数理解析研究所講究録
第1857巻 2013年 109-119

投資モデルに基づく逐次決定問題について

千葉大学教育学部　中井　達 (Tōru Nakai)
Faculty of Education, Chiba University

1　はじめに

Nakai(2009) などにおいて、評価と関連する状態をもとに、予算の範囲内で支出を決定する逐次決定問題を部分観測可能なマルコフ決定過程の一つとして考えた。このとき、状態は支出によって変化する。ここでは、この問題を一般化することを考える。

消防活動や警察活動といった公共サービスに対する支出を、毎年度の予算の範囲内で行うとし、これらの公共サービスに対する充足度合を1つの指標でとらえるとする。この指標を状態と考え、この状態は確率的に推移するとともに、予算から支出することによっても状態が変化する。このモデルを状態空間が $(-\infty,\infty)$ のマルコフ決定過程と考え、この状態に関する情報を、状態空間 $(-\infty,\infty)$ 上の確率変数で表す。この状態 $s\in(-\infty,\infty)$ が大きくなれば、このサービスにより満足を感じることを示し、小さくなれば満足を感じなくなるとする。

状態が s のとき、決定 x を取れば、状態を $\sigma(s,x)$ とでき $(x\geq 0)$、この決定に対する費用を $C(x)$ とする。さらに、決定によって変化した状態はこの決定にかかわらず、マルコフ過程にしたがって推移する。このとき、計画期間内で利得 (効用) を最大化する最適政策と最適政策にしたがったときに得られる最適値を考える。

2　Stocahstic Convexity and Concavity

X と Y を2つの確率変数とする。つぎのような確率的な順序関係を考える。

定義 1 確率密度関数 $f_X(x)$ と $f_Y(x)$ を持つ2つの確率変数 X と Y に対して、$x\geq y$ となる任意の x と y に対して、$f_X(y)f_Y(x)\leq f_X(x)f_Y(y)$ であるとき、X は Y より尤度比の意味で大きいといい、$X\geq_{LRD}Y$ あるいは $X\succeq Y$ と表す。

この定義を用いて導入される確率変数のあいだの順序が半順序であることは、簡単に示すことができる。

Shaked and Shanthikumar(1994) にしたがって、Stocahstic Convexity と Stocahstic Concavity を考える。$\{X(s)|s\in(-\infty,\infty)\}$ を s をパラメータとする確率変数列とする。

(1) $\{X(s)|s\in(-\infty,\infty)\}$ が SI(stocahstically increasing)$\Longleftrightarrow$ 任意の増加 関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の増加 関数である。

(2) $\{X(s)|s \in (-\infty,\infty)\}$ がSICX(stocahstically increasing and convex)$\Longleftrightarrow$ 任意の増加凸関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の増加凸 関数である。

(3) $\{X(s)|s \in (-\infty,\infty)\}$ がSICV(stocahstically increasing and concave)$\Longleftrightarrow$ 任意の増加凹関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の増加凹関数である。

つぎに、$s_1 \leq s_2 \leq s_3 \leq s_4$ で $s_1 + s_4 = s_3 + s_2$ のとき、$X_i = X(s_i)$ とおく $(i = 1,2,3,4)$。$(s_4 - s_3 = s_2 - s_1)$

(1) $\{X(s)|s \in (-\infty,\infty)\}$ がSICX(sp)(stocahstically increasing and convex in sample path sense)$\Longleftrightarrow \max\{X_2, X_3\} \leq X_4$ であり (a.s.)、$X_2+X_3 \leq X_1+X_4$ である。

(2) $\{X(s)|s \in (-\infty,\infty)\}$ がSICV(sp)(stocahstically increasing and concave in sample path sense)$\Longleftrightarrow X_1 \leq \max\{X_2, X_3\}$ であり (a.s.)、$X_2+X_3 \geq X_1+X_4$ である。

補題 1 (1) $\{X(s)|s \in (-\infty,\infty)\}$ が *SICX(sp)* ならば、*SICX* である。
(2) $\{X(s)|s \in (-\infty,\infty)\}$ が *SICV(sp)* ならば、*SICV* である。

補題 2 (1) $\{X(s)|s \in (-\infty,\infty)\}$ が *SICX(sp)* であり、$u(\cdot)$ を増加凸関数 とする。このとき、$\{u(X(s))|s \in (-\infty,\infty)\}$ もまた *SICX(sp)* である。
(2) $\{X(s)|s \in (-\infty,\infty)\}$ が *SICV(sp)* であり、$u(\cdot)$ を増加凹関数 とする。このとき、$\{u(X(s))|s \in (-\infty,\infty)\}$ もまた *SICV(sp)* である。

例 1 $X(\mu)$ を正規分布 $N(\mu,\sigma^2)$ とする。$Y(\mu) = e^{X(\mu)}$ とおけば、$u(x) = e^x$ が増加凸関数 だから $\{Y(\mu)|\mu \in (-\infty,\infty)\}$ は *SICX(sp)* である。したがって、$Y(\mu)$ は対数正規分布であり、*SICX(sp)* であり、*SICX* である。

3 投資モデルに基づく逐次決定モデル

消防活動や警察活動といった公共サービスに対する支出を、年度ごとの予算の範囲内で行う。これらの公共サービスに対する充足度合を $(-\infty,\infty)$ に含まれる実数で表し、この値が大きくなれば、このサービスにより満足を感じることを示し、小さくなれば満足を感じなくなるとする。ここでは、この指標を状態と考え、確率的に推移するとともに、予算から支出することによっても状態を変化できる。

このモデルを状態空間が $(-\infty,\infty)$ のマルコフ過程と考え、状態を表す値 s が大きくなれば充足度合が大きくなるとする。また、この状態は決定を取ることにより変化し、この新しい状態からこのマルコフ過程の推移法則にしたがって推移する。このとき、計画期間内で利得を最大化する最適政策と最適政策にしたがったときに得られる最適値について考える。また、[8] などと同様に、決定により状態を変化できることからマルコフ決定過程の一つと考えることができる。

状態空間を $(-\infty,\infty)$ とし、状態が s のとき決定 x を取れば、状態を $\sigma(s,x)$ とできる $(x\geq 0)$。ただし、$\sigma(s,0)=s$ とする $(s\in(-\infty,\infty))$。このときの決定に対応する費用を $C(x)$ とする $(C(0)=0)$。決定期間が n のとき、$u(s)$ を最後の状態が s のときの終端利得とし、$u(s)$ は凹関数とする。また、決定を取ったあと、推移法則を $\boldsymbol{P}=(p_s(t))_{s,t\in(-\infty,\infty)}$ とするマルコフ過程にしたがって状態が推移する。

はじめに、関数 $\sigma(s,x)$ につぎの仮定を置く。

仮定 1 s と x の関数 $\sigma(s,x)$ は凹関数とする。すなわち、$x<y$ および $s<t$ となる任意の $(t,y),(s,x)$ と $0\leq\lambda\leq 1$ に対して

$$\sigma(\lambda(t,y)+(1-\lambda)(s,x))\geq\lambda\sigma(t,y)+(1-\lambda)\sigma(s,x)$$

とする。

仮定 2 s と x の関数 $\sigma(s,x)$ は submodular function とする。すなわち、$x<y$ および $s<t$ となる任意の x,y と s,t に対して

$$\sigma(t,y)-\sigma(t,x)\leq\sigma(s,y)-\sigma(s,x) \tag{1}$$

とする。

つぎに、関数 $\overline{v}(s)=\max_{x\geq 0}\{-C(x)+u(\sigma(s,x))\}$ の性質を考える。このとき、$u(s)$ が s の増加関数であれば、$\overline{v}(s)$ も増加関数であることは明らかである。

補題 3 $C(x)$ が凸関数のとき、$u(s)$ が凹関数ならば、$\overline{v}(s)$ も凹関数である。ただし、$C(x)$ は増加関数とする。

証明: $\overline{v}(s)=C(x^*)+u(\sigma(s,x^*))$ および $\overline{v}(t)=C(x^{**})+u(\sigma(t,x^{**}))$ とおく。$0\leq\lambda\sigma(s,x^*)+(1-\lambda)\sigma(t,x^{**})\leq\sigma(\lambda s+(1-\lambda)t,\lambda x^*+(1-\lambda)x^{**})$ であり、任意の λ $(0<\lambda<1)$ と $s<t$ に対して $u(\lambda s+(1-\lambda)t)\geq\lambda u(s)+(1-\lambda)u(t)$ だから $C(x)$ に関する仮定を用いて

$$\begin{aligned}
\overline{v}(\lambda s+(1-\lambda)t) &= \max_{x\geq 0}\{-C(x)+u(\sigma(\lambda s+(1-\lambda)t,x)\} \\
&\geq -C(\lambda x^*+(1-\lambda)x^{**}) \\
&\quad +u(\sigma(\lambda s+(1-\lambda)t,\lambda x^*+(1-\lambda)x^{**})) \\
&\geq -C(\lambda x^*+(1-\lambda)x^{**})+u(\lambda\sigma(s,x^*)+(1-\lambda)\sigma(t,x^{**})) \\
&\geq -(\lambda C(x^*)+(1-\lambda)C(x^{**})) \\
&\quad +\lambda u(\sigma(s,x^*))+(1-\lambda)u(\sigma(t,x^{**})) \\
&= \lambda\overline{v}(s)+(1-\lambda)\overline{v}(t)
\end{aligned}$$

となる。したがって、$u(s)$ は凹関数となる。□

状態がマルコフ過程にしたがって推移し、推移法則が $\boldsymbol{P}=(p_s(t))_{s,t\in(-\infty,\infty)}$ のとき、決定を $x\geq 0$ とする。このとき、計画期間が n で、状態が s のとき、利得最大化問題において、最適に振る舞って得られる総期待利得を $u_n(s)$ とすれば、決定 x により状態は $\sigma(s,x)$ となり、この状態からマルコフ過程にしたがって推移するから、最適性の原理より最適方程式はつぎのようになる。ここで、$T(s)$ を状態が s のときつぎの状態を表す確率変数とすれば、$E[u_{n-1}(T(s))]=\int_0^\infty p_s(t)u_{n-1}(t)dt$ である。

$$u_n(s) = \max_{x\geq 0}\{-C(x)+E[u_{n-1}(T(\sigma(s,x)))]\}, \tag{2}$$

ここで、

$$u_1(s)=\max_{x\geq 0}\{-C(x)+E[u(T(\sigma(s,x)))]\}$$

とする。ただし、$u(s)$ は s の増加関数とし、$C(x)$ は x の増加関数とする。

推移法則が $(p_s(t))_{0\leq s\leq 1}$ だから、確率変数列 $\{T(s)|s\in(-\infty,\infty)\}$ に対して、つぎの仮定を設ける。

仮定 3 t に関する増加凹関数を $u(t)$ とすれば、$E[u(T(s))]$ は s に関する増加凹関数となっている。すなわち、確率変数列 $\{T(s)|s\in(-\infty,\infty)\}$ は、$SICV$ である。

補題 4 $u_n(s)$ は、s に関する増加関数である。

証明: n に関する帰納法を用いる。$u_0(s)=u(s)$ だから、$u_0(s)$ は増加凹関数である。$u_{n-1}(s)$ が増加凹関数と仮定すると、仮定 3 より $E[u_{n-1}(T(\sigma(s,x)))]$ もまた s に関する増加関数でである。$E[u_{n-1}(T(\sigma(s,x)))]$ が s の増加関数なので、$u_n(s)=\max_{x\geq 0}\{-C(x)+E[u_{n-1}(T(\sigma(s,x)))]\}$ も s の増加関数である。さらに、補題 3 より、$u_n(s)$ が凹関数となる。□

補題 5 仮定 3 のもとで、$u_n(s)$ は凹関数である。

計画期間が n であり、状態が s のときの、最適な決定を $x_n^*(s)$ とする。

性質 1 仮定 3 のもとで、$x_n^*(s)$ は s に関して減少する。

証明: n に関する帰納法を用いる。$n=1$ の場合は、一般のときと同じように導ける。$t\geq s$ とする。$n(>1)$ のとき、$x_n^*(s)=x^*$ とおけば、(2) 式より

$$\begin{aligned} u_n(s) &= \max_{x\geq 0}\{-C(x)+E[u_{n-1}(T(\sigma(s,x)))]\} \\ &= -C(x^*)+E[u_{n-1}(T(\sigma(s,x^*)))] \end{aligned} \tag{3}$$

となる。$0 < x^* \leq x$ となる任意の x に対して、不等式

$$-C(x) + E[u_{n-1}(T(\sigma(t,x)))] \leq -C(x^*) + E[u_{n-1}(T(\sigma(t,x^*)))]$$

が成り立てば、$x_n^*(s) = x^* \geq x_n^*(t)$ となることが示される。

(3) 式より、任意の $x \geq 0$ に対して

$$-C(x) + E[u_{n-1}(T(\sigma(s,x)))] \leq -C(x^*) + E[u_{n-1}(T(\sigma(s,x^*)))] \qquad (4)$$

だから、

$$-C(x) + C(x^*) \leq E[u_{n-1}(T(\sigma(s,x^*)))] - E[u_{n-1}(T(\sigma(s,x)))] \qquad (5)$$

となる。いっぽう、仮定3より、$E[u_{n-1}(T(\sigma(s,x)))]$ は s に関する増加凹関数である。したがって、仮定2より、$\sigma(t,x^*) - \sigma(s,x^*) \geq \sigma(t,x) - \sigma(s,x)$ である。

いっぽう、$\sigma(t,x) - \sigma(s,x) = \sigma(t,x^*) - (\sigma(t,x^*) - (\sigma(t,x^*) - \sigma(t,x)) + \sigma(s,x))$ であり $0 < x^* \leq x, s < t$ である。このことから

$$\begin{aligned} & E[u_{n-1}(T(\sigma(t,x)))] - E[u_{n-1}(T(\sigma(s,x))) \\ & \leq \quad E[u_{n-1}(T(\sigma(t,x^*)))] - E[u_{n-1}(T(\sigma(s,x^*) + (\sigma(t,x^*) - \sigma(s,x^*) \qquad (6) \\ & \qquad\qquad -(\sigma(t,x) - \sigma(s,x)))))] \\ & \leq \quad E[u_{n-1}(T(\sigma(t,x^*)))] - E[u_{n-1}(T(\sigma(s,x^*)))] \qquad (7) \end{aligned}$$

となる。(5) 式とこの不等式から、(4) 式が導かれ、この性質が成り立つ。□

仮定 4 $t \geq s$ のとき任意の凹関数 $u(s)$ に対して、$E[u(T(t))] - E[u(T(s))] \leq u(t) - u(s)$ である。

$u_n(s)$ が増加凹関数だから、仮定4より $s < t$ なら $E[u_n(T(t))] - E[u_n(T(s))] \leq u_n(t) - u_n(s)$ となる。

$s < t$ のとき、$E[u_n(T(t))] - E[u_n(T(s))]$ と $E[u_{n-1}(T(t))] - E[u_{n-1}(T(s))]$ の関係を考える $(n \geq 1)$。$x^* = x_n^*(t)$ とおけば、

$$\begin{aligned} u_n(t) - u_n(s) \quad & = \\ -C(x^*) \quad & + \quad E[u_{n-1}(T(\sigma(t,x^*)))] - \max_{x \geq 0}\{-C(x) + E[u_{n-1}(T(\sigma(s,x)))]\} \\ & \leq \quad E[u_{n-1}(T(\sigma(t,x^*)))] - E[u_{n-1}(T(\sigma(s,x^*)))] \end{aligned}$$

となる。いっぽう、補題5より、仮定3のもとで $E[u_{n-1}(T(s))]$ は s に関する増加凹関数である。

$s < t, 0 < x^*$ だから、$\sigma(t,x^*) - \sigma(s,x^*) \leq \sigma(t,0) - \sigma(s,0) = t - s$ であり、(7) 式と同じように

$$E[u_{n-1}(T(\sigma(t,x^*)))] - E[u_{n-1}(T(\sigma(s,x^*)))] \leq E[u_{n-1}(T(t))] - E[u_{n-1}(T(s))]$$

となる。これらの不等式から

$$u_n(t) - u_n(s) \leq E[u_{n-1}(T(t))] - E[u_{n-1}(T(s))]$$

となる。

補題5より、$u_n(s)$ が凹関数なので、仮定4より $E[u_n(T(t))] - E[u_n(T(s))] \leq u_n(t) - u_n(s)$ となる。これらの不等式から、任意の $n \geq 1$ に対して

$$E[u_n(T(t))] - E[u_n(T(s))] \leq E[u_{n-1}(T(t))] - E[u_{n-1}(T(s))] \tag{8}$$

となる。

性質 2 仮定4のもとで、$x_n(s)$ は n に関して減少する。

証明: n に関する帰納法を用いる $n = 1$ の場合は、一般のときと同じように導ける。$n(> 1)$ のとする。$s \leq t$ のとき、$x_n^*(s) = x^*$ とおけば、任意の $1 < x^* < x$ に対して、

$$-C(x) + E[u_{n-1}(T(\sigma(s,x)))] \leq -C(x^*) + E[u_{n-1}(T(\sigma(s,x^*)))]$$

である。いっぽう、(8) 式より

$$\begin{aligned} E[u_{n-1}(T(\sigma(s,x^*)))] \quad - \quad & E[u_{n-1}(T(\sigma(s,x)))] \\ & \leq E[u_n(T(\sigma(s,x^*)))] - E[u_n(T(\sigma(s,x)))], \end{aligned}$$

だから、

$$-C(x) + E[u_n(T(\sigma(s,x)))] \leq -C(x^*) + E[u_n(T(\sigma(s,x^*)))] \tag{9}$$

となる。このことから、n に関する帰納法より、$x^* \geq x_{n+1}^*(s)$ となる。□

したがって、任意の $n \geq 1$ に対して、$x_n^*(s) \geq x_{n+1}^*(s)$ なので、仮定4のもとで $x_n^*(s)$ の n に関する単調性が示される。

ところで、最適政策にしたがったときの最適値 $u_n(s)$ の n に関する単調性について考える。基本的に、公的サービスは、将来の満足度や充足度による期待効用が現時点に比べて悪くなったとしても、これらのサービスを打ち切ることはできず、続けて行う必要がある。したがって、状態の関数である効用関数と、推移法則によっては、$u_n(s)$ は n に関して増加することもあれば、減少することも考えられる。ところで、任意の s に対して $u_{n-1}(s) \leq u_{n-2}(s)$ ならば、$E[u_{n-1}(T(\sigma(s,x)))] \leq E[u_{n-2}(T(\sigma(s,x)))]$ となるので、

$$\begin{aligned} u_n(s) &= \max_{x \geq 0} \{-C(x) + E[u_{n-1}(T(\sigma(s,x)))]\} \\ u_{n-1}(s) &= \max_{x \geq 0} \{-C(x) + E[u_{n-2}(T(\sigma(s,x)))]\} \end{aligned}$$

より、$u_n(s) \leq u_{n-1}(s)$ となることがわかる。反対に、任意の s に対して $u_{n-1}(s) \geq u_{n-2}(s)$ ならば、$u_n(s) \geq u_{n-1}(s)$ となる。したがって、帰納法を用いれば、$n=1$ のときの性質によって、$u_n(s)$ の n に関する単調性が定まる。すなわち、$n=1$ のときは、$u_1(s) = \max_{x \geq 0}\{-C(x) + E[u_0(T(\sigma(s,x)))]\}$ であり、$u_0(s) = u(s)$ だから、$\sigma(s,0) = s$ より、$E[u_0(T(s))] \geq u_0(s)$ であれば $u_n(s)$ は n に関する非減少関数であり、$E[u_0(T(s))] \leq u_0(s)$ であれば $u_n(s)$ は n に関する非増加関数となることがわかる。

4 部分観測可能なマルコフ過程と情報

不完備情報のマルコフ過程のとき、この逐次支出モデルの最適政策の単調性を示すことは困難であるが、最適政策にしたがったときの総期待利得の性質について考えることができる。ここでは、Nakai[8] にしたがって簡単に結果をまとめる。また、ベイズの定理にしたがった学習プロセスを考えることから、尤度比順序に基づいて解析する。

つぎに、状態を直接観測できないとする。すなわち、部分観測可能なマルコフ連鎖における多段決定問題を考える。観測できない状態に関する情報は、状態空間 $(-\infty, \infty)$ 上の確率分布 $\boldsymbol{\mu}$ として表す。$\mathcal{S}$ に含まれる情報のあいだに、半順序を $\boldsymbol{\mu} \geq_{LRD} \boldsymbol{\nu}$ によって定義する。いっぽう、$T(t) \geq_{LRD} T(s)$ と仮定する。このとき、補題6が得られる。

補題 6 $\boldsymbol{\mu} \geq_{LRD} \boldsymbol{\nu}$ ならば $(\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{S})$、$x$ の非減少な非負関数 $h(x)$ に対して、$E_{\boldsymbol{\mu}}[h(X)] \geq E_{\boldsymbol{\nu}}[h(X)]$ となる。

状態 s 対して、この状態に依存する確率変数 Y_s を情報プロセスとする。すなわち、それぞれの状態に関する情報を確率変数 Y_s を通して得ることができる観測過程とする。学習プロセスはベイズ学習にしたがって解析することから、仮定5を設ける。この仮定は、Nakai [6] にしたがって一般化でき、多段決定問題へ応用できる。

仮定 5 $s \leq t$ ならば、$Y_t \geq_{LRD} Y_s$ である $(s, t \in (-\infty, \infty))$。

仮定5において、$Y_s \geq_{LRD} Y_t$ としたから、確率変数 Y_s は s の値が小さくなるにしたがって小さな値をとり、s が大きくなるにしたがって良くなる。推移法則に関する仮定から、状態を表す s が大きくなれば、より良い状態に推移する確率は大きくなるのである。

確率過程の観測できない状態に関して、確率変数 $\{Y_s\}_{s \in (-\infty,\infty)}$ を観測することによって、状態に関してベイズの定理を用いて学習を行う。その後、状態は推移し新しい状態になると考える。もちろん、この順序を変えても同じように解析できる。y を観測したとき、事後情報を $\boldsymbol{\mu}_y$ とし、その後で推移法則 $\boldsymbol{P}$ にしたがって状態が推移し、つぎの新しい状態に関する情報は $\overline{\boldsymbol{\mu}_y}$ となる。

このとき、集合値関数 $\boldsymbol{h}(y,s)$ に対して、定義2によって単調性を定義する。

定義 2 任意の $s,x \in \Re$ に関する非負の集合値関数 $\boldsymbol{h}(x) = (h(x,s))_{s\in(-\infty,\infty)}$ に対して、任意の t と $s(s \leq t$ かつ $s,t \in (-\infty,\infty))$ について、$x < y$ ならば $\boldsymbol{h}(y) \geq_{LRD} \boldsymbol{h}(x)$ $(\boldsymbol{h}(x) \geq_{LRD} \boldsymbol{h}(y))$ とする。このとき、関数 $\boldsymbol{h}(x,s)$ を x に関する増加関数 (減少関数) という。

事前情報が $\boldsymbol{\mu}$ のとき、推移後の事後情報を $\overline{\boldsymbol{\mu}}$ とする。事前情報 $\boldsymbol{\mu}$ と事後情報 $\overline{\boldsymbol{\mu}(x)}$ のあいだには、つぎの基本的な性質が成り立つ (Nakai [6] など)。

補題 7 $\boldsymbol{\mu} \geq_{LRD} \boldsymbol{\nu}$ ならば、任意の y に対して、$\boldsymbol{\mu}_y \geq_{LRD} \boldsymbol{\nu}_y$ および $\overline{\boldsymbol{\mu}_y} \geq_{LRD} \overline{\boldsymbol{\nu}_y}$ である。任意の $\boldsymbol{\mu}$ に対して、$\boldsymbol{\mu}_y$ と $\overline{\boldsymbol{\mu}_y}$ は y に関する増加関数である。

補題7から、事前情報 $\boldsymbol{\mu}$ における順序関係は、$\boldsymbol{\mu}_y$ と事後情報 $\overline{\boldsymbol{\mu}_y}$ に対して保たれる。さらに、同じ事前情報 $\boldsymbol{\mu}$ であれば、観測した値 y が大きくなれば、事後情報 $\overline{\boldsymbol{\mu}_y}$ もまたよくなる。

不完備情報の最適決定問題を考えるために、いくつかの準備をする。ここで、

$$\boldsymbol{\mu}^x(t) = \int_0^\infty \mu(s) p_{\sigma(s,x)}(t) ds \tag{10}$$

とおく。以下では、$\sigma(s,x) = s + \sigma(x)$ と表される場合を考える。このとき、(10) 式は、事前情報が $\boldsymbol{\mu}$ のとき、決定 x をとったときの、状態空間上の事後分布を表す。また、$\overline{\boldsymbol{\mu}} = \boldsymbol{\mu}^0$ である。

状態全体の集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ が

> $s < t, t < t'$ と $s - t = t - t' = c < 0$ を満たす任意の $s < t$, $t \leq t'$ に対して、
>
> $$\frac{\mu(s)}{\mu(t)} \geq \frac{\mu(t)}{\mu(t')}$$

となるとき、この $\boldsymbol{\mu}$ は性質 (G) を満たすということにする。

例 2 状態空間上の正規分布 $\mu(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-a)^2}{2\sigma^2}}$ はこの性質を満足する。

補題 8 $\boldsymbol{\mu} \in \mathcal{S}$ が性質 (G) を満たすとき、$x > x'$ ならば、$\boldsymbol{\mu}^x \geq_{LRD} \boldsymbol{\mu}^{x'}$ である。ただし、$\boldsymbol{\mu}^x = (\mu^x(t))$ とする。

補題 9 状態全体の集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ と $\boldsymbol{\nu}$ が性質 (G) を満たすとき、$\boldsymbol{\mu} \geq_{LRD} \boldsymbol{\nu}$ ならば、任意の $x(\geq 0)$ に対して、$\boldsymbol{\mu}^x \geq_{LRD} \boldsymbol{\nu}^x$ である。

仮定 6 任意の $s < t, t \leq t'$ および $u < v$ となる s,t,t',u,v に対して $p_u(s)p_v(t') - p_u(t)p_v(t) \geq p_v(s)p_u(t') - p_v(t)p_u(t)$ とする。

補題 10 $\boldsymbol{\mu} \in \mathcal{S}$ が性質 (G) を満たすならば、$\overline{\boldsymbol{\mu}}$ もまた性質 (G) を満たす。

例 3 正規分布による推移法則 $p_v(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-v)^2}{2\sigma^2}}$ は、仮定 6 の条件を満足する。

確率変数 Y_s の密度関数 $f_s(y)$ が $(s \in (-\infty, \infty))$、任意の $s < t, t < t'$ で $t - s = t' - t > 0$ となる s, t, t, t' に対して、性質

$$\frac{f_s(y)}{f_t(y)} \geq \frac{f_t(y)}{f_{t'}(y)}$$

が成り立つと仮定する。

補題 11 $\boldsymbol{\mu} \in \mathcal{S}$ が性質 (G) を満たすならば、任意の y に対して $\overline{\boldsymbol{\mu}_y}$ もまた性質 (G) を満たす。

補題 12 $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{S}$ が性質 (G) を満たすとする。$\overline{\boldsymbol{\mu}^x}$ もまた性質 (G) を満たす。$\boldsymbol{\mu} \geq_{LRD} \boldsymbol{\nu}$ ならば、任意の $x(\geq 0)$ に対して $\overline{\boldsymbol{\mu}^x} \geq_{LRD} \overline{\boldsymbol{\nu}^x}$ である。$x > x'$ ならば $\overline{\boldsymbol{\mu}^x} \geq_{LRD} \overline{\boldsymbol{\mu}^{x'}}$ である。

5 部分観測可能なマルコフ決定過程としての逐次支出モデル

不完備情報のマルコフ決定過程として逐次支出モデルを考え、最適政策にしたがったときの総期待利得の性質について考える。状態に関する情報は、状態空間上の確率分布として表され、状態に関する情報は、情報プロセスを通して得られる。事前情報と事後情報の関係などについては、4 節の部分観測可能なマルコフ過程に関する性質を用いることが出来る。したがって、部分観測可能なマルコフ決定過程として定式化するとき、以下では $\sigma(s, x) = s + \sigma(x)$ と表されることを仮定する。

4 節と同様に、観測できない状態に関する情報は、状態空間上の確率分布として表され、情報プロセスから得られた観測値をもとにベイズの定理にしたがって学習を行う。情報プロセスは、それぞれの状態 s $(\in (-\infty, \infty))$ に対して、確率変数 Y_s を考え、それらの値を観測することを観測過程とする。観測できない状態に関する情報が $\boldsymbol{\mu}$ で、計画期間が n のとき、最適政策にしたがって得られる総期待利得を $v_n(\boldsymbol{\mu})$ とすれば、最適性の原理より、つぎの再帰方程式が得られる。

$$\begin{aligned} v_n(\boldsymbol{\mu}) &= E[v_n(\boldsymbol{\mu}|Y)] \\ v_n(\boldsymbol{\mu}|y) &= \max_{x \geq 0} \left\{ -c(x) + v_{n-1}(\overline{\boldsymbol{\mu}_y^{\sigma(x)}}) \right\} \end{aligned} \tag{11}$$

ここで、$v_0(\boldsymbol{\mu}) = E[u(S)]$ とする。事前情報が $\boldsymbol{\mu}$ のとき、まず始めに状態 s に依存する確率変数 Y_s を観測し、y が得られたとき状態に関する情報をベイズの定理に

したがって $\boldsymbol{\mu}_y$ と改良する。状態が s のとき、決定 x を取れば状態は $s+\sigma(x)$ となるので、情報は $\boldsymbol{\mu}_y^{\sigma(x)}$ となる。そのあと、この状態から推移法則 $(p_{s+\sigma(x)}(t))_{0\leq s\leq 1}$ にしたがって状態が推移し1期間先に進む。この確率過程の状態は新しい状態となり、この新しい状態に関する情報は $\overline{\boldsymbol{\mu}_y^{\sigma(x)}}$ となる。それ以降、最適政策にしたがって得られる残り計画期間での総期待利得は $v_{n-1}(\overline{\boldsymbol{\mu}_y^{\sigma(x)}})$ である。したがって、n に関する帰納法を用いることにより、3節の仮定の下でつぎの性質が得られる。.

性質 3 $\boldsymbol{\mu},\boldsymbol{\nu}\in\mathcal{S}$ が性質 (G) を満たすとき、$\boldsymbol{\mu}\geq_{LRD}\boldsymbol{\nu}$ ならば、$v_n(\boldsymbol{\mu})\geq v_n(\boldsymbol{\nu})$ である。

$\boldsymbol{\mu}\geq_{LRD}\boldsymbol{\nu}$ であれば、補題7より観測値 y に対して、$\boldsymbol{\mu}_y\geq_{LRD}\boldsymbol{\nu}_y$ であり、補題9から、決定 x に対して、$\overline{\boldsymbol{\mu}_y^{\sigma(x)}}\geq_{LRD}\overline{\boldsymbol{\nu}_y^{\sigma(x)}}$ である。これらの事後情報に関する単調性から、任意の決定 x と観測値 y に対して、$\boldsymbol{\mu}\geq_{LRD}\boldsymbol{\nu}$ ならば、$\overline{\boldsymbol{\mu}_y^{\sigma(x)}}\geq_{LRD}\overline{\boldsymbol{\nu}_y^{\sigma(x)}}$ であり、このことから性質3が n に関する帰納法によって示される。

このように、不完備情報のマルコフ過程における逐次支出モデルにおいて、最適政策にしたがったときの総期待利得に関する単調性を求めることが出来る。しかし、性質1や2と同様の性質が、不完備情報のマルコフ決定過程と考えたときの最適政策に成り立つかどうかは課題となっている。

参考文献

[1] Albright, S. C., Structural results for partially observable Markov decision processes. Oper. Res. 27 (1979), 1041–1053.

[2] Cao, X. and Guo, X., Partially observable Markov decision processes with reward information: basic ideas and models. IEEE Trans. Automat. Control 52 (2007), 677–681.

[3] Fernandez-Gaucherand, E., Arapostathis, A., and Marcus, S. I. On the average cost optimality equation and the structure of optimal policies for partially observable Markov decision processes. Ann. Oper. Res. 29 (1991), 439–469.

[4] Grosfeld-Nir, A., A two-state partially observable Markov decision process with uniformly distributed observations. Oper. Res. 44 (1996), 458–463.

[5] Itoh, H. and Nakamura, K., Partially observable Markov decision processes with imprecise parameters. Artificial Intelligence 171 (2007), 453–490.

[6] T. Nakai, A Generalization of Multivariate Total Positivity of Order Two with an Application to Bayesian Learning Procedure, *Journal of Information & Optimization Sciences*, vol. 23, 163–176, 2002.

[7] T. Nakai, A Sequential Expenditure Problem for Public Sector Based on the Outcome, *Recent Advances in Stochastic Operations Research* (Eds. T. Dohi, S. Osaki and K. Sawaki), World Scientific Publishing, 277–295, 2007.

[8] T. Nakai, A Sequential Decision Problem based on the Rate Depending on a Markov Process, *Recent Advances in Stochastic Operations Research 2* (Eds. T. Dohi, S. Osaki and K. Sawaki), World Scientific Publishing, 11–30, 2009.

[9] Ohnishi, M., Kawai, H. and Mine, H., An optimal inspection and replacement policy under incomplete state information. European J. Oper. Res. 27 (1986), 117–128.

[10] Shaked, M. and Shanthikumar, J. G., *Stochastic Orders and Their Applications* (Probability and mathematical statistics : a series of monographs and textbooks), Academic Press, Boston, Massachusetts, 1994.

[11] White, D. J. Structural properties for contracting state partially observable Markov decision processes. J. Math. Anal. Appl. 186 (1994), 486–503