

ダイナミック・プログラミングとベイズ 適応制御系

阪大・基礎工 坂口 実

§1. 適応制御系

線型制御系

$$(1.1) \quad x_n = \alpha x_{n-1} + v_n + Y_n \quad (n=1, 2, \dots; x_0=c)$$

を考える。ここに $\alpha \neq 0$ は与えられた定数, $\{Y_n\}$ は, 独立確率変数列である, α , 同一分布 $N(\theta, \frac{1}{2})$ に従い, 平均値 θ の値は未知である。制御は

$$(1.2) \quad v_n = v_n(x_{n-1}, r^{n-1}), \quad (n=2, 3, \dots)$$

すなわち $r^{n-1} = \{r_1, \dots, r_{n-1}\}$ を情報として利用し, θ を推定しながら制御する。

θ が未知だからベイズ的接近をとる。すなわち θ の事前分布を $N(\mu, \sigma^2)$ とし, 情報 r^n を得た後の θ の事後分布は, やはり $N(\mu_n, \sigma_n^2)$, にだし

$$(1.3) \quad \mu_n = t_n \bar{Y}_n + (1-t_n)\mu, \quad \sigma_n^2 = \frac{1}{2n+\sigma^{-2}} \quad (n \geq 1)$$

$$(1.4) \quad t_n = 2n(2n + \sigma^{-2})^{-1}, \quad \bar{y}_n = n^{-1} \sum_{j=1}^n y_j$$

で与えられる。 $n=1$ に対して y_1 をえらぶときには θ に関する情報は何も無いわけであるが、仮定

$$(A.1) \quad \left\{ \begin{array}{l} \text{はじめに pilot observation } y^0 \sim N(\theta, \frac{1}{2}) \text{ を} \\ \text{とることが許されている, 第一段では } y_1 \text{ の分布} \\ \text{があたかも } N(y^0, \frac{1}{2}) \text{ として既知のときのように} \\ \text{ふるまうものとする。第二段以後の制御は(1.} \\ \text{2) であって } y^{n-1} \text{ には } y^0 \text{ を含ませない。} \end{array} \right.$$

を設けておく。

$V = \{v_n\}_{n=1}^N$ を(条件つき)制御政策という。制御の目的は

$$(1.5) \quad \int E_0 \left[\sum_{n=1}^N (\lambda v_n^2 + a_n x_n^2) \mid y^0, V \right] \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta \rightarrow \min_V$$

であるとする。ここに入, a_1, \dots, a_N は与えられた非負定数であって, E_0 は y^0 を観察した条件のもとで, y^N についてとった期待値を表わす。

この問題に対し, (a) ベイズ制御政策, (b) $\lambda=0$ のときの minimax 制御政策を求め, (c) $\{x_n\}_{n=1}^N$ が直接観察できなくて $y_n = x_n + \eta_n$, ただし $\eta_n \sim N(0, b^2)$ (b^2 は既知) は独立な random noise, が観察可能である場合について

も考える。

§2. ベイズ制御政策

ダイナミック・プログラミングの常とう的手段により

$f_{k,N}(c, Y^{k-1})$ ----- 時刻 k における情報が $X_{k-1} = c$
 および Y^{k-1} のとき、以後ベイズ制御政策
 ($N(\mu_{k-1}, \sigma_{k-1}^2)$ についての) を用いるとき
 得られるべき期待値

とおくと、明らかに $k \geq 2$ に対し

$$f_{k,N}(c, Y^{k-1}) = \min_{v_k} E^{(k)} \left[\lambda v_k^2 + a_k (c + v_k + Y_k)^2 + f_{k+1,N}(c + v_k + Y_k, Y^k) \right];$$

ここに

$$E^{(k)}[\phi(Y_k)] \equiv \int \frac{1}{\sqrt{2\pi}\sigma_{k-1}} e^{-\frac{(0-\mu_{k-1})^2}{2\sigma_{k-1}^2}} d\theta \int \phi(Y_k) \frac{1}{\sqrt{\pi}} e^{-(Y_k-0)^2} dY_k$$

$$= \int \phi(Y_k) dY_k \cdot \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{k-1}^2 + \frac{1}{2}}} e^{-\frac{(Y_k - \mu_{k-1})^2}{2(\sigma_{k-1}^2 + \frac{1}{2})}}$$

を表わす。(A.1) により、もしも

$$(2.2) \quad \mu_0 = Y^0, \quad \sigma_0^2 = 0$$

と定義しておけば、(2.1) はすべての $k \geq 1$ に対して成
 立するわけである。

よって (2.1) を $k=N$ のときに解くと ($f_{N+1,N}(\cdot) \equiv 0$),

$$(2.3) \quad v_{N,N}^*(x_{N-1}, y^{N-1}) = -\frac{a_N}{\lambda + a_N} (dx_{N-1} + E^{(N)}(y_N))$$

$$(2.4) \quad f_{N,N}(x_{N-1}, y^{N-1}) = \frac{\lambda a_N}{\lambda + a_N} (dx_{N-1} + E^{(N)}(y_N))^2 + a_N V^{(N)}(y_N).$$

いま一般に

$$(2.5) \quad f_{k,N}(c, y^{k-1}) = A_k c^2 - 2B_k(y^{k-1})c + C_k(y^{k-1}),$$

($k=1, \dots, N$) とおくと, (2.1) よりつぎの諸式が得られる (もしも分母が0でない限り).

$$(2.6) \quad v_{k,N}^*(c, y^{k-1}) = \frac{-(dc + E^{(k)}(y))(a_k + A_{k-1}) + E^{(k)}\{B_{k+1}(y^k)\}}{\lambda + a_k + A_{k+1}}$$

$$(k=1, \dots, N; E^{(1)}(y) = \mu_0 = y^0)$$

$$(2.7) \quad A_k = \frac{d^2 \lambda (a_k + A_{k+1})}{\lambda + a_k + A_{k+1}}$$

$$(2.8) \quad B_k(y^{k-1}) = \frac{d\lambda}{\lambda + a_k + A_{k+1}} \cdot [-(a_k + A_{k+1})E^{(k)}(y) + E^{(k)}(y) + E^{(k)}\{B_{k+1}(y^k)\}]$$

$$C_k(y^{k-1}) = E^{(k)}\{C_{k+1}(y^k)\} + (a_k + A_{k+1})E^{(k)}(y^2) - 2E^{(k)}\{B_{k+1}(y^k) \cdot y_k\} - \frac{[(a_k + A_{k+1})E^{(k)}(y) - E^{(k)}\{B_{k+1}(y^k)\}]^2}{\lambda + a_k + A_{k+1}}.$$

(2.4), (2.5) から

$$(2.7d) \quad A_N = \frac{d^2 \lambda a_N}{\lambda + a_N}$$

$$(2.8d) \quad B_N(Y^{N-1}) = -\frac{\alpha \lambda a_N}{\lambda + a_N} E^{(N)}(Y),$$

$$(2.9d) \quad C_N(Y^{N-1}) = \frac{\lambda a_N}{\lambda + a_N} (E^{(N)}(Y))^2 + a_N V^{(N)}(Y)$$

であるから、数列 $\{A_k\}_1^N$, $\{B^{(k)}(Y^{k-1})\}_1^N$, $\{C_k(Y^{k-1})\}_1^N$ を定めることができる。かく

[定理 1] 想定 (A.1) のもとで, (1.1) に対する Bayes 適応制御政策 $(N|\mu, \sigma^2)$ についての (2.6)~(2.9) が与えられる。

これはもちろん non-adaptive case (Sakaguchi, 1962) を特別な場合として含む。

(注意 1) $\lambda = 0$, $a_1, \dots, a_N > 0$ のときは Bayes 政策は one-step optimal である。何となれば, このとき (2.7)~(2.9d) より

$$(2.10) \quad \begin{cases} A_k \equiv 0, & B_k(Y^{k-1}) \equiv 0 & (k=1, \dots, N) \\ C_k(Y^{k-1}) = \sum_{i=k}^N a_i V^{(i)}(Y_i) = \sum_{i=k}^N a_i (\sigma_{i-1}^2 + \frac{1}{2}) \end{cases}$$

となり, (2.6) より

$$(2.11) \quad v_{k,N}^*(x_{k-1}, Y^{k-1}) = -(\alpha x_{k-1} + E^{(k)}(Y_k))$$

となつて, N に無関係である。

さらにこのとき

$$(2.12) \quad f_{1N}(x_0, Y^0) = A_1 x_0^2 - 2B_1(Y^0)x_0 + C_1(Y^0)$$

$$\begin{aligned}
 &= \frac{a_1}{2} + \sum_{i=2}^N a_i (\sigma_{i-1}^2 + \frac{1}{2}) \\
 &= \frac{a_1}{2} + \sum_{i=2}^N a_i \left(\frac{1}{2} + \frac{1}{2(i-1) + \sigma^2} \right)
 \end{aligned}$$

となつて, x_0, Y^0 および事前分布における μ に無関係であり, σ^2 にのみ依存するのである。

(注意2). $\lambda > 0$ のときの terminal control:

$$(2.13) \quad a_1 = a_2 = \dots = a_{N-1} = 0, \quad a_N = 1$$

を考へよう。このとき (2.7), (2.7d), (2.8), (2.8d) より)

$$(2.14) \quad \begin{cases} A_k = \frac{\alpha^{2(N-k+1)} \lambda}{\lambda + 1 + \alpha^2 + \dots + \alpha^{2(N-k)}} \\ B_k(Y^{k-1}) = \frac{-\alpha^{N-k+1} \lambda E^{(k)}(Y_k)}{\lambda + 1 + \alpha^2 + \dots + \alpha^{2(N-k)}} \cdot \sum_{i=k}^N \alpha^{N-i} \end{cases}$$

さらに (2.6) より

$$(2.15) \quad v_{k,N}^*(c, Y^{k-1}) = \frac{-\alpha^{N-k}}{\lambda + \sum_{i=k}^N \alpha^{2(N-k)}} \left(E^{(k)}(Y_k) \sum_{i=k}^N \alpha^{N-i} + \alpha^{N-k+1} c \right)$$

がでる。何となれば, つぎの補題があるからである:

(補題1). $1 \leq k \leq i \leq N$ に対し

$$E^{(k)} E^{(k+1)} \dots E^{(i-1)} E^{(i)}(Y_i) = E^{(k)}(Y_k) = \mu_{k-1}$$

証明: $E^{(k)} E^{(k+1)}(Y_{k+1}) = \mu_{k-1}$

を示せばよい。(1.3) より

$$\begin{aligned}
E^{(k)} E^{(k+1)}(Y_{k+1}) &= E^{(k)}(\mu_k) = E^{(k)}\{t_k \bar{Y}_k + (1-t_k)\mu\} \\
&= \frac{t_k}{k}((k-1)\bar{Y}_{k-1} + E^{(k)}(Y_k)) + (1-t_k)\mu \\
&= \frac{t_k}{k}(k-1+t_{k-1})\bar{Y}_{k-1} + \left(\frac{t_k(1-t_{k-1})}{k} + 1-t_k\right)\mu,
\end{aligned}$$

ところが (1.4) より

$$(2.16) \quad t_k(k-1+t_{k-1})/k = t_{k-1}$$

であるから証明された。▣

(2.15) は N に依存するから terminal control の Bayes 政策は one-step optimal でない。しかも (2.11) と (2.15) とを比較すればわかるように、制御問題 (1.5) の Bayes 政策は λ につき ($\lambda=0$ において) 連続でない。

§3. $\lambda=0$ のときの minimax 制御政策

前節注意1で述べたように、 $\lambda=0$ のときの Bayes 政策は (2.11) で与えられるが、いま

$$(3.1) \quad \begin{cases} \hat{v}_{1,N}(c, Y^0) = -(dc + Y^0), \\ \hat{v}_{k,N}(c, Y^{k-1}) = -(dc + \bar{Y}_{k-1}), \quad 2 \leq k \leq N \end{cases}$$

で定義される政策を

$$\hat{V} = \left\{ \hat{v}_{k,N}(c, Y^{k-1}) \right\}_{k=1}^N$$

とおく。つぎに想定 (A.1) のもとで *minimax* 制御問題:

$$(3.2) \quad \sup_{\theta} E_{\theta} \left[\sum_{n=1}^N a_n x_n^2 \mid \gamma^{\circ}, \nabla \right] \rightarrow \min_{\nabla}$$

に対し、 $\hat{\nabla}$ が *minimax* なことを示そう。

(1.1) に (3.1) を代入すると $\hat{\nabla}$ による軌道は

$$\begin{aligned} x_n &= \alpha x_{n-1} + \hat{v}_n + \gamma_n \quad (n=1, 2, \dots; x_0 = c) \\ &= \alpha x_{n-1} + \gamma_n - (\alpha x_{n-1} + \bar{\gamma}_{n-1}) \\ &= \gamma_n - \bar{\gamma}_{n-1} \quad (\text{ただし } \bar{\gamma}_0 = \gamma^{\circ} \text{ と定義する}) \end{aligned}$$

したがって θ に無関係に

$$(3.3) \quad E_{\theta} \left[\sum_{n=1}^N a_n x_n^2 \mid \gamma^{\circ}, \hat{\nabla} \right] = \frac{a_1}{2} + \sum_{n=2}^N a_n \left(\frac{1}{2} + \frac{1}{2(n-1)} \right)$$

が成立する。

一方 (1.1) に Bayes 制御 (2.11) を代入すると ∇^* による軌道は

$$\begin{aligned} x_n &= \alpha x_{n-1} + v_n^* + \gamma_n \\ &= \alpha x_{n-1} + \gamma_n - (\alpha x_{n-1} + \mu_{n-1}) \\ &= \gamma_n - \mu_{n-1} \quad (\text{ただし } \mu_0 = \gamma^{\circ} \text{ と定義がある}) \end{aligned}$$

したがって $n \geq 2$ に対し

$$E_0(x_n^2) = E_0\{(Y_n - t_{n-1}\bar{Y}_{n-1} - (1-t_{n-1})\mu)^2\}$$

$$= \frac{1}{2} + t_{n-1}^2 \frac{1}{2(n-1)} + (1-t_{n-1})^2 (\theta - \mu)^2,$$

ゆえに

$$E_0\left[\sum_1^N a_n x_n^2 \mid Y^0, V^*\right] = \frac{a_1}{2} + \sum_{n=2}^N a_n \left\{ \frac{1}{2} \left(1 + \frac{t_{n-1}^2}{n-1}\right) + (1-t_{n-1})^2 (\theta - \mu)^2 \right\},$$

したがって

$$(3.4) \quad \int E_0\left[\sum_1^N a_n x_n^2 \mid Y^0, V^*\right] \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta - \mu)^2}{2\sigma^2}} d\theta$$

$$= \frac{a_1}{2} + \sum_2^N a_n \left\{ \frac{1}{2} \left(1 + \frac{t_{n-1}^2}{n-1}\right) + (1-t_{n-1})^2 \sigma^2 \right\}$$

$$= \frac{a_1}{2} + \sum_2^N a_n \left(\frac{1}{2} + \frac{1}{2(n-1) + \sigma^{-2}} \right)$$

となつて、当然のことながらこれは (2.12) と一致する。

かくして

$$(3.5) \quad \lim_{\sigma^2 \rightarrow \infty} (3.4) = \frac{a_1}{2} + \sum_2^N \frac{a_n}{2} \left(1 + \frac{1}{n-1}\right)$$

を得る。(3.3) と (3.5) およびよく知られた補題

(補題2) > decision rule δ^0 が minimax なるための一

つの充分条件は

$\exists \{\xi^{(n)}\}_{n=1}^{\infty} : \xi^{(n)}$ に対する Bayes decision rule $\delta^{(n)}$

が $\xi^{(n)}$ を満足する。すなわち、すべての θ

に対し

$$\overline{\lim}_{n \rightarrow \infty} \int V(\theta, \delta^{(n)}) d\zeta^{(n)} \geq V(\theta, \delta^0)$$

(ここに $V(\theta, \delta^0)$ は δ^0 の risk $fct.$ である).

により, \hat{V} は minimax 制御政策であることがわかった。

さらに

(補題 3) \hat{V} は admissible control policy である。

証明: 簡単のため control policy V に対し

$$E_{\theta} \left[\sum_{n=2}^{\infty} a_n x_n^2 \mid V^0, V \right] = R(\theta, V) \text{ とおこう。また } N(\mu, \sigma^2)$$

の分布を ζ とおこう。また (3.3), (3.4) から

$$\begin{aligned} (3.6) \quad & \sigma \int (R(\theta, V^*) - R(\theta, \hat{V})) d\zeta(\theta) \\ &= \sum_{n=2}^{\infty} a_n \sigma \left(\frac{1}{2(n-1) + \sigma^{-2}} - \frac{1}{2(n-1)} \right) \\ &= \sum_{n=2}^{\infty} a_n \frac{-\sigma^{-1}}{2(n-1)(2(n-1) + \sigma^{-2})} \xrightarrow{(\sigma \rightarrow \infty)} 0 \end{aligned}$$

である。いま \hat{V} がかりに admissible ではないとすれば

$\exists V': R(\theta, V') \leq R(\theta, \hat{V}), \text{ for all } \theta \text{ and } < \text{ holds for}$
at least one θ .

$R(\theta, V)$ の θ についての連続性から

$\exists (A, B), \Delta > 0; R(\theta, V') \leq R(\theta, \hat{V}) - \Delta, \text{ in } A \leq \theta \leq B.$

すると,

$$(3.7) \quad \sigma \int (R(\theta, \hat{V}) - R(\theta, V')) d\xi(\theta) \geq \sigma \int_A^B \\ \geq \frac{\Delta}{\sqrt{2\pi}} \int_A^B e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta \xrightarrow{(\sigma \rightarrow \infty)} \frac{\Delta}{\sqrt{2\pi}} (B-A) > 0$$

とこらぞ”

$$\sigma \int (R(\theta, V^*) - R(\theta, V')) d\xi(\theta) = \sigma \int (R(\theta, V^*) - R(\theta, \hat{V})) d\xi \\ + \sigma \int (R(\theta, \hat{V}) - R(\theta, V')) d\xi$$

の右辺は (3.6), (3.7) により充分大きな σ に対し正となる。これは V^* の Bayes policy なることに矛盾する。▣

以上まとめ

[定理2] 想定 (A.1) のもとぞ” minimax control problem (3.2) に対し、一つの minimax かつ admissible な control policy は (3.1) ぞ” 与えられる。

(注意) もしも (A.1) を想定しなければ、すなわち、pilot observation r^0 をとることが許されず、最初の r_1 の分布を $N(\theta, \frac{1}{2})$, $\theta \sim N(\mu, \sigma^2)$ とすると、Bayes policy V^+ はやはり (2.6) - (2.9) ぞ” 与えられる。

ただし、このとき (2.2) の代わりに

$$(3.8) \quad \mu_0 = \mu, \quad \sigma_0^2 = \sigma^2$$

と定義しておかぬはならない。そしてこのとき (2.4) は

$$\int E_{\theta} \left[\sum_{n=1}^N a_n x_n^2 \mid Y^0, V^+ \right] \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} d\theta$$

$$= a_1 \left(\frac{1}{2} + \sigma^2 \right) + \sum_{n=2}^N a_n \left(\frac{1}{2} + \frac{1}{2(n-1) + \sigma^{-2}} \right) \xrightarrow{(\sigma^2 \rightarrow \infty)} \infty$$

となり、 σ 収束しないのである。

§4. $\{x_n\}$ が直接 observe できない場合

control system (1.1) において、 $\{x_n\}$ が直接 observe できなくて、その代わりに

$$(4.1) \quad y_n = x_n + \eta_n \quad (n=0, 1, 2, \dots)$$

$\eta_n \sim N(0, b^2)$ indep., 同-分布, b^2 は既知
が observable, つまり x_n は y_n を通して 推定される
のみとすると, optimal control はどうなるか?

(4.1) を (1.1) に代入すると

$$y_n = \alpha y_{n-1} + v_n + \gamma_n + \eta_n - \alpha y_{n-1},$$

y_0 is known.

となる。

$$\xi_n = \gamma_n + \eta_n - \alpha \eta_{n-1}, \quad n=1, 2, \dots$$

とおくと, $\xi_n \sim N(0, \frac{1}{2} + (1+\alpha^2)b^2)$ と 独立, 同-分布.

$\{\eta_n\}$ が observable だから $\{\xi_n\}$ もそうである。ゆえに

$$v_n = v_n(y_{n-1}, \xi^{n-1}).$$

また

$$\begin{aligned}
 J[v_1, \dots, v_N] &\equiv E\left[\sum_{i=1}^N (\lambda v_n^2 + a_n (y_n - \gamma_n)^2) \mid \gamma^0, \mathcal{V}\right] \\
 &= E\left[\sum_{i=1}^N (\lambda v_n^2 + a_n \gamma_n^2) + \sum_{i=1}^N a_n \{(-2x_n \gamma_n - 2\gamma_n^2) + \gamma_n^2\} \mid \gamma^0, \mathcal{V}\right] \\
 &= E\left[\sum_{i=1}^N (\lambda v_n^2 + a_n \gamma_n^2) \mid \gamma^0, \mathcal{V}\right] - b^2 \sum_{i=1}^N a_n
 \end{aligned}$$

(x_n と γ_n とは独立にから) 2"あるから §2 の議論はほとんど全部そのまま通用する。

(文献)

M. Sakaguchi, Dynamic programming for the optimal control of linear stochastic systems, Mathematica Japonica, 7 (1961/62), 79-86.