推移法則未知の区間型マルコフ決定過程におけるパーセンタイル型リスク評価について
(On a risk measurement by percentiles under controlled Markov set-chains with unknown transition probability)

神奈川大学・理学部・堀口正之
Masayuki HORIGUCHI
高知大学・理工学部・阪口昌彦
Masahiko SAKAGUCHI
Professor of Department of Mathematics,
Faculty of Science,
Kanagawa University
Associate Professor of Department of Information Science,
Faculty of Science and Technology,
Kochi University

Abstract

We are concerned with Markov decision processes with unknown transition matrices. In this talk, we derive the estimated intervals of transition matrices which has the specified posterior probability of parameters as the true transition matrices. From the data set of state observations, transition matrices are estimated as closed convex set in each component of matrices. By solving the integral equations for the set of measures of priors, we have lower and upper values of intervals as α-percentile credible interval. Posterior intervals are obtained by the values of inverse beta distribution functions. Then, we can formulate interval estimated Markov decision processes with α-percentile credibility and evaluate the information from the obtained dataset. Through the numerical examples we show the complete inferences of transition law based on the dataset and consider the risk measurement for Markov decision processes with uncertainty.

1 Preliminaries

Finite Markov decision processes(MDPs) consists of four objects:

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

where $S = \{1, 2, \ldots, n\}$ and $A = \{a_1, a_2, \ldots, a_k\}$ are finite state and action spaces and $Q = (q_{ij}(a))$ is transition probability matrices such that $q_{ij}(a) \in P(S|S \times A)$ and $r = r(i, a)$ is reward function on $S \times A$. When the system is in state $i \in S$ and we take an action $a \in A$, we move to a new state $j \in S$ selected according to probability $q(\cdot|(i, a))$ and receive an immediate reward $r(i, a)$. In uncertain MDPs transition probability matrices $Q = (q_{ij}(a))$ is unknown, so that we estimate $q_{i\cdot}(a)$ to be chosen from interval $[\underline{q}_{i\cdot}(a), \overline{q}_{i\cdot}(a)]$. The decision model with intervals of stochastic transition matrices, called controlled Markov set-chain, has developed by Kurano et al(e.g. [8]). For simplicity, we consider MDPs with stationary transition law.

Let $Q = (q_{ij})$ be unknown transition matrix. We derive interval estimations for each row of Q by prior intervals of measures and Bayesian inference. We estimate some fixed ith row of Q since transition occurs according to probability $p_{i\cdot}$ at the present state i and other rows can be estimated as Bayesian intervals similarly.

Let $P_n = P(S) = \{p = (p_1, p_2, \ldots, p_n) | p_i \geqq 0, \sum_{i=1}^{n} p_i = 1\}$. We denote an observed data set by $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ whose jth element σ_j is the number of outcomes of the next transition to state

j from the same fixed state. Then, for parameter $p = (p_1, p_2, \ldots, p_n) \in P_n$ a data σ has p.d.f. of multinomial distribution as follows:

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

where $N = \sum_{i=1}^{n} \sigma_i$.

Let $L(\cdot)$ be Lebesgue measure on P_n and $[L, kL]$ $(k \geqq 1)$ the interval of prior measure. Upper bound measure kL is a proportional measure of lower bound L. For σ posterior interval of measures $[L_\sigma, kL_\sigma]$ is constructed from the following measure([11]):

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

where $\mathcal{B}$ denotes σ-field of subsets of P_n.

2 Controlled Markov set-chain

Let $Q = (q_{ij})$ be unknown transition matrix. We derive a method of Bayesian interval estimation for each row of Q by prior intervals of measures. We estimate some fixed ith row of Q since transition occurs according to probability distribution $q_{i\cdot}$ at the present state i and other rows can be estimated as Bayesian intervals similarly.

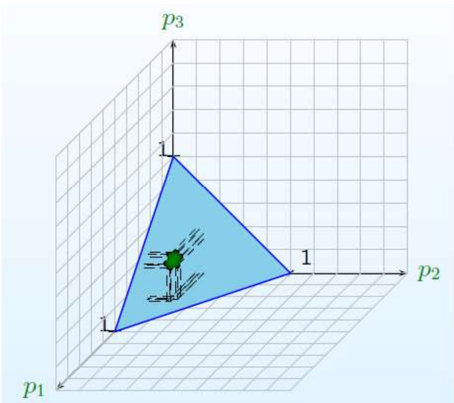
$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} & \cdots\cdots & q_{1n} \\ q_{21} & q_{22} & q_{23} & \cdots\cdots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ q_{i1} & q_{i2} & q_{i3} & \cdots\cdots & q_{in} \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots \\ q_{n1} & q_{n2}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

We define partial order $\preceq, \prec$ on $\mathbb{R}^{m \times n}$:
For $\mathbb{R}^{m \times n} \ni A = (a_{ij}), B = (b_{ij})$

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

For any $\underline{A} \preceq \overline{A}$, define $\langle \underline{A}, \overline{A} \rangle$ as follows:

$$\langle \underline{A}, \overline{A} \rangle = \left\{ Q = (q_{ij}) \in \mathbb{R}_+^{m \times n} \middle| \underline{a}_{ij} \leqq q_{ij} \leqq \overline{a}_{ij}, q_{ij} \geqq 0, \sum_{j=1}^{n} q_{ij} = 1 \ (1 \leqq i \leqq m, 1 \leqq j \leqq n) \right\}. \tag{4}$$



Let $\mathcal{M}_n$ be the set of all interval matrices with $n \times n$ elements.

$$\mathcal{M}_n = \left\{ \langle \underline{Q}, \overline{Q} \rangle \middle| \langle \underline{Q}, \overline{Q} \rangle \neq \emptyset, \underline{Q} \preceq \overline{Q}, \underline{Q}, \overline{Q} \in \mathbb{R}_+^{n \times n} \right\} \tag{5}$$

For $\mathcal{Q}_1, \mathcal{Q}_2 \in \mathcal{M}_n$, define the product of $\mathcal{Q}_1, \mathcal{Q}_2$ by

$$\mathcal{Q}_1 \mathcal{Q}_2 = \left\{ Q_1 Q_2 \middle| Q_1 \in \mathcal{Q}_1, Q_2 \in \mathcal{Q}_2 \right\}. \tag{6}$$

Also, the multiproduct of $\mathcal{Q} \in \mathcal{M}_n$ are defined inductively by

$$\mathcal{Q}^k = \mathcal{Q}^{k-1} \mathcal{Q} \ (k \geqq 2). \tag{7}$$

We denote by $C(\mathbb{R}_+)$ the set of all bounded and closed intervals in $\mathbb{R}_+$ and $C(\mathbb{R}_+)^n$ the set of all n-dimensional column vectors whose elements are in $C(\mathbb{R}_+)$:

$$C(\mathbb{R}_+)^n = \left\{ D = (D_1, D_2, \ldots, D_n)' \middle| D_i \in C(\mathbb{R}_+) \ (1 \leqq i \leqq n) \right\} \tag{8}$$

where $\boldsymbol{d}'$ denotes the transpose of a vector $\boldsymbol{d}$.

For $D = (D_1, D_2 \ldots, D_n)', E = (E_1, E_2, \ldots, E_n)' \in C(\mathbb{R}_+)^n$, $h \in \mathbb{R}_+^n, \lambda \in \mathbb{R}_+$,

$$\begin{aligned} D + E &= \{d + e | d \in D, e \in E\}, \\ h + D &= \{h + d | d \in D\}, \\ \lambda D &= \{\lambda d | d \in D\}. \end{aligned} \tag{9}$$

We set $D = [\underline{d}, \overline{d}] = ([\underline{d}_1, \overline{d}_1], [\underline{d}_2, \overline{d}_2], \ldots, [\underline{d}_n, \overline{d}_n])' \in C(\mathbb{R}_+)^n$, where $\underline{d} = (\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_n) \in \mathbb{R}_+^n, \overline{d} = (\overline{d}_1, \overline{d}_2, \ldots, \overline{d}_n) \in \mathbb{R}_+^n$. For any $D = (D_1, D_2, \ldots, D_n)' \in C(\mathbb{R}_+)^n$ and $G \subset \mathbb{R}_+^{1 \times n}$, the product GD is defined by

$$GD = \{gd | g = (g_1, g_2, \ldots, g_n) \in G, d = (d_1, d_2, \ldots, d_n)' \in D, d_i \in D_i \ (1 \leqq i \leqq n)\} \tag{10}$$

Then, we have the following:

Lemma 2.1. *([4, 7])*

(i) Any $\mathcal{Q} \in \mathcal{M}_n$ is a convex polytope in $\mathbb{R}^{n \times n}$.

(ii) For any compact subset $G \subset \mathbb{R}_+^{1 \times n}$ and $D \in C(\mathbb{R}_+)^n$, it holds $GD \in C(\mathbb{R}_+)$.

Partial order $\preceq, \prec$ on $C(\mathbb{R}_+)$: For $[c_1, c_2], [d_1, d_2] \in C(\mathbb{R}_+)$,

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

Partial order $\preceq, \prec$ on $C(\mathbb{R}_+)^n$ are defined by using the partial order on $C(\mathbb{R}_+)$ as follows: For $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)', \boldsymbol{w} = (w_1, w_2, \ldots, w_n)' \in C(\mathbb{R}_+)^n$,

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

Let $\mathbb{R}_+^n \supset D_1, D_2$ be bounded and closed set. We denote by ρ Hausdorff metric, i.e.,

$$\rho(D_1, D_2) = \max\{ \sup_{x \in D_1} \inf_{y \in D_2} \|x - y\|, \sup_{y \in D_2} \inf_{x \in D_1} \|x - y\| \}, \tag{11}$$

where $\|\cdot\|$ is Euclid metric in $\mathbb{R}^n$.

We set $S = \{1, 2, \ldots, n\}$ state space and $A = \{1, 2, \ldots, k\}$ action space. We set

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

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

Let $B_+(D)$ be the set of all non-negative real functions on finite set D. For a finite D ($n = \#D$), $B_+(D)$ is identified with $\mathbb{R}_+^n$.

We consider standard MDPs $\{S, A, Q, \boldsymbol{r}\}$(cf. [10]). For the simplicity of the problem, we treat the case of deterministic and stationary policy. We set F the set of all map $f : S \to A$. For any $f \in F$, discounted total expected reward $\phi(f|Q) \in \mathbb{R}_+^n$ with discount factor β ($0 < \beta < 1$) is defined as a function of stochastic matrix $Q \in P(S|S \times A)$ as:

$$\phi(f|Q) = \sum_{t=0}^{\infty} (\beta Q(f))^t \boldsymbol{r}(f), \tag{12}$$

where, $\boldsymbol{r}(f) = (r(1, f(1)), r(2, f(2)), \ldots, r(n, f(n)))' \in \mathbb{R}_+^n, Q(f) = (q_{ij}(f(i))) \in P(S|S)$. or $f \in F$, the map $L(f) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is defined as:

$$L(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta Q(f)\boldsymbol{x}, \ \boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}_+^n. \tag{13}$$

Then, the following fundamental lemma is known.

Lemma 2.2. *(cf. [10])*

(i) $L(f)$ *is monotone increasing and contractive mapping, i.e.,*

$$\boldsymbol{x} \leqq \boldsymbol{x}' \text{ implies } L(f)\boldsymbol{x} \leqq L(f)\boldsymbol{x}' \text{(componentwise)},$$
$$\|L(f)\boldsymbol{x} - L(f)\boldsymbol{x}'\| \leqq \beta \|\boldsymbol{x} - \boldsymbol{x}'\| \ (\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}_+^n),$$

where, $\|\cdot\|$ *means* sup-*norm.*

(ii) $\phi(f|Q)$ *is the unique fixed point of* $L(f)$, *i.e., for any* $\boldsymbol{x} \in \mathbb{R}_+^n$, *we have*

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

True transition matrix Q is estimated by $\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle$, where

$$\begin{aligned} &\underline{Q} = (\underline{q}_{ij}(a) : i, j \in S, a \in A) \in \mathbb{R}_+^{kn \times n}, \\ &\overline{Q} = (\overline{q}_{ij}(a) : i, j \in S, a \in A) \in \mathbb{R}_+^{kn \times n}, \\ &\mathcal{Q} = \langle \underline{Q}, \overline{Q} \rangle = \left\{ Q \in P(S|S \times A) \middle| \underline{Q} \leqq Q \leqq \overline{Q} \right\}. \end{aligned} \tag{14}$$

We define interval estimated MDPs $\{S, A, \mathcal{Q}, \boldsymbol{r}\}$ as follows. For $f \in F$, we define discounted total expected-set valued value function $\phi(f|\mathcal{Q})$ as follows:

$$\phi(f|\mathcal{Q}) = \left\{ \phi(f|Q) \middle| Q \in \mathcal{Q} \right\} \subset \mathbb{R}_+^n \tag{15}$$

where, the value $\phi(f|Q)$ of standard MDPs is defined in (12).

It can be shown that $\phi(f|\mathcal{Q}) \in C(\mathbb{R}_+)^n$:
map $\mathcal{L} : C(\mathbb{R}_+)^n \to C(\mathbb{R}_+)^n$:

$$\mathcal{L}(f)\boldsymbol{v} = \boldsymbol{r}(f) + \beta \mathcal{Q}(f)\boldsymbol{v}, \ \boldsymbol{v} \in C(\mathbb{R}_+)^n, \tag{16}$$

where, $\mathcal{Q}(f) = \langle \underline{Q}(f), \overline{Q}(f) \rangle, \underline{Q}(f) = (\underline{q}_{ij}(f(i))) \in \mathbb{R}_+^{n\times n}, \overline{Q}(f) = (\overline{q}_{ij}(f(i))) \in \mathbb{R}_+^{n\times n}$.

From Lemma 2.1, we have $\mathcal{L}(f)\boldsymbol{v} \in C(\mathbb{R}_+)^n$ ($\boldsymbol{v} \in C(\mathbb{R}_+)^n$). Moreover, we define $\underline{L}(f) : \mathbb{R}_+^n \to \mathbb{R}_+^n, \overline{L}(f) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ as follows: For $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}_+^n$,

$$\underline{L}(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta \min_{Q\in\mathcal{Q}(f)} Q\boldsymbol{x}, \tag{17}$$

$$\overline{L}(f)\boldsymbol{x} = \boldsymbol{r}(f) + \beta \max_{Q\in\mathcal{Q}(f)} Q\boldsymbol{x}. \tag{18}$$

Then, we have the followings:

Lemma 2.3. *For any $f \in F$,*

(i) $\mathcal{L}(f)$ is monotone increasing and contractive mapping.

(ii) $\underline{L}(f)$ and $\overline{L}(f)$ are both monotone increasing and contractive mapping with respect to sup*-norm.*

Applying Lemma 2.2 and Lemma 2.3, we have the following.

Theorem 2.1. *For any $f \in F$, it holds that*

(i) $\phi(f|\mathcal{Q}) \in C(\mathbb{R}_+)^n$ and $\phi(f|\mathcal{Q})$ is the unique fixed point of $\mathcal{L}(f)$. Moreover, for any $\boldsymbol{v} \in C(\mathbb{R}_+)^n$, we have

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

(ii) Let $\phi(f|\mathcal{Q}) = [\underline{\phi}(f), \overline{\phi}(f)]$. Then, $\underline{\phi}(f)$ and $\overline{\phi}(f)$ are the unique fixed point of $\underline{L}(f)$ and $\overline{L}(f)$, respectively.

We call $f^* \in F$ Pareto-optimal if there does not exists $f \in F$ such that $\phi(f^*|\mathcal{Q}) \prec \phi(f|\mathcal{Q})$.

For $D \subset C(\mathbb{R}_+)^n$, a point $\boldsymbol{v} \in D$ is called an *efficient point* of D iff it holds that there does not exist $\boldsymbol{u} \in D$ such that $\boldsymbol{v} \prec \boldsymbol{u}$. We denote by eff$(D)$ the set of all efficient points in D. For each element vector

$$\underline{Q}_{i,a} = (\underline{q}_{i1}(a), \underline{q}_{i2}(a), \ldots, \underline{q}_{in}(a)), \text{ and } \overline{Q}_{i,a} = (\overline{q}_{i1}(a), \overline{q}_{i2}(a), \ldots, \overline{q}_{in}(a))$$

of $\underline{Q}$ and $\overline{Q}$ respectively in (14), define $\mathcal{Q}_{i,a} = \langle \underline{Q}_{i,a}, \overline{Q}_{i,a} \rangle$ ($i \in S, a \in A$).
For $\boldsymbol{u} \in C(\mathbb{R}_+)^n$, let

$$\mathcal{L}(\boldsymbol{u}) := (\mathcal{L}(\boldsymbol{u})_1, \mathcal{L}(\boldsymbol{u})_2, \ldots, \mathcal{L}(\boldsymbol{u})_n)', \tag{19}$$

where, $\mathcal{L}(\boldsymbol{u})_i := \text{eff}(\{r(i,a) + \beta\mathcal{Q}_{i,a}\boldsymbol{u} | a \in A\})$ ($i \in S$).

Lemma 2.4. *For $f, g \in F$, if $\phi(f|\mathcal{Q}) \prec \mathcal{L}(g)\phi(f|\mathcal{Q})$, then $\phi(f|\mathcal{Q}) \prec \phi(g|\mathcal{Q})$.*

Then, we have the following.

Theorem 2.2. *$f^* \in F$ is Pareto-optimal if and only if $\phi(f^*|\mathcal{Q})$ is a maximal solution to the optimality inclusion*

$$\boldsymbol{u} \in \mathcal{L}(\boldsymbol{u}), \boldsymbol{u} \in C(\mathbb{R}_+)^n. \tag{20}$$

3 Interval estimation

We estimate each ith element p_i of parameter $p = (p_1, p_2, \ldots, p_n) \in P_n$ to be $[\underline{\lambda}_i, \overline{\lambda}_i]$ by applying the method of Bayesian inference([11]) with posteriar measures $Q_\sigma \in [L_\sigma, kL_\sigma]$.

For posterior measures $Q_\sigma \in [L_\sigma, kL_\sigma]$, first we shall consider intervals of mean value for Q_σ. Posterior interval for each p_i is given as the range of integral ratios:

$$\left\{ \int_{P_n} p_i \, Q_\sigma(dp) / \int_{P_n} Q_\sigma(dp) \middle| L_\sigma \leqq Q_\sigma \leqq U_\sigma \right\}. \tag{21}$$

Also, it follows that posterior interval $[\underline{\lambda}_i, \overline{\lambda}_i]$ is given by unique solutions of following equations:

$$\underline{\lambda}_i = \frac{B(s+1,t) + (k-1)B(s+1,t,\underline{\lambda}_i)}{B(s,t) + (k-1)B(s,t,\underline{\lambda}_i)}, \quad \overline{\lambda}_i = \frac{kB(s+1,t) - (k-1)B(s+1,t,\overline{\lambda}_i)}{kB(s,t) - (k-1)B(s,t,\overline{\lambda}_i)}, \tag{22}$$

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 $B(s,t,\lambda) = \int_0^\lambda x^{s-1}(1-x)^{t-1}dx$.

Theorem 3.1. *Lower bound $\underline{\lambda}_i$ and upper bound $\overline{\lambda}_i$ of posterior intervals $[\underline{\lambda}_i, \overline{\lambda}_i]$ $(i \in S)$ are unique solutions as following equations:*

$$U_\sigma(p_i - \underline{\lambda}_i)^- + L_\sigma(p_i - \underline{\lambda}_i)^+ = 0, \tag{23}$$

$$U_\sigma(p_i - \overline{\lambda}_i)^+ + L_\sigma(p_i - \overline{\lambda}_i)^- = 0, \tag{24}$$

where $Q(f)$ denotes the integral of function f w.r.t. measure Q, $x^+ = \max\{0,x\}, x^- = x - x^+ = \min\{0,x\}$.

For another posterior intervals $[\underline{\lambda}_i, \overline{\lambda}_i]$, we consider lower and upper α-percentile of p_i from posterior measures $Q_\sigma \in [L_\sigma, kL_\sigma]$. Let $\underline{g}_{i,a}, \overline{g}_{i,a}$ be measurable functions on P_n as follows:

$$\underline{g}_{i,a}(p) = I_{\{p_i \leqq a\}}(p), \quad \overline{g}_{i,a}(p) = I_{\{p_i \geqq a\}}(p), \tag{25}$$

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

$$\underline{\lambda}(a|\sigma) = \sup\left\{\frac{Q_\sigma(\underline{g}_{i,a})}{Q_\sigma(I_{P_n})}|Q_\sigma \in [L_\sigma, kL_\sigma]\right\}, \overline{\lambda}(a|\sigma) = \sup\left\{\frac{Q_\sigma(\overline{g}_{i,a})}{Q_\sigma(I_{P_n})}|Q_\sigma \in [L_\sigma, kL_\sigma]\right\}, \tag{26}$$

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

Now, lower α-percentile $\underline{p}_i(\alpha)$ and upper α-percentile $\overline{p}_i(\alpha)$ are defined as follows:

$$\underline{\lambda}(\underline{p}_i(\alpha)|\sigma) = \alpha, \quad \overline{\lambda}(\overline{p}_i(\alpha)|\sigma) = \alpha. \tag{27}$$

Then,

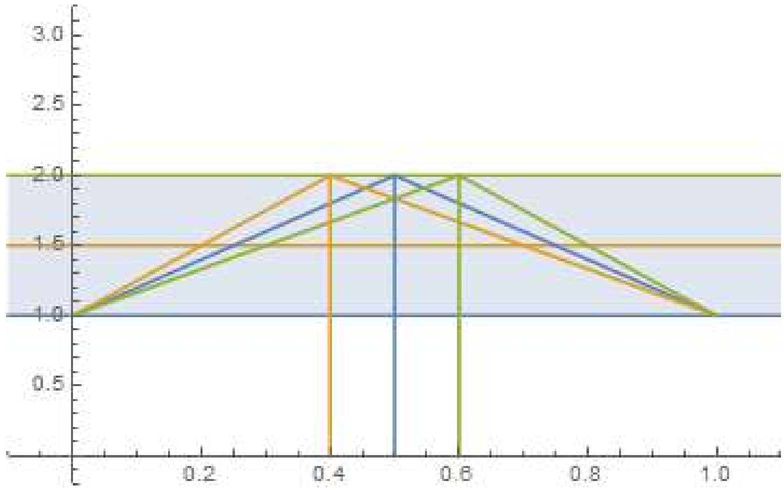
Theorem 3.2. *The values of lower α-percentile $\underline{p}_i(\alpha)$ and upper α-percentile $\overline{p}_i(\alpha)$ satisfies following equations respectively.*

$$\frac{B(s,t|\underline{p}_i(\alpha))}{B(s,t)} = \frac{\alpha}{\alpha + (1-\alpha)k}, \quad \frac{B(s,t|\overline{p}_i(\alpha))}{B(s,t)} = \frac{(1-\alpha)k}{\alpha + (1-\alpha)k} \tag{28}$$

By applying this theorem we have several types of $100(1-\alpha)\%$ credible interval for p_i, such as lower $[0, \overline{p}_i(\alpha)]$, central $[\underline{p}_i(\alpha/2), \overline{p}_i(\alpha/2)]$ and upper $[\underline{p}_i(\alpha), 1]$, and moreover we can consider credible decision model with stochastic transition matrices in their intervals.

4 Examples of prior interval measures by proportion

We show some set of prior distributions. Let $Q \in [L, 2L]$, where L is Lebesgue measure on $[0,1]$. (Uniformly distribution $U(0,1)$).



$$f(x) = \begin{cases} 2(x-\frac{1}{2})+2 & (0 \leqq x \leqq \frac{1}{2}) \\ -2(x-\frac{1}{2})+2 & (\frac{1}{2} < x \leqq 1) \end{cases},$$

$$g(x) = \begin{cases} \frac{5}{2}(x-\frac{2}{5})+2 & (0 \leqq x \leqq \frac{2}{5}) \\ -\frac{5}{3}(x-\frac{2}{5})+2 & (\frac{2}{5} < x \leqq 1) \end{cases},$$

$$h(x) = \begin{cases} \frac{5}{3}(x-\frac{3}{5})+2 & (0 \leqq x \leqq \frac{3}{5}) \\ -\frac{5}{2}(x-\frac{3}{5})+2 & (\frac{3}{5} < x \leqq 1) \end{cases},$$

$$\int_0^1 f(x)\,dx = \frac{3}{2}, \int_0^1 g(x)\,dx = \frac{5}{3}, \int_0^1 h(x)\,dx = \frac{5}{3}$$

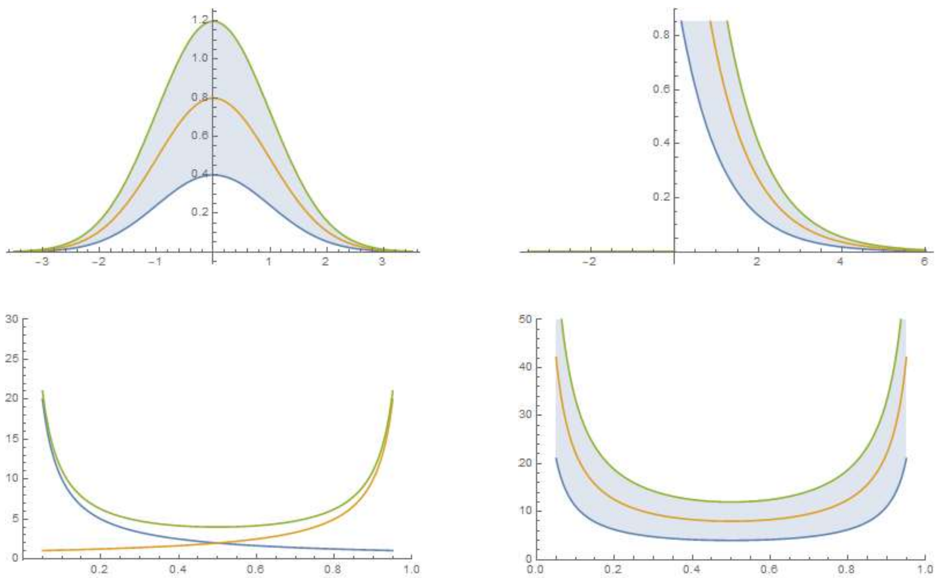


Figure 1: examples of intervals by proportional measure (normal (Gauss), exponential and heavy tails on $x = 0, 1$ distributions)

5 Numerical Examples

Here, we shall give numerical examples which illustrates interval estimated MDPs.

The numberof states $n := 3$, $S = \{1,2,3\}, P_3 := \{p = (p_1,p_2,p_3) | \sum_{i=1}^{3} p_i = 1, p_i \geq 1, i = 1,2,3\}$. Interval of prior measure:$[L, 2L]$ $(k = 2)$, where L: Lebesgue measure. For some fixed state i_0, we repeat the experiment $\hat{\sigma} = 6$ times and count the transition from state i_0 to the next state $i = 1,2,3$. Let σ_i be the number of outcomes of transition to state i from the fixed state i_0.

(Case 1: mean of $Q_\sigma \in [L_\sigma, kL_\sigma]$)
We denote by $\sigma = (\sigma_1, \sigma_2, \sigma_3) = (3,1,2)$ the results of experiments(data set). Let $\hat{\sigma} = \sum_{i=1}^{3} \sigma_i = 6, s = \sigma_1 + 1 = 4, t = \sigma_2 + \sigma_3 + (n-1) = 5$. Posterior intervals $[\underline{\lambda}_i, \overline{\lambda}_i]$ of p_i are given by solving $\hat{\sigma} + n$-degree polynomial equation. For example, $\underline{\lambda}_1$ is the solution of the following equation:

$$\left(\frac{4}{6+3} - \lambda\right) B(4,5) + \left(\sum_{i=0}^{4} \binom{4}{i} (-1)^{i+1} \lambda^{5+i} \left(\frac{1}{(4+i)(5+i)}\right)\right) = 0. \tag{29}$$

It is simplified $4 - 9\lambda - \lambda^5(126 - 336\lambda + 360\lambda^2 - 180\lambda^3 + 35\lambda^4) = 0$. We have the solution $\underline{\lambda} \risingdotseq 0.400$. Similarly, $\overline{\lambda}_1$ is the solution of polynomial equation as follows:

$$2\left(\frac{4}{6+3} - \lambda\right) B(4,5) - \left(\sum_{i=0}^{4} \binom{4}{i} (-1)^{i+1} \lambda^{5+i} \left(\frac{1}{(4+i)(5+i)}\right)\right) = 0, \tag{30}$$

or equivalently, $8 - 18\lambda + \lambda^5(126 - 336\lambda + 360\lambda^2 - 180\lambda^3 + 35\lambda^4) = 0$ Then, we have solution $\overline{\lambda}_1 \risingdotseq 0.489$.

(Case 2: lower and upper α-percentile of p_i from posterior measures $Q_\sigma \in [L_\sigma, kL_\sigma]$) The frequencies for each state under the experiments of step numbers $n = 100, 600$ and 2000 are as follows:

$$\begin{pmatrix} 21 & 7 & 17 \\ 8 & 7 & 8 \\ 15 & 9 & 8 \end{pmatrix}, \begin{pmatrix} 175 & 50 & 98 \\ 55 & 19 & 25 \\ 93 & 30 & 55 \end{pmatrix} \text{ and } \begin{pmatrix} 442 & 129 & 286 \\ 182 & 193 & 183 \\ 232 & 237 & 116 \end{pmatrix}.$$

From Theorem 3.2, we have the data table of estimated $[\underline{p}_i(\alpha/2), \overline{p}_i(\alpha/2)]$ interval matrices as follows:
Step number $n = 100$:

$$\begin{pmatrix} [0.3218, 0.5980] & [0.0768, 0.2823] & [0.2457, 0.5140] \\ [0.1805, 0.5340] & [0.1502, 0.4928] & [0.1805, 0.5339] \\ [0.2986, 0.6203] & [0.1516, 0.4427] & [0.1294, 0.4108] \end{pmatrix}. \tag{31}$$

Step number $n = 600$:

$$\begin{pmatrix} [0.4618, 0.5909] & [0.1034, 0.1945] & [0.2682, 0.3895] \\ [0.2756, 0.4255] & [0.3507, 0.5064] & [0.3301, 0.4968] \\ [0.3301, 0.4968] & [0.3301, 0.4968] & [0.1159, 0.2445] \end{pmatrix}. \tag{32}$$

Step number $n = 2000$:

$$\begin{pmatrix} [0.4819, 0.5483] & [0.1281, 0.1757] & [0.3028, 0.3654] \\ [0.2883, 0.3653] & [0.3072, 0.3854] & [0.2900, 0.3671] \\ [0.3573, 0.4359] & [0.3657, 0.4445] & [0.1679, 0.2320] \end{pmatrix}. \tag{33}$$

Figure 2 shows first row of interval matrices in the case of $\alpha = 0.05, n = 100, 600, 2000$.

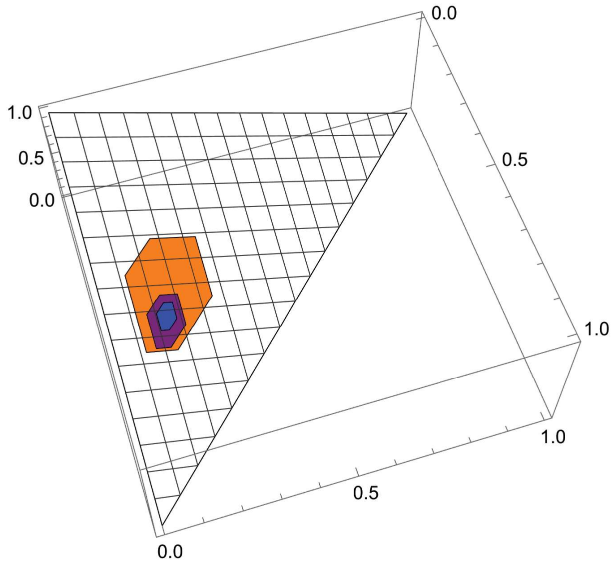


Figure 2: $\alpha = 0.05$, first row of interval matrices (n=100,600,2000)

For each interval matrices (31), (32) and (33), it can be formulated optimality equations (17) and (18). And from Theorem 20, we have Pareto-optimal policy $f^* \in F$ by solving those equations (17) and (18).

In the reminder of this section, the principal significance of the central $100(1-\alpha)$-credible interval $[\underline{p}_{1i}(\alpha), \overline{p}_{1i}(\alpha)]$ for $p_{ij} (i, j \in S)$. For example, the first row $([0.3218, 0.5980], [0.0768, 0.2823], [0.2457, 0.5140])$ of interval matrix (31) is given by (28). The first element $[0.3218, 0.5980]$ is solutions of following equations: Set $\alpha = 0.05, s-1 = 21, t-1 = 25, k = 2$ and

$$\frac{B(s, t|\underline{p}_{1i}(\alpha))}{B(s,t)} = \frac{\alpha}{\alpha + (1-\alpha)k}, \tag{34}$$

$$\frac{B(s, t|\overline{p}_{1i}(\alpha))}{B(s,t)} = \frac{(1-\alpha)k}{\alpha + (1-\alpha)k}. \tag{35}$$

Since $f(x) = \dfrac{B(s,t|x)}{B(s,t)}$ is incomplete beta function so that the solution of equations (34) and (35) are obtained by the values of right hand side values $\dfrac{\alpha}{\alpha + (1-\alpha)k}$ and $\dfrac{(1-\alpha)k}{\alpha + (1-\alpha)k}$.

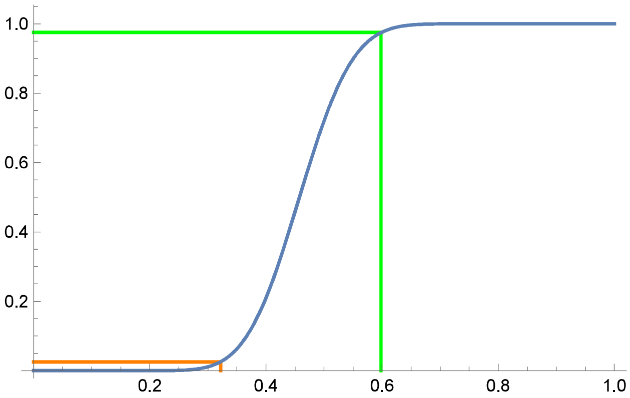


Figure 3: Incomplete beta function $f(x)$ for $\alpha = 0.05, s-1 = 21, t-1 = 25, k = 2$.

It can be also shown that the inverse function $f_{1i}^{-1}(x)$ of $f_{1i}(x)$ for getting each element $\underline{p}_{1i}(\alpha)$ have their curves of $f_{1i}^{-1}(x)$ as follows:

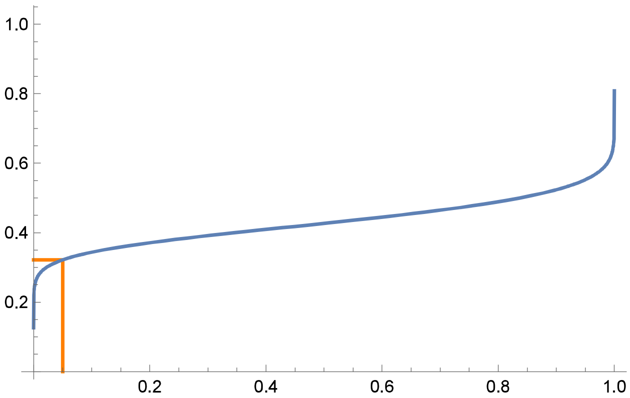


Figure 4: Inverse function $f_{11}^{-1}(x)$ of incomplete beta function $f_{11}(x)$ for $\alpha = 0.05, s-1 = 21, t-1 = 25, k = 2$.

In addition, the values $\overline{p}_{1i}(\alpha)$ are characterized by the value $f_{11}^{-1}(1-x)$ and the following curve of its function $f_{11}^{-1}(1-x)$.

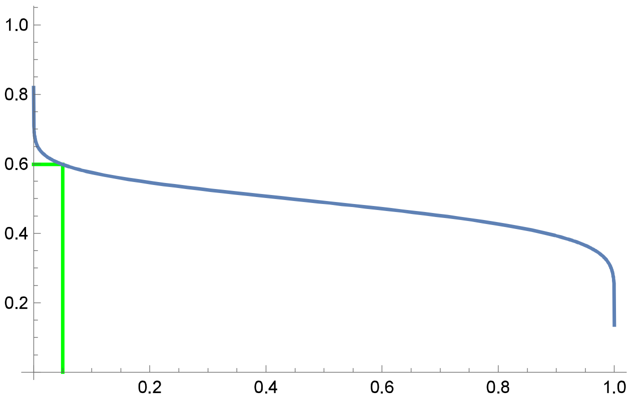


Figure 5: Inverse function $f_{11}^{-1}(1-x)$ of incomplete beta function $f_{11}(x)$ for $\alpha = 0.05, s-1 = 21, t-1 = 25, k = 2$.

Finally, it is noted that the 95% highest density interval for Beta distribution $Beta(21, 25)$ is $[0.315349, 0.598731]$ and our 95% central percentile $[0.3218, 0.5980]$ is a very close interval so that it may be applicable to Bayesian inference and testing hypothesis problems.

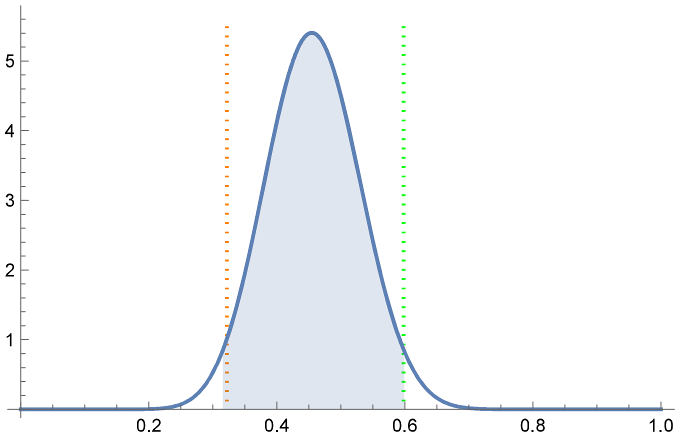


Figure 6: 95% highest density interval $[0.315349, 0.598731]$ and our estimated interval $[0.3218, 0.5980]$(between the dotted vertical lines).

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

[8] 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.

[9] 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.

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

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

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

[13] 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.)

Masayuki Horiguchi
Department of Mathematics,
Faculty of Science, Kanagawa University
Address: Rokkakubashi 3-27-1, Kanagawa-ku, Yokohama
Kanagawa Prefecture, 221-8686, Japan
E-mail address: horiguchi@kanagawa-u.ac.jp

Masahiko Sakaguchi
Department of Information Science,
Faculty of Science and Engineering, Kochi University
Address:2-5-1akebono-cho, kochi-shi,
Kochi Prefecture 780-8520, Japan
E-mail address: sakaguchim@kochi-u.ac.jp

神奈川大学・理学部　堀口正之
高知大学・理工学部　阪口昌彦