数理解析研究所講究録
第1682巻 2010年 70-77

未知の推移確率行列の事前・事後区間表現とマルコフ決定過程について (Bayes estimated intervals and uncertain MDPs)

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

1　はじめに

推移確率行列が未知のマルコフ決定過程 (Uncertain MDPs) において, 状態推移の観測から得られる情報に基づいて推移確率行列の推定を行い, モデルの価値関数 (value function) を予測する. 我々の先行研究 ([7]) で, 区間ベイズ法 (DeRobertis& Hartigan[1]) を適用することで, 事前測度区間から推移確率行列を事後区間として推定し, 区間推定 MDPs の構成とその解析について述べた. 例えば, 推移確率行列 $P=(p_{ij})$ の推定は, その成分が $[\underline{\lambda}_{ij}, \overline{\lambda}_{ij}]$ と区間表現される.

本発表では, ベータ関数の多項方程式の解として得られる区間の下限値 $\underline{\lambda}_{ij}$ と上限値 $\overline{\lambda}_{ij}$ について, 不完全ベータ関数の性質と Newton-Raphson 法から近似解を得る方法について考察する. また, この解法と分数計画問題との関係や解の収束の具体例について示す.

ここで扱う区間推定 MDPs は, 定常なマルコフ集合連鎖 (Markov set-chain) によって構成される. 区間で表される推移確率行列が各期で変動する "Controlled Markov set-chain model" については, Kurano et al.[8, 9] などを参照されたい.

2　準備

有限状態マルコフ決定過程は

$$\{S, A, Q, \boldsymbol{r}\}$$

の 4 つから成り立つ. 状態空間を $S:=\{1,2,\ldots,n\}$, 決定空間を $A:=\{a_1,a_2,\ldots,a_k\}$ とする. 次の集合を定義する:

$$P(S) :=\{p=(p_1,p_2,\ldots,p_n)\in\mathbb{R}_+^n \,|\, \sum_{i\in S} p_i=1\}, \tag{2.1}$$

$$P(S|S) :=\{q=(q_{ij}: i,j\in S)\in\mathbb{R}_+^{n\times n} \,|\, \sum_{j\in S} q_{ij}=1 \ (i\in S)\}, \tag{2.2}$$

$$P(S|S\times A) :=\{Q=(q_{ij}(a): i,j\in S, a\in A)\in\mathbb{R}_+^{kn\times n} | q_{i\cdot}(a)\in P(s) \ (i\in S, a\in A)\}. \tag{2.3}$$

ただし, $\mathbb{R}_+^n, \mathbb{R}_+^{m\times n}$ は, それぞれ非負の n 次元実ベクトルと (m,n) 実行列を表す. 推移確率行列を $Q=(q_{ij}(a))\in P(S|S\times A)$, 利得関数を $\boldsymbol{r}=(r(i,a))\in B_+(S\times A)$ で表す. ただし, $B_+(D)$ は集合 D 上の非負実数値関数の全体を表す. 我々は, 推移確率行列 $Q=(q_{ij}(a))$ $(i,j\in S, a\in A)$ が未知であるマルコフ決定過程 (Uncertain MDPs) を考える. 推移確率行列の推定を取り入れたマルコフ決定過程として, 最尤推定法 (cf. [2, 5, 6, 11]) やベイズ推定法 (cf. [5, 12, 17]) が良く知られている. 本研究では, ベイズ推定法において, その事前分布の設定に区間測度を用いる.

本発表では, 区間推定 MDPs の構成を簡単にするために, 確定的 (deterministic) かつ定常 (stationary) な政策のもとでのマルコフ決定過程について考察する. 従って, 以後, ある固定された deterministic stationary policy での推定される推移確率行列を $P=(p_{ij})$ と表すことにする.

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:

$$\phi(f|Q)=\sum_{t=0}^{\infty}(\beta Q(f))^t\boldsymbol{r}(f), \tag{2.4}$$

ただし, $\boldsymbol{r}(f)=(r(1,f(1)),r(2,f(2)),\ldots,r(n,f(n)))'\in\mathbb{R}_+^n, Q(f)=(q_{ij}(f(i)))\in P(S|S)$.

区間推定 MDPs の構成について, 推移確率行列は各行ごとに区間ベイズ推定を行う. 従って, ここでは, ある固定された状態から次の期に推移する確率 $\{p_i\}_{i\in S}$ について議論して行く.

$P_n := P(S) = \{p = (p_1, p_2, \ldots, p_n) | p_i \geq 0, \sum_{i=1}^n p_i = 1\}$ とおく. $\mathbb{R}^n$ のルベーグ可測集合の全体を $\mathcal{B}$ で表す. $\mathcal{B}$ 上の 2 つの測度 L, U が任意の $A \in \mathcal{B}$ に対して $L(A) \leqq U(A)$ であるとき, 単に $L \leqq B$ と表すことにする. このような $L \leqq U$ である 2 つの測度を用いて, 事前区間測度を $[L, U]$ と表す.

ここで, P_n 上のルベーグ測度 $L(\cdot)$ から, 上側の測度 U は, $U(\cdot) = kL(\cdot)$ となるような測度 L の $k(k \geqq 1)$ に関する比例測度 (proportional measure) と仮定する, すなわち, 事前区間測度を, $[L, kL] = [dp, k\,dp]$ とする.

独立試行実験を $\hat{\sigma}$ 回行い, 次の期に状態 i へ推移した回数を σ_i で表すとする. この時, それぞれの状態に関するデータを $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ と表すと $\hat{\sigma} = \sum_{k=1}^n \sigma_k$ である. 状態 i へ推移する確率を p_i とするとき, σ の確率密度関数はパラメータ $p = (p_1, p_2, \ldots, p_n)$ に関する多項分布で表されて

$$f(\sigma_1, \sigma_2, \ldots, \sigma_n | p) = \frac{(\sigma_1 + \cdots + \sigma_n)!}{\sigma_1! \cdots \sigma_n!} p_1^{\sigma_1} p_2^{\sigma_2} \cdots p_n^{\sigma_n} \tag{2.5}$$

となる.

ここで, DeRobertis/Hartigan[1] の区間ベイズ推定の手法を事前区間 $[L, kL]$ に適用すると, 事後測度区間 $[L_\sigma, U_\sigma] := [L_\sigma, kL_\sigma]$ は多次元ベータ関数によって表され, さらに p_i に関する事後測度区間 $[\underline{\lambda}_i, \overline{\lambda}_i]$ (簡単のため $[\underline{\lambda}, \overline{\lambda}]$ と表すことにする) は, 事後区間測度 $Q \in [L_\sigma, kL_\sigma]$ による次のような p_i との積分比の範囲から作られる:

$$\left\{ \frac{\int_{P_n} p_i Q(dp)}{\int_{P_n} Q(dp)} \middle| L_\sigma \leqq Q \leqq U_\sigma \right\} \tag{2.6}$$

さらに, $[\underline{\lambda}, \overline{\lambda}]$ はそれぞれ次の方程式の一意の解である.

$$U_\sigma (p_i - \underline{\lambda})^- + L_\sigma (p_i - \underline{\lambda})^+ = 0 \tag{2.7}$$

$$U_\sigma (p_i - \overline{\lambda})^+ + L_\sigma (p_i - \overline{\lambda})^- = 0 \tag{2.8}$$

ただし, $x^+ = \max\{0, x\}, x^- = x - x^+ = \min\{0, x\}$ である. ここで, $U_\sigma = kL_\sigma$ であることと多次元ベータ関数 (ディリクレ関数) を用いれば, 上述の (2.7),(2.8) は次のようにディリクレ積分による方程式として表すことができる:

$$\text{(lower bound } \underline{\lambda}\text{):}\quad k \int_{0 \leq p_i \leq \lambda, p \in P_n} \cdots \int (p_i - \lambda) p_1^{\sigma_1} \cdots p_n^{\sigma_n} dp + \int_{\lambda \leq p_i \leq 1, p \in P_n} \cdots \int (p_i - \lambda) p_1^{\sigma_1} \cdots p_n^{\sigma_n} dp = 0 \tag{2.9}$$

$$\text{(upper bound } \overline{\lambda}\text{):}\quad k \int_{\lambda \leq p_i \leq 1, p \in P_n} \cdots \int (p_i - \lambda) p_1^{\sigma_1} \cdots p_n^{\sigma_n} dp + \int_{0 \leq p_i \leq \lambda, p \in P_n} \cdots \int (p_i - \lambda) p_1^{\sigma_1} \cdots p_n^{\sigma_n} dp = 0 \tag{2.10}$$

ガンマ関数 $\Gamma(x)$ $(x > 0)$, ベータ関数 $B(x, y)$ $(x, y > 0)$, 不完全ベータ関数 $B(x, y|\lambda)$ $(x, y > 0, 0 \leqq \lambda \leqq 1)$ をそれぞれ次のように定義する:

ガンマ関数: $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \quad (x > 0)$

ベータ関数: $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1}\, dt \quad (x, y > 0)$

不完全ベータ関数: $B(x, y|\lambda) = \int_0^\lambda t^{x-1} (1-t)^{y-1} dt \quad (x, y > 0, 0 \leqq \lambda \leqq 1)$

また, ディリクレ積分 $D(\nu_1, \nu_2, \ldots, \nu_k; \nu_{k+1})$ および $D(\nu_1, \nu_2, \ldots, \nu_k; \nu_{k+1}|\lambda)$ $(k \geqq 1, 0 \leqq \lambda \leqq 1)$ を次のように定める.

$$\begin{aligned} &D(\nu_1, \nu_2, \ldots, \nu_k; \nu_{k+1}) \\ &:= \int \cdots \int_{S_k} x_1^{\nu_1 - 1} x_2^{\nu_2 - 1} \cdots x_k^{\nu_k - 1} (1 - x_1 - x_2 - \cdots - x_k)^{\nu_{k+1} - 1}\, dx_1 dx_2 \cdots dx_k \\ &D(\nu_1, \ldots, \nu_k; \nu_{k+1}|\lambda) \\ &:= \int \cdots \int_{S_k \cap \{0 \leqq x_1 \leq \lambda\}} x_1^{\nu_1 - 1} \cdots x_k^{\nu_k - 1} (1 - x_1 - \cdots - x_n)^{\nu_{k+1} - 1}\, dx_1 \cdots dx_k \end{aligned}$$

ただし, $S_k := \{(x_1, \ldots, x_k) : x_i \geqq 0, i = 1, \ldots, k, \sum_{i=1}^{k} x_i \leqq 1\}$.

ここで, ディリクレ積分とベータ関数の関係から

$$D(\nu_1, \nu_2, \ldots, \nu_k; \nu_{k+1}) = B(\nu_1, \nu_2 + \cdots + \nu_{k+1}) D(\nu_2, \nu_3, \ldots, \nu_k; \nu_{k+1}) \tag{2.11}$$

$$D(\nu_1, \nu_2, \ldots, \nu_k; \nu_{k+1}|\lambda) = B(\nu_1, \nu_2 + \cdots + \nu_{k+1}|\lambda) D(\nu_2, \nu_3, \ldots, \nu_k; \nu_{k+1}) \tag{2.12}$$

であって, さらにベータ関数の性質を利用して, 式 (2.9) と式 (2.10) は次の λ に関する多項方程式の解であることが示される.

$$K(s,t,\lambda) := \left(\frac{s}{s+t} - \lambda\right) B(s,t) + (k-1)\left(B(s+1,t|\lambda) - \lambda B(s,t|\lambda)\right) = 0 \tag{2.13}$$

$$G(s,t,\lambda) := k\left(\frac{s}{s+t} - \lambda\right) B(s,t) - (k-1)\left(B(s+1,t|\lambda) - \lambda B(s,t|\lambda)\right) = 0 \tag{2.14}$$

と表される. ただし, $\hat{\sigma} = \sum_{i=1}^{n} \sigma_i, s = \sigma_1 + 1, t = \hat{\sigma} - \sigma_1 + n - 1$ である. この式 (2.13) と (2.14) は, 具体的には $(\hat{\sigma} + n)$ 次多項方程式である $(s + t = \hat{\sigma} + n)$. 以下に定理としてまとめておく:

Theorem 2.1. パラメータ $p = (p_1, p_2, \ldots, p_n)$ について, 各 p_i に関する事後測度区間 $[\underline{\lambda}, \overline{\lambda}]$ は次の多項方程式の解として得られる.

$$K(s,t,\underline{\lambda}) := B(s+1,t) - \underline{\lambda} B(s,t) + (k-1)\left(B(s+1,t,\underline{\lambda}) - \underline{\lambda} B(s,t,\underline{\lambda})\right) = 0 \tag{2.15}$$

$$G(s,t,\overline{\lambda}) := k(B(s+1,t) - \overline{\lambda} B(s,t)) - (k-1)\left(B(s+1,t,\overline{\lambda}) - \overline{\lambda} B(s,t,\overline{\lambda})\right) = 0 \tag{2.16}$$

ただし, $s = \sigma_i + 1, t = \hat{\sigma} - \sigma_i + n - 1$ である.

$K(s,t,\lambda), G(s,t,\lambda)$ は, 変数 λ に関して, ともに狭義単調関数で $K(s,t,\lambda)$ は上に凸, $G(s,t,\lambda)$ は下に凸であり, K, G はともに $[0,1]$ に 1 つの解をもつことが容易に示される.

3 $K(s,t,\lambda) = 0$ と $G(s,t,\lambda) = 0$ の解

前節の式 (2.15) と (2.16) はともに $(\hat{\sigma} + n)$ 次多項方程式である. 本節では, 関数 K, G の狭義単調性と凸性から Newton-Raphson 法を適用したアルゴリズムを示す. さらに, 漸化式を与える関数 ϕ と分数計画問題 (Fractional programming problem) との関係についても述べる.

以後, $K(s,t,\lambda) = 0$ の解 $\underline{\lambda}$ について述べる. $G(s,t,\lambda) = 0$ の解 $\overline{\lambda}$ も同様のアルゴリズムと性質を持つことが容易に示される.

Proposition 3.1. $K(s,t,\lambda)$ は次の性質をもつ:

- $K(s,t,0) = B(s+1,t) = \frac{s}{s+t} B(s,t) > 0, K(s,t,1) = k\left(B(s+1,t) - B(s,t)\right) = -\frac{kt}{s+t} B(s,t) < 0$
- $K'(s,t,\lambda) = -B(s,t) - (k-1)B(s,t|\lambda) < 0,$
- $K''(s,t,\lambda) = -(k-1)\lambda^{s-1}(1-\lambda)^{t-1} < 0$

この命題から, 初期値として $K(s,t,\lambda) < 0$ となる $\lambda (0 < \lambda < 1)$ を選び Newton-Raphson 法を適用したアルゴリズム (**Algorithm A**) は以下のようになる.

Algorithm A:

Step 1. Set $n := 0$ and specify $\varepsilon > 0$. Select $\lambda (0 < \lambda < 1)$ such that $K(s,t,\lambda) < 0$. Set $x_n := \lambda$.

2. Let $W(s,t,x_n) := \dfrac{K(s,t,x_n)}{K'(s,t,x_n)}$. Compute $x_{n+1} := x_n - W(s,t,x_n)$.

3. If $|x_{n+1} - x_n| < \varepsilon$, set $\underline{\lambda}_i := x_{n+1}$ and stop. Otherwise increase n by 1 and go back to Step 2.

ここで, $W(s,t,x_n)=\dfrac{K(s,t,x_n)}{K'(s,t,x_n)}$ は具体的には

$$W(s,t,x_n):=-\frac{(\frac{s}{s+t}-x_n)B(s,t)+(k-1)(\frac{s}{s+t}-x_n)B(s,t,x_n)-\frac{k-1}{s+t}x_n^s(1-x_n)^t}{B(s,t)+(k-1)B(s,t,x_n)} \tag{3.1}$$

と, ベータ関数 $B(s,t)$ と不完全ベータ関数 $B(s,t,x_n)$ によって表されるが, この計算は煩雑である.

一般に, Nerwton-Raphson 法の漸化式 $x_{n+1}=x_n-\dfrac{f(x_n)}{f'(x_n)}$ について, $\phi(x)=x-\dfrac{f(x)}{f'(x)}$ とすれば $f'(\alpha)\neq 0$ である α について, $f(\alpha)=0$ であることと $\phi(\alpha)=\alpha$ であることは同値である. さらに, K について次の性質を得る:

Proposition 3.2.

$$K(s,t,\lambda)=\lambda K'(s,t,\lambda)-K'(s+1,t,\lambda) \tag{3.2}$$

$$\frac{K(s,t,\lambda)}{K'(s,t,\lambda)}=\lambda-\frac{K'(s+1,t,\lambda)}{K'(s,t,\lambda)} \tag{3.3}$$

上述の命題から,

$$\phi(\lambda)=\frac{K'(s+1,t,\lambda)}{K'(s,t,\lambda)} \tag{3.4}$$

であって,

$$x_{n+1}=\phi(x_n)=\frac{K'(s+1,t,x_n)}{K'(s,t,x_n)}=\frac{B(s+1,t)+(k-1)B(s+1,t,x_n)}{B(s,t)+(k-1)B(s,t,x_n)} \tag{3.5}$$

を得る. また, $\phi(\lambda)$ の 1 次導関数について

$$\phi'(\lambda)=\frac{-(k-1)\lambda^{s-1}(1-\lambda)^{t-1}}{K'(s,t,\lambda)}(\lambda-\phi(\lambda)) \tag{3.6}$$

であることから, 次の命題と定理を得る:

Proposition 3.3. $\phi(\lambda)=K'(s+1,t,\lambda)/K'(s,t,\lambda)$ とするとき, 次が成り立つ.

(i) $\phi(0)=\phi(1)=B(s+1,t)/B(s,t)$

(ii) $\phi'(\lambda)=0$ となる λ は $\lambda=0,1$ と $\phi(\alpha)=\alpha$ を満たす α である, すなわち, この α は $K(s,t,\lambda)=0$ の解である.

(iii) $\phi(\lambda)$ の増減は $0<\lambda<\alpha$ で狭義単調減少, $\alpha<\lambda<1$ で狭義単調増加である.

Theorem 3.1. 次は同値である. *(i)* $K(s,t,\alpha)=0$ *(ii)* $\phi(\alpha)=\alpha$

上述の命題と定理から $y=\phi(x)$ のグラフ (図 3.1) の特徴を考えると, Newton-Raphson 法を適用する初期値 x_1 は $[0,1]$ 上に任意に取ることができる. 特に, $\underline{\lambda}<x_n\Rightarrow\underline{\lambda}<x_{n+1}=\phi(x_n)<x_n$ であるから点列 $\{x_n\}$ は $n=2$ 以降で $\phi(\lambda)$ の不動点 $\underline{\lambda}$ に単調に収束する.

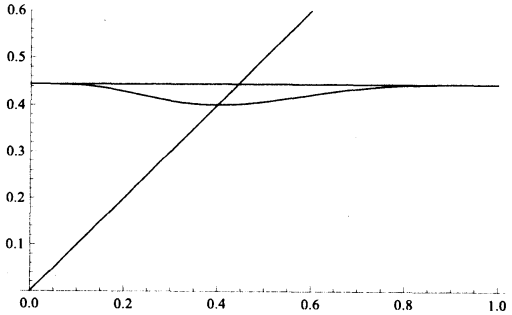

Figure 3.1: $y=\phi(x), y=x, y=B(s+1,t)$ のグラフ

次に, $\phi(\lambda)$ とその不動点 $\phi(\underline{\lambda}) = \underline{\lambda}$ について, $K(s,t,\lambda)$ の性質 (命題 3.2) から分数計画問題 (fractional programming problem) とそのパラメトリック問題 (parametric problem) の解との関係を次のように示せる.

Fractional programming problem: (P) $\min_{\lambda\in[0,1]} \dfrac{K'(s+1,t.\lambda)}{K'(s,t,\lambda)}$,

Parametric problem: (P_q) $F(q) = \min_{\lambda\in[0,1]}(K'(s+1,t,\lambda) - qK'(s,t,\lambda))$ とするとき,

Proposition 3.4.

(i) $\phi(\underline{\lambda}) = \min_{\lambda\in[0,1]} \dfrac{K'(s+1,t,\lambda)}{K'(s,t,\lambda)} = \dfrac{K'(s+1,t,\underline{\lambda})}{K'(s,t,\underline{\lambda})} = \underline{\lambda}$

(ii) $F(p)$ は狭義単調減少, 凹関数である.

(iii) $\underline{\lambda}$ は $F(p) = 0$ を満たす $p \in [0,1]$ の一意の解である.

上述の命題 3.4 から, 多項方程式 $K(s,t,\lambda) = 0$ の解と分数計画問題 (P), パラメトリック問題 (P_q) には次のような関係があることがわかる:

Corollary 3.1. $F(\underline{\lambda}) = K'(a+1,b,\underline{\lambda}) - \underline{\lambda}K'(a,b,\underline{\lambda}) = -K(a,b,\underline{\lambda}) = 0$

以上のことは, $G(s,t,\lambda) = 0$ の解として得られる事後測度区間の上限値 $\overline{\lambda}$ についても同様の議論ができる. すなわち,

$$\psi(\lambda) := \frac{G'(s+1,t,\lambda)}{G'(s,t,\lambda)} = \frac{kB(s+1,t) - (k-1)B(s+1,t,\lambda)}{kB(s,t) - (k-1)B(s,t,\lambda)} \tag{3.7}$$

とするとき, $\psi(\overline{\lambda}) = \psi'(\overline{\lambda}) = \overline{\lambda}$ であって, $\psi(\lambda)$ は $\psi(0) = \psi(1) = \frac{ks}{s+t}B(s,t) > 0, 0 < \lambda < \overline{\lambda}$ で狭義単調増加, $\overline{\lambda} < \lambda < 1$ で狭義単調減少である. $y = \psi(x)$ のグラフは図 3.2 のような概形をもつ.

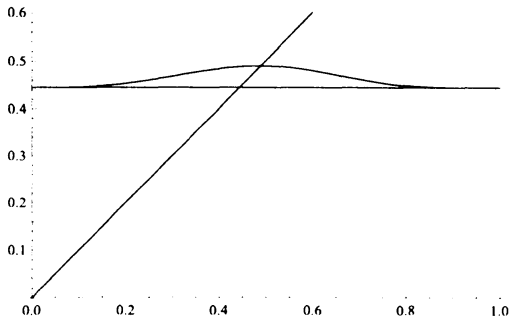

Figure 3.2: $y = \psi(x), y = x, y = kB(s+1,t)$ のグラフ

この節の最後に, $\underline{\lambda}, \overline{\lambda}$ を求めるアルゴリズム (**Algorithm B**) を以下のようにまとめておく.

Algorithm B:

Step 1. Set $m := 0, n := 0$ and specify $\varepsilon > 0$. Select $0 < x, y < 1$ such that $K(s,t,x) < 0$ and $G(s,t,y) > 0$. Set $x_m := x, y_n := y$.

2(a). Compute $x_{m+1} := \phi(x_m)$.

2(b). If $|x_{m+1} - x_m| < \varepsilon$, set $\underline{\lambda} := x_{m+1}$ and go to Step 3(a). Otherwise increase m by 1 and go back to Step 2(a).

3(a). Compute $y_{n+1} := \psi(y_n)$.

3(b). If $|y_{n+1} - y_n| < \varepsilon$, set $\overline{\lambda} := y_{n+1}$ and stop. Otherwise increase n by 1 and go back to Step 3(a).

4 数値例

前節の **Algorithm B** について, 例えば, $s=4,t=5,k=2$ として, $x_1=1,x_{m+1}=\phi(x_m),y_1=0,y_{n+1}=\psi(y_n)$ の反復計算を実行してみると, $\{x_n\}=\{1.,0.444444,0.401901,0.400395,0.400394,0.400394,\ldots\}$, $\{y_n\}=\{0.,0.444444,0.487562,0.48911,0.489112,0.489112,\ldots\}$ を得る. また, 状態数 $n=3$, データ観測数 $\hat{\sigma}=10$, 測度比例定数 $k=2$ としたときの s と t に関して次のような表を得る (表 4.1, ただし $s=\sigma_i+1,t=\hat{\sigma}+n-s$ であることに注意する):

Table 4.1: $n=3,\hat{\sigma}=10,k=2$ のときの p_i の事後測度区間 $[\underline{\lambda},\overline{\lambda}]$

$s=1,t=12$	$[0.060,0.097]$	$s=5,t=8$	$[0.349,0.421]$	$s=9,t=4$	$[0.657,0.726]$
$s=2,t=11$	$[0.129,0.182]$	$s=6,t=7$	$[0.424,0.499]$	$s=10,t=3$	$[0.734,0.799]$
$s=3,t=10$	$[0.201,0.263]$	$s=7,t=6$	$[0.501,0.576]$	$s=11,t=2$	$[0.818,0.871]$
$s=4,t=9$	$[0.274,0.343]$	$s=8,t=5$	$[0.579,0.651]$		

この表 4.1 ようなデータテーブルを用意しておけば, 任意の観測データから区間推定 MDPs を構成することができる. s と t は, $1\leqq s\leqq\hat{\sigma}+1,n-1\leqq t\leqq\hat{\sigma}+n-1$ である. 例えば $n=3,k=2,\hat{\sigma}\leqq 10$ とした時, s,t のそれぞれの値について **Algorithm B** を適用して得られる事後区間の両端点 $\underline{\lambda},\overline{\lambda}$ のデータテーブルを表 4.2 と表 4.3 にまとめておく. また, $k=1.5$ から $k=5$ まで 0.5 ずつ変化させたときの事後区間を表 4.4 と表 4.5 にまとめておく. これらの表を利用して構成される区間推定 MDPs の解析例は, 例えば先行研究の [7] を参照されたい.

Table 4.2: Lower bound for estimated inverval($n=3,k=2,\hat{\sigma}\leqq 10$)

	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
$s=1$	0.268	0.198	0.158	0.131	0.112	0.097	0.087	0.078	0.071	0.065	0.060
$s=2$	0.435	0.344	0.284	0.242	0.211	0.187	0.168	0.153	0.140	0.129	0.119
$s=3$	0.542	0.446	0.379	0.330	0.292	0.262	0.238	0.218	0.201	0.186	0.173
$s=4$	0.615	0.521	0.453	0.400	0.359	0.325	0.297	0.274	0.254	0.237	0.222
$s=5$	0.668	0.579	0.511	0.458	0.414	0.379	0.349	0.323	0.301	0.282	0.265
$s=6$	0.708	0.624	0.558	0.505	0.461	0.424	0.393	0.366	0.343	0.322	0.304
$s=7$	0.740	0.660	0.597	0.545	0.501	0.464	0.432	0.404	0.380	0.358	0.339
$s=8$	0.765	0.690	0.629	0.579	0.535	0.498	0.466	0.438	0.413	0.390	0.370
$s=9$	0.786	0.716	0.657	0.608	0.565	0.529	0.496	0.468	0.443	0.420	0.399
$s=10$	0.804	0.737	0.681	0.633	0.592	0.556	0.524	0.495	0.470	0.447	0.426
$s=11$	0.818	0.755	0.702	0.656	0.615	0.580	0.548	0.520	0.494	0.471	0.450

Table 4.3: Upper bound for estimated inverval($n=3,k=2,\hat{\sigma}\leqq 10$)

	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$
$s=1$	0.404	0.307	0.248	0.208	0.179	0.157	0.140	0.126	0.115	0.105	0.097
$s=2$	0.565	0.458	0.385	0.332	0.292	0.260	0.235	0.214	0.196	0.182	0.169
$s=3$	0.656	0.554	0.479	0.421	0.376	0.340	0.310	0.284	0.263	0.245	0.229
$s=4$	0.716	0.621	0.547	0.489	0.442	0.403	0.371	0.343	0.319	0.298	0.280
$s=5$	0.758	0.670	0.600	0.542	0.495	0.455	0.421	0.392	0.367	0.344	0.325
$s=6$	0.789	0.708	0.641	0.586	0.539	0.499	0.465	0.435	0.408	0.385	0.364
$s=7$	0.813	0.738	0.675	0.621	0.576	0.536	0.502	0.471	0.444	0.420	0.399
$s=8$	0.832	0.762	0.703	0.651	0.607	0.568	0.534	0.504	0.476	0.452	0.430
$s=9$	0.847	0.782	0.726	0.677	0.634	0.596	0.562	0.532	0.505	0.480	0.458
$s=10$	0.860	0.799	0.746	0.699	0.657	0.620	0.587	0.557	0.530	0.506	0.484
$s=11$	0.871	0.814	0.763	0.718	0.678	0.642	0.610	0.580	0.553	0.529	0.507

Table 4.4: Examples of posterior intervals for k :(1)

	$k=1.5$	$k=2$	$k=2.5$	$k=3$
$s=1, t=12$	[0.066,0.088]	[0.060,0.097]	[0.055,0.104]	[0.051,0.110]
$s=2, t=11$	[0.139,0.170]	[0.129,0.182]	[0.121,0.191]	[0.115,0.199]
$s=3, t=10$	[0.213,0.249]	[0.201,0.263]	[0.191,0.274]	[0.184,0.283]
$s=4. t=9$	[0.288,0.328]	[0.274,0.343]	[0.264,0.354]	[0.255,0.364]
$s=5. t=8$	[0.363,0.406]	[0.349,0.421]	[0.337,0.433]	[0.328,0.443]
$s=6. t=7$	[0.440,0.484]	[0.424,0.499]	[0.413,0.511]	[0.403,0.521]
$s=7, t=6$	[0.516,0.560]	[0.501,0.576]	[0.489,0.587]	[0.479,0.597]
$s=8. t=5$	[0.594,0.637]	[0.579,0.651]	[0.567,0.663]	[0.557,0.672]
$s=9. t=4$	[0.672,0.712]	[0.657,0.726]	[0.646,0.736]	[0.636,0.745]
$s=10, t=3$	[0.751,0.787]	[0.737,0.799]	[0.726,0.809]	[0.717,0.816]
$s=11. t=2$	[0.830,0.861]	[0.818,0.871]	[0.809,0.879]	[0.801,0.885]

Table 4.5: Examples of posterior intervals for k :(2)

	$k=3.5$	$k=4$	$k=4.5$	$k=5$
$s=1, t=12$	[0.048,0.116]	[0.045,0.120]	[0.043,0.125]	[0.041,0.128]
$s=2, t=11$	[0.111,0.206]	[0.106,0.212]	[0.103,0.217]	[0.100,0.221]
$s=3, t=10$	[0.178,0.290]	[0.173,0.297]	[0.168,0.303]	[0.164,0.308]
$s=4, t=9$	[0.248,0.372]	[0.242,0.379]	[0.237,0.385]	[0.233,0.391]
$s=5, t=8$	[0.321,0.451]	[0.314,0.459]	[0.309,0.465]	[0.304,0.470]
$s=6, t=7$	[0.395,0.529]	[0.388,0.536]	[0.382,0.542]	[0.377,0.548]
$s=7, t=6$	[0.471,0.605]	[0.464,0.612]	[0.458,0.618]	[0.452,0.623]
$s=8. t=5$	[0.549,0.679]	[0.541,0.686]	[0.535,0.691]	[0.530,0.696]
$s=9. t=4$	[0.628,0.752]	[0.621,0.758]	[0.615,0.763]	[0.609,0.767]
$s=10. t=3$	[0.710,0.822]	[0.703,0.827]	[0.697,0.832]	[0.692,0.836]
$s=11, t=2$	[0.794,0.889]	[0.788,0.894]	[0.783,0.897]	[0.779,0.900]

References

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

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

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

[4] Darald 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] 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.

[7] T. Iki, M. Horiguchi, M. Yasuda, and M. Kurano. An interval bayesian method for uncertain MDPs.(Japanese). *Surikaisekikenkyusyo Kokyuroku*, 1636:1–8, 2009.04.

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

[9] Masami Kurano, Masami Yasuda, and Jun-ichi Nakagami. Interval methods for uncertain Markov decision processes. In *Markov processes and controlled Markov chains (Changsha, 1999)*, pages 223–232. Kluwer Acad. Publ., Dordrecht, 2002.

[10] K. Kuratowski. *Topology. Vol. I.* New edition, revised and augmented. Translated from the French by J. Jaworowski. Academic Press, New York, 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., New York, 1967.

[13] A. M. Ostrowski. *Solution of equations and systems of equations.* Second edition. Pure and Applied Mathematics, Vol. 9. Academic Press, New York, 1966.

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

[15] Moshe Sniedovich. *Dynamic programming*, volume 154 of *Monographs and Textbooks in Pure and Applied Mathematics.* Marcel Dekker Inc., New York, 1992.

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

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

[18] Samuel S. Wilks. *Mathematical statistics.* A Wiley Publication in Mathematical Statistics. John Wiley & Sons Inc., New York, 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.