

Markovian Sequential Control Process
- Average cost criterion -

阪大 基礎工 藏 野 正美

§1 はじめに

まず考察する Markovian Sequential Control Process を定め、次に問題の定式化を行う。

State Space X を R^N の Borel 部分集合, Action Space A を有限集合とする。 $\mathcal{B} = (\mathcal{B}_X; X$ の上の σ -field, $A = \{1, 2, \dots, n(A)\}$, $\{X_t; t=0, 1, 2, \dots\}$, $\{\Delta_t; t=0, 1, 2, \dots\}$ をそれぞれ, state の系列, Action の系列とし, $\tilde{S}_t = \{(X_s, \Delta_s), s=0, 1, \dots, t\}$ と, t 時点の history を表わす。

$$\Xi = \{\xi; \xi = \langle \xi_1, \xi_2, \dots, \xi_{n(A)} \rangle, \xi_j \geq 0, \sum \xi_j = 1 \}$$

= { A の上の確率分布の全体 }

Ξ の中に値をとり, history $S_{t+1} = S_{t+1}$, と 現時点 $X_t = x$ との, 関数 $R(S_{t+1}, x) = \langle R_1(S_{t+1}, x), \dots, R_{n(A)}(S_{t+1}, x) \rangle \in$

Sequential Control Rule と呼ぶ, R を使って, 表わす。即ち, t 時点 $\tilde{S}_t = S_t$, $X_t = x$ のとき, action $j \in A$

をとり確率は, $R_j(s_{t+1}, x)$ である。

仮定 $\equiv \equiv$ $R(s_{t+1}, x)$ は, 各 argument について, Baire 函数であるとする。Control Rule のみを考える。

次に, $x \in X$, $a \in A$ を任意に与えよるとき, \mathcal{B} の上の probability measure $Q(\cdot, x, a)$ が存在して,

$$P_n \{X_{t+1} \in B \mid S_{t+1}, X_t = x, \Delta t = a\} = Q(B, x, a)$$

for every $B \in \mathcal{B}$, history S_{t+1} , が成立する

との仮定をする。

仮定

$Q(B, x, a)$ は $B \in \mathcal{B}$, $a \in A$ を任意に固定するとき, x の σ -function であり, μ がある σ -finite measure μ (on \mathcal{B}) に関して, 絶対連続, その p.d.f. を $g(\cdot, x, a)$ とすれば, x について, Baire 函数である。

初期状態 $X_0 = x$ と Control Rule R を与えよると, Sequence $\{X_t, \Delta t\}$, $t=0, 1, 2, \dots$ は Stochastic Process である。この Process を "Markovian Sequential Control Process", (M.S.C.P) と呼ぶ。

$V(x, a, x')$ を任意の M.S.C.P において, State が x で, Action a をとり, 次の時刻は State が x' に移行したときの, immediate Cost とする。

仮定 1

$V(x, a, x')$ は $X \times A \times X$ の上の非負値有界連続関数とし、かつ

$V(\cdot, a, x')$ は x' に関して、一様連続。

次に、

$$P_t(B, a, B' / x, R) = \Pr \{ X_t \in B, \Delta_t = a, X_{t+1} \in B' / X_0 = x, R \}$$

for $B, B' \in \mathcal{B}, x \in X, a \in A$

その density ; $P_t(\cdot, \cdot, \cdot / x, R)$ と定義し、

Two common measures of effectiveness of MSCP は次の通りとする。

(i) Discounted Case, $d \in (0, 1)$

$$\psi(x, d, R) = \sum_{t=0}^{\infty} d^t \int_{X \times X} \sum_{a \in A} V(v, a, u) P_t(v, a, u / x, R) \cdot \mu(dv) \times \mu(du)$$

(ii) Average Cost Criterion

$$\begin{aligned} \varphi(x, R) &= \liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \int_{X \times X} \sum_{a \in A} V(v, a, u) P_t(v, a, u / x, R) \cdot \mu(dv) \times \mu(du) \\ &= \liminf_{T \rightarrow \infty} \frac{1}{T} \varphi_T(x, R) \end{aligned}$$

定義

$\psi(x, d, R_d) \leq \psi(x, d, R)$ for $R, x \in X$, for $R_d \in$ Discounted Case における optimal Control Rule を示す。

$\varphi(x, R^*) \leq \varphi(x, R)$ for $R, x \in X$ for $R^* \in$ Average Cost Criterion における optimal Control Rule を示す。

問題は それぞれの場合における, optimal Control Rule の存在, その性質, 及び構成法の研究である。 §2 では, §3, §4 の証明に必要な Discounted Case における [1], [2] で得られる結果をのべる。 §3 では, [3] の手法を利用して, ある条件のもとでは, average Cost Criterion において, optimal Control Rule が存在する \Rightarrow を示す。 §4 では, Control Rule の改良法を示す。

§2 Optimality in discounted Case

定義

Control Rule R が Stationary であるとは, 全ての $S_{t+1}, x \in X$ に対して, $R(S_{t+1}, x) = R(x) \in A$ なる \Rightarrow である。

定理 2.1

仮定 1 のもとでは, optimal stationary Control Rule R_α が, 存在し, optimal Cost $\psi(\alpha, \alpha, R_\alpha)$ は次の方程式を満す。

$$\psi(\alpha, \alpha, R_\alpha) = \min_{a \in A} \int_X (V(\alpha, a, v) + \alpha \psi(v, \alpha, R_\alpha)) f(v, x, a) \mu(dv)$$

証明は [1] [2] 参照

$B(X) = \{X \text{ の上 に 定義 された 実数値 有界連続関数の全体} \}$

そのノルム $\|g\| = \sup_{x \in X} |g(x)|$ と定める。

$B(X)$ の上の operator T_α を次のように定める。

$$(T_\alpha g)(x) = \min_{a \in A} \int (r(x, a, v) + \alpha g(v)) g(v, x, a) \mu(dv)$$

for $g \in B(X)$

仮定 2

任意の $x \in X$, 任意の $a \in A$ に対し,

$$\lim_{x' \rightarrow x} \int_X |g(v, x, a) - g(v, x', a)| \mu(dv) = 0$$

補題 1

仮定 1, 仮定 2 のもとで, Operator T_α は縮小写像である。

証明

$$(i) |T_\alpha g(x) - T_\alpha g(x')| \leq (\|g\| + \|r\|) \left(\max_{a \in A} \int |g(v, x, a) - g(v, x', a)| \mu(dv) \right. \\ \left. + \max_{a \in A} \int |r(x, a, v) - r(x', a, v)| g(v, x, a) \mu(dv) \right)$$

$$(ii) g_1 \geq g_2 \text{ ならば, } T_\alpha g_1 \geq T_\alpha g_2 \quad (iii) T_\alpha(g+c) = T_\alpha g + \alpha c$$

$$\text{以上より, } T_\alpha g \in B(X), \|T_\alpha g_1 - T_\alpha g_2\| \leq \alpha \|g_1 - g_2\| \quad \text{証明終}$$

補題 2

$$g_0(x, \alpha) = 0 \text{ for all } x \in X, g_n(x, \alpha) = (T_\alpha g_{n-1})(x, \alpha), n=1, 2, \dots$$

とすれば, $\{g_n(x, \alpha)\}$ は $g(x, \alpha) \in B(X)$ に収束し, $g(x, \alpha)$

$$= T_\alpha g(x, \alpha), \text{ すなわち, } \lim_{n \rightarrow \infty} g_n(x, \alpha) = \psi(x, \alpha, R_\alpha)$$

証明 不動点定理と定理 2.1 (ii) による。

証明終

§ 3 Optimality in Average Cost Criterion

Average Cost Criterion について, 次の定理が成立する。

5

定理 3.1

次の等式を満足する $f(x, y) \in B(X \times X)$, $\gamma(y) \in B(X)$ が存在するとは。

$$f(x, y) + \gamma(y) = (\min_{a \in A} \int (V(x, a, v) + f(v, y)) g(v, x, a) \mu(dv))$$

for every $x \in X$, some $y \in X$.

$$R_y(a) = \left\{ x; f(x, y) + \gamma(y) = \int (V(x, a, v) + f(v, y)) g(v, x, a) \mu(dv) \right. \\ \left. \text{and } x \notin R_y(a'), a' \neq a \right\}$$

ある X の分割 $\{R_y(a), a \in X\}$ により、Control rule $R_y \in x \in R(a)$ ならば、 $R_y(S_{t+1}, x) = a$ なる Stationary Control rule である。

そのとき、

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \mathcal{J}_T(x, R) = \varphi(x, R) \geq \varphi(x, R_y) = \gamma(y)$$

for all $R, x \in X$.

注意 R_y を 上式 により構成される Control rule であることは確かである。

証明

$Z_T(y) = \sum_{t=1}^T \{ f(X_t, y) - E[f(X_t, y) / S_{t-1}] \}$ とすれば、
 $R, X_0 = x$ を任意に与えるときの M.S.C.P $\{X_t, \Delta_t, t=0, 1, 2, \dots\}$

により、 $Z_T(y)$ の期待値 $E(Z_T(y)) = 0$ である。

$$E \left[\sum_{t=1}^T [f(X_t, y) - \{ \gamma(y) + E \langle (V(X_{t+1}, \Delta_{t+1}, X_t) + f(X_t, y)) / S_{t-1} \rangle \} \right. \right. \\ \left. \left. + E(V(X_{t+1}, \Delta_{t+1}, X_t) / S_{t-1}) + \gamma(y) \right] \right] = 0$$

$$f(X_{t+1}, y) \leq \gamma(y) + E[(V(X_{t+1}, \Delta_{t+1}, X_t) + f(X_t, y)) / S_{t-1}]$$

等式が成立するのは、Rule R_y による M.S.C.P である。

故に次の不等式が成立する。

$$E\left\{\sum_{t=1}^T [f(X_t, y) - f(X_{t+1}, y)] + \sum_{t=1}^T E(r(X_{t+1}, a_t, X_t)) + T\gamma(y)\right\} \geq 0$$

$$= 0 \text{ であり, } E\{f(X_T, y) - f(X_0, y)\} + \varphi_T(\alpha, R) \geq T\gamma(y)$$

f の有界性により, $\liminf_{T \rightarrow \infty} \frac{1}{T} \varphi_T(\alpha, R) \geq \gamma(y)$ 証明終

次に, $g_\alpha(x) = \varphi(x, \alpha, R_\alpha)$, $f_\alpha(x, y) = g_\alpha(x) - g_\alpha(y)$ とおけば,
次の等式が成立する。

$$\begin{aligned} f_\alpha(x, y) + \gamma_\alpha(y) &= \left(\min_{a \in A} \int (r(x, a, v) + \alpha f(v, y)) g(v, x, a) \mu(dv) \right) \\ &= (T_\alpha f(\cdot, y))(x), \quad \text{但し } \gamma_\alpha(y) = (1-\alpha) g_\alpha(y) \end{aligned}$$

$x = z$, 関数族 $\{f_\alpha(x, y), 0 < \alpha < 1\}$ の性質を調べる

仮定3

$$\sup_{x \in X, y \in X, a, a' \in A} \int |g(v, x, a) - g(v, x, a')| \mu(dv) = \beta < 1$$

補題3

仮定1, 仮定3 のもとで, 関数族 $\{f_\alpha, \alpha \in (0, 1)\}$ は一様有界

証明

$$f_\alpha^{(m)}(x, y) = g_\alpha^{(m)}(x, x) - g_\alpha^{(m)}(x, y)$$

$$g_\alpha^{(m)}(x, y) = \int (r(y, a(y), v) + \alpha g_{\alpha}^{(m-1)}(v, y)) g(v, y, a(y)) \mu(dv) \quad \forall y \in A$$

が存在する。従って, $f_\alpha^{(m)}(x, y) = \min_{a \in A} \left[\int r(x, a, v) g(v, x, a) \mu(dv) \right.$

$$\left. - \int r(x, a, v) g(v, y, a(y)) \mu(dv) + \alpha \int f_\alpha^{(m-1)}(v, y) (g(v, x, a) - g(v, y, a(y))) \mu(dv) \right]$$

$\|r\| \leq K$ とし, $B > \frac{K}{1-\beta} > K$ かつ $B \in \mathbb{N}$, 証明は帰納法。

$f_\alpha^{(0)}(x, y) = 0 < B$, $|f_\alpha^{(n-1)}(x, y)| \leq B$ とおくと, 前式より,

7.

$$-B < -(K+\beta B) \leq f_\alpha^{(m)}(\alpha, \gamma) \leq K+\beta B < B$$

結局 $|f_\alpha^{(m)}(\alpha, \gamma)| \leq B$. $\alpha \neq 1$, $|f_\alpha(\alpha, \gamma)| \leq B$ ($\alpha \rightarrow \infty$)

仮定4 State Space X は Compact Set である。

補題4

仮定1. 2. 4. のもとで, 関数族 $\{f_\alpha, \alpha \in (0, 1)\}$ が一様有界,

(i.e. $\|f_\alpha\| \leq M$) ならば, 族 $\{f_\alpha\}$ は同程度連続である。

証明

$$\begin{aligned} |f_\alpha(\alpha, \gamma) - f_\alpha(\alpha', \gamma')| &\leq (K+M) \left(\max_{a \in A} \int |g(v; \alpha, a) - g(v; \alpha', a)| u(dv) \right) \\ &\quad + (K+M) \left(\max_{a \in A} \int |g(v, \gamma, a) - g(v, \gamma', a)| u(dv) \right) \\ &\quad + \left(\max_{a \in A} \int |r(\alpha, a, v) - r(\alpha', a, v)| g(v, \alpha, a) u(dv) \right) \\ &\quad + \left(\max_{a \in A} \int |r(\gamma, a, v) - r(\gamma', a, v)| g(v, \alpha, a) u(dv) \right) \end{aligned}$$

$\alpha' \rightarrow \alpha, \gamma' \rightarrow \gamma$ のとき右辺は $\rightarrow 0$ となる。 証明

以上の結果と Ascoli-Arzelà の定理により, 次の定理が得られる。

定理3.2

$$f_\alpha(\alpha, \gamma) = \phi(\alpha, \alpha, R_\alpha) - \psi(\gamma, \alpha, R_\alpha), \quad \gamma_\alpha(\gamma) = (1-\alpha)\psi(\gamma, \alpha, R_\alpha)$$

仮定1. 2. 3. 4 (仮定3は族が一様有界なる条件であった) のもとで,

次の (i) (ii) が成立する。

(i) (a), (b) が成立する $f \in B(X \times X), \gamma \in B(X)$ が存在する。

$$(a) \gamma_\alpha(\gamma) \rightarrow \gamma(\gamma), \quad \text{as } \alpha \rightarrow 1$$

$$(b) f(x, y) + \gamma(y) = \min_{a \in A} \int (r(x, a, v) + f(v, y)) g(v, x, a) \mu(dv)$$

(i) $\gamma_0(y) = \inf_R \int \varphi(y, R) g$ とおくと、(b), (c), (d) のうち (i) 成立

$$(c) \gamma_0(y) = \gamma(y) = \text{一定}$$

(d) optimal stationary rule R^* が存在する。

証明

$d_n \rightarrow 1$ となる d_n が存在する。 $\{f_{d_n}, n=1, 2, \dots\}$ は一様有界、かつ同程度連続だから、Ascoli-Arzelà の定理により、

$f_{d_n} \rightarrow f$ (一様収束) となる $f \in B(X \times X)$, d_n が d_n が存在する。

同様にして、 $a \in (0, 1)$ が一様有界、同程度連続だから、明らかに a があるから、 $\gamma_{d_n} \rightarrow \gamma$ (一様収束) となる $\gamma \in B(X)$ が存在する (2) 成立。

$$f_{d_n}(x, y) + \gamma_{d_n}(y) = \min_{a \in A} \int (r(x, a, v) + d_n f_{d_n}(v, y)) g(v, x, a) \mu(dv)$$

(2) により、 $\{f_{d_n}\}$ の一様収束性により、

$$f(x, y) + \gamma(y) = \min_{a \in A} \int (r(x, a, v) + f(v, y)) g(v, x, a) \mu(dv) \quad \text{om,}$$

成立する。

上式により、 γ が構成される stationary rule $R_\gamma \in R_\gamma$ とすれば、定理

3.1 により、 $\gamma(x) = \varphi(x, R_\gamma)$ for all $x \in X$.

一方、 $\liminf_{d \rightarrow 1} (1-d)\psi(x, R) = \varphi(x, R)$ [6] 参照。従って、

任意の Control rule R に対し、

$$\gamma(x) = \lim_{n \rightarrow \infty} (1-d_n)\psi(x, d_n, R_{d_n}) \leq \lim_{n \rightarrow \infty} (1-d_n)\psi(x, d_n, R) = \varphi(x, R).$$

故に $\gamma_0(x) = \inf_R \varphi(x, R) \geq \gamma(x)$, $\gamma(x) \geq \gamma_0(x)$ は明しおたさ
 $\gamma_0(x) = \gamma(x)$

$\{\gamma_n, \gamma_{n+1}\}$ の任意の収束列 γ_n には, $\lim_{n \rightarrow \infty} \gamma_n = \gamma_0$ である,

$\lim_{\alpha \rightarrow 1} \gamma_\alpha = \gamma_0$ を意味する。

$$\gamma(\gamma_0) = \min_{y \in X} \gamma(y) \text{ である。}$$

$$f(x, \gamma_0) + \gamma(\gamma_0) = \min_{a \in A} \int (r(x, a, v) + f(v, \gamma_0)) g(v, x, a) \mu(dv)$$

よって、この構成された Stationary rule R_{γ_0} は optimal である。

§4 改良法

Stationary Control rule の class を C で表わす。任意の $R \in C$ に対して M, S, P は Discrete parameter を持つ Markov Process である。以後 μ は dehasque 測度とする。

補題

仮定 1. 2. 3. 4 のもとで、任意の $R \in C$ に対して、次の等式が、 μ -almost all x に対して成立する有界な可測函数 $f^R(x)$, 実数 γ^R ($-\infty < \gamma^R < \infty$) が存在する。

$$f^R(x) + \gamma^R = \int (r(x, R(x), v) + f^R(v)) g(v, x, R(x)) \mu(dv)$$

証明

補題 3, 補題 4, 定理 3, 2 の (i) の証明方法と dehasque 測度の性質によって証明できる。省略する。

定理3.1の証明を補題5の等式を存おめ子 = とはまって

, $\varphi(x, R) \geq \gamma$ for all $x \in X$ 存子 = とお容易に知れ子。

$$E_R(x, a) = \int (r(x, R(x), v) + f^R(v)g(v; x, R(x))) \mu(dv) \\ - \int (r(x, a, v) + f^R(v)g(v; x, a)) \mu(dv)$$

$$E_R(x) = \max_{a \in A} E_R(x, a) \text{ とおければ, } E_R(x) \geq 0$$

$$X_R = \{x; E_R(x) = 0\}$$

$$R'_a = \{x; x \in \overline{X_R}, E_R(x) = E_R(x, a), x \notin R'(a'), a' < a\}$$

とおければ $\{R'_a, a \in A\}$ は $\overline{X_R}$ の分割である。 $\epsilon = \epsilon'$, 次の様な Stationary Control Rule R' を構成する。

$$R': \begin{cases} R'(x) = R(x) & x \in X_R \\ R'(x) = a & x \in \overline{X_R}, \text{ and } x \in R'_a \end{cases}$$

定理4.1

補題5の仮定のもとで, $\varphi(x, R') \leq \varphi(x, R)$ for all $x \in X$

もし $Q^*(\overline{X_R}; R') > 0$ 存らば, $\varphi(x, R') < \varphi(x, R)$ for all $x \in X$.

注意

$R \in C$ にお出来る Markov Process は 仮定のもとでは, ergodic

set は たた \rightarrow っ, かつ cyclically moving subset を たた 11。

従って 極限分布 $Q^*(\cdot, x, R)$ は 初期値 x に 依存 $L T \rightarrow 11$ 。 [7]

証明

R' にお出来る Markov Process の $X_0 = x'$ にお出る t -Step Transition

Probability $\in P_t(\cdot, x', k')$ \in あり。

$$E_R(x) = f^R(x) + \gamma^R - \int_X (r(x, k(x), v) + f^R(v)) g(v, x, k(x)) \mu(dv)$$

従って。

$$\int_X E_R(x) P_t(dx, x', k') = \gamma^R + \int_X f^R(x) P_t(dx, x', k') - \int_X f^R(x) P_{t+1}(dx, x', k') \\ - \int_X P_t(dx, x', k') \int_X (r(x, k(x), v) + f^R(v)) g(v, x, k(x)) \mu(dv)$$

故に

$$\frac{1}{T} \sum_{t=0}^{T-1} \int_X E_R(x) P_t(dx, x', k') = \gamma^R + \frac{1}{T} \left(\int_X f^R(x) P_0(dx, x', k') \right. \\ \left. - \int_X f^R(x) P_T(dx, x', k') \right) - \frac{1}{T} \varphi(x', k')$$

$\frac{1}{T} \sum_{t=0}^{T-1} P_t(B, x', k')$ は $Q^*(B, k')$ に $B \in \mathcal{B}$ と $x' \in X$ に関し
 一樣に収束する。

$$\text{従って} \int_X E_R(x) Q^*(dx, k') = \gamma^R - \varphi(x', k')$$

$$E_R(x) > 0, x \in \bar{X}_R \quad \text{従って} \quad Q^*(\bar{X}_R, k) > 0 \text{ 存在する。} \quad \gamma^R = \varphi(x', k) \\ > \varphi(x', k') \\ \text{〈証明終〉}$$

Reference

- [1] D. Blackwell, Discounted dynamic Programming
Ann Math Statist., 36 (1965) 226-235
- [2] R. F. Strauch, Negative Dynamic Programming,
Ann Math Statist., 37 (1966), 871-890.

- [3] Taylor Howard, Markovian Sequential Replacement Process, *Ann. math. Statist.*, 36 (1965) 1697-1694
- [4] Ugras Perman, Denumerable State Markovian Decision Process - average Cost Criterion -
ann. math. Statist., Vol 37, No 6 (1966)
- [5] Ugras Perman and A. F. Veinatt
a solution to a countable system of equalities arising
in Markovian decision process
ann. math. Statist., Vol 38, No 2 (1967)
- [6] Hardy "Divergent Series" 1949 Oxford
- [7] Doob "Stochastic Process"