数理解析研究所講究録
第1802巻 2012年 92-98

Chance 理論を利用した decision process の構成

布和額尔敦 [1], 侯 平軍 [2], 影山 正幸 [3]

[1,2] 千葉大学理学研究科　基盤理学専攻

[3] 統計数理研究所

摘要. 離散時間のマルコフタイプのハイブリッド決定過程を定義し，割引された総利得の最適化問題を考察する．そして動的計画法の考え方を適用して，最適利得関数はある作用素の不動点として与えられ，最適政策の特徴付けがなされる．

1 はじめに

可信性測度は Liu[7] によって提案された．自己双対性の公理を含んだ理論であり，不確実性の問題を分析する一つの手法として研究がなされている．一方， Chance 理論は可信性理論と確率理論の統合理論であり [6] よって提案されている．可信性理論に基づいて，M. Kageyama and K. Iwamura[4] は，可信性核の概念を与えて離散時間の可信性過程を構成した．さらに，2011 年に，Chnace 理論の動的システムとして，M. Kageyama ら [5] が離散時間のハイブリッド過程を定義してその性質を分析した．ファジー動的システムについてはたとえば [3] を参照のこと．

この論文は，Chance 理論を利用した M. Kageyama らの論文 [5] のモデルに action を採り入れて，マルコフタイプの決定過程を構成した．

まず，必要な記号と基礎となる Chance 測度とハイブリッド変数の期待値などを紹介しておく．

任意の非空集合 X，関数 g: $X \mapsto [0,0.5]$ が下記の式を満たすとき関数 g は X に関して条件 $\mathscr{K}$ を満たすということにする．

$$\sup_{x\in X} g(x) = 0.5, \tag{1.1}$$

$$\sup_{x\neq x^*, x\in X} g(x) = 0.5 \quad \text{if } g(x^*) = 0.5. \tag{1.2}$$

そのような関数の全体を $\mathscr{K}(X)$ で表す．

ボレル集合 X に対して，$\mathscr{P}(X)$ を X 上の確率測度の全体とする．X の部分集合の全体を $\mathcal{P}(X)$ で表す．$\Theta := \{\theta_1, \theta_2, \cdots, \theta_l\}$ をパラメータ空間で有限集合とする．可算集合 $S := \{1, 2, \cdots\}$ を状態空間とする．

$\mathscr{K}(\Theta) := \{g = (g(\theta_1), g(\theta_2), \cdots, g(\theta_l)) | 0 \leq g(\theta_i) \leq 0.5,\ g \in \mathscr{K}(\Theta)\}$，$\mathscr{P}(S) := \{p = (p(1), p(2), \cdots) | p(i) \geq 0,\ \sum_{i=1}^{\infty} p(i) = 1\}$，そして，$\boldsymbol{X} := \mathscr{K}(\Theta) \times \mathscr{P}(S)$ と定義する．

$(g, p) \in \boldsymbol{X}$ にたいして，$\Theta \times S$ 上の関数 δ を下記の通り定義する．

$$\delta(\Lambda | g, p) = \max_{\theta \in \Theta} g(\theta) \wedge p(\Lambda(\theta)). \tag{1.3}$$

ただし

$$\Lambda(\theta) = \{x \in S | (\theta, x) \in \Lambda\}. \tag{1.4}$$

$\Theta \times S$ 上の chance 測度 $\mathrm{Ch}(\cdot | g, p)$ は，次で定義される (cf.[7])．任意の $\Lambda \in \Theta \times S$ に対して

$$\mathrm{Ch}(\Lambda | g, p) = \begin{cases} \delta(\Lambda | g, p) & \text{if } \delta(\Lambda | g, p) < 0.5, \\ 1 - \delta(\bar{\Lambda} | g, p) & \text{if } \delta(\Lambda | g, p) = 0.5. \end{cases} \tag{1.5}$$

$\{\Theta \times S, \mathscr{P}(\Theta \times S), \mathrm{Ch}(\cdot | g, p)\}$ は $(g, p) \in \boldsymbol{X}$ から構成された chance 空間という．

関数 $\tilde{r}$: $\Theta \times S \mapsto \mathbb{R}^+ := [0, \infty)$ をハイブリッド変数といい，その期待値 $E(\tilde{r} | g, p)$ は Choquet 積分により下の式で定義する．

$$E(\tilde{r} | g, p) = \int_0^{\infty} \mathrm{Ch}(\{\tilde{r} \geq t\} | g, p)\, dt.$$

2 ハイブリッド決定過程

任意の空でない集合 X, Y に対して，$\boldsymbol{q}(y|x) : X \times Y \to [0, 0.5]$ が，各 $x \in X$ に対して $\boldsymbol{q}(\cdot | x) \in \mathscr{K}(Y)$ を満たすとき可信性核 (Credibilistic kernel) といい，その全体を $\mathscr{K}(Y|X)$ で表す (cf.[4])．任意のボレル集合 X, Y に対して X が与えられたときの Y の確率核 (stochastic kernel) といい，その全体を $\mathscr{P}(Y|X)$ で表す (cf.[1])．

ここで，マルコフタイプのハイブリッド決定過程を定義する．$A = \{a_1, a_2, \cdots, a_k\}$ を有限な行動空間，可信性核を $q(\theta' | \theta, a) \in \mathscr{K}(\Theta | \Theta \times A)$，確率核を $p(j | i, a) \in \mathscr{P}(S | S \times A)$，上に有界な利得関数を $\tilde{r} : \Theta \times S \times A \to R^+ = [0, \infty)$，とする．状態空間 $\boldsymbol{X} = \mathscr{K}(\Theta) \times \mathscr{P}(S)$ の上の離散時間のマルコフタイプの決定過程を下記の通りに定義する．

初期状態 $(g_0,p_0)\in\boldsymbol{X}$ に対して行動 $a_1\in A$ を選択すると次の二つのことが起こる.
(i) 状態の推移：$(g_0,p_0)\to(g_1,p_1)$ は下の式 (2.1) と (2.2) で与えられる.

$$g_1(\theta')=\max_{\theta\in\Theta}g_0(\theta)\wedge\boldsymbol{q}(\theta'|\theta,a_1)=:T_{\boldsymbol{q}}(g_0|a_1)\ \ \theta'\in\Theta, \tag{2.1}$$

$$p_1(j)=\sum_{i\in S}\boldsymbol{p}(j|i,a_1)p_0(i)=:T_{\boldsymbol{p}}(p_0|a_1). \tag{2.2}$$

(ii) 利得 $\tilde{r}$ 発生し，その期待値は次で与えられる.

$$r(g_0,p_0,a_1):=\int\mathrm{Ch}(\{(\theta,x)|\tilde{r}(\theta,x,a)\geq t\}|g_0,p_0)\,dt. \tag{2.3}$$

以下状態 (g_1,p_1) から決定と状態の推移，利得の発生が次々と繰り返す.
状態空間 $\boldsymbol{X}$ の距離 ρ を次により定義する. $(g,p),(g',p')\in\boldsymbol{X}$ に対して

$$\rho((g,p),(g',p'))=\max_{\theta\in\Theta}|g(\theta)-g'(\theta)|+\sup_{i\in S}|p(i)-p'(i)|. \tag{2.4}$$

ρ による $\boldsymbol{X}$ の Borel 部分集合の全体を $\mathcal{B}(\boldsymbol{X})$ と表す.
定常政策 f を，可測関数 $f:\boldsymbol{X}\to A$ とする，i.e., 任意の $a\in A$ に対して

$$B_a:=\{(g,p)\in\boldsymbol{X}|f(g,p)=a\}\in\mathcal{B}(\boldsymbol{X}). \tag{2.5}$$

定常政策 f の全体を $\mathbb{F}$ で表す.
初期状態 $(g,p)\in\boldsymbol{X}$ のとき $f\in\mathbb{F}$ を用いたときの割引された総利得 $(0<\beta<1)$ を

$$\varphi_f(g,p)=\sum_{t=0}^{\infty}\beta^t r(g_t,p_t,f(g_t,p_t)) \tag{2.6}$$

と定義する. 但し

$$\begin{cases}g_0=g,p_0=p,\\ g_t=T_{\boldsymbol{q}}(g_{t-1}|f(g_{t-1},p_{t-1})),p_t=T_{\boldsymbol{p}}(p_{t-1}|f(g_{t-1},p_{t-1})), & t\geq 1.\end{cases} \tag{2.7}$$

問題は $\varphi_f(g,p)$ を最大にする $f\in\mathbb{F}$ を求めることである.
$\bar{\varphi}(g,p):=\sup_{f\in\mathbb{F}}\varphi_f(g,p)$ を値関数 (Value function) といい，すべての $(g,p)\in\boldsymbol{X}$ に対して $\bar{\varphi}(g,p)=\varphi_{f^*}(g,p)$ を満たす $f^*\in\mathbb{F}$ が存在するとき f^* を最適政策という.

3 解析

補題 3.1 任意の $f \in \mathbb{F}$ に対して，$r(g_t, p_t, f(g_t, p_t))$ は初期状態 $(g_0, p_0) = (g, p) \in \boldsymbol{X}$ の関数として可測である.

証明:$t = 1$ のときを証明する. 任意の $d \in \mathbb{R}^+$ に対して, $\mathbb{C}_d := \{(\theta, x) | r(g_1, p_1, f(g_1, p_1)) \le d\} \in \mathcal{B}(\boldsymbol{X})$ を示せばよい，ただし g_1, p_1 はそれぞれ (2.1) と (2.2) によって与えられる. 各 $a \in A$ に対して $\mathrm{Ch}(\{(\theta, x) | \tilde{r}(\theta, x, f(g_1, p_1)) \ge t\} | g, p)$ は (1.3)-(1.5), (2.1),(2.2) および $\Theta \times X$ が可算集合であることから明らかに $(g, p, t) \in \boldsymbol{X} \times \mathbb{R}^+$ の可測関数である. 従って (2.3) より $r(g_1, p_1, a)$ は $(g, p) \in \boldsymbol{X}$ の可測関数である. $\mathbb{C}_d$ は

$$\mathbb{C}_d = \bigcup_{a \in A} \{(g, p) | r(g_1, p_1, a) \le d\} \bigcap (T_{\boldsymbol{q}}(\cdot|a), T_{\boldsymbol{p}}(\cdot|a))^{-1}(\mathcal{B}_a)$$

と表される. A は有限集合なので (2.5) より $\mathbb{C}_d \in \mathcal{B}_a$ が示された.

$t = 0$ および $t \ge 2$ の場合も同様に $r(g_t, p_t, f(g_t, p_t))$ の可測性が示される. □

定理 3.1 任意の $f \in \mathbb{F}$ に対して，割引された総利得 $\varphi_f(g, p)$ は $(g, p) \in \boldsymbol{X}$ の関数として可測である.

証明：補題 3.1 から明らかである. □

$\bar{\boldsymbol{X}}$ を $\boldsymbol{X}$ 上の有界な可測関数の全体とし，$\varphi, \varphi' \in \bar{\boldsymbol{X}}$ に対して，$\bar{\boldsymbol{X}}$ 上の距離 $\bar{\rho}$ を下記の通り定義する.

$$\bar{\rho}(\varphi, \varphi') = \sup_{(g,p) \in \boldsymbol{X}} |\varphi(g, p) - \varphi'(g, p)|. \tag{3.1}$$

明らかに $(\bar{\boldsymbol{X}}, \bar{\rho})$ は完備である.

任意の $f \in \mathbb{F}$ に対して $\bar{\boldsymbol{X}}$ 上の作用素 U_f を次で定義する.

$$U_f \varphi(g, p) = r(g, p, f(g, p)) + \beta \varphi(T_{\boldsymbol{q}}(g | f(g, p)), T_{\boldsymbol{p}}(p | f(g, p))). \tag{3.2}$$

補題 3.2 U_f は $\bar{\boldsymbol{X}}$ 上の縮小写像である.

証明：任意の (g, p) に対して

$$\begin{aligned}
&|U_f \varphi(g, p) - U_f \varphi'(g, p)| \\
&= \beta |\varphi(T_{\boldsymbol{q}}(g | f(g, p)), T_{\boldsymbol{p}}(p | f(g, p))) - \varphi'(T_{\boldsymbol{q}}(g | f(g, p)), T_{\boldsymbol{p}}(p | f(g, p)))| \\
&\le \beta \sup_{(g,p) \in \boldsymbol{X}} |\varphi(g, p) - \varphi'(g, p)| = \beta \bar{\rho}(\varphi, \varphi')
\end{aligned}$$

従って，$\bar{\rho}(U_f\varphi, U_f\varphi') \leq \beta\bar{\rho}(\varphi, \varphi')$. □

定理 3.2 φ_f は U_f の唯一の不動点として与えられる．つまり，

$$\varphi_f = U_f\varphi_f. \tag{3.3}$$

証明：r は非負の有界な関数だから $M > 0$ が存在して $0 \leq r \leq M$ を満たす．(2.6) よりすべての $(g,p) \in \boldsymbol{X}$ について $0 \leq \varphi_f \leq M/(1-\beta)$，故に $\varphi_f \in \bar{\boldsymbol{X}}$.

$$\begin{aligned}\varphi_f(g,p) &= r(g_0,p_0,f(g_0,p_0)) + \sum_{t=1}^{\infty}\beta^t r(g_t,p_t,f(g_t,p_t))\\ &= r(g_0,p_0,f(g_0,p_0)) + \beta\sum_{t=1}^{\infty}\beta^{t-1}r(g_t,p_t,f(g_t,p_t))\\ &= r(g_0,p_0,f(g_0,p_0)) + \beta\varphi_f(g_1,p_1)\end{aligned}$$

上の式から $\varphi_f = U_f\varphi_f$ がいえる． □

$\bar{\boldsymbol{X}}$ 上の作用素 $U\varphi$ を下記の通り定義する．

$$U\varphi(g,p) = \max_{a\in A}\{r(g,p,a) + \alpha\varphi(T_{\boldsymbol{q}}(g|a), T_{\boldsymbol{p}}(p|a))\},\quad (g,p) \in \boldsymbol{X},\quad \varphi \in \bar{\boldsymbol{X}}. \tag{3.4}$$

補題 3.3 U は $\bar{\boldsymbol{X}}$ 上の縮小写像である．

証明：$\bar{\boldsymbol{X}}$ と U の定義から明らかに U は $\bar{\boldsymbol{X}}$ から $\bar{\boldsymbol{X}}$ への写像である．また，任意の $(g,p) \in \boldsymbol{X}$ に対して，任意の $\varphi, \varphi' \in \bar{\boldsymbol{X}}$ が

$$\begin{aligned}|U\varphi(g,p) - U\varphi'(g,p)| &\leq \alpha\max_{a\in A}|\varphi(T_{\boldsymbol{q}}(g|a), T_{\boldsymbol{p}}(p|a)) - \varphi'(T_{\boldsymbol{q}}(g|a), T_{\boldsymbol{p}}(p|a))|\\ &\leq \alpha\sup_{(g,p)\in\boldsymbol{X}}|\varphi(g,p) - \varphi'(g,p)|\\ &= \alpha\bar{\rho}(\varphi,\varphi').\end{aligned}$$

これより $\bar{\rho}(U\varphi, U\varphi') \leq \alpha\bar{\rho}(\varphi,\varphi')$ が成り立つ． □

補題 3.4 値関数 $\bar{\varphi}$ は有界な可測関数である．つまり，$\bar{\varphi} \in \bar{\boldsymbol{X}}$.

証明：定理 3.1 より $\varphi_f \in \bar{\boldsymbol{X}}$. $n \to \infty$ のとき $U^n\varphi_f \to \bar{\varphi}$ (cf.[2, 9]). $U^n\varphi_f$ は可測関数であるから $\bar{\varphi}$ も可測関数である．また，$0 \leq U^n\varphi_f \leq M/(1-\beta)$ $(n \geq 1)$ より $0 \leq \bar{\varphi} \leq M/(1-\beta)$ $(n \geq 1)$. 以上より $\bar{\varphi} \in \bar{\boldsymbol{X}}$ となる． □

定理 3.3 次の (i) – (iii) が成り立つ.

(i) 値関数 $\bar{\varphi}$ は U の唯一の不動点である.
(ii) 最適政策 $f^* \in \mathbb{F}$ が存在する.
(iii) $U\bar{\varphi} = U_f\bar{\varphi}$ を満たす $f \in \mathbb{F}$ は最適政策である.

証明：(i) $\bar{\boldsymbol{X}}$ はバナッハ空間であり，任意の $(g,p) \in \boldsymbol{X}$ に対して

$$\begin{aligned}
\varphi_f(g,p) &= \sum_{t=0}^{\infty} \beta^t r(g_t, p_t, f(g,p)) \\
&= r(g_0, p_0, f(g_0,p_0)) + \sum_{t=1}^{\infty} \beta^t r(g_t, p_t, f(g_t,p_t)) \\
&\le r(g_0, p_0, f(g_0,p_0)) + \beta\varphi_f(g_1,p_1) \\
&\le r(g_0, p_0, f(g_0,p_0)) + \beta\bar{\varphi}(g,p) \\
&\le \max_{a\in A}[r(g_0,p_0,a) + \beta\bar{\varphi}(g_1,p_1)] \\
&= U\bar{\varphi}(g,p).
\end{aligned}$$

したがって，$\sup_{f\in\mathbb{F}} \varphi_f(g,p) = \bar{\varphi}(g,p) \le U\bar{\varphi}$.

次は，$\bar{\varphi} \ge U\bar{\varphi}$ を証明する．A が有限なので次を満たす $f^* \in \mathbb{F}$ が存在する

$$U\bar{\varphi}(g,p) = r(g,p,f^*(g,p)) + \beta\bar{\varphi}(g_1,p_1),\quad (g,p) \in \boldsymbol{X}.$$

$\bar{\varphi} \le U\bar{\varphi}$ より

$$U\bar{\varphi}(g,p) \le r(g,p,f^*(g,p)) + \beta U\bar{\varphi}(g_1,p_1),\quad (g,p) \in \boldsymbol{X}.$$

これを繰り返して

$$U\bar{\varphi}(g,p) \le \sum_{t=0}^{n} \beta^t r(g_t,p_t,f^*(g_t,p_t)) + \beta^{n+1}\bar{\varphi}(g_{t+1},p_{t+1}).$$

$n \to \infty$ より $U\bar{\varphi} \le \varphi_{f^*} \le \bar{\varphi}$．すなわち，$\bar{\varphi} = U\bar{\varphi}$ を得る

(ii) と (iii) については (i) の証明の中ですでに示されている.

□

参考文献

[1] J.L.Doob, Stochastic Processes, Jhon Wiley, New York, 1953.

[2] M. O. Hernandez-Lerma, R. Montes-de-Oca and R. Cavazos-Cadena, Recurrence conditions for Markov decision processes with Borel state space: A survey, Ann. Oper. Res. 28 (1991), 29-46.

[3] M. Kurano, M. Yasuda, J. Nakagami and Y. Yoshida, A limit theorem in some dynamic fuzzy systems. Fuzzy Sets and Systems 51 (1992), pp. 83-88.

[4] M. Kageyama and K. Iwamura, Discrete time credibilistic processes: Construction and convergences, Inform. Sci., 179, 4277-4283, 2009.

[5] M. Kageyama, B. Yang and P. Hou. Discrete-time hybrid processes and discounted total expected values, Fuzzy Optimization Decision Making (2011) 10:341-355.

[6] X.Li and B. Liu, Chance measure for hybrid events with fuzziness and randomness, Sobt Computing, Vol.13, No.2, 2009.

[7] B. Liu, Uncertain Theory, second ed., springer-Verlag, Berlin, 2007

[8] B.Liu, Uncertain Theory, third ed., UTLAB, 2009.

[9] M.Puterman, Markov Decision Processes, John Wiley and Sons, INC, 1994.