数理解析研究所講究録
978巻 1997年 160-172

負値乗法型関数の期待値最適化

九大・数理　藤田敏治　(Toshiharu Fujita)
九大・経済　津留崎和義　(Kazuyoshi Tsurusaki)

Abstract

確率的推移過程において加法型評価系を扱った問題はマルコフ決定過程として世界中で広く研究されている。この種の問題については、その議論において最適政策のマルコフ性が暗黙のうちに認められているようである。しかし加法型以外の評価系（乗法型、最小型、...）を考えた場合、最適政策は必ずしもマルコフ政策の中に存在するとは限らない ([1], [5], [4])。この事実は、マルコフ過程に適用されてきた通常の動的計画の手法がそのままでは通用しないことを物語っている。そこで我々は特に乗法型評価に的を絞り、最適政策がマルコフでない場合にも適用しうる解法を与える。その手法としては両決定過程と不変埋没原理の二つを用いて詳しく議論する ([2], [3], [6])。最後に、全ての場合を列挙する方法として多段確率決定ツリー法を紹介する。

1　乗法型関数の期待値最適化問題

有限段確率的推移過程における乗法型評価系の期待値最大化問題を考える。ここでは各段において得られる利得の積の期待値を最大化し、そのときの最適値、及び最適政策を求める。問題は次のように定式化される：

$$\begin{aligned} &\text{Maximize } E[\,r_0(x_0,u_0)r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)\,] \\ &\text{subject to } \quad \text{(i)} \quad x_{n+1}\sim p(\bullet\,|x_n,u_n) \\ &\qquad\qquad\quad \text{(ii)} \quad u_n\in U \quad n=0,1,\ldots,N-1 \end{aligned} \tag{1}$$

ここで、

$N\geq 2$	**終端時刻**
$X=\{s_1,s_2,\cdots,s_n\}$	**状態集合**
$U=\{a_1,a_2,\cdots,a_m\}$	**決定集合**
$x_n\in X$	時刻 $n\in\{0,1,\cdots,N\}$ における**状態**
$u_n\in U$	時刻 $n\in\{0,1,\cdots,N-1\}$ における**決定**
$r_n: X\times U\to\mathbf{R}$	時刻 n における $X\times U$ 上の**利得**
$r_G: X\to\mathbf{R}$	**終端利得**
p	**マルコフ推移法則** $p(y\|x,u)\geq 0\ {}^\forall(x,u,y)\in X\times U\times X$ $\sum_{y\in X}p(y\|x,u)=1\ {}^\forall(x,u)\in X\times U$

ただし、$y\sim p(\cdot|x,u)$ は現時刻の状態が x，決定が u であるとき，次の時刻で状態 y へ確率 $p(y|x,u)$ で推移することをあらわす．また問題 (1) における期待値は、初期状態 x_0 と

$\{u_0, u_1, \ldots, u_{N-1}\}$ を定める政策に依存した次のような N 重和である：

$$\begin{aligned} & E[\, r_0(x_0,u_0) r_1(x_1,u_1) \cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)\,] \\ = & \sum_{(x_1,x_2,\ldots,x_N)\in X\times\cdots\times X} \sum \cdots \sum \{[r_0(x_0,u_0) r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)] \\ & \qquad \times p(x_1|x_0,u_0) p(x_2|x_1,u_1)\cdots p(x_N|x_{N-1},u_{N-1})\} \end{aligned}$$

この種の問題に対しては、通常マルコフ政策全体に対して最大化が考えられる。しかし、ここではマルコフに限らない一般の政策（一般政策）とマルコフ政策とを考え、それぞれに対して最大化問題を考えることにより両者を比較する。

一般政策 $\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}$ とは、次のような決定関数列からなる政策である：

$$\sigma_0 : X \to U, \quad \sigma_1 : X \times X \to U, \quad \ldots \quad , \quad \sigma_{N-1} : X \times \cdots \times X \to U$$

この σ による決定と状態の交互列 $\{u_0, x_1, u_1, x_2, \ldots, u_{N-1}, x_N\}$ は次のように生成される：

$$\begin{array}{lll} \sigma_0(x_0) = u_0 & \to \ p(\cdot|x_0,u_0) \sim x_1 & \to \\ \sigma_1(x_0,x_1) = u_1 & \to \ p(\cdot|x_1,u_1) \sim x_2 & \to \\ \quad \vdots & \quad \vdots & \\ \sigma_{N-1}(x_0,\ldots,x_{N-1}) = u_{N-1} & \to \ p(\cdot|x_{N-1},u_{N-1}) \sim x_N & \end{array}$$

一方、**マルコフ政策** $\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$ とは、次のような決定関数列からなる政策である：

$$\pi_n : X \to U, \quad n = 0, 1, \ldots, N-1$$

この π による決定と状態の交互列 $\{u_0, x_1, u_1, x_2, \ldots, u_{N-1}, x_N\}$ は次のように生成される：

$$\begin{array}{lll} \pi_0(x_0) = u_0 & \to \ p(\,\cdot\,|x_0,u_0) \sim x_1 & \to \\ \pi_1(x_1) = u_1 & \to \ p(\,\cdot\,|x_1,u_1) \sim x_2 & \to \\ \quad \vdots & \quad \vdots & \\ \pi_{N-1}(x_{N-1}) = u_{N-1} & \to \ p(\,\cdot\,|x_{N-1},u_{N-1}) \sim x_N & \end{array}$$

最初に利得がすべて非負である場合を考える。すなわち

$$r_n(x,u) \geq 0 \qquad {}^{\forall}(x,u) \in X \times U, \quad 0 \leq^{\forall} n \leq N-1 \tag{2}$$

を仮定する。このとき

一般問題：
$$\begin{aligned} & \underset{\sigma:\text{一般政策}}{\text{Maximize}} \ E[\, r_0(x_0,u_0) r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)\,] \\ & \text{subject to} \quad \text{(i)} \ \ x_{n+1} \sim p(\bullet\,|x_n,u_n) \\ & \qquad\qquad \text{(ii)} \ \ u_n \in U \quad n = 0,1,\ldots,N-1 \end{aligned} \tag{3}$$

マルコフ問題：
$$\begin{aligned} & \underset{\pi:\text{マルコフ政策}}{\text{Maximize}} \ E[\, r_0(x_0,u_0) r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)\,] \\ & \text{subject to} \quad \text{(i)} \ \ x_{n+1} \sim p(\bullet\,|x_n,u_n) \\ & \qquad\qquad \text{(ii)} \ \ u_n \in U \quad n = 0,1,\ldots,N-1 \end{aligned} \tag{4}$$

の最適値は等しく、**最適政策はマルコフ政策の中に存在する**ことが示される。

次に、仮定 (2) を取り除いた場合を考える。このとき一般問題 (3) とマルコフ問題 (4) の最適値は一般に異なることが示される。従って、マルコフ政策のクラスのみを考えていては真の最適値・最適政策を得ることはできない。すなわち、**最適政策がマルコフ政策の中に存在するとは限らない**のである。

2 両決定過程

両決定過程 (Bidecision Process) の手法を用いて、一般問題 (1) の最適値、及び最適政策を与える。両決定過程においては、最適化問題を考える際に最大化と最小化の対を用いる。よってここでは、一般問題 (1) で n 段以降に限定した部分問題群を最大化、最小化の対で考える。まず、初期状態 x_n と一般政策 $\sigma = \{\sigma_n, \sigma_{n+1}, \ldots, \sigma_{N-1}\}$ により定まる決定列 $\{u_n, u_{n+1}, \ldots, u_{N-1}\}$ に依存した乗法型評価の期待値を I^n とおく：

$$I^n(x_n;\sigma) = \sum\sum\cdots\sum_{(x_{n+1}x_{n+2},\ldots,x_N)\in X\times\cdots\times X} \{[r_n(x_n,u_n), r_{n+1}(x_{n+1},u_{n+1})\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \times p(x_{n+1}|x_n,u_n)p(x_{n+2}|x_{n+1},u_{n+1})\cdots p(x_N|x_{N-1},u_{N-1})\}$$

この時、期待値を一般政策に関して最大化する**最大化部分問題群**、および一般政策に関して最小化する**最小化部分問題群**は次のように与えられ、その最適値をそれぞれ値関数 V^n, W^n とおく。

最大化部分問題群:

$$\begin{aligned} V^N(x_N) &= r_G(x_N) \qquad x_N \in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma=\{\sigma_n,\ldots,\sigma_{N-1}\}} I^n(x_n;\sigma) \qquad x_n \in X,\ \ 0 \le n \le N-1 \end{aligned}$$

最小化部分問題群:

$$\begin{aligned} W^N(x_N) &= r_G(x_N) \qquad x_N \in X \\ W^n(x_n) &= \operatorname*{min}_{\sigma=\{\sigma_n,\ldots,\sigma_{N-1}\}} I^n(x_n;\sigma) \qquad x_n \in X,\ \ 0 \le n \le N-1 \end{aligned}$$

ただし、一般政策 $\sigma = \{\sigma_n, \sigma_{n+1}, \ldots, \sigma_{N-1}\}$ は次のような決定関数からなる列であり：

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

決定と状態の交互列 $\{u_n, x_{n+1}, u_{n+1}, x_{n+2}, \ldots, u_{N-1}, x_N\}$ は次のように生成される：

$$\begin{array}{lll} \sigma_n(x_n) = u_n & \to \ p(\cdot|x_n,u_n) \sim x_{n+1} & \to \\ \sigma_{n+1}(x_n,x_{n+1}) = u_{n+1} & \to \ p(\cdot|x_{n+1},u_{n+1}) \sim x_{n+2} & \to \\ \quad\vdots & \quad\vdots & \\ \sigma_{N-1}(x_0,\ldots,x_{N-1}) = u_{N-1} & \to \ p(\cdot|x_{N-1},u_{N-1}) \sim x_N & \end{array}$$

定理 2.1 最大化部分問題群と最小化部分問題群の値関数 V^n, W^n に関し次の再帰式が成り立つ：

$$
\begin{aligned}
V^N(x) &= r_G(x) \qquad x \in X && (5)\\
V^n(x) &= \operatorname*{Max}_{u\in U(n,x,-)}\Big[r_n(x,u)\sum_{y\in X} W^{n+1}(y)p(y|x,u)\Big] && (6)\\
&\quad \vee \operatorname*{Max}_{u\in U(n,x,+)}\Big[r_n(x,u)\sum_{y\in X} V^{n+1}(y)p(y|x,u)\Big]
\end{aligned}
$$

$$
\begin{aligned}
W^N(x) &= r_G(x) \qquad x \in X && (7)\\
W^n(x) &= \operatorname*{min}_{u\in U(n,x,-)}\Big[r_n(x,u)\sum_{y\in X} V^{n+1}(y)p(y|x,u)\Big] && (8)\\
&\quad \wedge \operatorname*{min}_{u\in U(n,x,+)}\Big[r_n(x,u)\sum_{y\in X} W^{n+1}(y)p(y|x,u)\Big]\\
&\qquad\qquad\qquad\qquad x \in X,\ \ 0 \le n \le N-1
\end{aligned}
$$

ただし、$U(n,x,-) = \{u \in U | r_n(x,u) < 0\}$, $U(n,x,+) = \{u \in U | r_n(x,u) \ge 0\}$ である。

V^n（最大化部分問題群）を解くことにより得られる最適戦略を $\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$ W^n（最小化部分問題群）を解くことにより得られる最適戦略を $\sigma = \{\sigma_0, \sigma_0, \ldots, \sigma_{N-1}\}$ とする。このとき、もとの最大化問題 (1) の最適一般政策 $\mu = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ は π, σ から次のように構成される。

$$
\begin{aligned}
\mu_0(x_0) &:= \pi_0(x_0)\\
\mu_1(x_0,x_1) &:= \begin{cases}\sigma_1(x_0,x_1)\\ \pi_1(x_0,x_1)\end{cases} \quad \text{for } r_0(x_0,u_0)\begin{cases}\le 0\\ > 0\end{cases} \quad u_0 = \pi_0(x_0)\\
\mu_2(x_0,x_1,x_2) &:= \begin{cases}\pi_2(x_0,x_1,x_2)\\ \sigma_2(x_0,x_1,x_2)\\ \sigma_2(x_0,x_1,x_2)\\ \pi_2(x_0,x_1,x_2)\end{cases} \quad \text{for } r_1(x_1,u_1)\begin{cases}\le 0\\ \le 0\\ > 0\\ > 0\end{cases} \quad u_1 = \begin{cases}\sigma_1(x_0,x_1)\\ \pi_1(x_0,x_1)\\ \sigma_1(x_0,x_1)\\ \pi_1(x_0,x_1)\end{cases}\\
&\ \ \vdots\\
\mu_N(x_0,\ldots,x_N) &:= \begin{cases}\pi_N(x_0,\ldots,x_N)\\ \sigma_N(x_0,\ldots,x_N)\\ \sigma_N(x_0,\ldots,x_N)\\ \pi_N(x_0,\ldots,x_N)\end{cases} \quad \text{for } r_{N-1}(x_{N-1},u_{N-1})\begin{cases}\le 0\\ \le 0\\ > 0\\ > 0\end{cases}\\
&\qquad\qquad u_{n-1} = \begin{cases}\sigma_{N-1}(x_0,\ldots,x_{N-1})\\ \pi_{N-1}(x_0,\ldots,x_{N-1})\\ \sigma_{N-1}(x_0,\ldots,x_{N-1})\\ \pi_{N-1}(x_0,\ldots,x_{N-1}).\end{cases}
\end{aligned}
$$

ここまで π, σ 共に一般政策に関する最大（小）化で求めてきたが、実際は π, σ のどちらもマルコフ政策として得られることが示される。しかし、そこから得られる一般問題 (1) に対する最適政策はやはりマルコフではなく一般政策になる。

3 不変埋没原理

ここでは、不変埋没原理の手法を用いて一般問題 (1) の最適値・最適政策を求める。不変埋没原理の考えでは、まず解くべき問題をそれを含む問題群の中に埋め込み（この問題を埋め込み問題と呼ぶ)、その埋め込み問題の解法を導いて、それを解く。そして得られた最適値・最適政策をもとに元の問題の最適値・最適政策を求めるのである。簡単のため

$$\begin{aligned} -1 \leq\ & r_n(x,u) \leq 1 \quad {}^{\forall}(x,u) \in X \times U,\ 0 \leq^{\forall} n \leq N-1 \\ -1 \leq\ & r_G(x) \leq 1 \quad {}^{\forall}x \in X \end{aligned}$$

を仮定する（一般性を失うものではない)。この条件の下、乗法型関数の期待値最大化問題 (1) における被期待値関数として、利得の積とさらに $\lambda_0 \in [-1,1]$ との積を取ったものを考える：

$$\begin{aligned} &\text{Maximize } E[\,\lambda_0 r_0(x_0,u_0) r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)\,] \\ &\text{subject to } \quad \text{(i)} \ \ x_{n+1} \sim p(\bullet\,|x_n,u_n) \\ &\qquad\qquad\ \ \text{(ii)} \ \ u_n \in U \quad n=0,1,\ldots,N-1 \end{aligned} \tag{9}$$

そして、パラメーター $\lambda_n \in [-1,1]$ を状態空間に付加した拡大状態空間 $(X \times [-1,1])$ 上に埋め込んで考える。この埋め込み問題 (9) は $\lambda_0 = 1$ とおくことにより、もとの問題 (1) と等価なものになる。

最初に、問題 (9) の一般政策全体に関する最大化を考える（一般問題と呼ぶ)。ここでいう一般政策 $\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}$ は次のような決定関数からなる列になる。

$$\begin{aligned} \sigma_0 &: X \times [-1,1] \to U \\ \sigma_1 &: (X \times [-1,1]) \times (X \times [-1,1]) \to U \\ &\ \ \vdots \\ \sigma_{N-1} &: (X \times [-1,1]) \times \cdots \times (X \times [-1,1]) \to U \end{aligned}$$

この一般政策 σ により、決定と状態の交互列 $\{u_0, (x_1,\lambda_1), u_1, (x_2,\lambda_2), \ldots, u_{N-1}, (x_N,\lambda_N)\}$ は次のように生成される：

$$\begin{array}{lcll} \sigma_0(x_0) = u_0 & \to & p(\cdot|x_0,u_0) \sim x_1 & \to \\ & & \lambda_0 r_0(x_0,u_0) = \lambda_1 & \\ \sigma_1(x_0,x_1) = u_1 & \to & p(\cdot|x_1,u_1) \sim x_2 & \to \\ & & \lambda_1 r_1(x_1,u_1) = \lambda_2 & \\ \quad\vdots & & \quad\vdots & \\ \sigma_{N-1}(x_0,\ldots,x_{N-1}) = u_{N-1} & \to & p(\cdot|x_{N-1},u_{N-1}) \sim x_N & \\ & & \lambda_{N-1} r_{N-1}(x_{N-1},u_{N-1}) = \lambda_N & \end{array}$$

さて、一般問題 (9) に対する再帰式を導くために、部分問題群を定義しよう。まず x_n, λ_n が初期状態として与えられた場合の一般政策 $\sigma = \{\sigma_n, \sigma_{n+1}, \ldots, \sigma_{N-1}\}$ に対する期待値は

$$\begin{aligned} K^n(x_n,\lambda_n;\sigma) = \sum_{(x_{n+1},x_{n+2},\ldots,x_N)\in X\times X\times\cdots\times X} \sum \cdots \sum \{&[\lambda_n r_n(x_n,u_n) r_{n+1}(x_{n+1},u_{n+1})\cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)] \\ &\times p(x_{n+1}|x_n,u_n) p(x_{n+2}|x_{n+1},u_{n+1})\cdots p(x_N|x_{N-1},u_{N-1})\} \end{aligned}$$

と定義される。このとき、次のような部分問題群を考え、値関数を V^n とおく。
一般部分問題群：

$$\begin{aligned} V^N(x_N,\lambda_N) &= \lambda_N r_G(x_N) \qquad x_N \in X,\ -1 \le \lambda_N \le 1 \\ V^n(x_n,\lambda_n) &= \underset{\sigma=\{\sigma_n,\ldots,\sigma_{N-1}\}}{\mathrm{Max}} K^n(x_n,\lambda_n;\sigma) \qquad x_n \in X,\ -1 \le \lambda_n \le 1 \\ & \qquad\qquad\qquad\qquad\qquad\qquad 0 \le n \le N-1 \end{aligned}$$

この値関数 V^n 間に次の定理が成り立つ。

定理 3.1

$$\begin{aligned} V^N(x,\lambda) &= \lambda r_G(x) \qquad x \in X,\ \lambda \in [-1,1] \\ V^n(x,\lambda) &= \underset{u\in U}{\mathrm{Max}} \sum_{y\in X} V^{n+1}(y,\lambda r_n(x,u))p(y|x,u) \qquad x \in X,\ \lambda \in [-1,1] \\ & \qquad\qquad\qquad\qquad\qquad\qquad 0 \le n \le N-1 \end{aligned}$$

次に、問題 (9) のマルコフ政策全体に関する最大化を考える（マルコフ問題と呼ぶ)。ここでいうマルコフ政策 $\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$ は次のような決定関数からなる列である。

$$\pi_n : X \times [-1,1] \to U \qquad 0 \le n \le N-1$$

なおマルコフ政策により生成される決定と状態の交互列は次のようになる：

$$\begin{array}{lcl} \pi_0(x_0) = u_0 & \to & \begin{array}{l} p(\cdot|x_0,u_0) \sim x_1 \\ \lambda_0 r_0(x_0,u_0) = \lambda_1 \end{array} \to \\ \pi_1(x_1) = u_1 & \to & \begin{array}{l} p(\cdot|x_1,u_1) \sim x_2 \\ \lambda_1 r_1(x_1,u_1) = \lambda_2 \end{array} \to \\ \vdots & & \vdots \\ \pi_{N-1}(x_{N-1}) = u_{N-1} & \to & \begin{array}{l} p(\cdot|x_{N-1},u_{N-1}) \sim x_N \\ \lambda_{N-1} r_{N-1}(x_{N-1},u_{N-1}) = \lambda_N \end{array} \end{array}$$

このとき、次のような部分問題群を考え、値関数を v^n とおく。
マルコフ部分問題群：

$$\begin{aligned} v^N(x_N,\lambda_N) &= \lambda_N r_G(x_N) \qquad x_N \in X,\ -1 \le \lambda_N \le 1 \\ v^n(x_n,\lambda_n) &= \underset{\pi=\{\pi_n,\ldots,\pi_{N-1}\}}{\mathrm{Max}} K^n(x_n,\lambda_n;\pi) \qquad x_n \in X,\ -1 \le \lambda_n \le 1 \\ & \qquad\qquad\qquad\qquad\qquad\qquad 0 \le n \le N-1 \end{aligned}$$

この値関数 v^n 間に次の定理が成り立つ。

定理 3.2

$$v^N(x,\lambda) = \lambda r_G(x) \qquad x \in X,\ \lambda \in [-1,1] \tag{10}$$

$$v^n(x,\lambda) = \underset{u\in U}{\mathrm{Max}} \sum_{y\in X} v^{n+1}(y,\lambda r_n(x,u))p(y|x,u) \qquad x \in X,\ \lambda \in [-1,1] \tag{11}$$

$$0 \le n \le N-1$$

定理 3.1と定理 3.2により、V^n と v^n は共に全く同じ形の再帰式を満たすことが示された。よってこの二つはどちらを考えても同じであろうことが予想される。実際、V^n と v^n の関係として次の定理が成り立つ。

定理 3.3 (i) マルコフ政策が値関数 $V^0(\cdot)$ に到達する．すなわち最適マルコフ政策 π^* が存在して次を満たす：

$$V^0(x_0,\lambda_0) = K^0(x_0,\lambda_0\,;\pi^*) \quad \text{for all } (x_0,\lambda_0) \in X \times [0,1]$$

(ii) マルコフ部分問題群の最適値関数は一般部分問題群の最適値関数に等しい:

$$v^n(x,\lambda) = V^n(x,\lambda) \qquad (x,\lambda) \in X \times [0,1],\ \ 0 \le n \le N$$

(iii) 埋め込み問題 (9) の最適マルコフ政策 $\pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\}$ を用いて，元の問題 (1) の最適一般政策 $\sigma^* = \{\sigma_0^*, \sigma_1^*, \ldots, \sigma_{N-1}^*\}$ は次のように構成される:

$$\begin{aligned}
&\sigma_0^*(x_0) := \pi_0^*(x_0,\lambda_0),\ \ \lambda_0 := 1\\
&\sigma_1^*(x_0,x_1) := \pi_1^*(x_1,\lambda_1) \qquad \text{ただし } \lambda_1 = \lambda_0 r_0(x_0,u_0),\ \ u_0 = \pi_0^*(x_0,\lambda_0)\\
&\sigma_2^*(x_0,x_1,x_2) := \pi_2^*(x_2,\lambda_2) \qquad \text{ただし } \lambda_2 = \lambda_1 r_1(x_1,u_1),\ \ u_1 = \pi_1^*(x_1,\lambda_1),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \lambda_1 = \lambda_0 r_0(x_0,u_0),\ \ u_0 = \pi_0^*(x_0,\lambda_0)\\
&\quad\vdots\\
&\sigma_{N-1}(x_0,x_1,\ldots,x_{N-1}) := \pi_{N-1}(x_{N-1},\lambda_{N-1})\\
&\qquad\qquad \text{ただし } \lambda_{N-1} = \lambda_{N-2} r_{N-2}(x_{N-2},u_{N-2}),\ \ u_{N-2} = \pi_{N-2}(x_{N-2},\lambda_{N-2}),\\
&\qquad\qquad\qquad\ \ \lambda_{N-2} = \lambda_{N-3} r_{N-3}(x_{N-3},u_{N-3}),\ \ u_{N-3} = \pi_{N-3}(x_{N-3},\lambda_{N-3}),\\
&\qquad\qquad\qquad\quad\ \vdots\\
&\qquad\qquad\qquad\ \ \lambda_1 = \lambda_0 r_0(x_0,u_0),\ \ u_0 = \pi_0(x_0,\lambda_0).
\end{aligned}$$

4 数値例

多段確率決定過程上における乗法型評価の最適化問題の例として、次の 2 段 3 状態 2 決定問題を考える。前節までに述べた両決定過程、及び不変埋没を用いた解法と、多段確率決定ツリー法という所謂列挙法による解法の 3 通りの方法で解いてみる。そして、いずれの方法でも解が一致することと最適政策がマルコフでないことを確かめる。

【問題】

$$\begin{aligned}
&\text{Maximize} \quad E[\,r_0(u_0)r_1(u_1)r_G(x_2)\,]\\
&\text{subject to} \quad \text{(i)} \quad x_{n+1} \sim p(\,\cdot\,|x_n,u_n) \quad n = 0,1\\
&\qquad\qquad\ \ \text{(ii)} \quad u_0 \in U,\ u_1 \in U
\end{aligned} \tag{12}$$

【利得】

$$r_G(s_1) = 0.3 \qquad r_G(s_2) = 1.0 \qquad r_G(s_3) = -0.8$$

$$r_1(a_1) = -1.0 \qquad r_1(a_2) = 0.6 \qquad r_0(a_1) = -0.7 \qquad r_0(a_2) = 1.0$$

【推移確率】

$u_t = a_1$

$x_t \backslash x_{t+1}$	s_1	s_2	s_3
s_1	0.8	0.1	0.1
s_2	0.0	0.1	0.9
s_3	0.8	0.1	0.1

$u_t = a_2$

$x_t \backslash x_{t+1}$	s_1	s_2	s_3
s_1	0.1	0.9	0.0
s_2	0.8	0.1	0.1
s_3	0.1	0.0	0.9

4.1 両決定過程による解法

まず V^2, W^2 を求める。定理 2.1の (5), (7) より、$V^2(x) = W^2(x) = r_G(x)$ なので

$$V^2(s_1) = 0.3 \qquad V^2(s_2) = 1.0 \qquad V^2(s_3) = -0.8$$
$$W^2(s_1) = 0.3 \qquad W^2(s_2) = 1.0 \qquad W^2(s_3) = -0.8$$

次に V^1 を求める。定理 2.1の (6) より、

$$\begin{aligned}
V^1(s_1) &= [(-1.0) \times \{0.3 \times 0.8 + 1.0 \times 0.1 + (-0.8) \times 0.1\}] \\
&\qquad \vee [0.6 \times \{0.3 \times 0.1 + 1.0 \times 0.9 + (-0.8) \times 0.0\}] \\
&= (-0.26) \vee 0.558 \\
&= 0.558 \qquad \pi_1^*(s_1) = a_2 \\
V^1(s_2) &= [(-1.0) \times \{0.3 \times 0.0 + 1.0 \times 0.1 + (-0.8) \times 0.9\}] \\
&\qquad \vee [0.6 \times \{0.3 \times 0.8 + 1.0 \times 0.1 + (-0.8) \times 0.1\}] \\
&= 0.62 \vee 0.156 \\
&= 0.62 \qquad \pi_1^*(s_2) = a_1 \\
V^1(s_3) &= [(-1.0) \times \{0.3 \times 0.8 + 1.0 \times 0.1 + (-0.8) \times 0.1\}] \\
&\qquad \vee [0.6 \times \{0.3 \times 0.1 + 1.0 \times 0.0 + (-0.8) \times 0.9\}] \\
&= (-0.26) \vee (-0.414) \\
&= -0.26 \qquad \pi_1^*(s_3) = a_1
\end{aligned}$$

(6), (8) を用いて同様に再帰式を計算すると

$$\begin{aligned}
W^1(s_1) &= -0.26 & W^1(s_2) &= 0.156 & W^1(s_3) &= -0.414 \\
\hat{\sigma}_1(s_1) &= a_1 & \hat{\sigma}_1(s_2) &= a_2 & \hat{\sigma}_1(s_3) &= a_2
\end{aligned}$$

$$\begin{aligned}
V^0(s_1) &= 0.6138 & V^0(s_2) &= 0.4824 & V^0(s_3) &= 0.16366 \\
\pi_0^*(s_1) &= a_2 & \pi_0^*(s_2) &= a_2 & \pi_0^*(s_3) &= a_1
\end{aligned}$$

$$\begin{aligned}
W^0(s_1) &= -0.33768 & W^0(s_2) &= -0.2338 & W^0(s_3) &= -0.3986 \\
\hat{\sigma}_0(s_1) &= a_1 & \hat{\sigma}_0(s_2) &= a_2 & \hat{\sigma}_0(s_3) &= a_2
\end{aligned}$$

よって、問題 (12) の最適値は

$$\begin{pmatrix} V_0(s_1) \\ V_0(s_2) \\ V_0(s_3) \end{pmatrix} = \begin{pmatrix} 0.6138 \\ 0.4824 \\ 0.16366 \end{pmatrix}$$

最後に最適政策を構成する。$\mu_0^*(x_0) := \pi_0^*(x_0)$ より、$\mu_0^*(x_0)$ は

$$\mu_0^*(s_1) = a_2, \quad \mu_0^*(s_2) = a_2, \quad \mu_0^*(s_3) = a_1$$

また

$$\mu_1^*(x_0, x_1) := \begin{cases} \hat{\sigma}_1(x_1) \\ \pi_1^*(x_1) \end{cases} \quad \text{for} \quad r_0(u_0) \begin{cases} < 0 \\ \geq 0 \end{cases} \qquad \text{ただし } u_0 = \pi_0^*(x_0)$$

であったので $\mu_1^*(x_0, x_1)$ は次のように構成される：

- $r_0(\pi_0^*(s_1)) = r_0(a_2) = 1.0 > 0$ より

$$\mu_1^*(s_1, s_1) = \pi_1^*(s_1) = a_2, \qquad \mu_1^*(s_1, s_2) = \pi_1^*(s_2) = a_1, \qquad \mu_1^*(s_1, s_3) = \pi_1^*(s_3) = a_1$$

- $r_0(\pi_0^*(s_2)) = r_0(a_2) = 1.0 > 0$ より

$$\mu_1^*(s_2, s_1) = \pi_1^*(s_1) = a_2, \qquad \mu_1^*(s_2, s_2) = \pi_1^*(s_2) = a_1, \qquad \mu_1^*(s_2, s_3) = \pi_1^*(s_3) = a_1$$

- $r_0(\pi_0^*(s_3)) = r_0(a_1) = -0.7 < 0$ より

$$\mu_1^*(s_3, s_1) = \hat{\sigma}_1(s_1) = a_1, \qquad \mu_1^*(s_3, s_2) = \hat{\sigma}_1(s_2) = a_2, \qquad \mu_1^*(s_3, s_3) = \hat{\sigma}_1(s_3) = a_2$$

4.2 不変埋没原理による解法

問題 (12) を (9) の形に埋め込んで考える。そして、定理 3.1の再帰式 (10)(11) を用いて計算する。まず、$v^2(x_2; \lambda_2) = \lambda_2 \times r_G(x_2)$ より

$$v^2(s_1; \lambda_2) = \lambda_2 \times 0.3, \qquad v^2(s_2; \lambda_2) = \lambda_2 \times 1.0, \qquad v^2(s_3; \lambda_2) = \lambda_2 \times (-0.8)$$

次に $v^1(x_1; \lambda_1) = \underset{u_1 \in U}{\mathrm{Max}} \sum_{x_2 \in X} v^2(x_2;\ \lambda_1 \times r_1(x_1, u_1)) p(x_2 | x_1, u_1)$ より

$$\begin{aligned}
& v^1(s_1; \lambda_1) \\
= & \sum_{x_2 \in X} v^2(x_2; \lambda_1 \times r_1(a_1)) p(x_2 \mid s_1, a_1) \vee \sum_{x_2 \in X} v^2(x_2; \lambda_1 \times r_1(a_2)) p(x_2 \mid s_1, a_2) \\
= & [\lambda_1 \times (-1.0) \times 0.3 \times 0.8 + \lambda_1 \times (-1.0) \times 1.0 \times 0.1 + \lambda_1 \times (-1.0) \times (-0.8) \times 0.1] \\
& \qquad \vee [\lambda_1 \times 0.6 \times 0.3 \times 0.1 + \lambda_1 \times 0.6 \times 1.0 \times 0.9 + \lambda_1 \times 0.6 \times (-0.8) \times 0.0] \\
= & [\lambda_1 \times (-0.26)] \vee [\lambda_1 \times 0.558] \\
= & \begin{cases} \lambda_1 \times (-0.26) & \text{for} \quad -1 \leq \lambda_1 \leq 0 \\ \lambda_1 \times 0.558 & \text{for} \quad 0 \leq \lambda_1 \leq 1 \end{cases} \qquad \pi_1^*(s_1; \lambda_1) = \begin{cases} a_1 & \text{for} \quad -1 \leq \lambda_1 \leq 0 \\ a_2 & \text{for} \quad 0 \leq \lambda_1 \leq 1 \end{cases}
\end{aligned}$$

同様に $v^1(s_2; \lambda_1), v^1(s_3; \lambda_1)$ を計算し、次の結果を得る。

	$-1 \leq \lambda_1 \leq 0$	$0 \leq \lambda_1 \leq 1$
$v^1(s_1; \lambda_1),\ \pi_1^*(s_1; \lambda_1)$	$\lambda_1 \times (-0.26),\ a_1$	$\lambda_1 \times 0.558,\ a_2$
$v^1(s_2; \lambda_1),\ \pi_1^*(s_2; \lambda_1)$	$\lambda_1 \times 0.156,\ a_2$	$\lambda_1 \times 0.62,\ a_1$
$v^1(s_3; \lambda_1),\ \pi_1^*(s_3; \lambda_1)$	$\lambda_1 \times (-0.414),\ a_2$	$\lambda_1 \times (-0.26),\ a_1$

さらに $v^0(x_0;\lambda_0)$ を計算すると次を得る。

	$-1 \le \lambda_0 \le 0$	$0 \le \lambda_0 \le 1$
$v^0(s_1;\lambda_0),\ \pi_0^*(s_1;\lambda_0)$	$\lambda_0 \times (-0.33768),\quad a_1$	$\lambda_0 \times 0.6138,\quad a_2$
$v^0(s_2;\lambda_0),\ \pi_0^*(s_2;\lambda_0)$	$\lambda_0 \times (-0.2338),\quad a_2$	$\lambda_0 \times 0.4824,\quad a_2$
$v^0(s_3;\lambda_0),\ \pi_0^*(s_3;\lambda_0)$	$\lambda_0 \times (-0.3986),\quad a_2$	$\lambda_0 \times 0.16366,\quad a_1$

これより $\lambda_0 = 1$ とおいて、元の問題 (12) の最適値を得る：

$$\begin{pmatrix} v_0(s_1;1) \\ v_0(s_2;1) \\ v_0(s_3;1) \end{pmatrix} = \begin{pmatrix} 0.6138 \\ 0.4824 \\ 0.16366 \end{pmatrix}$$

最後に、ここで得られた埋め込み問題の最適マルコフ政策 $\pi^* = \{\pi_0^*, \pi_1^*\}$ を用いて、元の問題 (12) に対する最適一般政策 $\tilde{\gamma} = \{\tilde{\gamma}_0, \tilde{\gamma}_1\}$ を構成する。最初の決定は $\pi_0^*(x_0, \lambda_0)$ で $\lambda_0 = 1$ とおいて

$$\tilde{\gamma}_0(s_1) = \pi_0^*(s_1, 1) = a_2, \qquad \tilde{\gamma}_0(s_2) = \pi_0^*(s_2, 1) = a_2, \qquad \tilde{\gamma}_0(s_3) = \pi_0^*(s_3, 1) = a_1$$

次の決定は、

$$\tilde{\gamma}_1(x_0, x_1) \;=\; \pi_1^*(x_1, \lambda_1) = \pi_1^*(x_1, \lambda_0 \times r_0(u_0)) \qquad u_0 = \pi_0^*(x_0, \lambda_0),\ \ \lambda_0 = 1.0$$

より

$$\begin{aligned} \tilde{\gamma}_1(s_1, x_1) &= \pi_1^*(x_1, r_0(a_2)) = \pi_1^*(x_1, 1.0) \\ \tilde{\gamma}_1(s_2, x_1) &= \pi_1^*(x_1, r_0(a_2)) = \pi_1^*(x_1, 1.0) \\ \tilde{\gamma}_1(s_3, x_1) &= \pi_1^*(x_1, r_0(a_1)) = \pi_1^*(x_1, -0.7) \end{aligned}$$

であるので

$$\begin{aligned} &\tilde{\gamma}_1(s_1, s_1) = a_2, \quad \tilde{\gamma}_1(s_2, s_1) = a_2, \quad \tilde{\gamma}_1(s_3, s_1) = a_1 \\ &\tilde{\gamma}_1(s_1, s_2) = a_1, \quad \tilde{\gamma}_1(s_2, s_2) = a_1, \quad \tilde{\gamma}_1(s_3, s_2) = a_2 \\ &\tilde{\gamma}_1(s_1, s_3) = a_1, \quad \tilde{\gamma}_1(s_2, s_3) = a_1, \quad \tilde{\gamma}_1(s_3, s_3) = a_2 \end{aligned}$$

こうして得られた最適値及び最適政策は、両決定過程による解法で得られたものと一致していることが見て取れる。さらに $\tilde{\gamma}_1 (= \mu_1^*)$ をみると、最適政策は x_1 だけでなく x_0 にも依存していることがわかる。

4.3 確率決定ツリー法による解法

図 1 では、次の簡略記号を用いている：

$$\begin{aligned} &\text{履歴} = x_0 \;\; r_0(u_0) \,/\, u_0 \;\; p(x_1 \mid x_0, u_0) \;\; x_1 \;\; r_1(u_1) \,/\, u_1 \;\; p(x_2 \mid x_1, u_1) \;\; x_2 \\ &\text{終端} = \text{終端評価値} = r_G(x_2) \end{aligned}$$

経路 = 経路確率 = $p(x_1 \mid x_0, u_0)p(x_2 \mid x_1, u_1)$

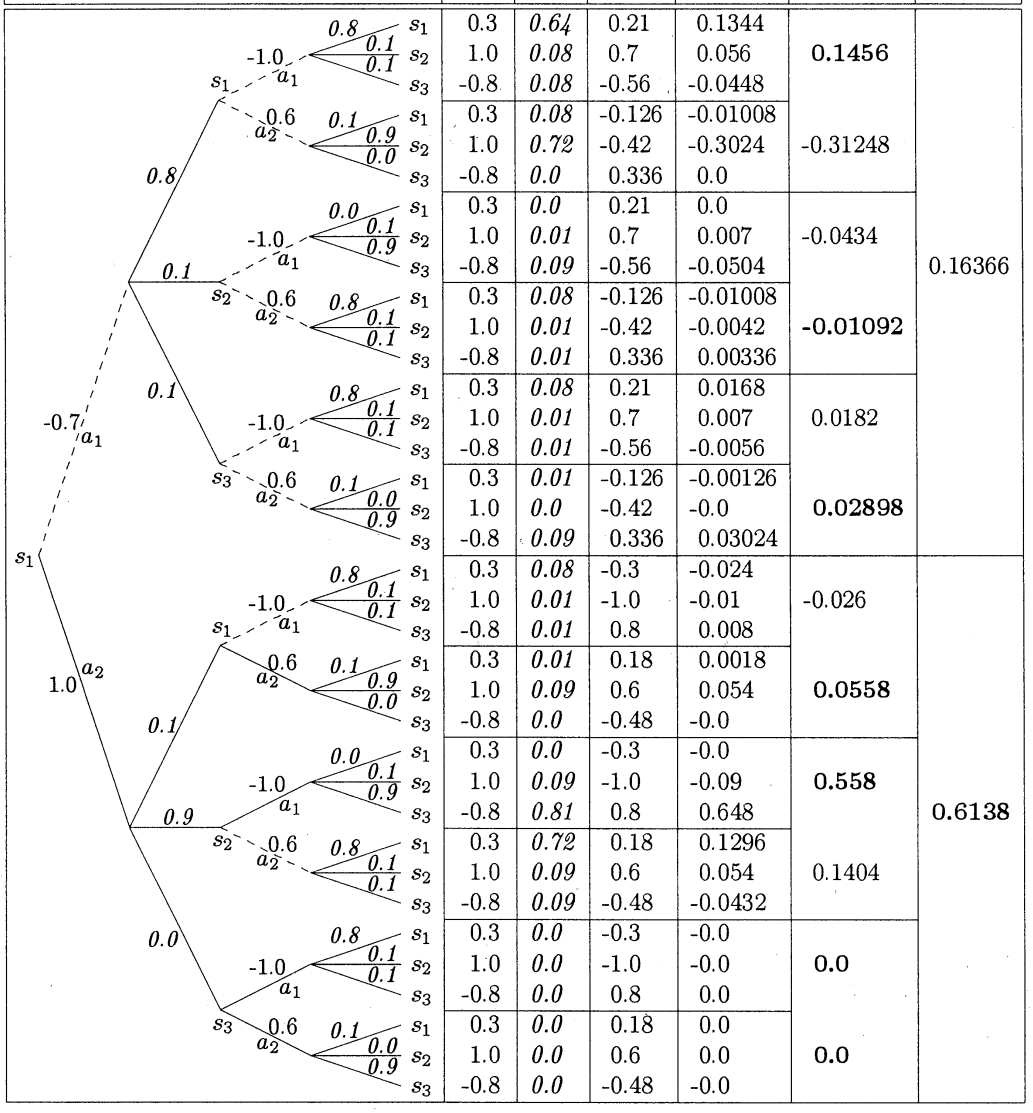
評価 = 評価値の積 = $r_0(u_0)r_1(u_1)r_G(x_2)$

積 = 経路 × 評価

部分期 = 部分期待値

全期待 = 全期待値

さらにイタリック体は確率を、ボールド体は上下の期待値の最大(大きい方)の数値を表す。



履歴	終端	経路	評価	積	部分期	全期待
s_1 a_1 s_1	0.3	*0.64*	0.21	0.1344		
	1.0	*0.08*	0.7	0.056	**0.1456**	
	-0.8	*0.08*	-0.56	-0.0448		
s_1 a_2 s_1	0.3	*0.08*	-0.126	-0.01008		
	1.0	*0.72*	-0.42	-0.3024	-0.31248	
	-0.8	*0.0*	0.336	0.0		
s_1 a_1 s_2	0.3	*0.0*	0.21	0.0		
	1.0	*0.01*	0.7	0.007	-0.0434	
	-0.8	*0.09*	-0.56	-0.0504		0.16366
s_2 a_2 s_1	0.3	*0.08*	-0.126	-0.01008		
	1.0	*0.01*	-0.42	-0.0042	**-0.01092**	
	-0.8	*0.01*	0.336	0.00336		
s_3 a_1 s_1	0.3	*0.08*	0.21	0.0168		
	1.0	*0.01*	0.7	0.007	0.0182	
	-0.8	*0.01*	-0.56	-0.0056		
s_3 a_2 s_1	0.3	*0.01*	-0.126	-0.00126		
	1.0	*0.0*	-0.42	-0.0	**0.02898**	
	-0.8	*0.09*	0.336	0.03024		
s_1 a_1 s_1	0.3	*0.08*	-0.3	-0.024		
	1.0	*0.01*	-1.0	-0.01	-0.026	
	-0.8	*0.01*	0.8	0.008		
s_1 a_2 s_1	0.3	*0.01*	0.18	0.0018		
	1.0	*0.09*	0.6	0.054	**0.0558**	
	-0.8	*0.0*	-0.48	-0.0		
s_1 a_1 s_2	0.3	*0.0*	-0.3	-0.0		
	1.0	*0.09*	-1.0	-0.09	**0.558**	
	-0.8	*0.81*	0.8	0.648		**0.6138**
s_2 a_2 s_1	0.3	*0.72*	0.18	0.1296		
	1.0	*0.09*	0.6	0.054	0.1404	
	-0.8	*0.09*	-0.48	-0.0432		
s_3 a_1 s_1	0.3	*0.0*	-0.3	-0.0		
	1.0	*0.0*	-1.0	-0.0	**0.0**	
	-0.8	*0.0*	0.8	0.0		
s_3 a_2 s_1	0.3	*0.0*	0.18	0.0		
	1.0	*0.0*	0.6	0.0	**0.0**	
	-0.8	*0.0*	-0.48	-0.0		

図 1 : 状態 s_1 からの 2 段確率決定ツリー

初期状態が s_2, s_3 の時も同様な図により問題 (12) に対する最適値と最適政策を求めることができる。

4.4 マルコフ政策による期待値一覧

次の表は問題 (12) に対し全てのマルコフ政策 $\pi = \{\pi_0, \pi_1\}$ による期待値ベクトル：

$$\begin{pmatrix} I^0(s_1;\pi) \\ I^0(s_2;\pi) \\ I^0(s_2;\pi) \end{pmatrix}$$

を一覧表にしたものである。この表からどのマルコフ政策も一般最適政策により得られる期待値に達していないことがわかる。

π_0 \ π_1	$\begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_1 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} a_1 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_1 \\ a_2 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_2 \\ a_2 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ 0.1204 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ 0.21742 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ 0.15288 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ 0.2499 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.1204 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.21742 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.15288 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.2499 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ 0.1204 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ 0.21742 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ 0.15288 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ 0.2499 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.1204 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.21742 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.15288 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.2499 \\ -0.3168 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ -0.172 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ -0.1874 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ -0.2184 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ -0.2338 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.4824 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.467 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.436 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.4206 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_2 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ -0.172 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ -0.1874 \\ -0.1986 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ -0.2184 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ -0.2338 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.4824 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.467 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.436 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.4206 \\ -0.3168 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.1204 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.21742 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.15288 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.2499 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.1204 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.21742 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.15288 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.2499 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.1204 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.21742 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.15288 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.2499 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.1204 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.21742 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.15288 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.2499 \\ -0.3168 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.172 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.1874 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2184 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2338 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.4824 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.467 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.436 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.4206 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_2 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.172 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.1874 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2184 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2338 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.4824 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.467 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.436 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.4206 \\ -0.3168 \end{pmatrix}$

参考文献

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

[2] S. Iwamoto, From dynamic programming to bynamic programming, *J. Math. Anal. Appl.* **177**(1993), 56-74.

[3] S. Iwamoto, On bidecision processes, *J. Math. Anal. Appl.* **187**(1994), 676-699.

[4] S. Iwamoto, Associative dynamic programs, *J. Math. Anal. Appl.*, **201**(1996), 195-211.

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

[6] S. Iwamoto, K. Tsurusaki and T. Fujita, On Markov Policies for Minmax Decision Process, submitted.