2000年度日本オペレーションズ・リサーチ学会
秋季研究発表会

最小型評価系の閾値確率制御

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

本論文では、有限段制御マルコフ連鎖において最小型評価値が所定の値以上になる閾値確率を最大化する問題を考える。この閾値確率制御問題に対して2つの動的計画法 —— (1) パラメトリック法、(2) 全履歴法 —— によって共通の最適解が得られることを示す。いわゆる加法型評価の期待値最適化問題ではマルコフ政策クラスの中で最適政策が得られる。すなわち、マルコフ政策は十分である ([3])。これに対して、「閾値確率」制御問題ではより広い一般政策クラスに最適解が得られる。すなわち、閾値確率制御問題ではマルコフ政策は十分でない。(1) パラメトリック法では「拡大」マルコフ政策クラスで最適化し、(2) 全履歴法 では原始政策クラスで最適化する。原始政策クラスで得られた最適政策を一般政策クラスに圧縮して、求める最適政策が得られる。さらに、3状態2決定2段モデルに対して2つの方法によって具体的に最適解を求め、一致することが示される。

2 確率制御問題

本節では、不確実性の下で最小型評価の確率制御を考える。以後全体を通して次のデータが与えられているものとする：

$N \geq 2$	**終端時刻**
$X = \{s_1, \cdots, s_p\}$	**状態集合**
$U = \{a_1, \cdots, a_k\}$	**決定集合**
$x_n \in X$	時刻 n における**状態**
$u_n \in U$	時刻 n における**決定**
$r_n : X \times U \to R^1$	**第 n 段利得**
$r_N : X \to R^1$	**終端利得**
$p = \{p(\cdot\|\cdot,\cdot)\}$	**マルコフ推移法則** $p(y\|x,u) \geq 0$, $\sum_{y \in X} p(y\|x,u) = 1$
$c \in R^1$	**基準値**
$\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}$	**一般政策**
$\mu = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$	**原始政策**

このとき、次の問題を考える：

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

ただし $r_n = r_n(X_n, U_n)$, $r_N = r_N(X_N)$ で、$Y \sim p(\cdot|x,u)$ 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 $P^{\sigma}_{x_0}$ は条件付き確率 $p(\cdot|\cdot,\cdot)$、政策 $\sigma = \{\sigma_1, \sigma_2, ..., \sigma_N\}$ 及び初期状態 $x_1 \in X$ に依存して定まる全履歴空間 $X \times U \times X \times U \times \cdots \times U \times X$ 上の確率測度を表す。したがって、意志決定者が一般 (現在までの状態列に依存する) 政策 σ を採用すると、最大化問題 (1) の閾値確率は「部分」多重和

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

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 $(x_1, x_2, \ldots, x_N) \in X \times X \times \cdots \times X$ 全体にわたる多重和である。ここに、式 (2),(3) における決定列 $\{u_0,\ u_1,\ \cdots,\ u_{N-1}\}$ は一般政策 $\sigma = \{\sigma_0, \ldots, \sigma_{N-1}\}$ を通して定まっている：

$$\begin{aligned} & u_0 = \sigma_0(x_0),\ u_1 = \sigma_1(x_0, x_1),\ \ldots, \\ & \qquad u_{N-1} = \sigma_{N-1}(x_0, x_1, \ldots, x_{N-1}). \end{aligned}$$

一般に、確率変数 Y が c 以上になる確率 $P(Y \geq c)$ は、数直線 R^1 上の区間 $[c, \infty)$ の定義関数

$$\psi(y) := 1_{[c,\infty)}(y) := \begin{cases} 1 & y \geq c \\ 0 & \text{その他} \end{cases}$$

を通した確率変数 $\psi(Y)$ の期待値 $E[\psi(Y)]$ で表わされる：

$$P(Y \geq c) = E[\psi(Y)].$$

このことに注意すると、一般問題 (1) の閾値確率は定義関数 $\psi = \psi(y)$ を通した期待値になる：

$$\begin{aligned} & P^{\sigma}_{x_0}(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N \geq c) \\ = \ & E^{\sigma}_{x_0}[\psi(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N)]. \end{aligned}$$

すなわち、「部分」多重和は定義関数を通した「全」多重和に等しい。

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.

3 拡大マルコフ政策クラス問題

問題 (1) に対して、**過去値集合列** $\{\Lambda_n\}$ を

$$\begin{aligned}\Lambda_0 &:= \{\lambda_0\} \quad \lambda_0 \text{ は、} r_n, r_N \text{の取り得る最大値}\\ \Lambda_n &:= \{\lambda_n \,|\, \lambda_n = r_0 \wedge \cdots \wedge r_{n-1},\\ &\qquad (x_0, u_0, \ldots, x_{n-1}, u_{n-1})\\ &\qquad \in X\times U\times\cdots\times X\times U\}\\ &\qquad\qquad n = 1, \ldots, N\end{aligned}$$

で定義すると、次の前向きの再帰式が成り立つ：

補題 3.1 (前向きの再帰式)

$$\begin{aligned}&\Lambda_0 = \{\lambda_0\}\\ &\Lambda_n = \{\lambda \wedge r_{n-1}(x,u) \,|\, \lambda \in \Lambda_{n-1},\ (x,u) \in X\times U\}\\ &\qquad\qquad n = 1, 2, \ldots, N.\end{aligned}$$

さらに、第 n 段までの**過去値確率変数** $\tilde{\Lambda}_n$:

$$\begin{aligned}\tilde{\Lambda}_0 &:= \lambda_0 \quad \text{ただし } \lambda_0 \text{ は十分大きな定数}\\ \tilde{\Lambda}_n &:= r_0(X_0, U_0) \wedge \cdots \wedge r_{n-1}(X_{n-1}, U_{n-1})\end{aligned}$$

を導入すると、拡大状態空間列 $\{X\times\Lambda_n\}_0^N$ 上の**終端型期待値**評価問題

$$\begin{aligned}\text{Max} \quad & \tilde{E}^{\gamma}_{y_0}[\psi(\tilde{\Lambda}_N \wedge r_N(X_N))]\\ \text{s.t.} \quad & \text{(i), (ii)} \ \ n = 0, 1, \ldots, N-1\\ & \text{(i)}' \ \ \tilde{\Lambda}_{n+1} = \tilde{\Lambda}_n \wedge r_n(X_n, U_n)\end{aligned}$$

が考えられる。ただし、 $y_0 = (x_0; \lambda_0)$. ここに $\tilde{E}^{\gamma}_{y_0}$ は、初期状態 y_0 、拡大マルコフ政策 γ および新マルコフ推移法則 q によって拡大状態空間列上に定まる確率測度 $\tilde{P}^{\gamma}_{y_0}$ に基づく期待値作用素である。

この終端型問題に対して、拡大状態 $y_n = (x_n; \lambda_n)$ から始まる部分問題

$$\begin{aligned}\text{Max} \quad & \tilde{E}^{\gamma}_{y_n}[\psi(\tilde{\Lambda}_N \wedge r_N(X_N))]\\ \text{s.t.} \quad & \text{(i), (ii), (i)}' \ \ n \sim N\text{–}1\end{aligned}$$

の最大値を $u^n(y_n)$ とすると、次の再帰式が成り立つ ([2]) :

定理 3.1 (後向きの再帰式)

$$\begin{aligned}u^N(x;\lambda) &= \psi(\lambda \wedge r_N(x)) \qquad x \in X,\ \lambda \in \Lambda_N\\ u^n(x;\lambda) &= \operatorname*{Max}_{u\in U} \sum_{y\in X} u^{n+1}(y; \lambda \wedge r_n(x,u)) p(y|x,u)\\ &\qquad x \in X,\ \lambda \in \Lambda_n,\ 0 \le n \le N-1.\end{aligned}$$

4 原始政策クラス問題

さらに、(1) に対しては、原始 (全履歴に依存する) 政策クラス上の問題が考えられる：

$$\begin{aligned}\text{Max} \quad & P^{\mu}_{x_0}(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N \ge c)\\ \text{s.t.} \quad & \text{(i), (ii)}\end{aligned}$$

これに対しては、後向きの再帰式が成り立つ：

定理 4.1 (後向きの再帰式)

$$\begin{aligned}w_n(h) &= \operatorname*{Max}_{u\in U} \sum_{y\in X} w_{n+1}(h,u,y) p(y|x,u)\\ &\qquad h \in H_n,\ \ 1 \le n \le N-1\\ w_{N+1}(h) &= \psi(r_0 \wedge \cdots \wedge r_{N-1} \wedge r_N)\\ &\qquad h \in H_N.\end{aligned}$$

5 3－2－2モデル

ここでは3状態2決定2段問題を考える：

$$\begin{aligned}\text{Max} \quad & P^{\mu}_{x_0}(r_0(U_0) \wedge r_1(U_1) \wedge r_2(X_2) \ge 0.7)\\ \text{s.t.} \quad & \text{(i)} \ \ X_{n+1} \sim p(\cdot\,|x_n, u_n) \qquad n = 0, 1\\ & \text{(ii)} \ \ u_0 \in U,\ u_1 \in U\end{aligned}$$

数値例として、Bellman & Zadeh([1]) の問題を考え、2つの方法で共通の最適解が得られることを示す。

【各段評価】

$$r_2(s_1) = 0.3 \quad r_2(s_2) = 1.0 \quad r_2(s_3) = 0.8$$

$$r_1(a_1) = 1.0 \quad r_1(a_2) = 0.6$$

$$r_0(a_1) = 0.7 \quad r_0(a_2) = 1.0$$

【推移確率】 $p(x_{n+1}|x_n, u_n)$, $n = 1, 2$

$u_n = a_1$

$x_n \backslash x_{n+1}$	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

$u_n = a_2$

$x_n \backslash x_{n+1}$	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

References

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

[2] Iwamoto, S. and Fujita, T.: *Stochastic decision-making in a fuzzy environment, J. Operations Res. Soc. Japan*, **38**, 1995

[3] 岩本 誠一 : 多段確率決定樹表について, 日本OR学会秋季研究発表会アブストラクト集, pp. 58-59, 1999.

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.