

A TWO-ARMED BANDIT PROBLEM WITH ONE ARM KNOWN AND
WITH BATCH SAMPLING

姫路短期大学 濱田年男 (Toshio Hamada)

1. 緒言

次のような問題を考える. 2種類の実験 e_0, e_1 があるものとする. e_0 を行うとパラメータ u の指数分布に従う確率変数を観察し, また e_1 を行うとパラメータ v の指数分布に従う確率変数を観察するものとする. ここで, v は既知であるが, u の値は未知であり, 事前分布として, パラメータ (w, a) のガンマ分布に従うものとする. 一般性を失うことなく $v=1$ とする. e_0 を m 回行って観察値 $z_1, z_2, \dots, z_m$ を得た後の事後分布はパラメータ $(w+z_1+z_2+\dots+z_m, a+m)$ のガンマ分布に従う (たとえば, DeGroot[1]). n 期間問題を考え, 各期において実験 e_0, e_1 をあわせて m 回行うものとする. すなわち各期に e_0 を k 回, e_1 を $m-k$ 回行うものとする. ここで $0 \leq k \leq m$ であり, k の値は各期において意思決定者が決めることができるものとする. 目的は n 期間の nm 個の観察値の和を最大にすることであり, そのためには各期において e_0, e_1 を何回行えばよいかを決定することである. この問題を $P_A(m)$ で表すことにする.

この問題の特殊な場合として, $k=0$ または $k=m$ のみが許される場合は, 各期にいずれか一方の実験から一度に m 個の観察値を得ることになる. この問題を $P_B(m)$ とする. 特に $m=1$ のとき, この問題 $P_B(1)$ は $P_A(1)$ でもあり, これを $P(1)$ で表すことにすると, これは従来のバンディット問題であり, この種類の問題に対して多くの結果が得られている. たとえば一様分布に対しては Hamada [2], Kalin and Theodorescu [3] 等がある.

2. 動的計画法による定式化

残り n 期間あり, 現在の事前分布のパラメータが (w, a) のとき, 以後最適政策を用いたときの問題 $P_A(m), P(1), P_B(m)$ の最大期待総利得をそれぞれ $V_n(w, a), U_n(w, a), W_n(w, a)$ とする. このとき, つぎのような3種類の再帰方程式が成立する. $P_A(m)$ に対して

$$V_n(w, a) = \max\{V_n^0(w, a), V_n^1(w, a), \dots, V_n^m(w, a)\}$$

$$(n=1, 2, 3, \dots; V_0(w, a)=0)$$

ここに,

$$V_n^0(w, a) = m + V_{n-1}(w, a)$$

であり, また $k=1, 2, \dots, n$ に対して

$$V_n^k(w, a) = E[Z_1 + Z_2 + \dots + Z_k + V_{n-1}(w + Z_1 + Z_2 + \dots + Z_k, a+k) \mid w, a]$$

である. $P(1)$ に対しては,

$$U_n(w, a) = \max\{U_n^0(w, a), U_n^1(w, a)\}$$

$$(n=1, 2, 3, \dots; U_0(w, a)=0)$$

ここに,

$$U_n^0(w, a) = 1 + U_{n-1}(w, a)$$

および

$$U_n^1(w, a) = E[Z_1 + U_n(w + Z_1, a+1) \mid w, a]$$

である. さらに, $P_B(m)$ に対しては,

$$W_n(w, a) = \max\{W_n^0(w, a), W_n^m(w, a)\}$$

$$(n=1, 2, 3, \dots; W_0(w, a)=0)$$

ここに,

$$W_n^0(w, a) = m + W_{n-1}(w, a)$$

および

$$W_n^m(w, a) = E[Z_1 + Z_2 + \dots + Z_m + W_n(w + Z_1 + Z_2 + \dots + Z_m, a+1) \mid w, a],$$

である.

【定理 1】 $n \geq 1, w > 0, a > 1, m \geq 2$ に対して

$$U_{mn}(w, a) \geq V_n(w, a) \geq W_n(w, a) \geq mU_n(w, a)$$

【系 1】 $n \geq 1, w > 0, a > 1, m \geq 2$ に対して

$$V_n^m(w, a) \geq W_n^m(w, a)$$

【定理 2】 $P_A(m), P(1), P_B(m)$ のそれぞれに対して, もしある期において a_0 が最適ならば, 以後の期において a_0 が最適になる.

この定理により, $V_n^0(w, a) = mn, U_n^0(w, a) = n, W_n^0(w, a) = mn$ となる. さらに, 次の 3つの補題が成立する.

【補題1】 $n \geq 1, w > 0, a > 1, m \geq 2, 1 \leq k \leq m$ に対して, $V_n^k(w, a)$ および $W_n^m(w, a)$ は w について連続で狭義単調増加, a について連続で狭義単調連続である.

【補題2】 $n \geq 1, w > 0, a > 1, m \geq 2, 1 \leq k \leq m$ に対して,

$$(i) V_n^k(w, a) \leq (mn - k) + (mn - m + k)w / (a - 1)$$

$$(ii) W_n^m(w, a) \leq (mn - m) + mnw / (a - 1)$$

【補題3】 $n \geq 1, w > 0, a > 1, m \geq 2$ に対して,

$$(i) 1 \leq k \leq m \text{ に対して, } w < k(a - 1) / (mn - m + k) \text{ ならば } V_n^k(w, a) < mn$$

$$(ii) 1 \leq k \leq m \text{ に対して, } w > a - 1 \text{ ならば } V_n^k(w, a) > mn$$

$$(iii) w < (a - 1) / n \text{ ならば } W_n^m(w, a) < mn$$

$$(iv) w > a - 1 \text{ ならば } W_n^m(w, a) > mn$$

補題1, 補題3より, 次の定理が得られる.

【定理3】 $n \geq 1, a > 1, m \geq 2$ に対して

(i) $1 \leq k \leq m$ なる k に対して, m 個の w についての方程式 $V_n^k(w, a) = mn$ は区間 $[k(a - 1) / (mn - m + k), a - 1]$ 内に唯一の根を持つ.

(ii) w についての方程式 $W_n^m(w, a) = mn$ は区間 $[(a - 1) / n, a - 1]$ 内に唯一の根を持つ.

この定理3と系1により, 「 $n \geq 1, w > 0, a > 1, m \geq 2$ に対して, もし $P_B(m)$ において a_0 が最適でなければ, $P_A(m)$ においても a_0 は最適でない」ということがわかる.

3. $P(1)$ の最適解

(i) $n=1$ のとき,

$$U_1(w, a) = \begin{cases} 1 & 0 \leq w \leq s_1(a), \\ w / (a - 1), & s_1(a) \leq w. \end{cases}$$

(ii) $n=2$ のとき,

$$U_2(w, a) = \begin{cases} 2, & 0 \leq w < s_2(a), \\ 1 + w / (a - 1) + B_{2,1}(a)w^a, & s_2(a) \leq w < s_1(a+1), \\ 2w / (a - 1), & s_1(a+1) \leq w \end{cases}$$

ここに

$$B_{2,1}(a) = -\{s_1(a+1)\}^{-a} + (a-1)^{-1}\{s_1(a+1)\}^{-a+1}$$

であり $s_2(a)$ は次の w についての方程式の唯一の根である.

$$B_{2,1}(a)w^a + (a-1)^{-1}w - 1 = 0.$$

(iii) $n \geq 3$ のとき,

$$U_n(w, a) = \begin{cases} n, & 0 \leq w < s_n(a), \\ k + (n-k)w/(a-1) + \sum_{i=1}^{n-k} \sum_{j=1}^k B_{n-i+1, j}(a+i-1) \binom{a+i-2}{i-1} (-1)^{i-1} w^{a+i-1}, & s_{k+1}(a+n-k-1) \leq w < s_k(a+n-k) \quad (k=1, 2, \dots, n-1) \\ nw/(a-1), & s_1(a+n-1) \leq w \end{cases}$$

ここに, 実数 a と非負整数 k に対して,

$$\binom{a}{k} = \begin{cases} a(a-1)(a-2)\cdots(a-k+1)/(1 \cdot 2 \cdot 3 \cdots k), & (k \geq 1) \\ 1, & (k=0) \end{cases}$$

とし, また

$$\begin{aligned} B_{n, k}(a) = & -\{s_k(a+n-k)\}^{-a} + (a-1)^{-1} \{s_k(a+n-k)\}^{-a+1} \\ & - (1 - \delta_{n-1, k}) \sum_{i=1}^{n-k-1} B_{n-i, k}(a+i) \binom{a+i-1}{i} \{-s_k(a+n-k)\}^i \\ & + (1 - \delta_{1, k}) \sum_{j=1}^{k-1} B_{k, j}(a+n-k) \binom{a+n-k-1}{n-k} \{-s_k(a+n-k)\}^{n-k} \end{aligned}$$

であり, $s_n(a)$ は次の w についての方程式の唯一の根である.

$$\sum_{j=1}^{n-1} B_{n, j}(a) w^a + w/(a-1) - 1 = 0$$

4. $P_B(2)$ の最適解

(i) $n=1$ のとき, $t_1(a)=a-1$ とすると,

$$W_1(w, a) = \begin{cases} 2, & 0 < w \leq t_1(a), \\ 2w/(a-1), & t_1(a) \leq w. \end{cases}$$

(ii) $n=2$ のとき,

$$W_2(w, a) = \begin{cases} 4, & 0 < w < t_2(a), \\ 2+2w/(a-1)+C_{2, 1}(a)w^a - D_{2, 1}(a)w^{a+1}, & t_2(a) \leq w < t_1(a+2), \\ 4w/(a-1), & t_1(a+2) \leq w \end{cases}$$

ここに

$$C_{2, 1}(a) = -2(a+1)\{t_1(a+2)\}^{-a} + 2a(a-1)^{-1}\{t_1(a+2)\}^{-a+1}$$

$$D_{2, 1}(a) = -2a\{t_1(a+2)\}^{-a} + 2\{t_1(a+2)\}^{-a+1}$$

であり $t_2(a)$ は次の w についての方程式の唯一の根である.

$$-2 + 2w/(a-1) + C_{2,1}(a)w^a - D_{2,1}(a)w^{a+1} = 0.$$

(iii) $n \geq 3$ のとき,

$$W_n(w, a) = \begin{cases} 2n, & 0 < w < t_n(a), \\ 2k + 2(n-k)w/(a-1) + \sum_{i=1}^{n-k} \sum_{j=1}^k C_{n-i+1,j}(a+2i-2) \binom{a+2i-3}{2i-2} w^{a+2i-2} \\ \quad - \sum_{i=1}^{n-k} \sum_{j=1}^k D_{n-i+1,j}(a+2i-2)(2i-1)^{-1} \binom{a+2i-3}{2i-2} w^{a+2i-1} \\ t_{k+1}(a+2n-2k-2) \leq w < t_k(a+2n-2k) \quad (k=1, 2, \dots, n-1) \\ 2nw/(a-1), & t_1(a+2n-2) \leq w \end{cases}$$

ここに,

$$\begin{aligned} C_{n,k}(a) = & -2(a+1)\{t_k(a+2n-2k)\}^{-a} + 2a(a-1)^{-1}\{t_k(a+2n-2k)\}^{-a+1} \\ & + \sum_{i=1}^{n-k-1} C_{n-i,k}(a+2i) \binom{a+2i-1}{2i} (2i-1) \{t_k(a+2n-2k)\}^{2i} \\ & - \sum_{i=1}^{n-k-1} D_{n-i,k}(a+2i) \binom{a+2i-1}{2i} 2i(2i+1)^{-1} \{t_k(a+2n-2k)\}^{2i+1} \\ & - \sum_{i=1}^{k-1} C_{k,j}(a+2n-2k) \binom{a+2n-2k-1}{2n-2k} \{t_k(a+2n-2k)\}^{2n-2k} \\ & + \sum_{i=1}^{k-1} D_{k,j}(a+2n-2k) \binom{a+2n-2k-1}{2n-2k} (2n-2k)(2n-2k+1)^{-1} \{t_k(a+2n-2k)\}^{2n-2k+1} \end{aligned}$$

および

$$\begin{aligned} D_{n,k}(a) = & -2a\{t_k(a+2n-2k)\}^{-a-1} + 2\{t_k(a+2n-2k)\}^{-a} \\ & + \sum_{i=1}^{n-k-1} C_{n-i,k}(a+2i) \binom{a+2i-1}{2i} (2i) \{t_k(a+2n-2k)\}^{2i-1} \\ & - \sum_{i=1}^{n-k-1} D_{n-i,k}(a+2i) \binom{a+2i-1}{2i} \{t_k(a+2n-2k)\}^{2i} \\ & - \sum_{i=1}^{k-1} C_{k,j}(a+2n-2k) \binom{a+2n-2k-1}{2n-2k} (2n-2k) \{t_k(a+2n-2k)\}^{2n-2k-1} \\ & + \sum_{i=1}^{k-1} D_{k,j}(a+2n-2k) \binom{a+2n-2k-1}{2n-2k} \{t_k(a+2n-2k)\}^{2n-2k} \end{aligned}$$

であり, $t_n(a)$ は次の w についての方程式の唯一の根である.

$$-2 + 2w/(a-1) + \sum_{j=1}^{n-1} C_{n,j}(a)w^j - \sum_{j=1}^{n-1} D_{n,j}(a)w^{j+1} = 0.$$

参 考 文 献

- [1] DeGroot, M. H. (1970). Optimal Statistical Decisions. McGraw-Hill, New York.
- [2] Hamada, T. (1978). A uniform two-armed bandit problem: the parameter of one distribution is known. J. Japan Statist. Soc. Vol.8, 29-35.
- [3] Kalin, D. and Theodorescu, R. (1990). A uniform two-armed bandit problem with one arm known revisited. J. Japan Statist. Soc. Vol.20, pp.159-168.