On a Sequential Allocation Problem with Partial Observations*

Tsuneyuki Namekata

Abstract

A sequential allocation problem in which a decision maker cannot observe appearing targets directly, is considered. There are finite types of targets, and the types of actually appearing targets (core process) constitute a temporally nonhomogeneous Markov chain. During a finite number of periods, the decision maker observes not the core process directly but another process (observation process) which is stochastically related to the core process, and allocates some of his resources on hand to acquire some reward dependent on the number of resources expended and the type of actually appearing target. The objective is to find a sequence of number of resources to be expended that maximizes the total expected reward. Some properties of an optimal policy are investigated.

I. Introduction

A decision maker will allocate his resources to appearing targets during the given periods. The target appears one by one at the beginning of each period and the types of actually appearing targets (core process)

The manuscript was submitted for publication on April 1, 1985.

* This paper is based on my study at Osaka University. I am very grateful to Professor T. Nishida and Associate Professor Y. Tabata for helpful suggestions and encouragement.

constitute a temporally nonhomogeneous Markov chain. The decision maker obtains the reward which depends on the number of resources expended and the actually appearing target. How does he allocate his resources to maximize the total expected reward?

Sequential allocation problems described above were studied by many authors when the decision maker can observe the appearing target directly [for example, 1,2,3,5,8 (2 and 8 treat the continuous-time version)]. The present paper considers the sequential allocation problem in which the decision maker observes not the core process directly but another process (observation process) that is stochastically related to the core process. This type of problem is modeled as a partially observable Markov decision process [9]. For example, in [6,7,10] the machine replacement problem with partial information are considered and some structural properties of an optimal policy are derived. In [6,7] the actions available at each decision epoch are 'replace', 'inspect', and 'do nothing' such that 'replace' and 'inspect' actions give the decision maker the true state of the machine, and 'do nothing' action gives him no information about the machine. In [10] two actions, 'replace' and 'do nothing', such that 'replace' indicates the true state and 'do nothing' gives some information about the machine, are available. And two extreme cases, complete information case and no information case, are considered in detail. In [4] the problem of optimal stopping with imperfect information was investigated. Three actions, 'accept', 'continue', and 'inspect' are available at each decision epoch. 'Continue' action gives no information. 'Inspect' action indicates some information about the core process, and this action can be taken many times successively without any change of core process.

In the present paper the sequential allocation problem with partial information is investigated. This generalizes the model in Section 4 of [5] since it is the case where the transition probability matrix of the core process has the same rows and the Q matrix is an identity matrix (full information). Actions consist of the number of resources expended from the resources on hand. The case is considered where all actions give the same information about the core process. Later the case is discussed where all actions but 'expend no unit' indicate the true state of the core process.

An outline of the paper is as follows: In Section 2 our problem is formulated as a partially observable Markov decision process. In Section 3 some properties of an optimal policy are discussed. Two-state-case and the case of transition probability matrix with the same rows are of special interest. In Section 4 some examples are presented to illustrate our model.

2. Model and Formulation

The decision maker will allocate M units of his resources to appearing targets during a given period N. There are I types of targets. We assume that the target appears one by one at the beginning of each period and the types of actually appearing targets (core process) constitute a temporally nonhomogeneous Markov chain. Let $X_n=$ the type of target actually appearing at the n-th period, then $\{X_n, n=1,\dots, N\}$ is a Markov chain with known transition matrix $P(n)=(p_{ij}(n))$, $i, j=0,\dots, I$, $n=2,\dots, N$, where $p_{ij}(n)=\Pr\{X_{N-n+2}=j | X_{N-n+1}=i\}$ (n denotes the number of remaining periods.) State 0 means no appearance of target. The decision maker cannot observe the core process $\{X_n, n=1,\dots, N\}$, instead can observe the observation process $\{Y_n, n=1,\dots, N\}$ which takes a value in $\{0,1,\dots, L\}$ and is stochastically related to the core process as follows: $Q=(q_{ij})$, $i=0,\dots, I$, $j=0,\dots, L$. $\Pr\{Y_n=j | X_n=i\}=q_{ij}$ for $i=0,\dots, I$, $j=0,\dots, L$, $n=$

1,..., N. If the decision maker allocates j units from his resources on hand, and the type $i(i=1,\dots, I)$ target is actually appearing, then he obtains the immediate reward $R_i(j)$. It is assumed that $R_i(j)$ is a nondecreasing and concave function of j with $R_i(0)=0$ for $i=1,\dots, I$. The objective is to find an optimal sequential allocation procedure which maximizes the total expected reward by allocating M units of resources to the appearing targets during N periods.

Consider, for example, the targets some types of which resemble one another in look but differ in essence. So the time needed to identify the appearing target is long. And suppose that the reward acquired at each period does not directly become known to the decision maker until the end of planning horizon. Therefore the model discussed in the above paragraphs may apply to this situation if the decision maker has to determine his choice within a fairly short period of time.

A main result of the theory of partially observable Markov decision process is that the process with incompletely known states can be transformed into a 'usual' Markov decision process by enlarging the state space. The new state space is the set of probability distribution over the unobservable states in the original process. Therefore the enlarged new state space is denoted by $S=\{(x_0,\dots, x_I);\ x_i \geqq 0, i=0,\dots, I,\ x_0+\cdots+x_I=1\}$, and $x=(x_0,\dots, x_I)\varepsilon S$ means that the probability of appearance of type i target is x_i just before the decision-making. The sequence of events, transition of core process, observation, and decision-making, is given in Fig. 1. The following quantities are calculated immediately by Bayes' rule:

(1) $\theta_k(n, x)$ = probability of observing output k at the next period given that there are n periods remaining and the current state is denoted by x

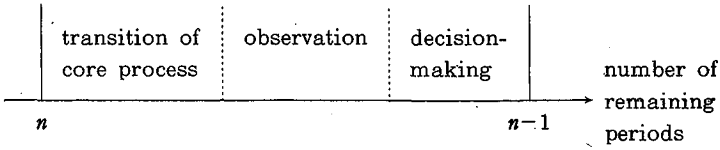

Fig. 1

$$=\sum_{j=0}^{I}\left(\sum_{i=0}^{I}x_i p_{ij}(n)\right)q_{jk},\ n=2,\dots,N,\ k=0,\dots,L,\ x\epsilon S.$$

$T_i(n, k, x)$ = probability of appearance of type i target at the next period given that there are n periods remaining, the current state is denoted by x, and output k is observed

$$=\left(\sum_{j=0}^{I}x_j p_{ji}(n)q_{ik}\right)/\theta_k(n, x),\ n=2,\dots,N,\quad i=0,\dots,I,\ k=0,\dots,L,\ x\epsilon S.$$

Let

(2) $$T(n, k, x)=(T_0(n, k, x),\dots,T_I(n, k, x)),\ n=2,\dots,N,\quad k=0,\dots,L,\ x\epsilon S,$$

then $T(n, k, x)$ denotes the state in the next period provided that there are n periods remaining, the current state is x, and output k is observed. Also define

$V_n(m, x)$ = the maximum total expected reward when there are n periods remaining, m units of resources are on hand, the current state is given by x, and an optimal policy follows.

Then the following recursive relation follows [for example, 9].

(3) $$V_n(m, x)=\max_{j=0,\dots,m}\left\{\sum_{i=1}^{I}x_i R_i(j)+\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m-j,\ T(n, k, x))\right\},\quad n=1,\dots,N,\ m=0,\dots,M,\ x\epsilon S.$$

where $V_0(\cdot,\cdot)=0$. If $j (j=0,\dots,m)$ units of resources are expended, then the immediate expected reward is given by $\sum_{i=1}^{I} x_i R_i(j)$, output k is observed at the next period with probability $\theta_k(n, x)$, and the maximum total expected reward from the resultant state is denoted by $V_{n-1}(m-j, T(n, k, x))$. Therefore (3) follows.

If the recursive relation (3) is solved, an optimal policy is presented.

3. Structure of Optimal Policy

We can find an optimal policy if the equation (3) is solved. Unfortunately we cannot solve them explicitly in general. So we develop some properties of an optimal policy.

To express one of the optimal policies we define $k(n, m, x)$ as the smallest value of j that maximizes the braces of the right hand side of equation (3), that is,

$$(4)\quad k(n, m, x)=\min\{t;\ \textstyle\sum_{i=1}^{I} x_i R_i(t)+\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m-t, T(n, k, x)) \\ = \max_{j=0,\dots,m} \{\textstyle\sum_{i=1}^{I} x_i R_i(j) \\ +\textstyle\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m-j, T(n, k, x))\}\}, \\ n=1,\dots, N,\ m=0,\dots, M,\ x\varepsilon S.$$

Using this notation it is the optimal policy that allocates $k(n, m, x)$ units of resources when there are n periods remaining, m units of resources are on hand, and the current subjective probability about the types of appearing target is denoted by x.

The monotonicity of $k(n, m, x)$ with respect to m is proved in Theorem 1. The following lemma needed in Theorem 1 is introduced without proof since it is obvious.

Lemma 1. $V_n(m, x)$ is a nondecreasing and concave function of m for $n=1,\dots, N$, $x\varepsilon S$.

The following theorem presents the monotonicity of the optimal number of resources to be expended with respect to the number of resources on hand. The result seems to meet with our intuition that the more units of resources are on hand the more units of resources should be expended. The proof is similar to Proposition 2 and Theorem 3 in [1] since $R_i(\cdot)$ and $\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(\cdot, T(n, k, x))$ are concave.

Theorem 1. $k(n, m, x) \leqq k(n, m+1, x) \leqq k(n, m, x)+1$ for $n=1,\dots, N$, $m=0,\dots, M-1$, and $x \epsilon S$.

Next we consider whether $k(n, m, x)$ is monotone in x with some partial order on S. This conjecture does not always hold in general even if P (transition matrix of core process) is increasing failure rate (IFR) or totally positive of order two (TP_2) as will be shown in Section 4. So we restrict our attention to the following two cases:

(i) $I=1$ and

$$P(n)=\begin{pmatrix} 1-p_0(n) & p_0(n) \\ 1-p_1(n) & p_1(n) \end{pmatrix} \text{ with } p_0(n) \leqq p_1(n) \text{ and } 0<p_1(n),$$

(this transition matrix is TP_2).

(ii) $P(n)$ with the same rows, that is,

$$P(n)=\begin{pmatrix} r_0(n) & \cdots & r_1(n) \\ \cdot & \cdot & \cdot \\ r_0(n) & \cdots & r_1(n) \end{pmatrix}$$

with $r_i(n) \geqq 0$, $i=0,\dots, I$, $r_0(n)+\cdots+r_I(n)=1$,

(this transition matrix is also TP_2).

3.1. Case (i)

We consider the first case (i) in this subsection.

Let

$$P(n)=\begin{pmatrix} 1-p_0(n) & p_0(n) \\ 1-p_1(n) & p_1(n) \end{pmatrix} \text{ with } p_0(n) \leqq p_1(n) \text{ and } 0<p_1(n),$$

$$Q=\begin{pmatrix}\alpha_0 & \cdots & \alpha_L \\ \beta_0 & \cdots & \beta_L\end{pmatrix} \text{ with } \alpha_i\beta_i \neq 0 \text{ for } i=0,\ldots,L$$

Then (1), (2), and (3) are rewritten as

(1′) $\theta_k(n, x)=(\beta_k-\alpha_k)(p_1(n)-p_0(n))x+\alpha_k+(\beta_k-\alpha_k)p_0(n),$

$n=2,\ldots, N,\ k=0,\ldots, L,$

(2′) $T_k(n, x)=[\beta_k((p_1(n)-p_0(n))x+p_0(n))]/\theta_k(n,x),$

$n=2,\ldots, N,\ k=0,\ldots, L,$

(3′) $V_n(m, x)=\max_{j=0,\ldots,m}\{xR_1(j)+\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m-j, T(n, k, x))\},$

$n=1,\ldots, N,\ m=0,\ldots, M,$

where $x\varepsilon$ [0,1] represents the probability that the appearing target is of type 1. The following lemma will be used in the proof of Theorem 2.

Lemma 2. $[\sum_{k=0}^{L}\theta_k(n,x)(V_n(m+1,T(n, k, x))-V_n(m, T(n, k, x)))]/x$

is nonincreasing in $x\varepsilon(0,1]$ for $n=1,\ldots, N$, $m=0,\ldots, M-1$.

Proof: First we show that

(5) $[V_n(m+1, x)-V_n(m, x)]/x$

is nonincreasing in $x\varepsilon(0,1)$ for $n=1,\ldots, N$, $m=0,\ldots, M-1$, by induction on n. For $n=1$, (5) becomes $R_1(m+1)-R_1(m)$, so (5) is nonincreasing in $x\varepsilon$ (0,1) for $n=1$. Suppose that (5) is nonincreasing in $x\varepsilon$ (0,1) for n.

$$[V_{n+1}(m+1, x)-V_{n+1}(m, x)]/x= \max\{R_1(j_0+1)-R_1(j_0), \sum_{k=0}^{L}(\theta_k(n, x)T(n, k, x)/x)\cdot [V_n(m+1-j_0, T(n, k, x))-V_n(m-j_0, T(n, k, x))]/T(n, k, x)\},$$

where $j_0=k(n+1, m, x)$ and the equality follows from Theorem 1. Since $\theta_k(n, x)T(n, k, x)/x$ is nonnegative and nonincreasing in $x\varepsilon(0, 1)$,

$T(n, k, x)$ is nondecreasing in $x\varepsilon(0, 1)$, and $[V_n(m+1-j_0, T(n, k, x)) - V_n(m-j_0, T(n, k, x))]/T(n, k, x)$ is nonnegative and nonincreasing by induction hypothesis, (5) is nonincreasing in $x\varepsilon(0,1)$ for $n+1$. Therefore,

$$\begin{aligned}&[\textstyle\sum_{k=0}^{L}\theta_k(n, x)(V_n(m+1, T(n, k, x))-V_n(m, T(n, k, x)))]/x\\&=\textstyle\sum_{k=0}^{L}(\theta_k(n, x)T(n, k, x)/x)[V_n(m+1, T(n, k, x))\\&\qquad\qquad -V_n(m, T(n, k, x))]\ /T(n, k, x)\end{aligned}$$

is nonincreasing in $x\varepsilon(0,1)$, and is continuous in $x\varepsilon(0,1]$. So the expression is nonincreasing in $x\varepsilon(0,1]$. Q. E. D.

The following theorem says that the more strongly the decision maker believes that the appearing target is of type 1, the more units of resources should be expended.

Theorem 2. $k(n, m, x)\leqq k(n, m, x')$ for $0\leqq x<x'\leqq 1$, $n=1,\dots, N$, $m=0,\dots, M$.

Proof: Since $k(n, m, 0)=0$, it is sufficient to prove the theorem for $0<x<x'\leqq 1$. Let $j_0=k(n, m, x)$. If $j_0=0$, then the result is obvious. Suppose $j_0>0$. Since

$$\begin{aligned}&xR_1(j_0)+\textstyle\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m-j_0, T(n, k, x))\\&\quad >xR_1(j)+\textstyle\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m-j, T(n, k, x)) \text{ for } 0\leqq j<j_0,\\&R_1(j_0)-R_1(j)>\textstyle\sum_{k=0}^{L}\theta_k(n, x)[V_{n-1}(m-j, T(n, k, x))\\&\quad -V_{n-1}(m-j_0, T(n, k, x))]/x\geqq\textstyle\sum_{k=0}^{L}\theta_k(n, x')[V_{n-1}(m-j, T(n, k, x'))\\&\quad -V_{n-1}(m-j_0, T(n, k, x'))]/x' \text{ for } 0\leqq j<j_0,\end{aligned}$$

by Lemma 2. The first and the last terms imply

$k(n, m, x')\geqq j_0=k(n, m, x)$ for $0<x<x'\leqq 1$. Q. E. D.

3.2. Case (ii)

The second case (ii) is considered in this subsection. Since

$$P(n)=\begin{pmatrix} r_0(n) & \cdots & r_I(n) \\ \cdot & \cdot & \cdot \\ r_0(n) & \cdots & r_I(n) \end{pmatrix} \text{with } r_i(n)\geqq 0,\ i=0,\ldots,\ I,\ r_0(n)+\cdots+r_I(n)=1,$$

$\theta_k(n, x)$ and $T(n, k, x)$ are independent of x and are abbreviated as $\theta_k(n)$ $T(n, k)$ respectively. Then the recursive relation (3) is rewritten as

$$(3'')\quad V_n(m, x)=\max_{j=0,\ldots,m}\{\textstyle\sum_{i=1}^{I} x_i R_i(j) + \sum_{k=0}^{L}\theta_k(n)V_{n-1}(m-j,\ T(n, k))\}.$$

As in [10], we introduce the partial order $\subset$ on S as follows:

$$x=(x_0,\ldots,\ x_I)\subset x'=(x'_0,\ldots,\ x'_I)$$

if and only if

$$x_0+\cdots\cdots+x_I\leqq x'_0+\cdots\cdots+x'_I,$$
$$x_1+\cdots+x_I\leqq x'_1+\cdots+x'_I,$$
$$\cdots\cdots\cdots$$
$$x_I\leqq x'_I.$$

The following theorem shows the monotonicity of the optimal number of resources to be expended with respect to the degree of belief of the decision maker. This is a generalization of Remark (2) in [5].

Theorem 3. If $R_i(m+1)-R_i(m)\leqq R_{i+1}(m+1)-R_{i+1}(m)$ for $i=1,\ldots,$ $I-1$, $m=0,\ldots,\ M-1$, then $x\subset x'$ implies $k(n, m, x)\leqq k(n, m, x')$ for $n=1,\ldots,\ N$, $m=0,\ldots,\ M$.

Proof: As in the proof of Theorem 2, suppose $j_0=k(n, m, x)>0$. Then,

$$\sum_{i=1}^{I} x_i R_i(j_0)+\sum_{k=0}^{L}\theta_k(n)V_{n-1}(m-j_0, T(n, k))$$
$$>\sum_{i=1}^{I} x_i R_i(j)+\sum_{k=0}^{L}\theta_k(n)V_{n-1}(m-j, T(n, k)) \text{ for } 0\leqq j<j_0.$$

Since $x\subset x'$,

$$\sum_{i=1}^{I} x'_i (R_i(j_0)-R_i(j))\geqq\sum_{i=1}^{I} x_i(R_i(j_0)-R_i(j)) \text{ for } 0\leqq j<j_0.$$

The above two inequalities yield

$$\sum_{i=1}^{I} x'_i R_i(j_0)+\sum_{k=0}^{L}\theta_k(n)V_{n-1}(m-j_0, T(n, k))$$
$$>\sum_{i=1}^{I} x'_i R_i(j)+\sum_{k=0}^{L}\theta_k(n)V_{n-1}(m-j, T(n, k)) \text{ for } 0\leqq j<j_0.$$

This implies $k(n, m, x')\geqq j_0=k(n, m, x)$. Q. E. D.

The following theorem which presents the monotonicity of the optimal number of resources to be expended with respect to the number of remaining periods is proved similarly to Theorem 5 (ii) in [5].

Theorem 4. $k(n, m, x)\geqq k(n+1, m, x)$ for $n=1,\dots, N-1$, $m=0,\dots, M$, $x\varepsilon S$.

3.3

We discuss in this subsection the case in which the decision maker knows the true type of appearing target from the knowledge of the expenditure of resources. The recursive relation corresponding to (3) becomes

$$(6)\quad V_n(m, x)= \max \{\sum_{k=0}^{L}\theta_k(n, x)V_{n-1}(m, T(n, k, x)),$$
$$\max_{j=1,\dots, m} \{\sum_{i=1}^{I} x_i R_i(j)$$
$$+\sum_{i=1}^{I} x_i \sum_{k=0}^{L}\theta_k(n, e_i)V_{n-1}(m-j, T(n, k, e_i))\}\},$$

where $e_i\varepsilon S$ has 1 as its $(i+1)$-st element and $V_0(\cdot, \cdot)=0$. Since $V_n(m, x)$ is a piecewise-linear and convex function of x (see [9]) and the terms corresponding to $j=1,\ldots,m$ in the right hand side of (6) are linear functions of x, it readily follows that for each fixed (n, m) $(n=1,\ldots,N,$ $m=1,\ldots,M)$ the set $\{x \varepsilon S;\ k(n, m, x)=t\}$ is convex for $t=1,\ldots,m$. Furthermore if $I=1$, the following property is derived.

Theorem 5. For $n=1,\ldots,N$, $m=1,\ldots,M$,

(i) if $0<x<x'$, $k(n, m, x)>0$, and $k(n, m, x')>0$, then $k(n, m, x)\leqq k(n, m, x')$,

(ii) if $p_1(n)<1$, then $k(n, m, 1)>0$.

Proof: (i) Note that $V_n(m, x)$ is nondecreasing in m.
Let $j_0=k(n, m, x')$, then

$$
\begin{aligned}
&x'(R_1(j_0)+\textstyle\sum_{k=0}^{L}\theta_k(n, 1)V_{n-1}(m-j_0, T(n, k, 1)))\\
&\qquad+(1-x')(\textstyle\sum_{k=0}^{L}\theta_k(n, 0)V_{n-1}(m-j_0, T(n, k, 0)))\\
&\geqq x'(R_1(j)+\textstyle\sum_{k=0}^{L}\theta_k(n, 1)V_{n-1}(m-j, T(n, k, 1)))\\
&\qquad+(1-x')(\textstyle\sum_{k=0}^{L}\theta_k(n, 0)V_{n-1}(m-j, T(n, k, 0))) \quad \text{for } j_0\leqq j\leqq m.
\end{aligned}
$$

This yields

$$
\begin{aligned}
&[R_1(j)+\textstyle\sum_{k=0}^{L}\theta_k(n, 1)V_{n-1}(m-j, T(n, k, 1))\\
&\qquad-\textstyle\sum_{k=0}^{L}\theta_k(n, 0)V_{n-1}(m-j, T(n, k, 0))]\\
&-[R_1(j_0)+\textstyle\sum_{k=0}^{L}\theta_k(n, 1)V_{n-1}(m-j_0, T(n, k, 1))\\
&\qquad-\textstyle\sum_{k=0}^{L}\theta_k(n, 0)V_{n-1}(m-j_0, T(n, k, 0))]\\
&\leqq[\textstyle\sum_{k=0}^{L}\theta_k(n, 0)(V_{n-1}(m-j_0, T(n, k, 0))-V_{n-1}(m-j, T(n, k, 0)))]/x'\\
&\leqq[\textstyle\sum_{k=0}^{L}\theta_k(n, 0)(V_{n-1}(m-j_0, T(n, k, 0))-V_{n-1}(m-j, T(n, k, 0)))]/x\\
&\qquad \text{for } j_0\leqq j\leqq m.
\end{aligned}
$$

The first and the third terms imply that $k(n, m, x)\leqq j_0=k(n, m, x')$.

To prove (ii), it is sufficient to show that

$$
\begin{aligned}
&R_1(1)+\textstyle\sum_{k=0}^{L}\theta_k(n, 1)V_{n-1}(m-1, T(n, k, 1))\\
&\qquad>\textstyle\sum_{k=0}^{L}\theta_k(n, 1)V_{n-1}(m, T(n, k, 1)),
\end{aligned}
$$

that is,

$$\sum_{k=0}^{L}\theta_k(n,1)[V_{n-1}(m,T(n,k,1))-V_{n-1}(m-1,T(n,k,1))]<R_1(1)$$
$$\text{for } n=2,\dots,N,\ m=1,\dots,M.$$

Since $p_1(n)<1$ implies $T(n,k,1)<1$ for some k, it suffices to show that for $n=1,\dots,N$, $m=0,\dots,M$,

$$(7)\quad V_n(m+1,x)-V_n(m,x)<(\leqq)R_1(1) \text{ for } x<(=)1.$$

For $n=1$, since $V_1(m+1,x)-V_1(m,x)=x(R_1(m+1)-R_1(m))\leqq xR_1(1)$, inequality (7) is satisfied. Suppose (7) holds for $n-1$.
Let $j_0=k(n,m+1,x)$, then

$$V_n(m+1,x)-V_n(m,x)$$
$$\leqq\begin{cases}\sum_{k=0}^{L}\theta_k(n,x)[V_{n-1}(m+1,T(n,k,x)) \\ \qquad -V_{n-1}(m,T(n,k,x))],\ \text{if } j_0=0, \\ x\sum_{k=0}^{L}\theta_k(n,1)[V_{n-1}(m+1-j_0,T(n,k,1)) \\ \qquad -V_{n-1}(m-j_0,T(n,k,1))] \\ \quad +(1-x)\sum_{k=0}^{L}\theta_k(n,0)[V_{n-1}(m+1-j_0,T(n,k,0)) \\ \qquad -V_{n-1}(m-j_0,T(n,k,0))]\ \text{if } 0<j_0<m+1, \\ x(R_1(m+1)-R_1(m)),\ \text{if } j_0=m+1,\end{cases}$$
$$<(\leqq)R_1(1),\ \text{for } x<(=)1. \qquad \text{Q. E. D.}$$

By the preceding facts and $k(n,m,0)=0$, the interval $[0,1]$ may be devided as is shown in Fig. 2.

As for the property of $k(n,m,x)$ with respect to m, the following theorem is proved.

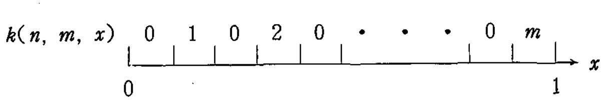

Fig. 2

Theorem 6. $k(n, m+1, x) \leqq k(n, m, x)+1$ for $n=1,\ldots, N, m=0,\ldots, M$, and $x \varepsilon S$.

Proof: Let $j_0=k(n, m, x)$. If $j_0>0$,

$$\sum_{i=1}^{I} x_i R_i(j_0)+\sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m-j_0, T(n, k, e_i))$$
$$\geqq \sum_{i=1}^{I} x_i R_i(j)+\sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m-j, T(n, k, e_i))$$
$$\text{for } j_0 \leqq j \leqq m.$$

Since $R_i(\cdot)$ is concave,

$$\sum_{i=1}^{I} x_i (R_i(j_0+1)-R_i(j_0)) \geqq \sum_{i=1}^{I} x_i (R_i(j+1)-R_i(j)) \text{ for } j_0 \leqq j \leqq m.$$

These two inequalities yield

$$\sum_{i=1}^{I} x_i R_i(j_0+1)$$
$$+\sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m+1-(j_0+1), T(n, k, e_i))$$
$$\geqq \sum_{i=1}^{I} x_i R_i(j+1)$$
$$+\sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m+1-(j+1), T(n, k, e_i))$$
$$\text{for } j_0+1 \leqq j+1 \leqq m+1.$$

This implies $k(n, m+1, x) \leqq j_0+1=k(n, m, x)+1$. If $j_0=0$,

$$\sum_{k=0}^{L} \theta_k(n, x) V_{n-1}(m, T(n, k, x))$$
$$\geqq \sum_{i=1}^{I} x_i R_i(j)+\sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m-j, T(n, k, e_i))$$
$$\text{for } 1 \leqq j \leqq m.$$

From the convexity of $V_{n-1}(m, \cdot)$,

$$\sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m, T(n, k, e_i))$$
$$\geqq \sum_{k=0}^{L} \theta_k(n, x) V_{n-1}(m, T(n, k, x)).$$

And from the concavity of $R_i(\cdot)$

$$\sum_{i=1}^{I} x_i R_i(1) \geqq \sum_{i=1}^{I} x_i (R_i(j+1)-R_i(j)) \text{ for } 1 \leqq j \leqq m.$$

Therefore,

$$\sum_{i=1}^{I} x_i R_i(1) + \sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m+1-1, T(n, k, e_i))$$
$$\geqq \sum_{i=1}^{I} x_i R_i(j+1)$$
$$+ \sum_{i=0}^{I} x_i \sum_{k=0}^{L} \theta_k(n, e_i) V_{n-1}(m+1-(j+1), T(n, k, e_i))$$
$$\text{for } 2 \leqq j+1 \leqq m+1.$$

So $k(n, m+1, x) \leqq j_0 + 1 = k(n, m, x) + 1$. Q. E. D.

4. Examples

In this section two examples are presented. One shows that the total positivity of order two (TP_2) is not sufficient to the monotonicity of $k(n, m, x)$ with respect to x. The other is a numerical example of 3.3 for $I=1$.

4.1.

Let $N=2$, $M=1$, $I=2$, and

$$P(2) = \begin{pmatrix} 1 & 0 & 0 \\ 1-R & R/2 & R/2 \\ 0 & 0 & 1 \end{pmatrix}$$

($0<R<1$, this transition matrix is TP_2).

And $R_i(\cdot)$ ($i=1,2$) are listed in Table 1.

Then it is easily checked that $V_2(1, e_1)=R$, $V_2(1, e_2)=1$, $k(2, 1, e_1)=1$, and $k(2, 1, e_2)=0$. This concludes that $e_1=(0, 1, 0) \subset e_2=(0, 0, 1)$ but $k(2, 1, e_1) > k(2, 1, e_2)$.

Table 1. $R_i(\cdot)$, $i=1,2$

i \ j	0	1
1	0	$R(0<R<1)$
2	0	1

4.2.

As an example of 3.3 for $I=1$, $N=5$, $M=5$,

$$P(n)=P=\begin{pmatrix}0.9 & 0.1\\ 0.5 & 0.5\end{pmatrix} \text{ for } n=2,\ldots,5, \text{ and } Q=\begin{pmatrix}0.3 & 0.7\\ 0.8 & 0.2\end{pmatrix}.$$

And $R_1(\cdot)$ is given in Table 2. Then the optimal policy is obtained in Fig. 1–Fig. 4. For example, if there are 5 periods remaining, 5 units of resources are on hand, and the decision maker has his subjective probability 0.5 that the target is appearing after the observation of Y, then it is optimal to allocate 2 units. Fig. 1–Fig. 4 show that $k(n, m, x) \leq k(n, m+1, x)$ for $n=1,\ldots, 5$, $m=0,\ldots, 4$, $x \varepsilon [0, 1]$ do hold, which we could not prove in 3.3. Also note that for (n, m), $n=1,\ldots, 5$, $m=0,\ldots, 5$, the set $\{x \varepsilon [0, 1]; k(n, m, x)=0\}$ is convex, which we could not prove in 3.3 either.

Table 2. $R_1(\cdot)$

j	0	1	2	3	4	5
$R_1(j)$	0	1.0	1.5	1.8	1.9	1.91

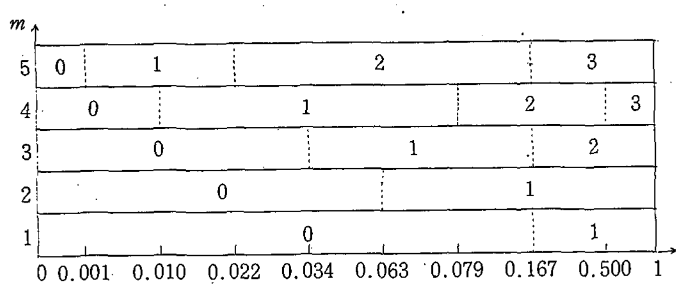

Fig. 1. $k(2, m, x)$

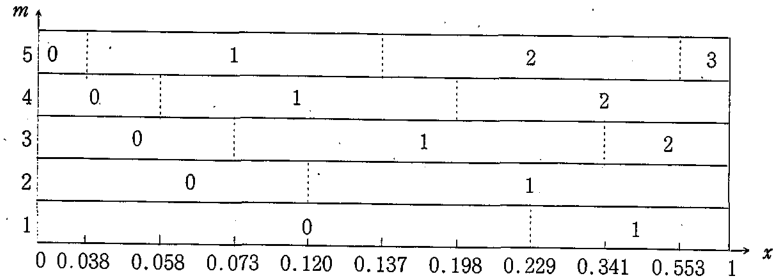

Fig. 2. $k(3, m, x)$

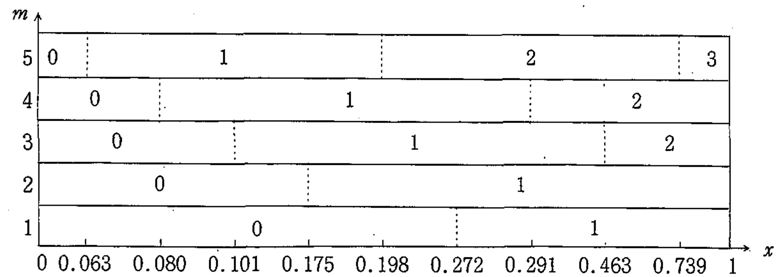

Fig. 3. $k(4, m, x)$

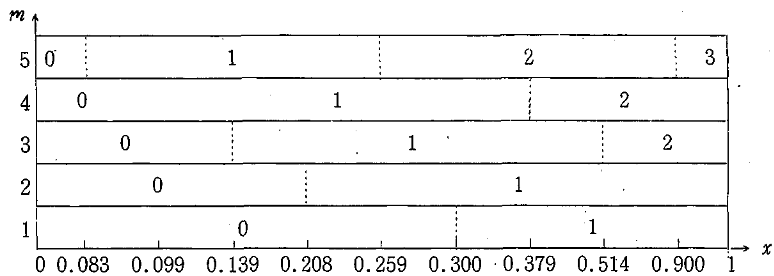

Fig. 4. $k(5, m, x)$

5. Conclusion

In this paper, we considered the discrete-time sequential allocation problem with partial observations. First we developed the structural properties of an optimal policy for the case where all actions gave the same information about the unobservable core process. Secondly the case where all actions but 'expend no unit' indicated the true state of the unobservable core process. We gave the properties of an optimal policy with respect to the number of resources on hand and with respect to the decision maker's belief about the appearing target. As for the property of an optimal policy with respect to the decision maker's belief about the appearing target, two special cases were of special interest: One was of only one type of target. The other assumed the transition probability matrix had the same rows.

It is a future problem to consider the more realistic case where the true type of actually appearing target gets known to the decision maker with some probability (the probability with which the expended resources hit the appearing target in the context of the example in [5]) instead of with certainty if some resources on hand are expended.

References

[1] Derman, C. G., Lieberman, G. J., and Ross, S. M.: A Stochastic Sequential Allocation Model. *Operations Research*, Vol. 23 (1975), 1120–1130.

[2] Donis, J. M. and Pollock, S. M.: Allocation of Resources to Randomly Occuring Opportunities. *Naval Research Logistics Quarterly*, Vol. 14 (1967), 513–527.

[3] Mastran, D. V. and Thomas, G. J.: Decision Rules for Attacking Targets of Opportunity. *Naval Research Logistics Quarterly*, Vol. 20 (1973), 661–672.

[4] Monahan, G. E.: Optimal Stopping in a Partially Observable Markov Process with Costly Information. *Operations Research*, Vol. 28 (1980), 1319–1334.

[5] Namekata, T., Tabata, Y., and Nishida, T.: A Sequential Allocation Problem with Two Kinds of Targets. *Journal of the Operations Research Society of Japan*, Vol. 22 (1979), 16–28.

[6] Rosenfield, D.: Markovian Deterioration with Uncertain Information. *Operations Research*, Vol. 24 (1976), 141–155.

[7] Ross, S. M.: Quality Control under Markov Deterioration. *Management Science*, Vol. 17 (1971), 587–596.

[8] Sakaguchi, M.: A Sequential Allocation Problem for Randomly Appearing Targets. *Mathematica Japonica,* Vol. 21 (1976), 89–103.

[9] Smallwood, R. D. and Sondik, E. J.: The Optimal Control of Partially Observalbe Markov Processes over a Finite Horizon. *Operations Research,* Vol. 21 (1973), 1071–1088.

[10] White, C. C.: Bounds on Optimal Cost for a Replacement Problem with Partial Observations. *Naval Research Logistics Quarterly,* Vol. 26 (1979), 415–422.