数理解析研究所講究録
第 1990 巻 2016 年 222-229

決定回数が未知の多段決定問題について

中井 達

千葉大学教育学部

1 逐次支出問題

Nakai[8, 11] などにおいて、公共サービスに対する評価とそれを基にした支出問題に関して、部分観測可能なマルコフ決定過程における逐次決定問題として学習過程と最適政策・最適値との関係を考えた。このモデルでは、状態空間が $(-\infty, \infty)$ のマルコフ過程を考え、その状態を s とし、s の値が大きくなるにしたがって状態は良くなると考える。この状態を改善するために支出を行い、状態はマルコフ過程の推移法則にしたがっても推移する。したがって、効用最大化問題を考え、状態を良くするために、どのくらい支出すれば良いかを決定する問題であるり、マルコフ過程における多段決定問題として定式化する。

そのため、$(-\infty, \infty)$ を状態空間とし、s をその状態とする。決定変数として支出額を x とし、$f(s, x)$ を状態が s のとき、決定 x により移る状態とする。また、$C(x)$ を決定 x に伴う費用とし、$u(s)$ 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。つぎに、$\boldsymbol{P} = (p_s(t))$ をマルコフ過程の推移法則とするとき、$T(s)$ 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。

2 劣モジュラー関数と確率的凸性

ここで用いる性質についてまとめる。

2.1 劣モジュラー関数と優モジュラー関数

任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}^n$ に対して

$$
\begin{aligned}
\boldsymbol{x} \vee \boldsymbol{y} &= (\max\{x_1, y_1\}, \cdots, \max\{x_n, y_n\}) \\
\boldsymbol{x} \wedge \boldsymbol{y} &= (\min\{x_1, y_1\}, \cdots, \min\{x_n, y_n\})
\end{aligned}
$$

とおく。このとき、

定義 1 X を $\mathcal{R}^n$ の部分集合で、n 変数関数を $f(\boldsymbol{x})$ とする。この関数 $f(\boldsymbol{x})$ が U で優モジュラー (supermodular) であるとは、任意の $\boldsymbol{x}, \boldsymbol{x}' \in X$ に対して

$$
f(\boldsymbol{x}) + f(\boldsymbol{x}') \leq f(\boldsymbol{x} \vee \boldsymbol{x}') + f(\boldsymbol{x} \wedge \boldsymbol{x}'), \tag{1}
$$

となることである。ただし、$\boldsymbol{x} \vee \boldsymbol{x}', \boldsymbol{x} \wedge \boldsymbol{x}' \in X$ とする。f が strictly supermodular であるとは、不等式 (1) が厳密に成り立つ $\boldsymbol{x}$ と $\boldsymbol{x}'$ が存在することである。関数 $f(\boldsymbol{x})$ が (*strictly*) 劣モジュラー (submodular) であるとは、$-f(\boldsymbol{x})$ が (*strictly*) 優モジュラー (supermodular) となることである。

したがって、劣モジュラ関数 (優モジュラ関数)(submodular (supermodular) function) であるとは、s と x の関数 $f(s,x)$ が、$x < y$ および $s < t$ となる任意の x, y と s, t に対して

$$f(t,y) - f(t,x) \leq (\geq) f(s,y) - f(s,x)$$

のときと同等である。

このとき次の性質が成り立つ。

補題 1 s と x の関数 $f(s,x)$ を凹な劣モジュラ関数とし、$u(s)$ を凹関数とすれば、$u(f(s,x))$ は凹関数である。

つぎに、任意の $i, j \in \{1, 2, \cdots, n\}$ に対して

$$\hat{\boldsymbol{x}}_{ij} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) \in \mathcal{R}^{n-2},$$

とおき、任意の関数 r $f : \mathcal{R}^n \to \mathcal{R}$ にたいし

$$f_{\hat{\boldsymbol{x}}_{ij}}(x_i . x_j) = f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_{j-1}, x_j, x_{j+1}, \cdots, x_n)$$

とおく。

$y \leq y'(y < y')$ となる任意の $y, y' \in \mathcal{R}$ に対し、$f(x,y') - f(x,y)$ が x に関して (strictly) increasing のとき、この 2 変数関数 f は (strictly) increasing differences を持つという。

$f_{\hat{\boldsymbol{x}}_{ij}}(x_i . x_j)$ が、任意の $i, j \in \{1, 2, \cdots, n\}$ と $\hat{\boldsymbol{x}}_{ij} \in \mathcal{R}^{n-2}$ に対して increasing differences を持つとき、この n 変数関数 $f(\boldsymbol{x})$ は increasing differences を持つと定義する。 n 変数関数 $f(\boldsymbol{x})$ が decreasing differences を持つ場合も同様に定義する。

定理 1 (Simchi-Levi, Chen, Bramel [16]) n 変数関数 $f(\boldsymbol{x})$ が (strictly) supermodular であることと、$f(\boldsymbol{x})$ が (strictly) increasing differences を持つことは同値である。

2.2 確率的凸性と凹性

Shaked and Shanthikumar(1994) にしたがって、確率的凸性と凹性を考える。

$\{X(s) | s \in \Theta\}$ を s をパラメータとする確率変数列とする。ただし、Θ は $(-\infty, \infty)$ または $(-\infty, \infty)$ に含まれる凸集合とする。

(1) $\{X(s) | s \in (-\infty, \infty)\}$ が SI(stocahstically increasing) であることと、任意の増加関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の増加関数であることは同値である。

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

(3) $\{X(s)|s \in (-\infty,\infty)\}$ が SICV(stocahstically increasing and concave) であることと、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ことは同値である。

(4) $\{X(s)|s \in (-\infty,\infty)\}$ が SDCX(stocahstically decreasing and convex) であることと、任意の増加凸関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の減少凸 関数であることは同値である。

(5) $\{X(s)|s \in (-\infty,\infty)\}$ が SDCV(stocahstically decreasing and concave) であることと、任意の増加凹関数 $u(s)$ に対して、$E[u(X(s))]$ が、s の減少凹関数であることは同値である。

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

定義 2 (1) $\{X(s)|s \in \Theta\}$ が SICX(sp)(stocahstically increasing and convex in sample path sense) であることと、$\max\{X_2, X_3\} \le X_4$ であり *(a.s.)*、$X_2 + X_3 \le X_1 + X_4$ であることは同値である。

(2) $\{X(s)|s \in \Theta\}$ が SICV(sp)(stocahstically increasing and concave in sample path sense) であることと、$X_1 \le \max\{X_2, X_3\}$ であり *(a.s.)*、$X_2 + X_3 \ge X_1 + X_4$ であることは同値である。

補題 2 (1) $\{X(s)|s \in \Theta\}$ が SICX(sp) ならば、SICX である。

(2) $\{X(s)|s \in \Theta\}$ が SICV(sp) ならば、SICV である。

補題 3 (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。

2.3 尤度比順序

X と Y を 2 つの確率変数とする。

定義 3 (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、$x \ge y$ となる任意の x と y に対して、$f_X(y)f_Y(x) \le f_X(x)f_Y(y)$ であるとき、X は Y より尤度比の意味で大きいといい、$X \ge_{LRD} Y$ あるいは $X \succeq Y$ と表す。

つぎの補題 4 は、よく知られた性質である。(Kijima and Ohnishi[5] など)

補題 4 $X \succeq Y$ ならば、非減少非負関数 $h(x)$ に対して $E[h(X)] \ge E[h(X)]$ となる。

2.4 決定回数が未知の決定過程

決定回数が未知の決定過程を考える。$\boldsymbol{q} = (q_0, q_1, q_2, \cdots)$ を残りの決定回数 N に関する事前情報とし、それぞれの決定機会が現れるまでの経過時間を表す確率変数 Z は互いに独立で、指数分布にしたがうものとする。Z_j を j 番目の事象が現れるまでの時間とすれば、

$$P(Z_j \leq t) = 1 - e^{-\lambda t}$$

となる。このとき、確率変数 Y を残りの決定機会 N 個のうち最初の決定機会があらわれるまでの時間とすれば、

$$P(Y \leq t | N = k) = 1 - (e^{-\lambda t})^k = 1 - e^{-k\lambda t}$$

となる。

つぎに、$\overline{\boldsymbol{q}} = (\overline{q}_0, \overline{q}_1, \overline{q}_2, \cdots)$ を最後に決定を行ってから t 時間後に決定機会が現れたとき、残り決定機会の回数に関する事後情報とすれば、

$$\overline{q}_k = c q_{k+1} e^{-k\lambda t}$$

となる。ただし、$\sum_{k=0}^{n-1} \overline{q}_k = 1$ である。また、$\boldsymbol{q}^* = (q_0^*, q_1^*, q_2^*, \cdots)$ を最後に決定をしてから t 時間のあいだに新たな決定機会が現れないとき、残り決定機会の回数に関する事後情報は、

$$q_k^* = d q_k e^{-k\lambda t}$$

となる。ただし、$\sum_{k=0}^{n} q_k^* = 1$ である。

3 有限期間の逐次支出問題

3.1 逐次支出問題とマルコフ決定過程

計画期間が有限の逐次支出問題について簡単にまとめる。決定期間を n とし、s を確率過程の状態とする。$u(s)$ を終端利得とし、$C(x)$ を決定 x に対する費用としたとき、$v_n(s)$ を n 期間問題の最適値とし、$x_n^*(s)$ を最適決定とする。

このとき、最適方程式は

$$v_n(s) \quad = \quad \max_{x \geq 0}\{-C(x) + E[v_{n-1}(T(f(s,x)))]\} \tag{2}$$

となる。ただし、$v_1(s) = \max_{x \geq 0}\{-C(x) + E[u(T(f(s,x)))]\}$ である。

仮定 1

(1) $u(s)$ は s の増加凹関数とする。

(2) $C(x)$ は x の増加凸関数で $C(0) = 0$ とする。

(3) $f(s,x)$ は s,x の劣モジュラー凹関数で $f(s,0)=s$ とし、s と x の厳密な増加関数とする。
(4) $s \le s'$ となる s,s' に対して $T(s') \succeq T(s)$ とする。
(5) $\{T(s)|s \in (-\infty,\infty)\}$ は SICV とする。

つぎに、決定と推移の順序をつぎのように考える。状態が s のとき、決定 x をとり、この決定により状態は $f(s,x)$ 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 $T(f(s,x))$ となる。このとき、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。

このことから、つぎの性質が導かれる。

補題 6 $v_n(s)$ は、s に関する増加凹関数である。

性質 1 仮定1のもとで、$x_n^*(s)$ は s に関して減少する。

仮定 2 $t \ge s$ のとき任意の凹関数 $u(s)$ に対し、$E[u(T(t))]-E[u(T(s))] \le u(t)-u(s)$ である。

仮定2より、任意の $n \ge 1$ に対して

$$E[v_n(T(t))] - E[v_n(T(s))] \le E[v_{n-1}(T(t))] - E[v_{n-1}(T(s))] \tag{3}$$

となり、最適政策 $x_n(s)$ と計画期間 n に関するつぎの性質が成り立つ。

性質 2 仮定2のもとで、$x_n(s)$ は n に関して減少する。

4 決定回数が未知の多段決定問題

決定回数が未知の多段決定問題として、逐次支出問題を考える。マルコフ過程の状態が s のとき、$\boldsymbol{q}=(q_0,q_1,q_2,\cdots)$ を残りの決定回数に関する事前情報とし、n を残り決定回数の最大値とする。決定変数 x に対して、支出の上限 K の範囲で決定を行う。状態が s のときの終端利得を $u(s)$ とし、決定 x を取ったときの費用を $C(x)$ とする。このとき、$v_n(s;T,t,\boldsymbol{q})$ を最適に振る舞ったときに得られる総期待利得とし、$x_n^*(s;T,t,\boldsymbol{q})$ をこの多段決定問題すなわち逐次支出問題の最適政策とする。

4.1 最適方程式

いま、$v_n(s;T,t,q)$ を最後の決定を行ったときの残存時間が T のとき、残りの決定回数に関する情報が $\boldsymbol{q}$ で残り決定回数の最大値が n とする。t 時間経過後に決定機会が現れたときに最適に振る舞って得られる総期待利得とする。さらに、$v_n^k(s;T,t,q)$

を最後の決定を行ったときの残存時間が T で残りの決定回数が k のとき、残存回数に関する情報が $\boldsymbol{q}$ で残り決定回数に関する情報の最大値が n のとき、t 時間経過後に決定機会が現れたときに最適に振る舞って得られる総期待利得とする。

このとき、最適性の原理よりつぎの再帰方程式が得られる。

$$\begin{aligned} v_n(s;T,t,\boldsymbol{q}) &= E_N[v_n^N(s;T,t,\boldsymbol{q})] \\ v_n^k(s;T,t,\boldsymbol{q}) &= -k\lambda\triangle t \max_{0\le x\le K}\{-C(x)+E_N[E_S[v_{n-1}^{N-1}(T(f(s,x));T-t-\triangle t,0,\overline{\boldsymbol{q}}))]]\} \\ &\quad +(1-k\lambda\triangle t)v_n^k(s;T,t+\triangle t,\boldsymbol{q})+o(\triangle t) \end{aligned} \tag{4}$$

ただし、E_N は N に関する期待値で有り、E_S は推移後の状態に関する期待値であり、

$$E_N[E_S[v_{n-1}^{N-1}(T(f(s,x));T-t-\triangle t,0,\overline{\boldsymbol{q}}))]] = E_S[v_{n-1}(T(f(s,x));T-t-\triangle t,0,\overline{\boldsymbol{q}})$$

である。このことから、つぎの関係式が成り立つ。

$$\begin{aligned} \frac{\partial}{\partial t}v_n^k(s;T,t,\boldsymbol{q}) &= -k\lambda \max_{0\le x\le K}\{-C(x)+E_S[v_{n-1}(T(f(s,x));T-t,0,\overline{\boldsymbol{q}})\} \\ &\quad +k\lambda v_n^k(s;T,t,\boldsymbol{q}) \end{aligned} \tag{5}$$

ただし、

$$v_0^k(s;T,t,\boldsymbol{q}) = u(s)$$

である。

はじめに、$n=1$ のときを考える。

$$\begin{aligned} v_1^1(s;T,T,\boldsymbol{q}) &= 0 \\ v_1^1(s;T,t,\boldsymbol{q}) &= \max_{0\le x\le K}\{-C(x)+E_S[u(T_{f(s,x)})]\}(1-e^{-\lambda(T-t)}) \end{aligned}$$

だから、

$$v_1(s;T,t,\boldsymbol{q}) = E_N[v_1^N(s;T,t,\boldsymbol{q})] = q_1^* \max_{0\le x\le K}\{-C(x)+E_S[u(T_{f(s,x)})]\}$$

となり、

$$\frac{\partial}{\partial t}v_1^1(s;T,t,\boldsymbol{q}) = -\lambda \max_{0\le x\le K}\{-C(x)+E_S[u(T_{f(s,x)})]\} + \lambda v_1^1(s;T,t,\boldsymbol{q})$$

となる。

つぎに、

$$\frac{\partial}{\partial t}v_n^k(s;T,t,\boldsymbol{q}) = -k\lambda \max_{0\le x\le K}\{-C(x)+E_S[v_{n-1}^{k-1}(T_{f(s,x)};T-t,0,\overline{\boldsymbol{q}})]\}+k\lambda v_n^k(s;T,t,\boldsymbol{q})$$

かつ

$$v_n^k(s;T,T,\boldsymbol{q}) = 0$$

より、

$$v_n^k(s;T,t,\boldsymbol{q}) = \max_{0\le x\le K}\{-C(x) + E_S[v_{n-1}^{k-1}(T_{f(s,x)};T-t,0,\overline{\boldsymbol{q}})]\}(1-e^{-k\lambda(T-t)})$$

である。また、

$$\begin{aligned}
&v_n(s;T,t,\boldsymbol{q})\\
&= E_N[v_n^N(s;T,t,\boldsymbol{q})]\\
&= \sum_{k=0}^{n} q_k^* \max_{0\le x\le K}\{-C(x) + E_N[E_S[v_{n-1}^{N-1}(T_{f(s,x)};T-t,0,\overline{\boldsymbol{q}})]]\}(1-e^{-k\lambda(T-t)})\\
&= \left\{\sum_{k=0}^{n} q_k^*(1-e^{-k\lambda(T-t)})\right\} \max_{0\le x\le K}\{-C(x) + E_N[E_S[v_{n-1}^{N-1}(T_{f(s,x)};T-t,0,\overline{\boldsymbol{q}})]]\}\\
&= \alpha(T,t,q) \max_{0\le x\le K}\{-C(x) + E_N[E_S[v_{n-1}^{N-1}(T_{f(s,x)};T-t,0,\overline{\boldsymbol{q}})]]\}
\end{aligned}$$

となる。ここで、$\sum_{k=0}^{n} q_k^*(1-e^{-k\lambda(T-t)}) = \alpha(T,t,\boldsymbol{q})$ とおく。このとき、つぎの性質を示すことが出来る。

補題 7 $v_n^k(s;T,t,\boldsymbol{q})$ は、T に関する増加関数であり t に関して減少する。

補題 8 $v_n^k(s;T,t,\boldsymbol{q})$ は、s に関する増加凹関数である。

性質 3 $x_n^*(s;T,t,\boldsymbol{q})$ は s に関する増加関数である。

注 1 有限期間問題の場合は、$x_n^*(s)$ は n に関する増加関数であったが、決定回数が未知の場合には最適政策 $x_n^*(s;T,t,\boldsymbol{q})$ の単調性は、事前情報 $\boldsymbol{q}$ の性質により異なり、同じような性質を示すことが出来ない。

参考文献

[1] A. R. Abdel-Hamid, J. A. Bather and G. B. Trustrum, Secretary Problem with unknown Number of Candidates, JAP, 19, 619-630, 1982.

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

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

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

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

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

[7] T. Nakai, Optimal Assignment for a Random Sequence with an Unknown Number of Jobs, *Journal of the Operations Research Society of Japan*, vol. 28, 179–194, 1985.

[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, マルコフ決定過程における学習プロセスと決定について, 京都大学数理解析研究所講究録「不確実性の下での数理モデルとその周辺」, vol. 1939, 79–87, 2015.4.

[12] 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.

[13] 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.

[14] 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

[15] M. Sakaguchi and M. Tamaki, On the Optimal Parking Problem in which Spaces Appear Randomly, Bulletin of Informatics and Cybernetics, 20, 1-10, 1982.

[16] 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

[17] T. J. Stewart, The Secretary Problem with an Unknown Number of Options, Operations Research, 29, 130-145, 1981.