数理解析研究所講究録
611巻 1987年 54-68

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るにとどまっている。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。

$I = \{1, 2, \cdots, M\}$：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 $f = (f_1, \cdots, f_M)$ の集合 $(f_i \in K_i, i \in I)$

$r(f) = (r_1(f_1), \cdots, r_M(f_M))^T$ （T は転置を表わす）

$P(f) = (p_{ij}(f_i))$：f のもとでの遷移確率行列

$P^*(f)$：$P(f)$ のチェザロ和 $\left(\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} P(f)^n\right)$

$g(f) = P^*(f) r(f)$：f の利得 (gain)

$v(f)$：f の相対値

$g(f)$ は

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

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

の一意解であり，$v(f)$ は $P(f)$ の各エルゴード連鎖に属する 1

つの状態 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, V, 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 $v^* = v(f^*)$ のみたす最適方程式は

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

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

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 f は，$e=(1,\cdots,1)^T$ にたいして

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

をみたすとき，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) $$w = g - P(f)g \ , \quad \gamma = g + v - r(f) - P(f)v$$

とおく。$P(f)$ が

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

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) $$w_c \leqq (\geqq)\ 0 \ , \quad w_t \leqq (\geqq)\ \alpha_t \ , \quad \gamma_c \leqq (\geqq)\ b_c$$

がなりたてば.

(9) $g_c - g_c(f) \leqq (\geqq)\ Q^*(f) b_c$

(10) $g_t - g_t(f) \leqq (\geqq)\ (I - S(f))^{-1}(a_t + R(f)Q^*(f)b_c)$

である。ここで I は単位行列を表わす。

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

(11) $\Delta g = \psi + P(f)\Delta g$

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

である。まず, (8), (11) 式より $\Delta g_c \leqq (\geqq) Q^*(f)\Delta g_c$ である。一方, (8), (12) 式より

$$\Delta g_c + \Delta v_c \leqq (\geqq)\ b_c + Q(f)\Delta v_c$$

であり, $Q^*(f)$ を左からかければ $Q^*(f)Q(f) = Q^*(f)$ であるから, (9) 式が得られる。また (8), (11) 式より

$$\Delta g_t \leqq (\geqq)\ a_t + R(f)\Delta g_c + S(f)\Delta g_t$$

であり, $S(f)$ は過渡行列であるから (9) 式より (10) 式が得られる。

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, $(I - S(f))^{-1}R(f)e_c = e_t$ であるから次の系が得られる。

[系 1]

(7) 式で与えられる $P(f)$ にたいして

(13) $g_c^* \leqq Q(f) g_c^*$, $g_t^* - (R(f)g_c^* + S(f)g_t^*) \leqq \varepsilon_1 e_t$

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

がなりたてば

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

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

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

がなりたてば

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

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 へ分解する。

(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 $C(r;\ell)$ $(r=1,\cdots,R_\ell)$ を定める。

3. $T_\ell=\phi$, $J=I_\ell-\sum_r C(r;\ell)$ 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，$D_i=\phi$ となれば 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. $J=\phi$ となれば $L=\ell$ とおいて 5. へ。さもなければ，$I_{\ell+1}=J$, $\ell=\ell+1$ とおいて 2. へ。

5. $R=\sum_{\ell=1}^{L}R_\ell$, $T=\sum_{\ell=1}^{L}T_\ell$ とおき，$\ell=1,\cdots,L$, $r=1,\cdots,R_\ell$ にたいして $C(\sum_{m=1}^{\ell-1}R_m+r)=C(r;\ell)$ とおく。

上記アルゴリズムから与えられる各 $C(r)$ $(r\in R=\{1,\cdots,R\})$ 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 $(C(r),r,P,F(r))$ $(r\in R)$ が与えられる。$(C(r),r,P,F(r))$ 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) $$\sigma^*(r)+u_i^*(r)=\max_{k\in D_i}\{r_i(k)+\sum_{j\in C(r)}p_{ij}(k)u_j^*(r)\}\quad(i\in C(r))$$

の解である。

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

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

となる。アルゴリズム (17) により $r \leqq R_1$ にたいしては $\tilde{K}_i = \phi$ $(i \in C(r))$ であるから

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

である。$r > R_1$ にたいしては $C(r)$ の少くとも 1 つの i で $\tilde{K}_i \neq \phi$ であり，もし

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

となれば $i \in C(r)$ で (20) 式が成立する。一方，

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

となれば，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^* を吸収状態とするような政策 $f(r) \in F(r)$ をとれば全ての $i \in C(r)$ で利得 $\sum_{j \in I} p_{ij}(k^*) g_j^*$ が得られる。ゆえに $i \in C(r)$ $(r \in R)$ にたいして

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

とおき，$\hat{k} = (i,k) \in \tilde{K}(r)$ にたいして

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

とおく。他方 $i \in T$ にたいしては $k \in K_i$ にたいして

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

とおくことにする。このとき次の最適停止問題が導かれる。

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

$$g_i^* = \max \{ \sum_{s=1}^{R} \hat{p}_{is}(k) g^*(s) + \sum_{j \in T} \hat{p}_{ij}(k) g_j^* \mid k \in K_i \} \qquad (i \in T)$$

ここで $\tilde{K}(r) = \phi$ $(r \leqq R_1)$ 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fは、決定$f(r)\in\hat{K}(r)\cup\{0\}$ $(r\in R)$, $f_i\in K_i$ $(i\in T)$の組$(f(r), f_i)$で与えられる。fにたいする停止領域を$R_s(f)=\{r\in R;\ f(r)=0\}(\supset\{r\leq R_1\})$とおき、$R_c(f)=R-R_s(f)$とおく。全ての定常政策$f$の集合を$\mathcal{F}$、問題(26)の最適政策を$f^*$で表わす。

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

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

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

$$Ug(r)=\max\{T(r;k)g\mid k\in\hat{K}(r)\} \tag{28}$$

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

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

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

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

$$g(i;f)=T(i;f_i)g(f)\quad(i\in T)$$

を満たしている。$r\in R_s(f)$にたいしてベクトル$g_s(f)=(g(r;f))$, $\sigma_s^*=(\sigma^*(r))$を導入し、$r\in R_c(f)$および$i\in T$にたいしてベクトル$g_{ct}(f)=(g(r;f), g(i;f))^T$, $T_{ct}(f)g=(T(r;f(r))g, T(i;f_i)g)^T$を導入すれば、(29), (30)式は各々

(29)′ $g_s(f) = \sigma^*_s$

(30)′ $g_{ct}(f) = T_{ct}(f)\, g(f)$

と書き直される。

[補題1]

任意の $f \in F$ について，$R_c(f)UT$ から $R_c(f)UT$ への遷移行列

(31) $$S(f) = \begin{pmatrix} (p_{rs}(f(r))) & (p_{rj}(f(r))) \\ (p_{is}(f_i)) & (p_{ij}(f_i)) \end{pmatrix}$$

は過渡行列である。$g(f)$ は $r \in R_s(f)$ について (29) 式，$r \in R_c(f)$，$i \in T$ について

(32) $$g_{ct}(f) = (I - S(f))^{-1} r(f)$$

で与えられる。ここで $r(f) = \left(\sum_{s \in R_s(f)} p_{rs}(f(r))\sigma^*(s), \sum_{s \in R_s(f)} p_{is}(f_i)\sigma^*(s)\right)^T$ である。

(証明) $R_c(f)UT$ が $g(f)$ のもとで閉じた状態集合 $\bar{T}$ を含むと仮定する。$\bar{T}$ 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 $R_s(f)$ 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 $g^* = g(f^*) = \max\{g(f) \mid f \in F\}$ をもつ。

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

(33) $$g^*(r) = \max\{\pi(r),\ v(r)g^*\} \quad (r \in R)$$
$$g^*_i = v(i)g^* \quad (i \in T)$$

の一意解 $\bar{g}^* = \bar{g}(\hat{f}^*)$ は $\bar{g}^* - \bar{g}^* \leqq \varepsilon_2 e$ をみたす。

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

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

をみたし，(29), (30)式より $\bar{g}^* = \bar{g}(\hat{f}^*) \leqq \max_{\hat{f} \in \hat{F}} \bar{g}(\hat{f})$ である。一方，

$\max_{\hat{f} \in \hat{F}} \bar{g}(\hat{f}) = \bar{g}(\hat{f}')$ とおけば (26)式より

(34) $\bar{g}^*_s \geqq \bar{\sigma}^*_s = \bar{g}_s(\hat{f}')$ $(r \in R_s(\hat{f}'))$, $\bar{g}^*_{ct} \geqq T_{ct}(\hat{f}')\bar{g}^*$ $(r \in R_c(\hat{f}'), i \in T)$

がなりたつ。(34)式より $r \in R_c(\hat{f}')$, $i \in T$ にたいして

(35) $\bar{g}^*_{ct} \geqq \bar{r}(\hat{f}') + \hat{S}(\hat{f}')\bar{g}^*_{ct}$

がなりたち，補題1より $\bar{g}^* \geqq \bar{g}(\hat{f}')$ 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，$r \in R_c(\hat{f})$, $i \in T$ 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。ここで $\bar{r}(\hat{f}) = \left(\sum_{s \in R_s(\hat{f})} \hat{P}_{rs}(\hat{f}(r))\bar{\sigma}(s), \sum_{s \in R_s(\hat{f})} \hat{P}_{is}(\hat{f}_i)\bar{\sigma}(s)\right)^T$ である。

ゆえに $r \in R_s(\hat{f}^*)$ にたいして

$$\bar{g}^*_s - \bar{g}^*_s \geqq \bar{\sigma}_s - \sigma^*_s \geqq -\varepsilon_2 e$$

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

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

$$\geqq (I - \hat{S}(\hat{f}^*))^{-1}\{-\varepsilon_2(I - \hat{S}(\hat{f}^*))e\} = -\varepsilon_2 e$$

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 $\varepsilon_2 > 0$ にたいして $\bar{f}(r) \geq f^*(r) - \varepsilon_2$ となる 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。このようにして得られた $\{\bar{f}(r); r \in R\}$ を用いた (33) 式を解く逐次近似法として，$n = 0, 1, \cdots$ にたいして

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

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

を考える。ただし初期値として $\bar{g}^0(r) = \bar{f}(r)$ $(r \in R)$, $\bar{g}_i^0 = \min_r \bar{f}(r)$ $(i \in T)$ をとることにする。(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。$r \in R_S^{n+1}$ にたいしては $\bar{g}_S^{n+1} = \bar{f}_S$ 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，$n = 0, 1, 2, \cdots$ にたいして

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

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

がなりたつ。

(証明) $\bar{f}_{max} = \max_{r\in R} \bar{f}(r)$ とおけば，$\bar{f}_{max} e \geq \bar{g}^1 \geq \bar{g}^0$ である。いま $\bar{f}_{max} e \geq \bar{g}^n \geq \bar{g}^{n-1}$ がなりたつものとすれば，

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

$$\bar{v}_{max} \geq v(i)\bar{g}^{n} \geq v(i)\bar{g}^{n-1} \qquad (i \in T)$$

である。ゆえに(36)式より

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

であり，$\bar{g}^{n}$ は単調に収束する。$\bar{g}^{n} \to \bar{g}$ とすれば $\bar{g}$ は(33)式をみたし，$\bar{g} = \bar{g}^{*}$ 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) $\quad \bar{g}_{ct}(\bar{f}^{n+1}) \geq \bar{g}_{ct}^{n} + \eta\mathbf{e} = \bar{g}_{ct}$

である。ここで $\eta = a_{min} / \{1 - \min_{r,i}[\hat{s}(\bar{f}^{n+1})\mathbf{e}]_{r,i}\}$ 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 $v_{ct}\bar{g} - \bar{g}_{ct} \leqq b$ ならば，$b > 0$ であり，

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

である。ここで $\bar{g}_s = \bar{v}_s$ 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 $v_s\bar{g} < \bar{v}_s - \varepsilon_1\mathbf{e}$，$r \in R_c^{n+1}$，$i \in T$ にたいして $\bar{g}_{ct}^{*} - \bar{g}_{ct} \leqq \varepsilon_1\mathbf{e}$ 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

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

$$= \hat{s}(\bar{f}^{n+1})(\bar{g}_{ct}(\bar{f}^{n+1}) - \bar{g}_{ct}^{n}) + a$$

であるから，

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

である。

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

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) $\quad \bar{v}(r) = \bar{g}(r) \leq g^*(r) = v(r) g^*$

がなりたつ。仮定より

$$\begin{aligned} v(r) g^* &= \sum_{r \in R_s(f^*)} \hat{p}_{rs}(f^*) \bar{v}(s) + \sum_{s \in R_s^{n+1} \cap R_c(f^*)} \hat{p}_{rs}(f^*) g^*(s) + \sum_{s \in R_c^{n+1}} \hat{p}_{rs}(f^*) g^*(s) \\ &\quad + \sum_{j \in T} \hat{p}_{rj}(f^*) g^*_j \\ &\leqq \sum_{r \in R_s} \hat{p}_{rs} \bar{v}(s) + \sum_{s \in R_s^{n+1} \cap R_c} \hat{p}_{rs} g^*(s) + \sum_{s \in R_c^{n+1}} \hat{p}_{rs}(g(s) + \varepsilon_1) + \sum_{j \in T} \hat{p}_{rj}(g_j + \varepsilon_1) \\ &< \bar{v}(r) + \sum_{s \in R_s^{n+1} \cap R_c} \hat{p}_{rs}(g^*(s) - \bar{v}(s)) \qquad - \varepsilon_1 \sum_{s \in R_s^{n+1}} \hat{p}_{rs}) \end{aligned}$$

であり，$c = \max_{r \in R_s^{n+1} \cap R_c} g^*(r) - \bar{v}(r)$ とおけば $c(1 - \sum_{s \in R_s^{n+1} \cap R_c} \hat{p}_{rs}) < -\varepsilon_1 \sum_{s \in R_s^{n+1}} \hat{p}_{rs}(f^*)$

となり矛盾である。

(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。

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

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 $n = 0, 1, \cdots$ にたいして

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

(45) $\mathbf{t}^{n+1} = \max_{f} \tilde{S}(f)\mathbf{t}^n \quad , \mathbf{t}^0 = \mathbf{e}$

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

(46) $\bar{t} = \max_{r,i} t^N_{r,i} < 1$

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

(47) $\mathbf{g} \leq \frac{\bar{g}}{1-\bar{t}}\mathbf{e}$

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 をもとめる。
2. 各 $r \in R$ にたいする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 $\bar{f}(r)$ をもとめる。
3. $g^0(r) = \bar{\sigma}(r)$ $(r \in R)$, $g^0_i = \min_r \bar{\sigma}(r)$ $(i \in T)$, $n=0$ とおく。
4. (36) 式から g^{n+1} を計算し，$r \in R^{n+1}_c$, $i \in T$ にたいして

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

   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 g_{ct} を計算し，$\varepsilon_1 = \varepsilon - \varepsilon_2 (>0)$ にたいして

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

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

として 4. へ。

6. (44)～(47)式より $\bar{t}$, $\bar{g}$ をもとめ,

$$\frac{\bar{g}}{1-\bar{t}} \leqq \varepsilon_1$$

となれば 7. へ。さもなければ $n=n+1$, $\delta=\alpha\delta$ として 4. へ。

7. $r \in R_s^{n+1}$ にたいしては $f_i = \bar{f}_i(r)$ $(i \in C(r))$, $i \in T$ にたいしては $f_i = \bar{f}_i^{n+1}$ 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 $f_{i^*} = k^*$ 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 $f(r) \in F(r)$ を求め, $f_i = \bar{f}_i(r)$ とおく。政策 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 Multi-chain Markov Renewal Programs," Zeit. Opns. Res. 28, 143-152.

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