不確実性の下でのマルコフ決定過程に対する区間ベイズ手法 (An Interval Bayesian Method for uncertain MDPs)

宮崎大学・教育文化学部　伊喜　哲一郎 (Tetsuichiro IKI)
Faculty of Education and Culture, Miyazaki University
神奈川大学・工学部　堀口　正之 (Masayuki HORIGUCHI)
Faculty of Engineering, Kanagawa University
千葉大学・理学部　安田　正實（Masami YASUDA）
Faculty of Science, Chiba University
蔵野　正美 (Masami KURANO)

1　はじめに

推移確率行列が未知のマルコフ決定過程 (Markov Decision Processes, MDPs) の解析は, 最尤推定法を用いる場合 (cf. [2, 5, 7, 11]) とベイズ推定法を用いる場合 (cf. [5, 12, 15]) がある. ベイズ推定法においては, 事前分布をいかに設定するかが一つの問題である. その設定において, 柔軟性と融通性に富んだ頑健なモデルを構成することは現実問題への応用において重要である.

本論文では De Robertis and Hartigan[1] が提唱した事前測度区間による区間ベイズ法の考え方を適応して, 推移確率行列が未知の MDPs の解析を試みる. そのために, 未知の推移確率行列をある区間で推定した場合のモデルとして, 区間推定 MDPs(Interval estimated MDPs) を定式化しその解析を行う. この解析結果を受けて, 事前情報を区間ベイズ法 [1] にもとづく処理から得られた区間を用いたモデルとして, 区間ベイズ MDPs(Interval Bayesian estimated MDPs) を構成する.

マルコフ連鎖の推移確率行列の区間ベイズ推定は, 基本的には, 多項分布の生起確率の区間推定に帰着されるので, これに関する計算法といくつかの数値例を与える. Kurano et al[8, 9] で考察された “Controlled Markov set-chain model” は, 推移確率行列を区間でとらえる考え方においては本論文と同じであるが, 前者においては各期で推移確率行列が区間内に変動することも可能な場合を取り扱っている. 区間推定 MDPs では, 全過程を通して推移確率行列は一定である場合を扱う.

2　記号と基本命題

ここでは, いくつかの記号と続く節で用いられる基本補題を与えておく.

$\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m\times n}$ をそれぞれ実数, n 次元実列ベクトル, $m\times n$ 型実行列の全体を表す. $\mathbb{R}=\mathbb{R}^{1\times 1}, \mathbb{R}^n=\mathbb{R}^{n\times 1}$ とする. また, $\mathbb{R}_+, \mathbb{R}_+^n, \mathbb{R}_+^{m\times n}$ はそれぞれ $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m\times n}$ の各成分が非負であるようなものの集合とする. $\mathbb{R}^{m\times n}$ の半順序 $\preceq, \prec$ は次で定める: $\mathbb{R}^{m\times n}\ni A=(a_{ij}), B=(b_{ij})$ に対して

$$A\preceq B (a_{ij}\leqq b_{ij}\ (1\leqq i\leqq m, 1\leqq j\leqq n) \text{ のとき}),\ A\prec B (A\preceq B \text{ かつ } A\neq B \text{ のとき}) \tag{2.1}$$

とする. $\underline{A}\preceq\overline{A}$ なる $\underline{A}=(\underline{a}_{ij}), \overline{A}=(\overline{a}_{ij})\in\mathbb{R}_+^{m\times n}$ に対して区間 $\langle\underline{A},\overline{A}\rangle$ を次で定める:

$$\langle\underline{A},\overline{A}\rangle=\Big\{Q=(q_{ij})\in\mathbb{R}_+^{m\times n}\Big|\underline{a}_{ij}\leqq q_{ij}\leqq\overline{a}_{ij}, q_{ij}\geqq 0, \sum_{j=1}^{n} q_{ij}=1\ (1\leqq i\leqq m, 1\leqq j\leqq n)\Big\}. \tag{2.2}$$

$n\times n$ 型の確率行列の区間集合全体を $\mathcal{M}_n=\big\{\langle\underline{Q},\overline{Q}\rangle\big|\langle\underline{Q},\overline{Q}\rangle\neq\emptyset, \underline{Q}\preceq\overline{Q}, \underline{Q},\overline{Q}\in\mathbb{R}_+^{n\times n}\big\}$ で表す. $\mathcal{M}_n\ni\mathcal{Q}_1,\mathcal{Q}_2$ に対する積 $\mathcal{Q}_1\mathcal{Q}_2$ を $\mathcal{Q}_1\mathcal{Q}_2=\big\{Q_1Q_2\big|Q_1\in\mathcal{Q}_1, Q_2\in\mathcal{Q}_2\big\}$ と定める. また, $\mathcal{Q}\in\mathcal{M}_n$ に対する多重積は逐次的に定義される: $\mathcal{Q}^k=\mathcal{Q}^{k-1}\mathcal{Q}\ (k\geqq 2)$.

$C(\mathbb{R}_+)$ を $\mathbb{R}_+$ の有界閉区間の全体とする. また, $C(\mathbb{R}_+)^n$ を $C(\mathbb{R}_+)$ の要素を成分に持つ n 次元列ベクトルの全体とする: $C(\mathbb{R}_+)^n=\big\{D=(D_1,D_2,\ldots,D_n)'\big|D_i\in C(\mathbb{R}_+)\ (1\leqq i\leqq n)\big\}$. ただし, $\boldsymbol{d}'$ はベクトル $\boldsymbol{d}$ の転置を表す. $C(\mathbb{R}_+)^n$ 上の算法 (加法, スカラー倍) は次で定める:
$D=(D_1,D_2\ldots,D_n)', E=(E_1,E_2,\ldots,E_n)'\in C(\mathbb{R}_+)^n, h\in\mathbb{R}_+^n, \lambda\in\mathbb{R}_+$ に対して,

$$D+E=\{d+e|d\in D, e\in E\}, h+D=\{h+d|d\in D\}, \lambda D=\{\lambda d|d\in D\}. \tag{2.3}$$

$D=([\underline{d}_1,\overline{d}_1],[\underline{d}_2,\overline{d}_2],\ldots,[\underline{d}_n,\overline{d}_n])'\in C(\mathbb{R}_+)^n$ を $D=[\underline{d},\overline{d}]$ と記す. ただし, $\underline{d}=(\underline{d}_1,\underline{d}_2,\ldots,\underline{d}_n)\in\mathbb{R}_+^n, \overline{d}=(\overline{d}_1,\overline{d}_2,\ldots,\overline{d}_n)\in\mathbb{R}_+^n$ とする. $D=(D_1,D_2,\ldots,D_n)'\in C(\mathbb{R}_+)^n$ と部分集合 $G\subset\mathbb{R}_+^{1\times n}$ に対し

て, その積 GD を $GD=\{gd|g=(g_1,g_2,\dots,g_n)\in G, d=(d_1,d_2,\dots,d_n)'\in D, d_i\in D_i\ (1\leqq i\leqq n)\}$ と定める.

次が成り立つ.

Lemma 2.1. *([4, 8]) (i)* 任意の $\mathcal{Q}\in\mathcal{M}_n$ は $n\times n$ 次元ベクトル空間 $\mathbb{R}^{n\times n}$ の凸多面体である. *(ii)* コンパクト凸部分集合 $G\subset\mathbb{R}_+^{1\times n}$ と $D=(D_1,D_2,\dots,D_n)\in C(\mathbb{R}_+)^n$ に対して $GD\in C(\mathbb{R}_+)$ である.

$C(\mathbb{R}_+)$ 上の半順序 $\preceq,\prec$ を次で定める: $[c_1,c_2],[d_1,d_2]\in C(\mathbb{R}_+)$ に対して

$$\begin{cases}[c_1,c_2]\preceq[d_1,d_2] & (c_i\leqq d_i\ (i=1,2)\ \text{のとき})\\ [c_1,c_2]\prec[d_1,d_2] & ([c_1,c_2]\preceq[d_1,d_2]\ \text{かつ}\ [c_1,c_2]\neq[d_1,d_2]\ \text{のとき})\end{cases}\tag{2.4}$$

とする. $C(\mathbb{R}_+)^n$ 上の半順序 $\preceq,\prec$ は $C(\mathbb{R}_+)$ 上の半順序を用いて次により定める: $\boldsymbol{v}=(v_1,v_2,\dots,v_n)', \boldsymbol{w}=(w_1,w_2,\dots,w_n)'\in C(\mathbb{R}_+)^n$ に対して

$$\boldsymbol{v}\preceq\boldsymbol{w}\ (v_i\preceq w_i\ (1\leqq i\leqq n)\ \text{のとき}),\ \boldsymbol{v}\prec\boldsymbol{w}\ (\boldsymbol{v}\prec\boldsymbol{w}\ \text{かつ}\ \boldsymbol{v}\neq\boldsymbol{w}\ \text{のとき})\tag{2.5}$$

$\mathbb{R}_+^n$ の 2 つの有界閉集合 D_1,D_2 の距離としてハウスドルフ距離 ρ を考える:

$$\rho(D_1,D_2)=\max\{\sup_{x\in D_1}\inf_{y\in D_2}\|x-y\|,\ \sup_{y\in D_2}\inf_{x\in D_1}\|x-y\|\}.\tag{2.6}$$

ただし, $\|\cdot\|$ は $\mathbb{R}^n$ におけるユークリッド距離とする.

次に, 次節以降の議論の準備として有限状態マルコフ決定過程について述べる. ある決定過程の状態空間を $S=\{1,2,\dots,n\}$, 行動空間を $A=\{1,2,\dots,k\}$ とする. 次の集合を定義する:

$$P(S):=\{p=(p_1,p_2,\dots,p_n)\in\mathbb{R}_+^n\big|\sum_{i\in S}p_i=1\},\tag{2.7}$$

$$P(S|S):=\{q=(q_{ij}:i,j\in S)\in\mathbb{R}_+^{n\times n}\big|\sum_{j\in S}q_{ij}=1\ (i\in S)\},\tag{2.8}$$

$$P(S|S\times A):=\{Q=(q_{ij}(a):i,j\in S,a\in A)\in\mathbb{R}_+^{kn\times n}|q_{i\cdot}(a)\in P(s)\ (i\in S,a\in A)\}.\tag{2.9}$$

有限集合 D 上の非負実数値関数の全体を $B_+(D)$ で表す. D が有限集合のとき $B_+(D)$ と $\mathbb{R}_+^n$ を同一視する. ただし $n=|D|$ であるとする. $Q=(q_{ij}(a))\in P(S|S\times A)$ と $\boldsymbol{r}=(r(i,a))\in B_+(S\times A)$ に対して, 通常のマルコフ決定過程 MDPs$\{S,A,Q,\boldsymbol{r}\}$ を考え (cf. [13]), ここでは簡単のために確定的 (deterministic) で定常 (stationary) な政策のみを考える. S から A への写像 f の全体を F で表す. 任意の $f\in F$ に対して, 割引率 β $(0<\beta<1)$ によって割り引かれた総期待利得ベクトル $\phi(f|Q)\in\mathbb{R}_+^n$ を確率行列 $Q\in P(S|S\times A)$ の関数として次で定める:

$$\phi(f|Q)=\sum_{t=0}^{\infty}(\beta Q(f))^t\boldsymbol{r}(f),\tag{2.10}$$

ただし, $\boldsymbol{r}(f)=(r(1,f(1)),r(2,f(2)),\dots,r(n,f(n)))'\in\mathbb{R}_+^n, Q(f)=(q_{ij}(f(i)))\in P(S|S)$. 各 $f\in F$ に対して写像 $L(f):\mathbb{R}_+^n\to\mathbb{R}_+^n$ を次で定める:

$$L(f)\boldsymbol{x}=\boldsymbol{r}(f)+\beta Q(f)\boldsymbol{x},\ \boldsymbol{x}=(x_1,x_2,\dots,x_n)'\in\mathbb{R}_+^n.\tag{2.11}$$

このとき, 次の基本補題が知られている.

Lemma 2.2. *(cf. [13]) (i)* $L(f)$ は単調増加および縮小写像である. すなわち, $\boldsymbol{x}\leqq\boldsymbol{x}'$ ならば $L(f)\boldsymbol{x}\leqq L(f)\boldsymbol{x}'$ *(componentwise),* $\|L(f)\boldsymbol{x}-L(f)\boldsymbol{x}'\|\leqq\beta\|\boldsymbol{x}-\boldsymbol{x}'\|$ $(\boldsymbol{x},\boldsymbol{x}'\in\mathbb{R}_+^n)$ が成り立つ. ただし, $\|\cdot\|$ は sup ノルムとする. *(ii)* $\phi(f|Q)$ は $L(f)$ の唯一の不動点である. すなわち任意の $\boldsymbol{x}\in\mathbb{R}_+^n$ に対して $L(f)^t\boldsymbol{x}\to\phi(f|Q)$ $(t\to\infty)$ が成り立つ.

3　区間推定 MDPs とパレート最適

本節では, MDPs$\{S, A, Q, \boldsymbol{r}\}$ の推移確率行列 Q を区間 $\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle$ で推定した場合を考察する. ただし,

$$\underline{Q} = (\underline{q}_{ij}(a) : i, j \in S, a \in A) \in \mathbb{R}_+^{kn \times n}, \overline{Q} = (\overline{q}_{ij}(a) : i, j \in S, a \in A) \in \mathbb{R}_+^{kn \times n}, \tag{3.1}$$

$$\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle = \left\{ Q \in P(S|S \times A) \middle| \underline{Q} \leqq Q \leqq \overline{Q} \right\} \tag{3.2}$$

とする. 推移確率行列 Q を $\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle$ で推定した決定モデルを区間推定 MDPs$\{\mathcal{Q}\}$(Interval estimated MDPs $\{\mathcal{Q}\}$) と呼ぶ. 以下, 区間推定 MDPs の利得関数を定義しその最適化について議論する.

$f \in F$ に対する割引された総期待-集合ベクトル $\phi(f|\mathcal{Q})$ を次で定める:

$$\phi(f|\mathcal{Q}) = \left\{ \phi(f|Q) \middle| Q \in \mathcal{Q} \right\} \subset \mathbb{R}_+^n. \tag{3.3}$$

ただし, $\phi(f|Q)$ は式 (2.10) で与えられている. ここで, $\phi(f|\mathcal{Q}) \in C(\mathbb{R}_+)^n$ であることを示そう. $\mathcal{L}$ を $C(\mathbb{R}_+)^n$ から $C(\mathbb{R}_+)^n$ への写像で次のように定める:

$$\mathcal{L}(f)\boldsymbol{v} = r(f) + \beta \mathcal{Q}(f)\boldsymbol{v}, \ \boldsymbol{v} \in C(\mathbb{R}_+)^n, \tag{3.4}$$

ただし, 式 (3.4) において $\mathcal{Q}(f) = \langle \underline{Q}(f), \overline{Q}(f) \rangle, \underline{Q}(f) = (\underline{q}_{ij}(f(i))) \in \mathbb{R}_+^{n \times n}, \overline{Q}(f) = (\overline{q}_{ij}(f(i))) \in \mathbb{R}_+^{n \times n}$ である. Lemma 2.1 により $\mathcal{L}(f)\boldsymbol{v} \in C(\mathbb{R}_+)^n$ $(\boldsymbol{v} \in C(\mathbb{R}_+)^n)$ であることが示されていることに注意する. さらに, $\underline{L}(f) : \mathbb{R}_+^n \to \mathbb{R}_+^n, \overline{L}(f) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ を次で定める: $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}_+^n$ に対して

$$\underline{L}(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta \min_{Q \in \mathcal{Q}(f)} Q\boldsymbol{x}, \overline{L}(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta \max_{Q \in \mathcal{Q}(f)} Q\boldsymbol{x}. \tag{3.5}$$

このとき, 次が成り立つ.

Lemma 3.1. 任意の $f \in F$ に対して, 次が成り立つ: *(i)* $\mathcal{L}(f)$ は単調増加かつ縮小写像である. *(ii)* $\underline{L}(f), \overline{L}(f)$ は, ともに単調増加かつ sup ノルムに関して縮小写像である.

Proof. [8] の定理 3.1 を参照. ∎

Lemma 2.2 と Lemma 3.1 を適用して次を得る.

Theorem 3.1. 任意の $f \in F$ に対して次が成り立つ: *(i)* $\phi(f|\mathcal{Q}) \in C(\mathbb{R}_+)^n$ かつ $\phi(f|\mathcal{Q})$ は $\mathcal{L}(f)$ の唯一の不動点である. さらに, 任意の $\boldsymbol{v} \in C(\mathbb{R}_+)^n$ に対して $\mathcal{L}(f)^\ell \boldsymbol{v} \to \phi(f|\mathcal{Q})$ $(\ell \to \infty)$. *(ii)* $\phi(f|\mathcal{Q}) = [\underline{\phi}(f), \overline{\phi}(f)]$ とするとき, $\underline{\phi}(f), \overline{\phi}(f)$ はそれぞれ $\underline{L}(f), \overline{L}(f)$ の唯一の不動点である.

Proof. 任意の $Q \in \mathcal{Q}$ に対して, $\phi(f|Q) = \boldsymbol{r}(f) + \beta Q(f)\phi(f|Q) \leqq \overline{L}(f)\phi(f|Q)$ これより, $\phi(f|Q) \leqq \overline{L}(f)^\ell \phi(f|Q) \to \overline{\phi}(f)$ $(\ell \to \infty)$. 同様にして, $\phi(f|Q) \geqq \underline{L}(f)^\ell \phi(f|Q) \to \underline{\phi}(f)$ $(\ell \to \infty)$. 故に, $\underline{\phi}(f) \leqq \phi(f|Q) \leqq \overline{\phi}(f)$ を得る. 明らかに, $\underline{\phi}(f), \overline{\phi}(f) \in \phi(f|\mathcal{Q})$ かつ $\phi(f|Q)$ は $Q \in \mathcal{Q}$ に関して連続 (cf. [14]) であるから $\phi(f|\mathcal{Q}) = [\underline{\phi}(f), \overline{\phi}(f)]$ が成り立つ. これで (ii) が示された.

$\mathcal{L}(f)$ の不動点を $\boldsymbol{u}(f) \in C(\mathbb{R}_+)^n$ とする. 任意の $\boldsymbol{v} = [\underline{v}, \overline{v}] \in C(\mathbb{R}_+)^n$ に対して次が成り立つ: $\mathcal{L}(f)\boldsymbol{v} = [\underline{L}\,\underline{v}, \overline{L}\,\overline{v}]$. 故に, $\ell \geqq 1$ に対して $\mathcal{L}(f)^{l\ell}\boldsymbol{v} = [\underline{L}^\ell \underline{v}, \overline{L}^\ell \overline{v}]$. これより, $\ell \to \infty$ とすることで $\boldsymbol{u}(f) = [\underline{\phi}(f), \overline{\phi}(f)]$ を得る. (ii) より $\boldsymbol{u}(f) = \phi(f|\mathcal{Q})$ となり (i) が示された. ∎

$f^* \in F$ がパレート最適であるとは, $\phi(f^*|\mathcal{Q}) \prec \phi(f|\mathcal{Q})$ なる $f \in F$ が存在しない場合を言う.

Lemma 3.2. $f, g \in F$ に対して, $\phi(f|\mathcal{Q}) \prec \mathcal{L}(g)\phi(f|\mathcal{Q})$ ならば $\phi(f|\mathcal{Q}) \prec \phi(g|\mathcal{Q})$.

Proof. $\mathcal{L}(g)$ の単調性と Theorem 3.1 から $\phi(f|\mathcal{Q}) \prec \mathcal{L}(g)\phi(f|\mathcal{Q}) \prec \mathcal{L}(g)\mathcal{L}(g)\phi(f|\mathcal{Q}) \prec \cdots$ $\prec (\mathcal{L}(g))^n \phi(f|\mathcal{Q}) \to \phi(g|\mathcal{Q})$ $(n \to \infty)$. 従って, $\phi(f|\mathcal{Q}) \prec \phi(g|\mathcal{Q})$ が示された. ∎

$D \subset C(\mathbb{R}_+)^n$ に対して点 $\boldsymbol{v} \in D$ が D の**有効点** (efficient point) であるとは, $\boldsymbol{v} \prec \boldsymbol{u}$ なる $\boldsymbol{u} \in D$ が存在していないを言う. D の有効点の全体を $\mathrm{eff}(D)$ で表す. 式 (3.1) の $\underline{Q}, \overline{Q}$ の成分ベクトル $\underline{Q}_{i,a} = (\underline{q}_{i1}(a), \underline{q}_{i2}(a), \ldots, \underline{q}_{in}(a)), \overline{Q}_{i,a} = (\overline{q}_{i1}(a), \overline{q}_{i2}(a), \ldots, \overline{q}_{in}(a))$ に対して $\mathcal{Q}_{i,a} = \langle \underline{Q}_{i,a}, \overline{Q}_{i,a} \rangle$ $(i \in S, a \in A)$ とする. $\boldsymbol{u} \in C(\mathbb{R}_+)^n$ に対して次を定める:

$$\mathcal{L}(\boldsymbol{u}) := (\mathcal{L}(\boldsymbol{u})_1, \mathcal{L}(\boldsymbol{u})_2, \ldots, \mathcal{L}(\boldsymbol{u})_n)', \tag{3.6}$$

ただし, $\mathcal{L}(\boldsymbol{u})_i := \mathrm{eff}(\{r(i,a) + \beta \mathcal{Q}_{i,a}\boldsymbol{u} | a \in A\})$ $(i \in S)$ である.

このとき, Lemma 3.2 を用いて次が示される.

Theorem 3.2. f^* がパレート最適であるための必要十分条件は, $\phi(f^*|\mathcal{Q})$ が次の最適包含式の最大解となることである.

$$\boldsymbol{u} \in \mathcal{L}(\boldsymbol{u}), \boldsymbol{u} \in C(\mathbb{R}_+)^n. \tag{3.7}$$

Proof. ($\Rightarrow$) $f^* \in F$ を Pareto-optimal とする. このとき, Theorem 3.1 から $\phi(f^*|\mathcal{Q})$ は $\mathcal{L}(f^*)$ の不動点である. よって, $\phi(f^*|\mathcal{Q}) \in \mathcal{L}(\phi(f^*|\mathcal{Q}))$ である. ここで, もし, ある $\boldsymbol{u} \in C(\mathbb{R}_+)^n$ が存在して $\boldsymbol{u} \in \mathcal{L}(\boldsymbol{u})$ かつ $\phi(f^*|\mathcal{Q}) \prec \boldsymbol{u}$ であるものが存在したとする. すなわち, $\exists g \in F, \exists i_0 \in S$ s.t. $\phi(f^*|\mathcal{Q})_{i_0} \prec \boldsymbol{u}_{i_0} = r(i_0, g(i_0)) + \beta\mathcal{Q}_{i_0,g(i_0)}\boldsymbol{u} = \phi(g|\mathcal{Q})_{i_0}, \phi(f^*|\mathcal{Q})_i \preceq \boldsymbol{u}_i = r(i, g(i)) + \beta\mathcal{Q}_{i,g(i)}\boldsymbol{u} = \phi(g|\mathcal{Q})_i \ \ (i \neq i_0, i \in S)$. ただし, $\boldsymbol{a}_i$ は $\boldsymbol{a} \in C(\mathbb{R}_+)^n$ の第 i 成分を表す. これは, f^* が Pareto-optimal であることに矛盾する.
($\Leftarrow$) $\phi(f^*|\mathcal{Q})$ を $\boldsymbol{u} \in \mathcal{L}\boldsymbol{u}$ の最大解であるが Pareto-optimal でないとする. このとき, $\exists g \in F$ s.t. $\phi(f^*|\mathcal{Q}) \prec \phi(g|\mathcal{Q})$ である. 特に, $\exists i \in S$ s.t. $\phi(f^*|\mathcal{Q})_i \prec \phi(g|\mathcal{Q})_i$ である. 一般に, $\phi(f|\mathcal{Q})_i \in C(\mathbb{R}_+)$ であって, $f \in F$ は高々有限個であるから $\phi(g|\mathcal{Q})_i \preceq \phi(\overline{g}|\mathcal{Q})_i \ \ (i \in S)$ となる $\overline{g}$ が存在する. すなわち $\phi(f^*|\mathcal{Q}) \prec \phi(g|\mathcal{Q}) \preceq \phi(\overline{g}|\mathcal{Q})$ が成り立つ. ここで, $\phi(\overline{g}|\mathcal{Q}) \notin \mathcal{L}(\phi(\overline{g}|\mathcal{Q}))$ であると仮定すると, 有効点の定義から $\exists i_0 \in S$ と $\exists a_0 \in A$ に対して

$$\phi(\overline{g}|\mathcal{Q})_{i_0} \prec r(i_0, a_0) + \beta\mathcal{Q}_{i_0,a_0}\phi(\overline{g}|\mathcal{Q})_{i_0} \tag{3.8}$$

が成り立つ. ここで $f^{(1)}(i) = a_0$ (if $i = i_0$), $\overline{g}(i)$ (if $i \neq i_0, i \in S$) とすれば, Lemma 3.2 から $\phi(\overline{g}|\mathcal{Q}) \prec \phi(f^{(1)}|\mathcal{Q})$ を得る. $\phi(f^*|\mathcal{Q}) \prec \phi(f^{(1)}|\mathcal{Q})$ であって, $\phi(f^{(1)}|\mathcal{Q}) \notin \mathcal{L}(\phi(f^{(1)}|\mathcal{Q}))$ であれば式 (3.8) と同様にして $\exists f^{(2)} \in F$ s.t. $\phi(f^{(1)}|\mathcal{Q}) \prec \phi(f^{(2)}|\mathcal{Q})$. $f \in F$ は高々有限個だから, $\exists f^{(k)} \in F$ s.t. $\phi(f^*|\mathcal{Q}) \prec \phi(f^{(1)}|\mathcal{Q}) \prec \cdots \prec \phi(f^{(k)}|\mathcal{Q})$ and $\phi(f^{(k)}|\mathcal{Q}) \in \mathcal{L}(\phi(f^{(k)}|\mathcal{Q}))$ が成り立つ. これは, $\phi(f^*|\mathcal{Q})$ が $\boldsymbol{u} \in \mathcal{L}(\boldsymbol{u})$ の最大解であることに矛盾する. ∎

4 ディリクレ分布

マルコフ連鎖の推移確率行列の区間ベイズ推定は, 行列の行成分に着目すれば, 多項分布の生起確率の区間推定に帰着される. そこで, 次節以降に用いられる区間ベイズ法による推移確率の事前・事後解析のためにディリクレ分布 (多次元ベータ分布) に関するいくつかの性質を示す.

ガンマ関数 $\Gamma(x)$ $(x > 0)$ とベータ関数 $B(x, y)$ $(x, y > 0)$ をそれぞれ次のように表すことにする.

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt \ \ (x > 0), B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt \ \ (x, y > 0).$$

k-変数ディリクレ分布の p.d.f. を次のように定義する:

$$f(x_1, \ldots, x_k) = \frac{\Gamma(\nu_1 + \cdots + \nu_{k+1})}{\Gamma(\nu_1)\cdots\Gamma(\nu_{k+1})} x_1^{\nu_1-1}\cdots x_k^{\nu_k-1}(1 - x_1 - x_2 - \cdots - x_k)^{\nu_{k+1}-1}. \tag{4.1}$$

但し, $x_1, \ldots, x_k$ は k 次元多面体 $S_k := \{(x_1, \ldots, x_k) : x_i \geqq 0, i = 1, \ldots, k, \sum_{i=1}^k x_i \leqq 1\}$ の各成分であり, f は S_k 上の点以外では 0, $\nu_i \in \mathbb{R}$ は $\nu_i > 0$ $(i\ 1, 2, \ldots, k+1)$ とする.

$$\tilde{D}(\nu_1, \ldots, \nu_k; \nu_{k+1}) = \int\cdots\int_{S_k} f(x_1, \ldots, x_k)\,dx_1\cdots dx_k \tag{4.2}$$

と表す. ディリクレ積分, すなわち, ディリクレ分布の定数係数を除いた被積分関数部分に関して

$$\begin{aligned} &D(\nu_1, \nu_2, \ldots, \nu_k; \nu_{k+1}) \\ &:= \int\cdots\int_{S_k} x_1^{\nu_1-1}x_2^{\nu_2-1}\cdots x_k^{\nu_k-1}(1 - x_1 - x_2 - \cdots - x_k)^{\nu_{k+1}-1}\,dx_1dx_2\cdots dx_k \\ &= \frac{\Gamma(\nu_1)\cdots\Gamma(\nu_{k+1})}{\Gamma(\nu_1 + \cdots + \nu_{k+1})} = \prod_{n=1}^{n=k} B\Big(\nu_n, \sum_{l=n+1}^{k+1}\nu_l\Big) \end{aligned} \tag{4.3}$$

を得る.

$0<\lambda<1$ に対して,

$$
\begin{aligned}
&D(\nu_1,\dots,\nu_k;\nu_{k+1}|\lambda)\\
&:=\int\cdots\int_{S_k\cap\{0<x_1\le\lambda\}} x_1^{\nu_1-1}\cdots x_k^{\nu_k-1}(1-x_1-\cdots-x_n)^{\nu_{k+1}-1}\,dx_1\cdots dx_k\ (k\geqq 1)
\end{aligned}
\tag{4.4}
$$

とする. 特に $B(\alpha,\beta|\lambda):=D(\alpha;\beta|\lambda)$ $(\alpha,\beta>0)$ と表すとき,

$$
\begin{aligned}
&D(\nu_1,\dots,\nu_k;\nu_{k+1}|\lambda)\\
&=B(\nu_1,\nu_2+\cdots+\nu_{k+1}|\lambda)B(\nu_2,\nu_3+\cdots+\nu_{k+1})B(\nu_3,\nu_4+\cdots+\nu_{k+1})\cdots B(\nu_k,\nu_{k+1})
\end{aligned}
\tag{4.5}
$$

が成り立つ. ここで, m,n を正の整数とするとき

$$
B(m,n|\lambda)=\int_0^\lambda x^{m-1}(1-x)^{n-1}dx=\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\lambda^{m+i}\frac{1}{m+i}\quad(m,n>0)
\tag{4.6}
$$

を得る.

5　区間ベイズ法による事前・事後解析

ここでは, De Robertis & Hartigan[1] による事前測度区間を用いた区間ベイズ法を定常マルコフ決定過程の推移確率行列の区間推定へ適用し, 区間推定 MDPs を構成する事後区間について考察する.

$P(S)=P_n=\{p=(p_1,p_2,\dots,p_n)|p_i\geq 0,\sum_{i=1}^n p_i=1\}$ とおく. $L(\cdot)$ を P_n 上のルベーグ測度 (*lower bound measure*), $U(\cdot):=kL(\cdot)$ (*upper bound measure*) を測度 L の $k(k>0)$ に関する比例測度 (proportional measure) とし, 事前測度区間を $[L,kL]=[dp,kdp]$ とする. データ $\sigma=(\sigma_1,\sigma_2,\dots,\sigma_n)$ はある状態における $\hat\sigma:=\sum_{k=1}^n\sigma_k$ 回の独立試行実験でそれぞれ状態 i が σ_i 回起きたことを表す. 状態 i の生起確率が p_i であるとき, $p=(p_1,\dots,p_n)\in P_n$ に対するデータ σ の p.d.f. は多項分布で表されて

$$
f(\sigma_1,\sigma_2,\dots,\sigma_n|p)=\frac{(\sigma_1+\cdots+\sigma_n)!}{\sigma_1!\cdots\sigma_n!}p_1^{\sigma_1}p_2^{\sigma_2}\cdots p_n^{\sigma_n}
\tag{5.1}
$$

となる.

データ σ における事後測度区間を $[L_\sigma,U_\sigma]=[L_\sigma,kL_\sigma]$ とする. 次の期に状態 i へ推移する確率 p_i のうち, まず, p_1 に関する事後測度区間

$$
\left\{\frac{\int_{P_n}p_1Q(dp)}{\int_{P_n}Q(dp)}\middle|L_\sigma\leqq Q\leqq U_\sigma\right\}
$$

について調べる. 論文 [1] から, 上の事後測度区間 $[\underline{\lambda},\overline{\lambda}]$ は次の方程式の一意の解である.

$$
U_\sigma(p_1-\underline{\lambda})^-+L_\sigma(p_1-\underline{\lambda})^+=0
\tag{5.2}
$$

$$
U_\sigma(p_1-\overline{\lambda})^++L_\sigma(p_1-\overline{\lambda})^-=0
\tag{5.3}
$$

ただし, $x^+=\max\{0,x\},x^-=x-x^+=\min\{0,x\}$ である.

$\hat\sigma=\sigma_1+\sigma_2+\cdots+\sigma_n,s=\sigma_1+1,t=\hat\sigma-\sigma_1+n-1$ とおくと, 式 (5.2) と式 (5.3) は結局,

$$
K(s,t,\lambda):=\left(\frac{s}{s+t}-\lambda\right)B(s,t)+(k-1)\left(B(s+1,t|\lambda)-\lambda B(s,t|\lambda)\right)=0
\tag{5.4}
$$

$$
G(s,t,\lambda):=k\left(\frac{s}{s+t}-\lambda\right)B(s,t)-(k-1)\left(B(s+1,t|\lambda)-\lambda B(s,t|\lambda)\right)=0
\tag{5.5}
$$

の方程式の解として表される. いずれも $(\hat\sigma+n)$ 次多項式による方程式の解である.

Theorem 5.1. データ $\sigma=(\sigma_1,\sigma_2,\dots,\sigma_n),\hat\sigma=\sum_{i=1}^n\sigma_i$ とする. 事前測度区間を $[L,kL]$ とするとき, $p=(p_1,p_2,\dots,p_n)$ の p_i についての事後測度区間 $[\underline{\lambda},\overline{\lambda}]$ は次のそれぞれの方程式の一意の実数解である.

$$
K(\sigma_i+1,\hat\sigma+n-\sigma_i-1,\lambda)=0,G(\sigma_i+1,\hat\sigma+n-\sigma_i-1,\lambda)=0.
$$

6 A numerical experiment

前節までの多項分布に関して状態の個数 $n=3$ のときを考える. $P_3=\{p=(p_1,p_2,p_3)|\sum_{i=1}^3 p_i=1, p_i\geq 1, i=1,2,3\}$ とおき, $k=2$ とする, すなわち事前測度区間を $[L,2L]$ とする. ある決まった状態から $\hat{\sigma}=6$ 回の試行がなされ, 6 回中, 状態 1 に 3 回, 状態 2 に 1 回, 状態 3 に 2 回推移したとする. よって, $\sigma_1=3,\sigma_2=1,\sigma_3=2$ であり, $\hat{\sigma}=\sigma_1+\sigma_2+\sigma_3=6, s=\sigma_1+1=4, t=\sigma_2+\sigma_3+(n-1)=5$ のデータが得られているとする. Theorem 5.1 より, $\overline{\lambda}$ に関する 9 次方程式 $8-18\lambda+\lambda^5(126-336\lambda+360\lambda^2-180\lambda^3+35\lambda^4)=0$ を解いて, 解 $\overline{\lambda}\fallingdotseq 0.489$ を得る. また, $\underline{\lambda}$ に関する方程式 $4-9\lambda-\lambda^5(126-336\lambda+360\lambda^2-180\lambda^3+35\lambda^4)=0$ を解いて, 解として $\underline{\lambda}\fallingdotseq 0.400$ を得る. よって p_1 の事後測度区間は $[0.400, 0.489]$ と考えられる.

$k=1$, すなわち, 事前測度区間としてルベーグ測度を考えたとき, 事後測度区間を求める方程式から $p_i=[\underline{p}_i,\overline{p}_i]=\frac{\sigma_i+1}{\hat{\sigma}+n}$ と 1 点で表される. これは, 一様事前分布を考えたときの観測値 $(\sigma_1,\sigma_2,\sigma_3)$ によるディリクレ分布 (多次元ベータ分布) のパラメータ p_i の事後分布での期待値に等しい.

さらに, 数値実験を行い事後測度区間をもとにした Markov set-chain の問題を解いてみる (cf. [6]). 状態数 $n=3, S=\{1,2,3\}$, policy は固定 (deterministic stationary policy) として初期状態 $x_1=1$ から第 20 期の状態 x_{20} を観測するまでのうちで, それぞれの状態から次の期に推移した頻度を調べたところ, Table 6.1 左上のような行列であったとする. 例えば, 状態 2 からの推移では, この行列の第 2 行目を見て, 6 回の試行実験で次の期にそれぞれ状態 1 に $\sigma_1=1$ 回, 状態 2 に $\sigma_2=3$ 回, 状態 3 に $\sigma_3=3$ 回の推移を観測したとする.

各状態 i における p_{i1},p_{i2},p_{i3} の事後測度区間は, 本文の Theorem 5.1 から以下のように得られる (Table 6.1). $\sigma_1,\sigma_2,\sigma_3$ はそれぞれ状態 i での観測値 (推移回数) とする.

Table 6.1: Intervals of posterior measures

状態の観測度数: $\begin{pmatrix}3&1&2\\1&3&2\\1&2&4\end{pmatrix}$

状態 1, $\hat{\sigma}=6$(実験回数), $\sigma_1=3,\sigma_2=1,\sigma_3=2$ のとき:

$\hat{p}_{11}=[\underline{p}_{11},\overline{p}_{11}]$	$\hat{p}_{12}=[\underline{p}_{12},\overline{p}_{12}]$	$\hat{p}_{13}=[\underline{p}_{13},\overline{p}_{13}]$
$[0.400, 0.489]$	$[0.187, 0.260]$	$[0.292, 0.376]$

状態 2, $\hat{\sigma}=6, \sigma_1=1,\sigma_2=3,\sigma_3=2$ のとき:

$\hat{p}_{21}=[\underline{p}_{21},\overline{p}_{21}]$	$\hat{p}_{22}=[\underline{p}_{22},\overline{p}_{22}]$	$\hat{p}_{23}=[\underline{p}_{23},\overline{p}_{23}]$
$[0.187, 0.260]$	$[0.400, 0.489]$	$[0.292, 0.376]$

状態 3, $\hat{\sigma}=7, \sigma_1=1,\sigma_2=2,\sigma_3=4$ のとき:

$\hat{p}_{31}=[\underline{p}_{31},\overline{p}_{31}]$	$\hat{p}_{32}=[\underline{p}_{32},\overline{p}_{32}]$	$\hat{p}_{33}=[\underline{p}_{33},\overline{p}_{33}]$
$[0.168, 0.235]$	$[0.262, 0.334]$	$[0.458, 0.542]$

$\mathcal{Q}=\langle\underline{Q},\overline{Q}\rangle$ の第 i 行目に関する凸多面体を $\hat{q}_i$ $(i=1,2,3)$ とおくとき, その端点の集合 $\mathrm{ext}(\hat{q}_i)$ はそれぞれ以下のようになる: $\mathrm{ext}(\hat{q}_1)=\{(0.437,0.187,0.376),(0.4,0.224,0.376),(0.448,0.26,0.292),(0.489,0.219,0.292),(0.4,0.26,0.34),(0.489,0.187,0.324)\}$, $\mathrm{ext}(\hat{q}_2)=\{(0.187,0.437,0.376),(0.224,0.4,0.376),(0.26,0.448,0.292),(0.219,0.489,0.292),(0.26,0.4,0.34),(0.187,0.489,0.324)\}$, $\mathrm{ext}(\hat{q}_3)=\{(0.196,0.262,0.542),(0.168,0.29,0.542),(0.208,0.334,0.458),(0.235,0.307,0.458),(0.168,0.334,0.498),(0.235,0.262,0.503)\}$ を得る.

$\beta=0.9, \boldsymbol{r}=(3,1,2)', F\ni f$(固定) として $\underline{L}(f)\boldsymbol{x}=\boldsymbol{r}(f)+\beta\min_{Q\in\mathcal{Q}(f)}Q\boldsymbol{x}, \overline{L}(f)\boldsymbol{x}=\boldsymbol{r}(f)+\beta\max_{Q\in\mathcal{Q}(f)}Q\boldsymbol{x}$ の不動点を求めてみると, $\underline{\phi}(f)=(20.003,17.508,18.643), \overline{\phi}(f)=(21.732,19.232,20.339)$ を得る. 従って, Theorem 3.1 から $\phi(f|\mathcal{Q}(f))=[\underline{\phi}(f),\overline{\phi}(f)]=\phi(f|\mathcal{Q}(f))=([20.003,21.732],[17.508,19.232],[18.643,20,339])$ を得る. 真の推移確率行列を $Q=(q_{1\cdot},q_{2\cdot},q_{3\cdot})', q_{1\cdot}=(\frac{1}{2},\frac{1}{6},\frac{1}{3}), q_{2\cdot}=(\frac{1}{3},\frac{1}{3},\frac{1}{3}), q_{3\cdot}=(\frac{2}{5},\frac{2}{5},\frac{1}{5})$ であるときの value fanction の値は $\phi=(22.469,20.116,21.135)$ である.

7 区間ベイズ MDPs

最初に, 区間推定 MDPs$\{\mathcal{Q}\}$ の $\mathcal{Q}\in\mathcal{M}_n$ に関する連続性を証明する. 次に, 事前情報を区間ベイズ法によって処理したデータを使って区間ベイズ MDPs を定義する.

まず, $\mathcal{Q}=\langle\underline{Q},\overline{Q}\rangle\in\mathcal{M}_n$ の $\underline{Q},\overline{Q}\in\mathbb{R}_+^{n\times n}$ の連続性について示そう. 次が成り立つ. ただし, 収束は各空間に対応してユークリッド距離とハウスドルフ距離に対応している.

Lemma 7.1. *(i)* $\underline{Q}_t \downarrow \underline{Q}, \overline{Q}_t \uparrow \overline{Q}$ $(t \to \infty), \langle \underline{Q}_t, \overline{Q}_t \rangle \neq \emptyset$ $(t \geqq 1)$ とするとき, $\langle \underline{Q}_t, \overline{Q}_t \rangle \xrightarrow{\rho} \langle \underline{Q}, \overline{Q} \rangle$ $(t \to \infty)$. *(ii)* $\underline{Q}_t \uparrow \underline{Q}, \overline{Q}_t \downarrow \overline{Q}$ $(t \to \infty), \langle \underline{Q}, \overline{Q} \rangle \neq \emptyset$ $(t \geqq 1)$ とするとき, $\langle \underline{Q}_t, \overline{Q}_t \rangle \xrightarrow{\rho} \langle \underline{Q}, \overline{Q} \rangle$ $(t \to \infty)$.

Proof. (i) $\langle \underline{Q}_t, \overline{Q}_t \rangle \uparrow$ より $\{\langle \underline{Q}_t, \overline{Q}_t \rangle\}$ は収束 (i.e., $\limsup_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle = \liminf_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle$) して, $\lim_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle = \overline{\cup_{t=1}^{\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle}$ (閉包) であって, $\langle \underline{Q}_t, \overline{Q}_t \rangle \subset \langle \underline{Q}, \overline{Q} \rangle$ for all $t \geqq 1$ より

$$\lim_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle \subset \langle \underline{Q}, \overline{Q} \rangle \tag{7.1}$$

である. $\forall Q \in \langle \underline{Q}, \overline{Q} \rangle$ を取る. $\langle \underline{Q}_t, \overline{Q}_t \rangle \neq$ $(t \geqq 1)$ より $\exists Q' \in \langle \underline{Q}_t, \overline{Q}_t \rangle$ $(t \geqq 1)$ であって $\langle \underline{Q}_t, \overline{Q}_t \rangle \subset \langle \underline{Q}, \overline{Q} \rangle$ より, $Q' \in \langle \underline{Q}, \overline{Q} \rangle$. $0 \leqq \alpha \leqq 1$ に対して, $Q(\alpha) = \alpha Q + (1-\alpha)Q'$ とする. このとき, $Q(0) = Q' \in \langle \underline{Q}, \overline{Q} \rangle, Q(1) = Q \in \langle \underline{Q}, \overline{Q} \rangle$ かつ Lemma 2.1(i) より $\langle \underline{Q}, \overline{Q} \rangle$ は凸集合であるから

$$Q(\alpha) \in \langle \underline{Q}, \overline{Q} \rangle \quad \text{for all } 0 \leqq \alpha \leqq 1. \tag{7.2}$$

$Q(0) = Q' \in \langle \underline{Q}_t, \overline{Q}_t \rangle$ for all $t \geqq 1$ に注意して, $\alpha_t := \sup\{\alpha | Q(\alpha) \in \langle \underline{Q}_t, \overline{Q}_t \rangle\}$ $(t \geqq 1)$ とすると, $\langle \underline{Q}_t, \overline{Q}_t \rangle$ はコンパクトな凸集合であり, かつ $\langle \underline{Q}_t, \overline{Q}_t \rangle \subset \langle \underline{Q}_{t+1}, \overline{Q}_{t+1} \rangle$ $(t \geqq 1)$ より, $Q(\alpha) \in \langle \underline{Q}_t, \overline{Q}_t \rangle$ for all $0 \leqq \alpha \leqq \alpha_t$ と $\alpha_t \leqq \alpha_{t+1} \leqq 1$ $(t \geqq 1)$ が成り立つ.

$\alpha^* := \lim_{t\to\infty} \alpha_t$ とする. $\alpha^* < 1$ とすると $\alpha^* < \alpha \leqq 1$ なる α に対して $Q(\alpha) \notin \langle \underline{Q}, \overline{Q} \rangle$ となるが, これは式 (7.2) に矛盾する. 故に, $\alpha^* = 1$. 従って, $Q(\alpha_t) \in \langle \underline{Q}_t, \overline{Q}_t \rangle \to Q(1) = Q$ $(t \to \infty)$. これは, $\limsup_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle \supset \langle \underline{Q}, \overline{Q} \rangle$ を意味する. $\limsup_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle = \lim_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle$ であるから

$$\lim_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle \supset \langle \underline{Q}, \overline{Q} \rangle. \tag{7.3}$$

式 (7.1) と式 (7.3) より $\lim_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle = \langle \underline{Q}, \overline{Q} \rangle$. これで (i) が示された. (ii) は明らかに成り立つ.∎

上の Lemma 7.1 を用いて次が示される.

Theorem 7.1. $\underline{Q}_t \to \underline{Q}, \overline{Q}_t \to \overline{Q}$ $(t \to \infty), \mathcal{Q}_t := \langle \underline{Q}_t, \overline{Q}_t \rangle \neq \emptyset$ $(t \geqq 1), \mathcal{Q} := \langle \underline{Q}, \overline{Q} \rangle$ とする. このとき, 次が成り立つ: *(i)* $\mathcal{Q}_t \to \mathcal{Q}$ $(t \to \infty)$. *(ii)* $\phi(f|\mathcal{Q}_t) \to \phi(f|\mathcal{Q})$ $(t \to \infty)$ $(f \in F)$.

Proof. (i) $\forall \varepsilon > 0$ に対して $\limsup_{t\to\infty} \mathcal{Q}_t = \limsup_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle \subset \langle \underline{Q} - \varepsilon E, \overline{Q} + \varepsilon E \rangle$, ただし, $E = (e_{ij})$ は $e_{ij} = 1$ $(1 \leqq i, j \leqq n)$ とする. このとき, $\liminf_{t\to\infty} \mathcal{Q}_t = \liminf_{t\to\infty} \langle \underline{Q}_t, \overline{Q}_t \rangle \supset \langle \underline{Q} + \varepsilon E, \overline{Q} - \varepsilon E \rangle$ であって, ここで $\varepsilon \to 0$ とすると Lemma 7.1 より $\lim_{\varepsilon\to 0} \langle \underline{Q} - \varepsilon E, \overline{Q} + \varepsilon E \rangle = \lim_{\varepsilon\to 0} \langle \underline{Q} + \varepsilon E, \overline{Q} - \varepsilon E \rangle = \langle \underline{Q}, \overline{Q} \rangle$. 故に, $\{\mathcal{Q}_t\}$ は収束して $\lim_{t\to\infty} \mathcal{Q}_t = \mathcal{Q}$ を得る. (ii) (i) より, $\mathcal{Q}_t \to \mathcal{Q}$ だから, $\boldsymbol{x} \in \mathbb{R}_+$ について $\min_{Q\in\mathcal{Q}_t} Q\boldsymbol{x} \to \min_{Q\in\mathcal{Q}} Q\boldsymbol{x}, \max_{Q\in\mathcal{Q}_t} Q\boldsymbol{x} \to \max_{Q\in\mathcal{Q}} Q\boldsymbol{x}$ $(t \to \infty)$. 従って $\underline{L}_t(f)\underline{\phi}_t(f) \to \underline{L}(f)\underline{\phi}(f), \overline{L}_t(f)\overline{\phi}_t(f) \to \overline{L}(f)\overline{\phi}(f)$ $(t \to \infty)$. 定理 3.1 から $\phi(f|\mathcal{Q}) = [\underline{\phi}(f), \overline{\phi}(f)]$ であるから $\phi(f|\mathcal{Q}_t) \to \phi(f|\mathcal{Q})$ $(t \to \infty)$ を得る.∎

真の推移確率行列 $Q \in P(S|S \times A)$ による MDPs$\{Q\}$ の t 期の状態と行動をそれぞれ X_t, Δ_T $(t \geqq 0)$ で表し, t 期までの履歴を $H_t = (X_0, \Delta_0, X_1, \Delta_1, \ldots, X_t)$ とする. 任意の $i, j \in S, a \in A$ に対して

$$N_T(j|i, a, H_T) := \sum_{t=0}^{T-1} I_{\{X_t=i, \Delta_t=a, X_{t+1}=j\}} \ (T \geqq 1) \tag{7.4}$$

とおく. 各 $i \in S, a \in A$ に対して, 多項分布の生起確率 $\{p_j = p_{ij}(a),\ (1 \leqq j \leqq n)\}$ に対する観測値 $\{N_T(j|i, a, H_T),\ 1 \leqq j \leqq n\}$ によるベイズ区間を $\mathcal{Q}(H_T) = \langle \underline{Q}(H_T), \overline{Q}(H_T) \rangle = \langle (\underline{q}_{ij}(a|H_T)), (\overline{q}_{ij}(a|H_T)) \rangle$ とする. すなわち, $\underline{Q}(H_T) := (\underline{q}_{ij}(a|H_T) :\ i, j \in S, a \in A) \in \mathbb{R}_+^{n\times nk}, \overline{Q}(H_T) := (\overline{q}_{ij}(a|H_T) :\ i, j \in S, a \in A) \in \mathbb{R}_+^{n\times nk}$ として, $\mathcal{Q}(H_T) = \langle \underline{Q}(H_T), \overline{Q}(H_T) \rangle$ とする. $Q \in P(S|S \times A)$ に対して, MDPs$\{Q\}$ を事前情報 H_T の区間ベイズ $\mathcal{Q}(H_T)$ で推定した MDPs を区間ベイズ MDPs$\{\mathcal{Q}(H_T)\}$ と言う. $N_T(i, a|H_T) := \sum_{j\in S} N_T(j|i, a, H_T)$ $(i \in S, a \in A)$ とおく.

区間ベイズの性質 ([1]) および Theorem 7.1 を用いて次の結果を得る.

Theorem 7.2. $\{X_0, \Delta_0, X_1, \Delta_1, \ldots\}$ を $MDPs\{Q\}$ からの過程とする. 任意の $i \in S, a \in A$ に対して, 確率 *1* で $N_T(i,a|H_T) \to \infty$ $(T \to \infty)$ とする. このとき, 確率 *1* で区間ベイズ $MDPs\{\mathcal{Q}(H_T)\}$ は $MDPs\{Q\}$ に収束する, すなわち, 次が成り立つ. *(i)* $\mathcal{Q}(H_T) \to \{Q\}$ $(T \to \infty)$, *(ii)* $\phi(f|\mathcal{Q}(H_T)) \to \phi(f|Q)$ $(T \to \infty)$, $(f \in F)$.

Proof. Theorem 7.2 の条件が成り立てば, [1] の定理 5.2 より $\underline{Q}(H_T), \overline{Q}(H_T) \to Q$ $(T \to \infty)$ を得る. 従って, Theorem 7.1 より (i),(ii) が成り立つことがわかる.∎

References

[1] L. De Robertis and J. A. Hartigan. Bayesian inference using intervals of measures. *Ann. Statist.*, 9(2):235–244, 1981.

[2] B. Doshi and S. E. Shreve. Strong consistency of a modified maximum likelihood estimator for controlled Markov chains. *J. Appl. Probab.*, 17(3):726–734, 1980.

[3] N. Furukawa. Characterization of optimal policies in vector-valued Markovian decision processes. *Math. Oper. Res.*, 5(2):271–279, 1980.

[4] D. J. Hartfiel. *Markov set-chains*, volume 1695 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.

[5] O. Hernández-Lerma. *Adaptive Markov control processes*, volume 79 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[6] M. Horiguchi. Examples for Bayesian approach to uncertain MDPs,
URL http://www.math.kanagawa-u.ac.jp/~horiguchi/

[7] T. Iki, M. Horiguchi, M. Yasuda, and M. Kurano. A learning algorithm for communicating markov decision processes with unknown transition matrices. *Bulletin of Informatics and Cybernetics*, 39:11–24, 2007.

[8] M. Kurano, J. Song, M. Hosaka, and Y. Huang. Controlled Markov set-chains with discounting. *J. Appl. Probab.*, 35(2):293–302, 1998.

[9] M. Kurano, M. Yasuda, and J. Nakagami. Interval methods for uncertain Markov decision processes. In *Markov processes and controlled Markov chains (Changsha, 1999)*, pages 223–232. Kluwer, 2002.

[10] K. Kuratowski. *Topology. Vol. I.* New edition, revised and augmented. Translated from the French by J. Jaworowski. Academic Press, 1966.

[11] P. Mandl. Estimation and control in Markov chains. *Advances in Appl. Probability*, 6:40–60, 1974.

[12] J. J. Martin. *Bayesian decision problems and Markov chains*. Publications in Operations Research, No. 13. John Wiley & Sons Inc., 1967.

[13] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons Inc., 1994.

[14] E. Solan. Continuity of the value of competitive Markov decision processes. *J. Theoret. Probab.*, 16(4):831–845 (2004), 2003.

[15] K. M. van Hee. *Bayesian control of Markov chains*, volume 95 of *Mathematical Centre Tracts*. Mathematisch Centrum, 1978.

[16] S. S. Wilks. *Mathematical statistics*. A Wiley Publication in Mathematical Statistics. John Wiley & Sons Inc., 1962. 田中英之, 岩本誠一 (訳),「数理統計学・増訂新版 1,2」, 1971,1972 年, 東京図書.