数理解析研究所講究録 1207 巻 2001 年 147-156

偏差行列とマルコフ決定過程の精密割引最適化基準へのその応用

和歌山大学　門田良信 (Yoshinobu Kadota)

概要: 可算状態空間と有界利得系をもつマルコフ決定過程に対して, マルコフ連鎖に関するある再帰条件を与える. そして, 定常政策に対する偏差行列の存在, ローラン展開の係数の連続性および Blackwell-最適定常政策の存在を導く. その条件のもとでは, マルコフ連鎖が周期的再帰クラスを可算個数もつことができる. また偏差行列の正確な表現を導く.

1. 導入

有限状態空間をもつマルコフ決定過程(以後 MDP と略す) に対して, Miller and Veinott[13] および Veinott[14] は割引総期待利得のローラン展開を使って Blackwell-最適定常政策を導いている. その後, 可算状態空間をもつ場合の研究が進み, Dietz and Nollau[7], Kadota[10], Mann[12] およびDekker and Hordijk[5, 6] を含む多くの結果が得られている. Yushkevich[15] は一般の有限状態空間をもつ場合にもこの問題を考えている.

それらの多くは定常政策に対応したマルコフ連鎖(以後 MC と略す) に関する再帰条件を与え, 有限個数の再帰クラスをもつ場合を研究している. [5] は可算個の非周期的クラスをもつ場合を考えているが, 彼等 [5, 6] は作用素論的接近を試みており MC の再帰的構造に関して十分な議論をしていない.

ここでは可算状態空間と有界利得系をもつ MDP に対してある再帰的条件(条件 (I)) を仮定して, 定常政策に対応した MC に対する偏差行列の存在, ローラン展開の係数の連続性および Blackwell-最適定常政策の存在を導く. 解析方法としては MC のエルゴード分解を用いる.

この報告の特徴は (i) 条件 (I) によってマルコフ連鎖が周期的再帰クラスを可算個数もつ場合を考察していること, (ii) 偏差行列の具体的な表現が得られることにある.

第 1, 2 節では基本的な補助定理を与え, 主要な結果は第 3, 4 節で導く.

標準的な MDP を (S, A, p, r) で表す. ここで, $S = \{1, 2, \cdots\}$ は可算状態空間, $A(i)$ は各状態 $i \in S$ に対応する距離空間 A の部分集合, $p = (p(a)_{ij})$ は推移確率で任意の

$i \in S$, $a \in A(i)$ に対して $\sum_{j \in S} p(a)_{ij} = 1$ を満たし, $r(i,a)$ は $\{(i,a); i \in S, a \in A(i)\}$ 上で有界な($|r(i,a)| \leq M$) 直接利得とする.

時刻 n での決定は, 履歴 $(i_0, a_0, \cdots, i_n) \in (S \times A)^n \times S$ に対して $\pi_n(A(i_n) | i_0, a_0, i_1, \cdots, i_n) = 1$ を満たす条件付確率 π_n であり, 政策 $\pi = (\pi_0, \pi_1, \cdots)$ は決定の列とする. いま, S 上の関数 f があって任意の $i \in S$ と n について $f(i) \in A(i)$ かつ $\pi_n(\{f(i)\} | i_n = i) = 1$ となるとき, π を定常政策とよび $\pi = f$ と表す. 定常政策の全体を $\mathcal{F}$ で表す.

標本空間を $\Omega = (S \times A)^\infty$ とする. 確率変数 X_n, Δ_n は, それぞれ時刻 n での状態と決定を表すものとする. このとき, 初期状態 $i_0 \in S$ と政策 π によって Ω 上の確率測度 $P_{i_0}^\pi$ が定まり, 任意の $n = 0, 1, \cdots$, $(i_0, a_0, \cdots, i_n) \in (S \times A) \times S$, $a_n \in A(i_n)$, $j \in S$ とボレル集合 $B \subset A(i_n)$ に対して,

$$P_{i_0}^\pi(\Delta_n \in B | i_0, a_0, \cdots, i_n) = \pi_n(B | i_0, a_0, \cdots, i_n) \quad \text{および}$$
$$P_{i_0}^\pi(X_{n+1} = j | i_0, a_0, \cdots, i_n, a_n) = p(a_n)_{i_n j}$$

が成立する. P_i^π による Y の期待値を $E_i^\pi(Y)$ で表す.

$\rho > 0$ とし, 割引率を $\beta = 1/(1+\rho)$ とする. 政策 π を使ったときの 各初期状態に対応した割引総期待利得のベクトルは,

$$V_\rho(\pi) = (E_i^\pi \{ \sum_{n=0}^{\infty} \beta^{n+1} r(X_n, \Delta_n) \} ; i \in S)$$

によって与えられる. 政策 π^* は, 任意の π に対して $V_\rho(\pi^*) \geq V_\rho(\pi)$ ならば ρ-割引最適であるという. (ただし, 不等式は成分毎にとるものとする.) 政策 π^* は, 任意の π に対して

$$\liminf_{\rho \to 0+} \rho^{-n}(V_\rho(\pi^*) - V_\rho(\pi)) \geq 0 \tag{1}$$

ならば n-割引最適であるという. また, 任意の n に対して n-割引最適ならば, Blackwell-最適とよぶ.

推移行列は任意の $f \in \mathcal{F}$ に対して $P(f) = (p(f(i))_{ij})$ によって与えられる. 以下においては記号 (f) は適宜省略して P, p_{ij} 等と表し, また後に定義されるものも P^*, R, T, E_a, $\mathcal{I}$, Q, H 等と表す.

$P^0 = I$ (単位行列) とし, P の n 個の積を $P^n = (p_{ij}^n; i, j \in S)$ とする. MC の平均エルゴード定理により, i, $j \in S$ に対して $p_{ij}^* = \lim_{n \to \infty} (n+1)^{-1} \sum_{k=0}^n p_{ij}^k$ が存在する. $P^* = (p_{ij}^*; i, j \in S)$ とすると, $PP^* = P^*P = P^*$ となる. 部分集合 $E \subset S$ に対して, $P^n(i, E) = \sum_{j \in E} p_{ij}^n$, $P^*(i, E) = \sum_{j \in E} p_{ij}^*$ と表す. いま, 任意の $i \in S$ に対して $P^*(i, S) = 1$ ならば, P を non-dissipative とよぶ.

空間 $\{P(f);\, f\in\mathcal{F}\}$ に次の再帰条件を定義する.

条件 (I). 任意の $i\in S$, $E\subset S$, $n=1,2,\ldots$, $f\in\mathcal{F}$ に対して, 次を満たす定数 B が存在する.

$$|\sum_{k=0}^{n}\{P(f)^k(i,E)-P(f)^*(i,E)\}|\leqq B. \tag{2}$$

$P(f)$ が Doeblin 条件を満たすかあるいは偏差行列をもてば, 各 $f\in\mathcal{F}$ に対して (2) を満たす B は存在する. (前者については [7], Hordijk[9] および [10] を参照. 後者については第 3節の等式 (10) を参照.) $\mathcal{F}$ 上での一様有界性は後の第 4節で偏差行列の連続性を導く. 条件 (I) に関する 2, 3 の性質は Kadota[11] に見られる.

$f\in\mathcal{F}$ に対して, $R(f)$, $T(f)$ をそれぞれ再帰的状態, 過渡的状態の集合とする. $\{E(f)_a; a\in\mathcal{I}(f)\}$ を再帰クラスの族とする. $d(f)_a$ を 各再帰クラス $E(f)_a$ の周期, ${}_aC(f)_0, {}_aC(f)_1,\ldots,{}_aC(f)_{d_a-1}$ をその周期的クラスとする. 次の補助定理は以後の議論において基本的である.

補助定理 1. 条件 (I) を仮定する. 各 $P(f)$ は non-dissipative であり, $d=$ l.c.m.$\{d(f)_a;$ $a\in\mathcal{I}(f), f\in\mathcal{F}\}$ は有限である.

証明. (2) を $n+1$ で割り $n\to\infty$ とすると, $i\in S$ and $E\subset S$ に関して一様に

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}P(f)^k(i,E)=P(f)^*(i,E) \tag{3}$$

が存在する. (3) で $E=S$ とすることにより, $P(f)$ は non-dissipative となる.

最小公倍数が存在しないと仮定すると, $2\leqq d(f_0)_{a_0}<d(f_1)_{a_1}<\ldots<d(f_n)_{a_n}<\ldots$ および $\lim_{n\to\infty}d(f_n)_{a_n}=\infty$ を満たす $\{f_n\}\subset\mathcal{F}$, $a_n\in\mathcal{I}(f_n)$ が存在する. $d_n=d(f_n)_{a_n}$ と表す. $E_n(f_n)$ を周期 d_n と周期的クラス$\{C(f_n)_r;\, r=0,1,\ldots,d_n-1\}$ をもつ再帰クラスとする. k_n を任意の n に対して $k_n+1\leqq d_n/2$ かつ $\lim_{n\to\infty}k_n=\infty$ となる自然数とする. $G_n(f_n)=\cup_{r=0}^{k_n}C(f_n)_r$ とすると, 任意の $i_n\in C(f_n)_0$ に対して $P(f_n)^*(i_n,G_n(f_n))=(k_n+1)/d_n\leqq 1/2$ である. 一方, 任意の $i_n\in C(f_n)_0$ と $r=0,1,\cdots,k_n$ に対して, $P(f_n)^r(i_n,G_n(f_n))=1$ となる. 従って, 任意の n に対して

$$|\sum_{r=0}^{k_n}\{P(f_n)^r(i_n,G_n(f_n))-P(f_n)^*(i_n,G_n(f_n))\}|\geqq(k_n+1)/2$$

. となる. これは (2) に反する. よって示された. □

2. 周期の最小公倍数に関する補助定理

この節では 補助定理 1 から得られる d-段階 MC に関するエルゴード的性質について, 2つの補助定理を準備する.

$f \in \mathcal{F}$ を固定しておく. 1単位時間の推移確率が p_{ij}^d によって与えられる MC を d-段階 MC とよび P^d で表す. p_{ij}^*, $P^*(i,E)$ の定義と同様にして, P^d に対して p_{ij}^{d*}, $P^{d*}(i,E)$ を定義する. $({}_RP^d)_{ij}^n$ を, P^d によって $i \in S$ を出発して途中 R を通ることなく時刻 n に $j \in S$ に到達する確率とする. $E \subset S$ に対して, $({}_RP^d)^n(i,E) = \sum_{j\in E}({}_RP^d)_{ij}^n$, $({}_RP^d)^+(i,E) = \sum_{n=1}^{\infty}({}_RP^d)^n(i,E)$ と表す.

行列 $A = (a_{ij}; i,j \in S)$ に対して, $\| A \| = \sup_{i\in S}\{\sum_{j\in S} | a_{ij} |\}$ とする. $\mathcal{N}(S)$ を $\| A \| < \infty$ となる A のノルム空間とする. このとき $A, B \in \mathcal{N}(S)$ について, $\| AB \| \leq \| A \| \| B \|$ が成り立つ. 以後省略して, $\sum_{j\in S} a_{ij}b(j,E)$ を $AB(i,E)$ で, $A(i,E) - B(i,E)$ を $(A-B)(i,E)$ で表すことがある.

次の補助定理 2の (5) は Chung [4, § I.6, pp32] の定理 4 の系の修正であり, (6) は Doob[8, pp208] (5.14) の修正である.

補助定理 2. MC P は non-dissipative であり, $d = \text{l.c.m.}\{d_a;\ a \in \mathcal{I}\}$ は有限と仮定する. このとき 任意の $i \in S$, $E \subset S$ に対して次の等式が成立する.

$$\sum_{r=0}^{d-1} P^r(P^{nd} - P^*)(i,E) = \sum_{r=0}^{d-1} P^r(P^{nd} - P^{d*})(i,E), \quad n = 0, 1, \cdots. \tag{4}$$

$$P^{d*}(i,E) = \sum_{j\in R}\sum_{n=1}^{\infty}({}_RP^d)_{ij}^n P^{d*}(j, E\cap R). \tag{5}$$

証明. P が non-dissipative より P^d もそうなる. Scheffé の定理(Hordijk[9] の Lemma 4.11 または Billingsley[2, pp224]) を使って $P^{nd+r}(i,E)$ は $E \subset S$ に関して一様に $P^rP^{d*}(i,E)$ に収束する. このことと (3) より,

$$P^*(i,E) = \frac{1}{d}\sum_{r=0}^{d-1} P^rP^{d*}(i,E) \tag{6}$$

が成立し, 従って (4) が成立する.

$i \in T$ と仮定して (5) を示せば十分である. R への到達時刻に関する分解定理により, 各 n について

$$P^{nd}(i,E) = \sum_{j\in R}\sum_{k=1}^{n}({}_RP^d)_{ij}^k P^{(n-k)d}(j, E\cap R) + P^{nd}(i, E\cap T) \tag{7}$$

が成立する. $n \to \infty$ とすると, (7) の右辺の第 1 項について極限と和に関する交換ができる. [4, §I.5, pp20] の Lemma A を $\sum_{k=1}^{\infty}({}_R p^d)_{ij}^k > 0$ となる各 $j \in R$ に適用すれば, (5) が導かれる. □

表記として, $Q_n(i,E) = \sum_{k=0}^{n}\sum_{r=0}^{d-1} P^r(P^{kd} - P^*)(i,E)$ および $Q(i,E) = \lim_{n\to\infty} Q_n(i, E)$ としておく.

[4] の意味での状態の閉集合に対して, 制限された MC's を $P_a = (p_{ij};\, i,j \in E_a)$ や $P_{(a,r)}^d = (p_{ij}^d; i,j \in {}_aC_r)$ と定義する. とくに $i \in E_a$ と $a \in \mathcal{I}$ に対して ${}_aQ(i,E) = Q(i,\, E \cap E_a)$ と表す.

補助定理 3. 条件 (I) を仮定する. 任意の $a \in \mathcal{I}$ に対して, $\| {}_aQ \| \leqq 2B$ を満たす ${}_aQ(i,\, E)$ が存在する.

証明. ${}_aQ(i,\, E)$ の存在をいうために, $\sum_{n=0}^{\infty} \| \sum_{r=0}^{d-1} P_a^r(P_a^{nd} - P_a^*) \| < \infty$ を示す. $\nu = 0, 1, \ldots, d_a - 1$ について $(p_{ij}^{d_a}; i,j \in {}_aC_\nu)$ を考える. このとき, $P_a^{d*} = P_a^{d_a *}$ である. $p_0 = \max\{p_{jj}^{d*};\, j \in {}_aC_\nu\} = p_{kk}^{d*} > 0$ ($k \in {}_aC_\nu$) と表しておく. (4) より, (3) において $n = md_a$, $E = \{k\}$ とおくと, $\varepsilon_0 = p_0/2$ と任意の $i \in {}_aC_\nu$ に対して $| N_0^{-1} \sum_{\ell=1}^{N_0}(p_{ik}^{\ell d_a} - p_{ik}^{d*}) | < \varepsilon_0$ を満たす自然数 N_0 が存在する. 従って, ある $1 \leqq n(i) \leqq N_0$ に対して, $p_{ik}^{n(i)d_a} > \varepsilon_0$ が成立する. 一方, $n \geqq N_1$ ならば $| p_{kk}^{nd_a} - p_{kk}^{d*} | < \varepsilon_0$ となる N_1 が存在する. また $p_{ik}^{d*} = p_{kk}^{d*}$ である. $N_2 = N_0 + N_1$, $\delta = \varepsilon_0^2$ にとる. 任意の $n \geqq N_2$, $i \in {}_\alpha C_\nu$ について, $\delta \leqq p_{ik}^{nd_a}, p_{ik}^{d*} \leqq 1$ だから, $| p_{ik}^{nd_a} - p_{ik}^{d*} | \leqq 1 - \delta$ を得る. この不等式は d_a の代りに d としても成り立つ.

任意の $0 < \varepsilon < 1$ に対して, $(1-\delta)^m < \varepsilon$ となる m をとり, $N_\nu = mN_2$ とおく. Doob[8, pp197] の Case(b) により, $\| P_{(a,\nu)}^{N_\nu d} - P_{(a,\nu)}^{d*} \| < \varepsilon$ となる. $N_a = \max\{N_\nu\}$ とおく. n と $r = 0, 1, \cdots, N_a - 1$ に対して, $\| P_a^{N_a d} - P_a^{d*} \| < \varepsilon$ かつ

$$\| P_a^{(nN_a + r)d} - P_a^{d*} \| \leqq \| P_a^{nN_a d} - P_a^{d*} \| \leqq \| P_a^{N_a d} - P_a^{d*} \|^n \tag{8}$$

を得る. (8) を r と n について加えると, 求める結果を得る. (2) より $\| Q_n(\cdot, \cdot \cap E_a) \| \leqq 2B$ だから $\| {}_aQ \| \leqq 2B$ となって有界性も示される. □

3. 偏差行列の存在

この節では偏差行列の存在およびその表現と性質について考察する.

割引率 $\beta = 1/(1+\rho)$ ($\rho > 0$) に対して,

$$H_\rho(i,E) = \sum_{n=0}^{\infty} \beta^n \{(P^n - P^*)(i,E)\} \tag{9}$$

とおく. いま, $H(i,E)=\lim_{\rho\to 0+}H_\rho(i,E)$ が存在して, $H=(H_{ij})\in\mathcal{N}(S)$ と $H(i,E)=\sum_{j\in E}H_{ij}$ (ただし $H_{ij}=H(i,\{j\})$) を満たすとき, H を偏差行列とよぶ.

定理 4. 条件 (I) を仮定する. このとき任意の P に対して,

(i) $\| H \| \leqq 2B$ を満たす偏差行列 H が存在し, 次の式が成立する.

$$H(i,E)=\sum_{n=0}^{\infty}\sum_{r=0}^{d-1}P^r(P^{nd}-P^{d*})(i,E)+\frac{1}{d}\sum_{r=0}^{d-1}(\frac{d-1}{2}-r)P^rP^{d*}(i,E). \tag{10}$$

(ii) 便宜上 H を H_0 と表す. このとき, $0\leqq\rho<\infty$ に対して (9) and (10) によって与えられた H_ρ は $\mathcal{N}(S)$ のなかで一意に定まり, $(I-\beta P)H_\rho=H_\rho(I-\beta P)=\beta(I-P^*)$ かつ $P^*H_\rho=H_\rho P^*=O$ を満たす. ここで O は零行列のこととする.

証明. 一般性を失うことなく $i\in T$ であり $\mathcal{I}$ は可算としてよい. (10) の存在をいうためには, $Q(i,E)$ のそれをいえばよい. $b(j)=\sum_{r=0}^{d-1}P^r\ (j,E)$ とおいて, 式 (4), (5), (7) を $Q_n(i,E)$ に代入すると,

$$\begin{aligned}Q_n(i,E)=&\sum_{s\in R}\sum_{k=1}^{n}\sum_{\ell=1}^{k}({}_RP^d)^\ell_{is}\{\sum_{j\in S}(P^{(k-\ell)d}-P^{d*})_{sj}b(j)\}+\sum_{j\in T}\sum_{k=1}^{n}p^{kd}_{ij}b(j)\\&+\sum_{j\in S}(\delta_{ij}-p^{d*}_{ij})b(j)-\sum_{s\in R}\sum_{k=1}^{n}\sum_{\ell=k+1}^{\infty}({}_RP^d)^\ell_{is}(\sum_{j\in S}p^{d*}_{sj}b(j))\end{aligned} \tag{11}$$

が得られる. ここで, δ_{ij} はクロネッカーのデルタである. P^d は non-dissipative だから,

$$\sum_{\ell=k+1}^{\infty}({}_RP^d)^\ell(i,R)=P^{kd}(i,T)$$

が成立する. (2) より, $0\leqq\sum_{k=0}^{\infty}P^{kd}(i,T)\leqq B$ である. よって, (11) の最後の 3 つの項は $n\to\infty$ のときに絶対収束する.

任意の $\varepsilon>0$ に対して, $K\subset\mathcal{I}$ を $({}_RP^d)^+(i,\cup_{a\in K}E_a)\geqq 1-\varepsilon/(3B)$ を満たす有限集合とする. 補助定理 3 より, すべての $j\in E_a$, $a\in K$, $E\subset S$ に対して, $|({}_aQ-Q_n)(j,E\cap R)|<\varepsilon/3$ となる n をとる. このとき

$$\lim_{n\to\infty}\sum_{a\in\mathcal{I}}\sum_{s\in E_a}(\sum_{k=1}^{n}({}_RP^d)^k_{is})Q_n(s,E)=\sum_{a\in\mathcal{I}}\sum_{s\in E_a}(\sum_{k=1}^{\infty}({}_RP^d)^k_{is})_aQ(s,E) \tag{12}$$

が成り立つ. コーシーの級数積の定理より, (11) の右辺の初項は (12) に収束する. よって $Q(i,E)$ は存在する.

式 (12) を $D(i,E)$ で表す. 補助定理 3 より, $\| D \| \leqq 2B$ を得る. 従って $D(i,E)=\sum_{j\in E}D_{ij}$ となり, (11) より $Q(i,\cdot)$ もそして $H(i,\cdot)$ も同じ性質をもつことが解る.

H_ρ が H に収束することを示す. (4) を使い, (9) の項を $\sum_{n=0}^{\infty}\beta^{nd}\{\sum_{r=0}^{d-1}\beta^r P^r P^{d*}\}$ と比較することにより, 次式を得る.

$$(13)\qquad H_\rho=\sum_{n=0}^{\infty}\sum_{r=0}^{d-1}(\beta P)^r\{\beta^{nd}(P^{nd}-P^{d*})\}+\frac{1}{1-\beta^d}\{\sum_{r=0}^{d-1}\beta^r(P^r-\frac{1}{d}\sum_{k=0}^{d-1}P^k)P^{d*}\}.$$

β を 1 に近づけると, アーベルの定理より (13) の右辺の初項は Q に収束する. (13) の最後の項にロピタルの定理を適用する. H_{ij} において $j\in E$ について和をとると (10) が導かれる.

(2) より $|H_\rho(i,E)|\leq B$ だから, $\|H\|\leqq 2B$ となることは明らかである. (ii) は容易にを示される. よって定理は示された. □

系 5. 条件 (I) を仮定すると, 次式が成立して, $i\in S$, $E\subset S$, $f\in\mathcal{F}$ に関して一様収束する.

$$(14)\qquad H(i,E)=\lim_{N\to\infty}\frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=0}^{n}\{P^k(i,E)-P^*(i,E)\},$$

証明. 定理 4(ii) の $H=I-P^*+PH$ を n 回反復代入して n について平均をとる. $P^*H=O$ より, 任意の $i\in S$, $E\subset S$ に対して次式を得る.

$$(15)\qquad H(i,E)=\frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=0}^{n}(P^k-P^*)(i,E)+\frac{1}{N+1}\sum_{n=1}^{N+1}(P^n-P^*)H(i,E)$$

定理 4(i) と (2) より, $\|\sum_{n=1}^{N+1}(P^n-P^*)H\|\leqq(2B)^2$ である. (15) において $N\to\infty$ とすると, (15) の最後の項は $i\in S$, $E\subset S$, $f\in\mathcal{F}$ に関して一様に 0 に収束する. よって (14) が成立する. □

Arapostathis 他 [1] の Theorem A.2 により, 系 5 は 定理 4(i) における $H_\rho(i,E)$ の収束が $i\in S$, $E\subset S$, $f\in\mathcal{F}$ に関して一様であることを示している.

4. Blackwell-最適定常政策の存在

この節では 連続な係数をもつローラン展開を求め, さらにBlackwell-最適な定常政策の存在を示す.

$y_{-1}(f)=P^*(f)r(f)$, $y_n(f)=H(f)^{n+1}r(f)$ とおく. ただし $r(f)=(r(i,f(i));i\in S)$ は列ベクトルとする. $x=(x_i)$ に対して, $\|x\|=\sup_{i\in S}|x_i|$ とすると, 任意の $n=-1,0,1,\cdots$ に対して, $\|y_n(f)\|\leqq(2B)^{n+1}M$ が成り立つ.

定理 4 を使い, 次の定理は [13] と同様にして示される.

定理 6. (Laurent 級数展開)　条件 (I) を仮定する. $0 < \rho_0 < 1/(2B)$ とすると, $0 < \rho \leqq \rho_0$, $f \in \mathcal{F}$ に対して次式が成立する.

$$V_\rho(f) = \rho^{-1} y_{-1}(f) + \sum_{n=0}^{\infty} (-\rho)^n y_n(f). \tag{16}$$

条件 (II). (i) 任意の $i \in S$ に対して, $A(i)$ は完備可分距離空間 A のコンパクト集合とする.

(ii) 任意の $i, j \in S$ に対して, $p(a)_{ij}$ と $r(i,a)$ は $a \in A(i)$ について連続である.

条件 (I), (II) のもとでは Scheffé の定理により, 任意の $i \in S$, $n = 1, 2, \ldots$ に対して $P(f)^n(i, E)$ は $\mathcal{F}$ 上で $E \subset S$ に関して一様連続である.

定理 7. 条件 (I), (II) を仮定する. このとき, 任意の $i \in S$ と $n = -1, 0, 1, \ldots$ に対して, $y_n(f)$ の第 i-成分 $y_n(f)_i$ は $\mathcal{F}$ 上で連続である.

証明. (3) より, $n^{-1} \sum_{k=1}^{n-1} P(f)^k r(f)$ の第 i-成分は $\mathcal{F}$ 上で連続で, $f \in \mathcal{F}$ に関して一様に $y_{-1}(f)_i$ に収束する. 従って, $y_{-1}(f)_i$ は連続である. $y_{-1}(f)_i$ の場合と同様にして, (14) より, $y_0(f)_i$ は連続となる. 残りの証明は容易であるので省略する. □

$\{\rho_n\}$ は減少列で $\lim_{n\to\infty} \rho_n = 0$ とする. Blackwell[3] によって得られる ρ_n-割引最適定常政策を $f_n \in \mathcal{F}$ とする. $\mathcal{F}$ はコンパクトだから, $f^* = \lim_{k\to\infty} f_{n_k} \in \mathcal{F}$ となる部分列 $\{f_{n_k}\}$ がある. このような f^* を ρ_k-割引最適(定常)政策の極限点とよぶ.

$Y_n(f) = (y_{-1}(f), y_0(f), \cdots, y_n(f))$ とする. $A, B \in \mathcal{N}(S)$ について $A - B$ の各行で 0 とならない最初の成分が正となるとき, $A \succeq B$ と表すことにする. $\mathcal{D}_n = \{f \in \mathcal{F};$ 任意の $g \in \mathcal{F}$について $Y_n(f) \succeq Y_n(g)\}$ と定義する.

$g, f \in \mathcal{F}$ と $n = -1, 0, 1, \cdots$ に対して, $r_0(g) = r(g)$, $n \neq 0$ のとき $r_n(g) = 0$ (零ベクトル), $y_{-2}(f) = 0$ かつ

$$\psi_n(g, f) = r_n(g) + P(g) y_n(f) - y_{n-1}(f) - y_n(f)$$

と定義する. $\Psi_n(g,f)_i$ は $\Psi_n(g,f) = (\psi_{-1}(g,f), \psi_0(g,f), \cdots, \psi_n(g,f))$ の第 i 行を表すものとする. 次の補助定理は (16) と定理 7 から容易に示される.

補助定理 8. 条件 (I), (II) を仮定する. $\mathcal{D}_n \subset \mathcal{F} \subset \mathcal{D}_{n-1}$ が成り立っているとすれば, $f^{(n+1)} \in \mathcal{D}_{n+1}$ となる. ただし, $f^{(n+1)}$ は ρ_k–割引最適政策の極限点とする.

定理 9.　条件 (I), (II) を仮定する.　ρ_k–割引最適政策の極限点によって得られる Blackwell-最適定常政策が存在する.

証明. n に関する帰納法によって, n-割引最適で $Y_n(f^{(n)}) = Y_n(f^*)$ を満たす $f^{(n)} \in \mathcal{D}_n$ の存在をいえばよい. $n = 0$ のとき, $f^* = f^{(0)}$ が帰納法の仮定を満たすことは容易に示される. (例えば [10] の Theorem 3 参照.)

$f^{(n)} \in \mathcal{D}_n$ と仮定する. $A_n(i) = \{a \in A(i)\,;\, a = g(i),\, \Psi_n(g, f^{(n)})_i = 0,\, g \in \mathcal{F}\}$ とおく. 制限された MDP $(S,\, A_n(i),\, p,\, r\,)$ において, $f^{(n+1)}$ が ρ_k-割引最適政策の極限点であるとする. $\mathcal{F}_n = \times_{i \in S} A_n(i)$ と表しておく. $g \in \mathcal{F}_n$ ならば $\Psi_n(g, f^{(n)}) = 0$ だから, $Y_{n-1}(g) = Y_{n-1}(f^{(n)})$ を得る. 従って, $\mathcal{D}_n \subset \mathcal{F}_n \subset \mathcal{D}_{n-1}$ となる. 補助定理 8 より, $f^{(n+1)} \in \mathcal{D}_{n+1}$ となることが解る. このことは, $Y_{n+1}(f^{(n+1)})$ の値が $\{\rho_k\}$ のとり方に無関係に定まることも示している.

式 (1) が $\pi^* = f^{(n)}$ としてある行で等号によって成立していると仮定する. 補助定理 8 と同様にして, (1) は $n = n+1$, $\pi^* = f^{(n+1)}$ として成立する. $Y_n(f^{(0)}) = Y_n(f^{(n)})$ と仮定すると, $V_{\rho_k}(f^{(0)}) \geqq V_{\rho_k}(f^{(n+1)})$ だから $Y_{n+1}(f^{(0)}) = Y_{n+1}(f^{(n+1)})$ となって帰納法が示される. □

References

[1] Arapostathis, A., Borkar, V. S., Ferdenandez-Gaucherand, E., Ghosh, M. K. and Marcus, S. I. (1993). Discrete–time controlled Markov processes with average cost criterion: A survey, *SIAM J. Control Optim.***31**, 282–344.

[2] Billingsley, P. (1968). Convergence of Probability Measures, John Wiley & Sons, Inc.

[3] Blackwell, D. (1965). Discounted dynamic programming, *Ann. Math. Statist.* **36**, 226–235.

[4] Chung, K. L. (1960). Markov Chains with Stationary Transition Probabilities, Springer-Verlag, Berlin.

[5] Dekker, R. and Hordijk, A. (1988). Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards, *Math. Oper. Res.* **13**, 783–809.

[6] Dekker, R. and Hordijk, A. (1992). Recurrence conditions for average and Blackwell optimality in denumerable state Markov decision chains, *Math. Oper. Res.* **17**, 271–289.

[7] Dietz, H. M. and Nollau, V. (1983). Markov Decision Problems with Countable State Spaces, Akademie-Verlag, Berlin.

[8] Doob, J. L. (1953). Stochastic Processes, John Wiley & Sons, Inc.

[9] Hordijk, A. (1974). Dynamic Programming and Potential Theory, *Math. Centre Tract* **51**(Mathematisch Centrum, Amsterdam).

[10] Kadota, Y. (1979). Countable state Markovian decision processes under the Doeblin Conditions, *Bull. Math. Statist.* **19**, 85–94.

[11] Kadota, Y. (1996). Simultaneous recurrent conditions on countable state Markov chains, *J. Inform. Optim. Sci.* **17**, 397–407.

[12] Mann, E. (1985). Optimality equations and sensitive optimality in bound-ed Markov decision processes, *Optimization* **16**, 767–781.

[13] Miller, B. L. and Veinott, A. F. Jr. (1969). Discrete dynamic programming with a small interest rate, *Ann. Math. Statist.* **40**, 366–370.

[14] Veinott, A. F. Jr. (1969). On discrete dynamic programming with sensitive discount optimality criteria, *Ann. Math. Statist.* **40**, 1635–1660.

[15] Yushkevich, A. A. (1994). Blackwell optimal policies in a Markov decision process with a Borel state space, *ZOR–Math. Meth. Oper. Res.* **40**, 253–288.