1996年度日本オペレーションズ・リサーチ学会
春季研究発表会

不確実性の下での最適危機管理について

01003676　九州大学　岩本誠一　IWAMOTO Seiichi

1　はじめに

最近の日本では危機意識の欠如がいろいろな型・状況で叫ばれているようである。この報告では、多期間にわたる不確実性の下での最適な危機管理という問題を捉え、その最適政策の構成と性質を考える。

以下では簡単のために2期間の不確実性の下でのミニマックス化を行う。一般に、危険度（リスク）、危機の度合や後悔の程度といった諸量は非加法的である。これらの量の複数個の総体としては単に足し合わせて総合計をだすのではなくて、それらの値のうちの最大の値ないしは最小の値で評価すべきであろう。以下では、値0が最も危機的な状態にあり、値1が最も平和的な状態にあることを示すものとする。ここでは多期間にわたって不確実性が存在する場合を対象にしているので、ランダムに変動する危険に対して最悪の危険（ワーストケース）を評価値と考え、この期待値をできるだけよくすることを考える。すなわち、制御パラメータ付きマルコフ推移システムにおいて各段評価の（危機の度合いを表す）値の最悪値の期待値を最大化する。

2　問題と定式化

いま、ある2段システムが、初期状態 $x_0 \in X$ から出発して制御マルコフ推移法則 $p(x_1|x_0,u_0)$, $p(x_2|x_1,u_1)$ に従って確率ツリーを構成しながら移り、ある確率で $x_2 \in X$ になり、そこで終了するとする。ただし、$u_0 \in U$ は状態 x_0 を観察して取る決定（行動ともいう）であり、$u_1 \in U$ は次の状態 x_1 に依存して取る決定である。このとき、第1段では決定 u_0 に依存した評価（リスク）$r_0(u_0)$ が課され、第2段では u_1 に関係した評価 $r_1(u_1)$ が課され、最終の第2段終了時点には終端状態 x_2 に依存したゴール（目標）評価 $r_G(x_2)$ が課されるとする。このとき、システム全体としてはこれら三つの評価値の最小値（意志決定者にとっての最悪値）$r_0(u_0)\wedge r_1(u_1)\wedge r_G(x_2)$ が総合評価として下されるもとのする。すなわち、システムとしての危機の度合いを各段でのリスクのうちで最も悪い値と考える。不確実な状況の下でこの総合評価値（最も悪い評価値）を最大にするように行動するには、意志決定者が各段での状態に応じてどのように決定を取っていけばよいかが問題である：

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

ここに E は、初期状態 x_0, マルコフ推移確率 $p(y|x,u)$ および 政策 $\pi=\{\pi_0,\pi_1\}$ から履歴の直積空間 $H=X\times U\times X\times U\times X$ 上に唯一定まる確率測度 $P^{\pi}_{x_0}$ による期待値作用素である。したがって、この評価関数の値は

$$\begin{aligned} &\mathrm{E}[r_0(u_0)\wedge r_1(u_1)\wedge r_G(x_2)] \\ =&\sum_{(x_1,x_2)\in X\times X}\{[r_0(u_0)\wedge r_1(u_1)\wedge r_G(x_2)] \\ &\qquad \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\} \end{aligned} \tag{2}$$

である。ただし、u_0,u_1 はそれぞれ**決定関数** π_0,π_1 を通して

$$u_0=\pi_0(x_0),\quad u_1=\pi_1(x_1) \tag{3}$$

で定まっている。

3　埋め込みと再帰式

さて、問題(1)を不変埋没原理を用いて解こう。すなわち、最小型評価系の直前に新しくパラメータ $\lambda\in[0,1]$ を導入して、部分問題群

$$\begin{aligned} v^2(x_2;\lambda) &= \lambda\wedge r_G(x_2) \\ v^1(x_1;\lambda) &= \operatorname*{Max}_{\pi_1}\sum_{x_2\in X}[\lambda\wedge r_1(u_1)\wedge r_G(x_2)]p(x_2|x_1,u_1) \\ v^0(x_0;\lambda) &= \operatorname*{Max}_{\pi_0,\pi_1}\sum_{x_1,x_2}\{[\lambda\wedge r_0(u_0)\wedge r_1(u_1)\wedge r_G(x_2)] \\ &\qquad \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\} \end{aligned} \tag{4}$$

を定義する。式(4)に $\lambda=1$ を代入した問題は原問題(1)になっている。したがって、$v^0(x_0;1)$ が求める最大値である。

このとき、2変数最大値関数列 $\{v^0, v^1, v^2\}$ 間には次の再帰式が成り立つ：

定理 1　$v^2(x;\lambda) = \lambda \wedge r_G(x)$

$$v^1(x;\lambda) = \operatorname*{Max}_{u\in U} \sum_{y\in X} v^2(y;\lambda \wedge r_1(u))p(y|x,u) \quad (5)$$

$$v^0(x;\lambda) = \operatorname*{Max}_{u\in U} \sum_{y\in X} v^1(y;\lambda \wedge r_0(u))p(y|x,u).$$

4　最適解の構成

原問題(1)の最適解を考えよう。まず、式(5)を解いて、最適値関数列 $\{v^0, v^1, v^2\}$ が得られる。同時に、最適政策 $\pi^* = \{\pi_0^*, \pi_1^*\}$ が得られる。従って、初期状態 $(x_0; 1)$ からの最適行動

$$(x_0;1) \to u_0^* \to (X_1^*;\lambda_1) \to U_1^* \to (X_2^*;\Lambda_2) \quad (6)$$

が、最適政策 π^* と状態推移法則 $p(y|x,u)$ によって拡大履歴空間

$$\tilde{H} = (X\times[0,1])\times U\times(X\times[0,1])\times U\times(X\times[0,1]) \quad (7)$$

上に確率的に定まる。ただし

$$u_0^* = \pi_0^*(x_0;1),\ \ X_1^* \sim p(\cdot|x_0,u_0^*),\ \ \lambda_1 = 1\wedge r_0(u_0^*)$$

$$U_1^* = \pi_1^*(X_1;\lambda_1),\ X_2^* \sim p(\cdot|X_1,U_1^*),\ \Lambda_2 = \lambda_1\wedge r_1(U_1^*). \quad (8)$$

ここに、小文字は確定数、大文字は確率変数を表している。

原問題(1)の最大期待値は $v^0(x_0;1)$ で与えられる。また、最大点（点とはいえ、ここでは確率過程）は拡大履歴空間上の最適行動(6)を元の履歴空間 $H = X\times U\times X\times U\times X$ 上に射影した行動

$$x_0 \to u_0^* \to X_1^* \to U_1^* \to X_2^* \quad (9)$$

で与えられる。これが本来の最適行動である。すなわち、各段評価の最悪値をシステム全体としての評価値としたとき、(9)がその期待値を最大にする行動である。この最適行動は推移確率 $p_0^*(x_1|x_0)$, $p_1^*(x_2|x_0,x_1)$ をもつ X 上の確率過程になっている。ただし

$$p_0^*(x_1|x_0) = p(x_1|x_0,u_0^*),\ p_1^*(x_2|x_0,x_1) = p(x_2|x_1,u_1^*) \quad (10)$$

ここに

$$u_0^* = \pi_0^*(x_0;1),\ \ \lambda_1 = r_0(u_0^*),\ \ u_1^* = \pi_1^*(x_1;\lambda_1) \quad (11)$$

だから、u_0^* は x_0 に依存し、u_1^* は x_0, x_1 の関数になっている。

さて、ここで注意すべきは、本来の制御マルコフ連鎖上で最大期待値に到達する（最適な）確率過程 $\{x_0, X_1^*, X_2^*\}$ は必ずしも X 上のマルコフ連鎖になっていないということである。すなわち、原問題(1)の最適政策はマルコフ政策の中には存在しないのである。事実、最適過程 $\{x_0, X_1^*, X_2^*\}$ を生じせしめる最適政策$\sigma^* = \{\sigma_0^*, \sigma_1^*\}$ の第2決定関数は初期状態 x_0 にも依存していて、X上のマルコフ政策になっていない。

定理 2　原問題*(1)*の最適政策 $\sigma^* = \{\sigma_0^*, \sigma_1^*\}$ は拡大過程*(5)*の最適政策 $\pi^* = \{\pi_0^*, \pi_1^*\}$ を用いて

$$\sigma_0^*(x_0) = \pi_0^*(x_0;1),\ \ \sigma_1^*(x_0,x_1) = \pi_1^*(x_1;\lambda_1) \quad (12)$$

で表される。

ここでも λ_1 は r_0, u_0^* を通して x_0 の関数になっていることに注意すべきである。

5　埋め込まない場合

評価系の集積値を表すパラメータ λ を導入しないで部分問題群を考えても（すなわち、加法型評価系の問題と同様な部分問題群を定義しても）、その（1変数）最大値関数列 $\{u^0, u^1, u^2\}$ 間には再帰式は一般に成立しない。

6　拡大過程上の終端評価

X 上の最小型評価である原問題(1)を埋め込んだ問題群(4)は拡大状態空間 $Y = X\times[0,1]$ 上で終端型評価をもつ決定過程に同値変換される。

7　定式化と最適性の再検討

いずれにしても、埋め込まないままで原問題を解こうとすると、無理が生じる。パラメータ λ を導入して解くことが重要である。この方法はいわば**確率的終端状態接近方法** (stochastic terminal state approach) である。

8　おわりに

結論的には、多期間にわたる不確実な状況の下で危険、危機、後悔、ショックというものを総体として和らげる**最適な方法は**、単に現在だけを見て行動するのではなく、**過去から現時点までの累積評価値を絶えず見据えて行動することである**ことを示している。