推移法則未知の区間型マルコフ決定モデルにおける確信集合について (On the credible set in controlled Markov set-chains with unknown transition probabilities)

神奈川大学・理学部・堀口正之
Masayuki HORIGUCHI
Professor of Department of Mathematics,
Faculty of Science,
Kanagawa University

Abstract

推移法則が未知のマルコフ決定過程において, 各期での行動によって生じる推移状態の観測によって推移法則を推定しながら適応型最適政策を求める学習問題がある. その推移法則の推定と最適化問題を構成するとき, 事前区間測度を用いた推定による事後区間測度から区間確率行列が得られる. そのことによって区間型マルコフ決定過程(controlled Markov set-chain)による解決アプローチを適用することができる. 本報告では, α-percentile に基づいて区間推移法則を推定するMDPの構成法と最適化問題を考察する.

1 はじめに

有限マルコフ決定過程(Finite MDPs)は, 次の4つの項目で構成される。

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

ここで, $S = \{1, 2, \ldots, n\}$ は状態空間を表し $A = \{a_1, a_2, \ldots, a_k\}$ は決定空間(行動空間) $Q = (q_{ij}(a))$ の各要素 $q_{ij}(a)$ は状態iにおいて決定aを選択した時の次の期の推移状態がjである確率を表し $q_{ij}(a) \in P(S|S \times A)$ であるような確率核の集まりでQを単に推移確率行列と呼ぶ. $r = r(i, a)$ は $S \times A$ 上の利得関数として定義される. システムの状態が $i \in S$ で $a \in A$ を選択した時, 次の期の状態へは $q(\cdot|(i, a))$ に従って推移し期待利得は $r(i, a)$ となる. また本稿のマルコフ決定過程(uncertain MDPs) では, 推移法則 $Q = (q_{ij}(a))$ が未知である場合を考察する. 状態観測によって未知の推移確率行列の各要素 $q_{i\cdot}(a)$ はそれぞれ区間表現 $[\underline{q}_{i\cdot}(a), \overline{q}_{i\cdot}(a)]$ される. ここでは推移法則推定の具体的な区間値の導出について考察するが, 区間型推移法則をもつマルコフ決定過程（controlled Markov set-chain models) については, Kurano et al. [7]などの先行研究が挙げられる.

推移法則の推定は, 各決定選択に応じた推移結果の状態観測数に基づいて行われるためここでは簡略して推定する行列を $Q = (q_{ij})$ と表すことにする. さらに, マルコフ決定過程での状態推移は, 現在の状態 i に対してその推移法則 $q_{i\cdot}$ に従って次の期の状態が確率的に定まる. そこで, 以後, 事前区間測度を用いて Q の第 i 行目 $q_{i\cdot}$ についてベイズ推定を行う.

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} & \cdots\cdots & q_{1n} \\ q_{21} & q_{22} & q_{23} & \cdots\cdots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ q_{i1} & q_{i2} & q_{i3} & \cdots\cdots & q_{in} \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ q_{n1} & q_{n2} & q_{n3} & \cdots\cdots & q_{nn} \end{pmatrix}$$

現在の状態がiのときに次の期に推移した状態jへの推移回数σ_jを記録したデータセットを $\hat{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ とおく. このとき, $P_n = P(S) = \{p = (p_1, p_2, \ldots, p_n) | p_i \geqq 0, \sum_{i=1}^{n} p_i = 1\}$に対して我々が知りたいのは以下のような σに関する多項分布の超パラメータ $p = (p_1, p_2, \ldots, p_n) \in P_n$ であってディリクレ分布とも呼ばれる多次元分布の推定の方法である.

$$f(\sigma|\hat{\sigma}, p) = \frac{(\sigma_1 + \cdots + \sigma_n)!}{\sigma_1! \cdots \sigma_n!} p_1^{\sigma_1} p_2^{\sigma_2} \cdots p_n^{\sigma_n}. \tag{1}$$

P_n上のルベーグ測度を$L(\cdot)$とし, 定数$k \geqq 1$に対して $[L, kL]$ によって事前区間測度を表す. データセットσから, $[L_\sigma, kL_\sigma]$による事後区間測度は以下のようにして得られる(cf. [9] DeRobertis and Hartigan(1981)):

$$L_\sigma(A) = \int_A f(\sigma|\hat{\sigma}, p) L(dp) \quad \text{for } A \in \mathcal{B}, \tag{2}$$

ただし $\mathcal{B}$ は P_n の部分集合による σ-集合体である.

DeRobertis and Hartigan ([9]) の結果により, 推定する推移確率の第i成分 p_i の区間表現は以下のような積分比の範囲として得られる:

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

このp_iの事後区間を$[\underline{\lambda}_i, \overline{\lambda}_i]$と表すことにする.

次が成り立つ.

Theorem 1 ([9] DeRobertis and Hartigan (1981)). 事後区間測度 $[L_\sigma, kL_\sigma]$ によって, 推移確率成分 p_i の下限値 $\underline{\lambda}_i$ と上限値 $\overline{\lambda}_i$ は次のそれぞれの方程式のただ一つの解である:

$$kL_\sigma(p_i - \underline{\lambda}_i)^- + L_\sigma(p_i - \underline{\lambda}_i)^+ = 0, \tag{3}$$

$$kL_\sigma(p_i - \overline{\lambda}_i)^+ + L_\sigma(p_i - \overline{\lambda}_i)^- = 0, \tag{4}$$

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

下限値 $\underline{\lambda}_i$ と上限値 $\overline{\lambda}_i$ はベータ関数および不完全ベータ関数を用いて, 以下のように具体的に表すことができる.

$$\underline{\lambda}_i = \frac{B(s+1,t)+(k-1)B(s+1,t,\underline{\lambda}_i)}{B(s,t)+(k-1)B(s,t,\underline{\lambda}_i)}, \quad \overline{\lambda}_i = \frac{kB(s+1,t)-(k-1)B(s+1,t,\overline{\lambda}_i)}{kB(s,t)-(k-1)B(s,t,\overline{\lambda}_i)}, \tag{5}$$

where $s=\sigma_i+1, t=\sum_{k=1}^{n}\sigma_k-\sigma_i+(n-1)$, $B(s,t)=\int_0^1 x^{s-1}(1-x)^{t-1}dx$ and $B(s,t,\lambda)=\int_0^\lambda x^{s-1}(1-x)^{t-1}dx$.

2 α-percentile による事後区間確率行列

事後区間$[\underline{\lambda}_i,\overline{\lambda}_i]$を,次のような α-パーセンタイルによって求める.

$\underline{g}_{i,a},\overline{g}_{i,a}$ を次のような P_n上の可測関数とする.

$$\underline{g}_{i,a}(p)=I_{\{p_i\leqq a\}}(p), \quad \overline{g}_{i,a}(p)=I_{\{p_i\geqq a\}}(p), \tag{6}$$

where $I_A(x)=1$ if $x\in A$, $=0$ if $x\notin A$.

さらに, $\underline{\lambda}(a|\sigma),\overline{\lambda}(a|\sigma)$を次のように定義する.

$$\underline{\lambda}(a|\sigma)=\sup\left\{\frac{Q_\sigma(\underline{g}_{i,a})}{Q_\sigma(I_{P_n})}\Big|Q_\sigma\in[L_\sigma,kL_\sigma]\right\}, \tag{7}$$

$$\overline{\lambda}(a|\sigma)=\sup\left\{\frac{Q_\sigma(\overline{g}_{i,a})}{Q_\sigma(I_{P_n})}\Big|Q_\sigma\in[L_\sigma,kL_\sigma]\right\}, \tag{8}$$

where $Q(g)=\int g(p)Q(dp)$ for $Q\in[L_\sigma,kL_\sigma]$ and mesurable function g on P_n.

ここで, 下側α-パーセンタイル $\underline{p}_i(\alpha)$ と上側α-パーセンタイル $\overline{p}_i(\alpha)$ を次のように定義する:

$$\underline{\lambda}(\underline{p}_i(\alpha)|\sigma)=\alpha, \quad \overline{\lambda}(\overline{p}_i(\alpha)|\sigma)=\alpha. \tag{9}$$

Proposition 1. 任意の$Q\in I(L_\sigma,kL_\sigma)$に対して

$$\frac{Q(I_{\{p_i\leqq\underline{p}_i(\alpha)\}})}{Q(1)}\leqq\alpha, \frac{Q(I_{\{p_i\geqq\overline{p}_i(\alpha)\}})}{Q(1)}\leqq\alpha,$$

が成り立つ.

このとき, 両側のα-パーセンタイルは,それぞれ次のようなベータ分布のパーセンタイルから求めることが出来る.

Theorem 2. 下側α-パーセンタイル $\underline{p}_i(\alpha)$ と上側α-パーセンタイル $\overline{p}_i(\alpha)$ は, それぞれ以下の方程式を満たす.

$$\frac{B(s,t|\underline{p}_i(\alpha))}{B(s,t)}=\frac{\alpha}{\alpha+(1-\alpha)k}, \quad \frac{B(s,t|\overline{p}_i(\alpha))}{B(s,t)}=\frac{(1-\alpha)k}{\alpha+(1-\alpha)k} \tag{10}$$

例えば, 両側の$\frac{\alpha}{2}$を端点にもつ区間$[\underline{p}_i(\alpha/2), \overline{p}_i(\alpha/2)]$ を 100(1-α)% 中央確信区間 (central credible interval)と呼ぶ.

小標本のデータセット$\hat{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$に対して, 母数パラメータ$p = (p_1, p_2, \ldots, p_n)$の区間推定と$p$に関する事後期待値に加えて, 確信区間の構成によって得られる区間推移確率行列の下限確率行列と上限確率行列によって, データセットの標本分布がどのように振る舞うのかを次の節でみていく.

3 数値例

ここで,具体的に$L(\cdot)$:ルベーグ測度, 事前測度区間$[L, kL]$(kは定数)について, $k = 2$と考えて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.([6]の数値例より引用)

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期目の状態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

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

であった.例えば,状態2からの推移では, 上の行列の第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, Theorem 1 から以下のように得られる(Table 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 1: Intervals of posterior measures(mean value)

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

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

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

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

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

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

このとき, 確率行列の区間集合$\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle = \{Q \in P(S|S \times A) | \underline{Q} \leqq Q \leqq \overline{Q}\}$ は次のように得られる. $\underline{Q} = \begin{pmatrix} .400 & .187 & .292 \\ .187 & .400 & .292 \\ .168 & .262 & .458 \end{pmatrix}$, $\overline{Q} = \begin{pmatrix} .489 & .260 & .376 \\ .260 & .489 & .376 \\ .235 & .334 & .542 \end{pmatrix}$ をそれぞれ下

限行列, 上限行列とする行列で確率行列を構成するものは, 以下の図のような凸集合として得られる.

例えば, 1行目の$q_1 = (q_{11}, q_{12}, q_{13})$の各成分は, $0.400 \leqq q_{11} \leqq 0.489$, $0.187 \leqq q_{12} \leqq 0.260$, $0.292 \leqq q_{13} \leqq 0.376$, $\sum_{j=1}^{3} q_{1j} = 1$を満たし, 具体的には, (q_{11}, p_{12}, p_{13})の凸集合の端点集合として, $(0.437, 0.187, 0.376)$, $(0.4, 0.224, 0.376)$, $(0.448, 0.26, 0.292)$, $(0.489, 0.219, 0.292)$, $(0.4, 0.26, 0.34)$, $(0.489, 0.187, 0.324)$ が得られる(1).

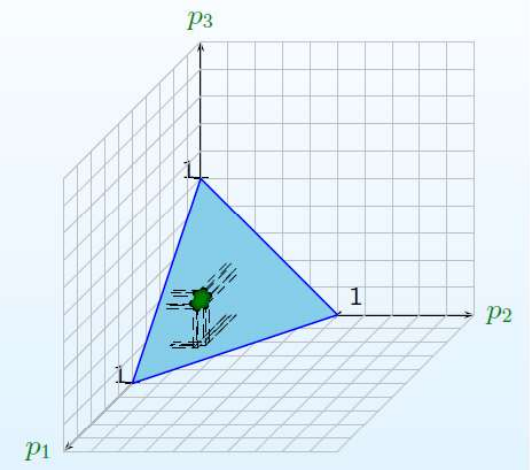

Figure 1: 確率行列の例

一般には, $\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle = \{Q \in P(S|S \times A) | \underline{Q} \leqq Q \leqq \overline{Q}\}$とするとき, $\mathcal{Q}$は凸多面体で表される(cf. [6])から,ある端点の集合 $\{Q^{(1)}, Q^{(2)}, \ldots, Q^{(l)}\}$によって$\mathcal{Q} = \mathrm{conv}\{Q^{(1)}, Q^{(2)}, \ldots, Q^{(l)}\}$ と表すことができる. $\mathcal{Q} \ni Q = (q_{ij})$について,各$i$行目ごとに推移確率行列の条件 $\sum_{j=1}^{3} q_{ij} = 1$ $(i = 1, 2, 3)$をみたす端点調べる. $\mathcal{Q}$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.

$$\begin{aligned}
\mathrm{ext}(\hat{q}_1) = \{&(0.437, 0.187, 0.376), (0.4, 0.224, 0.376), (0.448, 0.26, 0.292),\\
&(0.489, 0.219, 0.292), (0.4, 0.26, 0.34), (0.489, 0.187, 0.324)\},\\
\mathrm{ext}(\hat{q}_2) = \{&(0.187, 0.437, 0.376), (0.224, 0.4, 0.376), (0.26, 0.448, 0.292),\\
&(0.219, 0.489, 0.292), (0.26, 0.4, 0.34), (0.187, 0.489, 0.324)\},\\
\mathrm{ext}(\hat{q}_3) = \{&(0.196, 0.262, 0.542), (0.168, 0.29, 0.542), (0.208, 0.334, 0.458),\\
&(0.235, 0.307, 0.458), (0.168, 0.334, 0.498), (0.235, 0.262, 0.503)\}
\end{aligned}$$

を得る.

本報告では, 次の各αparcentile による区間表現に対して, $\alpha = 0.05$と$\alpha = 0.5$のときの, 各第一行目$q_1 = (q_{11}, q_{12}, q_{13})$の危険率$\alpha$と棄却域, 検出力について考察していく.

$\alpha = 0.05$のときの第一行目の端点集合は,

$\mathrm{ext}(\hat{q}_1) = \{(0.095, 0.641, 0.264)$,$(0.095, 0.155, 0.750)$, $(0.837, 0.013, 0.750)$,
$(0.237, 0.013, 0.750)$, $(0.837, 0.118, 0.045)$, $(0.314, 0.641, 0.045)\}$,

また, $\alpha = 0.5$のときの第一行目の端点集合は,
$\mathrm{ext}(\hat{q}_1) = \{(0.185, 0.483, 0.332)$,$(0.185, 0.204, 0.611), (0.720, 0.043, 0.237)$,
$(0.346, 0.043, 0.611), (0.720, 0.175, 0.105), (0.412, 0.483, 0.105)\}$ を得る. ここで, それぞれの端点$p = (p_1, p_2, p_3)$を母数パラメータとする多項分布を考えると, データセット$\sigma = (\sigma_1, \sigma_2, \sigma_3)$の各成分$\sigma_i$は, 成功確率$p_i$の2項分布に従う. そこで, σ_iの周辺分布をもとに, 生起確率を小さい順にして累積和をとることで下位5%未満を除いた観測予測集合を下記の表にまとめる(Table 2,3).

Table 2: $\alpha = 0.05$, predictable numbers of observation of σ_i

母数p_i	値	観測予測集合
p_1	0.095	$\{0, 1, 2\}$
p_1	0.237	$\{0, 1, 2, 3\}$
p_1	0.314	$\{0, 1, 2, 3, 4\}$
p_1	0.837	$\{3, 4, 5, 6\}$
p_2	0.013	$\{0, 1, 2\}$
p_2	0.118	$\{0, 1, 2, 3\}$
p_2	0.155	$\{0, 1, 2, 3, 4\}$
p_2	0.641	$\{3, 4, 5, 6\}$
p_3	0.045	$\{0, 1, 2\}$
p_3	0.150	$\{0, 1, 2, 3\}$
p_3	0.264	$\{0, 1, 2, 3, 4\}$
p_3	0.750	$\{3, 4, 5, 6\}$

Table 3: $\alpha = 0.5$, predictable numbers of observation of σ_i

母数p_i	値	観測予測集合
p_1	0.185	$\{0, 1, 2, 3\}$
p_1	0.346	$\{0, 1, 2, 3, 4\}$
p_1	0.412	$\{1, 2, 3, 4, 5\}$
p_1	0.720	$\{2, 3, 4, 5, 6\}$
p_2	0.043	$\{0, 1\}$
p_2	0.175	$\{0, 1, 2, 3\}$
p_2	0.204	$\{0, 1, 2, 3\}$
p_2	0.483	$\{1, 2, 3, 4, 5\}$
p_3	0.105	$\{0, 1, 2\}$
p_3	0.237	$\{0, 1, 2, 3\}$
p_3	0.332	$\{0, 1, 2, 3, 4\}$
p_3	0.611	$\{2, 3, 4, 5, 6\}$

また, 真値$p = (1/2, 1/6, 1/3)$に対する観測予測集合を次に示す(Table 4).

Table 4: $p = (1/2, 1/6, 1/3)$, predictable numbers of observation of σ_i

母数p_i	値	観測予測集合
p_1	1/2	$\{1, 2, 3, 4, 5\}$
p_2	1/6	$\{0, 1, 2, 3\}$
p_3	1/3	$\{0, 1, 2, 3, 4\}$
p_3	0.237	$\{0, 1, 2, 3\}$

多項分布に基づく観測では, 標本総数が少なくてもそれぞれのデータセット$\sigma = (\sigma_1, \sigma_2, \sigma_3)$の観測値の生起確率はそれほど大きくない. 例えば, $\sum_{i=1}^{3} \sigma_i = 6$の場合に, $(\sigma_1, \sigma_2, \sigma 3)$の観測パターンは28通り, $(\sigma_1, \sigma_2, \sigma_3) = (3, 1, 2)$となる事象が最大確率0.139である. そこで, 本報告で用いている区間型推定手法がどのくらい真値を含む可能性をもっているか, 真値に基づく観測予測集合を, 区間型確率行列を構成している端点の真値から離れているいくつかのパラメータ値に当てはめて検出力に相当するものをディリクレ分布をもとに求め, 以下の表（Table 5）のようにまとめておく.

Table 5: power of tests for some extream points

母数p_i	値	検出力
p_1	0.095	0.5494
p_1	0.837	0.3439
p_2	0.641	0.6289
p_3	0.750	0.5339

また, $\alpha = 0.5$percentileによる区間確率行列であっても, 真値に対する5%未満の棄却域を設定した時の検出力は, $p_1 = 0.185$に対して0.2931, $p_1 = 0.720$に対して0.1398, $p_2 = 0.483$に対して0.3124, $p_3 = 0.611$に対して0.2508 であって, 第一種の過誤≦第二種の過誤に類似の状況にあると考えられる.

References

[1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis, 2nd ed.*. Springer-Verlag, New York, 1988.

[2] M. H. DeGroot. *Optimal statistical decisions.* McGraw-Hill Book Co., New York, 1970.

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

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

[5] M. Horiguchi. Newton-Raphson Iteration for Uncertain Markov Decision Processes. In *Proceedings of the 2018 International Conference on Management and Operations Research, Yan Xianbin et al. Ed. ARPUB(2018)*, pages 42–52.

[6] 伊喜哲一郎, 堀口正之, 安田正實, 蔵野正美. 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. In 数理解析研究所講究録*1636*「不確実性と意思決定の数理」, pages 1–8.

[7] 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.

[8] Masami Kurano, Masayuki Horiguchi, and Minoru Sasaki. Flexibly structured Bayesian methods and their applications to quality control. (in Japanese) In *Shogaku Ronkyu*, Vol. 61(3), pages 181–192. Kwansei University, 2014.

[9] L. De Robertisand J. A. Hartigan, Bayesian inference using intervals of measures. *Ann. Statist.*, 9:235–244, 1981.

[10] M. Sasaki, M. Horiguchi and M. Kurano. Adaptive methods for multivariate Bayesian control chart. *RIMS kokyuroku No. 1912 (In Japanese)*, pages 181–192, 2014.

[11] Samuel S. Wilks. *Mathematical statistics*. A Wiley Publication in Mathematical Statistics. John Wiley & Sons Inc., New York, 1962.

Masayuki Horiguchi
Department of Mathematics,
Faculty of Science, Kanagawa University
Address: Tsuchiya 2946, Hiratsuka City,
Kanagawa Prefecture, 259-1293, Japan
E-mail address: horiguchi@kanagawa-u.ac.jp

神奈川大学・理学部　堀口正之