AN EXPONENTIAL TWO-ARMED BANDIT PROBLEM WITH ONE ARM KNOWN UNDER BATCH SAMPLING

Toshio Hamada*

There are two kinds of experiments e_0 and e_1 , and by performing e_0 or e_1 , an observation is obtained from the exponential distribution with a parameter 1 or u , respectively. Although the true value of u is unknown, u has a gamma distribution as the prior distribution. The action a_i ($i=0, 1$) is defined to select e_i and perform it simultaneously m times. An n -stage sequential decision problem, in which a_0 or a_1 is selected at each stage by considering the information obtained up to that stage in order to maximize the expected sum of mn observations, is constructed and formulated by dynamic programming and the optimal strategy is obtained.

The results of this paper illustrate how to calculate the critical value which has an important role in the optimal strategy numerically. The results also gives the optimal strategy for the integer-valued parameter case of the gamma two-armed bandit problem with one arm known.

Key words: Bandit problem. Dynamic programming. Bayesian learning. Optimal stopping. Exponential distribution. Gamma distribution.

1. Introduction

Suppose that there are two kinds of experiments e_0 and e_1 and by performing e_0 or e_1 an observation is obtained from the exponential distribution with parameter 1 or u , respectively. There is the prior information that u has a gamma distribution with parameters w and α as prior distribution. Since the gamma distribution is the conjugate prior distribution for the parameter of the exponential distribution, the posterior distribution after getting an observation is also the Gamma distribution.

We consider an n -stage sequential decision problem in which one of two arms a_0 and a_1 is selected and pulled, where to pull a_i ($i=0$ or 1) means to perform e_i simultaneously m times in parallel and to get observations of m independently, as well as identically distributed random variables. The reward obtained at each stage is the sum of m observations and the objective of this problem is to maximize the expected sum of mn observations over n stages. This means that if a_1 is pulled and m observations $Z_1, Z_2, \dots, Z_m$ are obtained when the prior distribution of u is the gamma distribution with parameters w and α , then the posterior distribution of u is the gamma distribution with parameters $w+Z_1+Z_2+\dots+Z_m$ and $\alpha+m$ (See, for example, DeGroot [5]).

The simplest case of $m=1$ for this problem is the exponential two-armed bandit problem with one arm known. The Bernoulli distribution version of

Received March, 1994. Revised July, 1995. Accepted August, 1995.

* Department of Management Science, Kobe University of Commerce, Gakuen-Nishimachi, Nishi-ku, Kobe 651-21 Japan

this problem with $m=1$ has been discussed in many papers. One of the earliest reports was by Bradt, Johnson and Karlin [4], Yakowitz [15] used a computer to derive the optimal strategy, and Berry and Fristedt [2] discussed this with arbitrary discount sequence. The uniform distribution version of the problem was solved in Hamada [10] and the optimal strategy was given by using the critical value function which was a solution of an equation derived by solving the recursive formula of dynamic programming. The case of the exponential distribution is also discussed in Hamada [11]. Infinite time horizon and discounted reward version of the problem with $m=1$ has also been discussed in many papers. For example, Bellman [1] solved this by dynamic programming, Gittins and Jones [8] proposed an index rule which has an important role in the analysis of this kind of problem. The importance and usefulness of the index rule are given in Gittins [6, 7] and Glazebrook [9] gives a new proof of the index rule. The case of uniform distribution is discussed in Woodroffe [14], Kalin and Theodorescu [13], and Hamada [12]. Many papers related in this area are listed in Berry and Fristedt [3] as a bibliography.

Since the sum of m random variables from the exponential distribution with a parameter u is also considered as a random variable from the gamma distribution with parameters m and u , this problem is also considered as a gamma two-armed bandit problem with one arm known. Since the conjugate prior distribution for an unknown parameter u of gamma distribution is also a gamma distribution, the problem is easily formulated by dynamic programming, however, it is difficult to deal with the gamma distribution case because of the difficulty of solving the recursive equations of this case. The recursive equations of the case of batch sampling of the exponential distributions is more tractable in analysis than the case of gamma distribution.

The model considered in this paper is useful in analyzing the model described as follows: Suppose that we have been using a light bulb e_0 for a long time and we know the distribution of its lifetime. A new kind of light bulb e_1 comes into the market quite recently and we want to use e_1 in place of e_0 . Suppose that the lifetime of e_i ($i=0, 1$) is distributed in an exponential distribution with parameter u_i ($i=0, 1$). Although the value of u_0 is known since it has been used quite frequently, the value of u_1 is unknown because it is unfamiliar to us. We assume that u_1 has a gamma distribution with parameters w and α as the prior distribution. There are two kinds of boxes, a_0 and a_1 , and the box a_i ($i=0, 1$) contains m identical parts e_i . We use light bulbs one at a time which need replacement when its life is over. At first, we decide which kind of box to buy. If we bought a_1 , we use e_1 for m times and obtain the information about its lifetime. After all the light bulbs of the box are broken, we have to decide which kind of box to buy again. We can repeat to make a decision after m observations have been obtained. This kind of model is useful in the analysis of this kind of replacement policy.

In Section 2, the exponential two-armed bandit problem with one arm known under batch sampling is formulated by the principle of optimality of dynamic programming and the optimal strategy is derived by defining the

critical value functions. In Section 3, some preliminary identical equations are derived. In Section 4, the recursive formula is solved explicitly as is calculation of the critical value. In Section 5, the gamma two-armed bandit problem with one arm known and an integer value parameter is also solved.

2. Formulation by dynamic programming

Let $F_{n,m}(w, \alpha)$ be the maximum expected total reward over n stages when the current information is (w, α) and there are n stages remaining. Then, the problem is formulated by the principle of optimality of dynamic programming as follows:

$$(2.1) \quad F_{n,m}(w, \alpha) = \max\{F_{n,m}^0(w, \alpha), F_{n,m}^1(w, \alpha)\}$$

for $n=1, 2, 3, \dots$ and $F_{0,m}(w, \alpha)=0$, where

$$(2.2) \quad F_{n,m}^0(w, \alpha) = m + F_{n-1,m}(w, \alpha)$$

and

$$(2.3) \quad F_{n,m}^1(w, \alpha) = \int_0^\infty \left[\int_0^\infty \int_0^\infty \cdots \int_0^\infty \left\{ \sum_{l=1}^m z_l + F_{n-1,m} \left(w + \sum_{l=1}^m z_l, \alpha + m \right) \right\} \right. \\ \left. \times \prod_{l=1}^m u \exp(-uz_l) dz_m \cdots dz_2 dz_1 \right] (\Gamma(\alpha))^{-1} w^\alpha u^{\alpha-1} e^{-wu} du.$$

Then, (2.3) can be rewritten as follows:

$$(2.4) \quad F_{n,m}^1(w, \alpha) = mw(\alpha-1)^{-1} \\ + \int_0^\infty \int_0^\infty \cdots \int_0^\infty F_{n-1,m} \left(w + \sum_{l=1}^m z_l, \alpha + m \right) {}_{\alpha+m-1}\mathbf{P}_m \\ \cdot w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1$$

where ${}_a\mathbf{P}_k = \alpha(\alpha-1) \cdots (\alpha-k+1)$ for any $\alpha > 1$ and a positive integer k . Furthermore, let ${}_{\alpha-1}\mathbf{P}_0 = 1$ and ${}_{\alpha-2}\mathbf{P}_{-1} = (\alpha-1)^{-1}$ for $\alpha > 1$.

THEOREM 1. For $w > 0$, $\alpha > 1$, $m \geq 1$ and $n \geq 1$, once a_0 is optimal at any stage, it is optimal thereafter.

PROOF. The proof is done by the same way as that of Theorem 1 of Hamada [10]. $\square$

From this theorem, $F_{n,m}^0(w, \alpha) = mn$ if $F_{n,m}^0(w, \alpha) > F_{n,m}^1(w, \alpha)$ and $F_{n,m}^1(w, \alpha) \geq mn$ if $F_{n,m}^0(w, \alpha) \leq F_{n,m}^1(w, \alpha)$. Therefore, mn is used in (2.1) in place of $F_{n,m}^0(w, \alpha)$ and (2.1) is rewritten as follows:

$$(2.5) \quad F_{n,m}(w, \alpha) = \max\{mn, F_{n,m}^1(w, \alpha)\}$$

Let $z_l = wy_l$ for $l=1, 2, \dots, m$. Then,

$$(2.6) \quad F_{n,m}^1(w, \alpha) = mw(\alpha-1)^{-1} \\ + \int_0^\infty \int_0^\infty \cdots \int_0^\infty F_{n-1,m} \left(w \left(1 + \sum_{l=1}^m y_l \right), \alpha + m \right) \\ \cdot {}_{\alpha+m-1}\mathbf{P}_m \left(1 + \sum_{l=1}^m y_l \right)^{-\alpha-m} dy_m \cdots dy_2 dy_1$$

From this equation, the following lemma is derived.

LEMMA 1. For $w > 0$, $\alpha > 1$, $m \geq 1$ and $n \geq 1$, $F_{n,m}^1(w, \alpha)$ is continuous and strictly increasing in w and continuous and strictly decreasing in α .

PROOF. The proof is done by induction on n in the same way as that of Theorem 1 of Hamada [11] by using the identity

$$\int_0^\infty \int_0^\infty \cdots \int_0^\infty {}_{\alpha+m-1}\mathbf{P}_m \left(1 + \sum_{i=1}^m y_i \right)^{-\alpha-m} dy_m \cdots dy_2 dy_1 = 1$$

for any $\alpha > 1$ and positive integer m . $\square$

LEMMA 2. For $w > 0$, $\alpha > 1$, $m \geq 1$ and $n \geq 1$, $F_{n,m}^1(w, \alpha) < mn$ if $w < (\alpha-1)/n$ and $F_{n,m}^1(w, \alpha) > mn$ if $w > \alpha-1$.

PROOF. The proof is done by induction on n by using (2.5) and (2.6). $\square$

THEOREM 2. For $w > 0$, $\alpha > 1$, $m \geq 1$ and $n \geq 1$,

- (i) $F_{n,m}^1(w, \alpha) = mn$ has a unique root $r_{n,m}(\alpha)$ in the interval $[(\alpha-1)/n, \alpha-1]$,
- (ii) $r_{n,m}(\alpha)$ is strictly increasing in α and nonincreasing in n , and
- (iii) $r_{n,m}(\alpha) < r_{n-1,m}(\alpha+m)$

PROOF. The existence and uniqueness of the root $r_{n,m}(\alpha)$ in (i) is proved by using Lemmas 1 and 2. The monotonicity (ii) in α is derived from Lemma 1 and monotonicity in n is proved as that of Theorem 2.3 of Hamada [10]. The inequality (iii) is the immediate consequence of (ii). $\square$

From this theorem,

$$F_{n,m}(w, \alpha) = \begin{cases} mn, & \text{if } 0 < w < r_{n,m}(\alpha) \\ F_{n,m}^1(w, \alpha), & \text{if } r_{n,m}(\alpha) \leq w \end{cases}$$

and the optimal strategy is derived as follows: "Pull a_1 if and only if $w \geq r_{n,m}(\alpha)$ ".

3. Calculation of critical values

In this section we solve the recursive equations (2.5) and (2.6) explicitly and show how to calculate the values of $r_{n,m}(\alpha)$ and $F_{n,m}(w, \alpha)$. We derive several identities which are useful in the proof of Theorem 3 which gives the optimal strategy for the bandit problem.

LEMMA 3. For $w > 0$, $\alpha > 1$ and $m \geq 1$,

$$\begin{aligned} & \int_0^\infty \int_0^\infty \cdots \int_0^\infty \left(w + \sum_{i=1}^m z_i \right) (\alpha + m - 1)^{-1} {}_{\alpha+m-1}\mathbf{P}_m w^\alpha \left(w + \sum_{i=1}^m z_i \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\ & = w(\alpha-1)^{-1}. \end{aligned}$$

PROOF. The proof is done by induction on n . $\square$

LEMMA 4. For $w > 0$, $\alpha > 1$ and $m \geq 1$,

$$\int_0^{c-w} \int_0^{c-w-z_1} \cdots \int_0^{c-w-z_1-\cdots-z_{m-1}} dz_m \cdots dz_2 dz_1 = \sum_{l=0}^m \frac{(c-b)^{m-l}}{(m-l)!} \frac{(b-w)^l}{l!} = \frac{(c-w)^m}{m!}.$$

PROOF. The proof is done by induction on n . $\square$

REMARK. Although the left hand side of the equation

$$\sum_{l=0}^m \frac{(c-b)^{m-l}}{(m-l)!} \frac{(b-w)^l}{l!} = \frac{(c-w)^m}{m!}$$

is more complex than the right hand side, we prefer the expression of the left hand side to that of the right one.

LEMMA 5. For $m \geq 1$, $c > w > 0$, $\alpha > 1$ and $b > 0$,

$$\begin{aligned} & \int_0^{c-w} \int_0^{c-w-z_1} \dots \int_0^{c-w-z_1-\dots-z_{m-1}} \left[m - m \left(w + \sum_{l=1}^m z_l \right) (\alpha + m - 1)^{-1} \right] \\ & \times {}_{a+m-1}\mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \dots dz_2 dz_1 = m w^\alpha \left[w^{-\alpha} - (\alpha-1)^{-1} w^{-\alpha+1} \right. \\ & \left. + \sum_{k=0}^{m-1} \sum_{l=k}^{m-1} \left\{ -\frac{a+l-1\mathbf{P}_l}{c^{\alpha+l}} + \frac{a+l-2\mathbf{P}_{l-1}}{c^{\alpha+l-1}} \right\} \frac{(c-b)^{l-k}}{(l-k)!} \frac{(b-w)^k}{k!} \right]. \end{aligned}$$

PROOF. Since ${}_{a-1}\mathbf{P}_0 = 1$ and ${}_{a-2}\mathbf{P}_{-1} = (\alpha-1)^{-1}$, it is easily found that this lemma is true for $m=1$. Suppose that this lemma is true for $m \geq 1$. The identity also holds for the case of a multiple integral of $m+1$ variables, $z_1, z_2, \dots, z_{m+1}$, integrating by z_{m+1} first and using the inductive hypothesis. $\square$

LEMMA 6. For $m \geq 1$, $c > w > 0$, and $\alpha > 1$,

$$\begin{aligned} & \int_0^{c-w} \int_0^{c-w-z_1} \dots \int_0^{c-w-z_1-\dots-z_{m-1}} \frac{\left(b-w - \sum_{l=1}^m z_l \right)^k}{k!} dz_m \dots dz_2 dz_1 \\ & = \frac{(b-w)^{k+m}}{(k+m)!} - \sum_{j=0}^{m-1} \sum_{l=j}^{m-1} \frac{(b-c)^{k+m-l}}{(k+m-l)!} \frac{(c-b)^{l-j}}{(l-j)!} \frac{(b-w)^j}{j!}. \end{aligned}$$

PROOF. The proof is done by induction on m and using Lemma 4. $\square$

By using these lemmas, the recursive equations (2.5) and (2.6) can be solved explicitly.

THEOREM 3. For $m \geq 1$, $w > 0$ and $\alpha > 1$,

$$F_{1,m}(w, \alpha) = \begin{cases} m, & \text{if } 0 < w < r_{1,m}(\alpha), \\ mw(\alpha-1)^{-1}, & \text{if } r_{1,m}(\alpha) \leq w, \end{cases}$$

and for $n \geq 2$,

$$F_{n,m}(w, \alpha) = \begin{cases} nm, & \text{if } 0 < w < r_{n,m}(\alpha), \\ A_{n,m,s}(w, \alpha), & \text{if } r_{s,m}(\alpha + (n-s)m) \leq w < r_{s-1,m}(\alpha + (n-s+1)m) \\ & \text{for } 2 \leq s \leq n, \\ nmw/(\alpha-1), & \text{if } r_{1,m}(\alpha + (n-1)m) \leq w, \end{cases}$$

where

$$\begin{aligned} A_{n,m,s}(w, \alpha) &= (n-s+1)mw/(\alpha-1) + (s-1)m \\ &+ mw^\alpha \sum_{k=0}^{(n-s+1)m-1} \sum_{i=1}^{s-1} C_{n,k,i,m}(\alpha) \{r_{1,m}(\alpha + (n-1)m) - w\}^k / k! \end{aligned}$$

for $2 \leq s \leq n$,

$$\begin{aligned} C_{n,k,1,m}(\alpha) &= -{}_{a+k-1}\mathbf{P}_k \{r_{1,m}(\alpha + (n-1)m)\}^{-\alpha-k} \\ &+ {}_{a+k-2}\mathbf{P}_{k-1} \{r_{1,m}(\alpha + (n-1)m)\}^{-\alpha-k+1} \end{aligned}$$

for $0 \leq k \leq m-1$,

$$\begin{aligned}
 C_{n,k,i,m}(\alpha) = & \sum_{l=k}^{m-1} \left[-_{a+l-1} \mathbf{P}_l \{r_{i,m}(\alpha + (n-i)m)\}^{-a-l} \right. \\
 & +_{a+l-2} \mathbf{P}_{l-1} \{r_{i,m}(\alpha + (n-i)m)\}^{-a-l+1} \\
 & - \sum_{j=0}^{(n-1-i)m-1} \sum_{t=1}^i C_{n-1,j,t,m}(\alpha+m) \{r_{1,m}(\alpha + (n-1)m) \\
 & - r_{i,m}(\alpha + (n-i)m)\}^{j+m-l} {}_{a+m-1} \mathbf{P}_m / (j+m-l)! \\
 & + \sum_{j=0}^{(n-i)m-1} \sum_{t=1}^{i-1} C_{n-1,j,t,m}(\alpha+m) \{r_{1,m}(\alpha + (n-1)m) \\
 & - r_{i,m}(\alpha + (n-i)m)\}^{j+m-l} {}_{a+m-1} \mathbf{P}_m / (j+m-l)! \left. \right] \\
 & \times \{r_{i,m}(\alpha + (n-i)m) - r_{1,m}(\alpha + (n-1)m)\}^{l-k} / (l-k)!
 \end{aligned}$$

for $2 \leq i \leq n-2$ and $0 \leq k \leq m-1$,

$$\begin{aligned}
 C_{n,k,n-1,m}(\alpha) = & \sum_{l=k}^{m-1} \left[-_{a+l-1} \mathbf{P}_l \{r_{n-1,m}(\alpha+m)\}^{-a-l} \right. \\
 & +_{a+l-2} \mathbf{P}_{l-1} \{r_{n-1,m}(\alpha+m)\}^{-a-l+1} \\
 & + \sum_{j=0}^{m-1} \sum_{t=1}^{n-2} C_{n-1,j,t,m}(\alpha+m) \{r_{1,m}(\alpha + (n-1)m) \\
 & - r_{n-1,m}(\alpha+m)\}^{j+m-l} {}_{a+m-1} \mathbf{P}_m / (j+m-l)! \left. \right] \\
 & \times \{r_{n-1,m}(\alpha+m) - r_{1,m}(\alpha + (n-1)m)\}^{l-k} / (l-k)!,
 \end{aligned}$$

$0 \leq k \leq m-1$, and

$$C_{n,k,i,m}(\alpha) = C_{n-1,k-m,i,m}(\alpha+m) {}_{a+m-1} \mathbf{P}_m$$

for $1 \leq i \leq n-2$ and $m \leq k \leq (n-i)m-1$. Furthermore,

$$r_{1,m}(\alpha) = \alpha - 1$$

and for $n \geq 2$, $r_{n,m}(\alpha)$ is the unique root of the following equation of w ,

$$-1 + w/(\alpha-1) + w^a \sum_{k=0}^{m-1} \sum_{i=1}^{n-1} C_{n,k,i,m}(\alpha) \{r_{1,m}(\alpha + (n-1)m) - w\}^k / k! = 0.$$

PROOF. It is trivial for $n=1$. For $n=2$, since

$$\begin{aligned}
 F_{2,m}^1(w, \alpha) = & mw(\alpha-1)^{-1} \\
 & + \int_0^\infty \int_0^\infty \cdots \int_0^\infty F_{1,m} \left(w + \sum_{l=1}^m z_l, \alpha+m \right) {}_{a+m-1} \mathbf{P}_m w^a \left(w + \sum_{l=1}^m z_l \right)^{-a-m} dz_m \cdots dz_2 dz_1.
 \end{aligned}$$

and

$$F_{1,m} \left(w + \sum_{l=1}^m z_l, \alpha+m \right) = \begin{cases} m, & \text{if } 0 < w + \sum_{l=1}^m z_l < r_{1,m}(\alpha+m) \\ m \left(w + \sum_{l=1}^m z_l \right) / (\alpha+m-1), & \text{if } r_{1,m}(\alpha+m) \leq w + \sum_{l=1}^m z_l, \end{cases}$$

two cases, (i) $r_{1,m}(\alpha+m) \leq w$ and (ii) $0 < w < r_{1,m}(\alpha+m)$, have to be considered. For case (i),

$$\begin{aligned}
F_{2,m}^1(w, \alpha) &= mw(\alpha-1)^{-1} \\
&+ \int_0^\infty \int_0^\infty \cdots \int_0^\infty m \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1} {}_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\
&= 2mw(\alpha-1)^{-1},
\end{aligned}$$

where the last equality is derived from Lemma 3. For case (ii), we use $r_{1,m}$ in place of $r_{1,m}(\alpha+w)$. Then,

$$\begin{aligned}
F_{2,m}^1(w, \alpha) &= mw(\alpha-1)^{-1} \\
&+ \int_0^{r_{1,m}-w} \int_0^{r_{1,m}-w-z_1} \cdots \int_0^{r_{1,m}-w-z_1-\cdots-z_{m-1}} \left[m - m \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1} \right] \\
&\times {}_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\
&+ \int_0^\infty \int_0^\infty \cdots \int_0^\infty m \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1} {}_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\
&= 2mw(\alpha-1)^{-1} + mw^\alpha \left[w^{-\alpha} - (\alpha-1)^{-1} w^{-\alpha+1} + \sum_{k=0}^{m-1} \sum_{l=k}^{m-1} \left\{ -\frac{a+l-1 \mathbf{P}_l}{r_{1,m}^{\alpha+l}} + \frac{a+l-2 \mathbf{P}_{l-1}}{r_{1,m}^{\alpha+l-1}} \right\} \right. \\
&\quad \left. \cdot \frac{(r_{1,m}-r_{1,m})^{l-k}}{(l-k)!} \frac{(r_{1,m}-w)^k}{k!} \right] \\
&= mw(\alpha-1)^{-1} + m + \sum_{k=0}^{m-1} \sum_{i=1}^1 C_{2,k,i,m}(\alpha) \frac{(r_{1,m}-w)^k}{k!},
\end{aligned}$$

where Lemmas 3 and 5 are used and $C_{2,k,i,m}(\alpha)$ in the last equality is defined by

$$C_{2,k,1,m}(\alpha) = -\frac{a+k-1 \mathbf{P}_k}{r_{1,m}^{\alpha+k}} + \frac{a+k-2 \mathbf{P}_{k-1}}{r_{1,m}^{\alpha+k-1}}.$$

Since it is derived from (iii) of Theorem 2 that $r_{2,m}(\alpha) < r_{1,m}(\alpha+m)$, this theorem is true for $n=2$. Suppose that it is true for $n \geq 2$. Then,

$$\begin{aligned}
F_{n+1,m}^1(w, \alpha) &= mw(\alpha-1)^{-1} \\
&+ \int_0^\infty \int_0^\infty \cdots \int_0^\infty F_{n,m} \left(w + \sum_{l=1}^m z_l, \alpha+m \right) {}_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1
\end{aligned}$$

and

$$F_{n,m} \left(w + \sum_{l=1}^m z_l, \alpha+m \right) = \begin{cases} nm, & \text{if } 0 < w + \sum_{l=1}^m z_l < r_{n,m}(\alpha+m), \\ A_{n,m,s} \left(w + \sum_{l=1}^m z_l, \alpha+m \right), & \text{if } r_{s,m}(\alpha+(n-s+1)m) \leq w < r_{s-1,m}(\alpha+(n-s+2)m) \text{ for } 2 \leq s \leq n, \\ nm \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1}, & \text{if } r_{1,m}(\alpha+nm) \leq w, \end{cases}$$

$n+1$ cases, $r_{1,m}(\alpha+nm) \leq w$, $r_{s,m}(\alpha+(n-s+1)m) \leq w < r_{s-1,m}(\alpha+(n-s+2)m)$ for $2 \leq s \leq n$, and $0 < w < r_{n,m}(\alpha+m)$, have to be considered. For the first case, $r_{1,m}(\alpha+nm) \leq w$

$$\begin{aligned}
F_{n+1,m}^1(w, \alpha) &= mw(\alpha-1)^{-1} \\
&+ \int_0^\infty \int_0^\infty \cdots \int_0^\infty nm \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1} {}_{\alpha+m-1}P_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\
&= (n+1)mw(\alpha-1)^{-1},
\end{aligned}$$

where the last equality is derived from Lemma 3. For $r_{2,m}(\alpha+(n-1)m) \leq w < r_{1,m}(\alpha+nm)$, we use $r_{1,m}$ in place of $r_{1,m}(\alpha+nm)$. Then,

$$\begin{aligned}
F_{n+1,m}^1(w, \alpha) &= mw(\alpha-1)^{-1} \\
&+ \int_0^{r_{1,m}-w} \int_0^{r_{1,m}-w-z_1} \cdots \int_0^{r_{1,m}-w-z_1-\cdots-z_{m-1}} \left[A_{n,m,2} \left(w + \sum_{l=1}^m z_l, \alpha+m \right) \right. \\
&\quad \left. - nm \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1} \right] \times {}_{\alpha+m-1}P_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\
&+ \int_0^\infty \int_0^\infty \cdots \int_0^\infty nm \left(w + \sum_{l=1}^m z_l \right) (\alpha+m-1)^{-1} {}_{\alpha+m-1}P_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} dz_m \cdots dz_2 dz_1 \\
&= (n+1)mw(\alpha-1)^{-1} \\
&+ \int_0^{r_{1,m}-w} \int_0^{r_{1,m}-w-z_1} \cdots \int_0^{r_{1,m}-w-z_1-\cdots-z_{m-1}} \left[\left\{ m-m \left(w + \sum_{l=1}^m z_l \right)^{\alpha+m} \right. \right. \\
&\quad \cdot (\alpha+m-1)^{-1} \Big\} {}_{\alpha+m-1}P_m w^\alpha \left(w + \sum_{l=1}^m z_l \right)^{-\alpha-m} + mw^\alpha \sum_{k=0}^{(n-1)m-1} \sum_{i=1}^1 C_{n,k,i,m}(\alpha+m) \\
&\quad \cdot \left(r_{1,m}-w-\sum_{l=1}^m z_l \right)^k {}_{\alpha+m-1}P_m / k! \Big] dz_m \cdots dz_2 dz_1.
\end{aligned}$$

By using Lemmas 5 and 6,

$$\begin{aligned}
F_{n+1,m}^1(w, \alpha) &= nmw(\alpha-1)^{-1} + m + mw^\alpha \sum_{k=0}^{m-1} \sum_{l=k}^{m-1} \left\{ -\frac{a+l-1 P_l}{r_{1,m}^{\alpha+l}} + \frac{a+l-2 P_{l-1}}{r_{1,m}^{\alpha+l-1}} \right\} \\
&\quad \cdot \frac{(r_{1,m}-r_{1,m})^{l-k}}{(l-k)!} \frac{(r_{1,m}-w)^k}{k!} + mw^\alpha \sum_{k=0}^{(n-1)m-1} \sum_{i=1}^1 C_{n,k,i,m}(\alpha+m) \frac{(r_{1,m}-w)^{k+m}}{k!} {}_{\alpha+m-1}P_m \\
&\quad - mw^\alpha \sum_{k=0}^{(n-1)m-1} \sum_{i=1}^1 \sum_{j=0}^{m-1} \sum_{l=j}^{m-1} C_{n,k,i,m}(\alpha+m) \frac{(r_{1,m}-r_{1,m})^{k+m-l}}{(k+m-l)!} \\
&\quad \times {}_{\alpha+m-1}P_m \frac{(r_{1,m}-r_{1,m})^{l-j}}{(l-j)!} \frac{(r_{1,m}-w)^j}{j!}.
\end{aligned}$$

We use k and j in place of j and k in the last term of the right hand side, respectively,

$$\begin{aligned}
F_{n+1,m}^1(w, \alpha) &= nmw(\alpha-1)^{-1} + m + mw^\alpha \sum_{k=0}^{m-1} \left[\sum_{l=k}^{m-1} \left\{ -\frac{a+l-1 P_l}{r_{1,m}^{\alpha+l}} + \frac{a+l-2 P_{l-1}}{r_{1,m}^{\alpha+l-1}} \right\} \frac{(r_{1,m}-r_{1,m})^{l-k}}{(l-k)!} \right] \\
&\quad - \sum_{j=0}^{(n-1)m-1} \sum_{l=1}^1 C_{n,j,l,m}(\alpha+m) \frac{(r_{1,m}-r_{1,m})^{j+m-l}}{(j+m-l)!} {}_{\alpha+m-1}P_m \Big\} \frac{(r_{1,m}-r_{1,m})^{l-k}}{(l-k)!} \Big] \\
&\quad \cdot \frac{(r_{1,m}-w)^k}{k!} + mw^\alpha \sum_{k=0}^{(n-1)m-1} \sum_{i=1}^1 C_{n,k,i,m}(\alpha+m) \frac{(r_{1,m}-w)^{k+m}}{(k+m)!} {}_{\alpha+m-1}P_m \\
&= nmw(\alpha-1)^{-1} + m + mw^\alpha \sum_{k=0}^{m-1} \left\{ -\frac{a+k-1 P_k}{r_{1,m}^{\alpha+k}} + \frac{a+k-2 P_{k-1}}{r_{1,m}^{\alpha+k-1}} \right\} \frac{(r_{1,m}-w)^k}{k!} \\
&\quad + mw^\alpha \sum_{k=m}^{nm-1} \sum_{i=1}^1 C_{n,k-m,i,m}(\alpha+m) {}_{\alpha+m-1}P_m \frac{(r_{1,m}-w)^k}{k!}.
\end{aligned}$$

Let

$$C_{n+1,k,1,m}(\alpha) = -\frac{\alpha+k-1\mathbf{P}_k}{r_{1,m}^{\alpha+k}} + \frac{\alpha+k-2\mathbf{P}_{k-1}}{r_{1,m}^{\alpha+k-1}}$$

for $0 \leq k \leq m-1$ and

$$C_{n+1,k,1,m}(\alpha) = C_{n,k-m,t,m}(\alpha+m)_{\alpha+m-1}\mathbf{P}_m$$

for $m \leq k \leq nm-1$. Then, the equation

$$F_{n+1,m}^1(w, \alpha) = A_{n+1,m,2}(w, \alpha)$$

holds for $r_{2,m}(\alpha + (n-1)m) \leq w < r_{1,m}(\alpha + nm)$. Suppose that

$$F_{n+1,m}^1(w, \alpha) = A_{n+1,m,s}(w, \alpha)$$

for $r_{s,m}(\alpha + (n-s+1)m) \leq w < r_{s-1,m}(\alpha + (n-s+2)m)$ holds for $2 \leq s \leq n-2$. For $r_{s+1,m}(\alpha + (n-s)m) \leq w < r_{s,m}(\alpha + (n-s+1)m)$, $r_{s,m}$ is used in place of $r_{s,m}(\alpha + (n-s+1)m)$. Then,

$$\begin{aligned} F_{n+1,m}^1(w, \alpha) &= A_{n+1,m,s}(w, \alpha) \\ &+ \int_0^{r_{s,m}-w} \int_0^{r_{s,m}-w-z_1} \dots \int_0^{r_{s,m}-w-z_1-\dots-z_{m-1}} \left[A_{n,m,s+1}\left(w + \sum_{l=1}^m z_l, \alpha+m\right) \right. \\ &\quad \left. - A_{n,m,s}\left(w + \sum_{l=1}^m z_l, \alpha+m\right) \right]_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l\right)^{-\alpha-m} dz_m \dots dz_2 dz_1. \end{aligned}$$

By using Lemmas 3, 5 and 6 and the definition of $A_{n+1,m,s}(w, \alpha)$, it is derived that

$$F_{n+1,m}^1(w, \alpha) = A_{n+1,m,s+1}(w, \alpha).$$

For $0 < w < r_{n,m}(\alpha+m)$,

$$\begin{aligned} F_{n+1,m}^1(w, \alpha) &= A_{n+1,m,n}(w, \alpha) \\ &+ \int_0^{r_{n,m}-w} \int_0^{r_{n,m}-w-z_1} \dots \int_0^{r_{n,m}-w-z_1-\dots-z_{m-1}} \left[nm - A_{n,m,n}\left(w + \sum_{l=1}^m z_l, \alpha+m\right) \right] \\ &\quad \times_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w + \sum_{l=1}^m z_l\right)^{-\alpha-m} dz_m \dots dz_2 dz_1. \end{aligned}$$

By using Lemmas 3, 5 and 6, it is derived that

$$F_{n+1,m}^1(w, \alpha) = A_{n+1,m,n+1}(w, \alpha)$$

which completes the proof since $r_{n+1,m}(\alpha) < r_{n,m}(\alpha+m)$. $\square$

This theorem gives how to calculate the critical value $r_{n,m}(\alpha)$ numerically by high speed computer. The values of $r_{n,m}(\alpha)$ for $n=1, 2, \dots, 7$; $m=1, 2, 3, 4$; $\alpha=2, 3, \dots, 10$ are calculated and tabulated in Table 1.

4. Applications to a Gamma two-armed bandit problem with one arm known

There are two arms A_0 and A_1 and by performing A_0 or A_1 an observation is derived from the gamma distribution with parameter vector $(m, 1)$ or (m, u) , respectively, where m is a given positive integer. The true value of u is unknown, however there is the prior information that u has a gamma distribution with parameter vector (w, α) . Suppose that there are n stages remaining and at each stage we select A_0 or A_1 , perform this we obtain an ob-

Table 1. Values of $r_{n,m}(\alpha)$ for $n=1, 2, \dots, 7$, $m=1, 2, 3, 4$, and $\alpha=2, 3, \dots, 10$

$m \backslash \alpha$	n	1	2	3	4	5	6	7
1	2	1.0000	0.8284	0.7279	0.6585	0.6065	0.5654	0.5319
	3	2.0000	1.7882	1.6523	1.5523	1.4737	1.4093	1.3550
	4	3.0000	2.7700	2.6162	2.4996	2.4057	2.3273	2.2600
	5	4.0000	3.7596	3.5950	3.4679	3.3640	3.2762	3.2000
	6	5.0000	4.7529	4.5810	4.4466	4.3357	4.2411	4.1585
	7	6.0000	5.7482	5.5710	5.4314	5.3152	5.2154	5.1278
	8	7.0000	6.7447	6.5636	6.4199	6.2996	6.1958	6.1042
	9	8.0000	7.7421	7.5579	7.4109	7.2873	7.1802	7.0855
2	10	9.0000	8.7399	8.5533	8.4037	8.2774	8.1677	8.0703
	2	1.0000	0.8038	0.6958	0.6238	0.5711	0.5302	0.4972
	3	2.0000	1.7417	1.5859	1.4758	1.3916	1.3241	1.2680
	4	3.0000	2.7092	2.5255	2.3916	2.2869	2.2012	2.1291
	5	4.0000	3.6889	3.4867	3.3363	3.2168	3.1179	3.0338
	6	5.0000	4.6751	4.4596	4.2970	4.1663	4.0572	3.9637
	7	6.0000	5.6651	5.4395	5.2674	5.1280	5.0107	4.9096
	8	7.0000	6.6574	6.4240	6.2444	6.0978	5.9738	5.8664
3	9	8.0000	7.6514	7.4116	7.2258	7.0733	6.9437	6.8310
	10	9.0000	8.6465	8.4016	8.2106	8.0531	7.9188	7.8015
	2	1.0000	0.7935	0.6829	0.6103	0.5576	0.5170	0.4844
	3	2.0000	1.7190	1.5549	1.4412	1.3553	1.2870	1.2307
	4	3.0000	2.6767	2.4791	2.3380	2.2291	2.1411	2.0676
	5	4.0000	3.6489	3.4278	3.2668	3.1408	3.0376	2.9507
	6	5.0000	4.6292	4.3906	4.2142	4.0747	3.9595	3.8617
	7	6.0000	5.6145	5.3622	5.1736	5.0230	4.8979	4.7911
4	8	7.0000	6.6030	6.3397	6.1411	5.9814	5.8480	5.7335
	9	8.0000	7.5938	7.3215	7.1145	6.9471	6.8065	6.6854
	10	9.0000	8.5863	8.3065	8.0923	7.9182	7.7715	7.6446
	2	1.0000	0.7879	0.6760	0.6031	0.5505	0.5101	0.4777
	3	2.0000	1.7053	1.5368	1.4213	1.3347	1.2662	1.2099
	4	3.0000	2.6560	2.4504	2.3054	2.1946	2.1055	2.0314
	5	4.0000	3.6225	3.3900	3.2229	3.0934	2.9881	2.8999
	6	5.0000	4.5980	4.3448	4.1604	4.0158	3.8974	3.7974
4	7	6.0000	5.5792	5.3097	5.1110	4.9540	4.8246	4.7147
	8	7.0000	6.5644	6.2814	6.0709	5.9034	5.7645	5.6461
	9	8.0000	7.5523	7.2582	7.0375	6.8610	6.7139	6.5880
	10	9.0000	8.5424	8.2387	8.0094	7.8249	7.6706	7.5382

servation as a reward. After obtaining the observation, we select A_0 or A_1 again at the next stage. We repeat this cycle for n times in order to maximize the sum of n observations.

Let $G_{n,m}(w, \alpha)$ be the maximum expected total reward over n stages when the current prior knowledge is (w, α) . The problem can then be formulated by dynamic programming as follows:

$$(4.1) \quad G_{n,m}(w, \alpha) = \max\{G_{n,m}^0(w, \alpha), G_{n,m}^1(w, \alpha)\}$$

for $n=1, 2, \dots$ and $G_{0,m}(w, \alpha)=0$, where

$$G_{n,m}^0(w, \alpha) = m + G_{n-1,m}(w, \alpha)$$

and

$$G_{n,m}^1(w, \alpha) = \int_0^\infty \left[\int_0^\infty \{z + G_{n-1,m}(w+z, \alpha+m)\} (\Gamma(m))^{-1} \alpha^m z^{m-1} e^{-uz} dz \right] \\ \times (\Gamma(\alpha))^{-1} w^\alpha u^{\alpha-1} e^{-wu} du$$

that is,

$$(4.2) \quad G_{n,m}^1(w, \alpha) = mw(\alpha-1)^{-1} \\ + \int_0^\infty G_{n-1,m}(w+z, \alpha+m) \frac{\Gamma(\alpha+m)}{\Gamma(\alpha)\Gamma(m)} w^\alpha z^{m-1} (w+z)^{-\alpha-m} dz$$

Since the gamma distribution version of Theorem 1 also holds in this problem, (4.1) is rewritten as follows:

$$(4.3) \quad G_{n,m}(w, \alpha) = \max\{mn, G_{n,m}^1(w, \alpha)\}$$

It is difficult to solve (4.2) and (4.3) directly. Now, we use the results of Section 3 to solve the case of gamma distribution.

LEMMA 7. For $\alpha > 1$, $m \geq 2$ and $f(\cdot)$,

$$\int_0^\infty f(z) \frac{\Gamma(\alpha+m)}{\Gamma(\alpha)\Gamma(m)} w^\alpha z^{m-1} (w+z)^{-\alpha-m} dz \\ = \int_0^\infty \int_0^\infty \cdots \int_0^\infty f\left(\sum_{i=1}^m y_i\right)_{\alpha+m-1} \mathbf{P}_m w^\alpha \left(w - \sum_{i=1}^m y_i\right)^{-\alpha-m} dy_m dy_{m-1} \cdots dy_1.$$

PROOF. The proof is done by the change of variables, $z = y_1 + y_2 + \cdots + y_m$, $z_2 = y_2, \dots$, and $z_m = y_m$, on the right hand side. $\square$

From this lemma, the following theorem is easily derived.

THEOREM 4. For $m \geq 1$, $n \geq 0$, $w > 0$, and $\alpha > 1$,

$$F_{n,m}(w, \alpha) = G_{n,m}(w, \alpha).$$

PROOF. The proof is done by induction on n by using Lemma 7. $\square$

Therefore, Table 1 is also useful for the gamma two-armed bandit problem with one arm known as discussed in this section.

REMARKS: (i) The problem is also solved in the same way in the case where the parameter vectors $(m, 1)$ and (m, u) are replaced by $(m_0, 1)$ and (m_1, u) .

(ii) The gamma two-armed bandit problem is difficult to solve in the case where m is not integer. In this case, we consider two paradigms of replacing m by $[m]$ and $[m]+1$, where $[m]$ is the largest integer which is not

greater than m . Then, we calculate

$$(1-\theta)r_{n,[m]}(\alpha) + \theta r_{n,[m]+1}(\alpha),$$

where

$$\theta = m - [m]$$

and use it as the critical value. This approximation is useful from a practical point of view.

Acknowledgement. The author appreciates the helpful comments provided by the referee.

REFERENCES

- [1] Bellman, R. (1956). A problem in the sequential design of experiments, *Sankhyā*, **A 16**, 221–229.
- [2] Berry, D. A. and Fristedt, B. (1979). Bernoulli one-armed bandits—Arbitrary discount sequences, *Ann. Statist.*, **7**, 1086–1105.
- [3] Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*, Chapman and Hall, New York.
- [4] Bradt, R. N., Johnson, S. M. and Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.*, **27**, 1060–1074.
- [5] DeGroot, M. H. (1970). *Optimal Statistical Decisions*, McGraw-Hill, New York.
- [6] Gittins, J. C. (1979). Bandit processes and dynamic allocation indices (with discussions), *J. R. Statist. Soc.*, **B 41**, 148–177.
- [7] Gittins, J. C. (1989). *Multi-armed Bandit Allocation Indices*, John Wiley & Sons, New York.
- [8] Gittins, J. C. and Jones, D. M. (1974). A dynamic allocation index for the sequential design of experiments, In *Progress in Statistics* (ed. J. Gani et al), 241–266, North-Holland, Amsterdam.
- [9] Glazebrook, K. D. (1983). Optimal strategy for families of alternative bandit processes, *IEEE Trans. Automatic Control*, **AC-28**, 856–861.
- [10] Hamada, T. (1978). A uniform two-armed bandit problem: the parameter of one distribution is known, *J. Japan Statist. Soc.*, **8**, 29–35.
- [11] Hamada, T. (1987). A two-armed bandit problem with one arm known including switching costs and terminal rewards, *J. Japan Statist. Soc.*, **17**, 21–30.
- [12] Hamada, T. (1992). A discounted uniform one-armed bandit problem, *Sequential Analysis*, **11**, 1–15.
- [13] Kalin, D. and Theodorescu, R. (1990). A uniform two-armed bandit problem with one arm known revisited, *J. Japan Statist. Soc.*, **20**, 159–168.
- [14] Woodroffe, M. (1976). On the one arm bandit problem, *Sankhyā*, **A 38**, 79–91.
- [15] Yakowitz, S. J. (1969). *Mathematics of Adaptive Control Processes*, American Elsevier Publishing Co., New York.