数理解析研究所講究録
第460巻 1982年 88-106



# 精密割引系をもつマルコフ決定過程について

和歌山大教育　門田良信

## §1. 精密割引系について

無限段階のマルコフ決定過程において最もよく研究されている最適系に、β-割引系と平均割引系がある。これらの系ではβ-割引期待利得（$0 \leq \beta < 1$）が有限段階のそれの極限によって表わされるのに対して、平均期待利得は、最初の有限段階の期待利得とは無関係に、推移確率の時間に関する極限分布に依存して決まる値である。従って多くの場合、β-割引最適政策の存在は平均最適政策のそれよりも緩い条件から導かれる事になる。更に、前者の存在定理を使って後者の存在を示す事も可能である。Blackwell [3]は、状態空間と決定空間が有限な場合には、1に十分近いすべてのβに対してβ-割引最適定常政策が存在する事を示した。この政策による時刻に関する利得の総和は、他の政策によるそれよりも小さくはならない事から、この政策が平均最適となる事は勿論、それよりも厳しい条件を持つ系に属するものである事がわかる。

Miller-Veinott [5]およびVeinott [6]はこの政策を∞-(割引)最適政策と呼び、存在定理の別証明とそれを求めるためのア

ルゴリズムを与えている。彼等の方法の特徴は、任意の定常政策に対応した利得を持つマルコフ連鎖をLaurent展開し、その係数を辞書式順序に従って大小を比較する事により、定常政策の優劣を決める事にある。初項を$-1$として第$n$項目までの係数比較において最大な政策を$n$-割引最適政策($n=-1,0,1,\cdots$)と呼び、その$n$に関する極限をもって$\infty$-割引最適政策を得たのである。これらの系を総称してsensitive discount optimality criteriaと言う。(この報告では仮りに精密割引系と呼ぶ。)

但し、状態や決定の空間が有限でない場合には、事情は複雑になってくる。各系における最適政策は存在しない事もあるし、Flynn[10]によれば$\infty$-割引最適であっても平均最適とは限らない。精密割引系の最適政策の存在を示すすべての文献に共通した方法は、何らかの意味でエルゴード条件を設定する事によってLaurent展開を導き、これを利用する事である。状態空間が可算な場合にはこのようにして、Hordijk[8], Hordijk-Sladky[11], Kadota[14,15]によってそれぞれ、0-割引最適、$n$-割引最適、および$\infty$-割引最適政策の存在が導かれている。状態空間が一般の場合にも、Taylor[9]は非周期的マルコフ連鎖のLaurent展開を与え、Wijngaard[13]は0-割引最適定常政策の存在を導いている。またShen-Farn[16]は

決定空間をコンパクト空間として、0-割引最適定常政策の存在を示している。これらの存在定理の相異は、先に述べたエルゴード条件と、Laurent 展開の係数の政策に関する連続性を保障する条件とに基づいていると思われる。その意味では[8], [11], [13]が与えた推移確率に関する条件は、[5], [6]の自然な拡張にはなっていない。

以下においては、精密割引最適政策に関する1つの存在定理を与える事を目的とする。取り扱うマルコフ決定過程の状態空間は、第2可算公理を満たす完備距離空間のボレル部分集合であり、決定空間は有限集合とする。諸定義と基礎的概念を§2で与えた後に、§3ではLaurent 展開を導く。その方法は基本的には[5]と同じである。§4では∞-割引最適定常政策の存在を導く。また平均最適政策は-1-割引最適政策に一致する事になる。証明の方法は、β-割引最適政策のβ→1とした極限をLaurent展開を使って考える事による。§3, 4で設定される条件は[14]によって与えられたものである。これは有限状態空間では常に満たされているから、存在定理は[5]のそれを特別な場合として含む事になる。

## §2. 準備

初めに記号の説明をする。可測空間を $(X, \mathcal{F})$ とする。$X$ が距離空間になっていれば、$\mathcal{F}$ は $X$ のボレル集合族 $\mathcal{B}(X)$ にとる。$\mathcal{F}$-可測な実数値有界関数の全体を $\mathbb{B}(X)$ と表わす。任意の $r \in \mathbb{B}(X)$ のノルムを $\|r\| = \sup\{|r(x)| ; x \in X\}$ によって与える。2つの可測空間 $(X, \mathcal{F})$, $(Y, \mathcal{G})$ が与えられた時、$\mathcal{F}$ $\times$ $Y$ 上の実数値有界関数で、任意の $y \in Y$ を固定すれば $\mathcal{F}$ 上の $\sigma$-加法的集合関数、任意の $E \in \mathcal{F}$ を固定すれば $\mathcal{G}$ 可測関数となるものの全体を $\mathbb{B}(X|Y)$ と表わす。任意の $H \in \mathbb{B}(X|Y)$ のノルムを $\|H\| = \sup\{|H(E|y)| ; E \in \mathcal{F}, y \in Y\}$ によって与える。特に $Y = X$, $\mathcal{F} = \mathcal{G}$ のとき $P \in \mathbb{B}(X|X)$ が、任意の $x \in X$, $E \in \mathcal{F}$ に対して $P(E|x) \geq 0$ かつ $P(X|x) = 1$ を満たせば、$P$ を推移確率関数と呼ぶ。推移確率関数の全体を $\mathbb{P}(X|X)$ と表わす。また、$X$ が距離空間の時、$2^X$ は $X$ の空でない閉部分集合の族とする。いま可測空間 $(S, \mathcal{F})$ と距離空間 $A$ が与えられた時、その直積空間 $\mathcal{H}_1 = SA$, $\mathcal{H}_n = SAS\cdots AS$ ($2n-1$ 個) が得られる。上記の表わし方により、空間 $\mathcal{B}(A)$, $\mathcal{B}(2^A)$, $\mathbb{B}(S|S)$, $\mathbb{B}(S|SA)$, $\mathbb{P}(S|S)$ $\mathbb{P}(S|SA)$, $\mathbb{B}(S)$, $\mathbb{B}(SA)$, $\mathbb{P}(A|\mathcal{H}_n)$ 等が定義される。

また $H, G \in \mathbb{B}(X|Y)$ が、すべての $E \in \mathcal{F}$, $y \in Y$ について $H(E|y) = G(E|y)$ を満たせば、これを簡単に $H = G$ と表わす事にする。同様の表記を $\mathbb{B}(X)$ の元についても用いる。不等号につ

いても等号の場合に準ずる。

任意の $H, G \in \mathbb{B}(S|S)$ に対して $HG(E|\lambda)$ を

$$HG(E|\lambda) = \int_S G(E|t)\, H(dt|\lambda)$$

(Lebesgue-Stiltjes 積分)によって定義する。$HG \in \mathbb{B}(S|S)$ である。特に $H, G \in \mathbb{P}(S|S)$ ならば $HG \in \mathbb{P}(S|S)$ である。同様にして $H \in \mathbb{B}(S|S)$ と $r \in \mathbb{B}(S)$ に対して $Hr \in \mathbb{B}(S)$ が定義される。次の補助定理によって示される $\mathbb{B}(S|S)$, $\mathbb{B}(S)$ に関する性質は、以下のセクションにおいて断りなしによく使われる。

補助定理 2.1.

(a) $\mathbb{B}(S|S)$, $\mathbb{B}(S)$ は共に Banach 空間となり、$\mathbb{P}(S|S)$ は $\mathbb{B}(S|S)$ の閉部分集合となる。

(b) $G, H, K \in \mathbb{B}(S|S)$ に対して、$(GH)K = G(HK)$ かつ $H(G+K) = HG + HK$, $(G+K)H = GH + KH$ が成立する。

(c) $I(E|\lambda)$ の値を $\lambda \in E$ ならば 1, $\lambda \notin E$ ならば 0 と定義すれば、$I \in \mathbb{B}(S|S)$ であり、任意の $H \in \mathbb{B}(S|S)$ に対して $HI = IH = H$ が成立する。

(d) $H, G \in \mathbb{B}(S|S)$ に対して $\|HG\| \leq \|H\| \cdot \|G\|$.

(e) $H_n, G_n \in \mathbb{B}(S|S)$, $n = 1, 2, \cdots$ に対して、$\|H_n - H\| \to 0$ $\|G_n - G\| \to 0$ as $n \to \infty$ ならば $\|H_n G_n - HG\| \to 0$ as $n \to \infty$

証明. (a), (d) は明らか。(e) は (d) を使って示される。(b), (c) は、$K$ あるいは $H$ が階段関数となる時に成立する事

を示した後にその極限をとればよい。□

任意の $H \in \mathbb{B}(S|S)$ に対して $H^1 = H$, $H^n = H \cdot H^{n-1} = H^{n-1} H$ とすれば、(b)により $H^n$ は定義されて $H^n \in \mathbb{B}(S|S)$ となる。特に $P \in \mathbb{P}(S|S)$ ならば $P^n \in \mathbb{P}(S|S)$ であり、$P^n(E|\lambda)$ は状態 $\lambda \in S$ から出発して $n$ 時刻後に状態の集合 $E$ にシステムが到達する確率を示す。

マルコフ決定過程を、順序づけられた4つの因子の組 $\mathcal{M} = (S, A, P, r)$ によって定義する。$S$ は第2可算公理を満たす完備距離空間のボレル部分集合、その上の $\sigma$-加法族 $\mathcal{B}(S)$ は相対位相に関するボレル集合族とする。$A$ は有限集合(離散距離空間と考える)とし、$A(\cdot)$ を $S$ から $2^A$ への集合値関数とする。$P \in \mathbb{B}(S|SA)$ かつ $r \in \mathbb{B}(SA)$ とする。任意の $\lambda \in S$ に対して $A(\lambda)$ は状態 $\lambda$ における使用可能な決定の集合を示す。$a \in A(\lambda)$ が定まれば、利得 $r(\lambda, a)$ が与えられシステムは推移確率 $P(\cdot|\lambda, a)$ に従って次の段階 $\lambda' \in S$ へと進む。この決定を各段階において行い利得のある意味での和が最大となるような決定の列を捜す事が、マルコフ決定過程の問題である。

任意の $(\lambda_1, a_1, \lambda_2, \cdots, \lambda_n) \in \mathcal{H}_n$ に対して $\pi(A(\lambda_n)|\lambda_1, a_1, \cdots, \lambda_n) = 1$ を満たす $\pi \in \mathbb{P}(A|\mathcal{H}_n)$ の全体を $\mathbb{P}(\{A(\lambda)\}|\mathcal{H}_n)$ と表わす。$n = 1, 2, \cdots$ に対して $\pi_n \in \mathbb{P}(\{A(\lambda)\}|\mathcal{H}_n)$ を任意に取ってできる列 $\pi = (\pi_1, \pi_2, \cdots)$ を政策と呼ぶ。各 $\pi_n$ は時刻 $n$ における決定を

それまでの履歴 $(s_1, a_1, s_2, \cdots, s_n)$ を参考にしながら確率的に定めるものである。すべての $n$, $(s_1, a_1, s_2, \cdots, s_n) \in \mathcal{H}_n$ に対して $\pi_n(\{a_n\} \mid s_1, a_1, s_2, \cdots, s_n) = 1$ となる $a_n \in A(s_n)$ が $s_1, a_1, s_2, \cdots, a_{n-1}$ に無関係に取れる時、$f_n(s_n) = a_n$ と定義する。$f_n$ は $S$ から $A$ への関数で任意の $s \in S$ に対して $f_n(s) \in A(s)$ を満たす。このような $f_n$ の列 $\pi = (f_1, f_2, \cdots)$ をマルコフ政策と呼ぶ。特に $f_n$ が $n$ に無関係になっていれば $\pi = f$ と表わし、これを定常政策と呼ぶ。定常政策の全体を $F$ で表わす。

$A(\cdot)$ が $\mathcal{B}(S)$ と $\mathcal{B}(2^A)$ に関して可測、即ち任意の $B \in 2^A$ に対して $A^{-1}(B) \in \mathcal{B}(S)$ ならば、上記の $f$ も可測となる。従って $F$ は $S$ から $A$ への $f(s) \in A(s)$ を満たす可測関数の全体に一致する。この時与えられた $P \in \mathbb{B}(S \mid SA)$, $r \in \mathbb{B}(SA)$ に対して、$P(f)(E \mid s) = P(E \mid s, f(s))$, $r(f)(s) = r(s, f(s))$ と表わせば、$P(f) \in \mathbb{B}(S \mid S)$, $r(f) \in \mathbb{B}(S)$ となる。

$r \in \mathbb{B}(SA)$ と $\pi_n \in \mathbb{P}(\{A(s)\} \mid \mathcal{H}_n)$ に対して $\pi_n r = \sum_{a \in A(s_n)} r(s_n, a)\, \pi(\{a\} \mid s_1, a_1, \cdots, s_n)$ とすれば、$\pi_n r \in \mathbb{B}(\mathcal{H}_n)$ となる。$P\pi_n r = \int_S \pi_n r(s_1, a_1, \cdots, s_n) P(ds_n \mid s_{n-1}, a_{n-1})$ とすれば、$P\pi_n r \in \mathbb{B}(\mathcal{H}_{n-1}A)$ となる。$\pi_{n-1} P \pi_n r = \sum_{a_{n-1} \in A(s_{n-1})} P\pi_n r(s_1, a_1, \cdots s_{n-1}, a_{n-1})\, \pi_{n-1}(\{a_{n-1}\} \mid s_1, a_1, \cdots, s_{n-1})$ とすれば、$\pi_{n-1} P \pi_n r \in \mathbb{B}(\mathcal{H}_{n-1})$ となる。このようにして $\pi_1 P \pi_2 P \cdots P \pi_n r \in \mathbb{B}(S)$ を得る。$\pi_1 P \pi_2 \cdots P \pi_n r(s)$ は、状態 $s \in S$ から出発して政策 $\pi = (\pi_1, \pi_2, \cdots)$ を使った時、時刻 $n$ における期待利得を

示す。これらの定義の詳細は Blackwell [4], Furukawa [7] にある。任意の政策 $\pi=(\pi_1,\pi_2,\cdots)$ に対して、$\rho$-割引総期待利得および平均期待利得をそれぞれ、

$$V_\rho(\pi)=\sum_{n=0}^{\infty}\beta^{n+1}\pi_1P\pi_2P\cdots\pi_{n-1}P\pi_n r,\quad \chi(\pi)=\liminf_n \frac{1}{n}\sum_{k=1}^{n}\pi_1P\pi_2P\cdots P\pi_k r$$

によって定義する。但し $0\le\beta<1$ かつ $\beta=\frac{1}{1+\rho}$ とし、$\rho$-割引とは実際には $\beta$-割引の事である。$V_\rho(\pi),\ \chi(\pi)\in B(S)$ である。

政策 $\pi^*$ が $\rho$-割引最適であるとは、任意の政策 $\pi$ に対して $V_\rho(\pi^*)\ge V_\rho(\pi)$ が成立する事を言う。同様に平均最適であるとは $\chi(\pi^*)\ge\chi(\pi)$ となる事を言う。$n$-割引最適であるとは

(1) $$\liminf_{\rho\to 0}\ \rho^{-k}\left(V_\rho(\pi^*)-V_\rho(\pi)\right)\ge 0,\quad k=-1,0,1,\cdots,n$$

が成立する事であり、$\infty$-割引最適であるとはすべての $n=-1,0,1,\cdots$ に対して $n$-割引最適となる事である。従って $\infty$-割引最適であるとは、任意の政策 $\pi$ と $\Delta\in S$ に対して $0<\sigma<\rho_0$ ならば $V_\sigma(\pi^*)(\Delta)\ge V_\sigma(\pi)(\Delta)$ となる定数 $\rho_0=\rho_0(\pi,\Delta)>0$ が存在する事である。状態空間 $S$ が有限ならば、上記の $\rho_0$ は $\pi,\Delta$ に無関係に存在する事が [3], [5], [6] によって知られている。また式 (1) の意味は Laurent 展開を得た段階で更に明らかとなろう。

## §3. Laurent 展開.

このセクションでは $(S,\mathcal{F})$ を可測空間とし、$P\in\mathbb{P}(S|S)$,

$r \in \mathbb{B}(S)$ を1つづつ取って固定しておく。Pに関する次の仮定は以下ずっと用いられる。

仮定I. 次の(2)を満たす正の数Mと $\mathcal{F} \times S$ から実数への関数 $P^*$ が存在する。すべての $n=1,2,\cdots$, $A \in S$, $E \in \mathcal{F}$ に対して

$$(2) \qquad \left| \sum_{k=0}^{n} \left( P^k(E|A) - P^*(E|A) \right) \right| \leqq M .$$

但し $P^0 = I$ とする。

(2)の両辺を $n+1$ で割って $n \to \infty$ にとれば $P^*$ は極限分布 $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} P^k$ となり、その収束は $A, E$ に関して一様である。また

$$(3) \qquad P^* = PP^* = P^*P = (P^*)^2$$

が成立し、$P^* \in \mathbb{P}(S|S)$ である。

Doob [2] によるとPがDoeblin条件を満たしていれば、またYosida-Kakutani [1] によればそれと同値な命題として、$r \to Pr$ が $\mathbb{B}(S)$ から $\mathbb{B}(S)$ への線型作用素として quasi-strongly completely conti. ならば、仮定Iは満たされる。

$0 \leqq \beta < 1$, $\beta = \frac{1}{1+\rho}$ とし、任意の $A \in S$, $E \in \mathcal{F}$ に対して

$$H_\rho(E|A) = \sum_{n=0}^{\infty} \beta^{n+1} \left( P^n(E|A) - P^*(E|A) \right) \quad \text{かつ} \quad h_\rho(A) = (H_\rho r)(A)$$

と定義すれば、$H_\rho \in \mathbb{B}(S|S)$, $h_\rho \in \mathbb{B}(S)$ である。$V_\rho = \sum_{n=0}^{\infty} \beta^{n+1} P^n$ のLaurent展開を得るために、まず $\rho \to 0$ のとき $\| H_\rho - H \| \to 0$ となる $H \in \mathbb{B}(S|S)$ の存在を言う。

補助定理3.1. 仮定Iのもとですべての $\rho > 0$ に対して、$\| H_\rho \| \leqq M$, $\| h_\rho \| \leqq M \cdot \| r \|$ が成立する。

証明は[14]と同じである。

補助定理3.2. 仮定Iのもとで $H_\rho$ は

$$(I-\beta P)H_\rho = H_\rho(I-\beta P)=\beta(I-P^*),\ P^*H_\rho=H_\rho P^*=0$$

を満たす $\mathbb{B}(S|S)$ の唯一の解である。

証明. (3)を使えば前半は明らか。$Y\in\mathbb{B}(S|S)$ が解とすれば $(I-\beta P)(Y-H_\rho)=0$. $\beta^k P^k$ を左から掛けて $k=0,1,\cdots,n-1$ までの和を取れば、$(I-\beta^n P^n)(Y-H_\rho)=0$ を得る。ノルムを取れば、$n\to\infty$ とした時、$\|Y-H_\rho\|=0$ を得る。よって $Y=H_\rho$。□

定理3.3. 仮定Iのもとで、$\rho\to 0$ ならば $\|H_\rho - H\|\to 0$ なる $H\in\mathbb{B}(S|S)$ が存在する。またこの時 $\|f_\rho - Hr\|\to 0$ が成立する。

証明. $\mathbb{B}(S|S)$ はBanach空間だから $\{\rho_k\}$ を $\rho_k\searrow 0$ as $k\to\infty$ に取った時、$\{H_{\rho_k}\}$ がコーシー列となる事を示せばよい。$H=\lim_{k\to\infty} H_{\rho_k}$ が $\{\rho_k\}$ の取り方によらず一意に定まる事は、次の補助定理3.4(a)によって示される。

任意の $\varepsilon>0$ に対して $N>\frac{9M^2}{\varepsilon}$ なる $N$ を取り、このNに対して $1-\beta_0 < \varepsilon N/2(N+2)^3$ を満たすように $\beta_0$ を取る。$\beta_0<\alpha<\beta<1$ なる $\alpha,\beta$ を取り $\alpha=\frac{1}{1+\rho}$, $\beta=\frac{1}{1+\rho}$ とする。補助定理3.2より $H_\rho=\beta(I-P^*)+\beta PH_\rho$. この式を $l$ 回反復して用いると、

$$H_\rho=\sum_{n=0}^{l}\beta^{n+1}(P^n-P^*)+\beta^{l+1}P^{l+1}H_\rho,\quad l=0,1,2,\cdots$$

$l=0,1,\cdots,N$ までを加えて $N+1$ で割ると、

(4) $$H_\rho = \frac{1}{N+1}\sum_{\ell=0}^{N}\sum_{n=0}^{\ell}\beta^{n+1}(P^n - P^*) + \frac{1}{N+1}\sum_{\ell=0}^{N}\beta^{\ell+1}P^{\ell+1}H_\rho$$

$\mathcal{S}_\ell = \sum_{j=0}^{\ell}(P^j - P^*)H_\rho$ とおくと補助定理3.1と仮定Iにより $\|\mathcal{S}_\ell\| \leqq M^2$. よって

$$\left\|\frac{1}{N+1}\sum_{\ell=0}^{N}\beta^{\ell+1}P^{\ell+1}H_\rho\right\| = \left\|\frac{1}{N+1}\sum_{\ell=0}^{N}\beta^{\ell+1}(P^{\ell+1}-P^*)H_\rho\right\|$$

$$\leqq \frac{1}{N+1}\left\|\sum_{\ell=0}^{N}\beta^{\ell+1}(\mathcal{S}_{\ell+1}-\mathcal{S}_\ell)\right\| < \frac{3M^2}{N} < \frac{\varepsilon}{3}$$

$H_\sigma$ についても(4)と同様の式が成立するから、結局

(5) $$\|H_\rho - H_\sigma\| < \frac{1}{N+1}\sum_{\ell=0}^{N}\sum_{n=0}^{\ell}|\beta^{n+1}-\alpha^{n+1}|\cdot\|P^n - P^*\| + \frac{2}{3}\varepsilon$$

$\|P^n - P^*\| \leqq 2$ および $\beta^{n+1}-\alpha^{n+1} < (1-\alpha)(\sum_{\ell=0}^{n}\alpha^\ell)$ を使って(5)の右辺第1項を評価してやれば

$$\|H_\rho - H_\sigma\| < (1-\alpha)\frac{2(N+2)^3}{3(N+1)} + \frac{2}{3}\varepsilon < \varepsilon$$

この事は $\{H_{\rho_k}\}$ がコーシー列となる事を示している。 □

以下の補助定理3.4、定理3.5に現われるHは定理3.3で求められたものである。3.4(a)は補助定理2.1(e)および3.2を使って示される。他の証明も基本的には[5]と同じだから省略する。

補助定理3.4. 仮定Iのもとで

(a) 補助定理3.2は $H_\rho$ のかわりにH, $\beta = 1$ として成立する。

(b) $\rho > 0$ に対して $M_\rho = \sum_{n=0}^{\infty}\beta^{n+1}P^n$ とおくと、$M_\rho \in \mathbb{B}(S|S)$ で $P^*M_\rho = \frac{1}{\rho}P^*$, $M_\rho = P^*M_\rho + H_\rho$ が成立する。

(c) $\rho > 0$ に対して $L_\rho = \sum_{n=0}^{\infty}\rho^n(-1)^n H^n$ とおくと、$L_\rho$ は

$$(I+\rho H)L_\rho = L_\rho(I+\rho H) = I$$

を満たす $\mathbb{B}(S|S)$ の唯一の解である。

(d) $\rho > 0$ に対して、$H_\rho = L_\rho H = H L_\rho$ が成立する。

定理 3.5. 仮定Iのもとで

(a) $x = P^* r$ とおくと $x$ は、$(I-P)x = 0$ かつ $P^* x = x$ を満たす $\mathbb{B}(S)$ の唯一の解である。また $y_0 = Hr$ とおくと $y_0$ は、$(I-P)y = r - x$ かつ $P^* y = 0$ を満たす $\mathbb{B}(S)$ の唯一の解である。

(b) $0 < \rho < \frac{1}{M}$ なる $\rho$ に対して

$$V_\rho = \sum_{n=0}^{\infty} \beta^{n+1} P^n r = \sum_{n=-1}^{\infty} y_n \rho^n \tag{6}$$

が成立する。ここで $y_{-1} = x = P^* r$, $y_0 = Hr$, $n = 1, 2, \cdots$ に対しては $y_n = (-1)^n H^{n+1} r$ とする。$M$ は(2)を満たす定数とする。

系 3.6. 仮定Iのもとで

(a) $n = -1, 0, \cdots$, $k = -1, 0, \cdots, n$ に対して定理 3.5 (b) によって定義された $\{y_{-1}, y_0, \cdots, y_n\}$ は

$$r_k = y_{k-1} + y_k - P y_k \,, \qquad k = -1, 0, \cdots, n \tag{7}$$

を満たす。但し、$y_{-2} = 0$ かつ $k \neq 0$ ならば $r_k = 0$, $k = 0$ ならば $r_k = r$ とする。

(b) 逆に $n = 0, 1, \cdots$ に対して $z_{-1}, z_0, \cdots z_n \in \mathbb{B}(S)$ が (7) を満たせば、$k = -1, 0, \cdots, n-1$ に対して $y_k = z_k$ が成立する。(定理 3.5 (a) は系 3.6 で $n = 0$ とした時に相当する。)

証明. (6) の左辺より $V_\rho = \beta r + \beta P V_\rho$ であり、$V_\rho$ に (6) の右辺を代入すればすべての $\rho > 0$ に対して、

$$0 = \beta\{ r + ( P - (1+\rho)I ) V_\rho \}$$

$$= \beta\{ \frac{1}{\rho}(Py_{-1} - y_{-1}) + ( r + Py_0 - y_0 - y_{-1} ) + \sum_{n=1}^{\infty} ( Py_n - y_n - y_{n-1} ) \rho^n \}$$

が成立する。よって $\rho^k$ の各係数は 0 となり (a) が示された。逆に $k = n-1, n$ に対して (7) が成立して、$y_{n-2} = z_{n-2}$ ならば、(7) の $n$ に関する式に $P^*$ を乗じて、$P^* r_n = P^* y_{n-1} = P^* z_{n-1}$。また

$$0 = r_{n-1} - r_{n-1} = y_{n-1} - z_{n-1} - P( y_{n-1} - z_{n-1} )$$

だから、$y_{n-1} - z_{n-1} = P^*( y_{n-1} - z_{n-1} ) = 0$ を得る。詳しくは $n$ に関する帰納法によって示される。 □

定理 3.5 の (a) は [3] の拡張であり、定理 3.5 (b) と系 3.6 は [5] の拡張である。[5], [6] により (6) の右辺は Laurent 展開と呼ばれている。$\beta \to 1$ のとき $\rho \to 0$ だから (6) の右辺は $V_\rho$ の漸近的性質を詳細に告げている。

## §4. 存在定理.

セクション 2 で定義したマルコフ決定過程 $\mathcal{M} = (S, A, P, r)$ に関して次の仮定をする。

仮定 II. $A(\cdot) ; S \to 2^A$ は $\mathcal{B}(S)$ と $\mathcal{B}(2^A)$ に関して可測。

仮定 III. すべての $f \in F$, $n = 1, 2, \cdots$, に対して $\| \sum_{k=0}^{n} ( P(f)^k - P^*(f)^* ) \| \leq M$ をみたす定数 $M$ が存在する。

仮定 III は、仮定 I で $P = P(f)$, $P^* = P(f)^*$ としたものに $f$

$\in F$ に関する一様有界性を付け加えたものである。$S$ が有限ならば仮定Ⅱ, Ⅲは消えてしまう。

$\{f_n\} \subset F$, $f \in F$ とし, 任意の $A \in S$ に対して $\lim_{n\to\infty} f_n(A) = f(A)$ が成立すれば, $\{f_n\}$ は $f$ に収束すると言う。決定空間 $A$ は有限だからこの事は, $n \geq N$ ならば $f_n(A) = f(A)$ となる自然数 $N$ が存在する事を意味する。

仮定Ⅱ, Ⅲより任意の $f \in F$ に対して $V_\rho(f) = \sum_{n=0}^{\infty} \beta^{n+1} P(f)^n r(f)$ の Laurent 展開が定理3.5 (b)によって導かれる。この時の $\rho^k$ の係数を $g_k = g_k(f)$ で表わす。次の補助定理は $\{g_k(f)\}$ の $F$ 上での連続性を示すものである。

補助定理4.1. 仮定Ⅱ, Ⅲのもとで $\{f_n\} \subset F$ が $f \in F$ に収束すれば, 任意の $A \in S$ に対して

(8) $$\lim_{n\to\infty} \sup_{E \in \mathcal{B}(S)} | P(f_n)^*(E|A) - P(f)^*(E|A) | = 0$$

かつ $k = -1, 0, 1, \cdots$ に対して $\lim_{n\to\infty} g_k(f_n)(A) = g_k(f)(A)$ が成立する。

証明. $P(f)(\cdot|A)$ と $f_n$ にエゴロフの定理を適用する事により任意の $\varepsilon > 0$, $k = 2, 3, \cdots$ に対して, $P(f)(S-K|A) < \frac{\varepsilon}{k-1}$, $A \in K$ かつ $t \in K$, $n \geq N$ ならば $f_n(t) = f(t)$ となる $K \in \mathcal{B}(S)$ と自然数 $N$ が存在する。これより $n \geq N$ ならば

$$| P(f_n)^k(E|A) - P(f)^k(E|A) | \leq \sum_{j=1}^{k-1} P(f)^j(S-K|A) < \varepsilon .$$

よって任意の $A \in S$ に対して $P(f_n)^k$ は $P(f)^k$ に $E \in \mathcal{B}(S)$ に関して一様収束する。$P(f)^*$ を仮定Ⅲを使って $\frac{1}{k+1} \sum_{j=0}^{k} P(f)^j$ で評価する事

により (8) が示される。またこれらの事より $H_\rho(f_n)(E|\lambda)$ が $n\to\infty$ とした時 $H_\rho(f)(E|\lambda)$ に $E\in\mathcal{B}(S)$ に関して一様収束する事が言える。定理3.3により $\rho>0$ を十分小さく取れば $\|H(f)-H_\rho(f)\|<\varepsilon$ だから、結局 $H(f_n)(E|\lambda)$ も $H(f)(E|\lambda)$ に $E\in\mathcal{B}(S)$ に関して一様収束する。これより $\lim_{n\to\infty} g_k(f_n)(\lambda)=g_k(f)(\lambda)$ は容易である。□

任意の $f\in F$, $\lambda\in S$ に対して $Y_n(f)(\lambda)=(g_{-1}(f)(\lambda), g_0(f)(\lambda),\cdots, g_n(f)(\lambda))$ と定める。$f,g\in F$ についてもし $g_k(f)(\lambda)\neq g_k(g)(\lambda)$ となる $k=-1,0,\cdots,n$ があれば、そのような最小の $k$ に対しては $g_k(f)(\lambda)>g_k(g)(\lambda)$ が成立する時、$Y_n(f)(\lambda)\succ Y_n(g)(\lambda)$ と表わす。すべての $g\in F$, $\lambda\in S$ に対して $Y_n(f)(\lambda)\succeq Y_n(g)(\lambda)$ を満たす $f\in F$ の全体を $\mathcal{D}_n$, $n=-1,0,\cdots$ によって表わす。任意の $\lambda\in S$, $a\in A(\lambda)$, $f\in F$ に対して、

$$\gamma_n(\lambda,a;f)=r_n(\lambda,a)+P g_n(\lambda,a)-g_n(f)(\lambda)-g_{n-1}(f)(\lambda)$$

$$\Psi_n(\lambda,a;f)=(\gamma_{-1}(\lambda,a;f),\gamma_0(\lambda,a;f),\cdots,\gamma_n(\lambda,a;f))$$

とする。$f\in\mathcal{D}_n$ のとき $A_n(\lambda)=\{a\in A(\lambda);\ \Psi_n(\lambda,a;f)=0\}$ とする。与えられた $\mathcal{M}$ から各 $\lambda\in S$ において使用可能な決定を $A_n(\lambda)$ に制限したマルコフ決定過程を考え、これを $\mathcal{M}_n$ と表わす。

系3.6(a)により $\Psi_n(\lambda,f(\lambda);f)=0$ だから $A_n(\lambda)\neq\emptyset$ である。また[7]の Lemma 4.5 により $A_n(\cdot)$ は $\mathcal{B}(S)$ と $2^A$ に関して可測となるから、$\mathcal{M}_n$ は仮定Ⅱ, Ⅲを満たす事になる。仮定Ⅱのもとでは任意の $\rho>0$ に対して $\rho$-割引最適定常政策が存在する事

が[7]によって知られている。Aの元に適当に順番をつけ、$\mathcal{M}_n$において $\rho_k \searrow 0$ に対して $F_n \ni f_k$ を $\rho_k$-割引最適定常政策とすれば、$\liminf_{k\to\infty} f_k(A) = f(A)$ が存在する。$f \in F_n$ となり、下極限政策と呼ぶ。ここで $F_n$ は $\mathcal{M}_n$ における定常政策の全体とする。

補助定理4.2. 仮定Ⅱ,Ⅲのもとで $f \in \mathcal{D}_n$ ならば、$f$ は $n$-割引最適定常政策である。

証明. $n-1$ の時を仮定する。$\liminf_{\rho} \rho^{-n}(V_\rho(f)(A) - V_\rho(\pi)(A)) = \alpha$ を与える $\rho_k \searrow 0$ をとり、$f_k$ を $\rho_k$-割引最適定常政策とする。$\gamma_n(f) \geq \gamma_n(f_k)$ だから $g_\ell(f_k)(A) = g_\ell(f)(A)$ となる $k$ が可算個あるような $\ell = -1, 0, \cdots, n$ が存在すれば、そのような $\ell$ の最大値と $\rho_k$, $f_k$ を改めて $\ell, \rho_k, f_k$ と表わす。

(9) $$\liminf_{k\to\infty} \rho_k^{-\ell}(V_{\rho_k}(f)(A) - V_{\rho_k}(f_k)(A)) + \liminf_{k\to\infty} \rho_k^{-\ell}(V_{\rho_k}(f_k)(A) - V_{\rho_k}(\pi)(A))$$

において第2項目は非負である。第1項目は、$-1 \leq \ell \leq n-1$ ならば帰納法の仮定より非負, $\ell = n$ ならば $k$ のとり方により0である。もし $g_{-1}(f_k)(A) < g_{-1}(f)(A)$ ならば第1項目は $\ell = -1$ の時に正、$\ell \geq 0$ の時、すべての $g_m$ が有界である事により $\infty$ となる。いずれの場合にも $\alpha \geq 0$ だから、$n$-割引最適である。□

補助定理4.3. 仮定Ⅱ,Ⅲのもとで任意の $f \in \mathcal{D}_n$ を取って $\mathcal{M}_n$ をつくると、$\mathcal{M}_n$ における下極限政策 $f_*$ は $\mathcal{D}_{n+1}$ に属する。但し $\mathcal{D}_{-2} = F$, $\mathcal{M}_{-2} = \mathcal{M}$ とする。

証明. 任意の $g \in F_n$ をとれば $\mathfrak{L}_n(\Lambda, g(\Lambda); f) = 0$ がすべての $\Lambda \in S$ について成立するから、系3.6により $\Upsilon_{n-1}(g) = \Upsilon_{n-1}(f)$. よって補助定理4.2により $g$ は $n-1$-割引最適となる。任意の $\Lambda \in S$ に対して $f_*(\Lambda) = \lim_{k\to\infty} f_k(\Lambda)$ となる $\{f_k\}$ を選ぶ。(9)において $l = n$, $f = f_*$ とした式を考えると、第2項目は非負であり第1項目は(6)を使って $\lim_{k\to\infty}(g_n(f_*)(\Lambda) - g_n(f_k)(\Lambda)) = 0$ に等しい。よって $f_* \in \mathfrak{D}_n$ となる。$\Upsilon_n(f_*) \geq \Upsilon_n(f_k)$ を使えば同様にして $f_* \in \mathfrak{D}_{n+1}$ を得る。□

定理4.4. 仮定Ⅱ, Ⅲを満たすマルコフ決定過程 $\mathcal{M}$ は $\infty$-割引最適定常政策をもつ。

証明. 補助定理4.3により $\mathfrak{D}_n \neq \phi$ であり、完備距離空間 $F$ の閉集合となる。また $\{\mathfrak{D}_n\}$ は単調非増加な集合列であるから、$\bigcap_{n=-2}^{\infty} \mathfrak{D}_n \neq \phi$ となる。補助定理4.2により、この事は $\infty$-割引最適定常政策の存在を示している。□

## 参考文献

[1] K. Yosida and S. Kakutani, "Operator-theoretical treatment of Markoff's process and mean ergodic theorem," Ann. math. 42. No.1. 1941.

[2] J. Doob, "Stochastic Processes," Wiley, New York. 1953.

[3] D. Blackwell, "Discrete dynamic programming," Ann. math. statist. 33. 1962.

[4] D. Blackwell, "Discounted dynamic programming," Ann. Math. Statist. 36. 1965.

[5] B. L. Miller and A. F. Veinott, Jr., "Discrete dynamic programming with a small interest rate", Ann. Math. Statist. 40. 1969.

[6] A. F. Veinott, Jr., "Discrete dynamic programming with sensitive discount optimality criteria", Ann. Math. Statist. 40. 1969.

[7] N. Furukawa, "Markovian decision processes with compact action spaces", Ann. Math. Statist. 43. 1972.

[8] A. Hordijk, "Regenerative Markov decision models," Math. Programming Study 6. 1976.

[9] H. M. Taylor, "A Laurent series for the resolvent of a strongly continuous stochastic semi-group," Math. Programming Study 6. 1976.

[10] J. Flynn, "Conditions for the equivalence of optimality criteria in dynamic programming," Ann. Statist. 4. 1976.

[11] A. Hordijk and K. Sladky, "Sensitive optimality criteria in countable state dynamic programming," Math. Oper. Res. 2. 1977.

[12] J. Wijngaard, "Stationary Markovian decision problems and perturbation theory of quasi-compact linear operators," Math. Oper. Res. 2. 1977.

[13] J. Wijngaard, "Sensitive optimality in stationary Markovian decision problems on a general state space," Math. Centre Tracts. 93. 1977.

[14] Y. Kadota, "Countable state Markovian decision processes under the

Doèblin conditions," Bull. math. Statist. 19. 1979.

[15] 門田.「可算状態マルコフ決定過程の sensitive discount 最適系について」数理解析研究所講究録358.「決定過程論とその周辺」1979.

[16] S. S. Sheu and K. L. Farn," A sufficient condition for the existence of a stationary 1-optimal plan in compact action Markovian decision processes," Recent Developements in Markov Decision Processes. Academic Press. 1980.