数理解析研究所講究録
第2078巻 2018年 187-193



# ベイズ最適化問題について

千葉大学大学院 教育学研究科 藤島 昂太
千葉大学教育学部 中井 達

## 1 はじめに

データを母集団から取り出し、逐次的に求めた事後分布に基づき母数を修正し、最適戦略を求める問題全般をベイズ最適化問題とよぶ。ここではベイズ最適化問題の中でも、特に TAB(Two-armed-bandit problem) について述べる。2 つの試行 $e_1, e_2$ はそれぞれベルヌーイ分布に従う確率変数 $X$,$Y$ を観測し、その成功確率がそれぞれ $p$,$q$ であるとする。事前情報 $I$ が与えられ、残り観測数が $n$ 個であるとき、次に試行 $e_a(a=1,2)$ をとり、残り $n-1$ 回の観測を最適戦略に従ったときの総期待利得を $V_n^a(I)$ と表す。問題の目的は、$V_n^a(I)$ を最大化するための行動の選び方、すなわち $\max_a V_n^a(I) = V_n(I)$ を得るような法則 $\pi^*$ を見つけることである。$e_1(e_2)$ からの観測で得られる事後情報を $I'(I'')$ とおくと、

$$V_n(I) = \max\{V_n^1(I), V_n^2(I)\}$$

$$V_n^1(I) = E[p + V_{n-1}(I')], V_n^2(I) = E[q + V_{n-1}(I'')]$$

となり、すべての $n$ における最適な戦略を再帰的に求めることができる。

本研究では、TAB を事前に与えられた情報の形態によって 2 種類に分け、それぞれに対する最適戦略を考えた。1 つ目は、パラメータについて 2 つの仮説が与えられており、仮説が正しい確率が事前情報として与えられている場合である。古くから Bradt,Jhonson,Karlin [1] によって提案されている TAB であり、その最適戦略については、Feldman [2] および Kelley [5] が差の関数を用いることで示している。ここでは、残り観測数に応じた最適な行動を決めるための点について、具体的に求めた。2 つ目は、確率 $p$,$q$ の分布が事前情報として与えられている場合である。この問題を考える際には、ベルヌーイ分布に対し自然な共役分布であるベータ分布を用いるとよい。片方のパラメータが既知である場合に対して、$n \leq 4$ における最適戦略を示した。また、最適戦略を決めるための関数 $r_n(a,b)$ は、未知パラメータが持つ母数 $a,b$ および残り観測数 $n$ について単調性を持つことを明らかにした。

## 2 パラメータについて 2 つの仮説がある TAB

ベルヌーイ分布にしたがう確率変数 $X$,$Y$ が下記のような仮説を持つとする。このとき $\xi$ は仮説 $H_1$ が正しいという事前確率である。

| | | $X$ | $Y$ |
|---|---|---|---|
| $\xi$ | $H_1$ | $c$ | $d$ |
| $1-\xi$ | $H_2$ | $a$ | $b$ |

ここで、$a<b$ かつ $c>d$ と仮定する。$e_1(e_2)$ から観測して得られる事後確率を $\xi_1(\xi_2)$ としたとき、

$$\begin{aligned} V_0(\xi) &= 0 \\ V_n(\xi) &= \max\{E[X + V_{n-1}(\xi_1)], E[Y + V_{n-1}(\xi_2)]\} \end{aligned} \tag{1}$$

となる。ここで、差の関数

$$d_n(\xi) = V_n^1(\xi) - V_n^2(\xi) \tag{2}$$

を定義する。このとき、補題1

**補題 1.** $\pi_n^{1,2}$ を、はじめ2回の観測を $e_1, e_2$ の順で行い、残りの $n-2$ 回の観測を最適戦略に従って行うという戦略とし、$\pi_n^{2,1}$ を、はじめ2回の観測を $e_2, e_1$ の順で行い、残りを最適戦略に従って観測する戦略とする。この時、すべての $n, \xi$ について

$$V_n^{1,2}(\xi) = V_n^{2,1}(\xi)$$

を利用することで、$d_n$ と $d_{n-1}$ の関係

$$\begin{aligned} d_n(\xi) &= E(X) + E[V_{n-1}(\xi_1)] - E(Y) - E[V_{n-1}(\xi_2)] \\ &= d_1(\xi) + E[E(Y|X) + E(V_{n-2}(\xi_{1,2})|X)] + E[d_{n-1}(\xi_1)^+] \\ &\quad -E[E(X|Y) + E(V_{n-2}(\xi_{2,1})|Y)] + E[d_{n-1}(\xi_2)^-] \\ &= E[d_{n-1}(\xi_1)^+] + E[d_{n-1}(\xi_2)^-] \end{aligned} \tag{3}$$

を得る。ただし $x^+ = \max\{0, x\}$,$x^- = \min\{0, x\}$ である。

**定理 1.** $n = 1, 2, ...$ に対して、以下が成り立つ。

**a** $d_n(\xi)$ は $\xi$ についての増加関数である。

**b** $d_n(\xi)$ は $\xi$ について連続な関数である。

**c** $d_n(0) < 0$ かつ $d_n(1) > 0$ である。

**d** $d_n(\alpha_n) = 0$ となるような固有の点 $\alpha \in (0, 1)$ が存在する。

証明は $n$ についての帰納法により示せる。この定理から、最適戦略 $\pi^*$ は、固有の点列 $\alpha_1, \alpha_2, ...\alpha_n$ を定めたとき、$i$ 回目の観測での事後確率 $\xi_i$ が $\xi_i \geq \alpha_i$ なら次の観測は $e_1$ から、そうでない場合は次の観測を $e_2$ から行うことである。

例：$a = 0.3, b = 0.8, c = 0.7, d = 0.2$ のとき

$n \leq 3$ における $\alpha_n$ を求めていく。

$$P(X=1) = 0.4\xi + 0.3, \quad \xi(X=1) = \frac{0.7\xi}{0.7\xi + 0.3(1-\xi)}$$

$$P(X=0) = 0.7 - 0.4\xi, \quad \xi(X=0) = \frac{0.3\xi}{0.3\xi + 0.7(1-\xi)}$$

$$P(Y=1) = 0.8 - 0.6\xi, \quad \xi(Y=1) = \frac{0.2\xi}{0.2\xi + 0.8(1-\xi)}$$

$$P(Y=0) = 0.6\xi + 0.2, \quad \xi(Y=0) = \frac{0.8\xi}{0.8\xi + 0.2(1-\xi)}$$

$$d_1(\xi) = 0.7\xi + 0.3(1-\xi) - 0.2\xi - 0.8(1-\xi) = \xi - 0.5$$

したがって、$\alpha_1 = 1/2$。(4) および (3) を利用し、

$$\begin{aligned}
d_1(\xi(X=1)) &= \frac{5\xi - 1.5}{4\xi + 3}, \quad d_1(\xi(X=1)) > 0 \Leftrightarrow \xi > \frac{3}{10} \\
d_1(\xi(X=0)) &= \frac{5\xi - 3.5}{7 - 4\xi} \quad d_1(\xi(X=0)) > 0 \Leftrightarrow \xi > \frac{7}{10} \\
d_1(\xi(Y=1)) &= \frac{2.5\xi - 2}{4 - 3\xi} \quad d_1(\xi(Y=1)) > 0 \Leftrightarrow \xi > \frac{4}{5} \\
d_1(\xi(Y=0)) &= \frac{2.5\xi - 0.5}{3\xi + 1} \quad d_1(\xi(Y=0)) > 0 \Leftrightarrow \xi > \frac{1}{5}
\end{aligned}$$

$$d_2(\xi) = \begin{cases} \xi - 0.5 & (0 \le \xi < \frac{1}{5}) \\ 0.5\xi - 0.4 & (\frac{1}{5} \le \xi < \frac{3}{10}) \\ \xi - 0.55 & (\frac{3}{10} \le \xi < \frac{7}{10}) \\ 1.5\xi - 0.9 & (\frac{7}{10} \le \xi < \frac{4}{5}) \\ \xi - 0.5 & (\frac{4}{5} \le \xi < 1) \end{cases}$$

したがって、$\alpha_2 = 55/100$。同様に、$\alpha_3 = 59/103$ となる。

**近視眼的戦略 (myopic strategy)**

$X$,$Y$ の分布がそれぞれ $F_i$,$G_i$ で与えられ、最適戦略 $\pi^*$ にしたがって得られる総期待利得を $V_n(\xi, \pi^*)$ としたとき、最初の観測を $e_1$ から行い、残りの観測を最適戦略に従って行ったときの総期待利得は

$$\begin{aligned}
V_n^1(\xi, \pi^*) &= \xi \int_{-\infty}^{\infty} t f_1(t) d\psi + (1-\xi) \int_{-\infty}^{\infty} t f_2(t) d\psi \\
&\quad + \int_{\infty}^{-\infty} V_{n-1}(\frac{\xi f_1(t)}{\xi f_1(t) + (1-\xi) f_2(t)}, \pi^*) \times [\xi f_1(t) + (1-\xi) f_2(t)] d\phi \qquad (4)
\end{aligned}$$

$e_2$ から始めた場合は

$$\begin{aligned}
V_n^2(\xi, \pi^*) &= \xi \int_{-\infty}^{\infty} t g_1(t) d\psi + (1-\xi) \int_{-\infty}^{\infty} t g_2(t) d\psi \\
&\quad + \int_{\infty}^{-\infty} V_{n-1}(\frac{\xi g_1(t)}{\xi g_1(t) + (1-\xi) g_2(t)}, \pi^*) \times [\xi g_1(t) + (1-\xi) g_2(t)] d\phi \qquad (5)
\end{aligned}$$

となる。このとき、近視眼的戦略 $\pi_1$ は、$j$ 回目の観測を終えた後の事後確率 $\xi_j$ が与えられときの、次における期待観測値

$$\int_{-\infty}^{\infty} t[\xi_j f_1(t) + (1-\xi_j) f_2(t)], \quad \int_{-\infty}^{\infty} t[\xi_j g_1(t) + (1-\xi_j) g_2(t)]$$

のうち、大きいほうを段階ごとに選んでいくという戦略である。ベルヌーイ分布をもつ TAB について、$\pi_1$ が最適となる条件は以下のようになる。

**定理 2.** パラメータの組が 2 点 $(a,b)$,$(c,d)$ にのみ分布していると仮定する。このとき、近視眼戦略は次の 4 つの条件のうち 1 つが満たされている時に限り、最適となる。

**(a)** $a \leq b$ かつ $c \leq d$。または $a \geq b$ かつ $c \geq d$。

**(b)** $a + b = c + d = 1$

**(c)** $(d, c) = (a, b)$

## 3 片方のパラメータが既知の場合の TAB

片方の成功確率 $q$ が既に知られているとする。もう片方の確率 $p$ はパラメータ $(a, b)$ のベータ分布に従っていると仮定する。このとき

$$\begin{aligned} V_n(a,b) &= \max\{V_n^1(a,b), V_n^2(a,b)\} \\ V_n^1(a,b) &= \frac{a}{a+b}(1 + V_{n-1}(a+1,b)) + \frac{b}{a+b}V_{n-1}(a,b+1) \\ V_n^2(a,b) &= q + V_{n-1}(a,b) \end{aligned}$$

が導ける。

**定理 3.** $n \geq 1, a > 0, b > 0$ に対し

**(a)** $V_n(a,b)$ は $a$ についての非減少関数である。

**(b)** $V_n(a,b)$ は $b$ についての非増加関数である。

ここで、

$$d_n(a,b) = \frac{a}{a+b}d_{n-1}^+(a+1,b) + \frac{b}{a+b}d_{n-1}^+(a,b+1) + d_{n-1}^-(a,b) \tag{6}$$

とすることで、最適な決定を行うための判断の基準となる関数をより具体的に求めることができる。$d_n(a,b)$ については次の定理が成り立つ。

**定理 4.** $n \geq 1, a > 0, b > 0$ に対し

**(a)** $d_n(a,b)$ は $n$ に関する非減少関数である。

**(b)** $d_n(a,b)$ は $a$ に関する増加関数である。

**(c)** $d_n(a,b)$ は $b$ に関する減少関数である。

この定理から、$d_n(r_n(a,b)) = 0$ となる関数 $r_n(a,b)$ が存在することがわかる。したがって、$r_n(a,b) \geq \dfrac{1-q}{q}$ ならば最適な選択は $e_1$、それ以外は $e_2$ である。

**数値計算**

ここでは、$n$ を具体的に定め、最適戦略を決めるための $r_n(a,b)$ を求める。$n = 1$ のときは

$$d_1(a,b) = \frac{a}{a+b} - q \tag{7}$$

より、$r_1(a,b) = \dfrac{b}{a}$ である。

$n=2$ のときは

$$d_2(a,b) = \frac{a}{a+b}\left(\frac{a+1}{a+b+1}-q\right)^+ + \frac{b}{a+b}\left(\frac{a}{a+b+1}-q\right)^+ + \left(\frac{a}{a+b}-q\right)^- \tag{8}$$

各項について

$$\begin{aligned}
\frac{a+1}{a+b+1} > q &\implies \frac{b}{a+1} < \frac{1-q}{q} \\
\frac{a}{a+b+1} > q &\implies \frac{b+1}{a} < \frac{1-q}{q} \\
\frac{a}{a+b} > q &\implies \frac{b}{a} < \frac{1-q}{q}
\end{aligned}$$

となるため、以下の場合分けに従って、最適戦略を考える。

**(i)** $\frac{b}{a+1} > \frac{1-q}{q}$ のとき

$$d_2(a,b) = \frac{a}{a+b} - q \tag{9}$$

$d_2(a,b) > 0 \implies \frac{b}{a} < \frac{1-q}{q}$ であるが、条件より、常に $d_2(a,b) < 0$ である。

**(ii)** $\frac{b}{a+1} < \frac{1-q}{q} < \frac{b}{a}$ のとき

$$d_2(a,b) = \frac{a(2a+b+2)}{(a+b)(a+b+1)} - \frac{2a+b}{a+b}q \tag{10}$$

$$d_2(a,b) > 0 \implies q < \frac{a(2a+b+1)}{(2a+b)(a+b+1)} \tag{11}$$

より、最適な選択は

$$\begin{cases} \dfrac{1-q}{q} > \dfrac{b(2a+b+1)}{a(2a+b+2)} \implies e_1 \\ otherwise \implies e_2 \end{cases} \tag{12}$$

**(iii)** $\frac{b}{a} < \frac{1-q}{q} < \frac{b+1}{a}$ のとき

$$d_2(a,b) = \frac{a}{a+b}\left(\frac{a+1}{a+b+1}-q\right) \tag{13}$$

従って、$d_2(a,b) > 0 \implies \frac{b}{a+1} < \frac{1-q}{q}$ であるが、条件より、常に $d_2(a,b) > 0$ である。

**(iv)** $\frac{b+1}{a} < \frac{1-q}{q}$ のとき

$$d_2(a,b) = \frac{a}{a+b} - q \tag{14}$$

従って、$d_2(a,b) > 0 \implies \frac{b}{a} < \frac{1-q}{q}$ であるが、条件より、常に $d_2(a,b) > 0$ である。

これにより、

$$r_2(a,b)=\frac{b(2a+b+1)}{a(2a+b+2)} \tag{15}$$

となり、$n=2$ における最適戦略は

$$\begin{cases} \dfrac{1-q}{q}>\dfrac{b(2a+b+1)}{a(2a+b+2)} \Longrightarrow e_1\text{から観測する} \\ \dfrac{1-q}{q}<\dfrac{b(2a+b+1)}{a(2a+b+2)} \Longrightarrow e_2\text{から観測する} \end{cases} \tag{16}$$

となる。

次に、$n=3$ のときは、(6) より

$$d_3(a,b)=\frac{a}{a+b}d_2^+(a+1,b)+\frac{b}{a+b}d_2^+(a,b+1)+d_2^-(a,b) \tag{17}$$

また、$n=2$ の時に行った場合分けの結果を用いることで

$$d_2(a+1)=\begin{cases} \dfrac{a+1}{a+b+1}-q & (\dfrac{1-q}{q}<\dfrac{b}{a+2}) \\ \dfrac{(a+1)(2a+b+4)}{(a+b+1)(a+b+2)}-\dfrac{2a+b+2}{a+b+1}q & (\dfrac{b}{a+2}<\dfrac{1-q}{q}<\dfrac{b}{a+1}) \\ \dfrac{(a+1)(a+2)}{(a+b+1)(a+b+2)}-\dfrac{a+1}{a+b+1}q & (\dfrac{b}{a+1}<\dfrac{1-q}{q}<\dfrac{b+1}{a+1}) \\ \dfrac{a+1}{a+b+1}-q & (\dfrac{b+1}{a+1}<\dfrac{1-q}{q}) \end{cases} \tag{18}$$

$$r_2(a+1,b)=\frac{b(2a+b+3)}{(a+1)(2a+b+4)} \tag{19}$$

を得る。$r_2(a+1,b)<\dfrac{1-q}{q}<r_2(a,b)$ のとき、

$$\begin{aligned} d_3(a,b) &= \frac{a(a+1)(a+2)}{(a+b)(a+b+1)(a+b+2)}-\frac{a(a+1)}{(a+b)(a+b+1)}q \\ &\quad +\frac{a(2a+b+2)}{(a+b)(a+b+1)}-\frac{2a+b}{a+b}q \end{aligned} \tag{20}$$

これを変形すると

$$d_3(a.b)>0\Longrightarrow\frac{1-q}{q}>\frac{b(3a^2+3ab+b^2+6a+3b+2)}{a(3a^2+3ab+b^2+9a+4b+6)} \tag{21}$$

従って、

$$r_3(a,b)=\frac{b(3a^2+3ab+b^2+6a+3b+2)}{a(3a^2+3ab+b^2+9a+4b+6)} \tag{22}$$

また、$r_2(a,b)>r_3(a,b)$ である。同様にして、$r_4(a,b)$ を求めると

$$r_4(a,b)=\frac{b(4a^3+b^3+6a^2b+4ab^2+18a^2+6b^2+22a+18ab+11b+6)}{a(4a^3+b^3+6a^2b+4ab^2+24a^2+7b^2+44a+22ab+18b+24)} \tag{23}$$

となり、$r_3(a,b)>r_4(a,b)$ である。帰納的に、次の定理が得られる。

**定理 5.** $a>0,b>0$ について

**(a)** $r_n(a,b)$ は $n$ に関する非増加関数である。

**(b)** $r_n(a,b)$ は $a$ に関する減少関数である。

**(c)** $r_n(a,b)$ は $b$ に関する増加関数である。

## 4 今後の課題

事前分布が情報として与えられているTABについては、パラメータが両方とも未知である場合の最適戦略を具体的に求めることが課題として挙げられる。このTABは、それぞれの確率変数がベータ分布に従うため、計4つのパラメータを有する。したがって、その大小関係による適切な場合分けを行った上で、最適戦略を考える必要がある。そのまま場合分けすることは容易ではないため、パラメータの組 $(a,b;c,d)$ に関し、例えばパラメータの比 $a/b$,$c/d$ の関係を用いることが有効であると考える。ここで述べた、片方のパラメータが既知の場合のTABにおける最適戦略を拡張させるという考え方も重要であると考える。また、今回は尤度関数としてベルヌーイ分布を仮定し、事前分布としてベータ分布を仮定したが、尤度関数を別のものに変えることで新たなTABを考えることができる。中でも、二項分布、幾何分布はその密度関数がベルヌーイ分布と共通する部分が多いため、最適戦略を求めることは難しくないだろう。これらのTABの最適戦略、また、具体的にどういった場面においてこのTABを活用できるかということも併せて、今後研究する必要があると考える。

## 参考文献


[1] R. N. Bradt, S. M. Johnson, S. Karlin(1956). On sequential designs for maximizing the sum of n observations, *Ann. Math. Statist*, Vol.27, p.1060-1074.

[2] D. Feldman(1962). Contributions on the Two-armed-bandit-problem, *Ann. Math. Statist*, Vol.33, p.847-856.

[3] T. Hamada(1978). A uniform two-armed-bandit problem:The parameter of one distribution is known, *J. Japan Statist. Soc*, Vol.8, No.1, p.29-35.

[4] T. Hamada(1984). On a uniform two-armed-bandit problem, *J. Japan Statist. Soc*, Vol.14, No.2, p.179-187.

[5] T.A. Kelley(1974). A note on the Bernoulli Two-armed-bandit problem, *Ann. Statist*, Vol.2, No.5, p.1056-1062.

[6] M. Kolonko, H. Benzing(1983). The Sequential design of Bernoulli experiments including switching cost,*Operations Research*, Vol.33, No. 2, p.412-426.

[7] M. Kolonko, H. Benzing(1985). On monotone optimal decision rules and the Stay-an-a-winner rule for the Two-armed bandit, *Metrika*, Vol.32, p.395-407.

[8] H. Robbins(1956). A sequential decision problem with a finite memory, *Proc.Nat.Acad.Sci*, Vol.42, p.920-923.