数理解析研究所講究録
48巻 1968年 39-52

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。

<u>State Space</u> X を R^N の Borel 部分集合，<u>Action Space</u> A を，有限集合とする。$\mathcal{B}=\mathcal{B}_X$; X の上の σ-field，$A=\{1,2,\cdots,n(A)\}$，$\{X_t\,;\,t=0,1,2,\cdots\}$，$\{\Delta_t\,;\,t=0,1,2,\cdots\}$ をそれぞれ，State の系列，Action の系列とし，$S_t=\{(X_\tau,\Delta_\tau),\ \tau=0,1,\cdots,t\}$ で，t 時点の history を表わす。

$$\Xi=\{\xi\,;\,\xi=\langle\xi_1,\xi_2,\cdots,\xi_{n(A)}\rangle,\ \xi_j\ge 0,\ \sum\xi_j=1\}$$
$$=\{A \text{ の上の確率分布の全体}\}$$

Ξ の中に値をもつ，history $S_{t-1}=s_{t-1}$ と現時点 $X_t=x$ との，関数 $R(s_{t-1},x)=\langle R_1(s_{t-1},x),\cdots,R_{n(A)}(s_{t-1},x)\rangle$ を <u>Sequential Control Rule</u> と呼び，R をもって，表わす。即ち，t 時点で，history $S_{t-1}=s_{t-1}$，$X_t=x$ のとき，action $j\in A$

を取る確率は，$R_j(S_{t-1}, x)$ である。

仮定　ここでは，$R(S_{t-1}, x)$ が，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，$\mathfrak{B}$ の上の Probability measure $Q(\cdot, x, a)$ が存在して，

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

for every $B \in \mathfrak{B}$, history S_{t-1}，が成立するものと仮定する。

仮定

$Q(B, x, a)$ は $B \in \mathfrak{B}$, $a \in A$ を任意に固定するとき，x のベール函数で，かつある σ-finite measure μ (on $\mathfrak{B}$) 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 を与えると，Sequence $\{(X_t, \Delta_t), t = 0, 1, 2, \cdots\}$ 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，次の時点に State が，x' に移行したときの，immediate Cost とする。

仮定1

$r(x,a,x')$ は $X\times A\times X$ の上の非負値有界連続関数とし，かつ $r(\cdot,a,x')$ は x' に関して，一様連続。

次に，

$$P_t(B,a,B'/x,R)=P_r\{X_t\in B,\ \Delta_t=a,\ X_{t+1}\in B'/X_0=x,R\}$$
$$\text{for } B,B'\in\mathcal{B},\ x\in X,\ a\in A$$

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 , $\alpha\in(0,1)$

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

(ii) Average Cost Criterion

$$\varphi(x,R)=\liminf_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\int_{X\times X}\sum_{a\in A}r(v,a,u)\,p_t(v,a,u/x,R)\cdot\mu(dv)\times\mu(du)$$

$$=\liminf_{T\to\infty}\frac{1}{T}\varphi_T(x,R)$$

定義

$\psi(x,\alpha,R_\alpha)\leqq\psi(x,\alpha,R)$ for $R,\ x\in X$, なる R_α を Discounted Case における optimal Control Rule と呼び，

$\varphi(x,R^*)\leqq\varphi(x,R)$ for $R,\ x\in X$ なる R^* を Average Cost Criterion における optimal Control Rule と呼ぶ。

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 においても，Optimal Control Rule が存在することを示す。§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，すべての S_{t-1}, $x \in X$ に対して，$R(S_{t-1}, x) = R(x) \in A$ なることである。

定理 2.1

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

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

証明は [1] [2] 参照

$B(X) = \{X$ の上に定義された実数値有界連続関数の全体$\}$

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

$B(X)$ の上の Operator T_α を次のように定める。

$$(T_\alpha g)(x) = \min_{a\in A}\left\{\int (r(x,a,v)+\alpha g(v))\, q(v,x,a)\,\mu(dv)\right.$$
$$\text{for } g\in B(X)$$

仮定2

任意の $x\in X$, 任意の $a\in A$ に対して,

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

補題1

仮定1, 仮定2 のもとで, Operator T_α は縮少写像である。

証明

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

$$+\max_{a\in A}\int |r(x,a,v)-r(x',a,v)|\,q(v,x,a)\,\mu(dv)$$

(ii) $g_1 \geqq g_2$ ならば, $T_\alpha g_1 \geqq T_\alpha g_2$ (iii) $T_\alpha(g+c) = T_\alpha g + \alpha c$

以上により, $T_\alpha g\in B(X)$, $\|T_\alpha g_1 - T_\alpha g_2\| \leq \alpha\|g_1-g_2\|$ 　証終

補題2

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

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

$= T_\alpha g(\alpha,\cdot)$, さらに, $\lim_{n\to\infty} g_n(\alpha,\cdot) = \psi(\cdot,\alpha,R_\alpha)$

証明 不動点定理と定理2,1より明らか 　証終

§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 $f(x,y)\in B(X\times X)$, $\gamma(y)\in B(X)$ が存在するとす。

$$f(x,y)+\gamma(y)=\min_{a\in A}\int(r(x,a,v)+f(v,y))g(v,x,a)\mu(dv)$$
$$\text{for every } x\in X,\ \text{some } y\in X.$$

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

なる X の分割 $\{R_y(a),\ a\in X\}$ において, Control Rule R_y を $x\in R(a)$ ならば, $R_y(S_{t-1},x)=a$ なる Stationary Control Rule とす。

そのとき,

$$\liminf_{T\to\infty}\frac{1}{T}\varphi_T(x,R)=\varphi(x,R)\geqq\varphi(x,R_y)=\gamma(y)$$
$$\text{for all } R,\ x\in X.$$

注意 R_y を上式によって構成される Control Rule とよぶことにする

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 $\{(X_t,\Delta_t),\ t=0,1,2\cdots\}$ において, $Z_T(y)$ の期待値 $E(Z_T(y))=0$ である。

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

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

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。

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

これより，$E\{f(X_T,y)-f(X_0,y)\}+\varphi_T(x,R)\geqq T\gamma(y)$

fの有界性により，$\liminf_{T\to\infty}\frac{1}{T}\varphi_T(x,R)\geqq\gamma(y)$ 証終

次に，$g_\alpha(x)=\psi(x,\alpha,R_\alpha)$，$f_\alpha(x,y)=g_\alpha(x)-g_\alpha(y)$ とすれば，次の等式が成立する。

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

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

仮定3

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

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

$$f_\alpha^{(n)}(x,y)=g_n(\alpha,x)-g_n(\alpha,y)$$

$g_n(\alpha,y)=\int(r(y,a(y),v)+\alpha g_{n-1}(\alpha,v))g(v,y,a(y))\mu(dv)$ なる$a(y)\in A$が存在する。従って，$f_\alpha^{(n)}(x,y)=\min_{a\in A}\Big[\int r(x,a,v)g(v;x,a)\mu(dv)$

$-\int r(x,a,v)g(v,y,a(y))\mu(dv)+\alpha\int f_\alpha^{(n-1)}(v,y)(g(v;x,a)-g(v;y,a(y))\cdot\mu(dv)$

$\|r\|\leq K$として，$B>\frac{K}{1-\beta}>K$なるBをとり，証明は帰納法。

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

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

結局 $|f_\alpha^{(n)}(x,y)| \leq B$. これより, $|f_\alpha(x,y)| \leq B$ (as $n \to \infty$)

仮定4　State Space X は Compact Set とする。

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

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

$x' \to x$, $y \to y'$ のとき右辺はことごとく0に収束する　証終

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

定理3.2

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

仮定1.2.3.4 (仮定3は $\{f_\alpha\}$ が一様有界の為の条件であった) 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) $\gamma_\alpha(y) \to \gamma(y)$, as $\alpha \to 1$

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

(ii) $\gamma_0(y)=\inf_R\{\varphi(y,R)\}$ とすれば，(c)，(d) が成り立つ

(c) $\gamma_0(y)=\gamma(y)=$ 一定

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

証明

$\alpha_n\to 1$ なる $\{\alpha_n\}$ が存在する。$\{f_{\alpha_n},\ n=1,2\cdots\}$ は一様有界，かつ同程度連続なる故に，Ascoli-Arzela の定理により，

$f_{\alpha_{n'}}\to f$（一様収束）なる $f\in B(X\times X)$，$\{\alpha_{n'}\}\subset\{\alpha_n\}$ が存在する。

関数族 $\{\gamma_\alpha,\ \alpha\in(0,1)\}$ は一様有界，同程度連続なることは明らかであるから，$\gamma_{\alpha_{n'}}\to\gamma$（一様収束）なる $\gamma\in B(X)$ が存在するとしてよい。

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

において，$\{f_{\alpha_{n'}}\}$ の一様収束性により，

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

成立する。

上式によって構成される stationary rule を R_y とすれば，定理 3.1 によって，$\gamma(y)=\varphi(x,R_y)$ for all $x\in X$.

一方，$\liminf_{\alpha\to 1}(1-\alpha)\psi(x,R)=\varphi(x,R)$ [6] 参照。従って任意の Control rule R に対して，

$$\gamma(x)=\lim_{n\to\infty}(1-\alpha_{n'})\psi(x,\alpha_n,R_{\alpha_n})\leqq\varliminf_{n\to\infty}(1-\alpha_n)\psi(x,\alpha_n,R)$$
$$=\varphi(x,R).$$

故に $\gamma_0(x)=\inf_R \varphi(x,R) \geqq \gamma(x)$, $\gamma(x) \geqq \gamma_0(x)$ は明らかだから

$\gamma_0(x)=\gamma(x)$

$\{\alpha_n, \gamma_{\alpha_n}\}$ が任意の収束列ならば, $\lim_{n\to\infty}\gamma_{\alpha_n}=\gamma_0$ これは,

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

$\gamma(y_0)=\min_{y\in X}\gamma(y)$ として,

$$f(x,y_0)+\gamma(y_0)=\min_{a\in A}\int(r(x,a,v)+f(v,y_0))\,g(v,x,a)\,\mu(dv)$$

によって構成される Stationary Rule R_{y_0} は Optimal である。

§4 改良法

Stationary Control Rule の class を C で表わす。任意の $R\in C$ に対して定まる M.S.C.P は Discrete parameter をもつ Markov Process である。以後 μ は Lebesgue 測度とする。

<u>補題</u>

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 $(-\infty<\gamma<\infty)$ が存在する

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

<u>証明</u>

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されるが, 省略する。

定理3.1の証明と補題5の等式をながめることによって，$\varphi(x, R) = \gamma$ for all $x \in X$ なることが容易に知れる。

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

$\varepsilon_R(x) = \max_{a \in A} \varepsilon_R(x,a)$ とすれば，$\varepsilon_R(x) \geqq 0$

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

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

とすれば $\{R'_a, a \in A\}$ は $\overline{X_R}$ の分割である。そこで，次の様な Stationary Control Rule R' を構成する。

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

<u>定理4.1</u>

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

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

<u>注意</u>

$R \in C$ によって出来る markov process は仮定3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]

<u>証明</u>

R' に対する markov process の $X_0 = x'$ における t-Step Transition probability を $P_t(\cdot, x', k')$ とする。

$$\varepsilon_k(x) = f^k(x) + \gamma^k - \int_X (r(x, k(x), v) + f^k(v))\, q(v, x, k'(x))\, \mu(dv)$$

従って

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

故に

$$\frac{1}{T}\sum_{t=0}^{T-1} \int_X \varepsilon_k(x) P_t(dx, x', k') = \gamma^k + \frac{1}{T}\Big(\int_X f^k(x) P_0(dx, x', k') - \int_X f^k(x) P_T(dx, x', k')\Big) - \frac{1}{T}\varphi_T(x', k')$$

$\frac{1}{T}\sum_{t=0}^{T-1} P_t(B, x', k')$ は $Q^*(B, k')$ に $B \in \mathcal{B}$ と、$x' \in X$ に関して一様に収束する。

従って $\int_X \varepsilon_k(x)\, Q^*(dx, k') = \gamma^k - \varphi(x', k')$

$\varepsilon_k(x) > 0, x \in \overline{X}_k$ 従って $Q^*(\overline{X}_k ; k) > 0$ ならば $\gamma^k = \varphi(x', k)$
$> \varphi(x', k')$

〈証終〉

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) 1677 – 1694

[4] Cyrus Derman, Denumerable State Markovian Decision Process — average Cost Criterion — ann. Math. Statist., Vol 37. No6 ((1966)

[5] Cyrus Derman and A. F Veinott a solution to a countable system of equalities arising in Markovian decision process ann. math. Statist., Vol 38, No2 ((1967)

[6] Hardy "Divergent Series" 1949 oxford

[7] Doob "Stochastic Process"