数理解析研究所講究録
第899巻 1995年 7-15

(u,v,w) 基準を持つベクトル値マルコフ決定過程について

宮崎大学教育学部　伊喜哲一郎 （ Tetuitirou IKI ）

要約：平均型基準u, 相対値v および準相対値 w をもつベクトル値のマルコフ決定過程について論じている。連鎖の状態数および各状態での選択肢数はともに有限個である。政策改良の収束後に於ける大域的最適性の判定法について述べてある。

§1. はじめに

p 次元ユークリッド空間をR^p とする。任意の有限集合 X 上で定義された R^p 値有界関数の全体を $M^p(X)$ とする。

離散時刻　$0,1,2,\cdots$　上のマルコフ決定過程

$$VMDP := (S, F, Q(F), R(F), K)$$

が与えられているとする。SはN 個の状態からなる状態空間を表し，$S := \{ 1, 2, \cdots, N \}$とする。各状態 $i \in S$ における選択肢をa_iとし，その集合を A_i とする。 $F := \underset{i \in S}{\times} A_i$ とおく。各 $f \in F$ によって決定される定常政策は f^∞ であるが，f^∞ を簡潔に f で表す。また定常政策の全体をも F と表す。各 $f \in F$に対し，$Q(f)$ は $N \times N$ の時間一様なマルコフ推移確率行列とし，その成分を $q_{ij}^{f(i)}$ とする。 $Q(F) := \{ Q(f), f \in F \}$ と表す。

$r(f)_i$ は状態 $i \in S$ における利得であり $r(f)_i \in M^p(S)$ であるとする。$r(f) := \left(r(f)_1, r(f)_2, \cdots, r(f)_N\right)^t$ とおき，さらに $R(F) := \{r(f), f \in F\}$ と表す。凸錘 K は $K \neq \phi$，$K \subset R^p$ かつ $K \cap (-K) = \{0\}$ を満足しているとするが，閉集合であるとは限らない。ここで $K_1 := K \cap (-K)^c$ とおく。

準備として§2において，B.L.Miller and A.F.Veinott,JR の結果を引用する。§3において主要結果を示す。政策改良が終了した後の最適性の判定を行う時には，エルゴード的部分連鎖数が直接的に影響している事を示す。

§2. 準備

割引率因子を $\beta := \dfrac{1}{(1+\rho)}$ $(0 \leq \rho \leq \infty)$ によって導入する。文献［1］B.L.Miller and A.F.Veinott,JR から以下の Lemma 1—3 の結果を引用する。

Q を $N \times N$ のマルコフ推移確率行列とすると

$$Q^* := \lim_{n \to \infty} \frac{1}{n+1} \sum_{k=0}^{n} Q^k$$

が存在する。各 $f \in F$ と行列 $Q(f)$ について

$$H(f) := \left[I - (Q(f) - Q^*(f))\right]^{-1} - Q^*(f)$$

$$\|H(f)\| := \max_i \sum_j |H(f)_{ij}|$$

$$u(f) := x_{-1}(f) := Q^*(f)r(f)$$

$$v(f) := x_0(f) := H(f)r(f)$$

$$x_n(f) := (-1)^n H(f)^{n+1} r(f), \quad n = 1, 2, \cdots$$

$$w(f) := x_1(f)$$

$$I_\beta(f^\infty) := \sum_{n=0}^{\infty} \beta^n Q^n(f)r(f) = [I - \beta Q(f)]^{-1} r(f)$$

とおく。さらに，　$f, g \in F$　に対し

$$L_\beta(g, f^\infty) := r(g) + \beta Q(g) I_\beta(f^\infty) - I_\beta(f^\infty)$$

$$\psi_n(g,f) := \begin{cases} Q(g)u(f) - u(f), & n = -1 \\ r(g) + Q(g)v(f) - u(f) - v(f), & n = 0 \\ Q(g)x_n(f) - x_{n-1}(f) - x_n(f), & n = 1,2,\cdots \end{cases}$$

とおくと，次の Lemma 1 の結果は良く知られている。

Lemma 1.　$f, g \in F$ に対し

(1) $u(g) - u(f) = \psi_{-1}(g,f) + Q(g)[u(g) - u(f)]$

(2) $u(g) - u(f) + v(g) - v(f) = \psi_0(g,f) + Q(g)[v(g) - v(f)]$

(3) $v(g) - v(f) + w(g) - w(f)$

$$= \psi_1(\mathsf{g}, f) + Q(\mathsf{g})[w(\mathsf{g}) - w(f)]$$

Lemma 2. もし $f \in F$ かつ $0 < \rho < \|H(f)\|^{-1}$ ならば

$$I_\beta(f^\infty) = (1+\rho)\sum_{n=-1}^{\infty} \rho^n x_n(f)$$

が成立する。□

Lemma 3. もし $f, \mathsf{g} \in F$ かつ $0 < \rho < \|H(f)\|^{-1}$ ならば

$$L_\beta(\mathsf{g}, f^\infty) = \sum_{n=-1}^{\infty} \rho^n \psi_n(\mathsf{g}, f)$$

が成立する。□

Lemma 4. ([2] 伊喜) 各 $f \in F$ に対して

$$|I - \beta Q(f)| > 0$$

が成立する。□

Lemma 5. ([2] 伊喜) $f, \mathsf{g} \in F$ に対して

$$|I - \beta Q(\mathsf{g})|(I_\beta(\mathsf{g}^\infty) - I_\beta(f^\infty)) = adj[I - \beta Q(\mathsf{g})]L_\beta(\mathsf{g}, f^\infty)$$

が成立する。□

Lemma 5 は割引率問題において政策改良が終了した後に $I_\beta(\mathsf{g}^\infty)$ と $I_\beta(f^\infty)$ の大域的な最適性の判定は，$L_\beta(\mathsf{g}, f^\infty)$ のみではなく $adj[I - \beta Q(\mathsf{g})]L_\beta(\mathsf{g}, f^\infty)$ で与えらるべき事を示している。

§ 3.主要結果

以下では $0 \leqq \beta < 1$ と仮定し，また 各 $f \in F$ に対応した $Q(f)$ の持つエルゴード的部分連鎖の数を $e(f)$ と表す。

Lemma 6.（[2] 伊喜，[3] 羽鳥・森）

各 $f \in F$ に対して，$e = e(f)$ とおくと

$$\exists\ \sigma > 0 \quad \text{s.t.} \quad \lim_{\beta \to 1} \frac{adj[I - \beta Q(f)]}{(1-\beta)^{e-1}}$$

$$= \lim_{\beta \to 1} \frac{|I - \beta Q(f)|}{(1-\beta)^{e-1}} \sum_{n=0}^{\infty} \beta^n Q^n(f) = \sigma Q^*(f)$$

となる正定数 σ が存在する。 とくに，$e = 1$ の場合に限って

$$adj[I - Q(f)] = \sigma Q^*(f)$$

が成立する。□

本稿では Lemma 2—3 と Lemma 5—6 の関連を調査する。

$f, g \in F$ に対して，$e = e(g)$ とおき Lemma 5 に Lemma 3 の結果を代入し両辺を $(1-\beta)^{e-1}$ で割ると

$$\frac{|I - \beta Q(g)|(I_\beta(g^\infty) - I_\beta(f^\infty))}{(1-\beta)^{e-1}} = \frac{adj[I - \beta Q(g)] \sum_{n=-1}^{\infty} \rho^n \psi_n(g, f)}{(1-\beta)^{e-1}}$$

となる。右辺に $\frac{1}{1-\beta}=\frac{1+\rho}{\rho}$ を代入して以下の定理を得る。

Lemma 7.（[2] 伊喜）$f,\mathsf{g}\in F$ に対して，$e=e(\mathsf{g})$ とおくと

$$\exists\ \sigma>0 \quad s.t.\ \lim_{\beta\to 1}\frac{|I-\beta Q(\mathsf{g})|(I_\beta(\mathsf{g}^\infty)-I_\beta(f^\infty))}{(1-\beta)^{e-1}}$$

$$=\ \sigma\ [u(\mathsf{g})-u(f)]$$

となる正定数 σ が存在する。 □

定理1　$f,\mathsf{g}\in F$ に対して，$e=e(\mathsf{g})$ とおくと

(1) $\psi_{-1}(\mathsf{g},f)=0$ ならば，

$$\lim_{\beta\to 1}\frac{adj[I-\beta Q(\mathsf{g})]}{(1-\beta)^{e-1}}\psi_0(\mathsf{g},f)=\sigma[u(\mathsf{g})-u(f)]$$

となる正定数 σ が存在する。　とくに $e=1$ の場合に限って

$$adj[I-Q(\mathsf{g})]\psi_0(\mathsf{g},f)=\sigma[u(\mathsf{g})-u(f)]$$

が成立する。

(2)　$u(\mathsf{g})=u(f)$ が成立するための必要かつ十分条件は

$$\psi_{-1}(\mathsf{g},f)=\lim_{\beta\to 1}\frac{adj[I-\beta Q(\mathsf{g})]}{(1-\beta)^{e-1}}\psi_0(\mathsf{g},f)=0$$

である。

（証明）$f,\mathsf{g}\in F$ に対して

(1) $\psi_{-1}(g,f)=0$ ならば, Lemma 3— Lemma 7 より

$$\lim_{\beta\to 1}\left(\begin{array}{c}\dfrac{1}{(1-\beta)^{e-1}}adj[I-\beta Q(g)]\times\\ \left[\dfrac{1}{\rho}\psi_{-1}(g,f)+\psi_0(g,f)+\rho\psi_1(g,f)+\cdots\right]\end{array}\right)$$

$$=\lim_{\beta\to 1}\frac{1}{(1-\beta)^{e-1}}adj[I-\beta Q(g)]\psi_0(g,f)$$

$$=\sigma Q^*(g)\psi_0(g,f)=\sigma[u(g)-u(f)]$$

(2) $u(g)=u(f)$ ならば

$$Q(g)u(f)=Q(g)u(g)=u(g)=u(f)$$

より $\psi_{-1}(g,f)=0$. また (1) によって

$$\lim_{\beta\to 1}\frac{adj[I-\beta Q(g)]}{(1-\beta)^{e-1}}\psi_0(g,f)=0$$

逆は (1) において $\sigma>0$ である事に注意すると 明らか。 □

定理 1 は平均型基準問題において，定常政策 f^* が大域的に最適であるための完結した判定法を与えている。

$\psi_{-1}(g,f^*)=\psi_0(g,f^*)=0$によって政策改良を終了させた後には，$f^*$ のまわりのすべての doubtful 政策 $\tilde{g}\in F$ と $e=e(\tilde{g})$ に対して

$$\sigma\left[u(\tilde{g})-u(f^*)\right]=\lim_{\beta\to 1}\frac{1}{(1-\beta)^{e-1}}adj\left[I-\beta Q(\tilde{g})\right]\psi_0(\tilde{g},f^*)$$

が成立している。各状態 i にたいする右辺の値を $M(\tilde{g},f^*)_i$ とする。このとき，$M(\tilde{g},f^*)_i\in K_1$ となっている状態 i が存在しない事を確認すればよい。　とくに，$e=1$ の場合に対しては

$$\sigma\left[u(\tilde{g})-u(f^*)\right]=adj\left[I-Q(\tilde{g})\right]\psi_0(\tilde{g},f^*)$$

によるべきである事を数値計算例を添えて発表した。

同様に，Lemma 1 によると，$(u(g),v(g))$ と $(u(f^*),v(f^*))$ を辞書式に比較する場合には $\psi_{-1}(g,f^*)=\psi_0(g,f^*)=\psi_1(g,f^*)=0$ によって政策改良を収束させた後に，f^* のまわりのすべての doubtful 政策 $\tilde{g}\in F$ と

$$\sigma\left[v(\tilde{g})-v(f^*)\right]=\lim_{\beta\to 1}\frac{1}{(1-\beta)^{e-1}}adj\left[I-\beta Q(\tilde{g})\right]\psi_1(\tilde{g},f^*)$$

の右辺に対して先と同様の事を確認すればよい事が示せる。ここで，$\psi_1(\tilde{g},f^*)$ の評価には $v(f^*)$ と $w(f^*)$ が同時に必要である。また右辺の極限値では，エルゴード的部分連鎖数 $e=e(\tilde{g})$ が直接的に影響を与えている事も分かる。　こうして (u,v,w) 基準を持つベクトル値マルコフ決定過程問題を解決する事ができる。

《訂正》 講演では，定理１を

$\psi_{-1}(\mathsf{g},f)=\psi_0(\mathsf{g},f)=\psi_1(\mathsf{g},f)=\cdots=\psi_{e-2}(\mathsf{g},f)=0$ ならば

$$\lim_{\rho\to 0}\left(\begin{array}{c} adj\left[I-\beta Q(\mathsf{g})\right](1+\rho)^{e-1}\times \\ \left[\begin{array}{c} \dfrac{1}{\rho^{e}}\psi_{-1}(\mathsf{g},f)+\dfrac{1}{\rho^{e-1}}\psi_0(\mathsf{g},f)+\dfrac{1}{\rho^{e-2}}\psi_1(\mathsf{g},f)+\cdots \\ \cdots+\dfrac{1}{\rho}\psi_{e-2}(\mathsf{g},f)+\psi_{e-1}(\mathsf{g},f)+\displaystyle\sum_{n=e}^{\infty}\rho^{n-e+1}\psi_n(\mathsf{g},f) \end{array}\right] \end{array}\right)$$

$$=adj\left[I-Q(\mathsf{g})\right]\psi_{e-1}(\mathsf{g},f)=\sigma\left[u(\mathsf{g})-u(f)\right]$$

が成立すると発表した。 その後の調査で $e\geqq 2$ の場合には恒等的に $adj\left[I-Q(\mathsf{g})\right]=0$ である事が判明したので上記の如く訂正する。

参 考 文 献

[１] B.L.Miller and A.F.Veinott,JR
Discrete Dynamic Programming with a small interest Rate.Ann.Math.Stat,vol.40.No.2,366-370,(1969)

[２] 伊喜哲一郎
"ベクトル値平均型マルコフ決定過程における非最適政策の除去について"
田中謙輔・安田正実 編集「統計的推測の数学的基礎とその応用に関する研究」平成４年度科学研究費総合（A）報告集，pp.13-23

[３] 羽鳥裕久・森 俊夫共著
有限マルコフ連鎖 培風館，昭和５７年