Some extension of the two-armed bandit problem with finite memory

By

Kensuke TANAKA and Kōichi INADA

(Received May 1, 1972)

1. Introduction and Summary

ROBBINS [10] investigated the two-armed bandit problem with finite memory. In his paper, under the memory r constraint he gave the rule which begins to change coins when sufficient negative information (r consecutive tails) is obtained. But the limiting proportion of heads equal to max $\{p_1, p_2\}$ was not achieved by such a rule. ISBELL [9] improved Robbins' rule as one becomes uniformly for $\{p_1, p_2\}$, and SMITH and PYKE [11] considered the family of rules, containing Isbell's as a special case, which obtained a further improvement. Finally SAMUELS [12] investigated the randomized versions of the rules in order to improve the above rules in each case. None of the above time-invariant rules with finite memory achieved max $\{p_1, p_2\}$.

Afterwards under the assumption of the time-varying decision rule, COVER [1] described a rule achieving max $\{p_1, p_2\}$ with a memory of $r=2$. His rule is that of interleaving trials and tests with trial lengths increasing in such a manner as to swamp out preceding trials and tests.

Here we shall state some extension of his rule. Because it does not seem general enough that these rules were investigated only in the case where two coins exist.

In Section 2 we shall at first describe a rule achieving max $\{p_1, p_2, p_3\}$ with a memory of $r=3$, using three coins. Next, in Section 3 we shall do it, achieving max $\{p_1, p_2, p_3, p_4\}$ with a memory of $r=4$, using four coins.

2. Three-armed bandit problem with finite memory

We are given three coins (coin ①, coin ②, coin ③) with unknown probabilities, $p_1=1-q_1$, $p_2=1-q_2$, and $p_3=1-q_3$ of coming up heads. We shall follow the procedure of interleaving test blocks $T_1, T_2, \cdots$ with trial blocks $U_1, U_2, \cdots$. Each test block T will be begun arbitrarily with coin ① as the favorite. (This precaution yields independence of the test blocks). A test block will be broken into m subblocks each consisting of $3s$ tosses. A subblock test will be said to be a success if $3s$ tosses yield the unbroken sequence of *THH*'s, *TTH*'s or *THT*'s.

At the termination of each subblock, the new favorite coin is used to begin the next subblock until m subblock tests have been performed. A test block consists of this collection of subblocks. Thus $6ms$ tosses of the coin are made in the test block T.

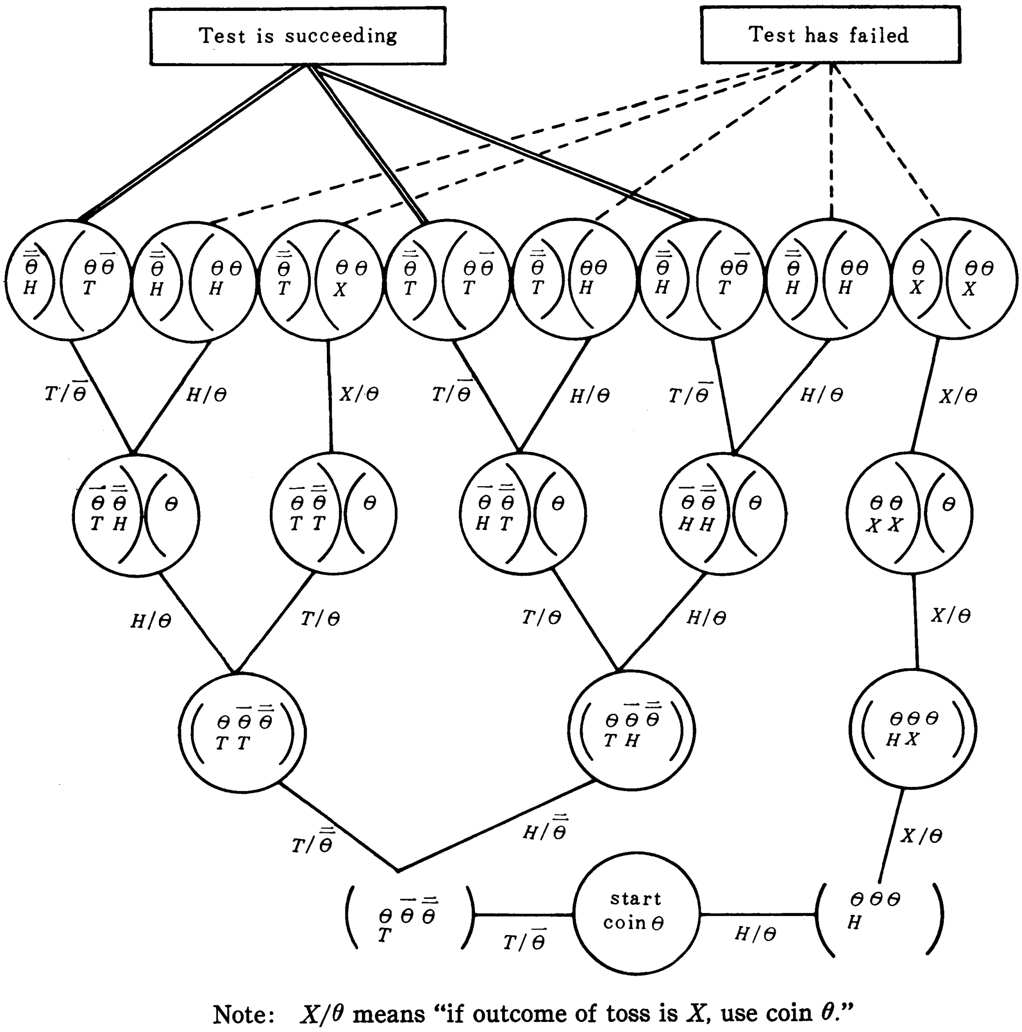
We shall state the details of the test subblock. Let the sequence of coins tossed $(\theta_1, \theta_2, \cdots \theta_{3s})$, $\theta_i \in \{①, ②, ③\}$, and outcomes observed $(x_1, x_2, \cdots, x_{3s})$, $x_i \in \{H, T\}$, be divided into pairs

$$\begin{pmatrix} \theta_1, & \theta_2, & \theta_3 \\ x_1, & x_2 & x_3 \end{pmatrix}\begin{pmatrix} \theta_4, & \theta_5, & \theta_6 \\ x_4, & x_5, & x_6 \end{pmatrix}\cdots\cdots\begin{pmatrix} \theta_{3s-2}, & \theta_{3s-1}, & \theta_{3s} \\ x_{3s-2}, & x_{3s-1}, & x_{3s} \end{pmatrix}.$$

The memory of the past at time n is the state

$$\begin{pmatrix} \theta_{n-2}, & \theta_{n-1}, & \theta_n \\ x_{n-2}, & x_{n-1}, & x_n \end{pmatrix} \text{ or } \left.\begin{matrix} \theta_{n-2}, & \theta_{n-1} \\ x_{n-2}, & x_{n-1} \end{matrix}\right)\begin{pmatrix} \theta_n \\ x_n \end{pmatrix} \text{ or } \left.\begin{matrix} \theta_{n-2} \\ x_{n-2} \end{matrix}\right)\left(\begin{matrix} \theta_{n-1}, & \theta_n \\ x_{n-1}, & x_n \end{matrix}\right.$$

according to $n=3n'$, $n=3n'-1$ or $n=3n'-2$ (n' : positive integer).
Thus the memory is of length $r=3$.



Note: X/θ means "if outcome of toss is X, use coin θ."

Fig. 1. State transition diagram (memory $r=3$).

Figure 1 exhibits a strategy for the coin selection as a function of the state.

In figure 2 we shall give the explicit description of this rule in which the details of the orderly transition from the current favorite coin to the new favorite coin are made clear.

coin ① : $p_1=1-q_1$ coin ② : $p_2=1-q_2$ coin ③ : $p_3=1-q_3$

①②③	①②③	①②③		
(THH)	(THH)······	(THH)	①→②	$(q_1\ p_2\ p_3)^s/2$
			↘③	$(q_1\ p_2\ p_3)^s/2$
(TTH)	(TTH)······	(TTH)	①→③	$(q_1\ q_2\ p_3)^s$
(THT)	(THT)······	(THT)	①→②	$(q_1\ p_2\ q_3)^s$
		otherwise	①→①	$1-\{(q_1\ p_2\ p_3)^s+(q_1\ q_2\ p_3)^s+(q_1\ p_2\ q_3)^s\}$
②③①	②③①	②③①		
(THH)	(THH)······	(THH)	②→①	$(q_2\ p_3\ p_1)^s/2$
			↘③	$(q_2\ p_3\ p_1)^s/2$
(TTH)	(TTH)······	(TTH)	②→①	$(q_2\ q_3\ p_1)^s$
(THT)	(THT)······	(THT)	②→③	$(q_2\ p_3\ q_1)^s$
		otherwise	②→②	$1-\{(p_1\ q_2\ p_3)^s+(p_1\ q_2\ q_3)^s+(q_1\ q_2\ p_3)^s\}$
③①②	③①②	③①②		
(THH)	(THH)······	(THH)	③→①	$(q_3\ p_1\ p_2)^s/2$
			↘②	$(q_3\ p_1\ p_2)^s/2$
(TTH)	(TTH)······	(TTH)	③→②	$(q_3\ q_1\ p_2)^s$
(THT)	(THT)······	(THT)	③→①	$(q_3\ p_1\ q_2)^s$
		otherwise	③→③	$1-\{(p_1\ p_2\ q_3)^s+(q_1\ p_2\ q_3)^s+(p_1\ q_2\ q_3)^s\}$

Fig. 2. Coin transition in the test subblock.

Let M be the coin transition probability matrix in which P_{ij} is the transition probability from the current favorite coin ⓘ to the new favorite coin ⓙ ($i, j=1, 2, 3$).

$$(2.1)\qquad M=\begin{pmatrix} P_{11}, & P_{12}, & P_{13} \\ P_{21}, & P_{22}, & P_{23} \\ P_{31}, & P_{32}, & P_{33} \end{pmatrix}$$

$$=\begin{pmatrix} 1-\{(q_1p_2p_3)^s+(q_1q_2p_3)^s+(q_1p_2q_3)^s\}, & (q_1p_2p_3)^s/2+(q_1p_2q_3)^s, & (q_1p_2p_3)^s/2+(q_1q_2p_3)^s \\ (p_1q_2p_3)^s/2+(p_1q_2p_3)^s, & 1-\{(p_1q_2p_3)^s+(p_1q_2q_3)^s+(q_1q_2p_3)^s\}, & (p_1q_2q_3)^s/2+(q_1q_2p_3)^s \\ (p_1p_2q_3)^s/2+(p_1q_2q_3)^s, & (p_1p_2q_3)^s/2+(q_1p_2q_3)^s, & 1-\{(p_1p_2q_3)^s+(q_1p_2q_3)^s+(p_1q_2q_3)^s\} \end{pmatrix}$$

And let P_i be the stationary probability of coin ⓘ being the favorite ($i=1, 2, 3$). In order to calculate P_i, we at first consider the following

(2.2)
$$\begin{aligned} P_1 &= P_1P_{11}+P_2P_{21}+P_3P_{31} \\ P_2 &= P_1P_{12}+P_2P_{22}+P_3P_{32} \\ P_3 &= P_1P_{13}+P_2P_{23}+P_3P_{33} \\ P_1&+P_2+P_3=1. \end{aligned}$$

From (2.1) and (2.2), we obtain

(2.3)
$$\begin{aligned} P_1 &= \frac{1}{1+a/c+b/c} \\ P_2 &= \frac{1}{1+b/a+c/a} \\ P_3 &= \frac{1}{1+a/b+c/b} \end{aligned}$$

where

(2.4)
$$\begin{aligned} a &= P_{12}P_{31}+P_{12}P_{32}+P_{13}P_{32} \\ b &= P_{12}P_{23}+P_{13}P_{21}+P_{13}P_{23} \\ c &= P_{21}P_{31}+P_{21}P_{32}+P_{23}P_{31}. \end{aligned}$$

Thus by using (2.4) we have

(2.5)
$$\begin{aligned} a &= (p_2q_1q_3)^s\Big\{\frac{3}{4}(p_1p_2p_3)^s+(p_2p_3q_1)^s+(p_1p_3q_2)^s+(p_1p_2q_3)^s+(p_2q_1q_3)^s \\ &\qquad +(p_1q_2q_3)^s+(p_3q_1q_2)^s\Big\} \\ b &= (p_3q_1q_2)^s\Big\{\frac{3}{4}(p_1p_2p_3)^s+(p_2p_3q_1)^s+(p_1p_3q_2)^s+(p_1p_2q_3)^s+(p_2q_1q_3)^s \\ &\qquad +(p_1q_2q_3)^s+(p_3q_1q_2)^s\Big\} \\ c &= (p_1q_2q_3)^s\Big\{\frac{3}{4}(p_1p_2p_3)^s+(p_2p_3q_1)^s+(p_1p_3q_2)^s+(p_1p_2q_3)^s+(p_2q_1q_3)^s \\ &\qquad +(p_1q_2q_3)^s+(p_3q_1q_2)^s\Big\} \end{aligned}$$

and then by using (2.5)

(2.6)
$$\begin{aligned} \frac{a}{c} &= \left(\frac{q_1p_2}{p_1q_2}\right)^s \\ \frac{b}{c} &= \left(\frac{q_1p_3}{p_1q_3}\right)^s \\ \frac{b}{a} &= \left(\frac{q_2p_3}{p_2q_3}\right)^s \end{aligned}$$

Substituting (2. 6) into (2. 3), P_1, P_2 and P_3 will be calculated.

That is, the stationary probabilities of coin ①, coin ② and coin ③ are then $1/(1+\alpha^s+\beta^s)$, $\alpha^s/(1+\alpha^s+\beta^s)$ and $\beta^s/(1+\alpha^s+\beta^s)$ respectively, where $\alpha=q_1p_2/p_1q_2$ and $\beta=q_1p_3/p_1q_3$. It is assured that $\beta\leqq\alpha<1$, if and only if, $p_1>p_2\geqq p_3$.

And we shall note the stationary probabilities of deciding on the first inferior coin and the second inferior coin in terms of t and t' respectively. By the circulation of the stationary probabilities in p_1, p_2 and p_3, we may, without loss of generality, assume $p_1>p_2\geqq p_3$. Thus $\beta\leqq\alpha<1$ and

$$t=\beta^s/(1+\alpha^s+\beta^s)<\beta^s<1$$

$$t'=\alpha^s/(1+\alpha^s+\beta^s)<\alpha^s<1.$$

Let $t_i(t'_i)$ be the probability of selecting the first (second) inferior coin in the test block T_i. Clearly $t_i(t'_i)$ depends on α, m_i, s_i and approaches $t(t')$ as $m_i\longrightarrow\infty$. First we shall choose $\{s_i\}$ so that $\sum\beta^{s_i}<\infty$ and $\sum\alpha^{s_i}<\infty$, e.g., $s_i=i$. We then choose $\{m_i\}$ large enough so that, for each $\beta\leqq\alpha<1$, $t_i<\beta^{s_i}$ and $t'_i<\alpha^{s_i}$ for sufficiently large m_i. This may easily be done. Thus it is assured that $\sum t_i<\infty$ and $\sum t'_i<\infty$.

As the block tests have been made independent, for the sake of the argument, by letting coin ① be the favorite at the beginning of each test block, thus ignoring all the previous information, we may conclude, from the Borel zero-one law and the finiteness of $\sum t_i$ and $\sum t'_i$, that with probability one only a finite number of test blocks T_i will result in an incorrect choice of coin.

Thus if we now use, for the next u_i trials, the coin chosen by block T_i, we may conclude that the proportion of heads obtained in the first n trials tends to $\max\{p_1, p_2, p_3\}$, with probability one, as $n\longrightarrow\infty$. This, of course, is the maximum achievable limit.

3. Four-armed bandit problem with finite memory

In this section four coins (coin ①, coin ②, coin ③, coin ④) are given with unknown probabilities, $p_1=1-q_1$, $p_2=1-q_2$, $p_3=1-q_3$ and $p_4=1-q_4$ of coming up heads. We shall follow the same procedure as used in Section 2.

In figure 3 we shall give the details of the orderly transition from the current favorite coin to the new favorite coin in the test subblock.

①②③④	①②③④	
(T T T H)······	(T T T H)	①→④
(T T H T)······	(T T H T)	①→③
(T H T T)······	(T H T T)	①→②
(T T H H)······	(T T H H)	①→③ ↘④
(T H T H)······	(T H T H)	①→② ↘④
(T H H T)······	(T H H T)	①→② ↘③
(T H H H)······	(T H H H)	①→② ↘③ ↘④
	otherwise	①→①

②③④①	②③④①	
(T T T H)······	(T T T H)	②→①
(T T H T)······	(T T H T)	②→④
(T H T T)······	(T H T T)	②→③
(T T H H)······	(T T H H)	②→① ↘④
(T H T H)······	(T H T H)	②→① ↘③
(T H H T)······	(T H H T)	②→③ ↘④
(T H H H)······	(T H H H)	②→① ↘③ ↘④
	otherwise	②→②

③④①②	③④①②	
(T T T H)······	(T T T H)	③→②
(T T H T)······	(T T H T)	③→①
(T H T T)······	(T H T T)	③→④
(T T H H)······	(T T H H)	③→① ↘②
(T H T H)······	(T H T H)	③→② ↘④
(T H H T)······	(T H H T)	③→① ↘④
(T H H H)······	(T H H H)	③→① ↘② ↘④
	otherwise	③→③

④①②③	④①②③	
(T T T H)······	(T T T H)	④→③
(T T H T)······	(T T H T)	④→②
(T H T T)······	(T H T T)	④→①
(T T H H)······	(T T H H)	④→② ↘③
(T H T H)······	(T H T H)	④→① ↘③
(T H H T)······	(T H H T)	④→① ↘②
(T H H H)······	(T H H H)	④→① ↘② ↘③
	otherwise	④→④

Fig. 3. Coin transition in the test subblock.

Let N be the coin transition probability matrix in which P_{ij} is the transition probability from the current favorite coin ⓘ to the new favorite coin ⓙ ($i, j = 1, 2, 3, 4$).

$$N = \begin{pmatrix} P_{11}, P_{12}, P_{13}, P_{14} \\ P_{21}, P_{22}, P_{23}, P_{24} \\ P_{31}, P_{32}, P_{33}, P_{34} \\ P_{41}, P_{42}, P_{43}, P_{44} \end{pmatrix} \tag{3.1}$$

where

$$
\begin{aligned}
P_{11}&=1-\{(q_1p_2p_3p_4)^s+(q_1p_2p_3q_4)^s+(q_1p_2q_3p_4)^s+(q_1q_2p_3p_4)^s+(q_1p_2q_3q_4)^s\\
&\qquad+(q_1q_2p_3q_4)^s+(q_1q_2q_3p_4)^s\}\\
P_{12}&=(q_1p_2q_3q_4)^s+\frac{1}{2}\{(q_1p_2q_3p_4)^s+(q_1p_2p_3q_4)^s\}+\frac{1}{3}(q_1p_2p_3p_4)^s\\
P_{13}&=(q_1q_2p_3q_4)^s+\frac{1}{2}\{(q_1q_2p_3p_4)^s+(q_1p_2p_3q_4)^s\}+\frac{1}{3}(q_1p_2p_3p_4)^s\\
P_{14}&=(q_1q_2q_3p_4)^s+\frac{1}{2}\{(q_1q_2p_3p_4)^s+(q_1p_2q_3p_4)^s\}+\frac{1}{3}(q_1p_2p_3p_4)^s\\
P_{21}&=(p_1q_2q_3q_4)^s+\frac{1}{2}\{(p_1q_2q_3p_4)^s+(p_1q_2p_3q_4)^s\}+\frac{1}{3}(p_1q_2p_3p_4)^s\\
P_{22}&=1-\{(p_1q_2q_3q_4)^s+(q_1q_2q_3p_4)^s+(q_1q_2p_3q_4)^s+(p_1q_2q_3p_4)^s+(p_1q_2p_3q_4)^s\\
&\qquad+(q_1q_2p_3p_4)^s+(p_1q_2p_3p_4)^s\}\\
P_{23}&=(q_1q_2p_3q_4)^s+\frac{1}{2}\{(p_1q_2p_3q_4)^s+(q_1q_2p_3p_4)^s\}+\frac{1}{3}(p_1q_2p_3p_4)^s\\
P_{24}&=(q_1q_2q_3p_4)^s+\frac{1}{2}\{(p_1q_2q_3p_4)^s+(q_1q_2p_3p_4)^s\}+\frac{1}{3}(p_1q_2p_3p_4)^s\\
P_{31}&=(p_1q_2q_3q_4)^s+\frac{1}{2}\{(p_1p_2q_3q_4)^s+(p_1q_2q_3p_4)^s\}+\frac{1}{3}(p_1p_2q_3p_4)^s\\
P_{32}&=(q_1p_2q_3q_4)^s+\frac{1}{2}\{(p_1p_2q_3q_4)^s+(q_1p_2q_3p_4)^s\}+\frac{1}{3}(p_1p_2q_3p_4)^s\\
P_{33}&=1-\{(q_1p_2q_3q_4)^s+(p_1q_2q_3q_4)^s+(q_1q_2q_3p_4)^s+(p_1p_2q_3q_4)^s+(q_1p_2q_3p_4)^s\\
&\qquad+(p_1q_2q_3p_4)^s+(p_1p_2q_3p_4)^s\}\\
P_{34}&=(q_1q_2q_3p_4)^s+\frac{1}{2}\{(q_1p_2q_3p_4)^s+(p_1q_2q_3p_4)^s\}+\frac{1}{3}(p_1p_2q_3p_4)^s\\
P_{41}&=(p_1q_2q_3q_4)^s+\frac{1}{2}\{(p_1q_2p_3q_4)^s+(p_1p_2q_3q_4)^s\}+\frac{1}{3}(p_1p_2p_3q_4)^s\\
P_{42}&=(q_1p_2q_3q_4)^s+\frac{1}{2}\{(p_1q_2p_3q_4)^s+(p_1p_2q_3q_4)^s\}+\frac{1}{3}(p_1p_2p_3q_4)^s\\
P_{43}&=(q_1q_2p_3q_4)^s+\frac{1}{2}\{(q_1p_2p_3q_4)^s+(p_1q_2p_3q_4)^s\}+\frac{1}{3}(p_1p_2p_3q_4)^s\\
P_{44}&=1-\{(q_1q_2p_3q_4)^s+(q_1p_2q_3q_4)^s+(p_1q_2q_3q_4)^s+(q_1p_2p_3q_4)^s+(p_1q_2p_3q_4)^s\\
&\qquad+(p_1p_2q_3q_4)^s+(p_1p_2p_3q_4)^s\}
\end{aligned}
\tag{3.2}
$$

And let P_i be the stationary probability of coin ⓘ being the favorite ($i=1, 2, 3, 4$). In order to calculate P_i, we at first consider the following

$$\mathrm{P}_i = \sum_{j=1}^{4} \mathrm{P}_j \mathrm{P}_{ji} \quad (i=1, 2, 3, 4)$$

(3. 3)

$$\mathrm{P}_1 + \mathrm{P}_2 + \mathrm{P}_3 + \mathrm{P}_4 = 1$$

From (3. 1), (3. 2) and (3. 3) we obtain

$$\mathrm{P}_1 = \frac{A_1}{A_1 + A_2 + A_3 + A_4}$$

$$\mathrm{P}_2 = \frac{A_2}{A_1 + A_2 + A_3 + A_4}$$

(3. 4)
$$\mathrm{P}_3 = \frac{A_3}{A_1 + A_2 + A_3 + A_4}$$

$$\mathrm{P}_4 = \frac{A_4}{A_1 + A_2 + A_3 + A_4}$$

where

$$A_1 = (p_1 q_2 q_3 q_4)^s A$$

(3. 5)
$$A_2 = (q_1 p_2 q_3 q_4)^s A$$

$$A_3 = (q_1 q_2 p_3 q_4)^s A$$

$$A_4 = (q_1 q_2 q_3 p_4)^s A$$

and

$$\begin{aligned}
\mathrm{A} = {} & \frac{16}{27}(p_1^2 p_2^2 p_3^2 p_4^2)^s + \frac{4}{3}\{(p_1^2 p_2^2 p_3^2 p_4 q_4)^s + (p_1^2 p_2^2 p_3 p_4^2 q_3)^s + (p_1^2 p_2 p_3^2 p_4^2 q_2)^s \\
& + (p_1 p_2^2 p_3^2 p_4^2 q_1)^s\} + \frac{55}{18}\{(p_1^2 p_2^2 p_3 p_4 q_3 q_4)^s + (p_1^2 p_2 p_3^2 p_4 q_2 q_4)^s + (p_1^2 p_2 p_3 p_4^2 q_2 q_3)^s \\
& + (p_1 p_2^2 p_3^2 p_4 q_1 q_4)^s + (p_1 p_2^2 p_3 p_4^2 q_1 q_3)^s + (p_1 p_2 p_3^2 p_4^2 q_1 q_2)^s\} \\
& + \frac{27}{4}\{(p_1^2 p_2 p_3 p_4 q_2 q_3 q_4)^s + (p_1 p_2^2 p_3 p_4 q_1 q_3 q_4)^s + (p_1 p_2 p_3^2 p_4 q_1 q_2 q_4)^s \\
& + (p_1 p_2 p_3 p_4^2 q_1 q_2 q_3)^s\} + \frac{7}{4}\{(p_1^2 p_2^2 p_3 q_3 q_4^2)^s + (p_1^2 p_2 p_3^2 q_2 q_4^2)^s + (p_1^2 p_2^2 p_4 q_3^2 q_4)^s \\
& + (p_1^2 p_2 p_4^2 q_2 q_3^2)^s + (p_1^2 p_3^2 p_4 q_2^2 q_4)^s + (p_1^2 p_3 p_4^2 q_2^2 q_3)^s + (p_1 p_2^2 p_3^2 q_1 q_4^2)^s \\
& + (p_1 p_2^2 p_4^2 q_1 q_3^2)^s + (p_1 p_3^2 p_4^2 q_1 q_2^2)^s + (p_2 p_3^2 p_4^2 q_1^2 q_2)^s + (p_2^2 p_3^2 p_4 q_1^2 q_4)^s \\
& + (p_2^2 p_3 p_4^2 q_1^2 q_3)^s\} + \frac{15}{4}\{(p_1^2 p_2 p_3 q_2 q_3 q_4^2)^s + (p_1^2 p_2 p_4 q_2 q_3^2 q_4)^s + (p_1^2 p_3 p_4 q_2^2 q_3 q_4)^s \\
& + (p_1 p_2^2 p_3 q_1 q_3 q_4^2)^s + (p_1 p_2^2 p_4 q_1 q_3^2 q_4)^s + (p_2^2 p_3 p_4 q_1^2 q_3 q_4)^s + (p_1 p_2 p_3^2 q_1 q_2 q_4^2)^s \\
& + (p_1 p_3^2 p_4 q_1 q_2^2 q_4)^s + (p_2 p_3^2 p_4 q_1^2 q_2 q_4)^s + (p_1 p_2 p_4^2 q_1 q_2 q_3^2)^s + (p_1 p_3 p_4^2 q_1 q_2^2 q_3)^s
\end{aligned}$$

$$+(p_2p_3p_4^2q_1^2q_2q_3)^s\} + \{(p_1^2p_2^2q_3^2q_4^2)^s+(p_1^2p_3^2q_2^2q_4^2)^s+(p_1^2p_4^2q_1^2q_3^2)^s$$

$$+(p_2^2p_3^2q_1^2q_4^2)^s+(p_3^2p_4^2q_1^2q_2^2)^s+(p_2^2p_4^2q_1^2q_3^2)^s\} +2\{(p_1^2p_2q_2q_3^2q_4^2)^s$$

$$+(p_1^2p_3q_2^2q_3q_4^2)^s+(p_1^2p_4q_2^2q_3^2q_4)^s+(p_1p_2^2q_1q_3^2q_4^2)^s+(p_2^2p_3q_1^2q_3q_4^2)^s$$

$$+(p_2^2p_4q_1^2q_3^2q_4)^s+(p_1p_3^2q_1q_2^2q_4^2)^s+(p_2p_3^2q_1^2q_2q_4^2)^s+(p_3^2p_4q_1^2q_2^2q_4)^s$$

$$+(p_1p_4^2q_1q_2^2q_3^{?})^s+(p_2p_4^2q_1^2q_2q_3^2)^s+(p_3p_4^2q_1^2q_2^2q_3)^s\} + \{(p_1^2q_2^2q_3^2q_4^2)^s$$

$$+(p_2^2q_1^2q_3^2q_4^2)^s+(p_3^2q_1^2q_2^2q_4^2)^s+(p_4^2q_1^2q_2^2q_3^2)^s\} +6\{(p_1p_2p_3q_1q_2q_3q_4^2)^s$$

$$+(p_1p_2p_4q_1q_2q_3^2q_4)^s+(p_1p_3p_4q_1q_2^2q_3q_4)^s+(p_2p_3p_4q_1^2q_2q_3q_4)^s\} +2\{(p_1p_2q_1q_2q_3^2q_4^2)^s$$

$$+(p_1p_3q_1q_2^2q_3q_4^2)^s+(p_1p_4q_1q_2^2q_3^2q_4)^s+(p_2p_3q_1^2q_2q_3q_4^2)^s+(p_2p_4q_1^2q_2q_3^2q_4)^s$$

$$+(p_3p_4q_1^2q_2^2q_3q_4)^s\} +13(p_1p_2p_3p_4q_1q_2q_3q_4)^s.$$

Substituting (3. 5) into (3. 4), P_1, P_2, P_3 and P_4 will be calculated.

That is, the stationary probabilities of coin ①, coin ②, coin ③ and coin ④ are then $1/(1+\alpha^s+\beta^s+\gamma^s)$, $\alpha^s/(1+\alpha^s+\beta^s+\gamma^s)$, $\beta^s/(1+\alpha^s+\beta^s+\gamma^s)$ and $\gamma^s/(1+\alpha^s+\beta^s+\gamma^s)$ respectively, where $\alpha=q_1p_2/p_1q_2$, $\beta=q_1p_3/p_1q_3$ and $\gamma=q_1p_4/p_1q_4$.
It is assured that $\gamma\leqq\beta\leqq\alpha<1$, if and only if, $p_1>p_2\geqq p_3\geqq p_4$.

In this section we shall follow the same rule as used in Section 2. Therefore we may conclude that the proportion of heads obtained in the first n trials tends to max $\{p_1, p_2, p_3, p_4\}$, with probability one, as $n\longrightarrow\infty$. This, of course, is the maximum achievable limit.

NIIGATA UNIVERSITY

References

[1] COVER, T. M.: *A note on the two-armed bandit problem with finite memory.* Information and Control, Vol. 12, (1968), 371–377.

[2] COVER, T. M.: *Hypothesis testing with finite statistics.* Ann. Math. Stat., Vol. 40, (1969), 828–835.

[3] COVER, T. M. and HELLMAN, M. E.: *The two-armed bandit problem with time-invariant finite memory,* IEEE Trans. Information Theory, Vol. IT-16, (1970), 185–195.

[4] FU, K, S.: *Stochastic automata as models of learning systems.* Computer and Information Sciences-II, Academic Press, New York, (1967), 177–191.

[5] FU, K. S.: Sequential methods in pattern recognition and machine learning. Academic Press, New York and London, (1968).

[6] FU, K. S. and MENDEL, J. M.: Adaptive, learning, and pattern recognition systems theory and applications. Academic Press, New York and London, (1970).

[7] HELLMAN, M. E. and COVER, T. M.: *Learning with finite memory.* Ann. Math. Stat., Vol. 41, (1970), 765–782.

[8] HELLMAN, M. E. and COVER, T. M.: *On memory saved by randomization.* Ann. Math. Stat., Vol. 42, (1971), 1075–1078.

[9] ISBELL, J. R.: *On a problem of Robbins,* Ann. Math. Stat., Vol. 30, (1959), 606–610.

[10] ROBBINS, H.: *A sequential decision problem with a finite memory.* Proc. Nat. Acad. Sci. Vol. 42, (1956), 920–923.

[11] SMITH, C. V. and PYKE, R.: *The Robbins-Isbell two-armed bandit probelm with finite memory.* Ann. Math. Stat., Vol. 36, (1965), 1375–1386.

[12] SAMUELS, S. M.: *Randomized rules for the two-armed bandit problem with finite memory.* Ann. Math. Stat., Vol. 39, (1968), 2103–2107.

[13] TSETLIN, M. L.: *On the behavior of finite automata in randon media.* Autom. and Remote Control, Vol. 22, No. 10. (1961), 1210–1219.

[14] VARSHAVSKII, V. I. and VORONTSOVA, I. P.: *On the behavior of stochastic automata with variable structure.* Autom. and Remote Control, Vol. 24, No. 3, (1963), 327–333.