数理解析研究所講究録
第1682巻 2010年 62-69

制御マルコフ連鎖における成長確率最大化について

九州大学大学院数理学府　吉良 知文 (Akifumi Kira)
Graduate School of Mathematics, Kyushu University
長崎県立大学経済学部　植野 貴之 (Takayuki Ueno)
Faculty of Economics, University of Nagasaki
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

用いる記号と用語

以下, 本稿で用いる記号と用語について述べておく.

1. $N\,(\geq 2)$ は**段 (期) の総数**.

2. $X=\{s_1, s_2, \ldots, s_m\}$ は有限**状態空間**. 時刻 n に確率的に生じる状態を $X_n\ (\in X)$ で表し, 実現した状態を x_n で表す $(n=0, 1, \ldots, N)$. ただし, 初期状態はあらかじめ確定的に与えられているものとする.

3. $U=\{a_1, a_2, \ldots, a_k\}$ は有限**決定空間**. $u_n\,(\in U)$ は時刻 n で選択する決定を表す $(n=0, 1, \ldots, N-1)$. $U_n : X \to 2^U \setminus \phi$ は点対集合値で, $U_n(x)$ を**可能決定空間**とよび, 時刻 n での状態が x であるときに実行可能な決定全体を表す.
$G_r(U_n)$ を $U_n(\cdot)$ のグラフとする :

$$G_r(U_n)=\{(x,u)\,|\,u\in U_n(x),\ x\in X\}$$

4. $r_n : G_r(U_n)\to\mathbb{R}$ $(n=0, 1, \ldots, N-1)$ は第 n **利得関数**. 時刻 n に状態 x_n において決定 $u_n\,(\in U_n(x_n))$ を選ぶと利得 $r_n(x_n, u_n)$ を得る.

   $r_G : X\to\mathbb{R}$ は**終端利得関数**. 最終時刻 N では状態 x_N で利得 $r_G(x_N)$ を得る.

5. $p=\{p_n(\cdot\,|x,u)\}$ は非定常**マルコフ推移法則**. $p_{n+1}(y\,|x_n, u_n)$ は時刻 n での状態 x_n において決定 u_n を選んだときに, 次状態 X_{n+1} が $y\ (\in X)$ になる条件付き確率である. この確率的推移を $X_{n+1}\sim p_{n+1}(\cdot\,|x_n, u_n)$ と表現する.

6. $\Sigma^{(0,N)}$ は時刻 0 に始まる N 期間の**一般政策全体**.

$$\Sigma^{(0,N)}:=\left\{\sigma=(\sigma_0, \ldots, \sigma_{N-1})\ \middle|\ \begin{array}{l}\sigma_n : X^{n+1}\to U,\\ \sigma_n(x_0, \ldots, x_n)\in U_n(x_n),\ \forall(x_0, \ldots, x_n)\in X^{n+1},\\ n=0, 1, \ldots, N-1\end{array}\right\}.$$

一般政策 $\sigma\in\Sigma^{(0,N)}$ を採用した意思決定者は時刻 n において, 過去の状態列 $(x_0, x_1, \ldots, x_{n-1})$ と現状態 x_n に依存した決定 $u_n=\sigma_n(x_0, x_1, \ldots, x_n)$ を選択する.

7. $\Pi^{(0,N)}$ ($\subset \Sigma^{(0,N)}$) は時刻 0 に始まる N 期間の**マルコフ政策全体**.

$$\Pi^{(0,N)} := \left\{ \pi = (\pi_0, \pi_1, \ldots, \pi_{N-1}) \;\middle|\; \begin{array}{l} \pi_n : X \to U, \\ \pi_n(x_n) \in U_n(x_n),\ \forall x_n \in X, \\ n = 0,\ 1,\ \ldots,\ N-1 \end{array} \right\}.$$

マルコフ政策 $\pi \in \Pi^{(0,N)}$ を採用した意思決定者は時刻 n において, 過去の状態列とは無関係に, 現状態 x_n のみに依存した決定 $u_n = \pi_n(x_n)$ を選択する.

1 はじめに

通常, 制御マルコフ連鎖 (= マルコフ決定過程) においては各段 (期) で得られる利得の総和の期待値最大化 (加法型評価) を扱うが, 動的計画法の伝統的手法である新たなパラメータの埋め込みによる状態空間の拡大を通して, より複雑な問題に対して再帰式を導くことができる (不変埋没原理 [2, 7]). 例えば, 各段で得られる利得の中で最小のものの期待値を最大にする最小型評価が挙げられる [8]. このような期待値基準の評価系については多くの研究がなされ, 範囲型, 分散型など多様な評価系が考えられ再帰式が導かれている.

一方、制御マルコフ連鎖において, 期待値だけではなく確率を最大化する**確率基準**の問題を考えるのは自然である. 現実に人間は期待値を最大化するよりも, 望ましい状況 (あるいは最低限の要求) が満たされるように行動することが多々ある. このような意思決定の原理は satisfying approach と呼ばれている (Simon [14]). 例えば, ポートフォリオの運用者はしばしば期待値よりも確率に興味がある. このような観点から, 各段で得られる利得の総和がある一定の水準以上になる確率を最大化する閾値確率問題が多くの研究者によって研究がなされてきた [4, 10, 12, 13, 15, 17]. しかしながら, 確率基準における非加法型評価としては, わずかに [16] が挙げられるのみである.

本稿では, 確率基準の新たな非加法型評価として, 各段で得られる利得が段の進行とともに単調に増加するという確率–**成長確率**を導入する. すなわち, 初期状態 x があらかじめ (確定的に) 与えられているときに成長確率を最大化する問題は以下で与えられる.

$$\mathrm{P}_0(x) \quad \begin{array}{ll} \text{Maximize} & \mathrm{P}^{\sigma}\big(r_0(X_0, U_0) \le \cdots \le r_{N-1}(X_{N-1}, U_{N-1}) \le r_G(X_N) \,\big|\, X_0 = x \big) \\ \text{subject to} & \text{(i) } X_{n+1} \sim p_{n+1}(\,\cdot\,|x_n, u_n), \quad n = 0,\ 1,\ \ldots,\ N-1 \\ & \text{(ii) } \sigma \in \Sigma^{(0,N)} \end{array}$$

原問題 $\mathrm{P}_0(x)$ に対して部分問題の族 $\{\mathrm{P}_n(x)\}$ を定義する. 時刻 n での状態があらかじめ与えられているときに残りの $(N-n)$ 期間における成長確率を最大化する問題である.

$$\mathrm{P}_n(x) \quad \begin{array}{ll} \text{Maximize} & \mathrm{P}^{\sigma}\big(r_n(X_n, U_n) \le \cdots \le r_{N-1}(X_{N-1}, U_{N-1}) \le r_G(X_N) \,\big|\, X_n = x \big) \\ \text{subject to} & \text{(i)}' \; X_{t+1} \sim p_{t+1}(\,\cdot\,|x_t, u_t), \quad t = n,\ n+1,\ \ldots,\ N-1 \\ & \text{(ii)}' \; \sigma \in \Sigma^{(n,N-n)} \end{array}$$

ただし, $\Sigma^{(n,N-n)}$ は時刻 n に始まる $(N-n)$ 期間の**一般政策全体**である.

$$\Sigma^{(n,N-n)} := \left\{ \sigma = (\sigma_n, \ldots, \sigma_{N-1}) \;\middle|\; \begin{array}{l} \sigma_t : X^{t+1-n} \to U, \\ \sigma_t(x_n, \ldots, x_t) \in U_t(x_t),\ \forall (x_n, \ldots, x_t) \in X^{t+1-n}, \\ t = n,\ n+1,\ \ldots,\ N-1 \end{array} \right\}.$$

ところが, これらの部分問題間に直接的に**再帰式を導くことはできない**. これは原問題の最大値を与える最適政策が一般にマルコフ政策ではないことを意味する. そこで, 新たなパラメータの埋め込みと定義関数の導入によって, 成長確率最大化問題を既知の期待値基準 (非負値乗法型) の問題に帰着させる. この非負値乗法型評価系の概要を第 2 節で述べる. また, 確率は定義関数の期待値であることから, 定義関数の導入により, 直ちに期待値基準の問題に帰着されるクラスを第 3 節で定義する. これらの結果を用いて第 4 節で成長型評価系の再帰式を導く.

2 非負値乗法型評価系

乗法型評価系においてはシステム全体の評価値として, 各段 (期) で得られる利得の積を考える. すなわち, 評価値の期待値最大化問題は次で与えられる.

$$\mathrm{P}_0(x)\quad \begin{array}{ll} \text{Maximize} & \mathrm{E}^{\sigma}\left[r_0(X_0,U_0)\times\cdots\times r_{N-1}(X_{N-1},U_{N-1})\times r_G(X_N)\,|\,X_0=x\right] \\ \text{subject to} & \text{(i) } X_{n+1}\sim p_{n+1}(\cdot\,|\,x_n,u_n),\quad n=0,\,1,\,\ldots,N-1 \\ & \text{(ii) } \sigma\in\Sigma^{(0,N)} \end{array}$$

ただし, この節では利得がすべて非負である場合を考える.

$$\forall(x,u)\in G_r(U_n),\ r_n(x,u)\geq 0,\quad n=0,\,1,\,\ldots,N-1$$
$$\forall x\in X,\ r_G(x)\geq 0$$

利得が負の値をとり得るときは別の再帰式を考える必要がある [5, 7, 9].
ここで, 原問題 $\mathrm{P}_0(x)$ に対して, 時刻 n に始まる $(N-n)$ 期間の部分問題の族を定義する.

$$\mathrm{P}_n(x)\quad \begin{array}{ll} \text{Maximize} & \mathrm{E}^{\sigma}\left[r_n(X_n,U_n)\times\cdots\times r_{N-1}(X_{N-1},U_{N-1})\times r_G(X_N)\,|\,X_n=x\right] \\ \text{subject to} & \text{(i)}' \ X_{t+1}\sim p_{t+1}(\cdot\,|\,x_t,u_t),\quad t=n,\,n+1,\,\ldots,N-1 \\ & \text{(ii)}' \ \sigma\in\Sigma^{(n,N-n)} \end{array}$$

定理 2.1. $V_n : X\to\mathbb{R}$ $(n=0,\,1,\,\ldots,N-1)$ をそれぞれ部分問題の最適値関数とする:

$$V_n(x):=\operatorname*{Max}_{\text{(i)}',\,\text{(ii)}'}\mathrm{E}^{\sigma}\left[r_n(X_n,U_n)\times\cdots\times r_{N-1}(X_{N-1},U_{N-1})\times r_G(X_N)\,\middle|\,X_n=x\right]$$

また, $V_N : X\to\mathbb{R}$ を

$$V_N(x):=\mathrm{E}^{\sigma}\left[r_G(X_N)\,\middle|\,X_N=x\right]$$

とする. このとき以下の**最適再帰式 (Bellman 方程式)** が成り立つ.

$$V_N(x)=r_G(x)$$
$$V_n(x)=\operatorname*{Max}_{u\in U_n(x)}r_n(x,u)\sum_{y\in X}V_{n+1}(y)p_{n+1}(y|x,u),\quad n=0,\,1,\,\ldots,N-1$$

定理 2.2. 再帰式を解くことにより得られる**マルコフ政策** $\pi^*=(\pi_0^*,\,\pi_1^*,\,\ldots,\pi_{N-1}^*)$:

$$\pi_n^*(x)\in\operatorname*{arg\,max}_{u\in U_n(x)}r_n(x,u)\sum_{y\in X}V_{n+1}(y)p_{n+1}(y|x,u),\quad n=0,\,1,\,\ldots,N-1$$

は原問題 $\mathrm{P}_0(x)$ に対する一般政策全体の中での**最適政策**である.

3 期待値基準に帰着される確率最適化問題

この節では定義関数の導入により直ちに期待値基準 (非負値乗法型) に帰着される確率基準の問題のクラスを定義する. $I_n \subset \mathbb{R}$ $(n = 0, 1, \ldots N)$ を所与の区間とし, 各段 (期) に得られる利得がそれぞれ区間 I_n に収まる閾値確率を最大化する問題を考える.

$$\mathrm{P}_0(x) \quad \begin{array}{ll} \text{Maximize} & \mathrm{P}^\sigma\big(\, r_0(X_0,U_0)\in I_0,\ r_1(X_1,U_1)\in I_1,\ \ldots,\ r_G(X_N)\in I_N \,\big|\, X_0 = x \,\big) \\ \text{subject to} & \text{(i) } X_{n+1} \sim p_{n+1}(\,\cdot\,|x_n,u_n), \quad n = 0,\ 1,\ \ldots, N-1 \\ & \text{(ii) } \sigma \in \Sigma^{(0,N)} \end{array}$$

特に, $I_0 = I_1 = \cdots = I_N = [c, \infty)$ とすると**最小型閾値確率最大化問題**[16] となる :

$$\text{Maximize} \quad \mathrm{P}^\sigma\big(\, r_0(X_0,U_0) \wedge r_1(X_1,U_1) \wedge \cdots \wedge r_G(X_N) \geq c \,\big|\, X_0 = x \,\big)$$

原問題 $\mathrm{P}_0(x)$ に対して, 時刻 n に始まる $(N-n)$ 期間の部分問題の族を定義する.

$$\mathrm{P}_n(x) \quad \begin{array}{ll} \text{Maximize} & \mathrm{P}^\sigma\big(\, r_n(X_n,U_n)\in I_n,\ \ldots,\ r_G(X_N)\in I_N \,\big|\, X_n = x \,\big) \\ \text{subject to} & \text{(i)}' \ X_{t+1} \sim p_{t+1}(\,\cdot\,|x_t,u_t), \quad t = n,\ n+1,\ \ldots, N-1 \\ & \text{(ii)}' \ \sigma \in \Sigma^{(n,N-n)} \end{array}$$

このとき, 原問題 $\mathrm{P}_0(x)$ の目的関数は次のように表現できる.

$$\mathrm{E}^\sigma\big[\, \mathbf{1}_{I_0}(\, r_0(X_0,U_0)\,) \times \mathbf{1}_{I_1}(\, r_1(X_1,U_1)\,) \times \cdots \times \mathbf{1}_{I_N}(\, r_G(X_N)\,) \,\big|\, X_0 = x \,\big]$$

ゆえに, $\bar{r}_n(x_n,u_n) := \mathbf{1}_{I_n}(\, r_n(x_n,u_n)\,)$ を利得関数とみなせば定理 2.1・2.2 より次を得る.

定理 3.1. $V_n : X \to \mathbb{R}$ $(n = 0,\ 1,\ \ldots, N-1)$ をそれぞれ部分問題の最適値関数とする:

$$V_n(x) := \operatorname*{Max}_{\text{(i)}',\,\text{(ii)}'} \mathrm{P}^\sigma\,\big(\, r_n(X_n,U_n)\in I_n,\ \ldots,\ r_G(X_N)\in I_N \,\big|\, X_n = x \,\big)$$

また、$V_N : X \to \mathbb{R}$ を

$$V_N(x) := \mathrm{P}^\sigma\,\big(\, r_G(X_N)\in I_N \,\big|\, X_N = x \,\big)$$

とする. このとき以下の**再帰式 (Bellman 方程式)** が成り立つ.

$$V_N(x) = \mathbf{1}_{I_N}(\, r_G(x)\,)$$

$$V_n(x) = \operatorname*{Max}_{u\in U_n(x)} \mathbf{1}_{I_n}(\, r_n(x,u)\,) \sum_{y\in X} V_{n+1}(y)\, p_{n+1}(y|x,u), \quad n = 0,\ 1,\ \ldots, N-1$$

定理 3.2. 再帰式を解くことにより得られる**マルコフ政策** $\pi^* = (\pi_0^*,\ \pi_1^*,\ \ldots, \pi_{N-1}^*)$:

$$\pi_n^*(x) \in \operatorname*{arg\,max}_{u\in U_n(x)} \mathbf{1}_{I_n}(\, r_n(x,u)\,) \sum_{y\in X} V_{n+1}(y)\, p_{n+1}(y|x,u), \quad n = 0,\ 1,\ \ldots, N-1$$

は原問題 $\mathrm{P}_0(x)$ に対する一般政策全体の中での**最適政策**である.

4 成長型評価系

ここで, 新たな所与のパラメータ $\lambda_0 \in \mathbb{R}$ が埋め込めこまれた**埋め込み問題**を考える.

$$\mathrm{P}_0(x;\lambda_0)\quad \begin{array}{ll} \text{Maximize} & \mathrm{P}^\sigma\big(\lambda_0 \le r_0(X_0,U_0) \le \cdots \le r_{N-1}(X_{N-1},U_{N-1}) \le r_G(X_N) \,\big|\, X_0 = x\big) \\ \text{subject to} & \text{(i) } X_{n+1} \sim p_{n+1}(\cdot\,|x_n,u_n), \quad n=0,\,1,\,\ldots,N-1 \\ & \text{(ii) } \sigma \in \Sigma^{(0,N)} \end{array}$$

定義 4.1. 過去値集合列 $\{\Lambda_n\}_{n=0}^N$ を次のように定義する.

$$\Lambda_n := \{\, r_{n-1}(x,u) \,|\, u \in U_{n-1}(x),\ x \in X \,\}, \quad n=1,\,2,\,\ldots,N, \quad \Lambda_0 := \mathbb{R}$$

m を Λ_1 の最小値よりも小さな定数とすると, $\mathrm{P}_0(x;m)$ は原問題 $\mathrm{P}_0(x)$ と同値である. 同様に新たな所与のパラメータ $\lambda_n \in \Lambda_n$ が埋め込まれた**埋め込み部分問題**の族を定義する.

$$\mathrm{P}_n(x;\lambda_n)\quad \begin{array}{ll} \text{Maximize} & \mathrm{P}^\sigma\big(\lambda_n \le r_n(X_n,U_n) \le \cdots \le r_{N-1}(X_{N-1},U_{N-1}) \le r_G(X_N) \,\big|\, X_n = x\big) \\ \text{subject to} & \text{(i)}' \ X_{t+1} \sim p_{t+1}(\cdot\,|x_t,u_t), \quad t=n,\,n+1,\,\ldots,N-1 \\ & \text{(ii)}' \ \sigma \in \Sigma^{(n,N-n)} \end{array}$$

定理 4.1. $V_n : X\times\Lambda_n \to \mathbb{R}$ $(n=0,\,1,\,\ldots,N-1)$ を埋め込み部分問題の最適値関数とする:

$$V_n(x;\lambda_n) := \operatorname*{Max}_{\text{(i)}',\,\text{(ii)}'} \mathrm{P}^\sigma\big(\lambda_n \le r_n(X_n,U_n) \le \cdots \le r_{N-1}(X_{N-1},U_{N-1}) \le r_G(X_N) \,\big|\, X_n = x\big)$$

また、$V_N : X\times\Lambda_N \to \mathbb{R}$ を

$$V_N(x;\lambda_N) := \mathrm{P}^\sigma\big(\lambda_N \le r_G(X_N) \,\big|\, X_N = x\big)$$

とする. このとき以下の**最適再帰式 (Bellman 方程式)** が成り立つ.

$$V_N(x;\lambda_N) = \mathbf{1}_{[\lambda_N,\infty)}(\,r_G(x)\,)$$

$$V_n(x;\lambda_n) = \operatorname*{Max}_{u\in U_n(x)} \mathbf{1}_{[\lambda_n,\infty)}(\,r_n(x,u)\,)\sum_{y\in X} V_{n+1}(y;r_n(x,u))p_{n+1}(y|x,u), \quad n=0,\,\ldots,N-1$$

定理 4.2. $\bar\sigma_n^* : X\times\Lambda_n \to U$ $(n=0,\,1,\,\ldots,N-1)$ を次のように定義する.

$$\bar\sigma_n^*(x,\lambda_n) \in \operatorname*{arg\,max}_{u\in U_n(x)} \mathbf{1}_{[\lambda_n,\infty)}(\,r_n(x,u)\,)\sum_{y\in X} V_{n+1}(y;r_n(x,u))p_{n+1}(y|x,u), \quad n=0,\,\ldots,N-1.$$

このとき, λ_0 に m $(\le \min\Lambda_1)$ を代入することで得られる**一般政策** $\sigma^* = (\sigma_0^*,\sigma_1^*,\,\ldots,\sigma_{N-1}^*)$:

$$\begin{aligned}
u_0^* &= \sigma_0^*(x_0) := \bar\sigma_0^*(x_0,\lambda_0), \quad \lambda_0 = m\ (m \le \min\Lambda_1) \\
u_1^* &= \sigma_1^*(x_0,x_1) := \bar\sigma_1^*(x_1,\lambda_1), \quad \lambda_1 = r_0(x_0,u_0^*) \\
u_2^* &= \sigma_2^*(x_0,x_1,x_2) := \bar\sigma_2^*(x_2,\lambda_2), \quad \lambda_2 = r_1(x_1,u_1^*) \\
&\ \ \vdots \\
u_n^* &= \sigma_n^*(x_0,\,\ldots,x_n) := \bar\sigma_n^*(x_n,\lambda_n), \quad \lambda_n = r_{n-1}(x_{n-1},u_{n-1}^*) \\
&\ \ \vdots \\
u_{N-1}^* &= \sigma_{N-1}^*(x_0,\,\ldots,x_{N-1}) := \bar\sigma_{N-1}^*(x_{N-1},\lambda_{N-1}), \quad \lambda_{N-1} = r_{N-2}(x_{N-2},u_{N-2}^*)
\end{aligned}$$

は原問題 $\mathrm{P}_0(x)$ に対する一般政策全体の中での**最適政策**である.

いくつかの定義をおき, 以下で定理4.1および定理4.2が得られるアウトラインを述べる.

定義 4.2. 過去値確率変数列 $\{\tilde{\Lambda}_n\}_{n=0}^N$ を次のように定義する.

$$\tilde{\Lambda}_n := r_{n-1}(X_{n-1}, U_{n-1}), \quad n = 1, 2, \ldots, N, \qquad \tilde{\Lambda}_0 := \lambda_0$$

定義 4.3. 拡大状態空間列 $\{W_n\}_{n=0}^N$ を X と Λ_n の直積で定義する:

$$W_n := X \times \Lambda_n, \quad n = 0, 1, \ldots, N$$

また、$\tilde{W}_n = (X_n, \tilde{\Lambda}_n)$ $(\in W_n)$ は確率的に生じる状態を表す.

定義 4.4. $\mathcal{A}_n : W_n \to 2^U \setminus \phi$ $(n = 0, 1, \ldots, N-1)$ は第 *1* 成分のみに依存して

$$\mathcal{A}_n(w_n) := U_n(x_n), \quad \forall w_n \big(= (x_n, \lambda_n) \big) \in W_n$$

と定義する. $\mathcal{A}_n(w_n)$ は拡大状態空間上の**可能決定空間**である.

定義 4.5. 新たな**利得関数** $\bar{r}_n : G_r(\mathcal{A}_n) \to \mathbb{R}$ $(n = 0, 1, \ldots, N-1)$ を次のように定義する.

$$\bar{r}_n(w_n, u_n) := \mathbf{1}_{[\lambda_n, \infty)}\big(r_n(x_n, u_n) \big), \quad (w_n, u_n) = (x_n, \lambda_n, u_n) \in G_r(\mathcal{A}_n).$$

また、**終端利得関数** $\bar{r}_G : W_N \to \mathbb{R}$ を次のように定義する.

$$\bar{r}_G(w_N) := \mathbf{1}_{[\lambda_N, \infty)}\big(r_G(x_N) \big), \quad w_N = (x_N, \lambda_N) \in W_N$$

定義 4.6. 拡大状態空間上の非定常**マルコフ推移法則** $q = \{q_n(\,\cdot\,|w, u)\}$ を次で定義する.

$$\begin{aligned} q_{n+1}\big(w_{n+1} | w_n, u_n \big) &= q_{n+1}\big((x_{n+1}, \lambda_{n+1}) | (x_n, \lambda_n), u_n \big) \\ &:= \begin{cases} p_{n+1}\big(x_{n+1} | x_n, u_n \big) & \lambda_{n+1} = r_n(x_n, u_n) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

定義4.1〜4.6より, 埋め込み問題 $\mathrm{P}_0(x; \lambda_0)$ の目的式および制約条件 (i) は $\bigcup_{n=0}^N W_n$ を状態空間とする次の非負値乗法型評価系の問題 $\bar{\mathrm{P}}_0(x; \lambda_0)$ の目的式および制約条件 (i) とそれぞれ同値であることが示される.

$$\bar{\mathrm{P}}_0(x; \lambda_0) \quad \begin{array}{ll} \text{Maximize} & \mathrm{E}^{\bar{\sigma}}\big[\bar{r}_0(\tilde{W}_0, U_0) \times \cdots \times \bar{r}_{N-1}(\tilde{W}_{N-1}, U_{N-1}) \times \bar{r}_G(\tilde{W}_N) \,\big|\, \tilde{W}_0 = (x, \lambda_0) \big] \\ \text{subject to} & \text{(i) } \tilde{W}_{t+1} \sim q_{n+1}(\,\cdot\,|w_n, u_n), \quad n = 0, 1, \ldots, N-1 \\ & \text{(ii)}'' \ \bar{\sigma} \in \bar{\Sigma}^{(0,N)} \end{array}$$

ただし, $\bar{\mathrm{P}}_0(x; \lambda_0)$ の制約条件 (ii)$''$ の $\bar{\Sigma}^{(0,N)}$ は拡大状態空間上の一般政策全体とし, X 上の一般政策全体 $\Sigma^{(0,N)}$ を含んでいる:

$$\bar{\Sigma}^{(0,N)} := \left\{ \bar{\sigma} = (\bar{\sigma}_0, \ldots, \bar{\sigma}_{N-1}) \;\middle|\; \begin{array}{l} \bar{\sigma}_n : W_0 \times \cdots \times W_n \to \mathcal{A}_n, \\ \bar{\sigma}_n(w_0, \ldots, w_n) \in \mathcal{A}_n(w_n), \forall (w_0, \ldots, w_n) \in W_0 \times \cdots \times W_n \\ n = 0, 1, \ldots, N-1 \end{array} \right\}.$$

定理2.1を $\bar{\mathrm{P}}_0(x; \lambda_0)$ へ適用することで $\bar{\mathrm{P}}_0(x; \lambda_0)$ の最適再帰式が導かれる. さらに, 定理2.2を適用することにより, $\bar{\mathrm{P}}_0(x; \lambda_0)$ の最適政策である拡大状態空間上のマルコフ政策が得られるが, これは X 上の一般政策であり, 埋め込み問題 $\mathrm{P}_0(x; \lambda_0)$ の制約条件 (ii) を満たしていることが示される. したがって, 定理4.1・4.2を得る.

5 最適政策の非マルコフ性

成長型評価系においては最適政策は一般政策クラスの中に存在し, 一般にマルコフ政策では最適化は達成されない. この事実を例を挙げて示す. ここでは 2 状態 2 決定 2 段モデルを考える. 図 1 に示したように, 時刻 1 での状態が s_1 であるときの最適な決定は初期状態 x_0 に依存している. すなわち, 最適政策はマルコフ政策ではない.

- **状態空間** : $X = \{s_1, s_2\}$
- **決定空間**および**可能決定空間** : $U \equiv U_0(x) \equiv U_1(x) \equiv \{a_1, a_2\}$
- **利得関数**および**終端利得関数** : r_0, r_1, r_G

$r_0(x_0, u_0)$

$x_0 \setminus u_0$	a_1	a_2
s_1	3	4
s_2	1	5

$r_1(x_1, u_1)$

$x_1 \setminus u_1$	a_1	a_2
s_1	4	2
s_2	3	2

$r_G(x_2)$

x_2	$r_G(x_2)$
s_1	5
s_2	3

- **推移法則** : $p = \{p_n(y\,|x, u)\}$

$p_1(x_1|x_0, a_1)$

$x_0 \setminus x_1$	s_1	s_2
s_1	1.0	0.0
s_2	1.0	0.0

$p_1(x_1|x_0, a_2)$

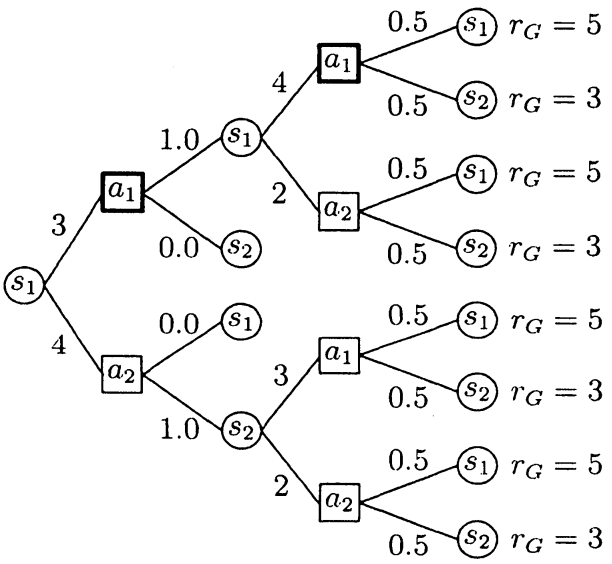
$x_0 \setminus x_1$	s_1	s_2
s_1	0.0	1.0
s_2	0.0	1.0

$p_2(x_2|x_1, a_1)$

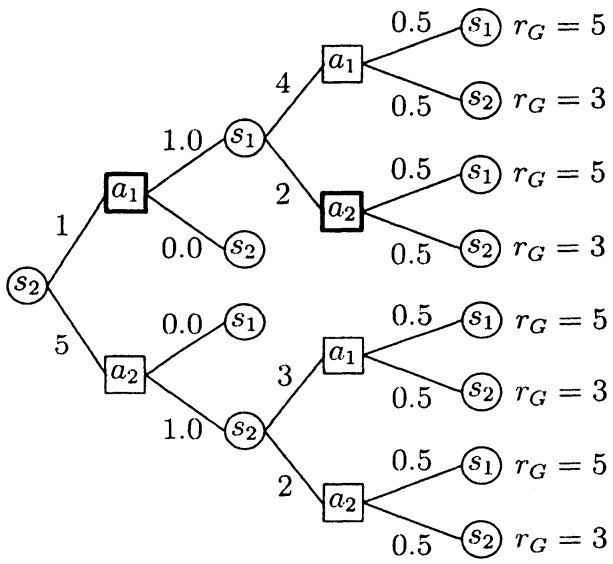
$x_0 \setminus x_1$	s_1	s_2
s_1	0.5	0.5
s_2	0.5	0.5

$p_2(x_2|x_1, a_2)$

$x_0 \setminus x_1$	s_1	s_2
s_1	0.5	0.5
s_2	0.5	0.5



初期状態 $x_0 = s_1$ の場合



初期状態 $x_0 = s_2$ の場合

太枠の □ は最適な決定を表す

図 1: 2 状態 2 決定 2 段モデル

参考文献

[1] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, New Jersey, 1957

[2] R. Bellman and E. Denman, Invariant Imbedding, *Lecture Notes in Operation Research and Mathematical Systems*, Vol.**52**, Springer-Verlag, Berlin, 1971.

[3] D. Blackwell, Discounted dynamic programming, *Ann. Math. Statist.* **36** (1965), pp.226–235.

[4] M. Bouakiz and Y. Kebir, Target-level criterion in Markov decision processes, *Journal of Optimization Theory and Applications*, **86**(1995), pp.1-15.

[5] T. Fujita, K. Tsurusaki, Stochastic optimization of multiplicative functions with negative value. *J. Oper. Res. Soc. Japan* **41**(1998), No. 3, pp.351–373.

[6] R. Howard, *Dynamic Programming and Markov Processes*, The M.I.T. Press, 1960.

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

[8] S. Iwamoto, Fuzzy decision-making through three dynamic programming approaches, *International Journal of Fuzzy Systems*, Vol. **3**, No. 4, December, 2001, 520.526.

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

[10] 岩本誠一, 確率最適化における再帰式と決定樹表, 京大数理研講究録 **1132**(2000), pp.15-23.

[11] S. Iwamoto and T. Ueno, A dual approach in optimizing threshold probabilities, 経済学研究（九州大学経済学会）, **73**, No.1(2006), pp.19–33.

[12] 大坪義夫, 目標集合を持つ非割引マルコフ決定過程における最適閾値確率, 京大数理研講究録 **1306**(2003), pp.83-90.

[13] Y. Ohtsubo, K. Toyonaga, Optimal policy for minimizing risk models in Markov decision processes. *J. Math. Anal. Appl.* **271**(2002), 66-81.

[14] H. Simon, *Models of man*, New York: Wiley, 1957.

[15] 植野貴之・岩本誠一, 確率最適化における過去集積値と未来閾値について, 京大数理研講究録 **1207**(2001), pp.79-100.

[16] 植野貴之・岩本誠一, 制御マルコフ連鎖上での閾値確率最適化の方法, 京大数理研講究録 **1194**(2001), pp.24-32

[17] C. Wu and Y. Lin, Minimizing risk models in Markov decision processed with policies depending on target values, *J. Math. Anal. Appl.*, **231**(1999), 47-67.