閾値確率制御における双対性

02004906　九州大学　植野 貴之　UENO Takayuki
01003676　九州大学　岩本 誠一　IWAMOTO Seiichi

1　はじめに

本報告では、不確実性の下で多段階にわたる最小型評価値が所定の基準値以上になる確率をマルコフ政策クラス上で最適化する問題を考える。
以下、記号と用語を記す。

$N \geq 2$	終端時刻
$X = \{s_1, s_2, \ldots, s_p\}$	状態空間
$U = \{a_1, a_2, \ldots, a_k\}$	決定空間
$x_n \in X$	時刻 n での 状態
$u_n \in U$	時刻 n での 決定
$r_n : X \times U \to R^1$	第 n 利得関数
$r_N : X \to R^1$	終端利得関数
$p = \{p(\cdot\|\cdot,\cdot)\}$	マルコフ推移法則
$\quad p(y\|x,u) \geq 0$	
$\quad \sum_{y\in X} p(y\|x,u) = 1$	
$c \in R^1$	基準値
$\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$	マルコフ政策
$\Pi = \{\pi\}$	マルコフ政策クラス
$\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}$	一般政策

2　マルコフ政策クラス問題

本報告での閾値確率最大化問題 $M_0(x_0)$ は

$$\begin{aligned} &\text{Max} \quad P^{\pi}_{x_0}(\, r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c) \\ &\text{s.t.} \quad (\text{i})_{\text{n}} \quad X_{n+1} \sim p(\,\cdot\,|x_n, u_n) \qquad (1) \\ &\qquad\quad (\text{ii})_{\text{n}} \quad u_n \in U \quad n = 0, 1, \ldots, N-1 \end{aligned}$$

で表される。意志決定者がマルコフ政策 $\pi = \{\pi_0, \ldots, \pi_{N-1}\}$ $(\in \Pi)$ を採用すると、最大化問題 (1) の閾値確率は「部分」多重和

$$\begin{aligned} &P^{\pi}_{x_0}(\, r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c\,) \\ =\ &\sum\sum\cdots\sum_{(x_1,x_2,\ldots,x_N)\in(*)} p_1 p_2 \cdots p_N \qquad (2) \\ &\qquad (\, p_n = p(x_n|x_{n-1}, u_{n-1})\,) \end{aligned}$$

で表わされる。ただし、多重和をとる領域 $(*)$ は

$$\begin{aligned} &r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c \qquad (3) \\ &\qquad (\, r_n = r_n(x_n, u_n),\ r_N = r_N(x_N)\,) \end{aligned}$$

を満たす $(x_1, x_2, \ldots, x_N) \in X \times X \times \cdots \times X$ 全体である。ここに、式 (2),(3) における決定列 $\{u_0,\ u_1,\ \cdots,\ u_{N-1}\}$ はマルコフ政策 π の決定関数列を通して定まっている：

$$u_0 = \pi_0(x_0),\ u_1 = \pi_1(x_1),\ \ldots,\ u_{N-1} = \pi_{N-1}(x_{N-1}).$$

この閾値確率最大化問題に対しては、期待値問題に変換することなく、閾値確率自身を直接最適化する。

さて、時刻 n で状態 $x_n (\in X)$ から始まる部分閾値確率問題

$$\begin{aligned} &\text{Max} \quad P^{\pi}_{x_n}(\, r_n \wedge \cdots \wedge r_N \geq c) \\ &\text{s.t.} \quad (\text{i})_{\text{m}},\ (\text{ii})_{\text{m}} \quad n \sim N\text{–}1 \end{aligned} \qquad (4)$$

のマルコフ政策 $\pi = \{\pi_n, \pi_{n+1}, \ldots, \pi_{N-1}\} \in \Pi(n)$ にわたる最大値を $f_n(x_n)$ とする。ただし

$$f_N(x_N) \overset{\triangle}{=} \phi(r_N(x_N)). \qquad (5)$$

ここに ϕ は区間 $[c, \infty)$ の定義関数である：

$$\phi(y) = \begin{cases} 1 & \text{if} \quad y \geq c \\ 0 & \text{otherwise.} \end{cases}$$

このとき、次の関係式を得る。

補題 2.1 任意のマルコフ政策 $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ と任意の $x_n \in X$ に対して、

$$P^{\pi}_{x_n}(\, r_n \wedge \cdots \wedge r_N \geq c)$$

$$= \begin{cases} \displaystyle\sum_{x_{n+1}\in X} P^{\pi'}_{x_{n+1}}(\, r_{n+1} \wedge \cdots \wedge r_N \geq c) p_n & \\ & \textit{if} \quad r_n \geq c \\ 0 & \textit{otherwise} \end{cases} \qquad (6)$$

が成り立つ。ここに

$$r_n = r(x_n, u_n), \qquad u_n = \pi_n(x_n),$$
$$\pi' = \{\pi_{n+1}, \ldots, \pi_{N-1}\},\ p_{n+1} = p(x_{n+1}|x_n, u_n).$$

したがって、上述の補題から後向きの再帰式が成り立つ：

定理 2.1

$$f_n(x) = \begin{cases} \displaystyle\operatorname*{Max}_{u;r(x,u)\geq c} \sum_{y\in X} f_{n+1}(y)p(y|x,u) \\ \qquad\qquad\qquad if\ \exists\, u\ ;\ r(x,u) \geq c \\ 0 \qquad\qquad\qquad otherwise \end{cases} \quad (7)$$

$$x \in X,\quad 0 \leq n \leq N-1$$

$$f_N(x) = \begin{cases} 1 & if\ \ r(x) \geq c \\ 0 & otherwise \end{cases} \qquad x \in X. \quad (8)$$

さて、式 (7) の最大（値に到達する）点の全体を $\pi_n^*(x)$ としよう。すなわち、

$$\pi_n^*(x) = \begin{cases} \text{Max に到達する } u; r(x,u) \geq c \text{ の全体} \\ \qquad\qquad\qquad \text{if } \exists\, u\ ;\ r(x,u) \geq c \\ \text{任意の } u \in U \qquad\qquad otherwise \end{cases} \quad (9)$$

$$x \in X,\quad 0 \leq n \leq N-1.$$

このようにして得られたマルコフ政策 $\pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\}$ は最適である。

3 双対問題

まず、最大化問題 $M_0(x_0)$ に対して最小化問題 $m_0(x_0)$

$$\begin{array}{ll} \min & P_{x_0}^{\pi}(r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c) \\ \text{s.t.} & (\mathrm{i})_\mathrm{n},\ (\mathrm{ii})_\mathrm{n} \quad n = 0,1,\ldots,N-1 \end{array} \quad (10)$$

を導入する。次に、「上限 c を閾値とする」確率最小化問題 $d_0(x_0)$

$$\begin{array}{ll} \min & P_{x_0}^{\pi}(r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N < c) \\ \text{s.t.} & (\mathrm{i})_\mathrm{n},\ (\mathrm{ii})_\mathrm{n} \quad n = 0,1,\ldots,N-1 \end{array} \quad (11)$$

および、この最大化問題 $D_0(x_0)$

$$\begin{array}{ll} \text{Max} & P_{x_0}^{\pi}(r_0 \wedge r_1 \wedge \cdots \wedge r_{N-1} \wedge r_N < c) \\ \text{s.t.} & (\mathrm{i})_\mathrm{n},\ (\mathrm{ii})_\mathrm{n} \quad n = 0,1,\ldots,N-1 \end{array} \quad (12)$$

を考える。このとき、$\mathrm{d}_0(x_0)$ を $\mathrm{M}_0(x_0)$ の、$\mathrm{D}_0(x_0)$ を $\mathrm{m}_0(x_0)$ のそれぞれ**双対問題**という。

以上、4つの閾値確率最適化問題群に対して次の双対性が成り立つ。

定理 3.1 （双対定理）

(i) 最大化問題群 $\mathrm{M} = \{M_n(x_n)\}$ と最小化問題群 $\mathrm{d} = \{d_n(x_n)\}$ の最適値関数の和は常に 1 である：

$$\begin{array}{c} f_n(x) + h_n(x) \ = \ 1 \\ n = 0,1,\ldots,N, \quad x \in X. \end{array} \quad (13)$$

マルコフ政策 π が M に対して最適である必要十分条件は、それが d に対して最適であることである。

(ii) 最小化問題群 $\mathrm{m} = \{m_n(x_n)\}$ と最大化問題群 $\mathrm{D} = \{D_n(x_n)\}$ の最適値関数の和は常に 1 である：

$$\begin{array}{c} g_n(x) + k_n(x) \ = \ 1 \\ n = 0,1,\ldots,N, \quad x \in X. \end{array} \quad (14)$$

マルコフ政策 π が m に対して最適である必要十分条件は、それが D に対して最適であることである。

定理 3.2 （一致定理）

(i) 最大化問題群 M の最適政策 π^* は、最小化問題群 d の最適政策 $\hat{\pi}$ に一致している：

$$\pi^* = \hat{\pi}. \quad (15)$$

(ii) 最小化問題群 m の最適政策 $\tilde{\pi}$ は、最大化問題群 D の最適政策 $\bar{\pi}$ に一致している：

$$\tilde{\pi} = \bar{\pi}. \quad (16)$$

参考文献

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

[2] S. Iwamoto, Maximizing threshold probability through invariant imbedding, *Ed. H.F. Wang and U.P. Wen, Proceedings of The Eighth BELLMAN CONTINUUM*, Hsinchu, ROC, Dec.2000, pp.17-22.

[3] 植野貴之・岩本誠一, 最小型評価系の閾値確率制御, 日本OR学会秋季研究発表会アブストラクト集, 2000, pp.124-125.