数理解析研究所講究録
第1636巻 2009年 9-16

全期間依存型制約つき決定過程
ー確率的推移システム上での加法型制約ー

九州工業大学・大学院工学研究院　藤田 敏治 (Toshiharu Fujita)
Graduate School of Engineering, Kyushu Institute of Technology
九州工業大学・大学院工学研究科　千布 裕樹 (Yuuki Chibu)
Graduate School of Engineering, Kyushu Institute of Technology

1　はじめに

本論文では，加法型の全期間依存型制約をもつ確率システム上の決定過程問題に対し，動的計画法による再帰式を導く．全期間依存型制約とは，制約中に，通常の目的関数と同形あるいは類似した形の関数が現れるものである．例えば，時間と費用が与えられた最適ルート問題において，費用を一定以下に抑えつつ，所要時間を最小化するといった問題である．ここでは，特に加法型制約を取り上げ，得られた結果を報告する．

動的計画法は，最適性の原理をその基本原理とし，様々な問題に対する再帰的アプローチを与える強力な手法である．R. E. Bellman ([1]) により創出され，幅広い分野において研究・応用がなされてきた．離散・連続，有限期間・無限期間，最適化・非最適化を問わず，また想定する推移も確定・確率のみならず，ファジィ([2, 5]) へ，そして最近では非決定性推移 ([3]) までも幅をひろげ，その応用分野は，理学，工学，経済学と多岐にわたる．この動的計画の適用範囲を広げることが我々の一連の研究の目的である．

以下，本文中で用いる記号と用語を述べる．ただし，$\mathbf{R}$ は実数全体を表すものとする．

(1) $N \geq 2$ は段の総数を表す正整数

(2) $X = \{s_1, s_2, \ldots, s_l\}$ は有限状態空間　　(3) $U = \{a_1, a_2, \ldots, a_k\}$ は有限決定空間

(4) p はマルコフ推移法則; $p(y|x,u) \geq 0 \quad \forall (x,u,y) \in X \times U \times X$, $\sum_{y \in X} p(y|x,u) = 1 \quad \forall (x,u) \in X \times U$;

$p(y|x,u)$ は，状態 x で決定 u をとったとき，次の状態が y になる条件付き確率を表す．
この p により表される確率的推移を $y \sim p(\cdot|x,u)$ と表現する．

(5) $r_n : X \times U \to \mathbf{R}$ は第 n 利得関数 $(n = 0, 1, 2, \ldots, N-1)$

(6) $q_n : X \times U \to \mathbf{R}$ は第 n 損失関数 $(n = 0, 1, 2, \ldots, N-1)$

(7) $r_G : X \to \mathbf{R}$ は終端利得関数　　(8) $q_G : X \to \mathbf{R}$ は終端損失関数

2　全期間依存型制約つき決定過程

2.1　定式化

初期状態 x_0 は与えられるものとする．決定 $u_0 \in U$ をとるとき，次期の状態 x_1 は確率的推移：$x_1 \sim p(\cdot|x_0, u_0)$ に従う．同様に繰り返し，終了期までの状態決定列：

$$u_1 \in U,\ x_2 \sim p(\cdot|x_1, u_1),\ u_2 \in U,\ x_3 \sim p(\cdot|x_2, u_2), \ldots,\ u_{N-1} \in U,\ x_N \sim p(\cdot|x_{N-1}, u_{N-1})$$

を得る．ここで，評価として利得の総和の期待値：

$$E[r_0(x_0, u_0) + r_1(x_1, u_1) + \cdots + r_{N-1}(x_{N-1}, u_{N-1}) + r_G(x_N)]$$

を考えるが，これと同時に，損失の総和の期待値に対し，次の制約を課す．

$$E[q_0(x_0, u_0) + q_1(x_1, u_1) + \cdots + q_{N-1}(x_{N-1}, u_{N-1}) + q_G(x_N)] \ \leq \ T$$

ただし，T は損失の最大許容値であり，定数として与えられるものとする．このとき，問題は次のように定式化される．

$$\text{Maximize} \quad E[r_0(x_0,u_0)+r_1(x_1,u_1)+\cdots+r_{N-1}(x_{N-1},u_{N-1})+r_G(x_N)]$$
$$\text{subject to} \begin{cases} x_{n+1}\sim p(\cdot|x_n,u_n) & n=0,1,2,\ldots,N-1 \\ \sigma=\{\sigma_0,\sigma_1,\cdots,\sigma_{N-1}\} & \\ E[q_0(x_0,u_0)+q_1(x_1,u_1)+\cdots+q_{N-1}(x_{N-1},u_{N-1})+q_G(x_N)] \le T & \end{cases}$$

これを全期間依存型制約付問題と呼ぶ．ここで $\sigma=\{\sigma_0,\sigma_1,\cdots,\sigma_{N-1}\}$ は一般政策 ([6]) をあらわす．一般政策とは，各期において，その時点までのすべての状態に依存して決定を定める一般決定関数：

$$\sigma_n \ : \ X^{n+1}\to U \qquad n=0,1,\ldots,N-1$$

の列からなる政策である．すなわち，全期間依存型制約付問題における決定は，一般政策 σ により

$$u_0=\sigma_0(x_0),\ u_1=\sigma_1(x_0,x_1),\ u_2=\sigma_2(x_0,x_1,x_2),\ldots,\ u_{N-1}=\sigma_{N-1}(x_0,x_1,\ldots,x_{N-1})$$

と定まり，最大化 (Maximize) は，加法型制約を満たすあらゆる一般政策に関してとられる．また E は，初期状態 x_0，マルコフ推移法則 p および一般政策 σ から履歴の直積空間

$$H=X\times U\times X\times U\times\cdots\times U\times X$$

上に唯一定まる確率速度 $P^{\sigma}_{x_0}$ による期待値作用素である．この意味で E には上下に添字をつけて $E^{\sigma}_{x_1}$ で表すべきだが，ここでは簡単のために E で表しておく．

2.2 再帰式

全期間依存型制約付問題において，最大許容値 T をパラメータ $t\in\mathbf{R}$ に置き換えた問題を考える．

$$\text{Maximize} \quad E[r_0(x_0,u_0)+r_1(x_1,u_1)+\cdots+r_{N-1}(x_{N-1},u_{N-1})+r_G(x_N)]$$
$$\text{subject to} \begin{cases} x_{n+1}\sim p(\cdot|x_n,u_n) & n=0,1,2,\ldots,N-1 \\ \sigma=\{\sigma_0,\sigma_1,\cdots,\sigma_{N-1}\} & \\ E[q_0(x_0,u_0)+q_1(x_1,u_1)+\cdots+q_{N-1}(x_{N-1},u_{N-1})+q_G(x_N)] \le t & \end{cases}$$

この問題は，$t=T$ とおくと，与問題と等価になる．すなわち，全期間依存型制約付問題はこの問題に埋め込まれている．この意味で，この問題を埋め込み問題と呼ぶ．

次に，埋め込み問題の部分問題群を考える．これは，開始段 n と始発状態 x_n を変化させて得られる一連の部分問題で，その最適値を最適値関数 v^n とおく．

$$v^n(x_n,t) = \max_{\substack{(\mathrm{i})_n,(\mathrm{ii})_n; \\ E[q_n(x_n,u_n)+q_{n+1}(x_{n+1},u_{n+1})+\cdots+q_G(x_N)]\le t}} E[(r_n(x_n,u_n)+\cdots+r_{N-1}(r_{N-1},u_{N-1})+r_G(x_N))]$$
$$x_n\in X,\ \ t\in\mathbf{R},\ \ n=0,1,2,\ldots,N-1$$

ここで，

$$(\mathrm{i})_n \qquad x_{m+1}\sim p(\cdot|x_m,u_m) \qquad m=n,n+1,\ldots,N-1$$
$$(\mathrm{ii})_n \qquad \sigma=\{\sigma_n,\sigma_{n+1},\cdots,\sigma_{N-1}\}$$

である．ただし，このときの一般政策 σ は，開始段が n であるので

$$\sigma_n:X\to U\ ,\quad \sigma_{n+1}:X\times X\to U\ ,\ \cdots\ ,\quad \sigma_{N-1}:X\times\cdots\times X\to U$$

からなる．また，終了期 $(n=N)$ に対する値関数は

$$v^N(x_N,t)=\begin{cases} r_G(x_N) & q_G(x_N)\le t \\ — & q_G(x_N)>t \end{cases} \qquad x_N\in X,\ \ t\in\mathbf{R}$$

である．これは，開始段 N に対する部分問題の制約は $q_G(x_N) \leq t$ であり，$q_G(x_N) > t$ のときは常に制約を満たさないので，値なしという意味で，“—”とおいた．同様に，開始段 n $(n = 0, 1, \ldots, N-1)$ に対する部分問題において，加法型制約を満たす一般政策 $\sigma = \{\sigma_n, \sigma_{n+1}, \cdots, \sigma_{N-1}\}$ が存在しない場合も

$$v^n(x_n, t) = —$$

と表す．

“—”と実数との2項演算および最大化演算については次のように定義する．$a \in \mathbf{R}$ に対して，

$$a + — = — \qquad — + — = —$$
$$\mathrm{Max}(a, —) = a \qquad \mathrm{Max}(—, —) = —$$

このとき次の再帰式が成り立つ

定理1

$$v^N(x, t) = \begin{cases} r_G(x) & q_G(x) \leq t \\ — & q_G(x) > t \end{cases} \qquad x \in X, \ t \in \mathbf{R}$$

$$v^n(x, t) = \max_{u \in U} \left[r_n(x, u) + \max_{t_1, t_2, \ldots, t_{l-1}} \left\{ \sum_{j=1}^{l-1} v^{n+1}(s_j, t_j) p(s_j|x, u) + v^{n+1}\left(s_l, \frac{t - q_n(x, u) - \sum_{j=1}^{l-1} t_j p(s_j|x, u)}{p(s_l|x, u)} \right) p(s_l|x, u) \right\} \right] \tag{1}$$

$$x \in X, \ t \in \mathbf{R}, \ n = 0, 1, 2, \ldots, N-1$$

なお，この再帰式の等号は，両辺ともに値なしの場合を含む．

最適政策の構成

定理1の再帰式(1)において，右辺の最大値 $v^n(x, t)$ を与える決定を $\pi_n^*(x, t)$ で表し，さらに，その最大値を与える t_j の値の集合を $I_{s_j}^{n+1}(x, t)$ で表す．このとき，与問題の最適政策 $\sigma^* = \{\sigma_0^*, \sigma_1^*, \ldots, \sigma_{N-1}^*\}$ は次のように構成される．

$$\sigma_0^*(x_0) = \pi_0^*(x_0, T) \tag{2}$$

$$\sigma_n^*(x_0, x_1, \cdots, x_{n-1}, s_j) = \pi_n^*(s_j, t_j), \quad t_j \in I_{s_j}^n(x_{n-1}, T_{n-1}), \quad j = 1, 2, \cdots, l-1 \tag{3}$$

$$\sigma_n^*(x_0, x_1, \cdots, x_{n-1}, s_l) = \pi_n^*\left(s_l, \frac{T_{n-1} - q_{n-1}(x_{n-1}, u_{n-1}^*) - \sum_{k=1}^{l-1} t_k p(s_k|x_{n-1}, u_{n-1}^*)}{p(s_l|x_{n-1}, u_{n-1}^*)} \right) \tag{4}$$

$$u_{n-1}^* = \sigma_{n-1}^*(x_0, \cdots, x_{n-1}), \quad t_k \in I_{s_k}^n(x_{n-1}, T_{n-1}) \ (k = 1, 2, \cdots, l-1), \quad n = 1, 2, \ldots, N-1$$

ただし $\{T_0, T_1, \ldots, T_{n-1}\}$ は $x_0, x_1, \cdots, x_{n-1}$ に依存する次の条件を満たす列とする．

$$T_0 = T \ , \ T_m \in I_{x_m}^m(x_{m-1}, T_{m-1}) \ , \ m = 1, 2, \cdots, n-1$$

2.3 再帰式の証明

簡単のため期数2 $(N = 2)$，状態数2 $(X = \{s_1, s_2\})$，決定数2 $(U = \{a_1, a_2\})$ とし，各状態への推移確率は正とする(一般の場合も同様)．このとき

$$v^1(x_1, t) = \max_{u_1 \in U} \left[r_1(x_1, u_1) + \max_{t_1} \left\{ v^2(s_1, t_1) p(s_1|x_1, u_1) + v^2\left(s_2, \frac{t - q_1(x_1, u_1) - t_1 p(s_1|x_1, u_1)}{p(s_2|x_1, u_1)} \right) p(s_2|x_1, u_1) \right\} \right] \tag{5}$$

$$v^0(x_0,\,t) \;=\; \underset{u_0\in U}{\mathrm{Max}}\Big[r_0(x_0,u_0)+\underset{t_1}{\mathrm{Max}}\Big\{v^1(s_1,\,t_1)p(s_1|x_0,u_0) \\ +v^1\left(s_2,\,\frac{t-q_0(x_0,u_0)-t_1p(s_1|x_0,u_0)}{p(s_2|x_0,u_0)}\right)p(s_2|x_0,u_0)\Big\}\Big] \quad (6)$$

を示せばよいが，式 (5) については v^2 を具体的に代入することで比較的容易に示せるので，式 (6) について考える．まず，部分問題の定義より，左辺 $v^0(x_0,\,t)$ は

$$v^0(x_0,\,t) \;=\; \underset{\sigma=\{\sigma_0,\sigma_1\};C_0(x_0,t_0)}{\mathrm{Max}}\sum_{x_1,x_2}\sum\{r_0(x_0,u_0)+r_1(x_1,u_1)+r_G(x_2)\}p(x_1|x_0,u_0)p(x_2|x_1,u_1)$$

ただし

$$C_0(x_0,\,t):\sum_{x_1,x_2}\sum\{q_0(x_0,u_0)+q_1(x_1,u_1)+q_G(x_2)\}p(x_1|x_0,u_0)p(x_2|x_1,u_1)\le t$$

となるので，示すべき式は

$$\underset{\sigma=\{\sigma_0,\sigma_1\};C_0(x_0,\,t)}{\mathrm{Max}}\sum_{x_1,x_2}\sum\{r_0(x_0,u_0)+r_1(x_1,u_1)+r_G(x_2)\}p(x_1|x_0,u_0)p(x_2|x_1,u_1)$$

$$=\; \underset{u_0\in U}{\mathrm{Max}}\Big[r_0(x_0,u_0)+\underset{t_1}{\mathrm{Max}}\Big\{v^1(s_1,\,t_1)p(s_1|x_0,u_0) \\ +v^1\left(s_2,\,\frac{t-q_0(x_0,u_0)-t_1p(s_1|x_0,u_0)}{p(s_2|x_0,u_0)}\right)p(s_2|x_0,u_0)\Big\}\Big] \quad (7)$$

である．ここで $\sigma=\{\sigma_0,\,\sigma_1\}$ に対し

$$u_0=\sigma_0(x_0)\;,\qquad \begin{cases} u_1^1=\sigma_1(x_0,\,s_1)\\ u_1^2=\sigma_1(x_0,\,s_2)\end{cases}$$

とおくと，左辺は

$$(\text{左辺}) \;=\; \underset{u_0,u_1^1,u_1^2;C_0(x_0,t)}{\mathrm{Max}}\Big[\sum_{x_2}\{r_0(x_0,u_0)+r_1(s_1,u_1^1)+r_G(x_2)\}p(s_1|x_0,u_0)p(x_2|s_1,u_1^1) \\ +\sum_{x_2}\{r_0(x_0,u_0)+r_1(s_2,u_1^2)+r_G(x_2)\}p(s_2|x_0,u_0)p(x_2|s_2,u_1^2)\Big]$$

となり，決定に課されている条件 $C_0(x_0,\,t)$ は

$$C_0(x_0,\,t) \;:\; \sum_{x_2}\{q_0(x_0,u_0)+q_1(s_1,u_1^1)+q_G(x_2)\}p(s_1|x_0,u_0)p(x_2|s_1,u_1^1) \\ +\sum_{x_2}\{q_0(x_0,u_0)+q_1(s_2,u_1^2)+q_G(x_2)\}p(s_2|x_0,u_0)p(x_2|s_2,u_1^2)\le t$$

である．一方，$C_1(x_1,\,t)$ が条件：

$$C_1(x_1,\,t) \;:\; (q_1(x_1,u_1)+q_G(s_1))p(s_1|x_1,u_1)+(q_1(x_1,u_1)+q_G(s_2))p(s_2|x_1,u_1)\le t$$

を表すとし

$$C_1^1(s_1,\,t_1) \;=\; C_1(s_1,\,t_1)\;,\quad C_1^2(s_2,\,t_1) \;=\; C_1\left(s_2,\,\frac{t-q_0(x_0,u_0)-t_1p(s_1|x_0,u_0)}{p(s_2|x_0,u_0)}\right)$$

とおくとき，右辺は $v^1(x_1,\,t)$ の定義より

$$(\text{右辺}) \;=\; \underset{u_0\in U}{\mathrm{Max}}\Big[r_0(x_0,u_0)+\underset{t_1}{\mathrm{Max}}\Big\{\underset{u_1^1;C_1^1(s_1,t_1)}{\mathrm{Max}}\sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}\Big\}p(x_2|s_1,u_1^1)\times p(s_1|x_0,u_0) \\ +\underset{u_1^2;C_1^2(s_2,t_1)}{\mathrm{Max}}\sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}p(x_2|s_2,u_1^2)\times p(s_2|x_0,u_0)\}\Big] \quad (8)$$

また，$C_1^1(s_1, t_1)$, $C_1^2(s_2, t_1)$ はそれぞれ

$$(q_1(s_1,u_1^1)+q_G(s_1))p(s_1|s_1,u_1^1)+(q_1(s_1,u_1^1)+q_G(s_2))p(s_2|s_1,u_1^1) \quad \leq \quad t_1$$

$$\begin{aligned} t_1 \quad \leq \quad & [t-q_0(x_0,u_0)-(q_1(s_2,u_1^2)+q_G(s_1))p(s_1|s_2,u_1^2)p(s_2|x_0,u_0) \\ & -(q_1(s_2,u_1^2)+q_G(s_2))p(s_2|s_2,u_1^2)p(s_2|x_0,u_0)]\frac{1}{p(s_1|x_0,u_0)} \end{aligned}$$

と書き換えられるので，ある t_1 に対し，決定の組 (u_1^1, u_1^2) が存在するためには

$$\begin{aligned} & q_0(x_0,u_0)+q_1(s_1,u_1^1)p(s_1|x_0,u_0)+q_G(s_1)p(s_1|x_0,u_0)p(s_1|s_1,u_1^1) \\ & \quad +q_G(s_2)p(s_1|x_0,u_0)p(s_2|s_1,u_1^1)+q_1(s_2,u_1^2)p(s_2|x_0,u_0)+q_G(s_1)p(s_2|x_0,u_0)p(s_1|s_2,u_1^2) \\ & \quad\quad +q_G(s_2)p(s_2|x_0,u_0)p(s_2|s_2,u_1^2) \leq t \end{aligned} \tag{9}$$

を満たさなければならず，しかもこれは，条件 $C_0(x_0, t)$ と一致する．従って，左辺において実行可能解が存在しない条件と，右辺の $\underset{u_1^1;C_1^1(s_1,t_1)}{\mathrm{Max}}$ と $\underset{u_1^2;C_1^2(s_2,t_1)}{\mathrm{Max}}$ の少なくとも一方に実行可能解が存在しない条件が一致する．よって，この場合，示すべき式 (7) は，両辺とも値なしという意味で一致する．

次に，ある t_1 に対し，条件 (9) を満たす決定の組 (u_1^1, u_1^2) が存在するときを考える．まず，左辺は

$$\begin{aligned} (\text{左辺}) \quad = \quad & \underset{u_0}{\mathrm{Max}} \underset{u_1^1,u_1^2;C_0(x_0,t)}{\mathrm{Max}} \left[r_0(x_0,u_0)+\sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}\times p(s_1|x_0,u_0)p(x_2|s_1,u_1^1) \right. \\ & \left. +\sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}\times p(s_2|x_0,u_0)p(x_2|s_2,u_1^2) \right] \\ = \quad & \underset{u_0}{\mathrm{Max}} \left[r_0(x_0,u_0)+\underset{u_1^1,u_1^2;C_0(x_0,t)}{\mathrm{Max}} \left\{ \sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}\times p(s_1|x_0,u_0)p(x_2|s_1,u_1^1) \right. \right. \\ & \left. \left. +\sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}\times p(s_2|x_0,u_0)p(x_2|s_2,u_1^2) \right\} \right] \end{aligned}$$

と変形できる．また，このとき，右辺 (8) に対し，$\hat{t}_1$ が存在し次を満たす．

$$\begin{aligned} (\text{右辺}) \quad = \quad & \underset{u_0\in U}{\mathrm{Max}} \left[r_0(x_0,u_0)+\underset{u_1^1;C_1^1(x_1,\hat{t}_1)}{\mathrm{Max}} \sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}\times p(x_2|s_1,u_1^1)\times p(s_1|x_0,u_0) \right. \\ & \left. +\underset{u_1^2;C_1^2(x_1,\hat{t}_1)}{\mathrm{Max}} \sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}\times p(x_2|s_2,u_1^2)\times p(s_2|x_0,u_0) \right] \end{aligned}$$

さらに，$C_1^1(s_1, t_1)$, $C_1^2(s_2, t_1)$ をみたす $(\hat{u}_1^1, \hat{u}_1^2)$ が存在し

$$\begin{aligned} (\text{右辺}) \quad = \quad & \underset{u_0\in U}{\mathrm{Max}} \left[r_0(x_0,u_0)+\sum_{x_2}\{r_1(s_1,\hat{u}_1^1)+r_G(x_2)\}\times p(x_2|s_1,\hat{u}_1^1)\times p(s_1|x_0,u_0) \right. \\ & \left. +\sum_{x_2}\{r_1(s_2,\hat{u}_1^2)+r_G(x_2)\}\times p(x_2|s_2,\hat{u}_1^2)\times p(s_2|x_0,u_0) \right] \end{aligned}$$

を満たす．ここで，$(\hat{u}_1^1, \hat{u}_1^2)$ は条件 (9) を満たしているので，$(u_1^1, u_1^2) = (\hat{u}_1^1, \hat{u}_1^2)$ は $C_0(x_0, t)$ を満たす．よって

$$\begin{aligned} (\text{右辺}) \quad \leq \quad & \underset{u_0\in U}{\mathrm{Max}} \left[r_0(x_0,u_0)+\underset{u_1^1,u_1^2;C_0}{\mathrm{Max}} \left\{ \sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}\times p(s_1|x_0,u_0)p(x_2|s_1,u_1^1) \right. \right. \\ & \left. \left. +\sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}\times p(s_2|x_0,u_0)p(x_2|s_2,u_1^2) \right\} \right] \quad = \quad (\text{左辺}) \end{aligned} \tag{10}$$

一方，左辺においては条件 (9) すなわち $C_0(x_0,t)$ を満たす $(\tilde{u}_1^1,\tilde{u}_1^2)$ が存在するので

$$(\text{左辺}) = \underset{u_0}{\text{Max}}\left[r_0(x_0,u_0)+\sum_{x_2}\{r_1(s_1,\tilde{u}_1^1)+r_G(x_2)\}\times p(s_1|x_0,u_0)p(x_2|s_1,\tilde{u}_1^1)\right.$$
$$\left.+\sum_{x_2}\{r_1(s_2,\tilde{u}_1^2)+r_G(x_2)\}\times p(s_2|x_0,u_0)p(x_2|s_2,\tilde{u}_1^2)\right\}\Bigg]$$

と表すことができる．また，$(\tilde{u}_1^1,\tilde{u}_1^2)$ が条件 (9) を満たしていることから，先の議論を逆にたどれば，ある $\tilde{t}_1$ が存在し，$u_1^1=\tilde{u}_1^1$，$u_1^2=\tilde{u}_1^2$ がそれぞれ $C_1^1(s_1,\tilde{t}_1)$，$C_1^2(s_2,\tilde{t}_1)$ を満たすことがわかる．従って

$$\begin{aligned}(\text{左辺}) &\le \underset{u_0\in U}{\text{Max}}\left[r_0(x_0,u_0)+\underset{u_1^1;C_1^1(s_1,\tilde{t}_1)}{\text{Max}}\sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}p(s_1|x_0,u_0)p(x_2|s_1,u_1^1)\right.\\
&\qquad\left.+\underset{u_1^2;C_1^2(s_2,\tilde{t}_1)}{\text{Max}}\sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}p(s_2|x_0,u_0)p(x_2|s_2,u_1^2)\right]\\
&\le \underset{u_0\in U}{\text{Max}}\left[r_0(x_0,u_0)+\underset{t_1}{\text{Max}}\left\{\underset{u_1^1;C_1^1(s_1,t_1)}{\text{Max}}\sum_{x_2}\{r_1(s_1,u_1^1)+r_G(x_2)\}p(s_1|x_0,u_0)p(x_2|s_1,u_1^1)\right.\right.\\
&\qquad\left.\left.+\underset{u_1^2;C_1^2(s_2,t_1)}{\text{Max}}\sum_{x_2}\{r_1(s_2,u_1^2)+r_G(x_2)\}p(s_2|x_0,u_0)p(x_2|s_2,u_1^2)\right\}\right]=(\text{右辺})\end{aligned}\qquad(11)$$

よって式 (10) と式 (11) より式 (3) が示された．

3 数値例

$N=2$，$X=\{s_1,s_2\}$，$U=\{a_1,a_2\}$，$T=1.0$ のとき，次の問題を考える．

$$\text{Maximize}\quad E[r_0(u_0)+r_1(u_1)+r_G(x_2)]$$
$$\text{subject to}\begin{cases}x_1\sim p(\cdot\,|x_0,u_0),\ x_2\sim p(\cdot\,|x_1,u_1)\\ u_0\in U,u_1\in U\\ E[q_0(u_0)+q_1(u_1)+q_G(x_2)]\le 1.0\end{cases}$$

$$\begin{aligned}r_G(s_1)=0.3,\quad r_G(s_2)=0.5,&\qquad q_G(s_1)=0.3,\quad q_G(s_2)=0.4\\ r_1(a_1)=0.7,\quad r_1(a_2)=0.8,&\qquad q_1(a_1)=0.2,\quad q_1(a_2)=0.5\\ r_0(a_1)=0.5,\quad r_0(a_2)=0.4,&\qquad q_0(a_1)=0.3,\quad q_0(a_2)=0.2\\ p(s_1|s_1,a_1)=0.4,\quad p(s_2|s_1,a_1)=0.6,&\qquad p(s_1|s_1,a_2)=0.5,\quad p(s_2|s_1,a_2)=0.5\\ p(s_1|s_2,a_1)=0.6,\quad p(s_2|s_2,a_1)=0.4,&\qquad p(s_1|s_2,a_2)=0.3,\quad p(s_2|s_2,a_2)=0.7\end{aligned}$$

まず $v^2(x_2,t)$ を求める．

$$v^2(s_1,t)=\begin{cases}r_G(s_1) & q_G(s_1)\le t\\ — & q_G(s_1)>t\end{cases}=\begin{cases}0.3 & 0.3\le t\\ — & 0.3>t\end{cases}$$
$$v^2(s_2,t)=\begin{cases}r_G(s_2) & q_G(s_2)\le t\\ — & q_G(s_2)>t\end{cases}=\begin{cases}0.5 & 0.4\le t\\ — & 0.4>t\end{cases}$$

次に $v^1(x_1,t)$ は

$$v^1(x_1,t)=\underset{u_1=a_1,a_2}{\text{Max}}\left[r_1(x_1,u_1)+\underset{t_1}{\text{Max}}\left\{v^2(s_1,t_1)p(s_1|x_1,u_1)\right.\right.$$
$$\left.\left.+v^2\left(s_2,\frac{t-q_1(x_1,u_1)-t_1p(s_1|x_1,u_1)}{p(s_2|x_1,u_1)}\right)p(s_2|x_1,u_1)\right\}\right]$$

より

$$v^1(s_1,t) = \left[0.7+\underset{t_1}{\mathrm{Max}}\left\{v^2(s_1,t_1)\times 0.4+v^2\left(s_2,\frac{t-0.2-0.4t_1}{0.6}\right)\times 0.6\right\}\right] \vee \left[0.8+\underset{t_1}{\mathrm{Max}}\left\{v^2(s_1,t_1)\times 0.5+v^2\left(s_2,\frac{t-0.5-0.5t_1}{0.5}\right)\times 0.5\right\}\right]$$

この前半部で値が存在する場合を考えると

$$0.7+\underset{t_1;\ 0.3\le t_1 \text{かつ } 0.4\le \frac{t-0.2-0.4t_1}{0.6}}{\mathrm{Max}}(0.3\times 0.4+0.5\times 0.6)$$

であり，この t_1 が存在するのは2つの条件が満たされるとき，すなわち $0.56\le t$ のとき．ゆえに前半部は

$$\begin{cases} 0.7+0.3\times 0.4+0.5\times 0.6 & 0.56\le t \\ — & \text{その他} \end{cases} = \begin{cases} 1.12 & 0.56\le t \\ — & \text{その他} \end{cases}$$

同様に後半部を求めると，$v^1(s_1,t)$ は

$$v^1(s_1,t)=\begin{Bmatrix} 1.12 & 0.56\le t \\ — & \text{その他} \end{Bmatrix}\vee\begin{Bmatrix} 1.2 & 0.85\le t \\ — & \text{その他} \end{Bmatrix} = \begin{cases} 1.2 & 0.85\le t \\ 1.12 & 0.56\le t<0.85 \end{cases}, \pi_1^*(s_1,t)=\begin{cases} a_2 & 0.85\le t \\ a_1 & 0.56\le t<0.85 \end{cases}$$

となる．(以後，値なしの記述は省略する．) 同様に $v^1(s_2,t)$ を求めると

$$v^1(s_2,t)=\begin{cases} 1.24 & 0.87\le t \\ 1.08 & 0.54\le t<0.87 \end{cases}, \pi_1^*(s_2,t)=\begin{cases} a_2 & 0.87\le t \\ a_1 & 0.54\le t<0.87 \end{cases}$$

次に $v^0(x_0,t)$ を求める．$x_0=s_1$ に対しては

$$v^0(s_1,t) = \left[0.5+\underset{t_1}{\mathrm{Max}}\left\{v^1(s_1,t_1)\times 0.4+v^1\left(s_2,\frac{t-0.3-0.4t_1}{0.6}\right)\times 0.6\right\}\right] \vee \left[0.4+\underset{t_1}{\mathrm{Max}}\left\{v^1(s_1,t_1)\times 0.5+v^1\left(s_2,\frac{t-0.2-0.5t_1}{0.5}\right)\times 0.5\right\}\right]$$

この前半部は

$$0.5+\underset{t_1}{\mathrm{Max}}\left[\begin{Bmatrix} 1.2\times 0.4 & 0.85\le t_1 \\ 1.12\times 0.4 & 0.56\le t_1<0.85 \end{Bmatrix}+\begin{Bmatrix} 1.24\times 0.6 & 0.87\le \frac{t-0.3-0.4t_1}{0.6} \\ 1.08\times 0.6 & 0.54\le \frac{t-0.3-0.4t_1}{0.6}<0.87 \end{Bmatrix}\right]$$

$$= 0.5+\underset{t_1}{\mathrm{Max}}\left[\begin{Bmatrix} 0.48 & 0.85\le t_1 \\ 0.448 & 0.56\le t_1<0.85 \end{Bmatrix}+\begin{Bmatrix} 0.744 & 0.822\le t-0.4t_1 \\ 0.648 & 0.624\le t-0.4t_1<0.822 \end{Bmatrix}\right]$$

ここで $0.4t_1=\tau$ とおいて整理すると，

$$\underset{\tau}{\mathrm{Max}}\begin{cases} 1.724 & 0.34\le\tau\le t-0.822 \\ 1.692 & 0.224\le\tau<0.34 \text{ かつ } \tau\le t-0.822 \\ 1.628 & 0.34\le\tau \text{ かつ } t-0.822\le\tau\le t-0.624 \\ 1.596 & 0.224\le\tau<0.34 \text{ かつ } t-0.822<\tau\le t-0.624 \end{cases} \tag{12}$$

最大値の対象となる各値が存在する（すなわち τ が存在する）t の範囲を考えることで，式 (12) は

$$\begin{cases} 1.724 & 1.162\le t & (0.85\le t_1\le 2.5t-2.055) \\ 1.692 & 1.046\le t<1.162 & (0.56\le t_1\le 2.5t-2.055) \\ 1.628 & 0.964\le t<1.046 & (0.85\le t_1\le 2.5t-1.56) \\ 1.596 & 0.848\le t<0.964 & (0.56\le t_1\le 2.5t-1.56 \text{ かつ } 2.5t-2.55<t_1<0.85) \end{cases}$$

となる．同様にして後半部を求め，$v^0(s_1,t)$ を求めた結果が次である．

$$v^0(s_1,t) = \begin{cases} 1.724 & 1.162 \le t \\ 1.692 & 1.046 \le t < 1.162 \\ 1.628 & 0.964 \le t < 1.046 \\ 1.596 & 0.848 < t < 0.964 \\ 1.5 & 0.75 < t \le 0.848 \end{cases} \qquad \pi_0^*(s_1,t) = \begin{cases} a_1 & 0.848 < t \\ a_2 & 0.75 < t \le 0.848 \end{cases}$$

$$I^1_{s_1}(s_1,t) = \begin{cases} [0.85, 2.5t - 2.055] & 1.162 \le t \\ [0.56, 2.5t - 2.055] & 1.046 \le t < 1.162 \\ [0.85, 2.5t - 1.56] & 0.964 \le t < 1.046 \\ [0.56, 2.5t - 1.56] & 0.848 < t < 0.964 \\ [0.56, 2t - 0.94] \cap (2t - 1.27, 0.85) & 0.75 < t \le 0.848 \end{cases}$$

$v^0(s_2,t)$ の計算の詳細について，ここでは省略するが，各状態 s_1, s_2 を初期状態とする際の最適値は，$v^0(x,t)$ において $t = T = 1.0$ とおくことにより

$$v^0(s_1,1.0) = 1.628\ , \quad v^0(s_2,1.0) = 1.668$$

と求めることができる．また，最適一般政策 $\sigma^* = \{\sigma_0^*, \sigma_1^*\}$ については，(2) において $T = 1.0$ より

$$\sigma_0^*(s_1) = \pi_0^*(s_1,T) = \pi_0^*(s_1,1.0) = a_1\ , \quad \sigma_0^*(s_2) = \pi_0^*(s_2,T) = \pi_0^*(s_2,1.0) = a_1$$

を得る．さらに (3), (4) における t_1 は，$I^1_{s_1}(s_1,1.0) = [0.85, 0.94]$ 内の任意の値をとってよいので，ここでは $t_1 = 0.9$ をとると

$$\begin{aligned} \sigma_1^*(s_1,s_1) &= \pi_1^*(s_1,t_1) = \pi_1^*(s_1,0.9) = a_2 \\ \sigma_1^*(s_1,s_2) &= \pi_1^*\left(s_2, \frac{T_0 - q_0(s_1,\sigma_0^*(s_1)) - t_1 p(s_1|s_1,\sigma_0^*(s_1))}{p(s_2|s_1,\sigma_0^*(s_1))}\right) \\ &= \pi_1^*\left(s_2, \frac{1.0 - 0.3 - 0.9 \times 0.4}{0.6}\right) = \pi_1^*(s_2,0.566) = a_1 \end{aligned}$$

同様にして

$$\sigma_1^*(s_2,s_1) = a_1\ , \quad \sigma_1^*(s_2,s_2) = a_2$$

を得る．なお，

$$\sigma_1^*(s_1,s_1) \neq \sigma_1^*(s_2,s_1)$$

であり，最適政策がマルコフではないことが確認できる．

参考文献

[1] R.E. Bellman, *Dynamic Programming*, NJ: Princeton Univ. Press, 1957.

[2] R.E. Bellman and L.A. Zadeh, Decision-making in a fuzzy environment, Management Science, **17**(1970), B141-B164.

[3] T. Fujita, On Nondeterministic Dynamic Programming, RIMS Kokyuroku 1488, Kyoto Univ. (2006), 15-24

[4] T. Fujita and K. Tsurusaki, Stochastic optimization of multiplicative functions with negative value, J. Oper. Res. Soc. Japan, **41**(1998), 351-373.

[5] S. Iwamoto and T. Fujita, Stochastic decision-making in a fuzzy environment, J. Oper. Res. Soc. Japan, **38**(1995), 467-482.

[6] S. Iwamoto, K. Tsurusaki and T. Fujita, On Markov Policies for Minimax Decision Processes, J. Math. Anal. Appl., **253**(2001), 58-78.

[7] M. Sniedovich, *Dynamic Programming*, Marcel Dekker, Inc. NY, 1992.