数理解析研究所講究録
第2044巻 2017年 199-206

区間型マルコフ決定モデルについて (On a Markov decision model with interval-valued transition matrices)

神奈川大学・理学部　堀口　正之 (Masayuki HORIGUCHI)
Faculty of Science, Kanagawa University

1　はじめに

本報告では、推移法則が未知の場合のマルコフ決定過程 (Markov decision Processes, MDP) について、推移法則を区間で推定し、推移確率行列の各成分が閉区間で表され、推移確率行列はその閉区間内に成分をもつ確率行列の集合として構成される区間型マルコフ連鎖とそのもとでのマルコフ決定過程の定式化を行う。割引総期待利得の評価基準での価値関数の表現と, α-パーセンタイル型の価値関数の表現について考察を行う.

先行研究として、区間型のマルコフ集合連鎖モデルを考察した Kurano et. al (1999) や Hartfiel (1998) がある. これらの研究では、マルコフ連鎖を構成する推移確率行列の各成分が単一閉区間のそれぞれで表され, その条件下での推移確率行列の集合は、その区間内にある任意の確率行列からなる集合と捉えられる. そして、価値関数に関しても区間表現され、最適方程式とそのもとでの最適性（パレート最適性）によって最適解 (最適政策) が特徴づけられる. 本報告では, これらについてその概略を述べ, 推移法則未知のマルコフ決定モデルにおいて, その評価基準としてパーセンタイル型について考察をする.

2　準備

本節では、割引平均期待利得を評価基準とするマルコフ決定過程についてまず述べて、その後、推移法則が各成分ごとに閉区間で表現されるマルコフ集合連鎖モデルでの最適化についてみていく.

マルコフ決定過程は, 次のような 4 つの要素からなる $\{S, A, Q, \boldsymbol{r}\}$:
$S = \{1, 2, \ldots, n\}$ は n 個の有限要素をもつ状態空間であり, $A = \{a_1, a_2, \ldots, a_K\}$ は K 個の有限要素をもつ決定空間である. また, $Q = (q_{ij}(a)) \in P(S|S \times A)$ は K 種の未知の推移確率行列からなるパラメータ空間とする. さらに, $\boldsymbol{r} = (r(i,a)) \in B_+(S \times A)$ は, 状態 $i \in S$ で決定 $a \in A$ を選択したときに生じる利得を表す関数である. 関数 $f : S \to A$ によって, 各状態 $i \in S$ に対する決定 $f(i)$ を表す. f は各期での確定的定常政策 $(f, f, \ldots,)$ も表し, 簡単に $f = (f, f, \ldots,)$ とも表す. 真の推移確率行列を Q とするとき, 確定的定常政策 f の下での価値関数 $\phi : S \to \boldsymbol{R}$ は次式で与えられる.

$$\phi(f|Q) = \sum_{t=0}^{\infty} (\beta Q(f))^t \boldsymbol{r}(f), \tag{1}$$

区間型の推移確率行列 $\langle \underline{A}, \overline{A} \rangle$ は, 次式のような推移確率行列 (q_{ij}) の集合として定義される.

$$\langle \underline{A}, \overline{A} \rangle := \Big\{ Q = (q_{ij}) \in \boldsymbol{R}_+^{m \times n} \Big| \underline{a}_{ij} \leqq q_{ij} \leqq \overline{a}_{ij}, \; q_{ij} \geqq 0, \sum_{j=1}^{n} q_{ij} = 1 \; (1 \leqq i \leqq m, 1 \leqq j \leqq n) \Big\}. \tag{2}$$

一般に, $\boldsymbol{R}^{m\times n}$ 上の半順序 $\preceq$ と $\prec$ を次のように定義する. 次のような 2 つの実数値行列 $\boldsymbol{R}^{m\times n}\ni A=(a_{ij}), B=(b_{ij})$ に対して,

$$\begin{cases} A\preceq B & \text{if } a_{ij}\leqq b_{ij}\ (1\leqq i\leqq m, 1\leqq j\leqq n) \\ A\prec B & \text{if } A\preceq B \text{ and } A\neq B \end{cases} \tag{3}$$

とおく. そして, 任意の順序のついた行列 $\underline{A}\preceq\overline{A}$ に対して, 区間型の推移確率行列 $\langle\underline{A},\overline{A}\rangle$ が定義されるのである.

$n\times n$ 行列を成分に持つ区間型非負値行列の集合を $\mathcal{M}_n$ と表す. すなわち, .

$$\mathcal{M}_n=\left\{\langle\underline{Q},\overline{Q}\rangle\middle|\langle\underline{Q},\overline{Q}\rangle\neq\emptyset,\underline{Q}\preceq\overline{Q},\underline{Q},\overline{Q}\in\boldsymbol{R}_+^{n\times n}\right\} \tag{4}$$

ここで, $\mathcal{Q}_1,\mathcal{Q}_2\in\mathcal{M}_n$ に対してその積を次のように定義する.

$$\mathcal{Q}_1\mathcal{Q}_2=\left\{Q_1Q_2\middle|Q_1\in\mathcal{Q}_1,Q_2\in\mathcal{Q}_2\right\} \tag{5}$$

同一の区間型行列 $\mathcal{Q}\in\mathcal{M}_n$ の k 回の積は次式によって帰納的に求められる.

$$\mathcal{Q}^k=\mathcal{Q}^{k-1}\mathcal{Q}\ (k\geqq 2). \tag{6}$$

非負実数値の集合 $\boldsymbol{R}_+$ におけるすべての有界かつ閉区間の集合を $C(\boldsymbol{R}_+)$ と表し, その n 次元列ベクトルをなす集合を $C(\boldsymbol{R}_+)^n$ と表す, すなわち,

$$C(\boldsymbol{R}_+)^n=\left\{D=(D_1,D_2,\ldots,D_n)'\middle|D_i\in C(\boldsymbol{R}_+)\ (1\leqq i\leqq n)\right\} \tag{7}$$

とする. ただし, ベクトル $\boldsymbol{d}$ の転置ベクトルの記号として $\boldsymbol{d}'$ を用いる.

区間型非負値行列における演算として, それぞれ次のように定義する. $D=(D_1,D_2\ldots,D_n)', E=(E_1,E_2,\ldots,E_n)'\in C(\boldsymbol{R}_+)^n$, $h\in\boldsymbol{R}_+^n,\lambda\in\boldsymbol{R}_+$ に対して,

$$\begin{aligned} D+E&=\{d+e|d\in D,e\in E\}, \\ h+D&=\{h+d|d\in D\}, \\ \lambda D&=\{\lambda d|d\in D\}. \end{aligned} \tag{8}$$

また, $D:=[\underline{d},\overline{d}]=([\underline{d}_1,\overline{d}_1],[\underline{d}_2,\overline{d}_2],\ldots,[\underline{d}_n,\overline{d}_n])'\in C(\boldsymbol{R}_+)^n$, where, $\underline{d}=(\underline{d}_1,\underline{d}_2,\ldots,\underline{d}_n)\in\boldsymbol{R}_+^n,\overline{d}=(\overline{d}_1,\overline{d}_2,\ldots,\overline{d}_n)\in\boldsymbol{R}_+^n$ とそれぞれおくとき, 任意の $D=(D_1,D_2,\ldots,D_n)'\in C(\boldsymbol{R}_+)^n$ と $G\subset\boldsymbol{R}_+^{1\times n}$ に対してその積 GD を次のように定める.

$$\begin{aligned} GD=\{gd|g=(g_1,g_2,\ldots,g_n)\in G,d=(d_1,d_2,\ldots,d_n)'\in D,\\ d_i\in D_i\ (1\leqq i\leqq n)\} \end{aligned} \tag{9}$$

この時, 次のような性質を得る.

Lemma 2.1 *([7, 16])*

(i) 任意の $\mathcal{Q}\in\mathcal{M}_n$ は $\boldsymbol{R}^{n\times n}$ における凸多面体である.

(ii) 任意のコンパクト部分集合 $G\subset\boldsymbol{R}_+^{1\times n}$ と集合 $D\in C(\boldsymbol{R}_+)^n$ に対して, 積 $GD\in C(\boldsymbol{R}_+)$ が成り立つ.

さらにいくつかの半順序について定義する. $C(\boldsymbol{R}_+)$ 上の半順序 $\preceq, \prec$ は次のように定められる. $[c_1, c_2], [d_1, d_2] \in C(\boldsymbol{R}_+)$ に対して,

$$\begin{cases} [c_1, c_2] \preceq [d_1, d_2] & \text{if } c_i \leqq d_i \ (i = 1, 2), \\ [c_1, c_2] \prec [d_1, d_2] & \text{if } [c_1, c_2] \preceq [d_1, d_2] \text{ and } [c_1, c_2] \neq [d_1, d_2]. \end{cases}$$

また, $C(\boldsymbol{R}_+)^n$ 上の半順序 $\preceq, \prec$ は上述の $C(\boldsymbol{R}_+)$ での半順序を用いて次のように定められる. 区間型のベクトル $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)'$ と $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)' \in C(\boldsymbol{R}_+)^n$ に対して,

$$\begin{cases} \boldsymbol{v} \preceq \boldsymbol{w} & \text{if } v_i \preceq w_i \ (1 \leqq i \leqq n), \\ \boldsymbol{v} \prec \boldsymbol{w} & \text{if } \boldsymbol{v} \preceq \boldsymbol{w} \text{ and } \boldsymbol{v} \neq \boldsymbol{w}. \end{cases}$$

3 区間型モデルでの評価関数

本節では, さらにいくつかの記法を準備して区間推定されたマルコフ決定モデルの定式化をする.

$\boldsymbol{R}_+^n$ における優界かつ閉部分集合 D_1 と D_2 に対してハウスドルフ距離 ρ を次のように定義する.

$$\rho(D_1, D_2) := \max\{ \sup_{x \in D_1} \inf_{y \in D_2} \|x - y\|, \sup_{y \in D_2} \inf_{x \in D_1} \|x - y\| \}. \tag{10}$$

ただし, $\|\cdot\|$ は $\boldsymbol{R}^n$ でのユークリッド距離である.

状態空間 $S = \{1, 2, \ldots, n\}$ と決定空間 $A = \{1, 2, \ldots, k\}$ に対して, 次のようにそれぞれの確率分布を定義する.

$$\begin{aligned} P(S) :=& \{p = (p_1, p_2, \ldots, p_n) \in \boldsymbol{R}_+^n | \sum_{i \in S} p_i = 1\}, \\ P(S|S) :=& \{q = (q_{ij} : i, j \in S) \in \boldsymbol{R}_+^{n \times n} | \sum_{j \in S} q_{ij} = 1 \ (i \in S)\}, \\ P(S|S \times A) :=& \{Q = (q_{ij}(a) : i, j \in S, a \in A) \in \boldsymbol{R}_+^{kn \times n} | q_{i\cdot}(a) \in P(S) \ (i \in S, a \in A)\}. \end{aligned}$$

また, 有限要素の集合 D 上のすべての非負実数値関数の集合を $B_+(D)$ とおく. D $(n = \#D)$ に対して, $B_+(D)$ は $\boldsymbol{R}_+^n$ と同一視できる.

前述した下記の割引平均期待利得 ϕ に対して, つぎの Lemma が成り立つことがよく知られている (cf. Puterman(1994)).

$$\phi(f|Q) = \sum_{t=0}^{\infty} (\beta Q(f))^t \boldsymbol{r}(f) \tag{11}$$

確定的定常政策の集合を F とおき, $f \in F$ に対して, 写像 $L(f) : \boldsymbol{R}_+^n \to \boldsymbol{R}_+^n$ は次のように定義される.

$$L(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta Q(f)\boldsymbol{x}, \ \boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \boldsymbol{R}_+^n. \tag{12}$$

このとき, つぎが成り立つ

Lemma 3.1 *(cf. Puterman[22])*

(i) $L(f)$ は単調増加かつ縮小写像である、すなわち、

$$\boldsymbol{x} \leqq \boldsymbol{x}' \text{ ならば } L(f)\boldsymbol{x} \leqq L(f)\boldsymbol{x}' \text{ が各成分ごとに成り立つ,}$$
$$\|L(f)\boldsymbol{x} - L(f)\boldsymbol{x}'\| \leqq \beta\|\boldsymbol{x} - \boldsymbol{x}'\| \ (\boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{R}_+^n),$$

ただし, $\|\cdot\|$ は sup-ノルムである.

(ii) $\phi(f|Q)$ は $L(f)$ ただ一つの不動点である, すなわち, 任意の $\boldsymbol{x} \in \boldsymbol{R}_+^n$ に対して,

$$L(f)^t\boldsymbol{x} \to \phi(f|Q) \ (t \to \infty)$$

が成り立つ.

真の推移確率行列 Q によるマルコフ決定過程を $\{S, A, Q, \boldsymbol{r}\}$ と表すことにし, Q を区間表現によって推定してその区間型推移確率行列を $\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle$ とおく, ただし, この行列集合は

$$\begin{aligned} &\underline{Q} = (\underline{q}_{ij}(a) : i, j \in S, a \in A) \in \boldsymbol{R}_+^{kn \times n}, \\ &\overline{Q} = (\overline{q}_{ij}(a) : i, j \in S, a \in A) \in \boldsymbol{R}_+^{kn \times n}, \\ &\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle = \left\{ Q \in P(S|S \times A) \middle| \underline{Q} \leqq Q \leqq \overline{Q} \right\}. \end{aligned} \tag{13}$$

によって表される. 今, $\mathcal{Q}$ によって構成されるマルコフ決定過程を $\{S, A, \mathcal{Q}, \boldsymbol{r}\}$ と表し, 区間によって表現されたマルコフ過程と呼ぶことにしよう. このとき, 評価関数も区間によって表現され次のように表される.

$$\phi(f|\mathcal{Q}) = \left\{ \phi(f|Q) \middle| Q \in \mathcal{Q} \right\} \subset \boldsymbol{R}_+^n \tag{14}$$

ただし, $\phi(f|Q)$ は通常の推移確率行列からなるマルコフ決定過程での価値関数 (式 (11)) である.

次に, $\phi(f|\mathcal{Q})$ の性質についてまとめておく.

まず, $\phi(f|\mathcal{Q}) \in C(\boldsymbol{R}_+)^n$ が成り立つ. 写像 $\mathcal{L} : C(\boldsymbol{R}_+)^n \to C(\boldsymbol{R}_+)^n$ を次のように定義する.

$$\mathcal{L}(f)\boldsymbol{v} = \boldsymbol{r}(f) + \beta\mathcal{Q}(f)\boldsymbol{v}, \ \boldsymbol{v} \in C(\boldsymbol{R}_+)^n, \tag{15}$$

ただし $\mathcal{Q}(f) = \langle \underline{Q}(f), \overline{Q}(f) \rangle, \underline{Q}(f) = (\underline{q}_{ij}(f(i))) \in \boldsymbol{R}_+^{n \times n}, \overline{Q}(f) = (\overline{q}_{ij}(f(i))) \in \boldsymbol{R}_+^{n \times n}$. このとき Lemma 2.1 から, $\mathcal{L}(f)\boldsymbol{v} \in C(\boldsymbol{R}_+)^n$ $(\boldsymbol{v} \in C(\boldsymbol{R}_+)^n)$ であることもわかる. また, $\underline{L}(f) : \boldsymbol{R}_+^n \to \boldsymbol{R}_+^n, \overline{L}(f) : \boldsymbol{R}_+^n \to \boldsymbol{R}_+^n$ とそれぞれ定義すると, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \boldsymbol{R}_+^n$ に対して以下の等式がそれぞれ成り立つ.

$$\underline{L}(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta \min_{Q \in \mathcal{Q}(f)} Q\boldsymbol{x} \tag{16}$$

$$\overline{L}(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta \max_{Q \in \mathcal{Q}(f)} Q\boldsymbol{x}. \tag{17}$$

よって, 以下を得る.

Lemma 3.2 任意の確定的定常政策 $f \in F$ に対して,

(i) $\mathcal{L}(f)$ は単調増加かつ縮小写像である.

(ii) $\underline{L}(f)$ と $\overline{L}(f)$ はともに sup-ノルムに関して単調増加かつ縮小写像である.

Lemma 3.1 と Lemma 3.2 から次を得る.

Theorem 3.1 任意の確定的定常政策 $f \in F$ に対して,

(i) $\phi(f|\mathcal{Q}) \in C(\boldsymbol{R}_+)^n$ であって $\phi(f|\mathcal{Q})$ は $\mathcal{L}(f)$ のただ一つの不動点である. さらに, 任意の $\boldsymbol{v} \in C(\boldsymbol{R}_+)^n$, に対して

$$\mathcal{L}(f)^\ell \boldsymbol{v} \to \phi(f|\mathcal{Q}) \ (\ell \to \infty).$$

が成り立つ.

(ii) $\phi(f|\mathcal{Q}) = [\underline{\phi}(f), \overline{\phi}(f)]$ とおく. このとき, $\underline{\phi}(f)$ と $\overline{\phi}(f)$ はそれぞれ $\underline{L}(f)$ と $\overline{L}(f)$ の不動点である.

$\boldsymbol{u} \in C(\boldsymbol{R}_+)^n$ に対して,

$$\mathcal{L}(\boldsymbol{u}) := (\mathcal{L}(\boldsymbol{u})_1, \mathcal{L}(\boldsymbol{u})_2, \ldots, \mathcal{L}(\boldsymbol{u})_n)', \tag{18}$$

とおく, ただし $\mathcal{L}(\boldsymbol{u})_i := \mathrm{eff}(\{r(i,a) + \beta \mathcal{Q}_{i,a}\boldsymbol{u} | a \in A\})$ $(i \in S)$ であってこれは有効点の集合を表す.

Lemma 3.3 *2* つの確定的定常政策 $f, g \in F$ に対して, もし $\phi(f|\mathcal{Q}) \prec \mathcal{L}(g)\phi(f|\mathcal{Q})$ が成り立つとき,

$$\phi(f|\mathcal{Q}) \prec \phi(g|\mathcal{Q})$$

を得る.

Theorem 3.2 $f^* \in F$ がパレート最適であるための必要かつ十分条件は次の最適包含関係の最大解であることである

$$\boldsymbol{u} \in \mathcal{L}(\boldsymbol{u}), \boldsymbol{u} \in C(\boldsymbol{R}_+)^n. \tag{19}$$

4 区間推定された MDP の考察

ここでは, 区間推定されたマルコフ決定過程の構成とパーセンタイル型の評価について考察する.

De Robertis と Hartigan のベイズ的区間推定手法の結果 ([25]) を, 我々の推移確率行列未知のマルコフ決定過程での真の推移確率行列の区間推定表現に適用する. 具体的には, 推移確率行列の各行において, その行ベクトルの成分の推定を行う. よって, $P_n := P(S) = \{p = (p_1, p_2, \ldots, p_n) | p_i \geqq 0, \sum_{i=1}^n p_i = 1\}$ とおき, $\mathcal{B}$ を $\boldsymbol{R}^N$ のすべての可測集合の全体とおく. 通常のベイズ推定では, 事前分布として事後分布の表現に便利な分布族を用いるが, ここでは, 事前分布として、$\mathcal{B}$ 上の区間で表現可能な測度の集合を事前分布として扱う. それらは, 次のようにして得られる. $\mathcal{B}$ 上の 2 つの測度 L と U がすべての可測集合 $A \in \mathcal{B}$ に対して $L(A) \leq U(A)$ であるとき, 単に $L \leq U$ と表すことにする. また, $[L, U]$ はすべての $A \in \mathcal{B}$ に対して $L(A) \leq Q(A) \leq U(A)$ を満たす測度 Q の凸集合であって, これを事前測度として扱う.

以下では, 議論の単純化のため $[L, kL]$ $(k \geqq 1)$ を仮定する, ただし $L(\cdot)$ は P_n 上のルベーグ測度である.

事前測度 $[L, U]$ によって, 推移確率行列の各成分 p_i は次のような区間 $[\underline{\lambda}_i, \overline{\lambda}_i]$ $(i \in S)$ によって事後推定され, とくにこの両端点は次の積分比の区間表現として得られる.

$$\left\{ \int_{P_n} p_i Q(dp) / \int_{P_n} Q(dp) \middle| L_\sigma \leqq Q \leqq U_\sigma \right\}, \tag{20}$$

ただし, L_σ と U_σ はそれぞれ観測データ $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ に対する事後測度の下限と上限であり, 以下の方程式 (21) と (22) のただ一つの解にそれぞれなっている.

Theorem 4.1 下限 $\underline{\lambda}_i$ と上限 $\overline{\lambda}_i$ はそれぞれ次の方程式のただ一つの解である.

$$U_\sigma(p_i - \underline{\lambda}_i)^- + L_\sigma(p_i - \underline{\lambda}_i)^+ = 0, \tag{21}$$

$$U_\sigma(p_i - \overline{\lambda}_i)^+ + L_\sigma(p_i - \overline{\lambda}_i)^- = 0, \tag{22}$$

ただし, $Q(f)$ は可測関数 f の測度 Q に関する積分値を表し, $x^+ = \max\{0, x\}, x^- = x - x^+ = \min\{0, x\}$ である.

α-パーセンタイル型の区間推定に関しては次のようになる. $\underline{g}_{i,a}$ と $\overline{g}_{i,a}$ を P_n 上の次のような可測関数とおく.

$$\underline{g}_{i,a}(p) = I_{\{p_i \leq a\}}(p), \quad \overline{g}_{i,a}(p) = I_{\{p_i \geq a\}}(p), \tag{23}$$

ただし, I_A は A の指示関数で $I_A(x) = 1$ if $x \in A$, $= 0$ if $x \notin A$ である. また,

$$\underline{\lambda}(a|\sigma) = \sup\left\{\frac{Q_\sigma(\underline{g}_{i,a})}{Q_\sigma(I_{P_n})} | Q_\sigma \in [L_\sigma, kL_\sigma]\right\}, \tag{24}$$

$$\overline{\lambda}(a|\sigma) = \sup\left\{\frac{Q_\sigma(\overline{g}_{i,a})}{Q_\sigma(I_{P_n})} | Q_\sigma \in [L_\sigma, kL_\sigma]\right\}. \tag{25}$$

とおく. $\underline{p}_i(\alpha)$ と $\overline{p}_i(\alpha)$ を次のように定義する.

$$\underline{\lambda}(\underline{p}_i(\alpha)|\sigma) = \alpha, \quad \overline{\lambda}(\overline{p}_i(\alpha)|\sigma) = \alpha. \tag{26}$$

このとき, $\underline{p}_i(\alpha)$ と $\overline{p}_i(\alpha)$ は次を満たす. $B(\alpha, \beta)$ はベータ関数, $B(\alpha, \beta|x)$ は不完全ベータ関数を表す.

Theorem 4.2 $\underline{p}_i(\alpha)$ と $\overline{p}_i(\alpha)$ について

$$\frac{B(s, t|\underline{p}_i(\alpha))}{B(s, t)} = \frac{\alpha}{\alpha + (1-\alpha)k}, \quad \frac{B(s, t|\overline{p}_i(\alpha))}{B(s, t)} = \frac{(1-\alpha)k}{\alpha + (1-\alpha)k} \tag{27}$$

が成り立つ.

例えば, 各行ごとの状態推移の回数を数えて次のような観測結果 Σ を得たとき

$$\Sigma = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix}$$

このデータセットから、$\alpha = 0.05$ としたときの区間推定された推移確率行列は

$$\begin{pmatrix} [0.095, 0.837] & [0.013, 0.641] & [0.045, 0.750] \\ [0.013, 0.641] & [0.095, 0.837] & [0.045, 0.750] \\ [0.011, 0.595] & [0.039, 0.701] & [0.140, 0.860] \end{pmatrix}$$

を得る. また, これをもとにした価値関数の区間表現は

$$\phi(f|\mathcal{Q}(f)) = ([14.0878, 27.9643], [11.8025, 25.66], [12.9396, 26.656])$$

である.

[α-パーセンタイル型評価の課題]

真の推移確率行列が既知の場合には, value $\phi(f|\mathcal{Q})$ の閾値評価問題として次のような考察が可能である (cf. J.Filar, D.Krass, K.W. Ross (1995) など) :

$$\mathrm{Prob}_u(\phi \geqq \tau | X_1 = s_1) \geqq \alpha$$

を満たす価値関数の値が τ 以上であることの政策 u の存在性の議論を行い, 最適解ではなくあるレベル α 以上の政策 u の存在において最適化問題を議論する. これらでは, Occupation measure や parametric LP での考察が最適政策を導出可能かさらに検討が必要となる. 一方, 推移法則未知のモデルでの α-percentile の意味は、統計数学での検出力に関するアプローチでもあり, 価値関数の区間表現とそれをもとにした意思決定のもつ意味についてさらに検討が必要で、これらが今後の課題の一つである.

References

[1] Bertsekas, D.P. and Shreve, S.E. (1978). *Stochastic Optimal Control.* New York: Academic Press.

[2] Dynkin, E.B. and Yushkevich A.A. (1979). *Controlled Markov Processes and their Applications.* New York - Berlin: Springer-Verlag.

[3] Ferguson, T.S. (1967). *Mathematical Statistics.* New York - London: Academic Press.

[4] DeGroot, M.H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill Book Co.

[5] Furukawa, N. and Iwamoto, S. (1970). Stopped decision processes on complete separable metric spaces. *J. Math. Anal. Appl.*, 31:615–658.

[6] van Hee, K.M. (1978). *Bayesian Control of Markov Chains.* Mathematical Centre Tracts, No. 95. Amsterdam: Mathematisch Centrum.

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

[8] Hernández-Lerma, O. and Marcus, S.I. (1985). Adaptive control of discounted Markov decision chains. *Journal of Optimization Theory and Applications* 46: 227–235.

[9] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*, volume 79 of *Applied Mathematical Sciences.* New York: Springer-Verlag.

[10] Hernandez-Lerma, O. and Lasserre, J.B. (1996). *Discrete-time Markov Control Processes.* New York: Springer.

[11] Hordijk, A. (1974). *Dynamic Programming and Markov Potential Theory.* Mathematical Centre Tracts, No. 51. Amsterdam: Mathematisch Centrum.

[12] Horiguchi, M. (2015). Adaptive methods for multivariate Bayesian Control Chart (II), *RIMS kokyuroku No. 1939 (In Japanese)*, 152–161.

[13] Horiguchi, M. (2015). Bayesian Inference in Markov Decision Processes. In, *Modern Trends in Controlled Stochastic Processes: Theory and Applications, Vol. 2 (A.B. Piunovskiy ed.)*, Luniver Press, 177–189.

[14] Horiguchi, M. and Piunovskiy, A.B. (2013). Optimal stopping model with unknown transition probabilities. *Control Cybernet.*, 42(3):593–612.

[15] Iki, T., Horiguchi, M., Yasuda, M. and Kurano, M. (2009). An interval Bayesian Method for uncertain MDPs.(In Japanese), *RIMS Kokyuroku* 1636:1–8.

[16] Masami Kurano, Jinjie Song, Masanori Hosaka, and Youqiang Huang. Controlled Markov set-chains with discounting. *J. Appl. Probab.*, 35(2):293–302, 1998.

[17] Kurano, M. (1972). Discrete-time Markovian decision processes with an unknown parameter. Average return criterion. *Journal of the Operations Research Society of Japan* 15: 67–76.

[18] Kurano, M. (1983). Adaptive policies in Markov decision processes with uncertain transition matrices. *Journal of Information and Optimization Sciences* 4: 21–40.

[19] Makis, V. (2008). Multivariate Bayesian control chart. *Oper. Res.*, 56(2):487–496, 2008.

[20] Mandl, P. (1974). Estimation and control in Markov chains. *Advances in Applied Probability* 6: 40–60.

[21] Martin, J.J. (1967). *Bayesian Decision Problems and Markov Chains.* Publications in Operations Research, No. 13. New York: John Wiley & Sons Inc.

[22] Martin L. Puterman. *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons Inc., New York, 1994. A Wiley-Interscience Publication.

[23] Raiffa, H and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Studies in Managerial Economics. Division of Research, Graduate School of Business Administration, Harvard University, Boston, Mass.

[24] Rieder, U. (1975). Bayesian Dynamic Programming. *Advances in Applied Probability* 7: 330–348.

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

[26] Ross, S.M. (1970). *Applied Probability Models with Optimization Applications.* San Francisco: Holden-Day.

[27] Ross, S.M. (1983). *Introduction to Stochastic Dynamic Programming.* San Diego, CA: Academic Press.

[28] Wald, A. (1950). *Statistical Decision Functions.* New York: John Wiley & Sons Inc.

[29] White, D.J. (1969). *Dynamic Programming.* Mathematical Economic Texts, 1. Edinburgh-London: Oliver & Boyd.