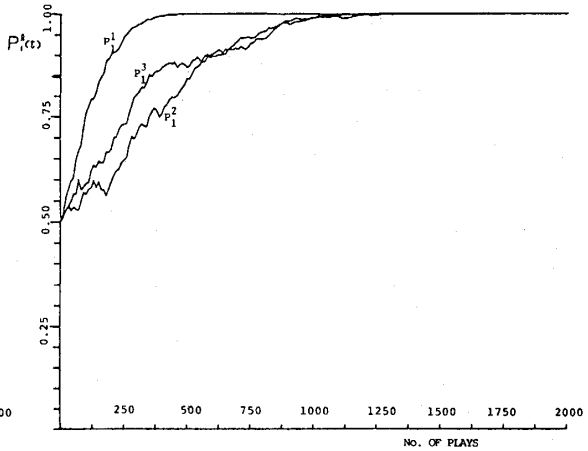

Fig. 5 Behavior of P_1^1, P_1^2 and P_1^3 in the game given by Table. 2 (Average of 10 exepts. with $\beta=0.04$).

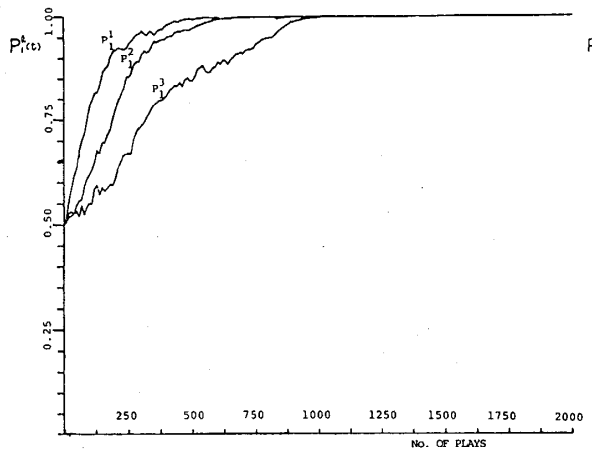

Fig. 6 Behavior of P_1^1, P_1^2 and P_1^3 in the game given by Table. 3 (Average of 10 exepts. with $\beta=0.04$).

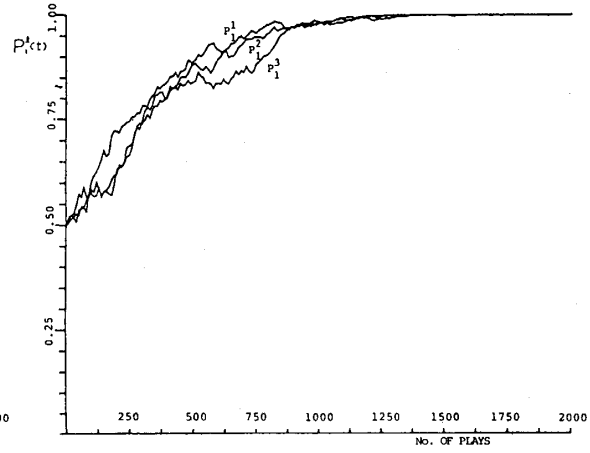

Fig. 7 Behavior of P_1^1, P_1^2 and P_1^3 in the game given by Table. 4 (Average of 10 exepts. with $\beta=0.04$).

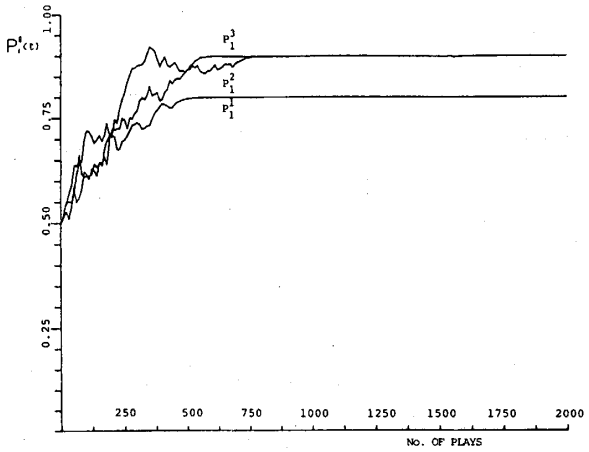

Fig. 8 Behavior of P_1^1, P_1^2 and P_1^3 in the game given by Table. 4 (Average of 10 exepts. with $\beta=0.08$).

6. Conclusions

In this paper, we defined some solutions in three person zero-sum game. And it has been shown that the players having no prior information about the game learn the solution when they update their own output probability vectors using L_{R-I} schemes on the basis of the response from the environment. These results will be easily extended to N person zero-sum game.

The problems stated below are left to study further.

1) The study of cooperative games of automata.
2) What about the scheme that an optimal strategy is a mixed one?

Appendix

Derivation of (21)

Let $S=(y_a^1, y_c^2, y_c^3)$ be a saddle point. When all the players update their own output probability vectors using L_{R-I} schemes with the same parameter β, the conditional expection of the random variable $p_a^1 p_b^2 p_c^3$ is given by

$$
\begin{aligned}
&E\{p_a^1(t+1)p_b^2(t+1)p_c^3(t+1)|\boldsymbol{P}_c^3(t+1)|\boldsymbol{P}(t)\}\\
&=p_a^1p_b^2p_c^3({}^1_{a,b,c}C^2C^3p_a^1p_b^2p_c^3+\overline{C^1_{a,b,c}}CC^3\bar{p}_a^1p_b^2p_c^3+C^1_{a,b,c}\bar{C}^2C^3p_a^1\bar{p}_b^2p_c^3\\
&+C^1_{a,b,c}C^2\bar{C}^3p_a^1p_a^1p_b^2\bar{p}_c^3+\overline{C^1_{a,b,c}}\bar{C}^2C^3\bar{p}_a^1p_b^2p_c^3+\overline{C^1_{a,b,c}}C^2\bar{C}^3\bar{p}_a^1p_b^2p_c^3+C^1_{a,b,c}\bar{C}^2\bar{C}^3p_a^1\bar{p}_b^2\bar{p}_c^3\\
&+\overline{C^1_{a,b,c}}\bar{C}^2\bar{C}^3\bar{p}_a^1\bar{p}_b^2\bar{p}_c^3)+p_a^1p_b^2\sum_{\substack{k=1\\k\neq c}}^{r_3}p_k^3\{C'_{a,b,k}C^2C^3p_a^1p_b^2p_c^3+\overline{C^1_{a,b,k}}C^2C^3\bar{p}_a^1p_b^2p_c^3\\
&+C^1_{a,b,k}\bar{C}^2C^3p_a^1\bar{p}_b^2p_c^3+C^1_{a,b,k}C^2\bar{C}^3p_a^1p_b^2\tilde{p}_c^3+\overline{C^1_{a,b,k}}\bar{C}^2C^3\bar{p}_a^1\bar{p}_b^2p_c^3+C^1_{a,b,k}C^2\bar{C}^3\bar{p}_a^1\tilde{p}_c^3\\
&+C^1_{a,b,k}\bar{C}^2\bar{C}^3p_a^1\bar{p}_b^2\tilde{p}_2^3+\overline{C^1_{a,b,k}}\bar{C}^2\bar{C}^3\bar{p}_a^1\bar{p}_b^2\tilde{p}_c^3\}+p_a^1p_c^3\sum_{\substack{j=1\\j\neq b}}^{r_2}p_j^2\{C^1_{a,j,c}C^2C^3p_a^1p_b^2p_c^3\\
&+C^1_{a,j,c}C^2C^3\bar{p}_a^1p_b^2p_c^3+C^1_{a,j,c}\bar{C}^2C^3p_a^1\tilde{p}_b^2p_c^3+C^1_{a,j,c}C^2\bar{C}^3p_a^1p_b^2\bar{p}_c^3+C^1_{a,j,c}C^2C^3\bar{p}_a^1\tilde{p}_b^2p_c^3\\
&+C^1_{a,j,c}C^2\bar{C}^3\bar{p}_a^1p_b^2\bar{p}_c^3+C^1_{a,j,c}\bar{C}^2\bar{C}^3p_a^1\tilde{p}_b^2\bar{p}_c^3+C^1_{a,j,c}\bar{C}^2\bar{C}^3\bar{p}_a^1\tilde{p}_p^2\bar{p}_c^3\}\\
&+p_b^2p_c^3\sum_{\substack{i=1\\i\neq a}}^{r_1}p_i^1\{C^1_{i,b,c}C^2C^3p_a^1p_b^2p_c^3+\overline{C^1_{i,b,c}}C^2C^3\tilde{p}_a^1p_b^2p_c^3+C^1_{i,b,c}\bar{C}^2C^3p_a^1\bar{p}_b^2p_c^3\\
&+C^1_{i,b,c}C^2\bar{C}^3p_a^1p_b^2\bar{p}_c^3+\overline{C^1_{i,b,c}}C^2C^3\tilde{p}_a^1\bar{p}_b^2p_c^3+\overline{C^1_{i,b,c}}C^2\bar{C}^3p_a^1p_b^2\bar{p}_c^3+C^1_{i,b,c}\bar{C}^2\bar{C}^3p_a^1\bar{p}_b^2\bar{p}_c^3\\
&+C^1_{i,b,c}\bar{C}^2\bar{C}^3\tilde{p}_a^1\bar{p}_b^2\bar{p}_c^3\}+p_a^1\sum_{\substack{j=1\\j\neq b}}^{r_2}\sum_{\substack{k=1\\k\neq c}}^{r_3}p_j^2p_k^3\{C^1_{aj,k}C^2C^3p_a^1p_b^2p_c^3+\overline{C^1_{a,,j,k}}C^2C^3\bar{p}_a^1p_b^2p_c^3\\
&+C^1_{a,j,k}\bar{C}^2C^3p_a^1\tilde{p}_b^2p_c^3+C^1_{a,j,k}C^2\bar{C}^3p_a^1p_b^2\tilde{p}_c^3+\overline{C^1_{a,j,k}}C^2C^3\bar{p}_a^1\tilde{p}_b^2p_c^3+C^1_{a,j,k}C^2\bar{C}^3\bar{p}_a^1p_b^2\tilde{p}_c^3
\end{aligned}
$$

$$+C^1_{a,j,k}\bar{C}^2\bar{C}^3p^1_a\tilde{p}^2_b\tilde{p}^3_c+\overline{C^1_{a,j,k}}C^2\bar{C}^3\bar{p}^1_a\tilde{p}^2_b\tilde{p}^3_c\}+p^2_b\sum_{\substack{i=1\\ i\neq a}}^{r_1}\sum_{\substack{k=1\\ k\neq c}}^{r_2}p^1_ip^3_k\{C^1_{i,b,k}C^2C^3p^1_ap^2_bp^3_c$$

$$+\overline{C^1_{i,b,k}}C^2C^3p^1_ap^2_bp^3_c+C^1_{i,b,k}\bar{C}^2C^3p^1_a\bar{p}^2_bp^3_c+\overline{C^1_{i,b,k}}C^2\bar{C}^3p^1_ap^2_b\bar{p}^3_c+\overline{C^1_{i,b,k}}C^2C^3\tilde{p}^1_a\bar{p}^2_bp^3_c$$

$$+\overline{C^1_{i,b,k}}C^2\bar{C}^3\tilde{p}^1_ap^2_b\tilde{p}^3_c+C^1_{i,b,k}\bar{C}^2\bar{C}^3p^1_a\bar{p}^2_b\tilde{p}^3_c+\overline{C^1_{i,b,k}}\bar{C}^2\bar{C}^3\tilde{p}^1_a\bar{p}^2_b\tilde{p}^3_c\}$$

$$+p^3_c\sum_{\substack{i=1\\ i\neq a}}^{r_1}\sum_{\substack{j=1\\ j\neq b}}^{r_2}p^1_ip^2_j\{C^1_{i,j,c}C^2C^3p^1_ap^2_bp^3_c+\overline{C^1_{i,j,c}}C^2C^3\tilde{p}^1_ap^2_bp^3_c+C^1_{i,j,c}\bar{C}^2C^3p^1_a\tilde{p}^2_bp^3_c$$

$$+C^1_{i,j,c}C^2\tilde{p}^3P^1_aP^2_b\bar{P}^3_c+\overline{C^1_{i,j,c}}\bar{C}^2C^3\tilde{p}^1_a\tilde{p}^2_bp^3_c+\overline{C^1_{i,j,c}}C^2\bar{C}^3\tilde{p}^1_ap^2_b\bar{p}^3_c+\overline{C^1_{i,j,c}}\bar{C}^2\bar{C}^3p^1_a\tilde{p}^2_b\bar{p}^3_c$$

$$+\overline{C^1_{i,j,c}}\bar{C}^2\bar{C}^3\tilde{p}^1_a\tilde{p}^2_b\bar{p}^3_c\}+\sum_{\substack{i=1\\ i\neq a}}^{r_1}\sum_{\substack{j=1\\ j\neq b}}^{r_2}\sum_{\substack{k=1\\ k\neq c}}^{r_3}p^1_ip^2_jp^3_c\{C^1_{i,j,k}C^2C^3p^1_ap^2_bp^3_c+\overline{C^1_{i,j,k}}C^2C^3\tilde{p}^1_ap^2_bp^3_c$$

$$+C^1_{i,j,k}\bar{C}^2C^3p^1_a\tilde{p}^2_bp^3_c+C^1_{i,j,k}C^2\bar{C}^3p^1_ap^2_b\tilde{p}^3_c+\overline{C^1_{i,j,k}}C^2C^3\tilde{p}^1_a\tilde{p}^2_bp^3_c+\overline{C^1_{i,j,k}}C^2\bar{C}^3\tilde{p}^1_ap^2_b\tilde{p}^3_c$$

$$+C^1_{i,j,k}C^2C^3p^1_a\tilde{p}^2_b\tilde{p}^3_c+\overline{C^1_{i,j,k}}C^2C^3\tilde{p}^1_aC^2_b\tilde{p}^3_c\}, \tag{A1}$$

where for simplicity we put $p^l_i=p^l_i(t)$and $C^1_{i,j,k}C^2C^3=C^1_{i,j,k}C^2_{i,j,k}$, so on and

$$\bar{p}^l_i=p^l_i(t)+\beta(1-p^l_i(t)), \tag{A2}$$

$$\tilde{p}^l_i=(1-\beta)p^l_j(t) \tag{A3}$$

Substituting (A2) and (A3) into (A1), let H_1 be the term with β.
Then,

$$H_1=p^1_ap^2_bp^3_c\{\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}\sum_{k=1}^{r_3}p^1_i\ p^2_jp^3_k(C^1_{i,j,k}C^2C^3-\bar{C}^1_{i,j,k}\bar{C}^2C^3-\bar{C}^1_{i,j,k}C^2\bar{C}^3-C^1_{i,j,k}\bar{C}^2\bar{C}^3$$

$$-2\bar{C}^1_{i,j,k}\bar{C}^2\bar{C}^3)+\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}p^1_jp^2_j(\overline{C^1_{i,j,c}}\bar{C}^2\bar{C}^3+C^1_{i,j,c}C^2\bar{C}^3+\bar{C}^1_{i,j,c}\bar{C}^3+C^1_{i,j,c}\bar{C}^2\bar{C}^3)$$

$$+\sum_{i=1}^{r_1}\sum_{k=1}^{r_3}p^1_ip^3_k(\overline{C^1_{i,b,k}}\bar{C}^2C^3+C^1_{i,b,k}\bar{C}^2C^3+\overline{C^1_{i,b,k}}\bar{C}^2C^3+C^1_{i,b,k}\bar{C}^2C^3)$$

$$+\sum_{j=1}^{r_2}\sum_{k=1}^{r_3}p^2_jp^3_k(\overline{C^1_{a,j,k}}\bar{C}^2\bar{C}^3+\overline{C^1_{a,j,k}}C^2\bar{C}^3+\overline{C^1_{a,j,k}}C^2\bar{C}^3+\overline{C^1_{a,j,k}}\bar{C}^2C^3)-1\} \tag{A4}$$

$$=p^1_ap^2_bp^3_c\Big(\sum_{j=1}^{r_2}\sum_{k=1}^{r_2}p^2_jp^3_k\overline{C^1_{a,j,k}}+\sum_{i=1}^{r_1}\sum_{k=1}^{r_2}p^1_ip^3_k\overline{C^2_{i,b,k}}+\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}p^1_ip^2_j\overline{C^3_{i,j,c}}-\frac{3}{2}\Big) \tag{A5}$$

Deleting the term $p^1_ap^2_bp^3_c$ of (A5), we have (21).

References

1) V. I. Varshavskii and I. P. Vorontsova: "On the behavior of stochastic automata with a variable structure", Automation and Remote Control, **24**, pp. 327–333 (1963).
2) B. Chandrasekaran and D. W. C. Shen: "Stochastic automata games", IEEE Trans. Syst. Sci. Cybern., SSC–5, pp. 145–149 (1969).
3) R. Viswanathan and K. S. Narendra: "Games of stochastic automata", IEEE Trans. Syst. Man Cybern., SMC–4, pp. 131–135 (1974).
4) N. Baba and Y. Sawaragi: "On the learning behavior of stochastic automata under a non-stationary random environment", IEEE Trans. Syst. Man Cybern., SMC–6, pp. 756–763 (1976).
5) K. S. Narendra and S. Lakshmivarahan: "Learning automata — a critique", Cybern. and Inf. Sci., **1**, pp. 53–66 (1978).
6) S. Lakshmivarahan and K. S. Narendra: "Learning algorithms for two person zero-sum stochastic game with incomplete information", S & IS Report, No. 7712, Yale Univ., (1978).
7) G. Langholz and E. Katz: "Learning automaton in a three-move zero-sum game", IEEE Trans. Syst. Man Cybern., SMC–9, pp. 304–309 (1979).
8) Y. M. Elfattah: "Stochastic automata modeling of certain problems of collective behavior", IEEE Trans. Cyst. Man Cybern., SMC–10, pp. 304–314 (1980)
9) J. L. Doob: "Stochastic process", John Wiely, New York, (1955).