日本オペレーションズ・リサーチ学会
2004年秋季研究発表会

部分観測可能なマルコフ過程における多段決定問題とTP_2

01402656　九州大学大学院経済学研究院　中井　達 NAKAI Tōru

1　はじめに

TP_2は、多段決定問題の性質を解析する上で基本的な性質であり、確率的逐次割り当て問題、最適選択問題、ジョブ・サーチ、取り替え問題など、多くの分野で応用されている。とくに、不完備情報の多段決定問題を、部分観測可能なマルコフ過程における決定問題で、ベイズ学習を考えるとき、重要な役割を果たすことは知られている。ここでは、Lippman and MacCall [1] で扱われたジョブ・サーチを考える。これは、仕事の賃金が、いくつかのクラスに分けている経済の状態に依存するモデルであり、ここでは状態が未知の場合を扱う。そのため、TP_2を用いた仮定の下で、最適政策やその政策にしたがったときに得られる期待利得などの性質を考える。とくに、学習過程の性質や、状態の推移確率に関する性質などについても考える (詳細は中井 [2]・Nakai [3] など)。最後に、TP_2を用いた多段決定問題への応用として、投資問題を考える。

2　ジョブ・サーチ

状態空間を$[0,S]$とし、任意の状態sに対して、状態空間上の確率分布の確率密度関数を$\boldsymbol{p}_s = (p_s(t))$とし、これらの$\boldsymbol{p}_s$は、状態が$s$のときのマルコフ過程の推移法則を表し、$\boldsymbol{P} = (p_s(t))_{s,t\in[0,S]}$とする。状態$s$が経済の状態を表すとすれば、この状態に依存する賃金を表す確率変数をX_sとする。ジョブ・サーチとは、期待賃金を最大にする最適政策を求めることである。いま、ある人が仕事を探していて、費用cを支払って一つの仕事が紹介され、最大でm個の仕事が出現するまで続けることができる。また、リコールはないとする。

また、定義1において、全順序$\geq$が定義された完備で可分な距離空間上の確率変数のあいだに、尤度比を用いてTP_2と呼ばれる確率的な順序関係を導入する。また、推移法則と確率変数X_sの分布に関して、2つの仮定を設ける。

定義 1 確率変数XとYが、それぞれ密度関数$f(x)$と$g(x)$を持ち、$x \geq y$となる任意のxとyに対して$f(y)g(x) \leq f(x)g(y)$であれば、XはYより大きいといい、$X \succeq Y$と表す。

定義 2 関数$\boldsymbol{P} = (p_s(t))$が、$s \leq t$および$u \leq v$となる任意の$s,t,u$と$v$に対して、$p_s(u)\,p_t(v) \geq p_t(u)\,p_s(v)$のとき、この$\boldsymbol{P}$を$TP_2$という。

仮定 1 確率変数$\{X_s\}$に対して、$s \leq t$ならば$X_s \succeq X_t$である。

仮定 2 推移法則$\boldsymbol{P}$はTP_2である。

状態に関する情報は、状態空間$[0,S]$上の確率分布$\boldsymbol{\mu}$で表され、$\mathcal{S}$を情報全体の集合とする。また、情報のあいだに、定義1による順序関係を導入する。ところで、確率変数$\{X_s\}$が未知の状態に依存するから、これらの確率変数を観測することを情報過程と考える。いま、事前情報が$\boldsymbol{\mu}$で、直面する仕事の賃金がxのとき、状態についての情報を$\boldsymbol{\mu}(x) = (\mu(x,s))$と改良し、そのあと状態が推移し、新しい状態へ移ると考える。このとき、つぎの決定時点における事前情報を$\overline{\boldsymbol{\mu}(x)} = (\overline{\mu(x,s)})$とする。

つぎに、マルコフ過程の未知の状態に関する事前情報を$\boldsymbol{\mu}$とする。いま、n個の仕事が残っていて、直面している仕事の賃金がxのとき、$v_n(\boldsymbol{\mu},x)$を、最適政策を用いたときのβで割引された総期待利得とする$(0<\beta<1)$。最適性の原理より、この$v_n(\boldsymbol{\mu},x)$は次の再帰方程式を満足する。

$$v_n(\boldsymbol{\mu},x)=\max\{u_n(x),\ c+\beta\int_0^\infty v_{n-1}(\overline{\boldsymbol{\mu}(x)},y)dF_{\overline{\boldsymbol{\mu}(x)}}(y)\}$$

ただし、$v_1(\boldsymbol{\mu},x) = \mathrm{E}_{\boldsymbol{\mu}}[u_1(X)]$とする。ここで$S(\boldsymbol{\mu},n)$と$C(\boldsymbol{\mu},n)$を、それぞれこのジョブ・サー

チにおける停止領域と継続領域とすれば、これらの領域に関して、次の性質が得られる。

補題 1 $\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば、$S(\boldsymbol{\nu}, n) \subset S(\boldsymbol{\mu}, n)$ および $S(\boldsymbol{\mu}, n+1) \subset S(\boldsymbol{\mu}, n)$ である。

3 状態への推移確率

つぎに、n 期間後の状態の確率分布を考える。はじめに、これらの確率を決定と未知の状態に関する学習過程を除いて考える。未知の状態に関する事前情報が $\boldsymbol{\mu}$ のとき、$\overline{\boldsymbol{P}}_m(\boldsymbol{\mu})$ を m 期間後の状態を表す確率変数の確率密度とする。また、$\boldsymbol{\mu} = (\mu(s))$ と $\boldsymbol{P}$ に対して、$\langle \boldsymbol{\mu}, \boldsymbol{P} \rangle = (\langle \boldsymbol{\mu}, \boldsymbol{P} \rangle(t))$ とする ($\langle \boldsymbol{\mu}, \boldsymbol{P} \rangle(t) = \int_0^S \mu(s) p_s(t) ds$)。このとき、$\overline{\boldsymbol{P}}(\boldsymbol{\mu}, m)$ は $\overline{\boldsymbol{P}}_{\boldsymbol{\mu},m} = \langle \boldsymbol{\mu}, \boldsymbol{P}^m \rangle$ となり、$\boldsymbol{P}$ が TP_2 だから、$\overline{\boldsymbol{P}}_m(\boldsymbol{\mu}) = \langle \boldsymbol{\mu}, \boldsymbol{P}^m \rangle$ もまた TP_2 であり、$\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば、$\overline{\boldsymbol{P}}_{\boldsymbol{\mu},m} \succeq \overline{\boldsymbol{P}}_{\boldsymbol{\nu},m}$ である。

つぎに、決定を除いて考える。状態に関する事前情報が $\boldsymbol{\mu}$ のとき、$\widehat{P}_{\boldsymbol{\mu},m}(t)$ を m 期間後の状態の確率密度とする。このとき、$\widehat{\boldsymbol{P}}_{\boldsymbol{\mu},m}$ は (1) 式を満足し、つぎの性質を持つ。

$$\widehat{\boldsymbol{P}}_{\boldsymbol{\mu},m} = \int_0^\infty \widehat{\boldsymbol{P}}_{\overline{\boldsymbol{\mu}(x)},m-1} dF_{\boldsymbol{\mu}}(x), \qquad (1)$$

ここで、$\widehat{\boldsymbol{P}}_{\boldsymbol{\mu},1} = \int_0^\infty \overline{\mu(x)} dF_{\boldsymbol{\mu}}(x)$ とする。

命題 1 $\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば、$\widehat{\boldsymbol{P}}_{\boldsymbol{\mu},m}$ は $\boldsymbol{\mu}$ の増加関数である。すなわち、$\widehat{\boldsymbol{P}}_{\boldsymbol{\mu},m} \succeq \widehat{\boldsymbol{P}}_{\boldsymbol{\nu},m}$ である。

最後に、同様の確率を、決定と学習過程を含めて考える。いま、$(\widetilde{P}_{\boldsymbol{\mu},n,m}(t))$ を最適政策にしたがったときの m 期間後の状態を表す確率変数の確率密度とする。このとき、新しい仕事が現れ、その賃金をもとに決定する。$x \in C(\boldsymbol{\mu}, n)$ のとき、つぎの仕事へと進むから、$\widetilde{\boldsymbol{P}}_{\boldsymbol{\mu},n,m} = \left(\widetilde{P}_{\boldsymbol{\mu},n,m}(t)\right)$ は再帰方程式

$$\widetilde{P}_{\boldsymbol{\mu},n,m}(t) = \int_{C(\boldsymbol{\mu},n)} \widetilde{P}_{\overline{\boldsymbol{\mu}(x)},n-1,m-1}(t) dF_{\boldsymbol{\mu}}(x)$$

を満たす。ここで、$\int_0^S \widetilde{P}_{\boldsymbol{\mu},n,m}(t) dt \leq 1$ であることは明らかである。さらに、$\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば $C(\boldsymbol{\mu}, n) \subset C(\boldsymbol{\nu}, n)$ となる。すなわち、見送ってつぎの仕事を探す確率は、$\boldsymbol{\mu}$ が増加するにしたがって減少する。いっぽう、より悪い状態へ推移する確率は、$\boldsymbol{\mu}$ が増加すれば増加する。したがって、この確率 $(\widetilde{P}_{\overline{\boldsymbol{\mu}(x)},n-1,m-1}(t))$ が観測した x によって変化するので、この場合には $\widetilde{\boldsymbol{P}}_{\boldsymbol{\mu},n,m}$ の性質をみることは難しい。

4 投資問題

消防や警察などの公的な部門に資源を投入することを考える。このような公的部門に対して、サービスに対する満足度は $[0, 1]$ 区間に含まれる値 s で表せるとする。一方、要求に応えるために新たな設備や人員の増加をしたとしても、おかれている状況が変化すれば要求がさらに大きくなり満足度が減少することも考えられる。このために、満足度を表す値を状態と考え、この状態は新たな設備や人員の増加することによっても変化するが、環境が変化するなど制御できない要素によっても変化するものとする。

一方、これらのサービスを実現するために資金 x を使って設備や人員を配置すれば、満足度が関数 $s(x)$ で表されるとする。この関数 $s(x)$ は (1)$s(x)$ は x に関する増加関数であり、(2)$s(x)$ は x に関して凹関数とする。さらに、満足度が状態を表すと考えたとき、これらの状態がマルコフ過程にしたがって推移するとする。また、推移法則 $(p_s(t))_{0 \leq s \leq 1}$ は TP_2 とする。このとき、残りの計画期間が n のとき、予算 K の範囲内で資本を投下して、最適に振る舞ったときに得られる満足度を $V_n(s)$ とすれば、最適方程式は

$$\begin{aligned} V_n(s) &= \max_{x \geq 0} \{-c(s, s+x(s)) \\ &\quad + \int_0^1 p_{s+x(s)}(t) V_{n-1}(t) dt\} \end{aligned}$$

となる。このときに $V_n(s)$ と最適な投資額を $x_n^*(s)$ に関して基本的な性質が成り立つ。

参考文献

[1] S. A. Lippman and J. J. McCall, (1976). *J. Econ. Theory*, **12**, 365–390.

[2] 中井　達, 京都大学数理解析研究所講究録「不確実性と意思決定数理の諸問題」, 2004.

[3] T. Nakai, *Sci. Math. Jap. Online*, **10**, 219–230, 2004.