

ある sequential assignment problem について

大阪府立大学総合科学部 中井 達

1. Introduction

S を state space, P を stochastic transition function とする stationary Markov chain を (S, P) とするとき. この Markov chain 上の sequential stochastic assignment problem $(A, S, P, r, \{P_0\}_{s \in S}, \beta)$ を考える. 但し A は action space であり, 二つの disjoint set $\{a_1, \dots, a_m\}$ と $\{p\}$ の和としてあらわせる. また全ての state $s \in S$ に対して non-negative random variable X_0 の値を観測することが出来る. この確率変数は独立で 分布関数 $F_0(x)$ は既知とする. $r(a, x)$ は 観測値 x に対し action a を用いたときの reward とし $E r(a, X_0) < \infty$ であると仮定する. β は discount factor であり $0 \leq \beta \leq 1$ とする.

現在まで考えられている sequential stochastic assignment problem と同様に. 次の形の sequential assignment problem を考える. ([1][3][4][5][7]). まず position (A, ρ) を現在の state が ρ であり. そのときの action space が A_ρ である状態とする. position が (A, ρ) であるとき. 確率変数 X_0 の

実現値 x_t を観測して action a_t を A_n の中から選ぶ。もし a_t が $\{a_1, \dots, a_n\}$ のうちの a_i であるならば, immediate reward $r(a_i, x_t)$ を得る。このとき次の stage での position は (A_{n+1}, t) である。 $A_{n+1} = \{a_1, \dots, a_i, a_{i+1}, \dots, a_n\}$ でありこれは Markov chain (S, P) の next stage である。また a_t が action "p" に等しければ, immediate reward は 0 であり, 次の stage での position は (A_n, t) である。即ち action "p" は pass することと同じであり, その回数には制限はない。以上の様な場合の, total expected discounted reward を最大にする問題を考える。ここで action " a_i " を取ることは即ち sequential assignment problem における action a_i を観測値 x_t に assign することと等しい。

2. Problem formulation.

Sequential stochastic assignment problem $\{A_n, S, P, r\}$ for $\beta \in (0, 1)$ において, 現在の position が $(A_n, 0)$ であるとき, optimal strategy の T での value を $U(A_n, 0)$, また, 観測値 x を知ったという条件付きの, この problem の value を $U(A_n, 0|x)$ とすれば, よく知られた dynamic programming formulation により次の recursive equation が得られる。

$$U(A_n, 0) = E_0 U(A_n, 0 | X_0)$$

(1)

$$U(A_n, 0|x) = \max_{1 \leq i \leq n} \{ \max_{a_i \in A_n} \{ r(a_i, x) + \beta \int U(A_{n+1}, t) P_0(dt) \},$$

$$\beta \int U(A_n, \omega) P(\omega, d\omega) \gamma$$

但し、ここにおいて、 $A_n = \{a_1, \dots, a_n \in U\} / p$, $A_n' = \{a_1, \dots, a_{n-1}, a_{n+1}, \dots, a_n \in U\} / p$ とする。(1)式と同じく、よく知られた, Yarkov decision process と同様に, optimal strategy の存在とその value の存在が示される。ここでは以下, optimal strategy と value が求められる様に, reward function に条件を付ける。即ち、 $r(a, x) = r_1(a)g(x)$ とあらわせる場合を考える。また、state が ω であるときの確率変数を X_0 の代りに $r_2(X_0)$ を考えることにより $r(a, x) = r(a)x$ と考える。また A_n に対し $r(a_1) \geq \dots \geq r(a_n)$ を仮定する。このとき、optimal strategy 及びその value について次の事が成り立つ。

Proposition 1. 全ての $\omega \in S$ に対して次の reward function の列 $\{g_i(\omega)\}_{i=1, \dots}, \{h_i(\omega)\}_{i=1, 2, \dots}$ が存在する。即ち、

$$g_1(\omega) \geq g_2(\omega) \geq g_3(\omega) \geq \dots \geq g_n(\omega) \geq \dots \geq 0$$

$$h_1(\omega) \geq h_2(\omega) \geq h_3(\omega) \geq \dots \geq h_n(\omega) \geq \dots \geq 0$$

であり、position (A_n, ω) において次の事が成り立つ。

1) 観測値の値が x であるときの optimal strategy は、

$$g_i(\omega) \leq x < g_{i-1}(\omega) \text{ のとき } i\text{-th action } a_i \text{ を選ぶ } (i=1, \dots, n)$$

$$0 \leq x < g_n(\omega) \text{ のとき action "p" を選ぶ事である。}$$

2) position (A_n, ω) のときの value は

$$U(A_n, \omega) = \sum_{i=1}^n r(a_i) h_i(\omega) \text{ である。}$$

3) $\{g_i(\omega)\}_{i=1,2,\dots}$ 及び $\{h_i(\omega)\}_{i=1,2,\dots}$ は次の recursive equation によって求めることが出来る。

$$g_i(\omega) = \beta \int h_i(t) P(\omega, dt).$$

$$h_i(\omega) = \int_{g_{i-1}(\omega)}^{\infty} g_{i-1}(\omega) dF_0(x) + \int_{g_i(\omega)}^{g_{i-1}(\omega)} x dF_0(x) + \int_0^{g_i(\omega)} g_i(\omega) dF_0(x),$$

$$\text{かつ } h_1(\omega) = \int_{g_1(\omega)}^{\infty} x dF_0(x) + \int_0^{g_1(\omega)} g_1(\omega) dF_0(x), \quad g_1(\omega) = \beta \int h_1(t) P(\omega, dt)$$

$$g_0(\omega) = \infty.$$

さて、次に $P_0 \in P(T_0 \leq t) = P(\omega, t)$ であるような確率変数とする。今 ω' が Λ に収束するとし、 $X_{\omega'}$ 及び $T_{\omega'}$ がそれぞれ X_{ω} 及び T_{ω} に確率収束すると仮定する。この仮定の下で次の事が成り立つ。

Corollary 1 $g_i(\omega)$ ($\omega \in S, i=1,2,\dots$) は S に関して連続である。

以下では上の事を仮定して話を進める。

3. Special case ($S \subset \mathbb{R}^n \times \mathbb{R}'$)

この節で、stochastic transition function P が、現在の state ω のみならず、観測値 x の値に depend する場合を考える。即ち前節の definition に従えば、 $S \subset \mathbb{R}^n \times \mathbb{R}'$ で、 $S \ni (\omega, x)$ と考える。また、 $dF_{(\omega, x)}(y) = I_x$ (I_x は indicator function) かつ、 $P((\omega, x), (t, y)) = P(\omega, x, t) F_t(y)$ である場合と相当する。この問題では、Proposition 1 における関数 $g_i(\omega, x)$ が、よ

り簡単な式で表わることが出来る。但し、ここには次の様な仮定を設ける。即ち $P(o, x, t)$ は x に関して stochastically increasing かつ、 $\partial P(o, x, t) / \partial x^2 \geq 0$ 、更に、 $H_0(x)$ は x に関して stochastically increasing であるとする。

今、 $U(A_n, o) = E V(A_n, t_0, X_0)$ かつ $U(A_n, o | x) = U(A_n, (o, x))$ と定義すれば、これらの値は、optimal strategy の下での value 及び、条件付の value であり、(1)式を用いて次の recursive equation を得ることが出来る。

$$U(A_n, o) = \int_0^\infty U(A_n, o | x) dH_0(x)$$

$$(2) \quad U(A_n, o | x) = \max \left\{ \max_{1 \leq i \leq n} \{ V(a_i) x + \beta \int U(A_{n-1}^i, t) dP(o, dt) \}, \beta \int U(A_n, t) dP(o, dt) \right\},$$

但し $A_n = \{a_1, \dots, a_n\} \cup \{p\}$ かつ、 $A_n^i = \{a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n\} \cup \{p\}$ 。

このとき次の事が成り立つ。

Proposition 2. 全ての $o \in S$ に対して次の非負関数列 $\{t_n(o)\}_{n=1,2,\dots}$ が存在する。

$$t_1(o) \geq t_2(o) \geq \dots \geq t_n(o) \geq \dots \geq 0.$$

このとき、position (A_n, o) における optimal strategy 及び value は次の形にあるわせる。

1) 観測値が x のとき optimal strategy は

$t_i(o) \leq x < t_{i+1}(o)$ ($i=1, \dots, n$) のとき i -th action a_i を選ぶ

$0 \leq x < t_1(o)$ のとき action "p" を選ぶ事である。

2) value $u(A_n, s)$ は

$$u(A_n, s) = \sum_{i=1}^n \pi(a_i) t_i(s) \text{ である。}$$

3) $\{t_i(s)\}_{i=1,2,\dots}$ は次の recursive equation を満足する。

$$t_n(s) = \int_{d_{n+1}, s}^{\infty} t_{n+1}(s, x) dF_0(x) + \int_{d_{n+1}, s}^{d_{n+1}, s} x dF_0(x) + \int_0^{d_{n+1}, s} t_n(s, x) dF_0(x)$$

$$t_n(s, x) = \beta \int t_n(t) P(s, x, dt),$$

但し $t(s) = \infty$, $-\infty \cdot 0 = 0 \cdot \infty = 0$ かつ d_{n+1}, s は $t_n(s, x) = x$ の unique root とする。

ここで、前節の Proposition 1 で述べられた $\{g_i(s, x)\}_{i=1,2,\dots}$ 及び $\{h_i(s, x)\}_{i=1,2,\dots}$ と Proposition 2 における $\{t_i(s)\}_{i=1,2,\dots}$ 及び $\{t_i(s, x)\}_{i=1,2,\dots}$ の関係は、次の様にあることが来る。

$$t_n(s) = \int h_n(s, x) dF_0(x)$$

$$\begin{aligned} t_n(s, x) &= \beta \int t_n(t) P(s, x, dt) = \beta \iint h_n(t, y) dF_t(y) P(s, x, dt) \\ &= g_n(s, x) \end{aligned}$$

上の Proposition 2 に関連して次の二つの Corollary を得ることが出来る。

Corollary 2. $t_i(s)$ は s に関して増加関数である。

Corollary 3. $t_i(s, x)$ は x に関して増加関数でありまた s に関して凹関数である。

4. Variants and example

4.1 2節で考えた問題を N -stage problem として考えれば.

Proposition 1 と同様の結果が得られ. recursive equation により逐次計算して optimal strategy 及び value を得ることが出来る.

Corollary 4. 全ての state $o \in S$ に対し. 関数列 $\{g_{i,N}(o)\}_{i=1, \dots, N}$ 及び $\{h_{i,N}(o)\}_{i=1, \dots, N}$ の組が存在する.

$$g_{1,N}(o) \geq g_{2,N}(o) \geq \dots \geq g_{N-1,N}(o) \geq g_{N,N}(o) = 0.$$

$$h_{1,N}(o) \geq h_{2,N}(o) \geq \dots \geq h_{N-1,N}(o) \geq h_{N,N}(o) = 0.$$

position が $N(A_n, o)$ であるとき次の事が成り立つ。(但し $N(A_n, o)$ は. 残り stage の数を N で. position が (A_n, o) である時を言う。またそのときの value を $U_N(A_n, o)$ であるとする。)

1) 報酬値の値が α であるときの optimal strategy は.

$$g_{i,N}(o) \leq \alpha < g_{i+1,N}(o) \text{ のとき } i\text{-th action } a_i \text{ を選ぶ } (i=1, \dots, n)$$

$$0 \leq \alpha < g_{n,N}(o) \text{ のとき action "p" をとる ことである。}$$

2) value $U_N(A_n, o)$ は

$$U_N(A_n, o) = \sum_{i=1}^n r(a_i) h_{i,N}(o) \text{ である。}$$

3) $\{g_{i,N}(o)\}_{i=1, \dots, N}$ 及び $\{h_{i,N}(o)\}_{i=1, \dots, N}$ は次の recursive equation を満たす。

$$g_{1,N}(o) = \beta \int h_{1,N}(x) P(o, dx)$$

$$h_{i,N}(o) = \int_{g_{i-1,N-1}(o)}^{\infty} g_{i-1,N-1}(o) dF_0(x) + \int_{g_{i-1,N-1}(o)}^{g_{i,N-1}(o)} x dF_0(x) + \int_0^{g_{i,N-1}(o)} g_{i,N-1}(o) dF_0(x)$$

但し $R_{1,N}(w) = \int_{g_{1,N-1}(w)}^{\infty} x dF_0(x) + \int_0^{g_{1,N-1}(w)} g_{1,N-1}(w) dF_0(x)$
 $g_{1,N}(w) = \beta \int R_{1,N}(t) P(w, dt)$. である。

4.2 $W \in \text{space } \Omega$ の値をとる parameter, X を実数値確率変数とし $f(x|w)$ を $W=w$ であるときの X の conditional g.p.d.f. とする。 $g(w)$ を W の prior g.p.d.f. とし, $g(w)$ と、観測値 $x \in \mathbb{R}^1$ に対して $g(w|x)$ を W の posterior g.p.d.f. とする。

現在まで m 個の値 $X_1=x_1, \dots, X_m=x_m$ を observe したときの posterior distribution を $G_m(w|x_1, \dots, x_m)$ として次の仮定を設ける。(但し $G_0(w) = G(w)$ とする。) 1) $\mu_w = E[X|w] < \infty (w \in \Omega)$ かつ $\mu = E[\mu_w] < \infty$. 2) 与えられた $x_1, \dots, x_m$ に対して w に対する充分統計量 $\bar{x}_m$ が存在し 3) $\bar{x}_m$ は $\bar{x}_{m-1}$ 及び x_m により recursively に生成される。 4) $E[W|\bar{x}_m]$ は $\bar{x}_m$ の non-decreasing function であり 5) $P(X_m \geq x | \bar{x}_{m-1})$ は $\bar{x}_{m-1}$ に関して stochastically increasing とする。

このとき 3 節と同様に、次の様子を sequential stochastic assignment problem と考える。 $A_n = \{a_1, \dots, a_n \in \mathcal{A}\}$, $S = N \times \mathbb{R}^1 \Rightarrow (m, \bar{x}_m)$, $r(a, x) = r(a) \cdot x$ ($r(a_1) \geq \dots \geq r(a_n)$), $\bar{H}(m, \bar{x}_m) = H(x|\bar{x}_m)$ かつ $P((m, \bar{x}_m), x, (m', y)) = I_{(m+1, \bar{x}_{m+1}(\bar{x}_m, x))}$, 但し $\bar{x}_{m+1}(\bar{x}_m, x)$ は $\bar{x}_m$ 及び w , $(m+1)$ stage における観測値 x により生成される充分統計量とする。 このとき Proposition 2

により optimal strategy 及び value が得られる。この結果は Nakai [3] における解と一致する。

4.3. Random termination である場合の sequential stochastic assignment problem を考える。この場合の problem は $\{A_n, S, P, r, \{T_{s^*+t}\}, \beta\}$ であらわされ、特に $S = S_1 \cup S^*$, $P(T = S^* | S^*) = 1$, かつ $P(X_{s^*+t} = 0) = 1$, 但し T は $P(T \leq t) = P(S^* + t)$ である確率変数である。また S^* は absorbing state であり, $S^* + t$ の観測値の値は常に 0 をとると考える。この場合、absorbing state S^* に入れば、reward は常に 0 であり、そのときこの問題は終わると考えられる。

4.4. この sequential stochastic assignment problem を game 的に扱う事も考えられる。即ち、 n 人の player が、逐次出現する観測値から m 個を選び、それぞれに、action space A_n の m 個の action を取る場合を考える。但し player I は、観測値の選択と同時に、どの action をとるかという選択を合わせて行ない、player II は、観測値の選択のみを行なう。但しこのとき、両 player 共に、ここで考えた problem と異なり、pass する回数は制限されているものとする。また player I は maximizing player として player II は minimizing player として行動をとるものとする。このゲームについては、Nakai [5] に詳しく述べられていて、同様に optimal strategy

及 U value を得る ことが出来る。

References

- [1] C. Derman, G. J. Lieberman and S. M. Ross "A sequential stochastic assignment problem" *Management Science* vol 18, p349-355, 1972.
- [2] G. H. Hardy, J. E. Littlewood and G. Polya "Inequalities" Cambridge University Press, 1934.
- [3] T. Nakai "Optimal assignment for a random sequence with an unknown parameter" *Journal of Information & Optimization Science*, vol 1, p214~228, 1980.
- [4] T. Nakai "Sequential stochastic assignment problem with rejection" *Journal of Information & Optimization Science*, vol 2, p169-180, 1981.
- [5] T. Nakai "A time sequential game related to the sequential stochastic assignment problem", *Journal of Operations Research Society of Japan*, vol 25, No. 2, 1982
- [6] H. Sakaguchi "A sequential assignment problem for randomly arriving jobs", *Rep. Stat. Appl. Res. JUSE*, vol 19, p99~109, 1972.