数理解析研究所講究録
第1912巻 2014年 193-201

確率的凸性と部分観測可能なマルコフ決定過程について

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

1　はじめに

Nakai[8, 11, 12] などにおいて、状態空間が $(-\infty,\infty)$ の部分観測可能なマルコフ決定過程における学習過程と最適政策・最適値との関係を考えた。この中で、状態に関する情報を、$(-\infty,\infty)$ 上の確率変数で表し、状態 $s\in(-\infty,\infty)$ が大きくなれば、良い状態と考えた。このモデルは、公的部門の最適支出問題と捉えた。すなわち、公的部門の活動の評価ではアウトカムは重要な要素であり、アウトカムに基づく公的部門に対する支出問題である。たとえば、消防などの公的サービスの活動を考え、年度ごとに予算の範囲で支出することを考えるが、このような公的サービスにおけるアウトカムの指標は公的な支出によって変化するだけでなく、マルコフ過程にしたがっても変化するマルコフ決定過程としたのである。

このモデルでは、アウトカムの指標を状態とする確率過程を考え、マルコフ過程での多段決定問題として支出モデルを定式化する。さらに、状態は確率的に推移するとともに、追加支出によっても変化する。このとき、アウトカムを改善するためにどのくらい支出すれば良いかを決める問題である。

このモデルでは、利得最大化問題として定式化をし最適政策や最適値について考えた。さらに、ここでは費用最小化問題としてのマルコフ決定過程についても考える。

2　確率的順序関係と凸性

まずはじめに、部分観測可能なマルコフ決定過程における、最適政策や最適値と学習や決定との関係を見るために必要となる確率的順序関係と確率的凸性について基本的な性質をまとめる。

2.1　確率的順序関係

X と Y を2つの確率変数とする。これら2つの確率変数のあいだの確率的な順序関係については、いろいろ知られているが、基本的な順序関係はつぎのようなものである。

定義 1 任意の $u\in(-\infty,\infty)$ に対して、$P(Y>u)\leq P(X>u)$ のとき、X は usual stochastic order の意味で Y より大きいと言い、$X\geq_{ST}Y$ と表す。

定義 2 (TP_2) 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。

2つの確率変数のあいだの確率的な順序関係について、関数とその期待値を使って定義することも出来る。その主なものはつぎのようなものである。

(1) $X\geq_{ST}Y\Longleftrightarrow$ 任意の増加関数 $u(s)$ に対して、$E[u(X)]\geq E[u(Y)]$ である。(stochastic order)

(2) $X \geq_{ICX} (\geq_{DCX}) Y \Longleftrightarrow$ 任意の増加 (減少) 凸 (convex) 関数 $u(s)$ に対して、$E[u(X)] \geq E[u(Y)]$ である。(increasing (decreasing) convex order)
(3) $X \geq_{ICV} (\geq_{DCV}) Y \Longleftrightarrow$ 任意の増加（減少）凹 (concave) 関数 $u(s)$ に対して、$E[u(X)] \geq E[u(Y)]$ である。(increasing (decreasing) concave order)

2.2 確率的凸性と凹性

$\{X(s)|s \in (-\infty, \infty)\}$ を s をパラメータとする確率変数列とするとき、Shaked and Shanthikumar[15] にしたがって、確率的凸性と凹性を定義する。

(1) $\{X(s)|s \in (-\infty, \infty)\}$ が SI(stocahstically increasing) とは、任意の増加 関数 $u(s)$ に対して $E[u(X(s))]$ が、s の増加関数となることをいう。
(2) $\{X(s)|s \in (-\infty, \infty)\}$ が SICX(stocahstically increasing and convex) とは、任意の増加凸関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の増加凸 関数となることをいう。
(3) $\{X(s)|s \in (-\infty, \infty)\}$ が SICV(stocahstically increasing and concave) とは、任意の増加凹関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の増加凹関数となることをいう。

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

(1) $\{X(s)|s \in (-\infty, \infty)\}$ が SICX(sp)(stocahstically increasing and convex in sample path sense) とは、$\max\{X_2, X_3\} \leq X_4$ であり (a.s.)、$X_2 + X_3 \leq X_1 + X_4$ となることをいう。
(2) $\{X(s)|s \in (-\infty, \infty)\}$ が SICV(sp)(stocahstically increasing and concave in sample path sense) とは、$X_1 \leq \max\{X_2, X_3\}$ であり (a.s.)、$X_2 + X_3 \geq X_1 + X_4$ となることをいう。

例 1 $X(\mu)$ を正規分布 $N(\mu, \sigma^2)$ とする。$\{X(\mu)|\mu \in (-\infty, \infty)\}$ は *SICX(sp)* であり *SICV(sp)* である。

このとき、つぎの性質が成り立つ。

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 2 $X(\mu)$ 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 凹関数と劣モジュラ関数

s と x の関数 $\sigma(s, x) = s(x)$ は凹関数 (凸関数) とは、$x < y$ 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

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

となることである。

定義 3 s と x の関数 $\sigma(s,x)$ が、$x<y$ および $s<t$ となる任意の x,y と s,t に対して

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

のとき、劣モジュラ関数 (優モジュラ関数)(*submodular (supermodular) function*) という。

凹関数と劣モジュラ関数 (凸関数と優モジュラ関数) はよく似た性質を持っており、特定の条件の下で同値であることが知られている (Lovasz[6])。

補題 3 s と x の関数 $\sigma(s,x)$ を凹な劣モジュラ関数とし、$u(s)$ を凹関数とすれば、$u(\sigma(s,x))$ は凹関数である。

4 逐次支出モデル

状態空間を $(-\infty,\infty)$ とするマルコフ過程を考え、状態 s をアウトカムを表す指標とする。指標は s の値が大きくなるにしたがって良くなると考え、この指標を改善するために支出を行う。また、状態は、支出によるだけでなく、マルコフ過程の推移法則にしたがっても推移する。この問題はアウトカムを良くするために、どのくらい支出すれば良いかを決定する問題であり、マルコフ過程における多段決定問題として定式化する。

ここで、$(-\infty,\infty)$ を状態空間とし、s を状態とする。決定する支出額が x とすれば、この決定により状態は $\sigma(s,x)$ となり、この決定に伴う支出を $C(x)$ とする。$C(x)=x$ のときは、費用は投入額に等しいことになる。$u(s)$ を状態が s のときの終端利得とし、非減少非負な凹関数とする。このマルコフ過程の推移法則を $\boldsymbol{P}=(p_s(t))$ とし、$T(s)$ を状態 s に対して、$p_s(t)$ を密度関数とする確率変数とすれば、マルコフ過程の推移後の状態を表す確率変数となる。

仮定 1 $s'>s$ となる任意の s',s に対して、$T(s')\geq_{LRD}T(s)$ とする。

補題 4 任意の非減少非負関数 $u(s)$ に対して、$s<s'$ なら $E[u(T(s))]\leq E[u(T(s'))]$ である。

仮定 2 s と x の関数 $\sigma(s,x)=s(x)$ を非減少非負関数で凹な劣モジュラ関数とする。

$\sigma(s,x)=s+\sigma(x)$ で、$\sigma(x)$ が増加非負凹関数ならば、この関数は仮定 2 を満たす。

このモデルでは、決定と推移の順序はつぎのように考える。このマルコフ決定過程の状態が s のとき、決定 x をとる。この決定により状態は $\sigma(s,x)$ となる。つぎに、推移法則 $\boldsymbol{P}$ にしたがって状態が推移し、状態は $T(\sigma(s,x))$ となる。

つぎに、

$$\overline{v}(s)=\max_{x\geq 0}\{-C(x)+u(\sigma(s,x))\}$$

とおけば、$u(s)$ 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。さらに、つぎの性質が成り立つ。

補題 5 $C(x)$ 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。

4.1 逐次支出モデル

n を決定期間とし、K を予算の上限とする。マルコフ過程の状態が s のときの終端利得を $u(s)$ とし、s の増加凹関数とする。$C(x)$ を決定 x を取ったときの費用とし、x に関する増加凸関数とする。このとき、$u_n(s)$ を最適値とし、$x_n^*(s)$ を最適決定とする。このとき、最適性の原理より、最適方程式

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

が得られる。ただし、$u_1(s) = \max_{x\geq 0}\{-C(x) + E[u(T(\sigma(s,x)))]\}$ とする。

つぎの仮定の下で、関数 $u_n(s)$ と最適政策 $x_n^*(s)$ の s に関する性質が得られる。

仮定 3 確率変数列 $\{T(s)|s \in (-\infty,\infty)\}$ は、$SICV$ である。

補題 6 $u_n(s)$ は、s に関する増加関数である。

補題 7 仮定3のもとで、$u_n(s)$ は凹関数である。

性質 1 仮定3のもとで、$x_n^*(s)$ は s に関して減少する。

$x_n(s)$ の n に関する性質をつぎの仮定の下で考える。

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

仮定4より、任意の $n \geq 1$ に対して

$$E[u_n(T(t))] - E[u_n(T(s))] \leq E[u_{n-1}(T(t))] - E[u_{n-1}(T(s))] \tag{3}$$

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。

5 費用最小化問題

逐次支出問題では効用最大化を考えた。つぎに、同様の問題を機会などのシステムを維持する費用最小化問題に適応する。状態空間が $(-\infty,\infty)$ のマルコフ過程を考え、状態 s が大きくなるにしたがってシステムの状態が悪くなると考える。このとき、この状態を改善するために支出を行う。状態は、マルコフ過程の推移法則にしたがって推移するとともに、決定によって改善される。状態を改善するために、どのくらい支出すれば良いかを決定する問題であり、マルコフ過程における多段決定問題として定式化する。

このため、$(-\infty,\infty)$ を状態空間、s を状態とし、x を支出額とする。状態が s のとき、決定 x により移る状態を $\sigma(s,x)$ とすれば、s の増加関数であり x の減少関数とする。$C(x)$ を決定 x に伴う費用とし、x の増加関数とする。$u(s)$ は状態が s のときの終端利得であり、非減少非負な凸関数とする。これまで同様、$\boldsymbol{P} = (p_s(t))$ をマルコフ過程の推移法則とし、$T(s)$ を任意の状態 s に対して $p_s(t)$ を密度関数とする確率変数とおく。

このとき、つぎの仮定を置く。

仮定 5 $s' > s$ となる任意の s', s に対して、$T(s') \geq_{LRD} T(s)$ である。

補題 8 任意の非減少非負関数 $u(s)$ に対して、$s < s'$ なら $E[u(T(s))] \leq E[u(T(s'))]$ である。

仮定 6 s と x の関数 $\sigma(s,x)$ を s の非減少非負凸関数で、x の非増加非負凸関数とし、$x < y$ および $s < t$ となる任意の x, y と s, t に対して、優モジュラ関数

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

とする。

$\sigma(s,x) = s - \sigma(x)$ で、$\sigma(x)$ が増加非負凸関数ならば、この関数は仮定 6 を満たす。

このとき、決定と推移の順序をつぎのように考える。状態を s のとき、決定 x をとる。状態はこの決定により $\sigma(s,x)$ となる。つぎに、推移法則 $\boldsymbol{P}$ にしたがって状態が推移し、状態は $T(\sigma(s,x))$ となる。

ここで、

$$\overline{w}(s) = \min_{x \geq 0}\{C(x) + u(\sigma(s,x))\}$$

とおく。このとき、$u(s)$ が s の増加関数であれば、$\overline{w}(s)$ も増加関数である。

補題 9 $C(x)$ が凸関数のとき、$u(s)$ が凸関数ならば、$\overline{v}(s)$ も凸関数である。

5.1 費用最小化モデル

この多段決定問題の決定期間を n とし、s をマルコフ過程の状態とする。$C(x)$ を決定 x に対する費用とし、x の増加凸関数とする。$u(s)$ は終端利得で、s の増加凸関数とする。このとき、最適値を $u_n(s)$、最適政策を $x_n^*(s)$ とすれば、最適性の原理より、つぎの最適方程式が得られる。

$$w_n(s) \quad = \quad \min_{x \geq 0}\{C(x) + E[w_{n-1}(T(\sigma(s,x)))]\}, \tag{4}$$

ただし、$w_1(s) = \min_{x \geq 0}\{C(x) + E[u(T(\sigma(s,x)))]\}$ である。

ここで、つぎの仮定を置くと、最適値 $w_n(s)$ と最適政策 $x_n^*(s)$ の s に関する単調性が求まる。

仮定 7 確率変数列 $\{T(s)|s \in (-\infty, \infty)\}$ は、$SICX$ である。

補題 10 $w_n(s)$ は、s に関する増加関数である。

補題 11 仮定 7 のもとで、$w_n(s)$ は凸関数である。

性質 3 仮定 7 のもとで、$x_n^*(s)$ は s に関して増加する。

つぎの仮定の下で、最適政策 $x_n(s)$ の n に関する単調性が求められる。

仮定 8 $t \geq s$ のとき任意の凸関数 $u(s)$ に対し、$E[u(T(t))] - E[u(T(s))] \leq u(t) - u(s)$ である。

仮定 8 より、任意の $n \geq 1$ に対して

$$E[w_n(T(t))] - E[w_n(T(s))] \geq E[w_{n-1}(T(t))] - E[w_{n-1}(T(s))] \tag{5}$$

となる。

性質 4 仮定 8 のもとで、$x_n(s)$ は n に関して増加する。

6 部分観測可能なマルコフ過程

状態空間を $(-\infty,\infty)$ とする部分観測可能なマルコフ過程を考える。このマルコフ過程の状態 s はひとつの指標とする。それぞれの状態 s ($s\in(-\infty,\infty)$) に対して確率変数 Y_s を考え、これを観測課程とする。すなわち、観測できない状態に関する情報は、これらの確率変数 Y を観測することで得る。ここでは、ベイズ学習にしたがって情報を改良する。

6.1 部分観測可能なマルコフ過程と事前・事後情報

観測できない状態に関する情報を状態空間上の確率分布 $\mu(s)$ で表し、$\mathcal{S}$ を情報全体の集合とする。この集合 $\mathcal{S}$ に含まれる情報のあいだには、LRD ($\geq_{LRD}$) に基づく順序関係を仮定する。また、Y_s を状態 s に対して依存する確率変数で、この確率変数の値を観測して状態に関する情報を得る。この確率変数は期待値有限で、分布関数を $F_s(y)$ とする。このとき、$s\leq t$ ならば、$Y_t\geq_{LRD} Y_s$ と仮定する ($s,t\in(-\infty,\infty)$)。したがって、つぎの性質が成り立つ。

補題 12 $h(y)$ を非減少非負関数とし、$F_{\boldsymbol{\mu}}(y)$ を $\boldsymbol{\mu}$ に関する *weighted distribution function* とする。このとき、$\boldsymbol{\mu}\geq_{LRD}\boldsymbol{\nu}$ 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。

ここで、つぎのような記号を用いる。$\boldsymbol{\mu}$ を事前情報としての状態空間上の確率分布とする。このとき、$\overline{\boldsymbol{\mu}}$ を推移法則に従って状態が推移したあとの状態空間上の確率分布とし、$\boldsymbol{\mu}_y$ を y を観測したあと、ベイズの定理にしたがって改良した事後情報とする。さらに、$\boldsymbol{\mu}^x$ を決定 x を取ったとき、この決定によって変化する状態空間上の確率分布とする。

ここで、観測、決定、推移の順序についてつぎのように考えることにする。事前情報を $\boldsymbol{\mu}$ のとき、値 y を観測し状態に関する情報を得る。この値をもとに、ベイズの定理にしたがって情報を $\boldsymbol{\mu}_y\in\mathcal{S}$ と改良する。この情報に基づき、決定 x をとり、その結果情報は $\boldsymbol{\mu}_y^x$ となる。最後に、推移法則 $\boldsymbol{P}$ にしたがって状態が推移し、事後情報は $\overline{\boldsymbol{\mu}_y^x}$ となると考える。

このとき、$x,s\in\Re$ に関する非負の集合値関数 $\boldsymbol{h}(x)=(h(x,s))_{s\in(-\infty,\infty)}$ の単調性をつぎのように定義する。

定義 4

任意の $s\in\Re$ と $x\in\Re$ の非負集合値関数 $\boldsymbol{h}(x)=(h(x,s))_{s\in(-\infty,\infty)}$ に対して、任意の t と s ($s\leq t$ かつ $s,t\in(-\infty,\infty)$) について、

(1) $x<y$ ならば $\boldsymbol{h}(y)\geq_{LRD}\boldsymbol{h}(x)$ とする。すなわち $h(x,t)\,h(y,s)\leq h(x,s)\,h(y,t)$ である。このとき、関数 $\boldsymbol{h}(x,s)$ を x に関する増加関数という。

(2) $x<y$ ならば $\boldsymbol{h}(x)\geq_{LRD}\boldsymbol{h}(y)$ とする。すなわち $h(x,t)\,h(y,s)\geq h(x,s)\,h(y,t)$ である。このとき、関数 $\boldsymbol{h}(x,s)$ を x に関する減少関数という。

この定義を用いれば、事前情報と事後情報のあいだには、つぎのような関係がある。

補題 13 $\boldsymbol{\mu}\geq_{LRD}\boldsymbol{\nu}$, $y<y'$ とする。

(1) 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。

(2) 任意の $\boldsymbol{\mu}$ に対して、$\boldsymbol{\mu}_{y'}\geq_{LRD}\boldsymbol{\mu}_y$ および $\overline{\boldsymbol{\mu}_{y'}}\geq_{LRD}\overline{\boldsymbol{\mu}_y}$ である。

6.2 事後情報と決定の関係

以下では、$\sigma(s,x) = s + \sigma(x)$ とし、$\sigma(x)$ は $x \geq 0$ の増加非負凹関数で $\sigma(0) = 0$ とする。このとき、μ^x を決定 x をとったあとの事後情報とすれば、$s = t - \sigma(s,x)$ より $\mu^x(t) = p_{t-\sigma(s,x)}(t)$ となる。ただし、$\mu = \mu^0$ である。

状態全体の集合 $\mathcal{S}$ に含まれる確率分布 μ が

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

となるとき、この μ は性質 (G) を満たすということにする。

例 3 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。

このとき、つぎの性質が成り立つ。

補題 14 μ が性質 (G) を満たせば、μ^x は任意の x に対して性質 (G) を満たす。

補題 15 μ, ν を性質 (G) を満たす $\mathcal{S}$ に含まれる 2 つの情報とすれば，つぎの性質を持つ。

(1) $\mu \geq_{LRD} \nu$ ならば、任意の x, y に対して $\mu^x \geq_{LRD} \nu^x, \overline{\mu^x} \geq_{LRD} \overline{\nu^x}$ である。
(2) $x > x'$ ならば、$\mu^x \geq_{LRD} \mu^{x'}, \overline{\mu^x} \geq_{LRD} \overline{\mu^{x'}}$ である。

推移法則と情報過程に関して、つぎの仮定を設けるれば、いくつかの性質が示される。

仮定 9 任意の $s < t,\ t \leq t'$ および $u < v$ となる s, t, 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。

例 4 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 9 の条件を満足する。

補題 16 $\mu \in \mathcal{S}$ が性質 (G) を満たすならば、$\overline{\mu}$ もまた性質 (G) を満たす。

仮定 10 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。

例 5 $f_s(y)$ を正規分布 $N(s, \sigma^2)$ の密度関数とすれば、仮定 10 を満たす。

補題 17 $\mu \in \mathcal{S}$ が性質 (G) を満たせば、任意の y に対して $\overline{\mu_y}$ もまた性質 (G) を満たす。

補題 18 仮定 9 と 10 のもとで、μ が性質 (G) を満たせば、任意の x と y に対して $\overline{\mu}$, μ_y および μ_y^x も性質 (G) を満たす。

補題 19 μ, ν : 性質 (G) を満たす $\mathcal{S}$ に含まれる事前情報

(1) $y > y'$ ならば、任意の x に対して $\mu_y^x \geq_{LRD} \mu_{y'}^x, \overline{\mu_y^x} \geq_{LRD} \overline{\mu_{y'}^x}$ である。
(2) $\mu \geq_{LRD} \nu$ なら、任意の x, y に対して $\mu_y \geq_{LRD} \nu_y, \mu_y^x \geq_{LRD} \nu_y^x, \overline{\mu_y^x} \geq_{LRD} \overline{\nu_y^x}$ である。
(3) $x > x'$ なら、任意の y に対して $\mu_y^x \geq_{LRD} \mu_y^{x'}, \overline{\mu_y^x} \geq_{LRD} \overline{\mu_y^{x'}}$ である。

7 部分観測可能なマルコフ過程での多段決定問題

状態空間を $(-\infty,\infty)$ とするマルコフ過程を考え、状態を s とする。この状態は部分観測可能なマルコフ過程にしたがって推移し、状態に関する情報は、それぞれの状態 s $(s\in(-\infty,\infty))$ に対する確率変数 Y_s を通して得る。すなわち、これらの Y を観測し、ベイズ学習にしたがって情報を改良する。さらに、状態を改良するため、決められた範囲で支出を行う。この問題は、状態を改良するため、どのくらい支出を行えば良いかを決定することである。

ここで、観測、決定、推移の順序をつぎのように考える。この順序を変えても同様の性質が得られる。事前情報を $\boldsymbol{\mu}$ とし、状態に依存する確率変数を観測し、観測値を y とする。この値を使って、ベイズの定理にしたがって情報を $\boldsymbol{\mu}_y\in\mathcal{S}$ と改良する。つぎに、決定 x をとり、状態 s を $s+\sigma(x)$ とする。この決定 x により、情報は $\boldsymbol{\mu}_y^{\sigma(x)}$ となる。最後に、推移法則 $\boldsymbol{P}$ にしたがって状態が推移し、事後情報は $\overline{\boldsymbol{\mu}_y^{\sigma(x)}}$ となる。

7.1 部分観測可能なマルコフ決定過程

n をこの問題の計画期間とし、決定を x とする $(0\le x\le K)$。このとき、$C(x)$ を決定 x に伴う費用とし、$\sigma(s,x)=s+\sigma(x)$ を状態が s のとき、この決定 x により変化する状態とする。K を予算の上限とする。$\boldsymbol{\mu}$ を事前情報とするとき、$w_n(\boldsymbol{\mu})$ を最適値とおく。

このとき、事前情報が $\boldsymbol{\mu}$ のとき、値 y を観測し、ベイズの定理にしたがって情報を $\boldsymbol{\mu}_y\in\mathcal{S}$ と改良する。この情報のもとで、決定 x をとり、状態に関する情報は $\boldsymbol{\mu}_y^{\sigma(x)}$ となる。最後に、推移法則 $\boldsymbol{P}$ に従って状態が推移し、事後情報は $\overline{\boldsymbol{\mu}_y^{\sigma(x)}}$ となる。このとき最適政策にしたがったときの最適値は $w_{n-1}(\overline{\boldsymbol{\mu}_y^{\sigma(x)}})$ である。したがって、最適性の原理より、最適方程式は

$$\begin{aligned} w_n(\boldsymbol{\mu}) &= E[w_n(\boldsymbol{\mu}|Y)] \\ w_n(\boldsymbol{\mu}|y) &= \min_{x\ge 0}\left\{c(x)+w_{n-1}(\overline{\boldsymbol{\mu}_y^{\sigma(x)}})\right\} \end{aligned} \tag{6}$$

となる。ただし、S を決定過程の状態を表す確率変数とすれば、$w_0(\boldsymbol{\mu})=E_{\boldsymbol{\mu}}[u(S)]$ とする。

性質 5 $\boldsymbol{\mu},\boldsymbol{\nu}$ が性質 (G) を満たすとき、$\boldsymbol{\mu}\ge_{LRD}\boldsymbol{\nu}$ ならば $w_n(\boldsymbol{\mu})\ge w_n(\boldsymbol{\nu})$ である。

状態空間を $(0,\infty)$ とするマルコフ過程として最適修理問題を展開することが出来る。この問題では、修理レベルを決定する問題で、費用はこのレベルに依存し、総期待費用を最小にする修理レベルを決定する問題である。とくに、[10] にあるように、選択したレベルによって状態は積の形で変化するものと考える。また、この問題は、Monahan[7], Grosfeld-Nir[2], Albright[1], White[17], Itoh and Nakamura[3], Ohnishi, Kawai and Mine[13] などのように、部分観測可能なマルコフ決定過程の一つである。

参考文献

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

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

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

[4] M. Kijima and M. Ohnishi, Stochastic Orders and Their Applications in Financial Optimization, *Math. Methods of Oper. Res.*, **50**, 351–372, (1999).

[5] David Simchi-Levi, Xin Chen, Julien Bramel, *Convexity and Supermodularity*, The Logic of Logistics, Theory, Algorithms, and Applications for Logistics and Supply Chain Management, Springer Series in Operations Research, 2005, pp 13-32

[6] L. Lovasz, Submodular functions and convexity, in: Mathematical Programming: the State of the Art (ed. A.Bachem, M.Grotschel, B.Korte), Springer (1983), 235-257.

[7] G. E. Monahan, Optimal selection with alternative information. *Naval Res. Logist. Quart.* 33 (1986), 293–307.

[8] 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.

[9] 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.

[10] T. Nakai, Sequential Decision Problem with Partial Maintenance on a Partially Observable Markov Process, *Scientiae Mathematicae Japonicae*, vol. 72, no. 1, 11–20, 2010.

[11] 中井　達, 多段決定問題と Stochastic Convexity について, 京都大学数理解析研究所講究録「不確実・不確定環境下における数理的意思決定とその周辺」, vol. 1802, 193–199, 2012.7.

[12] 中井　達, 投資モデルに基づく逐次決定問題について, 京都大学数理解析研究所講究録「決定過程に関わる数理モデルの新たな展開と応用」, vol. 1857, 109–120, 2013.10.

[13] 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.

[14] S. M. Ross, *Stochastic Processes*, John-Wiley and Sons, New York, New York, 1983.

[15] 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.

[16] Moshe Shaked and J. George Shanthikumar, Parametric stochastic convexity and concavity of stochastic processes, Annals of the Institute of Statistical Mathematics, September 1990, Volume 42, Issue 3, pp 509-531

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