Journal of the Operations Research
Society of Japan
Vol. 40, No. 2, June 1997

A STOCHASTIC SEQUENTIAL ALLOCATION PROBLEM WHERE THE RESOURCES CAN BE REPLENISHED

Masahiro Sato
Univ. of Tsukuba

(Received August 25, 1995; Revised July 31, 1996)

Abstract Suppose a hunter starts hunting over certain given t periods with i bullets in hand. A distribution of the value of each appearing target and the hitting probability of a bullet are known. For shooting, he takes a strategy of *shoot-look-shoot* scheme, implying that if a bullet just fired does not hit the target, then the hunter must decide whether or not to shoot an additional one. At the end of each period, it is allowed to replenish a given number of bullets by paying a certain cost. The objective here is to examine the properties of the optimal policy which maximizes the total expected net reward. We get the following main results: (1) the optimal policy for shooting is monotone in the number of bullets in hand if it is always optimal either to replenish a certain number of bullets every period or not to replenish them at all, (2) if only one bullet can be replenished per period, then both the optimal policies for shooting and replenishment are monotone in the number of bullets in hand, (3) if more than one can be replenished per period, then there exist examples where the optimal policies for shooting are not monotone in the number of remaining bullets.

1. Introduction

Consider a problem of allocating countable resources to investment opportunities appearing one by one over a given planning horizon. At the beginning of each period, an opportunity comes with a certain value which is a random sample from a known probability distribution. Assume the resources are allocated to the opportunities pursuant to the *shoot-look-shoot* policy, implying that, if investing one unit of resources yields an unsuccessful result, then it is decided whether or not to invest one more at once. At the end of each period, the resources can be replenished by paying a certain cost; it must be decided whether or not to replenish m units of resources then. The aim is to maximize the total expected reward obtained from the successful opportunities minus the total cost for replenishment.

In general, there exist two kinds of policies in sequential allocation problems: shoot-look-shoot and *volley* policies. In volley policy, it must be decided how much resources to invest in salvo. Mastran and Thomas [4] treat the problem as a target attacking one in which the computational method to obtain the optimal decision rules for both policies are shown. Kisi [3] considers a model of shoot-look-shoot policy and examines the relation between the approximate solution and the exact. Sakaguchi [9] investigates the continuous-time version of [4]. Namekata et al. [5] deal with a model of volley policy where there exist two kinds of targets in a sense that the necessary number of resources to get them are different. They also examine problems with volley policy in [6] and [7]. In [6], it is discussed how to allocate perishable resources, and in [7], a case with a random planning horizon is investigated. Derman et al. [2], and Prastacos [8] deal with the problems as investment ones with volley policy. In [10], a problem with shoot-look-shoot policy, in which the search cost must be paid to find an investment opportunity, is discussed, and it is derived that the critical value, at which investing or not become indifferent in the optimal decision, is not always decreasing† in the number of remaining resources.

In models such as stated above, if all of the resources are spent before the deadline, then later chances, which may be more attractive, will be unavailable. However, if the resources can be replenished by paying a certain cost, then the decision maker can continue investing activities in order to gain the total expected reward. In this paper, we discuss the problem where such replenishment is assumed.

In the following section, we exactly define our model and formulate fundamental equations. In Section 3, properties of the optimal policy are derived. In Sections 4 and 5 that follow, a case for which it is optimal to replenish the resources every period and a case for which it is optimal not to replenish at all are investigated. A case that only one unit can be replenished per period is considered in Section 6, and a case for more than one unit is examined and some numerical examples are shown in Section 7. Finally, in Section 8, we summarize conclusions obtained and examine the problem with volley policy roughly.

2. Model and Fundamental Equations

Now using the following hunting problem, we will explain the model treated in this paper. Suppose a hunter starts hunting over a given planning horizon t with i bullets in hand. At the beginning of each period, he goes to hunt and can find only one target. The case that he cannot find any target is regarded to be equivalent to finding a target of value 0. The value of a target, w, is a random variable having a known probability distribution function $F(w)$ with a finite expectation μ, continuous or discrete where $F(w) = 0$ for $w < 0$, $F(w) < 1$ for $w < 1$, and $F(w) = 1$ for $1 \leq w$. The distribution does not concentrate on only a point, i.e., $\Pr(w) < 1$ for any w. The values of successive targets are assumed to be stochastically independent.

He observes the value of a target as soon as finding it and has to immediately decide whether or not to shoot. If the value is rather small, then he may decide not to shoot and come home with no profit. Suppose the value is favorable and he decides to shoot a bullet. Then the bullet will hit the target with hitting probability q. If the bullet does not hit it, then two cases are further possible: either the target disappears immediately with escape probability r or still remains without any defense. If it stands still there, then he has to decide whether or not to fire an additional bullet. Assume that repeated firings waste no time. If he decides not to shoot any more at the present target, need not shoot (get it,) or cannot shoot (it flees or $i = 0$,) then he comes home. On his way home, he must furthermore decide whether or not to replenish m bullets by paying a cost a; it is not permitted to supply more or less than m bullets. Thus, the period ends and the next comes. The objective is to maximize the total expected discounted net reward over t periods. The decision process is illustrated in Figure 1.

Now we shall formulate the fundamental equations of the model. Let points of time be numbered backward from the final point of the planning horizon as 0, 1, and so on; an interval between time t and time $t-1$ is called period t. We define $u_t(i, w)$ to be the maximum of the total expected net reward starting from time t when i bullets are in hand

†Throughout this paper, the following terms are used in order to avoid the expressions of double negatives; "increasing (decreasing)" means "nondecreasing (nonincreasing)."

and a target of value w is found, and $v_t(i)$ to be its expectation in terms of w, that is;

$$v_t(i) = \int_0^1 u_t(i,\xi)dF(\xi), \quad t \geq 0. \tag{2.1}$$

Furthermore, $z_t(i)$ is defined as the maximum of the total expected net reward starting from time t when he decide not to shoot at the present target any more, provided that i bullets remain. Then we have the following relations:

$$\begin{aligned} u_t(i,w) &= \max\{z_t(i),\ q(w+z_t(i-1)) + (1-q)(rz_t(i-1) + (1-r)u_t(i-1,w))\} \\ &= \max\{z_t(i),\ pu_t(i-1,w) + qw + (1-p)z_t(i-1)\}, \quad t \geq 0,\ i \geq 1, \end{aligned} \tag{2.2}$$

$$u_t(0,w) = v_t(0) = z_t(0), \quad t \geq 0, \tag{2.3}$$

$$z_t(i) = \max\{\beta v_{t-1}(i),\ \beta v_{t-1}(i+m) - a\}, \quad t \geq 1, i \geq 0 \tag{2.4}$$

where $p = (1-q)(1-r) \in [0,1)$ and $\beta \in (0,1]$, a discount factor. The first (second) term inside the braces in the right hand side of (2.2) represents the maximum of the total expected reward when it is decided not to shoot (to shoot) at the present target, and the first (second) term inside the braces in the right hand side of (2.4) denotes the maximum of the total expected reward when it is decided not to replenish (to replenish) m bullets.

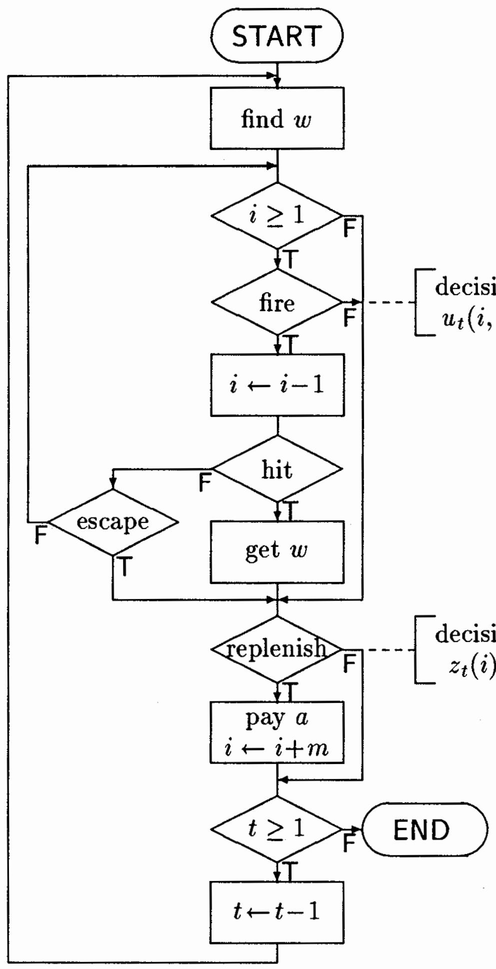

Figure 1. Flowchart of the Decision Process

Further, we immediately have the following final conditions:

$$u_0(i,w) = qw + pu_0(i-1,w) = \frac{1-p^i}{1-p}qw, \quad i \geq 1, \tag{2.5}$$

$$v_0(i) = \frac{1-p^i}{1-p}q\mu, \quad i \geq 1, \tag{2.6}$$

$$z_0(i) = 0, \quad i \geq 0. \tag{2.7}$$

Here, (2.5) and (2.6) hold for $i \geq 0$.

Below, we examine the properties of the fundamental equations.

Lemma 1.

(a) *$u_t(i,w)$, $v_t(i)$ and $z_t(i)$ are increasing in t for any i and w.*

(b) *If $p > 0$, then $u_t(i,w)$, $v_t(i)$ and $z_t(i)$ are strictly increasing in i for any t and w except $z_0(i)$ and $u_0(i,0)$. If $p = 0$, then they are increasing in i for any t and w.*

(c) *$u_t(i+1,w) - u_t(i,w) \leq q$ for any t, i and w where the equal sign holds only for $i = 0$ and $w = 1$. In addition, $v_t(i+1) - v_t(i) < q$ and $z_t(i+1) - z_t(i) < q$ also hold for any t and i.*

(d) *$u_t(i,w)$ is increasing in w for any t and i.*

Proof: All of the above statements can be proven from the definitions and assumptions of this model. First, (a) is true because a hunter who had $t+1$ periods left could simply follow the policy that he would have followed if he had only t periods left. If he decides not to replenish m bullets at $t = 1$, and since $v_0(i) \geq 0$, he will obtain a greater or equal reward. In a similar manner, (b) is immediate because a hunter with $i+1$ bullets at time t could easily put one bullet in his pocket, decide that it is never to be used until $t = 1$, and then follow until $t = 1$ the optimal policy for i bullets and t periods remaining. At the end of period 1, if he fishes the bullet out of his pocket, and since $v_0(i)$ is strictly increasing (increasing for $p = 0$) in i from (2.6), he will get a greater (greater or equal for $p = 0$) reward. Statement (c) follows because the hunter with i bullets can simply adopt the same decisions as a hunter with $i+1$ bullets, until at least one of the two events occurs: he has run out of bullets and the other has one bullet left; or $t = 1$. In the former, if the hunter with a bullet left decides to shoot once more at some target, he can gain at most q (the expected value of hitting a target of maximum worth 1;) so then $u_t(i+1,w) \leq u_t(i,w) + q$ where the equal sign holds only for $w = 1$ and $i = 0$ since $\Pr\{w = 1\} < 1$, and so also hold the similar relations for $v_t(i)$ and $z_t(i)$. In the latter, (c) is intuitive because $v_0(i+1) - v_0(i) = p^i q\mu < q$ from (2.6.) Statement (c) for $u_0(i,w)$ and $z_0(i)$ is trivial. Last, (d) is clear since a hunter whose current target has value w', greater than w, can simply pretend the target has value w, follow the optimal policy for the value w, but in fact obtain a greater or equal reward. ∎

Using these properties, we will discuss the structure of the optimal decision policy in the next section.

3. Properties of the Optimal Policy

Now define $g_t(i,w)$ and $\phi_t(i)$ as follows:

$$g_t(i,w) = pu_t(i-1,w) + qw + (1-p)z_t(i-1) - z_t(i), \quad i \geq 1,\ t \geq 0, \tag{3.1}$$

$$\phi_t(i) = \beta(v_{t-1}(i+m) - v_{t-1}(i)) - a, \quad i \geq 0,\ t \geq 1. \tag{3.2}$$

Then, the lemma below holds true.

Lemma 2.

(a) *For $t \geq 1$ and $i \geq 1$, $g_t(i,w)$ is strictly increasing in w, which is also true for $t \to \infty$.*

(b) *$g_t(i,w) = 0$ has a unique solution $w = h_t(i) \in (0,1)$ for $p > 0$ ($\in [0,1)$ for $p = 0$).*

Proof: (a) It is immediate from Lemma 1(d).

(b) Assume $p > 0$. It can be easily proven by induction that $u_t(i,0) = z_t(i)$ for any t and i. Accordingly we get

$$g_t(i,0) = pu_t(i-1,0)+(1-p)z_t(i-1)-z_t(i) = z_t(i-1)-z_t(i) < 0 \tag{3.3}$$

from Lemma 1(b). In addition, it is obvious from Lemma 1(c) that

$$g_t(i,1) = pu_t(i-1,1)+q+(1-p)z_t(i-1)-z_t(i) \geq q+z_t(i-1)-z_t(i) > 0 \tag{3.4}$$

for $i \geq 1$ and $t \geq 0$. From (3.3), (3.4) and the continuity of $g_t(i,w)$ in w, it follows that $g_t(i,w) = 0$ has a unique solution $h_t(i) \in (0,1)$ for $p > 0$. For $p = 0$, the proof is almost the same as above. ∎

Remark: We call $h_t(i)$ a critical value when i bullets and t periods remain. From Lemma 2, the optimal decision policy for shooting becomes as follows; if $g_t(i,w) \geq 0$ ($w \geq h_t(i)$,) then fire, or else don't fire. The optimal policy for replenishment becomes as follows; if $\phi_t(i) \geq 0$, then replenish m bullets, or else don't replenish them.

From Lemma 2(b), it follows that

$$\begin{aligned} 0 = g_t(i+1, h_t(i+1)) &= pu_t(i, h_t(i+1)) + qh_t(i+1) + (1-p)z_t(i) - z_t(i+1) && (3.5) \\ &\geq qh_t(i+1) + z_t(i) - z_t(i+1), && (3.6) \end{aligned}$$

or

$$h_t(i+1) \leq (z_t(i+1) - z_t(i))/q, \quad i \geq 0. \tag{3.7}$$

In particular for $i = 1$, it is true from $g_t(1, h_t(1)) = 0$ that

$$h_t(1) = (z_t(1) - z_t(0))/q. \tag{3.8}$$

The following lemma gives a more detailed description of the relation between $h_t(i)$ and $z_t(j)$.

Lemma 3.

(a) *If $p > 0$, then for $i \geq 1$ and $t \geq 1$,*

$$h_t(i) \geq (<)\, h_t(i+1) \iff h_t(i+1) = (<)\, (z_t(i+1) - z_t(i))/q.$$

When $p = 0$, it always holds true for $i \geq 1$ and $t \geq 1$ that $h_t(i+1) = (z_t(i+1) - z_t(i))/q$.

(b) *For $t \geq 1$ and $p > 0$,*

$$h_t(1) \left\{ \begin{matrix} \geq \\ = \\ < \end{matrix} \right\} h_t(2) \iff 2z_t(1) - z_t(0) - z_t(2) \left\{ \begin{matrix} \geq \\ = \\ < \end{matrix} \right\} 0.$$

(c) *For $i \geq 1$ and $t \geq 1$, if $h_t(i) > (=)\, h_t(i+1)$, then $2z_t(i) - z_t(i-1) - z_t(i+1) > (\geq) 0$.*

(d) *For $i \geq 1$ and $t \geq 1$, if $2z_t(i) - z_t(i-1) - z_t(i+1) < 0$, then $h_t(i) < h_t(i+1)$.*

(e) *Assume $h_t(i) = (z_t(i) - z_t(i-1))/q$ for $i \geq 1$ and $t \geq 1$. Then*

$$2z_t(i) - z_t(i+1) - z_t(i-1) > (=)\, 0 \implies h_t(i) > (\geq) h_t(i+1).$$

Proof: (a) It is immediate from Lemma 2(a) and (3.5).
(b) Using (3.8), we have

$$g_t(2, h_t(1)) = pu_t(1, h_t(1)) + qh_t(1) + (1-p)z_t(1) - z_t(2) = 2z_t(1) - z_t(0) - z_t(2), \quad (3.9)$$

which yields the statement.
(c) From the assumption and (3.7), we have

$$0 <(=)\, g_t(i+1, h_t(i)) = qh_t(i) + z_t(i) - z_t(i+1) \leq 2z_t(i) - z_t(i-1) - z_t(i+1). \quad (3.10)$$

(d) The statement is the contraposition of (c).
(e) Because $qh_t(i) = z_t(i) - z_t(i-1)$ for $i \geq 1$ from the assumption and $qh_t(i+1) \leq z_t(i+1) - z_t(i)$ from (3.7), we have

$$\begin{aligned} 0 &<(=)\, 2z_t(i) - z_t(i+1) - z_t(i-1) \\ &= qh_t(i) + z_t(i) - z_t(i+1) \leq q(h_t(i) - h_t(i+1)), \end{aligned} \quad (3.11)$$

from which we get the statement. ∎

Lemma 4. *The critical value $h_t(i)$ is strictly decreasing (decreasing) in i for a given t if and only if for all $i \geq 1$*

$$2z_t(i) - z_t(i-1) - z_t(i+1) > (\geq)\, 0.$$

Proof: If $h_t(i)$ is strictly decreasing in i, then $2z_t(i) - z_t(i-1) - z_t(i+1) > 0$ for all $i \geq 1$ from Lemma 3(c). The sufficient condition can be proven as follows. From Lemma 3(b), $h_t(1) > h_t(2)$ holds true, hence we have

$$h_t(2) = (z_t(2) - z_t(1))/q \quad (3.12)$$

due to Lemma 3(a). Accordingly, we get $h_t(2) > h_t(3)$ using Lemma 3(e), so

$$h_t(3) = (z_t(3) - z_t(2))/q. \quad (3.13)$$

Repeating the same procedure, we obtain $h_t(i) > h_t(i+1)$ for all $i \geq 1$. In a similar way, we can prove the case that $h_t(i)$ is decreasing in i. ∎

Next, we clarify the relation between $h_t(i)$ and $v_t(j)$.

Lemma 5. *If $h_t(i)$ is strictly decreasing (decreasing) in i for a given t, then for all $i \geq 1$*

$$2v_t(i) - v_t(i-1) - v_t(i+1) > (\geq)\, 0.$$

Proof: We only prove the case that $h_t(i)$ is strictly decreasing in i. For the case that $h_t(i)$ is decreasing in i, it can be proven in a similar way.

From Lemma 3(a) and the assumption of this lemma, we get $h_t(i) = (z_t(i) - z_t(i-1))/q$ for all $i \geq 1$. Hence, we can express $2v_t(i) - v_t(i-1) - v_t(i+1)$ as follows;

$$2v_t(i) - v_t(i-1) - v_t(i+1) = \int_0^{h_t(i+1)} A_t(i,\xi)dF(\xi) + \int_{h_t(i+1)}^{h_t(i)} B_t(i,\xi)dF(\xi) + \int_{h_t(i)}^{1} C_t(i,\xi)dF(\xi) \quad (3.14)$$

where

$$A_t(i,\xi) = 2z_t(i) - z_t(i-1) - z_t(i+1), \quad (3.15)$$

$$B_t(i,\xi) = 2z_t(i) - z_t(i-1) - (pz_t(i) + q\xi + (1-p)z_t(i)) = z_t(i) - z_t(i-1) - q\xi, \quad (3.16)$$

$$\begin{aligned} C_t(i,\xi) &= 2u_t(i,\xi) - u_t(i-1,\xi) - (pu_t(i,\xi) + q\xi + (1-p)z_t(i)) \\ &= (2-p)(pu_t(i-1,\xi) + q\xi + (1-p)z_t(i-1)) - u_t(i-1,\xi) - q\xi - (1-p)z_t(i) \\ &= -(1-p)^2 u_t(i-1,\xi) + (1-p)q\xi - (1-p)z_t(i) + (2-p)(1-p)z_t(i-1). \end{aligned} \quad (3.17)$$

Then, we get $A_t(i,\xi) > 0$ for $0 \leq \xi \leq h_t(i+1)$ from Lemma 4 and $B_t(i,\xi) > 0$ for $h_t(i+1) \leq \xi < h_t(i)$ from Lemma 3(a).

Below, using induction, we shall verify $C_t(i,\xi) > 0$ for $h_t(i) < \xi \leq 1$ and all $i \geq 1$. If $i = 1$, then we get for $h_t(1) < \xi \leq 1$

$$\begin{aligned} C_t(1,\xi) &= -(1-p)^2 u_t(0,\xi) + (1-p)q\xi - (1-p)z_t(1) + (2-p)(1-p)z_t(0) \\ &> -(1-p)^2 z_t(0) + (1-p)qh_t(1) - (1-p)z_t(1) + (2-p)(1-p)z_t(0) \\ &= (1-p)(qh_t(1) + z_t(0) - z_t(1)) = 0. \end{aligned} \quad (3.18)$$

Assume $C_t(i-1,\xi) > 0$ for $h_t(i-1) < \xi \leq 1$. Then, we have for $h_t(i-1) < \xi \leq 1$

$$\begin{aligned} C_t(i,\xi) &= -(1-p)^2(pu_t(i-2,\xi) + q\xi + (1-p)z_t(i-2)) \\ &\qquad +(1-p)q\xi - (1-p)z_t(i) + (2-p)(1-p)z_t(i-1) \\ &= pC_t(i-1,\xi) + (1-p)(2z_t(i-1) - z_t(i-2) - z_t(i)) > 0. \end{aligned} \quad (3.19)$$

Further, we obtain for $h_t(i) < \xi \leq h_t(i-1)$

$$\begin{aligned} C_t(i,\xi) &= -(1-p)^2 z_t(i-1) + (1-p)q\xi - (1-p)z_t(i) + (1-p)(2-p)z_t(i-1) \\ &= (1-p)(q\xi + z_t(i-1) - z_t(i)) > 0. \end{aligned} \quad (3.20)$$

Therefore, we get $C_t(i,\xi) > 0$ for $h_t(i) < \xi \leq 1$ and $i \geq 1$. In addition, it follows by direct calculation that $B_t(i,h_t(i)) = C_t(i,h_t(i)) = 0$. Finally, from the fact that the distribution does not concentrate on only $w = h_t(i)$, we get $2v_t(i) - v_t(i-1) - v_t(i+1) > 0$ for $i \geq 1$. ∎

We have investigated the basic structure of the optimal policy for shooting. In the following sections, the properties of the optimal policy for some special cases will be discussed.

4. Case for Which Replenishment is Always Optimal

In this section, suppose $\phi_t(i) \geq 0$ for all $t \geq 1$ and $i \geq 0$, implying that it is always optimal to replenish m bullets. Then we shall clarify the monotonicity of $h_t(i)$ in i and the condition for $\phi_t(i) \geq 0$ for all $t \geq 1$ and $i \geq 0$.

Theorem 1.

(a) *On the above condition, the critical value $h_t(i)$ is decreasing in i for any $t \geq 1$. Particularly for $p > 0$, it is strictly decreasing in i.*

(b) *It holds true if and only if $a = 0$ that $\phi_t(i) \geq 0$ for any $t \geq 1$ and $i \geq 0$.*

Proof: (a) We only verify the case for $p > 0$. The proof for $p = 0$ is almost the same as below. Since $\phi_t(i) \geq 0$ for all $t \geq 1$ and $i \geq 0$, $z_t(i) = \beta v_{t-1}(i+m) - a$ always holds true. Therefore, using (2.6), we get for $i \geq 1$

$$2z_1(i) - z_1(i-1) - z_1(i+1) = \beta(2v_0(i+m) - v_0(i-1+m) - v_0(i+1+m)) > 0. \quad (4.1)$$

Hence it follows that $2v_1(i) - v_1(i-1) - v_1(i+1) > 0$ for all $i \geq 1$ due to Lemmas 4 and 5. Accordingly we obtain

$$2z_2(i) - z_2(i-1) - z_2(i+1) = \beta(2v_1(i+m) - v_1(i-1+m) - v_1(i+1+m)) > 0 \quad (4.2)$$

for all $i \geq 1$. Repeating the procedure above yields $2z_t(i) - z_t(i-1) - z_t(i+1) > 0$ for all $t \geq 1$ and $i \geq 1$. Hence $h_t(i)$ is strictly decreasing in i for any $t \geq 1$ from Lemma 4.

(b) Now suppose $\phi_t(i) \geq 0$ for all $t \geq 1$ and $i \geq 0$. Because the number of targets the hunter gets over the whole planning horizon is at most $t+1$, $v_t(i)$ has an upper bound for any t, implying that $v_t(i)$ converges as $i \to \infty$. Hence we get

$$\lim_{i\to\infty} \phi_t(i) = -a \geq 0. \tag{4.3}$$

Therefore, it must be that $a = 0$. To go the other way, if $a = 0$, then $\phi_t(i) = \beta(v_{t-1}(i+m) - v_{t-1}(i)) \geq 0$ for all $t \geq 1$ and $i \geq 0$ since $v_t(i)$ is increasing in i. ∎

5. Case for Which No Replenishment is Always Optimal

Next, suppose $\phi_t(i) \leq 0$ for all $t \geq 1$ and $i \geq 0$, implying that it is always optimal not to replenish m bullets. The case is the same as the model in [7] with $c = 0$, in which the conclusion that $h_t(i)$ is strictly decreasing (decreasing) in i for $p > 0$ ($p = 0$) is obtained. Using this fact, we examine the condition for which it is always optimal not to replenish at all.

Theorem 2. *If $\beta mq \leq a$, then $\phi_t(i) \leq 0$ for all $t \geq 1$ and $i \geq 0$. In particular for $\beta = 1$, $\phi_t(i) \leq 0$ for all $t \geq 1$ and $i \geq 0$ if and only if $mq \leq a$.*

Proof: Now, we define the limits of $v_t(i)$, $z_t(i)$, $\phi_t(i)$ and $h_t(i)$ as $t \to \infty$, if they exist, by $v(i)$, $z(i)$, $\phi(i)$ and $h(i)$, respectively. Using Lemma 1(c), we obtain $v_t(i+m) - v_t(i) < mq$ for all t and i, from which we get for all $t \geq 1$ and $i \geq 0$

$$\phi_t(i) = \beta(v_{t-1}(i+m) - v_{t-1}(i)) - a < \beta mq - a \leq 0. \tag{5.1}$$

Thus, the former part of the theorem, which is also the sufficient condition for the latter part, is proven. Now assume $\beta = 1$ and $\phi_t(i) \leq 0$ for all $t \geq 1$ and $i \geq 0$. Then, noting that $z(i) = v(i)$, we get for $i \geq 1$

$$\begin{aligned} v(i) &= \int_0^1 \max\{v(i),\ pu(i-1,\xi) + q\xi + (1-p)v(i-1)\} dF(\xi) \\ &= \int_0^{h(i)} v(i) dF(\xi) + \int_{h(i)}^1 (pu(i-1,\xi) + q\xi + (1-p)v(i-1)) dF(\xi), \end{aligned} \tag{5.2}$$

which is rewritten

$$\int_{h(i)}^1 v(i) dF(\xi) = \int_{h(i)}^1 (pu(i-1,\xi) + q\xi + (1-p)v(i-1)) dF(\xi). \tag{5.3}$$

Now suppose $h(i) < 1$. Then from Lemma 2(a), we obtain $v(i) < pu(i-1,\xi) + q\xi + (1-p)v(i-1)$ for $h(i) < \xi \leq 1$, which contradicts (5.3) because $F(w) < 1$ for $w < 1$. Therefore, $h(i)$ must be equal to 1. Thus, it follows from Lemma 3(a) that

$$\begin{aligned} 1 = h(i) &= (z(i) - z(i-1))/q \\ &= (v(i) - v(i-1))/q, \quad i \geq 1, \end{aligned} \tag{5.4}$$

which yields $v(i+m) - v(i) = mq$ for any i. Thus, we have

$$\phi(i) = v(i+m) - v(i) - a = mq - a \leq 0, \tag{5.5}$$

that is, $mq \leq a$. ∎

Incidentally, as stated in Derman et al. [2], the no replenishment case for certain parameters can be reduced to the sequential assignment problem by Derman et al. [1]. In fact, the critical value for $p=0$ and $\beta=1$ can be expressed in the same expression as Equation (8) in [1], i.e.,

$$h_t(i)=\int_{h_{t-1}(i)}^{h_{t-1}(i-1)}\xi dF(\xi)+h_{t-1}(i)F(h_{t-1}(i))+h_{t-1}(i-1)(1-F(h_{t-1}(i-1))),\quad t\geq 1,\ i\geq 1 \quad (5.6)$$

where $h_t(0)$ is assumed to be 1. Further, we can obtain a relation for any p and β which is regarded as an extension of the above equation.

Corollary 1. *Assume $h_t(0)=1$. If it is always optimal not to replenish, then for $t\geq 1$ and $i\geq 1$*

$$h_t(i)=\beta\Big(\sum_{j=1}^{i}p^{i-j}\int_{h_{t-1}(j)}^{h_{t-1}(j-1)}\xi dF(\xi)+h_{t-1}(i)F(h_{t-1}(i))+(1-p)\sum_{j=1}^{i-1}p^{i-j-1}h_{t-1}(j)(1-F(h_{t-1}(j)))\Big)$$

where let $\sum_{j=1}^{0}x_j=0$ for any series of x_j.

Proof: We can see that if the hunter follows the shoot-look-shoot policy, he does not need to make a decision after each miss, but it suffices to decide up to how many bullets to consume when he finds a target. Because $h_t(i)$ is decreasing in i, if the hunter has $i\,(>j)$ bullets in hand and encounters a target of value w such that $h_t(j+1)\leq w<h_t(j)$, he should continue firing until at least one of the following three events occurs: he obtains the target; loses it in his sight; or the number of remaining bullets becomes j. Therefore, we get for $h_t(j+1)\leq w\leq h_t(j)$

$$\begin{aligned}u_t(i,w)&=\sum_{k=1}^{i-j}\Pr\{\text{the }k\text{th bullet for the present target hits it}\}(w+z_t(i-k))\\&\quad+\sum_{k=1}^{i-j}\Pr\{\text{the target flees just after the }k\text{th miss shot}\}z_t(i-k)\\&\quad+\Pr\{\text{the hunter decides not to shoot the }(i-j+1)\text{st bullet;}\\&\qquad\text{nevertheless it does not escape after the }(i-j)\text{th miss shot}\}z_t(j)\\&=\sum_{k=1}^{i-j}p^{k-1}q(w+z_t(i-k))+\sum_{k=1}^{i-j}p^{k-1}r(1-q)z_t(i-k)+p^{i-j}z_t(j)\\&=\frac{1-p^{i-j}}{1-p}qw+\sum_{k=1}^{i-j-1}p^{k-1}z_t(i-k)+p^{i-j-1}z_t(j).\end{aligned}\qquad(5.7)$$

It follows from Lemma 3(a) that

$$\begin{aligned}qh_{t+1}(i)/\beta&=(z_{t+1}(i)-z_{t+1}(i-1))/\beta\\&=v_t(i)-v_t(i-1)\\&=\int_0^1(u_t(i,\xi)-u_t(i-1,\xi))dF(\xi).\end{aligned}\qquad(5.8)$$

Substituting (5.7), we obtain

$$qh_{t+1}(i)/\beta=\int_0^{h_t(i)}(z_t(i)-z_t(i-1))dF(\xi)+\sum_{j=0}^{i-1}\int_{h_t(j+1)}^{h_t(j)}X_t(i,j,\xi)dF(\xi)\qquad(5.9)$$

where

$$\begin{aligned}X_t(i,j,\xi)&=u_t(i,\xi\mid h_t(j+1)\leq\xi<h_t(j))-u_t(i-1,\xi\mid h_t(j+1)\leq\xi<h_t(j))\\&=p^{i-j-1}q\xi+(1-p)\sum_{k=1}^{i-j-1}p^{k-1}(z_t(i-k)-z_t(i-k-1)).\end{aligned}\qquad(5.10)$$

Therefore, it follows that

$$h_{t+1}(i)/\beta = h_t(i)F(h_t(i)) + \sum_{j=0}^{i-1} p^{i-j-1}\int_{h_t(j+1)}^{h_t(j)} \xi dF(\xi) + (1-p)\sum_{j=0}^{i-1}\int_{h_t(j+1)}^{h_t(j)} \sum_{k=1}^{i-j-1} p^{k-1} h_t(i-k) dF(\xi)$$
$$= h_t(i)F(h_t(i)) + \sum_{j=1}^{i} p^{i-j}\int_{h_t(j)}^{h_t(j-1)} \xi dF(\xi) + (1-p)\sum_{j=1}^{i-1} p^{i-j-1} h_t(j)(1-F(h_t(j))). \quad (5.11)$$

Thus the proof is complete. ∎

6. Case of $m = 1$

Now suppose that only one bullet can be replenished each period. Then, the properties below can be stated. Here, if necessary, we will use the symbols $\phi_t(i,a)$, $u_t(i,w,a)$, $v_t(i,a)$ and $z_t(i,a)$ instead of $\phi_t(i)$, $u_t(i,w)$, $v_t(i)$ and $z_t(i)$ in order to emphasize them to be functions of a.

Theorem 3. *For any* $t \geq 1$,

(a) $h_t(i)$ *is decreasing in* i;

(b) $\phi_t(i)$ *is decreasing in* i;

(c) $\phi_t(i,a)$ *is decreasing in* a.

Proof: (a) It is clear for $m = 1$ that $\phi_t(i)$ is decreasing in i for any $t \geq 1$ if and only if $2v_t(i) - v_t(i-1) - v_t(i+1) \geq 0$ for any $t \geq 0$ and $i \geq 1$. From (2.6), it is true that

$$2v_0(i) - v_0(i-1) - v_0(i+1) = (1-p)p^{i-1}q\mu \geq 0, \quad i \geq 1, \quad (6.1)$$

accordingly for $i \geq 1$,

$$2z_1(i) - z_1(i-1) - z_1(i+1)$$
$$= 2\max\{\beta v_0(i), \beta v_0(i+1)-a\} - \max\{\beta v_0(i-1), \beta v_0(i)-a\} - \max\{\beta v_0(i+1), \beta v_0(i+2)-a\}$$
$$= \begin{cases} \beta(2v_0(i+1)-v_0(i)-v_0(i+2)), & 0 \leq a < \beta(v_0(i+2)-v_0(i+1)), \\ 2(\beta v_0(i+1)-a)-(\beta v_0(i)-a)-\beta v_0(i+1), & \\ & \beta(v_0(i+2)-v_0(i+1)) \leq a < \beta(v_0(i+1)-v_0(i)), \\ 2\beta v_0(i)-(\beta v_0(i)-a)-\beta v_0(i+1), & \beta(v_0(i+1)-v_0(i)) \leq a < \beta(v_0(i)-v_0(i-1)), \\ \beta(2v_0(i)-v_0(i-1)-v_0(i+1)), & \beta(v_0(i)-v_0(i-1)) \leq a. \end{cases} \quad (6.2)$$

From (6.1) and (6.2), we get $2z_1(i) - z_1(i-1) - z_1(i+1) \geq 0$ for any a and $i \geq 1$. Hence it follows from Lemma 4 that $h_1(i)$ is decreasing in i for any a. Now suppose $2z_t(i) - z_t(i-1) - z_t(i+1) \geq 0$, so $h_t(i)$ is decreasing in i and $h_t(i) = (z_t(i) - z_t(i-1))/q$. Therefore, it follows from Lemma 5 that

$$2v_t(i) - v_t(i-1) - v_t(i+1) \geq 0, \quad i \geq 1. \quad (6.3)$$

Hence we have for $i \geq 1$

$$2z_{t+1}(i) - z_{t+1}(i-1) - z_{t+1}(i+1)$$
$$= 2\max\{\beta v_t(i), \beta v_t(i+1)-a\} - \max\{\beta v_t(i-1), \beta v_t(i)-a\} - \max\{\beta v_t(i+1), \beta v_t(i+2)-a\}$$

$$= \begin{cases} \beta(2v_t(i+1)-v_t(i)-v_t(i+2)), & 0 \le a < \beta(v_t(i+2)-v_t(i+1)), \\ \beta(v_t(i+1)-v_t(i))-a, & \beta(v_t(i+2)-v_t(i+1)) \le a < \beta(v_t(i+1)-v_t(i)), \\ -\beta(v_t(i+1)-v_t(i))+a, & \beta(v_t(i+1)-v_t(i)) \le a < \beta(v_t(i)-v_t(i-1)), \\ \beta(2v_t(i)-v_t(i-1)-v_t(i+1)), & \beta(v_t(i)-v_t(i-1)) \le a. \end{cases} \quad (6.4)$$

From (6.3) and (6.4), we have $2z_{t+1}(i) - z_{t+1}(i-1) - z_{t+1}(i+1) \ge 0$ for any a. Thus by induction in terms of t, we obtain $2z_t(i) - z_t(i-1) - z_t(i+1) \ge 0$ for any $t \ge 1$, $i \ge 1$ and $a \ge 0$, so $h_t(i)$ is decreasing in i.

(b) It has been verified already in the proof of (a) that $2v_t(i) - v_t(i-1) - v_t(i+1) \ge 0$ for any $t \ge 0$, $i \ge 1$ and $a \ge 0$, which has yielded the statement.

(c) It is clear that $v_0(i+1,a) - v_0(i,a) - a$ is decreasing in a for any i. Assume that $v_{t-1}(i+1,a) - v_{t-1}(i,a) - a$ is decreasing in a for any i. Using Theorem 3(b), we get

$$z_t(i+1,a)-z_t(i,a)-a = \max\{\beta(v_{t-1}(i+2,a)-v_{t-1}(i+1,a)-a)-(1-\beta)a, \\ \beta v_{t-1}(i+1,a)-(\beta v_{t-1}(i+1,a)-a)-a,\ \beta(v_{t-1}(i+1,a)-v_{t-1}(i,a)-a)-(1-\beta)a\}. \quad (6.5)$$

Hence (6.5) is decreasing in a for any i. Now since $u_t(1,w,a)-u_t(0,w,a)-a = \max\{z_t(1,a)-z_t(0,a)-a,\ qw-a\}$, it is decreasing in a. Furthermore, assuming that $u_t(i,w,a)-u_t(i-1,w,a)-a$ is decreasing in a as the second inductive assumption, we have from Theorem 3(a)

$$u_t(i+1,w,a) - u_t(i,w,a) - a = \max\{z_t(i+1,a) - z_t(i,a) - a,\ qw - a, \\ p(u_t(i,w,a)-u_t(i-1,w,a) - a) + (1-p)(z_t(i,a) - z_t(i-1,a) - a)\}. \quad (6.6)$$

Therefore, it follows that $u_t(i,w,a) - u_t(i-1,w,a) - a$ is decreasing in a for $i \ge 1$, hence $v_t(i+1,a) - v_t(i,a) - a$ is decreasing in a for all i, which also holds for any t by double induction. Because $\phi_t(i,a) = \beta(v_{t-1}(i+1,a) - v_{t-1}(i,a) - a) - (1-\beta)a$, $\phi_t(i,a)$ is also decreasing in a for $t \ge 1$ and $i \ge 0$. ∎

Remark: (b) implies that the critical point for replenishment in terms of i where $\phi_t(i-1) \ge 0 > \phi_t(i)$ is at most one. Concretely speaking, if replenishment is optimal for $i = i'$, then it is also optimal for $i < i'$. Similarly, (c) says that if replenishment is optimal for $a = a'$, then it is also optimal for $a < a'$.

The monotonicity of $h_t(i)$ in i for any t and a is characteristic to the case for which $m = 1$, however, this does not always hold true for $m \ge 2$.

7. Case of $m \ge 2$ and Numerical Examples

Here we shall demonstrate an example that $h_t(i)$ is not always decreasing in i for $m \ge 2$. Let $p > 0$, $m = 2$ and $a = \beta(1+p)pq\mu$. Then, we get

$$z_1(0) = \max\{\beta v_0(0), \beta v_0(2) - a\} = \beta v_0(2) - a = \beta(1-p^2)q\mu, \quad (7.1)$$

$$z_1(1) = \max\{\beta v_0(1), \beta v_0(3) - a\} = \beta v_0(1) = \beta q\mu, \quad (7.2)$$

$$z_1(2) = \max\{\beta v_0(2), \beta v_0(4) - a\} = \beta v_0(2) = \beta(1+p)q\mu. \quad (7.3)$$

Accordingly, we have

$$2z_1(1) - z_1(0) - z_1(2) = -\beta(1-p)pq\mu < 0, \quad (7.4)$$

which means $h_1(1) < h_1(2)$ due to Lemma 3(d).

Below, we depict the results of several numerical examples where a discrete uniform distribution function with 101 mass points equally spaced on $[0,1]$ is used.

(a) When $m = 1$, $h_t(i)$ is decreasing in i even for $a > 0$ (Figure 2(a).)

(b) The non-monotonicity of $h_t(i)$ in i is shown in Figures 2(b,c,d), which also lead us to the conclusion that $h_t(i)$ is not always increasing in t. In [9], the monotonicity of $h_t(i)$ in t has been proven only for the case that it is always optimal not to replenish at all with $\beta = 1$.

(c) So far we have not investigated the relations between $h_t(i)$ and parameters a, q and r; it is quite intractable to reveal them theoretically. All of the numerical examples we calculate show that $h_t(i)$ is increasing in a and r and decreasing in q. Figure 2(e) is an example of the relation of $h_t(i)$ to a.

(d) Figure 2(f) tells us the fact that $h_t(i)$ is not always monotone in β. We also get examples where $h_t(i)$ is not monotone in m.

Such a non-monotonicity of $h_t(i)$ in i may fit our intuition in the following case where, for a certain j, $\phi_t(i) \geq 0$ if $i \leq j$ or else $\phi_t(i) < 0$. First, suppose the hunter has $j+1$ bullets in hand. If it is decided not to shoot, then he needs not replenish m bullets at the period, or else he must replenish them by paying cost a according to the optimal policy for replenishment. Therefore, his behavior for shooting may become a little cautious, that is, $h_t(j+1)$ may become a little high. Next, suppose he has j bullets. Then, his behavior may be more or less active since it is already decided to replenish them at the period whether he decides or not to shoot, so $h_t(j)$ may become a little low.

On the other hand, in terms of the optimal policy for replenishment, it has been clarified that $\phi_t(i, a)$ is monotone in i and a for $m = 1$. However, we have not been able to prove the property for $m \geq 2$ and find any counterexamples.

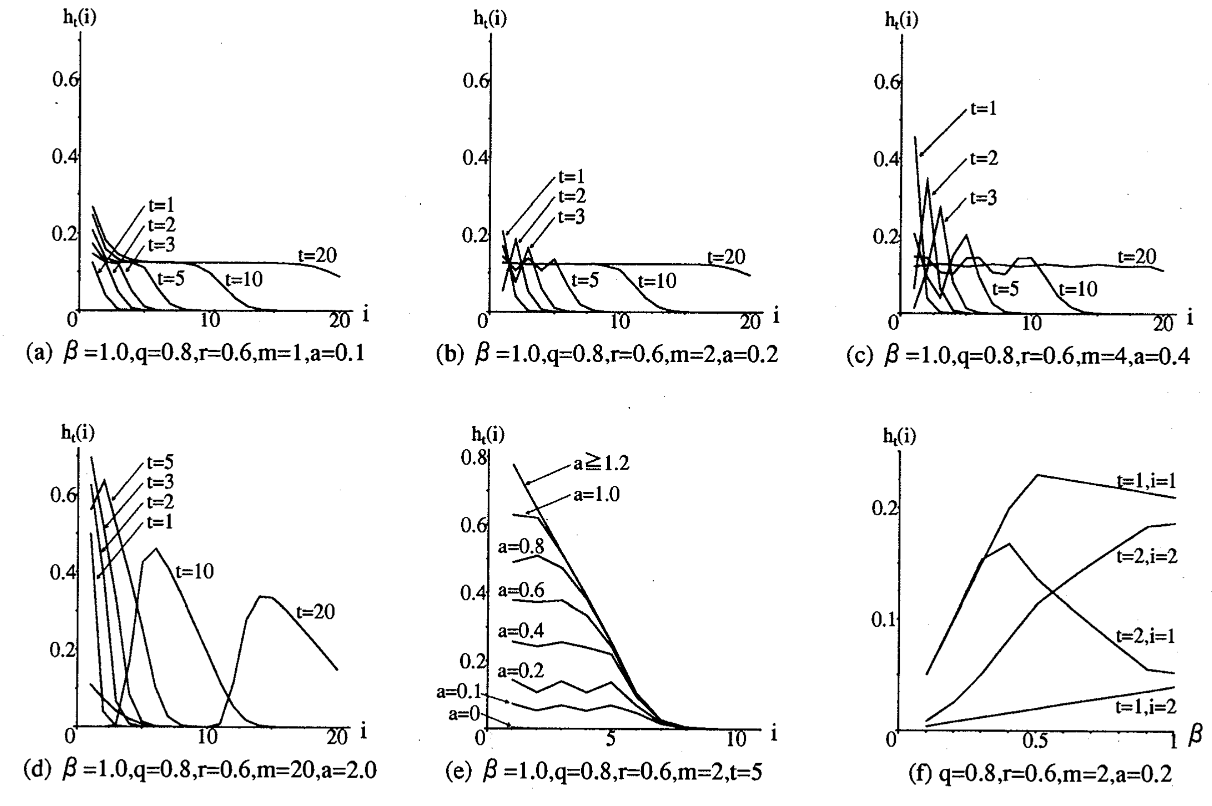

Figure 2. Numerical Examples

8. Conclusions and a Future Study

We have considered a discrete-time sequential allocation problem with countable resources which can be replenished each point in time, and the following conclusions are obtained:

(a) The necessary and sufficient condition for $\phi_t(i) \geq 0$ for all $t \geq 1$ and $i \geq 0$ is $a = 0$, for which $h_t(i)$ is always decreasing in i.

(b) If $\beta mq \leq a$, then $\phi_t(i) \leq 0$ for all $t \geq 1$ and $i \geq 0$, that is, it is optimal not to replenish at all. In particular for $\beta = 1$, $\phi_t(i) \leq 0$ for all $t \geq 1$ and $i \geq 0$ if and only if $mq \leq a$.

(c) If $m = 1$, $h_t(i)$ is always decreasing in i. Furthermore, when $m = 1$, if replenishment is optimal for $i = i'$ ($a = a'$,) then it is also optimal for $i < i'$ ($a < a'$.)

(d) If $m \geq 2$, then $h_t(i)$ is not always decreasing in i.

As a future study, it must be interesting to consider the problem with volley policy. Below, we will examine the volley problem roughly where clearly, escape probability r makes no sense. Assuming $r = 0$ as a matter of form, we get $p = 1 - q$. Then, (2.1)–(2.4) still hold if (2.2) is replaced by

$$u_t(i, w) = \max_{j=0,1,\ldots,i} \{(1 - p^j)w + z_t(i - j)\}. \tag{8.1}$$

Furthermore, final conditions (2.5) and (2.6) are replaced by $u_0(i, w) = (1 - p^i)w$ and $v_0(i) = (1 - p^i)\mu$, respectively. Let $j_t(i, w)$ be the smallest number of j maximizing the term inside the braces in the right hand side of (8.1). Because the hunter who has a rifle with multiple muzzles must decide how many bullets he fires in salvo, the optimal firing policy is characterized by $j_t(i, w)$.

It may be intuitive that $j_t(i, w)$ is increasing in i. However, we can also show a counterexample in this volley model. Now assume $(\beta, \mu, p, m, a) = (0.9, 0.5, 0.5, 2, 0.1)$. Then we get $z_1(i) = 19/80$, $47/160$, $27/80$, and $63/160$ for $i = 0, 1, 2$, and 3, respectively. Furthermore, supposing $w = 0.1$, we obtain

$$u_1(2, w) = \max\{z_1(2), (1-p)w+z_1(1), (1-p^2)w+z_1(0)\} = (1-p)w+z_1(1), \tag{8.2}$$
$$u_1(3, w) = \max\{z_1(3), (1-p)w+z_1(2), (1-p^2)w+z_1(1), (1-p^3)w+z_1(0)\} = z_1(3). \tag{8.3}$$

The above equations yield, respectively, $j_1(2, 0.1) = 1$ and $j_1(3, 0.1) = 0$, hence $j_1(2, 0.1) > j_1(3, 0.1)$. This is caused by a reason similar to that for the non-monotonicity of $h_t(i)$ in i.

Acknowledgments

The author would like to thank Associate Professor Seizo Ikuta of the University of Tsukuba and the referees for their comments and suggestions.

References

[1] C. Derman, G. J. Lieberman and S. M. Ross: A Sequential Stochastic Assignment Problem, *Management Science*, Vol.18, (1972), 349-355.

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

[3] T. Kisi: Suboptimal Decision Rule for Attacking Targets of Opportunity, *Naval Research Logistics Quarterly*, Vol.23, (1976), 525-533.

[4] D. V. Mastran and C. J. Thomas: Decision Rule for Attacking Targets of Opportunity, *Naval Research Logistics Quarterly*, Vol.20, (1973), 661-672.

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

[6] T. Namekata, Y. Tabata and T. Nishida: A Sequential Allocation Problem with Perishable Goods, *Journal of the Operations Research Society of Japan*, Vol.24, (1981), 202-212.

[7] T. Namekata, Y. Tabata and T. Nishida: Some Remarks on Sequential Allocation Problem over Unknown Number of Periods, *Journal of the Operations Research Society of Japan*, Vol.24, (1981), 339-349.

[8] G. P. Prastacos: Optimal Sequential Investment Decisions under Conditions of Uncertainty, *Management Science*, Vol.29, (1983), 118-134.

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

[10] M. Sato: A Sequential Allocation Problem with Search Cost where the Shoot-Look-Shoot Policy is Employed, *Journal of the Operations Research Society of Japan*, Vol.39, (1996), 435-454.

Masahiro SATO
Research Department
of Communications School
Ground Self Defense Force
Kubiri, Yokosuka, Kanagawa 203
Japan