

多重連鎖マルコフ決定過程の逐次近似法について

名古屋工業大学 大野 勝久 (Katsuhisa Ohno)

1. 序論

有限状態, 有限決定をもつマルコフ決定過程において, 時間平均利得を最大にする最適定常政策を決定するアルゴリズムとしては, 政策反復法, 逐次近似法, 線形計画法, 修正政策反復法が知られている。特に逐次近似法, 修正政策反復法は多状態問題にたいする有力な手法として多くの研究が行なわれているが, 多重連鎖問題については余り論じられていない。

Bather (1973) は多重連鎖問題にたいする communicating sets への分解と政策反復法に関連した逐次近似法を提案し, その収束を示している。また ε -最適政策の構成をも論じているが, 反復回数が十分大きければ ε -最適政策がえられると述べている。Schweitzer (1984) もまた communicating sets への分解 (unique chain decomposition と名づけている) にもと

づく逐次近似アルゴリズムを提案し、その収束を示しているが、 ε -最適政策を保証する停止基準は与えられていない。本論文では ε -最適政策を求める逐次近似法を Bather (1973) を参考に論ずる。

2. 準備

以下の記号を使用する。

$I = \{1, 2, \dots, M\}$: 状態空間

$K_i (i \in I)$: 状態 i でとりうる決定の有限集合

$r_i(k) (i \in I, k \in K_i)$: i において k をとってえられる平均利得

$p_{ij}(k) (i, j \in I, k \in K_i)$: i において k をとった時 j へ遷移する確率

F : 定常政策 $f = (f_1, \dots, f_M)$ の集合 ($f_i \in K_i, i \in I$)

$r(f) = (r_1(f_1), \dots, r_M(f_M))^T$ (T は転置を表す)

$P(f) = (p_{ij}(f_i))$: f のもとでの遷移確率行列

$P^*(f)$: $P(f)$ のアイガロ和 ($\lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N P(f)^n$)

$g(f) = P^*(f)r(f)$: f の利得 (gain)

$v(f)$: f の相対値

$g(f)$ は

$$(1) \quad g = P(f)g$$

$$(2) \quad g + v = r(f) + P(f)v$$

の一解であり、 $v(f)$ は $P(f)$ の各エルゴード連鎖に属する 1

この状態 i で $v_i(f) = 0$ とおけば一意に決定される。時間平均マルコフ決定過程は 4 つ組 (I, r, p, F) で与えられ、利得を最大にする政策 $f^* \in F$ を決定する問題である。

最大利得 $g^* = g(f^*) = \max_{f \in F} p^*(f) r(f)$ および相対値 $v^* = v(f^*)$ のみたる最適方程式は

$$(3) \quad g_i^* = \max_{k \in K_i} \left\{ \sum_{j \in I} p_{ij}(k) g_j^* \right\} \quad (i \in I)$$

$$(4) \quad g_i^* + v_i^* = \max_{k \in L_i} \left\{ r_i(k) + \sum_{j \in I} p_{ij}(k) v_j^* \right\} \quad (i \in I)$$

である。ここで

$$L_i = \{ k \in K_i; (3) \text{ 式右辺を最大にする } k \}$$

である。定常政策子 f は、 $e = (1, \dots, 1)^T$ にたいして

$$(5) \quad g^* - g(f) \leq \varepsilon e$$

を満たすとき、 ε -最適政策とよばれる。

[定理 1]

$f \in F$ および M 次元ベクトル g, v にたいして

$$(6) \quad \psi = g - p(f)g, \quad \gamma = g + v - r(f) - p(f)v$$

とおく。 $p(f)$ が

$$(7) \quad p(f) = \begin{pmatrix} Q(f) & 0 \\ R(f) & S(f) \end{pmatrix} \quad (S(f) \text{ は過渡行列})$$

で与えられ、 $Q(f), S(f)$ に対応する部分ベクトルを添字 c, t で表わしたとき

$$(8) \quad \psi_c \leq (\geq) 0, \quad \psi_t \leq (\geq) a_t, \quad \gamma_c \leq (\geq) b_c$$

がなり $T = T^*$ ば、

$$(9) \quad g_c - g_c(t) \leq (\geq) Q^*(t) b_c$$

$$(10) \quad g_t - g_t(t) \leq (\geq) (I - S(t))^{-1} (a_t + R(t) Q^*(t) b_c)$$

である。こゝで I は単位行列を表わす。

(証明) $\Delta g = g - g(t)$, $\Delta v = v - v(t)$ とおけば、(1), (2) 式より

$$(11) \quad \Delta g = \psi + P(t) \Delta g$$

$$(12) \quad \Delta g + \Delta v = \gamma + P(t) \Delta v$$

である。まず、(8), (11) 式より $\Delta g_c \leq (\geq) Q^*(t) \Delta g_c$ である。一方、(8), (12) 式より

$$\Delta g_c + \Delta v_c \leq (\geq) b_c + Q(t) \Delta v_c$$

であり、 $Q^*(t)$ を左からかければ $Q^*(t) Q(t) = Q^*(t)$ であるから、

(9) 式がえられる。また (8), (11) 式より

$$\Delta g_t \leq (\geq) a_t + R(t) \Delta g_c + S(t) \Delta g_t$$

であり、 $S(t)$ は過渡行列であるから (9) 式より (10) 式がえられる。

定理 1 において、 $g = g^*$, $v = v^*$, $a_t = \varepsilon_1 e_t$, $b_c = \varepsilon_2 e_c$ とおけば、

$(I - S(t))^{-1} R(t) e_c = e_t$ であるから次の系がえられる。

[系 1]

(7) 式で与えられる $P(t)$ にたいして

$$(13) \quad g_c^* \leq Q(t) g_c^*, \quad g_t^* - (R(t) g_c^* + S(t) g_t^*) \leq \varepsilon_1 e_t$$

$$g_c^* + v_c^* - r_c(t) - Q(t) v_c^* \leq \varepsilon_2 e_c$$

がなりたてば

$$(14) \quad g_c^* - g_c(f) \leq \varepsilon_2 e_c, \quad g_t^* - g_t(f) \leq \varepsilon_2 e_t + \varepsilon_1 (I - S(f))^{-1} e_t$$

である。

Whittle (1983, 定理 4.1 (b), p. 122) は最小化問題にたいして、系 1 と同様 f が ε -最適となる条件を示しているが、証明中にミスがあり、示された条件は正しくない。

定理 1 の () に対応する式において $f = f^*$ とおけば、系 1 と同様にして次の系がえられる。

[系 2]

$P(f^*)$ が (7) 式の形で与えられるとき、

$$(15) \quad g_c \geq Q(f^*) g_c, \quad g_t - (R(f^*) g_c + S(f^*) g_t) \geq -\varepsilon_1 e_t$$

$$g_c + v_c - r_c(f^*) - Q(f^*) v_c \geq -\varepsilon_2 e_c$$

がなりたてば

$$(16) \quad g_c^* - g_c \leq \varepsilon_2 e_c, \quad g_t^* - g_t \leq \varepsilon_2 e_t + \varepsilon_1 (I - S(f^*))^{-1} e_t$$

である。

3. 最適停止問題

Bather (1973), Federgrun and Schweitzer (1984), Schweitzer (1984) に従い、まず状態空間 I を communicating sets に分解する。

(17) Unique Chain Decomposition :

1. $I_1 = I$, $D_i = K_i$ ($i \in I_1$), $l=1$ とおく。

2. 各 $i \in I_l$ において D_i の全ての決定を正の確率でとる政策

にたいする I_l の全ての部分連鎖 $C(r;l)$ ($r=1, \dots, R_l$) を定める。

3. $T_l = \phi$, $J = I_l - \sum_r C(r;l)$ とおく。 $i \in J$ にたいして $\sum_{j \in J} p_{ij}(k) < 1$ となる k を D_i からとり除き, $D_i = \phi$ とすれば i を J から T_l に移す。この操作を $D_i (i \in J)$ が変化しなくなるまで反復する。

4. $J = \phi$ とすれば $L = l$ とおいて 5.へ。さもなければ,

$I_{l+1} = J$, $l = l+1$ とおいて 2.へ。

5. $R = \sum_{l=1}^L R_l$, $T = \sum_{l=1}^L T_l$ とおく, $l=1, \dots, L$, $r=1, \dots, R_l$ にたいして $C(\sum_{m=1}^{l-1} R_m + r) = C(r;l)$ とおく。

上記アルゴリズムからえられる各 $C(r)$ ($r \in R = \{1, \dots, R\}$) は, $D_i (i \in C(r))$ からの政策にたいして communicating set をなす。定常政策 $f(r) = (f_i)$ ($f_i \in D_i$, $i \in C(r)$) の集合を $F(r)$ とおけば, 部分マルコフ決定過程 $(C(r), r, P, F(r))$ ($r \in R$) がえられる。 $(C(r), r, P, F(r))$ は状態に依存しない最大利得 $\sigma^*(r)$ をもち, 相対値 $u_i^*(r)$ ($i \in C(r)$) は最適方程式

$$(18) \quad \sigma^*(r) + u_i^*(r) = \max_{k \in D_i} \{ r_i(k) + \sum_{j \in C(r)} p_{ij}(k) u_j^*(r) \} \quad (i \in C(r))$$

の解である。

(18) 式をみたす $\sigma^*(r)$, $u_i^*(r)$ および最適政策 $f^*(r)$ が $r \in R$ についてえられたものとある。このとき $i \in C(r)$ にたいして $\tilde{K}_i = K_i - D_i$ とおけば (3) 式は

$$(19) \quad g_i^* = \max \{ \sigma^*(r), \max \{ \sum_{j \in \tilde{K}_i} p_{ij}(k) g_j^* \mid k \in \tilde{K}_i \} \}$$

となる。アルゴリズム (17) により $r \leq R_1$ については $\tilde{K}_i = \phi$ ($i \in C(r)$) であるから

$$(20) \quad g_i^* = \sigma^*(r) \quad (i \in C(r), r \leq R_1)$$

である。 $r > R_1$ については $C(r)$ の少くとも 1 つの i で $\tilde{K}_i \neq \phi$ であり、もし

$$(21) \quad \sigma^*(r) > \max_{i \in C(r)} \max_{k \in \tilde{K}_i} \left\{ \sum_{j \in I} p_{ij}(k) g_j^* \right\} = \sum_{j \in I} p_{i^*j}(k^*) g_j^*$$

とすれば $i \in C(r)$ で (20) 式が成立する。一方、

$$(22) \quad \sum_{j \in I} p_{i^*j}(k^*) g_j^* \geq \sigma^*(r)$$

とすれば、状態 i^* で決定 $k^* \in \tilde{K}_{i^*}$ をとり、 i^* 以外の $i \in C(r)$ では i^* を吸収状態とあるような政策 $f(r) \in F(r)$ をとれば全ての $i \in C(r)$ で利得 $\sum_{j \in I} p_{ij}(k^*) g_j^*$ がえられる。ゆえに $i \in C(r)$ ($r \in R$) について

$$(23) \quad g_i^*(r) = g_i^*, \quad \tilde{K}(r) = \{(i, k) ; i \in C(r), k \in \tilde{K}_i\}$$

とあり、 $\tilde{k} = (i, k) \in \tilde{K}(r)$ について

$$(24) \quad \hat{p}_{rs}(\tilde{k}) = \sum_{j \in C(s)} p_{ij}(k) \quad (s \in R), \quad \hat{p}_{rj}(\tilde{k}) = p_{ij}(k) \quad (j \in T)$$

とある。他方 $i \in T$ については $k \in K_i$ について

$$(25) \quad \hat{p}_{is}(k) = \sum_{j \in C(s)} p_{ij}(k) \quad (s \in R), \quad \hat{p}_{ij}(k) = p_{ij}(k) \quad (j \in T)$$

とあることに注意。このとき次の最適停止問題が導かれる。

$$(26) \quad g_i^*(r) = \max \{ \sigma^*(r), \max_{\tilde{k} \in \tilde{K}(r)} \{ \sum_{s=1}^R \hat{p}_{rs}(\tilde{k}) g_s^* + \sum_{j \in T} \hat{p}_{rj}(\tilde{k}) g_j^* \} \} \quad (r \in R)$$

$$g_i^* = \max_{k \in K_i} \{ \sum_{s=1}^R \hat{p}_{is}(k) g_s^* + \sum_{j \in T} \hat{p}_{ij}(k) g_j^* \} \quad (i \in T)$$

ここで $\tilde{K}(r) = \phi$ ($r \leq R_1$) である。各 $r \in R$ において決定“停止”

(以後 "0" で表わす) を選べば、利得 $\sigma^*(r)$ をえて過程は停止する。問題 (26) にたいする定常政策 $\hat{f}$ は、決定 $\hat{f}(r) \in \hat{K}(r) \cup \{0\}$ ($r \in R$), $\hat{f}_i \in K_i$ ($i \in T$) の組 $(\hat{f}(r), \hat{f}_i)$ で与えられる。 $\hat{f}$ にたいする停止領域を $R_s(\hat{f}) = \{r \in R; \hat{f}(r) = 0\} (\supset \{r \in R_1\})$ とおき, $R_c(\hat{f}) = R - R_s(\hat{f})$ とおく。全ての定常政策 $\hat{f}$ の集合を $\hat{F}$, 問題 (26) の最適政策を $\hat{f}^*$ で表わす。

$R+|T|$ 次のベクトル $\hat{g}$ と $r \in R, \hat{g} \in \hat{K}(r), i \in T, k \in K_i$ にたいして次の写像 T, U を

$$(27) \quad T(r; \hat{g})\hat{g} = \sum_{s \in R} \hat{p}_{rs}(\hat{g})\hat{g}(s) + \sum_{j \in T} \hat{p}_{rj}(\hat{g})\hat{g}_j$$

$$T(i; k)\hat{g} = \sum_{s \in R} \hat{p}_{is}(k)\hat{g}(s) + \sum_{j \in T} \hat{p}_{ij}(k)\hat{g}_j$$

$$(28) \quad U(r)\hat{g} = \max \{T(r; \hat{g})\hat{g} \mid \hat{g} \in \hat{K}(r)\}$$

$$U(i)\hat{g} = \max \{T(i; k)\hat{g} \mid k \in K_i\}$$

で定義する。政策 $\hat{f}$ のもとでの利得を $\hat{g}(\hat{f}) = (\hat{g}(r; \hat{f}), \hat{g}(i; \hat{f}))$ で表わせば、 $\hat{g}(\hat{f})$ は

$$(29) \quad \hat{g}(r; \hat{f}) = \sigma^*(r) \quad (r \in R_s(\hat{f}))$$

$$(30) \quad \hat{g}(r; \hat{f}) = T(r; \hat{f}(r))\hat{g}(\hat{f}) \quad (r \in R_c(\hat{f}))$$

$$\hat{g}(i; \hat{f}) = T(i; \hat{f}_i)\hat{g}(\hat{f}) \quad (i \in T)$$

を満たしている。 $r \in R_s(\hat{f})$ にたいしてベクトル $\hat{g}_s(\hat{f}) = (\hat{g}(r; \hat{f}))$,

$\sigma^* = (\sigma^*(r))$ を導入し, $r \in R_c(\hat{f})$ および $i \in T$ にたいしてベクトル

$\hat{g}_{ct}(\hat{f}) = (\hat{g}(r; \hat{f}), \hat{g}(i; \hat{f}))^T$, $T_{ct}(\hat{f})\hat{g} = (T(r; \hat{f}(r))\hat{g}, T(i; \hat{f}_i)\hat{g})^T$ を導

入すれば, (29), (30) 式は各々

$$(29)' \quad \hat{g}_s(\hat{f}) = \sigma_s^*$$

$$(30)' \quad \hat{g}_{ct}(\hat{f}) = T_{ct}(\hat{f}) \hat{g}(\hat{f})$$

と書き直される。

[補題 1]

任意の $\hat{f} \in \hat{F}$ に対して, $R_c(\hat{f})UT$ から $R_c(\hat{f})UT$ への遷移行列

$$(31) \quad \hat{g}(\hat{f}) = \begin{pmatrix} (\hat{p}_{rs}(\hat{f}(r))) & (\hat{p}_{rj}(\hat{f}(r))) \\ (\hat{p}_{is}(\hat{f}_i)) & (\hat{p}_{ij}(\hat{f}_i)) \end{pmatrix}$$

は過渡行列である。 $\hat{g}(\hat{f})$ は $r \in R_s(\hat{f})$ に対して (29) 式, $r \in R_c(\hat{f})$, $i \in T$ に対して

$$(32) \quad \hat{g}_{ct}(\hat{f}) = (I - \hat{g}(\hat{f}))^{-1} \hat{f}(\hat{f})$$

で与えられる。ここで $\hat{f}(\hat{f}) = (\sum_{s \in R_s(\hat{f})} \hat{p}_{rs}(\hat{f}(r)) \sigma^*(s), \sum_{s \in R_s(\hat{f})} \hat{p}_{is}(\hat{f}_i) \sigma^*(s))^T$ である。

(証明) $R_c(\hat{f})UT$ が $\hat{g}(\hat{f})$ のもとで閉じた状態集合 $\hat{I}$ を含むと仮定する。 $\hat{I}$ に属する状態のうちで最小のレベルをもつものをとれば, 必ず $R_s(\hat{f})$ へ遷移する正の確率をもち, 矛盾である。

(32) 式は (30) 式から明らかである。

[定理 2]

(i) 問題 (26) は一意解 $\bar{g}^* = \hat{g}(\hat{f}^*) = \max \{ \hat{g}(\hat{f}) \mid \hat{f} \in \hat{F} \}$ をもつ。

(ii) $\{ \bar{\sigma}(r) : r \in R \}$ が $\bar{\sigma}(r) \geq \sigma^*(r) - \varepsilon_2 \quad (r \in R)$ をみたすものとする, このとき

$$(33) \quad \bar{g}^*(r) = \max \{ \bar{\sigma}(r), U(r) \bar{g}^* \} \quad (r \in R)$$

$$\bar{g}_i^* = U(i) \bar{g}^* \quad (i \in T)$$

の一急解 $\bar{g}^* = \hat{g}(\hat{f}^*)$ は $\bar{g}^* - \hat{g}^* \leq \varepsilon_2 e$ を満たす。

(証明) (i) $\hat{f}^*$ は (26) 式右辺を最大化する政策であり、

$$\hat{g}_s^* = \sigma_s^* \quad (r \in R_s(\hat{f}^*)), \quad \hat{g}_{ct}^* = T_{ct}(\hat{f}^*) \hat{g}^* \quad (r \in R_{ct}(\hat{f}^*), i \in T)$$

を満たし、(29), (30) 式より $\hat{g}^* = \hat{g}(\hat{f}^*) \leq \max_{\hat{f} \in \hat{F}} \hat{g}(\hat{f})$ である。一方、

$$\max_{\hat{f} \in \hat{F}} \hat{g}(\hat{f}) = \hat{g}(\hat{f}') \quad \text{と仮定すれば} \quad (26) \text{ 式より}$$

$$(34) \quad \hat{g}_s^* \geq \sigma_s^* = \hat{g}_s(\hat{f}') \quad (r \in R_s(\hat{f}')), \quad \hat{g}_{ct}^* \geq T_{ct}(\hat{f}') \hat{g}^* \quad (r \in R_{ct}(\hat{f}'), i \in T)$$

が成り立つ。(34) 式より $r \in R_{ct}(\hat{f}'), i \in T$ にたいして

$$(35) \quad \hat{g}_{ct}^* \geq \hat{r}(\hat{f}') + \hat{s}(\hat{f}') \hat{g}_{ct}^*$$

が成り立ち、補題 1 より $\hat{g}^* \geq \hat{g}(\hat{f}')$ がえられる。

(ii) 停止利得 $\{\bar{\sigma}(r); r \in R\}$ をもつ $\hat{f}$ の利得を $\bar{g}(\hat{f})$ で表わせば、

(29) 式より

$$\bar{g}_s(\hat{f}) = \bar{\sigma}_s = (\bar{\sigma}(r)) \quad (r \in R_s(\hat{f}))$$

であり、 $r \in R_{ct}(\hat{f}), i \in T$ にたいしては (32) 式より

$$\bar{g}_{ct}(\hat{f}) = (I - \hat{s}(\hat{f}))^{-1} \bar{r}(\hat{f})$$

である。こゝで $\bar{r}(\hat{f}) = (\sum_{\hat{s} \in R_{ct}(\hat{f})} \hat{p}_{rs}(\hat{f}(r)) \bar{\sigma}(s), \sum_{\hat{s} \in R_s(\hat{f})} \hat{p}_{rs}(\hat{f}_i) \bar{\sigma}(s))^T$ である。

ゆえに $r \in R_{ct}(\hat{f}^*)$ にたいして

$$\bar{g}_s^* - \hat{g}_s^* \geq \bar{\sigma}_s - \sigma_s^* \geq -\varepsilon_2 e$$

であり、 $r \in R_{ct}(\hat{f}^*), i \in T$ にたいしては補題 1 より

$$\begin{aligned} \bar{g}_{ct}^* - \hat{g}_{ct}^* &\geq \bar{g}_{ct}(\hat{f}^*) - \hat{g}_{ct}(\hat{f}^*) = (I - \hat{s}(\hat{f}^*))^{-1} (\bar{r}(\hat{f}^*) - \hat{r}(\hat{f}^*)) \\ &\geq (I - \hat{s}(\hat{f}^*))^{-1} \{-\varepsilon_2 (I - \hat{s}(\hat{f}^*)) e\} = -\varepsilon_2 e \end{aligned}$$

が成立する。

4. 逐次近似法

部分マルコフ決定過程は、Platzman (1977) 等により提案された逐次近似法を用いれば、任意の $\varepsilon_2 > 0$ に対して $\bar{\sigma}(r) \geq \sigma^*(r) - \varepsilon_2$ とする ε_2 -最適政策 $\bar{f}(r)$ を有限回の反復で求めることができる。このようにしてえられた $\{\bar{\sigma}(r) : r \in R\}$ を用いた (33) 式を解く逐次近似法として、 $n=0, 1, \dots$ に対して

$$(36) \quad \bar{g}^{n+1}(r) = \max \{ \bar{\sigma}(r), U(r) \bar{g}^n \} \quad (r \in R)$$

$$\bar{g}_i^{n+1} = U(i) \bar{g}^n \quad (i \in T)$$

を考える。ただし初期値として $\bar{g}^0(r) = \bar{\sigma}(r) \quad (r \in R)$, $\bar{g}_i^0 = \min_r \bar{\sigma}(r) \quad (i \in T)$ をとることとする。(36)式右辺の最大を与える政策を $\bar{f}^{n+1}$ とおき, $R_s^{n+1} = R_s(\bar{f}^{n+1})$, $R_c^{n+1} = R_c(\bar{f}^{n+1})$ とおく。 $r \in R_s^{n+1}$ に対しては $\bar{g}_s^{n+1} = \bar{\sigma}_s$ である。

[定理 3]

逐次近似法 (36) は (33) 式の解へ単調に収束する。すなわち、 $n=0, 1, 2, \dots$ に対して

$$(37) \quad \bar{g}^n \leq \bar{g}^{n+1}, \quad \lim_{n \rightarrow \infty} \bar{g}^n = \bar{g}^*$$

$$(38) \quad R_s^n \supset R_s^{n+1}, \quad \lim_{n \rightarrow \infty} R_s^n = R_s(\bar{f}^*)$$

がなりたつ。

(証明) $\bar{\sigma}_{\max} = \max_{r \in R} \bar{\sigma}(r)$ とおけば、 $\bar{\sigma}_{\max} \varepsilon \geq \bar{g}^1 \geq \bar{g}^0$ である。

いま $\bar{\sigma}_{\max} \varepsilon \geq \bar{g}^n \geq \bar{g}^{n-1}$ がなりたつものとすれば、

$$(39) \quad \bar{\sigma}_{\max} \geq U(r) \bar{g}^n \geq U(r) \bar{g}^{n-1} \quad (r \in R_c^{n-1})$$

$$\bar{\sigma}_{\max} \geq U(i) \bar{g}^n \geq U(i) \bar{g}^{n-1} \quad (i \in T)$$

である。ゆえに (36) 式より

$$\bar{\sigma}_{\max} e \geq \bar{g}^{n+1} \geq \bar{g}^n$$

であり, $\bar{g}^n$ は単調に収束する。 $\bar{g}^n \rightarrow \bar{g}$ とおけば $\bar{g}$ は (33) 式を満たし, $\bar{g} = \bar{g}^*$ である。(38) 式は (39) 式より明らかである。

[定理 4]

(i) $r \in R_c^{n+1}$, $i \in T$ にたいして $\bar{g}_{ct}^{n+1} - \bar{g}_{ct}^n = a (\geq 0)$ とおけば,

$$(40) \quad \hat{g}_{ct}(\bar{f}^{n+1}) \geq \bar{g}_{ct}^n + \eta e = \bar{g}_{ct}$$

である。ここで $\eta = a_{\min} / \{1 - \min_{r,i} [\hat{S}(\bar{f}^{n+1})e]_{r,i}\}$ である。

(ii) $R_s^{n+1} = R_s(\bar{f}^*)$ がなりたつものとする。 $r \in R_c^{n+1}$, $i \in T$ にたいして $U_{ct} \bar{g} - \bar{g}_{ct} \leq b$ ならば, $b > 0$ であり,

$$(41) \quad \bar{g}_{ct}^* - \bar{g}_{ct} \leq (I - \hat{S}(\bar{f}^*))^{-1} b$$

である。ここで $\bar{g}_s = \bar{\sigma}_s$ であり, $\bar{g}_{ct}$ は (40) 式右辺で与えられる。

(iii) $r \in R_s^{n+1}$ にたいして $U_s \bar{g} < \bar{\sigma}_s - \varepsilon_1 e$, $r \in R_c^{n+1}$, $i \in T$ にたいして $\bar{g}_{ct}^* - \bar{g}_{ct} \leq \varepsilon_1 e$ がなりたてば $R_s^{n+1} = R_s(\bar{f}^*)$ である。

(証明) (i) $r \in R_c^{n+1}$, $i \in T$ にたいして

$$\begin{aligned} \hat{g}_{ct}(\bar{f}^{n+1}) - \bar{g}_{ct}^n &= (F(\bar{f}^{n+1}) + \hat{S}(\bar{f}^{n+1}) \hat{g}_{ct}(\bar{f}^{n+1})) - (F(\bar{f}^{n+1}) + \hat{S}(\bar{f}^{n+1}) \bar{g}_{ct}^n) + a \\ &= \hat{S}(\bar{f}^{n+1}) (\hat{g}_{ct}(\bar{f}^{n+1}) - \bar{g}_{ct}^n) + a \end{aligned}$$

であるから,

$$\bar{f}_{ct}(f^{n+1}) - \bar{f}_{ct}^n = (I - \hat{S}(f^{n+1}))^{-1} a = y$$

である。

$$y = a + \hat{S}(f^{n+1}) y \geq (a_{\min} + y_{\min} \cdot \min_{r,i} \{[\hat{S}(f^{n+1})e]_{r,i}\}) e$$

より (40) 式をうる。

(ii) (33) 式と系 2 から導かれる。

(iii) $r \in R_S^{n+1} \cap R_c(f^*)$ が存在するものとすれば

$$(42) \quad \bar{\sigma}(r) = \bar{g}(r) \leq \bar{g}^*(r) = U(r) \bar{g}^*$$

が成り立つ。仮定より

$$\begin{aligned} U(r) \bar{g}^* &= \sum_{r \in R_S(f^*)} \tilde{P}_{rs}(f^*) \bar{\sigma}(s) + \sum_{s \in R_S^{n+1} \cap R_c(f^*)} \tilde{P}_{rs}(f^*) \bar{g}^*(s) + \sum_{s \in R_c^{n+1}} \tilde{P}_{rs}(f^*) \bar{g}^*(s) \\ &\quad + \sum_{j \in T} \tilde{P}_{rj}(f^*) \bar{g}_j^* \\ &\leq \sum_{r \in R_S} \tilde{P}_{rs} \bar{\sigma}(s) + \sum_{s \in R_S^{n+1} \cap R_c} \tilde{P}_{rs} \bar{g}^*(s) + \sum_{s \in R_c^{n+1}} \tilde{P}_{rs} (\bar{g}(s) + \varepsilon_1) + \sum_{j \in T} \tilde{P}_{rj} (\bar{g}_j + \varepsilon_1) \\ &< \bar{\sigma}(r) + \sum_{s \in R_S^{n+1} \cap R_c} \tilde{P}_{rs} (\bar{g}^*(s) - \bar{\sigma}(s)) - \varepsilon_1 \sum_{s \in R_S^{n+1}} \tilde{P}_{rs} \end{aligned}$$

であり, $C = \max_{r \in R_S^{n+1} \cap R_c} \bar{g}^*(r) - \bar{\sigma}(r)$ とおけば $C(1 - \sum_{s \in R_S^{n+1} \cap R_c} \tilde{P}_{rs}) < -\varepsilon_1 \sum_{s \in R_S^{n+1}} \tilde{P}_{rs}(f^*)$

と成り立つのである。

(41) 式の評価には $(I - \hat{S}(f^*))^{-1} b$ の上限 (y とおく) が必要である。

$$(I - \hat{S}(f^*))^{-1} b = \sum_{m=0}^{\infty} \hat{S}(f^*)^m b \leq \max_{f \in F} \sum_{m=0}^{\infty} \hat{S}(f)^m b$$

であるから、上限 y は次のマルコフ決定過程に帰着される。

$$(43) \quad y = \max_{f \in F} \{ b + \hat{S}(f) y \}.$$

この問題を解く逐次近似法として $n=0, 1, \dots$ にたいして

$$(44) \quad y^{n+1} = \max_f \{ b + \hat{S}(f) y^n \}, \quad y^0 = b$$

$$(45) \quad \pi^{n+1} = \max_f \tilde{g}(f) \pi^n, \quad \pi^0 = e$$

を考える。補題 1 より $N \leq |R^{n+1}| + |T|$ をみたすある N で

$$(46) \quad \bar{\epsilon} = \max_{r,i} t_{r,i}^N < 1$$

となり, $\bar{y} = \max_{r,i} y_{r,i}^{N-1}$ とおけば

$$(47) \quad y \leq \frac{\bar{y}}{1-\bar{\epsilon}} e$$

がなりたつ。

以上の結果から多重連鎖マルコフ決定過程 (I, r, p, F) の ε 最適政策をもとめる次の逐次近似法がえられる。

1. アルゴリズム (17) により $C(r)$ ($r \in R$), D_i ($i \in C(r)$), T をとめる。

2. 各 $r \in R$ について部分マルコフ決定過程 $(C(r), r, p, F(r))$ を逐次近似法 (たとえば Platzman (1977)) により解き、利得 $\bar{\sigma}(r)$, ε_2 最適政策 $\bar{f}(r)$ を求める。

3. $\bar{g}^0(r) = \bar{\sigma}(r)$ ($r \in R$), $\bar{g}_i^0 = \min_r \bar{\sigma}(r)$ ($i \in T$), $n=0$ とおく。

4. (36) 式から $\bar{g}^{n+1}$ を計算し, $r \in R^{n+1}$, $i \in T$ について

$$\bar{g}_{ct}^{n+1} - \bar{g}_{ct}^n = a \leq \delta e$$

となれば 5.へ。さもなければ $n=n+1$ として (36) 式の計算にもどる。

5. (40) 式右辺の $\bar{g}_{ct}$ を計算し, $\varepsilon_1 = \varepsilon - \varepsilon_2 (> 0)$ について

$$U(r) \bar{g} < \bar{\sigma}(r) + \varepsilon_1 \quad (r \in R_s^{n+1})$$

となれば 6.へ。さもなければ $n=n+1$, $\delta = \alpha \delta$ ($0 < \alpha < 1$)

といて 4. $\wedge$.

6. (44) ~ (47) より $\bar{\epsilon}$, $\bar{\gamma}$ を求め,

$$\frac{\bar{\gamma}}{1-\bar{\epsilon}} \leq \varepsilon_1$$

とすれば 7. $\wedge$. もしなければ $n=n+1$, $\delta=\alpha\delta$ といて 4.

$\wedge$.

7. $r \in R_s^{n+1}$ にたいしては $\bar{f}_i = \bar{f}_i(r)$ ($i \in C(r)$), $i \in T$ にたいして

は $\bar{f}_i = \bar{f}_i^{n+1}$ とおく。各 $r \in R_c^{n+1}$ にたいして, $\bar{f}^{n+1}(r) =$

(i^*, k^*) ならば $\bar{f}_{i^*} = k^*$ とおき, $i (\neq i^*) \in C(r)$ にたいして

は i^* を吸収状態とする政策 $\bar{f}(r) \in F(r)$ を求め, $\bar{f}_i = \bar{f}_i(r)$

とおく。政策 $\bar{f}$ は ε -最適政策である。

参考文献

Bather, J. (1973), "Optimal Decision Procedures for Finite Markov chains, III," Adv. Appl. Prob. 5, 541-553.

Federgrun, A. and P.J. Schweitzer (1984), "A Fixed Point Approach to Undiscounted Markov Renewal Programs," SIAM J. Alg. Dis. Math. 5, 539-550.

Platzman, L. (1977), "Improved Conditions for Convergence in Undiscounted Markov Renewal Programs," Opns. Res. 25, 529-533.

Schweitzer, P.J. (1984), "A Value-Iteration Scheme for Undiscounted Multichain Markov Renewal Programs," Zeit. Opns. Res. 28, 143-152.

Whittle, P. (1983), Optimization over Time, Vol.2, John Wiley, Chichester.