

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。 $\mathcal{B} = (\mathcal{B}_X; X \text{ の上 の } \sigma\text{-field}, A = \{1, 2, \dots, n(A)\}, \{X_t; t=0, 1, 2, \dots\}, \{\Delta_t; t=0, 1, 2, \dots\} \in \text{それぞれ, state の系列, Action の系列とし, } \tilde{S}_t = \{(X_s, \Delta_s), s=0, 1, \dots, t\} \text{ じ, } t \text{ 時点の history を表わす。}$

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

$= \{A \text{ の上 の 確率分布の全体} \}$

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

Sequential Control Rule と呼ぶ, R を使って, 表わす。即ち, t 時点じ, history $\tilde{S}_{t+1} = \tilde{S}_t, X_t = x$ のとき, action $j \in A$

を収める確率は, $R_j(s_{t+1}, x)$ である。

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

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

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

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

と仮定する。

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 -additive 数で, かつある 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 を与えたと, 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, 次の時刻に 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,

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

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

と 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, $\alpha \in (0, 1)$

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

(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

$\psi(x, \alpha, R_\alpha) \leq \psi(x, \alpha, R)$ for $R, x \in X$, かつ $R_\alpha \in$ Discounted Case における optimal Control Rule とする。

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

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] で得られた結果をのべる。§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, すべて $S_{t+1}, x \in X$ 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

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

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 $a \in A$ に対して,

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

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

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

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

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

証明終

補題 2

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

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

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

証明 不動点定理と定理 2.1 (i) 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。

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 $y \in X$.

$$R_y(a) = \{x; f(x, y) + \gamma(y) = \int (V(x, a, v) + f(v, y)) g(v, x, a) \mu(dv) \\ \text{and } x \in R_y(a'), a' < a\}$$

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

そのとき、

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

for all $R, x \in X$.

注意 R_y は 上式により構成される Control Rule であることがわかる。

証明

$$Z_T(y) = \sum_{t=1}^T \{f(X_t, y) - E[f(X_t, y) / S_{t-1}]\} \quad \text{とすれば、}$$

$R, X_0 = x$ 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。

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

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

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。

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

$$= 0 \neq 1, \quad E\{f(X_T, y) - f(X_0, y)\} + \varphi_T(\cdot, R) \geq T\gamma(y)$$

f の有界性により, $\liminf_{T \rightarrow \infty} \frac{1}{T} \varphi_T(\cdot, R) \geq \gamma(y)$ 証明終
 次に, $g_\alpha(x) = \varphi(x, \alpha, R)$, $f_\alpha(x, y) = g_\alpha(x) - g_\alpha(y)$ とおけば,
 次の等式が成立する。

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

$x = y$, 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

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

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

が存在する。従って, $f_\alpha^{(m)}(x, y) = \min_{a \in A} \left[\int r(x, a, v) g(v, x, a) \mu(dv) \right.$
 $\left. - \int r(x, a, v) g(v, y, a(y)) \mu(dv) + \alpha \int f_\alpha^{(m-1)}(v, y) (g(v, x, a) - g(v, y, a(y))) \mu(dv) \right]$

$\|r\| \leq K \in \mathbb{R}, \quad B > \frac{K}{1-\beta} > K$ かつ $B \in \mathbb{R}$, 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.

$$-B < -(K+\beta B) \leq f_\alpha^{(n)}(x, y) \leq K+\beta B < B$$

結局 $|f_\alpha^{(n)}(x, y)| \leq B$. $n \neq 1$, $|f_\alpha(x, y)| \leq B$ ($\text{as } n \rightarrow \infty$)

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

証明

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

$x' \rightarrow x, y \rightarrow y'$ のとき右辺は $\rightarrow 0$ である。 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

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

仮定1. 2. 3. 4 (仮定3は $\{f_\alpha\}$ が一様有界である条件であった) 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。

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

- (b) $f(x, y) + \gamma(y) = \min_{a \in A} \int (r(x, a, v) + f(v, y)) g(v, x, a) u(dv)$
- (i) $\gamma_0(y) = \inf_R \{ \varphi(y, R) \}$ とおき、(b), (c), (d) のどれか成立する。
- (c) $\gamma_0(y) = \gamma(y) = -\text{定数}$
- (d) optimal stationary rule R^* が存在する。

証明

$d_n \rightarrow 1$ なる $\{d_n\}$ が存在する。 $\{f_{d_n}, n=1, 2, \dots\}$ は一様有界、かつ同程度連続な故に、Ascoli-Arzelà の定理により、 $f_{d_n} \rightarrow f$ (一様収束) なる $f \in B(X \times X)$, $\{d_n\} \subset \{d_n\}$ が存在する。同数値 $\alpha \in (0, 1)$ は一様有界、同程度連続な $\gamma_{d_n} \rightarrow \gamma$ (一様収束) なる $\gamma \in B(X)$ が存在する (2.2 参照)。

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

において、 $\{f_{d_n}\}$ の一様収束性に注意して、

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

成立する。

上式により、2 構成される stationary rule $\pi \in R_y$ とすれば、定理 3.1 により、 $\gamma(y) = \varphi(x, R_y)$ 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 $\gamma_0(x) = \inf_R \varphi(x, R) \geq \gamma(x)$, $\gamma(x) \geq \gamma_0(x)$ は明かであるから
 $\gamma_0(x) = \gamma(x)$

$\{\gamma_n, \gamma_n\}$ の任意の収束列 γ_n 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。

$$\gamma(\gamma_0) = \min_{y \in X} \gamma(y) \leq \gamma_0.$$

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

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 $R \in C$ に対してある M, S, P は Discrete parameter を持つ Markov Process である。以後 μ は Lebesgue 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 Lebesgue 測度の性質によって証明できるのである。

定理3.1の証明と補題5の等式をたがめるとともに

, $\varphi(x, R) \geq \gamma$ for all $x \in X$ なることを容易に知れり。

$$\begin{aligned} E_R(x, a) &= \int (r(x, R(\omega), v) + f^R(v) g(v; x, R(\omega))) \mu(dv) \\ &\quad - \int (r(x, a, v) + f^R(v) g(v; x, a)) \mu(dv) \end{aligned}$$

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

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

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

とおけば $\{R'_a, a \in A\}$ は $\overline{X_R}$ の分割である。これより、次の様な Stationary Control Rule R' を構成する。

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

定理4.1

補題5の仮定のもとで, $\varphi(x, R') \leq \varphi(x, R)$ for all $x \in X$
 とし $Q^*(\overline{X_R}; R') > 0$ ならば, $\varphi(x, R') < \varphi(x, R)$ for all $x \in X$.

注意

$R \in C$ によって出来る Markov process は仮定のもとでは, ergodic set はただ一つで、かつ cyclically moving subset をもたない。
 従って 極限分布 $Q^*(\cdot, x, R)$ は初期値 x に依存しない。 [7]

証明

R' に対する Markov process の $X_0 = x'$ における t -Step Transition

//

probability $\in P_t(\cdot, x', k')$ である。

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

から、

$$\begin{aligned} \int_X E_k(\omega) P_t(dx, x', k') &= \gamma^k + \int_X f^k(\omega) P_t(dx, x', k') - \int_X f^k(\omega) P_{t+1}(dx, x', k') \\ &\quad - \int_X P_t(dx, x', k) \int_X (r(x, k(\omega), v) + f^k(v)) g(v, x, k(\omega)) \mu(dv) \end{aligned}$$

故に

$$\begin{aligned} \frac{1}{T} \sum_{t=0}^{T-1} \int_X E_k(\omega) P_t(dx, x', k') &= \gamma^k + \frac{1}{T} \left(\int_X f^k(\omega) P_0(dx, x', k') \right. \\ &\quad \left. - \int_X f^k(\omega) P_T(dx, x', k') \right) - \frac{1}{T} \varphi(x', k) \end{aligned}$$

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

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

$$E_k(x) > 0, x \in \bar{X}_k \text{ から } Q^*(\bar{X}_k, k) > 0 \text{ である。 } \gamma^k = \varphi(x', k) > \varphi(x', k')$$

(証明終)

Reference

- [1] D. Blackwell, Discounted dynamic Programming
Ann Math Statist., 36 (1965) 226-235
- [2] R. E. 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] Gyrgus Permon, Denumerable State Markovian Decision Process — average Cost Criterion —
ann. math. Statist., Vol 37. No 6 (1966)
- [5] Gyrgus Permon 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"

