数理解析研究所講究録
第835巻 1993年 153-164

ある Markov Chain の上での Optimal Selection Problem について

中井　達

九州大学経済学部

1 Introduction

最適停止問題・逐次割当問題を始めとする多くの確率的多段決定問題では、逐次に出現する確率変数を1度に1つづつ観測し、その観測値をもとにして決定を行う場合を扱うことが多い。ここでは、1度に1つではなく複数の確率変数を観測することができる場合を扱う。

ここでは、1度に複数の確率変数を観測できる確率的多段決定問題として、Nakai [4] で考えられたような optimal selection problem について考える。この optimal selection problem は、逐次に出現する確率変数を決められた期間にわたって観測し、それら確率変数の観測値の中からあらかじめ決められた数だけ選択し、選択した観測値の総和を最大にすることを目的とする。そこで、まず Nakai [4] において得られた結果について簡単に述べることにする。つぎに Markov chain に従って変化する state に、それぞれの期に観測できる確率変数が依存する場合について考える。これらの値の性質について考えるために、ここで考える Markov chain の遷移確率行列と、この Markov chain の state に依存する確率変数の確率密度関数についていくつかの仮定を設け、これらの仮定のもとでのこの問題の持つ性質についても考える。

つぎに、Markov chain に従って遷移する state に依存する確率変数を逐次に観測する問題を扱うが、部分観測可能な場合について考える。ここでは、いくつかの仮定のもとでの partially observable Markov chain の性質については既に知られているが、ここでは複数の値を一度に観測することからこの問題に必要な性質について考える。ここでは、前の2つの問題と同様に optimal policy および、この政策に従って最適に振る舞ったときに得られる total expected reward を求める。さらにこれらの partially observable optimal selection problem の性質について、確率変数の観測値と prior information との関係を中心として考える。

2 Optimal Selection Problem

この問題は逐次に出現する確率変数を決められた期間にわたって観測し、それら確率変数の観測値の中からあらかじめ決められた数だけ選択して、選択した観測値の総和を最大にすることを目的とする。すなわち、残り N期の計画期間の間に、m 個の i.i.d. 確率変数 $\{X_i\}_{i=1,\cdots,m}$が観測される。このとき、これら N期間の間に出現する m 個の確率変数の中から k個を選択し、選択した観測値の総和の期待値を最大にする問題を考える。ただし、これらの m 個のそれぞれの確率変数が計画期間の間のどの期に出現するかについては、互いに独立で一様に出現するものとする。すなわち、計画期間が Nのとき、ある確率変数が最初の期に出現する確率は $\frac{1}{N}$である。したがって、1度に観測することのできる確率変数の数は常に1個と限られたものではなく、複数の値を観測することになる。また、1度観測して選択しなかった観測値は、再び選択することはできないものと考える。

残り N期の時、m 個の確率変数が残っている場合、この期に出現する確率変数の数の分布を、$\{p_{N,m}(n)\}_{n=0,1,\cdots,m}$と表す。もし、$m$ 個の確率変数が、N期間の間に互いに独立に一様に出現する場合は、

$$p_{N,m}(n) = {}_mC_n \frac{(N-1)^{m-n}}{N^m} \quad (0 \leq n \leq m, p_{1,m}(m) = 1) \tag{1}$$

と表すことができる。ここでは、この場合について考えるが一般の場合についても同様の議論をする事ができる。

つぎに、残り N期の間に、m 個の独立で同一の分布に従う確率変数 $\{X_i\}_{i=1,\cdots,m}$ を観測し、それら m 個の確率変数の中から k個を選択して、選択した観測値の総和の期待値を最大にする問題を考えるとき、この (N,m,k) を optimal selection problem の状態と呼び、$P_{N,m,k}$と表す。また問題の状態が (N,m,k) であるとき、n 個の確率変数が観測できるという部分問題を $P_{N,m,k}(n)$ で表す。また、それら n 個の確率変数の観測値が $x_{(1)},\cdots,x_{(n)}$であるときの部分問題を $P_{N,m,k}(n;x_{(1)},\cdots,x_{(n)})$ と表す。ただし、$x_{(1)},\cdots,x_{(n)}$は n 個の確率変数 $X_1,\cdots,X_n$の観測値 $x_1,\cdots,x_n$の順序統計量とする。($x_{(1)} \geq \cdots \geq x_{(n)}$、$X_{(1)} \geq \cdots \geq X_{(n)}$) 以下では、$n$ 個の確率変数 $\{X_i\}_{i=1,\cdots,m}$に対して、それらの順序統計量を、$\{X_{(i)}\}_{i=1,\cdots,m}$で表す。この問題の目的は選択した k個の観測値の総和を最大にするような optimal policy と、そのもとで最適に振る舞ったときの total expected reward を求めることを考える。

つぎに、よく知られているように n 個の確率変数が観測できるとき、大きい方から i 番目の順序統計量 $X_{(i)}$の密度関数は、

$$g_{n,i}(x_{(i)}) = \frac{n!}{(i-1)!(n-1)!}(F(x_{(i)}))^{n-i}(1-F(x_{(i)}))^{i-1}f(x_{(i)}) \tag{2}$$

である。ここでは、観測できる m 個の確率変数 $\{X_i\}_{i=1,\cdots,m}$は絶対連続であり、すべての確率変数が独立かつ同一の分布に従うから、その確率密度関数を $f(x)$ とする。

このとき、問題 $P_{N,m,k}$、$P_{N,m,k}(n)$ および $P_{N,m,k}(n; x_{(1)}, \cdots, x_{(n)})$ において、最適に振る舞ったときに得られるこれらの問題の total expected reward をそれぞれ $v_{N,m,k}$、$v_{N,m,k}(n)$ および $v_{N,m,k}(n; x_{(1)}, \cdots, x_{(n)})$ とする。このとき、最適性の原理により次の最適方程式を満足する。

$$v_{N,m,k} = \sum_{n=0}^{m} v_{N,m,k}(n) p_{N,m}(n) \tag{3}$$

$$v_{N,m,k}(n) = E[v_{N,m,k}(n; X_{(1)}, \cdots, X_{(n)})] \tag{4}$$

$$v_{N,m,k}(n; x_{(1)}, \cdots, x_{(n)}) = \max_{1 \le i \le n} \{\sum_{j=1}^{i} x_{(j)} + v_{N-1,m-n,k-i}\} \tag{5}$$

これらの方程式の解は、次のようになることが知られている。(Nakai [4]) まず、i に関して減少する正数列 $\{a_i\}$ ($a_0 = \infty$) に対して新たに関数 $U_n(a_i, a_{i-1}|k, y)$ を

$$\begin{aligned} & U_n(a_i, a_{i-1}|k, y) \\ = & \int_0^{a_i \wedge y} a_i h_{n,k}(x_{(k)}) f(x_{(k)}) dx_{(k)} + \int_{a_i \wedge y}^{a_{i-1} \wedge y} x_{(k)} h_{n,k}(x_{(k)}) f(x_{(k)}) dx_{(k)} \\ + & \int_{a_{i-1} \wedge y}^{y} U_n(a_{i-1}, a_{i-2}|k+1, x_{(k)}) f(x_{(k)}) dx_{(k)} \end{aligned} \tag{6}$$

によって定義する。ただし、

$$U_n(a_i, a_{i-1}|n+1, y) = a_i, \quad (y \ge 0) \tag{7}$$

$$h_{n,k}(x_{(k)}) = \frac{n!}{(n-k)!}(F(x_{(k)}))^{n-k} \tag{8}$$

とする。つぎに数列 $\{a^i_{N,m}\}_{i=1,\cdots}$ および $\{a^i_{N,m}(n)\}_{i=1,\cdots}$ を次のように帰納的に定義する。$(0 \le n \le m)$

$$a^i_{N,m} = \sum_{n=0}^{m} a^i_{N,m}(n) p_{N,m}(n) \tag{9}$$

$$a^i_{N,m}(n) = U_n(a^i_{N-1,m-n}, a^{i-1}_{N-1,m-n}|1, \infty) \tag{10}$$

$$a^i_{N,m}(0) = a^i_{N-1,m} \tag{11}$$

ただし、$a^0_{N,m} = a^0_{N,m}(n) = \infty$, $a^i_{0,0} = 0$ とする。また、$a^i_{1,m} = a^i_{1,m}(m)$, $a^i_{1,m}(m) = \mathrm{E}[X_{(i)}]$ であることがわかる。このとき、つぎのような基本的な性質が求められる。(Nakai [4])

Proposition 1 問題 $P_{N,m,k}$ の optimal policy は次のようなる。

問題の状態が (N, m, k) であるとき、n 個の確率変数を観測することができ、それらの観測値の順序統計量が $x_{(1)}, \cdots, x_{(n)}$ であるとする。いま j を $x_{(j)} \ge a^{k-j+1}_{N-1,m-n}$ かつ

$1 \le j \le k \wedge n$ を満足する最大の数、すなわち $x_{(j+1)} < a^{k-j}_{N-1,m-n}$ または $j = k \wedge n$ を満足する最大の数とする。このとき、大きい方から j 個の値、すなわち $x_{(1)}, \cdots, x_{(j)}$ を select することが最適である。

Proposition 2 $v_{N,m,k}$ および $v_{N,m,k}(n)$ は、つぎのように求められる。

$$v_{N,m,k} = \sum_{i=1}^{k} a^i_{N,m}, \quad v_{N,m,k}(n) = \sum_{i=1}^{k} a^i_{N,m}(n) \tag{12}$$

Proposition 3 数列 $\{a^i_{N,m}\}_{i=1,2,\cdots}$ および $\{a^i_{N,m}(n)\}_{i=1,2,\cdots}$ は、つぎのようになる。

$$\begin{aligned} a^i_{N,m}(n) &= \sum_{j=1}^{m \wedge i} \int_{a^{i-j+1}_{N-1,m-n}}^{a^{i-j}_{N-1,m-n}} x_{(j)} g_{M,j}(x_{(j)}) dx_{(j)} \\ &+ \sum_{j=0}^{n \wedge (i-1)} a^{i-j}_{N-1,m-n} \, {}_nC_j (1 - F(a^{i-j}_{N-1,m-n}))^j (F(a^{i-j}_{N-1,m-n}))^{i-j} \end{aligned} \tag{13}$$

$$a^i_{N,m} = \sum_{m=0}^{m} a^i_{N,m}(n) p_{N,m}(n) \tag{14}$$

Corollary 1 $\{a^i_{N,m}\}_{i=1,2,\cdots}$ および $\{a^i_{N,m}(n)\}_{i=1,2,\cdots}$ は i に関して減少する数列である。

3 Markov chain の場合

ここでは観測できる確率変数がそれぞれの期によって、その確率変数の従う分布関数が変化する場合について考える。ここでは transition probability matrix が $\mathbf{P} = (p_{ss'})_{s,s'=1,2,\cdots}$ であるような Markov chain $\{Z_t\}_{t=1,2,\cdots}$ を考え、それぞれの期に観測することができる確率変数がこの Markov chain の state に依存するものとする。この場合は各々の確率変数は、互いに独立であるが、その期の state には依存する同一の確率分布に従うものとする。また、それらの確率変数がどの期に出現するかは各々互いに独立であり、その確率は一様であるものとし、それぞれの期の state とは独立であるとする。いま、残り計画期間 N 期の間に、m 個の確率変数 $\{X_{is_i}\}_{i=1,2,\cdots,m,s_i \in \{1,2,\cdots\}}$ が出現しそれら確率変数を観測することができるとする。ここで s_i は、i 番目の確率変数が出現した期の Markov chain の state を表すものと考える。それらの m 個の確率変数の観測値の中から k 個を選択する。目的は total expected reward を最大にするような optimal policy と、そのもとでの total expected reward を求めることを考える。ここでは、それらの確率変数がどの期に出現する

かは各々互いに独立であり、その確率は一様であるものとし、それぞれの期におけるMarkov chainのstateとは独立であると考え、したがって$\{p_{N,m}(n)\}_{n=0,1,2,\cdots,m}$は、(1)式によって与えられる。

いま、state spaceを$\{1,2,\cdots\}$、transition probability matrixを$\mathbf{P}=(p_{ss'})\{s,s'=1,2,3,\cdots\}$とする。このMarkov chainのstateは観測することができ、decision-makerには既知であるものとする。いま、stateが$s(\in\{1,2,\cdots\})$であるとき、n個の確率変数を観測した場合を考えると、これらのn個の確率変数$\{X_{is}\}_{i=1,2,\cdots,n}$は、state $s\in\{1,2,\cdots\}$に依存するが、互いに独立であり、同一の分布に従うものとする。この分布関数を$F_s(x)$とする。

このとき、残りの計画期間がN、残っている確率変数の数がm、それらの確率変数の観測値の内から選択できる数がk、Markov chainのstateがsであるとき、(N,m,k,s)をこの問題の状態と呼び、$P_{N,m,k,s}$で表す。また状態が(N,m,k,s)のとき、n個の確率変数が観測できるという条件付きの部分問題を$P_{N,m,k,s}(n)$で表し、それら観測値が$x_{(1)},\cdots,x_{(n)}$であるときの部分問題を、$P_{N,m,k,s}(n;x_{(1)},\cdots,x_{(n)})$で表す。このとき、これらの問題$P_{N,m,k,s}$、$P_{N,m,k,s}(n)$および$P_{N,m,k,s}(n;x_{(1)},\cdots,x_{(n)})$において最適に振る舞ったときに得られるtotal expected rewardをそれぞれ$v_{N,m,k,s}$、$v_{N,m,k,s}(n)$および$v_{N,m,k,s}(n;x_{(1)},\cdots,x_{(n)})$で表せば、次の最適方程式を満足することがわかる。

$$
\begin{aligned}
v_{N,m,k,s} &= \sum_{n=0}^{m} v_{N,m,k,s}(n)p_{N,m}(n) && (15)\\
v_{N,m,k,s}(n) &= \mathrm{E}[v_{N,m,k,s}(n;X_{(1)},\cdots,X_{(n)})] && (16)\\
v_{N,m,k,s}(n;x_{(1)},\cdots,x_{(n)}) &= \max_{1\le i\le k}\{\sum_{j=1}^{i} x_{(j)}+\sum_{t=1}^{\infty} p_{st}v_{N-1,m-n,k-i,t}\} && (17)
\end{aligned}
$$

これらの最適方程式を満足する解を求めるために、次のような数列$\{b^i_{N,s,m}\}_{i=1,2,\cdots}$、$\{b^i_{N,s,m}(n)\}_{i=1,2,\cdots}$および$\{c^i_{N,s,m}\}_{i=1,2,\cdots}$を帰納的に定義する。$(s\in\{1,2,\cdots\}),0\le n\le m)$

$$
\begin{aligned}
b^i_{N,s,m} &= \sum_{n=0}^{m} b^i_{N,s,m}(n)p_{N,m}(n) && (18)\\
b^i_{N,s,m}(n) &= U_n(c^i_{N-1,s,m-n},c^{i-1}_{N-1,s,m-n}|1,\infty) && (19)\\
b^i_{N,s,m}(0) &= c^i_{N-1,s,m} && (20)\\
c^i_{N-1,s,m} &= \sum_{t=1}^{\infty} p_{st}b^i_{N-1,t,m} \qquad (N\ge 2) && (21)
\end{aligned}
$$

とする。ただし、$b^0_{N,s,m}=b^0_{N,s,m}(n)=\infty$, $b^i_{0,s,0}=0$とする。またここで確率分布関数に関して、次の2つの仮定を設ける。ここで、SはMarkov chainのstateを表す確率変数とする。

Assumption 1 Markov chain の state が s であるとき、この state s に依存した確率変数の条件付き期待値$\mu_s = \mathrm{E}[X|S = s]$ は有界とする。また、確率分布関数 $\Pr\{X \le x|S = s\} = F_s(x)$ は絶対連続であり、確率密度関数 $f_s(x)$ を持つものとし、$dF_s(x) = f_s(x)\,dx$ とおく。

Assumption 2 もし、$t < s\,(s, t = 1, 2, \cdots)$ であれば $x \le y$に対して、

$$f_t(y)f_s(x) \ge f_s(y)f_t(x) \text{ すなわち } \begin{vmatrix} f_s(x) & f_t(x) \\ f_s(y) & f_t(y) \end{vmatrix} \ge 0 \tag{22}$$

を満足する。

このときつぎの性質が成り立つ。

Lemma 1 仮定 2のもとで、Markov chain の任意の state $s(s = 1, 2, \cdots)$ に対して、state が s であるという条件のもとでの確率変数 Xの期待値を$\mu_s = \mathrm{E}[X|S = s]$ とおけば、$\{\mu_s\}_{s=1,2,\cdots}$ は s に関して減少する数列である。

Assumption 3 任意の $t, t'(t \ge t', t, t' = 1, 2, \cdots)$ に対して、

$$p_{st'}p_{s't} \ge p_{s't'}p_{st} \text{ すなわち } \begin{vmatrix} p_{s't} & p_{s't'} \\ p_{st} & p_{st'} \end{vmatrix} \ge 0 \tag{23}$$

が　$s \le s'\,(s, s' = 1, 2, \cdots)$ を満足する全ての s と s'に対して成り立つ。

Lemma 2 state に依存する関数 $f(s)$ が s に関して減少する関数であれば、

$$\sum_{t=1}^{\infty} p_{st} f(t)$$

も s に関して減少する関数である。

Lemma 3 (18) 式から (21) 式で定義した、$\{b^i_{N,s,m}\}_{s=1,2,\cdots}$,$\{b^i_{N,s,m}(n)\}_{s=1,2,\cdots}$, $\{c^i_{N,s,m}\}_{s=1,2,\cdots}$ は、s に関して減少する数列である。

Lemma 4 (18) 式から (21) 式で定義した、$\{b^i_{N,s,m}\}_{i=1,2,\cdots}$,$\{b^i_{N,s,m}(n)\}_{i=1,2,\cdots}$, $\{c^i_{N,s,m}\}_{i=1,2,\cdots}$ は、i に関して減少する数列である。

Proposition 4 問題 $P_{N,m,k,s}$における optimal policy は次のように表せる。

すなわち 、この計画期間が Nであるような問題の最初の期で n 個の確率変数を観測し、それらの順序統計量が $x_{(1)}, \cdots, x_{(n)}$ であるとする。いま jを $x_{(j)} \ge c^{k-j+1}_{N-1,s,m-n}$ かつ $1 \le j \le k \wedge n$ を満足する最大の数、すなわち $x_{(j+1)} < c^{k-j}_{N-1,s,m-n}$ または $j = k \wedge n$ を満足する最大の数とするとき、optimal policy は大きい方から j個の値、すなわち $x_{(1)}, \cdots, x_{(j)}$を select することである。

Proposition 5 $v_{N,m,k,s}$および $v_{N,m,k,s}(n)$ は、つぎのように求められる。

$$v_{N,m,k,s} = \sum_{i=1}^{k} b^i_{N,s,m}, \quad v_{N,m,k,s}(n) = \sum_{i=1}^{k} b^i_{N,s,m}(n) \tag{24}$$

Lemma 5 (15) 式から (17) 式で定義した、$\{v_{N,m,k,s}\}$, $\{v_{N,m,k,s}(n)\}$ は、s に関して減少する。

Proposition 6 数列 $\{b^i_{N,s,m}\}_{i=1,2,\cdots,s=1,2,\cdots}$ および $\{b^i_{N,s,m}(n)\}_{i=1,2,\cdots,s=1,2,\cdots}$ は、

$$\begin{aligned} b^i_{N,s,m} &= \sum_{j=1}^{m \wedge i} \int_{c^{i-j+1}_{N-1,s,m-n}}^{c^{i-j}_{N-1,s,m-n}} x_{(i)} g_{M,j}(x_{(j)}) dx_{(j)} \\ &+ \sum_{j=0}^{n \wedge (i-1)} c^{i-j}_{N-1,s,m-n} \, {}_nC_j (1 - F(c^{i-j}_{N-1,s,m-n}))^j (F(c^{i-j}_{N-1,s,m-n}))^{i-j} \end{aligned} \tag{25}$$

$$b^i_{N,m} = \sum_{m=0}^{m} b^i_{N,m}(n) p_{N,m}(n) \tag{26}$$

のように求めることができる。

4 Partially Observable Markov Chain の場合

4.1 Partially Observable Markov Chain

Markov chain に従って変化する state に依存する確率変数の列を観測することができる場合の optimal selection problem について考える。ただしここでは、その Markov chain の state を直接知ることができず、その state についての部分情報を持っている場合について扱う。

そこで、$\{Z_t;\ t = 1, 2, \cdots\}$ を state space が $\{1, 2, 3, \cdots\}$ である Markov chain とし、その transition probability matrix を $\mathbf{P} = (p_{ss'})\{s, s' = 1, 2, \cdots\}$ とする。このとき、state が何であるかについての情報は、state space $\{1, 2, \cdots\}$ 上の確率分布によって与えられているものとする。すなわち、$\mathcal{S} = \{\mathcal{P} | \mathcal{P} = (p_1, p_2, \cdots), p_s \geq 0, \sum_{s=1}^{\infty} p_s = 1\}$ に含まれる確率分布 $\mathcal{P}$によって表されるものと考える。

ここで、それぞれの期に観測できる確率変数は、state に依存するから、これらの確率変数の観測値を通してその state に関する情報が得られる。したがって、1 つも確率変数を観測できないときは、何も情報を得ることはできない。

いま、Markov chain の state に関しての prior information が $\mathcal{P}(\in \mathcal{S})$ によって表され、残りの決定期間が Nであり、その間に m 個の確率変数を観測することがで

き、それらの中から k個の観測値を選択することができるとき、$(N, m, k, \mathcal{P})$ をこの optimal selection problem の状態といい、$P_{N,m,k,\mathcal{P}}$ と表す。

このとき、m 個の確率変数の中から、partially observable Markov chain の state s に依存した n 個の確率変数 $\{X_{is}\}_{i=1,\cdots,n}$ の観測値 $\{x_i\}_{i=1,\cdots,n}$ を得る。ここで state $s \in \{1, 2, \cdots\}$ を直接知ることはできない。そこで、それら n 個の確率変数の観測値から state に関して学習を行いその結果を $\overline{T(\mathcal{P}, \mathbf{x})}$ と表すことにする。そのとき n 個の観測値の中から i 個を選択し、残りのものは選択せず次の期に進む $(i = 0, 1, \cdots, n)$。そして次の期の状態は $(N-1, m-n, k-i, \overline{T(\mathcal{P}, \mathbf{x})})$ と変化する。ここで $\overline{T(\mathcal{P}, \mathbf{x})}$ は、n 個の確率変数 $\{X_{is}\}_{i=1,\cdots,n}$ の観測値 $\{x_i\}_{i=1,\cdots,n}$ を得たときの posterior information であり、ベイズの定理を用いて求められる。((25) 式と (26) 式)

いま、n 個の確率変数の観測値 $\{x_i\}_{i=1,\cdots,n}$ を得たとき、これら観測値 $\{x_i\}_{i=1,\cdots,n}$ ($x_i \in R = (0, \infty), i = 1, \cdots, n$) に対してベイズの定理により事後確率分布が

$$\left\{\begin{array}{ll} T_t(\mathcal{P}, \mathbf{x}) & = \dfrac{p_t f_t(\mathbf{x})}{\sum_{s=1}^{n} p_s f_s(\mathbf{x})} \\ T(\mathcal{P}, \mathbf{x}) & = (T_1(\mathcal{P}, \mathbf{x}), T_2(\mathcal{P}, \mathbf{x}), \cdots) \end{array}\right. \tag{27}$$

と求められる。($t = 1, 2, \cdots$) 以下では、$\mathbf{x} = (x_1, \cdots, x_n)$ と表すことにする。また、$f_t(\mathbf{x})$ は、Markov chain の state が $t(t = 1, 2, \cdots)$ であるとき n 個の確率変数 $\mathbf{X}$ の同時分布関数とする。すなわち、

$$f_t(\mathbf{x}) = \prod_{i=1}^{n} f_t(x_i) \tag{28}$$

である。次に、transition probability matrix $\mathbf{P}$ に従って state が推移するから、次の期の始めでの prior information $\overline{T(\mathcal{P}, \mathbf{x})}$ は、

$$\left\{\begin{array}{ll} \overline{T_t(\mathcal{P}, \mathbf{x})} & = \sum_{s=1}^{n} T_s(\mathcal{P}, \mathbf{x}) p_{st}, \\ \overline{T(\mathcal{P}, \mathbf{x})} & = (\overline{T_1(\mathcal{P}, \mathbf{x})}, \overline{T_2(\mathcal{P}, \mathbf{x})}, \cdots) \end{array}\right. \tag{29}$$

と求められる。

つぎに、information 全体の集合 $\mathcal{S}$ につぎのような尤度比を用いた順序を導入する。この順序は likelihood ratio ordering と呼ばれ Nakai [3]・[5]・[6] などでいろいろ性質が知られている。

Definition 1 $\mathcal{S}$ に含まれる任意の $\mathcal{P}$ と $\mathcal{Q}$ に対して、$\mathcal{P} >_l \mathcal{Q}$ であるとは、全ての s と t ($t \le s, s, t = 1, 2, \cdots$) に対して

$$p_s q_t \le p_t q_s, \text{すなわち} \begin{vmatrix} q_s & q_t \\ p_s & p_t \end{vmatrix} \ge 0 \tag{30}$$

が成立し、少なくとも一つの s と t の組み合わせに対して、$p_s q_t < p_t q_s$, が成り立つ場合を言う。もし、$p_s = q_s$、が任意の $s = 1, 2, \cdots$ に対して成り立つとき $\mathcal{P} =_l \mathcal{Q}$ であるとする。また、$\mathcal{P} \geq_l \mathcal{Q}$ であるとは、$\mathcal{P} =_l \mathcal{Q}$ かつ、$\mathcal{P} >_l \mathcal{Q}$ が成り立つことを言う。

ここで、$\mathcal{P}$上の関数 $u(\mathcal{P})$ が、定義 1 によって定めた尤度比順序に関して増加関数であるとは、$\mathcal{P} \geq_l \mathcal{Q}$ となる$\mathcal{P}$と $\mathcal{Q}$ に関して $u(\mathcal{P}) \geq u(\mathcal{Q})$ を満足するときを言うことにする。

確率分布関数に関してさらに次の仮定を設ける。

Assumption 4 $l \leq m < k$, であれば、 $x_{m\,k} = \sup\{x | f_m(x) \leq f_k(x)\}$, となる $x_{m\,k}$で、次の条件を満足するものが存在する。

(1) $x > x_{m\,k}$ ならば $f_m(k) > f_k(x) > f_{k+1}(x) > \cdots$ であり、

(2) $x \leq x_{m\,k}$ ならば $f_k(x) \geq f_m(x) \geq f_{m-1}(x) \geq \cdots \geq f_1(x)$ である。

仮定 1 から仮定 4 のもとで次の 2 つの性質が成り立つ。(Nakai [3] · [5] · [6])

Lemma 6 今、$\{f_s(x)\}_{s=1,2,\cdots}$ を、仮定 4 の性質を満足するような、確率密度関数の列と考える。次に、$\{a_s\}_{s=1,2,\cdots}$ を、$\sum_{s=1}^{\infty} a_s = 0$ を満足する実数列とする。

このとき、適当な $1 \leq k < n$ が存在し、 $a_s \geq 0, t \leq k$ かつ、$a_t \leq 0, t > k$ であるとする。以上の仮定のもとで、いま $g(x) = \sum_{s=1}^{\infty} a_s f_s(x)$ とおけば、

$$\int_0^{\infty} h(x) g(x) dx \geq 0$$

が成り立つ。

Lemma 7 定義 1 において定義した、$\mathcal{S}$における順序は半順序である。

Definition 2 2 つの k個の観測値の組 $\mathbf{x}, \mathbf{y} \in R^n$ に対して $x_{(i)} \leq y_{(i)}, (i = 1, 2, \cdots, k)$ であるとき、またそのときに限り $\mathbf{x} \prec \mathbf{y}$と表す。

Lemma 8 $\mathbf{x} \prec \mathbf{y}$を満たす任意の $\mathbf{x}$ と $\mathbf{y}$に対して

$$f_t(\mathbf{y}) f_s(\mathbf{x}) \geq f_s(\mathbf{y}) f_t(\mathbf{x}) \text{ すなわち } \begin{vmatrix} f_s(\mathbf{x}) & f_t(\mathbf{x}) \\ f_s(\mathbf{y}) & f_t(\mathbf{y}) \end{vmatrix} \geq 0 \tag{31}$$

$t < s$ ($s, t = 1, 2, \cdots$) が成り立つ。

Lemma 9 $\mathbf{x} \prec \mathbf{y}$を満たす任意の $\mathbf{x}$ と $\mathbf{y}$に対して $T(\mathcal{P}, \mathbf{x}) \leq_l T(\mathcal{P}, \mathbf{y})$ が全ての $\mathcal{P} \in \mathcal{S}$に対して成り立つ。

Proposition 7 $\mathbf{x} \prec \mathbf{y}$を満たす任意の$\mathbf{x}$と$\mathbf{y}$に対して$\overline{T(\mathcal{P},\mathbf{x})} \leq_l \overline{T(\mathcal{P},\mathbf{y})}$である。($\mathcal{P} \in \mathcal{S}$)

Lemma 10 任意の$\mathcal{P},\mathcal{Q}$に対して、$\mathcal{P} \geq_l \mathcal{Q}$ならば、$T(\mathcal{P},\mathbf{x}) \geq_l T(\mathcal{Q},\mathbf{x})$である。($\mathbf{x} \in R^n$)

Proposition 8 任意の$\mathcal{P},\mathcal{Q}$に対して、$\mathcal{P} \geq_l \mathcal{Q}$ならば、$\overline{T(\mathcal{P},\mathbf{x})} \geq_l \overline{T(\mathcal{Q},\mathbf{x})}$である。($\mathbf{x} \in R^n$)

4.2 Partially Observable Markov Chain の上での Optimal Selection Problem

ここでは partially observable Markov chain の上での optimal selection problem について考える。すなわち、それぞれの期に観測することができる確率変数が、ある Markov chain の state に依存する場合であるが、この state について直接知ることはできず、あらかじめ prior information のみがわかっている場合を考える。ただし、それぞれの期にいくつの確率変数を観測できるかは、この state とは独立であるものとする。このような問題について、optimal policy とその政策のもとで最適に振る舞て得られる total expected reward の性質について考えることにする。

いま、残り計画期間をN期とし、この決められた期間の間にm個の確率変数$\{X_{is_i}\}_{i=1,2,\cdots,m,s_i\in\{1,2,\cdots\}}$を観測することができるものとする。ここで$s_i$は、$i$番目の確率変数が観測されたときの state を表すものとし、またiは観測される順番を表すものではないとする。ここで考える問題においても、これまで同様それぞれの確率変数は他の確率変数とは独立に一様な確率で観測されるものと考える。一方、state space を$\{1,2,\cdots\}$、transition probability matrix を$P=(p_{ss'})$とし、この state が$s \in \{1,2,\cdots\}$であるとき、n個の確率変数を観測できるならば、これらのn個の確率変数$\{X_{is}\}_{i=1,2,\cdots,n}$は、state $s \in \{1,2,\cdots\}$に依存するが、互いに独立同一の分布に従うものとし、それらの分布関数は既知のものであるとするする。

この optimal selection problem の状態が$(N,m,k,\mathcal{P})$であるものを$P_{N,m,k,\mathcal{P}}$で表す。またこの状態で、n個の確率変数を観測できるという部分問題を$P_{N,m,k,\mathcal{P}}(n)$で表し、それらの観測値が$x_{(1)},\cdots,x_{(n)}$であるという部分問題を、$P_{N,m,k,\mathcal{P}}(n;x_{(1)},\cdots,x_{(n)})$で表す。このとき、この問題の目的は、total expected reward を最大にするような optimal policy と、そのもとで最適に振る舞ったときに得られる total expected reward を求めることを考える。

問題の状態が$(N,m,k,\mathcal{P})$であるとき、n個の確率変数$\{X_i\}_{i=1,\cdots,n}$を観測することができ、それらの観測値$\{x_i\}_{i=1,\cdots,n}$を得たとする。このとき、decision-maker はそれらのn個の確率変数の観測値の中からi個($0 \leq i \leq n$)を選択し、次の期に進む。このとき次の期のこの問題の状態は$(N-1,m-n,k-i,\overline{T(\mathcal{P},\mathbf{x})})$とな

る。ここで$\overline{T(\mathcal{P},\mathbf{x})}$は、$n$ 個の確率変数 $\{X_i\}_{i=1,\cdots,n}$を観測し、それら n 個の観測値 $\{x_i\}_{i=1,\cdots,k}$をもとにして得られる state についての posterior information を表し、(25) 式 と (26) 式で求められる。

このとき、$P_{N,m,k,\mathcal{P}}$、$P_{N,m,k,\mathcal{P}}(n)$ および $P_{N,m,k,\mathcal{P}}(n;x_{(1)},\cdots,x_{(n)})$ において、optimal policy に従って最適に振る舞ったときに得られるこれらの問題の total expected reward をそれぞれ $v_{N,m,k,\mathcal{P}}$、$v_{N,m,k,\mathcal{P}}(n)$ および $v_{N,m,k,\mathcal{P}}(n;x_{(1)},\cdots,x_{(n)})$ とすると、よく知られているように、次の最適方程式を満足する。

$$v_{N,m,k,\mathcal{P}} = \sum_{n=0}^{m} v_{N,m,k,\mathcal{P}}(n)p_{N,m}(n) \tag{32}$$

$$v_{N,m,k,\mathcal{P}}(n) = \mathrm{E}[v_{N,m,k,\mathcal{P}}(n;X_{(1)},\cdots,X_{(n)})] \tag{33}$$

$$v_{N,m,k,\mathcal{P}}(n;\mathbf{x}) = \max\{\sum_{j=1}^{k} x_{(j)} + v_{N-1,m-n,k-i,\overline{T(\mathcal{P},\mathbf{X})}}\} \tag{34}$$

数列 $\{d^i_{N,\mathcal{P},m}\}_{i=1,2,\cdots}$、$\{d^i_{N,\mathcal{P},m}(n)\}_{i=1,2,\cdots}$ および $\{e^i_{N,\mathcal{P},m}\}_{i=1,2,\cdots}$ を帰納的に次のように定義する。$(j\in\{1,2,\cdots\}, 0\le n\le m)$

$$d^i_{N,\mathcal{P},m} = \sum_{n=0}^{m} d^i_{N,j,m}(n)p_{N,m}(n) \tag{35}$$

$$d^i_{N,\mathcal{P},m}(n) = \mathrm{E}[d^i_{N,\mathcal{P},m}(n;\mathbf{X})] \tag{36}$$

$$d^i_{N,\mathcal{P},m}(n;\mathbf{x}) = U_n(e^i_{N-1,\overline{T(\mathcal{P},\mathbf{X})},m-n}, e^{i-1}_{N-1,\overline{T(\mathcal{P},\mathbf{X})},m-n}|1,\infty) \tag{37}$$

$$d^i_{N,\mathcal{P},m}(0) = e^i_{N-1,\overline{\mathcal{P}},m} \tag{38}$$

ただし、

$$e^i_{N-1,\mathcal{P},m} = d^i_{N-1,\overline{T(\mathcal{P},\mathbf{X})},m} \qquad (N\ge 2) \tag{39}$$

$$\overline{\mathcal{P}} = (\overline{p}_1,\overline{p}_2,\cdots) \tag{40}$$

$$\overline{p}_s = \sum_t p_s p_{st} \tag{41}$$

とする。また、$d^0_{N,\mathcal{P},m} = d^0_{N,\mathcal{P},m}(n) = \infty$, $d^i_{0,\mathcal{P},0} = 0$ であるとする。また、$\mathrm{E}_{\mathcal{P}}$は state space $\{1,2,\cdots\}$ の上の分布 $\mathcal{P}$に関する期待値を表すものとする。

このとき、つぎの性質が成り立つことがわかる。ただし、ここで $a_{\mathcal{P}}$が$\mathcal{P}$に関して減少するとは、すなわち $\mathcal{P}\le_l \mathcal{Q}$ であれば $a_{\mathcal{P}}\ge_l a_{\mathcal{Q}}$が成り立つことを言うものとする。

Lemma 11 (30) 式から (34) 式で定義した、$d^i_{N,\mathcal{P},m}$, $d^i_{N,\mathcal{P},m}(n)$, $e^i_{N,\mathcal{P},m}$ は、$\mathcal{P}$ に関して減少する。

Lemma 12 (30) 式から (34) 式で定義した、$\{d^i_{N,\mathcal{P},m}\}_{i=1,2,\cdots,m}$, $\{d^i_{N,\mathcal{P},m}(n)\}_{i=1,2,\cdots,m}$, $\{e^i_{N,\mathcal{P},m}\}_{i=1,2,\cdots,m}$ は、i に関して減少する数列である。

Proposition 9 問題 $P_{N,m,k,\mathcal{P}}$ における optimal policy は次のように表せる。

すなわち、この計画期間が N であるような問題の最初の期で n 個の確率変数を観測し、それらの順序統計量が $x_{(1)},\cdots,x_{(n)}$ であるとする。いま j を $x_{(j)} \geq e^{k-j+1}_{N-1,\mathcal{P},m-n}$ かつ $1 \leq j \leq k \wedge n$ を満足する最大の数、すなわち $x_{(j+1)} < e^{k-j}_{N-1,\mathcal{P},m-n}$ または $j = k \wedge n$ を満足する最大の数とするとき、optimal policy は大きい方から j 個の値、すなわち $x_{(1)},\cdots,x_{(j)}$ を select することである。

Proposition 10 $P_{N,m,k,\mathcal{P}}$ において、値 $v_{N,m,k,\mathcal{P}}$ および $v_{N,m,k,\mathcal{P}}(n)$ は、

$$v_{N,m,k,\mathcal{P}} = \sum_{i=1}^{k} d^i_{N,\mathcal{P},m}, \quad v_{N,m,k,\mathcal{P}}(n) = \sum_{i=1}^{k} d^i_{N,\mathcal{P},m}(n) \tag{42}$$

を満足する。

Lemma 13 (27) 式から (29) 式で定義した、$v_{N,m,k,\mathcal{P}}$, $v_{N,m,k,\mathcal{P}}(n)$ は、$\mathcal{P}$ に関して減少する。

参考文献

[1] M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, New York, 1970.

[2] G. Monahan, Optimal Stopping in a Partially Observable Markov Processes with Costly Information, *Operations Research*, vol. 28, pp. 1319–1334, 1980.

[3] T. Nakai, The Problem of Optimal Stopping in a Partially Observable Markov Chain, *Journal of Optimization Theory and Applications*, vol. 45, pp. 425–442, 1985.

[4] T. Nakai, An Optimal Selection Problem with a Random Number of Applicants per Period, *Operations Research*, vol. 34, pp. 478–485, 1986.

[5] T. Nakai, A Sequential Stochastic Assignment Problem in a Partially Observable Markov Chain, *Mathematics of Operations Research*, vol. 11, pp. 230–240, 1986.

[6] T. Nakai, A Stochastic Ordering and Related Sequential Decision Problems, *Journal of Information & Optimization Sciences* vol 11, 49–65, 1990.

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