区間ベイズ手法と分数計画問題について
(On an interval Bayesian method and fractional programming problems)

神奈川大学・理学部・堀口正之
Masayuki HORIGUCHI[(1)]
神奈川大学大学院・理学研究科・王 瀚東, 鈴木 陸斗, 源 隆哉
Wang Handong[(2)], Rikuto Suzuki[(2)], Ryuya Minamoto[(2)]

[(1)]Professor of Department of Mathematics, Faculty of Science, Kanagawa University
[(2)]Students of Graduate School of Science, Kanagawa University

Abstract

In Bayesian Markov decision model, the transition law of process is expressed as matrices in which each element is interval and estimated from the observation of pair of state-action transitions. In this report, we consider an optimization problems of fractional programming and show an algorithm of solving the equations in order to have the interval estimated transition matrices in Bayesian Markov decision model.

1 Introducution

Markov decision model consists of four tuple $\{S, A, Q, \boldsymbol{r}\}$. $S := \{1, 2, \ldots, n\}$ denotes finite state space and $A := \{a_1, a_2, \ldots, a_k\}$ denotes finite action space. We define the followings:

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

$$P(S|S) :=\{q = (q_{ij} : i, j \in S) \in \mathbb{R}_+^{n \times n} | \sum_{j \in S} q_{ij} = 1 \ (i \in S)\}, \tag{2}$$

$$P(S|S \times A) :=\{Q = (q_{ij}(a) : i, j \in S, a \in A) \in \mathbb{R}_+^{kn \times n} | q_{i\cdot}(a) \in P(s) \ (i \in S, a \in A)\}, \tag{3}$$

where $\mathbb{R}_+^n$ is the set of nonnegative n dimensional real vectors and and $\mathbb{R}_+^{m \times n}$ is the set of (m, n) nonnegative real matrices. $Q = (q_{ij}(a)) \in P(S|S \times A)$ denotes transition laws, $\boldsymbol{r} = (r(i, a)) \in B_+(S \times A)$ denotes reward function, where $B_+(D)$ is the set of nonnegative real valued functions on D. In our decision model, we consider an MDPs with unkown transition laws of matrices $Q = (q_{ij}(a))$ $(i, j \in S, a \in A)$ (Uncertain MDPs) and each estimated transition law has an element of interval s.t. $[\underline{p}_{ij}(a), \overline{p}_{ij}(a)]$.

We consider a Markov decision process whose true transition matrices are denoted by $P = (p_{ij}(a))$, $i, j \in S, a \in A$, where S denotes finite state space and A denotes finite action space. For each ith row of transition matrices, the data set of obserbations are denoted by

$$\sigma_i(a) = (\sigma_{i1}(a), \sigma_{i2}(a), \ldots, \sigma_{in}(a)) \quad a \in A,$$

and set $\hat{\sigma}_i(a) = \sum_j \sigma_{ij}(a), a \in A$.

For ith distribution of true transition matrices $p_i(a) = (p_{i1}(a), p_{i2}(a), \ldots, p_{in}(a)), a \in A$, the distribution of $\sigma_i(a)$ is expressed as multinomial s.t.

$$f(\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{in}) = \frac{\hat{\sigma}_{i1}!}{\sigma_{i1}!\sigma_{i2}! \cdots \sigma_{in}} p_{i1}^{\sigma_{i1}} p_{i2}^{\sigma_{i2}} \cdots p_{in}^{\sigma_{in}},$$

where $\sigma_{ij}(a)$ and $p_{in}^{\sigma_{in}(a)}(a)$ are denoted by σ_{ij} and $p_{in}^{\sigma_{in}}$, respectively. Therefore, from the prior information of the observed date set $\hat{\sigma}_i, i \in S$, we treat Dirichlet distribution of unknown parameters $(p_{ij}(a))$ in order to have estimated interval matrices of true transition matrices $(p_{ij}(a)) \in P(a), a \in A$.

For the simplicity, we treat deterministic and stationary MDPs since in this report we focus on an algorithms to have solutions of equations which gives interval estimated MDPs from the date set of observations. Hence, from now on, we denote by $P = (p_{ij})$ an estimated transition matrices under a fixed deterministic stationary policy.

Let the map $f : S \to A$ and denote by F the set of all maps f. For any $f \in F$, the expected discounted reward function is denoted by $\phi(f|Q) \in \mathbb{R}^n_+$ and defined as follows for the discount factor β $(0 < \beta < 1)$ with transition matrices $Q \in P(S|S \times A)$:

$$\phi(f|Q) = \sum_{t=0}^{\infty} (\beta Q(f))^t \boldsymbol{r}(f), \tag{4}$$

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

In practical decision problem, we must estimate the true transition Q from the data set of state-action observation of transitions in the system. It is called controlled Markov set-chain (cf. [10]) if the transition matrices are given by the set of interval matrices. In this report we consider how to estimate those interval matrices from the data set of observations.

We assume the process start some fixed initial state i and estimate transition law $\{p_i\}_{i\in S}$ from the date set of prior information.

Let $P_n := P(S) = \{p = (p_1, p_2, \ldots, p_n)|p_i \geq 0, \sum_{i=1}^n p_i = 1\}$. We denote by $\mathcal{B}$ the set of all Borel measurable set of $\mathbb{R}^n$. Let us denote by $L \leqq B$ if there exist two measure L and U on $\mathcal{B}$ which satisfy $L(A) \leqq U(A)$ for any $A \in \mathcal{B}$. Here and subsequently, for such two measures $L \leqq U$, we denote by $[L, U]$ and call them prior interval measure. We also call L lower measure and U upper measure to $[L, U]$. In addition, throughout the paper we assume upper measure U satisfy $U(\cdot) = kL(\cdot)$ for fixed real value $k(k \geqq 1)$. That is, U is assumed proportional measure of L and prior interval measure is denoted by $[L, kL] = [dp, k\,dp]$.

For fixed initial state $k \in S$, we execute trials independently and repeatedly by $\hat{\sigma}$ times and observe the transitions from fixed state k to each state i and record the number σ_i of transitions to i. Then we have the date set $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, where $\hat{\sigma} = \sum_{k=1}^n \sigma_k$. Let p_i the transition probability from fixed initial state $k \in S$ to $i \in S$. For $\hat{\sigma}$ and parameter $p = (p_1, p_2, \ldots, p_n)$ the distribution of $\hat{\sigma}$ is multinomial and represented as follows

$$f(\sigma_1, \sigma_2, \ldots, \sigma_n|p) = \frac{(\sigma_1 + \cdots + \sigma_n)!}{\sigma_1! \cdots \sigma_n!} p_1^{\sigma_1} p_2^{\sigma_2} \cdots p_n^{\sigma_n}. \tag{5}$$

By applying the result of method of interval Bayesian estimation by DeRobertis/Hartigan[1] to prior interval $[L, kL]$, we have posterior interval measures $[L_\sigma, U_\sigma] := [L_\sigma, kL_\sigma]$ which is expressed by multidimensional beta function and the posterior interval measure $[\underline{\lambda}_i, \overline{\lambda}_i]$ (for simplicity, we denote it by$[\underline{\lambda}, \overline{\lambda}]$) is constructed as intervals of integral proportions by posterior interval measures $Q \in [L_\sigma, kL_\sigma]$ such as

$$\left\{ \frac{\int_{P_n} p_i Q(dp)}{\int_{P_n} Q(dp)} \middle| L_\sigma \leqq Q \leqq U_\sigma \right\}. \tag{6}$$

Moreover, each value of end points of interval $[\underline{\lambda}, \overline{\lambda}]$ is unique solution of equations:

$$U_\sigma(p_i - \underline{\lambda})^- + L_\sigma(p_i - \underline{\lambda})^+ = 0, \tag{7}$$

$$U_\sigma(p_i - \overline{\lambda})^+ + L_\sigma(p_i - \overline{\lambda})^- = 0 \tag{8}$$

where $x^+ = \max\{0, x\}, x^- = x - x^+ = \min\{0, x\}$.

2 Dirichlet integral and the solutions of estimated parameters

We can rewrite above equations (7) and (8) as the form of Dirichlet integrals below from the assumption $U_\sigma = kL_\sigma$ and by multidimensional beta function(Dirichlet function):

(lower bound $\underline{\lambda}$):

$$k\int\cdots\int_{0\le p_i\le\lambda,p\in P_n}(p_i-\lambda)p_1^{\sigma_1}\cdots p_n^{\sigma_n}\,dp+\int\cdots\int_{\lambda\le p_i\le 1,p\in P_n}(p_i-\lambda)p_1^{\sigma_1}\cdots p_n^{\sigma_n}\,dp=0 \tag{9}$$

(upper bound $\overline{\lambda}$):

$$k\int\cdots\int_{\lambda\le p_i\le 1,p\in P_n}(p_i-\lambda)p_1^{\sigma_1}\cdots p_n^{\sigma_n}\,dp+\int\cdots\int_{0\le p_i\le\lambda,p\in P_n}(p_i-\lambda)p_1^{\sigma_1}\cdots p_n^{\sigma_n}\,dp=0 \tag{10}$$

Gamma function $\Gamma(x)$ $(x>0)$ and beta function $B(x,y)$ $(x,y>0)$, Incomplete beta function $B(x,y|\lambda)$ $(x,y>0, 0\leqq\lambda\leqq 1)$ are defined as follows: $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,dt\quad(x>0), B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}\,dt\quad(x,y>0), B(x,y|\lambda)=\int_0^\lambda t^{x-1}(1-t)^{y-1}dt\quad(x,y>0,0\leqq\lambda\leqq 1)$.

We will denote Dirichlet integral by $D(\nu_1,\nu_2,\ldots,\nu_k;\nu_{k+1})$ and $D(\nu_1,\nu_2,\ldots,\nu_k;\nu_{k+1}|\lambda)$ $(k\geqq 1, 0\leqq\lambda\leqq 1)$ which are defined as follows:

$$\begin{aligned}&D(\nu_1,\nu_2,\ldots,\nu_k;\nu_{k+1})\\&:=\int\cdots\int_{S_k}x_1^{\nu_1-1}x_2^{\nu_2-1}\cdots x_k^{\nu_k-1}(1-x_1-x_2-\cdots-x_k)^{\nu_{k+1}-1}\,dx_1dx_2\cdots dx_k\\&D(\nu_1,\ldots,\nu_k;\nu_{k+1}|\lambda)\\&:=\int\cdots\int_{S_k\cap\{0\leqq x_1\leqq\lambda\}}x_1^{\nu_1-1}\cdots x_k^{\nu_k-1}(1-x_1-\cdots-x_n)^{\nu_{k+1}-1}\,dx_1\cdots dx_k,\end{aligned}$$

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

By the relations of Dirichlet integrals and beta functions we have

$$D(\nu_1,\nu_2,\ldots,\nu_k;\nu_{k+1})=B(\nu_1,\nu_2+\cdots+\nu_{k+1})D(\nu_2,\nu_3,\ldots,\nu_k;\nu_{k+1}) \tag{11}$$

$$D(\nu_1,\nu_2,\ldots,\nu_k;\nu_{k+1}|\lambda)=B(\nu_1,\nu_2+\cdots+\nu_{k+1}|\lambda)D(\nu_2,\nu_3,\ldots,\nu_k;\nu_{k+1}). \tag{12}$$

Then, we can show that the solutions of the equations (7) and (8) are solutions λ of polynomial equations (13) and (14) in the followings:

$$K(s,t,\lambda):=\left(\frac{s}{s+t}-\lambda\right)B(s,t)+(k-1)\left(B(s+1,t|\lambda)-\lambda B(s,t|\lambda)\right)=0, \tag{13}$$

$$G(s,t,\lambda):=k\left(\frac{s}{s+t}-\lambda\right)B(s,t)-(k-1)\left(B(s+1,t|\lambda)-\lambda B(s,t|\lambda)\right)=0, \tag{14}$$

where $\hat{\sigma}=\sum_{i=1}^n\sigma_i, s=\sigma_1+1, t=\hat{\sigma}-\sigma_1+n-1$. The above equations (13) and (14) have the same dimension $(s+t=\hat{\sigma}+n)$of polynomials respectively. We have the following.

Theorem 1. *For a parameter $p=(p_1,p_2,\ldots,p_n)$, each p_i of posterior interval measure $[\underline{\lambda},\overline{\lambda}]$ is given by the unique solutions of the following polynomial equations.*

$$K(s,t,\underline{\lambda}):=B(s+1,t)-\underline{\lambda}B(s,t)+(k-1)\left(B(s+1,t,\underline{\lambda})-\underline{\lambda}B(s,t,\underline{\lambda})\right)=0, \tag{15}$$

$$G(s,t,\overline{\lambda}):=k(B(s+1,t)-\overline{\lambda}B(s,t))-(k-1)\left(B(s+1,t,\overline{\lambda})-\overline{\lambda}B(s,t,\overline{\lambda})\right)=0, \tag{16}$$

where $s=\sigma_i+1$ and $t=\hat{\sigma}-\sigma_i+n-1$.

It can be shown that [7] the functions $K(s,t,\lambda)$ and $G(s,t,\lambda)$ are concave and convex in λ and have the unique solutions $\underline{\lambda}$ and $\overline{\lambda}$ in the closed interval $[0,1]$ respectively.

It is noted that incomplete beta function $B(m, n, \lambda)$ can be rewritten by polynomial as the following: $B(m, n, \lambda) = \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \lambda^{m+i} \frac{1}{m+i}$.

In order to have the solution $\underline{\lambda}$, we apply the Newton-Raphson algorithm([7]) as follows:

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 $W(s, t, x_n) := \frac{K(s, t, x_n)}{K'(s, t, x_n)}$. 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,

where

$$W(s, t, x_n) := -\frac{(\frac{s}{s+t} - x_n)B(s, t) + (k-1)(\frac{s}{s+t} - x_n)B(s, t, x_n) - \frac{k-1}{s+t} x_n^s (1 - x_n)^t}{B(s, t) + (k-1)B(s, t, x_n)}. \tag{17}$$

Let f be a differentiable function and $\phi(x) = x - \frac{f(x)}{f'(x)}$. The recursive formula of Newton-Raphson method $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ gives approximate sequence of the root λ which satisfies $f(\lambda) = 0$. It is also known that a fixed point λ of $\phi(\lambda) = \lambda$ is also solution of the equation $f(\lambda) = 0$ if $f'(\lambda) \neq 0$.

For a differentiable function K, we have the followong:

Proposition 1.

$$K(s, t, \lambda) = \lambda K'(s, t, \lambda) - K'(s+1, t, \lambda) \tag{18}$$

$$\frac{K(s, t, \lambda)}{K'(s, t, \lambda)} = \lambda - \frac{K'(s+1, t, \lambda)}{K'(s, t, \lambda)} \tag{19}$$

From the above proposition, we have a fractional programming problem related to finding the fixed point of ϕ defined by

$$\phi(\lambda) = \frac{K'(s+1, t, \lambda)}{K'(s, t, \lambda)}, \tag{20}$$

and convergent sequence to fixed point λ is defined by

$$x_{n+1} = \phi(x_n) = \frac{K'(s+1, t, x_n)}{K'(s, t, x_n)} = \frac{B(s+1, t) + (k-1)B(s+1, t, x_n)}{B(s, t) + (k-1)B(s, t, x_n)}. \tag{21}$$

For the first derivative of $\phi(\lambda)$, it is easily shown that

$$\phi'(\lambda) = \frac{-(k-1)\lambda^{s-1}(1-\lambda)^{t-1}}{K'(s, t, \lambda)}(\lambda - \phi(\lambda)). \tag{22}$$

Let $\phi(\lambda) = K'(s+1, t, \lambda)/K'(s, t, \lambda)$ on $\lambda \in [0, 1]$. We have the followings.

Proposition 2. *(i)* $\phi(0) = \phi(1) = \frac{B(s+1, t)}{B(s, t)}$

(ii) The solutions λ *for equation* $\phi'(\lambda) = 0$ *are* $\lambda = 0, 1$ *and fixed point of* $\phi(\lambda) = \lambda$*, i.e., the fixed point* λ *is solution of equation (15)* $K(s, t, \lambda) = 0$.

(iii) The function $\phi(\lambda)$ *is monotone decreasing in* $0 < \lambda < \alpha$ *and monotone increasing in* $\alpha < \lambda < 1$.

We can define the optimization problems as follows.

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

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

Then,

Proposition 3. *(1)* $\phi(\underline{\lambda}) = \min_{\lambda\in[0,1]} \dfrac{K'(s+1,t,\lambda)}{K'(s,t,\lambda)} = \dfrac{K'(s+1,t,\underline{\lambda})}{K'(s,t,\underline{\lambda})} = \underline{\lambda}$

(2) $F(q)$ *is strictly monotone decreasing and concave in* q.

(3) $\underline{\lambda}$ *is unique solution to equation* $F(q) = 0$ $q \in [0,1]$.

By the proposition 3, we have a property of relation between a solution of polynomial $K(s,t,\lambda) = 0$ and fractional programing (P).

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

It can be shown similarly that the solution of the equation (16) $G(s,t,\lambda) = 0$ is characterised as the optimal value of fractional programming.

3 An approach by Dinkelbach algorithm

In this section, we summarise an approach to find the solutions $\underline{\lambda}_{ij}$ and $\overline{\lambda}_{ij}$ for $i,j \in S$ of $K = 0$ and $G = 0$ by applying Dinkelbach algorithm (cf. Stancu-Minasian 1997[17]).

Let $X \subset \mathbb{R}$ and $f, g : X \to \mathbb{R}$. We define optimization problems as follows.

Primary problem (P)

$$\max q(x) = \frac{f(x)}{g(x)}, x \in X,$$

where $g(x) > 0$, , $x \in X$ and $f(x) \geqq 0$ at least one $x \in X$.

Parametric problem $Q(\lambda)$

$$\max f(x) - \lambda g(x), x \in X.$$

Let $\overline{x}$ is optimal solution to the primal (P) and assume $\overline{\lambda}$ satisfy the condition $\overline{\lambda} = \dfrac{f(\overline{x})}{g(\overline{x})}$. Then, it is known that the following theorem follows (e.g. [17]).

Theorem 2. *(i)* $F(\lambda) > 0$ *if and only if* $\lambda < \overline{\lambda}$. *(ii)* $F(\lambda) = 0$ *if and only if* $\lambda = \overline{\lambda}$. *(iii)* $F(\lambda) < 0$ *if and only if* $\lambda > \overline{\lambda}$.

Proof. (i) ("only if" part) If $F(\lambda) > 0$, $\exists x' \in X$ s.t. $f'(x) - \lambda g(x') > 0$. Since we assume $g(x) > 0$ so that we have $\frac{f(x')}{g(x')} > \lambda$. Then, $\lambda < \overline{\lambda}$ follows. ("if" part) If $\lambda < \overline{\lambda}$, $\exists x'' \in X$ s.t. $\overline{\lambda} = \frac{f(x'')}{g(x'')} > \lambda$. Hence, we have $f(x'') - \lambda g(x'') = F(\lambda) > 0$. (ii), (iii) can be shown analogously.■

We have the following.

Corollary 2. *If* $F(\lambda) = 0$, *the optimal solution* x' *to Parametric problem* $Q(\lambda)$ *satisfies* $f(x') - \lambda g(x') = F(\lambda) = 0$. *Moreover, since* $\frac{f(x')}{g(x')} = \overline{\lambda}$, x' *is also optimal solution to Primly problem (P).*

Now by applying Newton-Raphson method to the function $F(\lambda)$ we have the following argorithm (cf. [17]).

Dinkelbach Algorithm

Step 1 Take $\lambda = \lambda_i$ such that $F(\lambda_i) \geqq 0$.

Step 2 Solve problem $Q(\lambda)$. If $|F(\lambda)| \leqq \delta$, stop. Otherwise, go to Step 3.

Step 3 Let $\lambda = \frac{f(x^*)}{g(x^*)}$, where x^* is an optimal solution of problem $Q(\lambda)$ obtained in Step 2. Repeat Step 2.

4 Numerical example

Let $[L, kL]$ $(k \geqq 1)$ be prior interval measure and set $s = \sigma_{ij} + 1$ and $t = \hat{\sigma}_i - \sigma_{ij} + n - 1$ where $\hat{\sigma}$ is observed data set for fixed initial state $i \in S$ and transition state $j \in S$ from i. For each $q_{ij}(a), a \in A$ of true transition matrix $Q(a) = (q_{ij}(a)), a \in A$, $q_{ij}(a)$ is estimated as the interval $[\underline{\lambda}_{ij}(a), \overline{\lambda}_{ij}(a)]$ by applying Bayesian interval method with observation data set $\hat{\sigma} = \hat{\sigma}(i, a) = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ for each $i \in S$ and $a \in A$.

For simplicity, we fixed initial state i and $a \in A$ in order to show our estimating algorithm works to each true transition law $q_i(a) = (q_{ij}(a)) \in \mathbb{R}^n$.

In this section we show the numerical example related to estimated upper values $\overline{\lambda}_{ij}$. For parametric problem $Q(\lambda)$, let

$$F(\lambda) = (kB(s+1, t) - (k-1)B(s+1, t, x) - \lambda\,(kB(s, t) - (k-1)B(s, t, x))\,.$$

If we set parameters as $k = 2, s = \sigma_{ij} + 1 = 11, t + 1 = \hat{\sigma}_i - \sigma_{ij} + n = 5 + 1$, $F(\lambda)$ have the form as follows:

$$F(\lambda) = (2B(12, 5) - B(12, 5, x)) - \lambda(2B(10, 5) - B(10, 5, x)).$$

Applying our algorithm (Algorithm A) as mentioned in preceding section, we have a solution $\overline{\lambda}_{ij} = 0.71826$. On the other hand, by applying Dinkelbach algorithm above, we have the sequences $\{\lambda_0, \lambda_1, \lambda_2, \ldots\}$ in the following.

$\lambda_0 = 0.7 \longrightarrow \lambda_1 = 0.7007, \longrightarrow \lambda_2 = 0.71790, \longrightarrow \lambda_3 = 0.720102$. It is noted that $\lambda_3 > \overline{\lambda}_{ij}$ is shown by $F(\lambda_3) < 0$.

Here is another example, if we set initial value $\lambda_0 = 0.6$, then $\lambda_0 = 0.6 \longrightarrow \lambda_1 = 0.60454, \longrightarrow \lambda_2 = 0.70769, \longrightarrow \lambda_3 = 0.70777, \longrightarrow \lambda_4 = 0.718131, \longrightarrow \lambda_5 = 0.718978$. We know that $\lambda_5 > \overline{\lambda}_{ij}$ from $F(\lambda_5) < 0$.

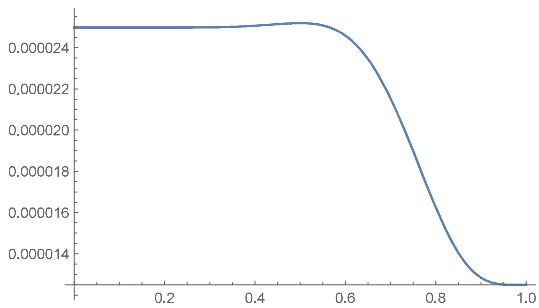

Figure 1: Trajectory of $F(\lambda) = f(x) - \lambda g(x)$ with fixed $\lambda = 0.5$ and $x \in [0, 1]$. Solution x' of parametric problem $Q(\lambda) = Q(0.5)$ gives maximal point of $F(\lambda)$.

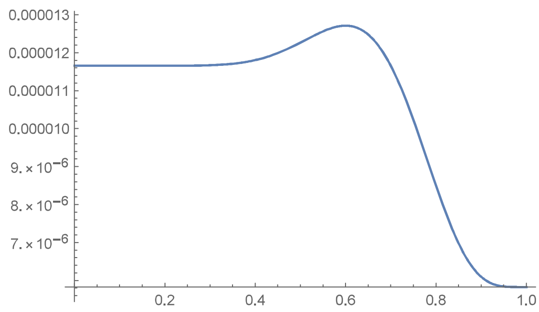

Figure 2: Trajectory of $F(\lambda) = f(x) - \lambda g(x)$ with fixed $\lambda = 0.6$ and $x \in [0, 1]$.

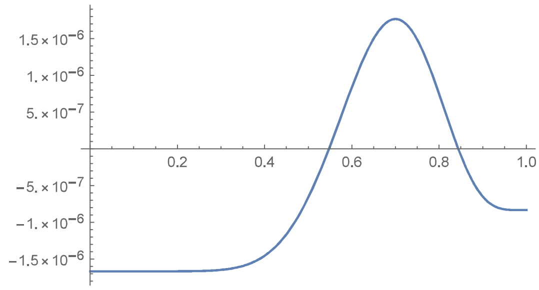

Figure 3: Trajectory of $F(\lambda) = f(x) - \lambda g(x)$ with fixed $\lambda = 0.7$ and $x \in [0, 1]$.

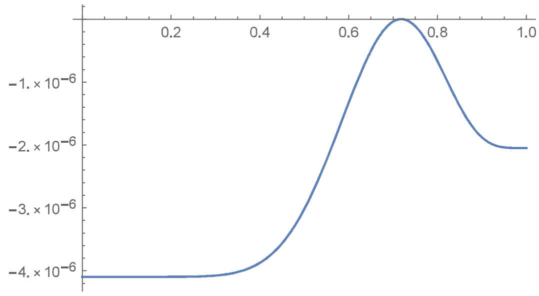

Figure 4: Trajectory of $F(\lambda) = f(x) - \lambda g(x)$ with fixed $\lambda = 0.71826$ and $x \in [0, 1]$. Solution $x' = 0.71826$ gives maximal point of $F(\lambda)$ and satisfying $F(\lambda) = 0$.

Figure 1 to 3 show trajectories of Function $F(\lambda)$ for each $\lambda = 0.5, 0.6, 0.7$. Figure 4 shows trajectories of $F(\lambda)$ with the optimal value of fractional programming (P) and the optimal value $\lambda = \overline{\lambda} = 0.71826$ is characterised as the solution of the equation $G(\lambda) = 0$ which is estimated upper bound of true transition matrix $q_{ij}(a)$.

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] Dynkin, E.B., Yushkevich, A.A.: Controlled Markov Processes Springer, New York (1979)

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

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

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

[7] M. Horiguchi. Bayes estimated intervals and uncertain MDPs. *Surikaisekikenkyusyo Kokyuroku*, 1682:70–77, 2010.

[8] 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.

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

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

[11] 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.

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

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

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

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

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

[17] I.M. Stancu-Minasian. *Fractional Programming: Theory, Methods and Applications* . Kluwer Academic Publishers, Dordrecht, 1997.

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

[19] 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

Wang Handong, Rikuto Suzuki and Ryuya Minamoto
Students of Graduate School of Science, Kanagawa University

神奈川大学・理学部　堀口正之
神奈川大学・理学研究科　王 瀚東, 鈴木 陸斗, 源 隆哉