数理解析研究所講究録
第 680 巻 1989 年 174-182

部分観測可能なセミマルコフ決定問題について

神戸大学教養部　中井　達 (Tōru Nakai)

1. Introduction

状態空間を $\{0,1,2,\cdots\cdots\}$ とし、$Q_{ij}(t)$ を state が i のとき j へ移って時刻 t までに transition の起こる確率とする regular な semi - Markov process における部分観測が可能な決定問題を考える。ここで、部分観測が可能であるとは、この semi - Markov process の state が何であるかを直接観測することが出来るのではなく、各 state に depend する分布が既知の確率変数を観測し、その実現値と予め得られている事前分布 (事前知識) をもとにして学習を行ないその結果を基ずいて決定を行なうことを言う。

ここでは、次のような決定問題を考える。この問題の決定期間を T とし、この定められた期間内で、考えている semi - Markov process の transition が起こる都度、その state に depend した確率変数の実現値を観測してその値をもとに決定を行なう。このとき取りうる決定の回数は最大 N 回と定められているものとする。ここで、決定を取ることは即ち各 state に depend した確率変数の実現値を観測して stop する事を意味し、そのときの利得は観測した実現値 x の関数 $u(x)$ として表わされるものとする。ここでは一般性を失う事なく $u(x)=x$ として考える。また stop は N 回のうち 1 回のみ決定することが出来るものとする。

いま、各 state i に depend する確率変数を X_i とし、X_i は絶対連続であると仮定しその密度関数を $f_i(x)$ とおく。(ここで X_i が絶対連続と仮定したが、そうでない場合についても同様の議論をすることが出来る。) またこれらの確

率変数 X_i は state i にのみ depend し他の時間 T 等に関しては、独立であるものとする。semi - Markov process の状態がどの様であるかについての information は state space 上の probability distribution μ によって表わされているものとする。(一般に state space が $\{0, 1, 2, 3, \cdots\cdot\}$ の様な可算集合でない場合でも、例えば R^1 の compact subset で連結なもの即ち閉区間のようなものであれば、同様の議論を行なうことが出来る。) このとき semi - Markov process の状態に関する情報は、transition が起こる毎にそれまでの state に depend した random variable の実現値を観測し、その値をもとにして学習を行なうものとする。従って値を観測してから transition が起こり、次の transition を待つことになる。ここでは、学習の方法は、通常の事後分布即ち Bayes の定理にしたがって行なうものと考え、その学習のもとでの決定問題の性質について考える。

2. Assumptions

第 1 節で述べたような部分観測が可能な semi - Markov process の上での決定問題を考えるとき Bayes の定理にしたがって学習を行なうときのいろいろな性質について考えるために、次のような仮定を設け、それらの仮定のもとで解析を行なう。まず始めに transition probability $Q_{ij}(t)$ について $q_{ij}(t) = p_{ij}\, g_j(t)$ (但し $dQ_{ij}(t) = q_{ij}(t)\,dt$) の形をしているものと仮定する。すなわち semi - Markov process の状態が i のとき確率 p_{ij} で state j に移り次の transition が起こるまでの時間は、density $g_j(t)$ に従うということを仮定する。より一般の場合については、解析が困難となる。

仮定 1

任意の $i, j\,(i < j)$ 及び $k, l\,(k < l)$ に対して $p_{ik}\,p_{jl} \geqq p_{il}\,p_{jk}$ が成り立つ。

即ち仮定1は (p_{ij}) に対して、TP_2 (totally positivity of order two) を仮定するものである。

仮定2

任意の $i, j\,(i<j)$ 及び $t, s\,(t<s)$ に対して $g_i(t)g_j(s) < g_i(t)g_j(s)$ が成り立つ。

この仮定は、$(g_{ij}(t))$ に関しても TP_2 であることを仮定する。

仮定3

任意の i に対して $E[X_i] < \infty$

仮定4

任意の $i, j\,(i<j)$ 及び $x, y\,(x<y)$ に対して $f_i(x)g_j(y) \geqq f_j(x)f_i(y)$ が成り立つ。

即ち確率変数 X_i の密度関数 $f_i(x)$ に対して同様に TP_2 を仮定するものであり、仮定3は X_i が有界な期待値を持つものと仮定するものである。

仮定5

任意の $i, j\,(i<j)$ に対して次のような性質を持つ x_{ij}^* が存在する。

$x < x_{ij}^*$ ならば $f_i(x) > f_j(x) > f_{j+1}(x) > \cdots\cdots$

$x \geqq x_{ij}^*$ ならば $f_j(x) \geqq f_i(x) \geqq f_{i-1}(x) \geqq \cdots\cdots$

次に state に関する information は state space $S = \{0, 1, 2, 3 \cdots\cdots\}$ 上の probability distribution μ として表わせるものと考える。いま、ある時点で state に関する information が μ であるときその時点から t 時間をへて始めて

transitionが起こり、それまでのstateにdependした確率変数Xの実現値xを観測した時のstateに関するposterior informationを$\bar{T}(\mu, t|x)$で表わすとすれば、$\bar{T}(\mu, t|x)$は次のように表わすことが出来る。

$$T_j(\mu, t) = \sum_{i=0}^{\infty} \mu_i p_{ij} g_j(t)$$

$$\bar{T}_j(\mu, t|x) = \frac{T_j(\mu, t) f_j(t)}{\sum_{j=0}^{\infty} T_j(\mu, t) f_j(x)}$$

但し、$\mu_i = \mathrm{P}(S = i|\ \mu)$ (Sはstateを表わす確率変数)

次にstate space S上の確率分布μ (stateに関するinformation)全体に次のような順序を導入する。この順序はlikeliood ratio orderingである。

定義

S上のprobability distribution μ及びνに対し$\mu < \nu$であるとは、任意の$i, j\,(i<j)$に対して$\mu_i \nu_j \geqq \mu_j \nu_i$が成り立つことである。

簡単な計算より、上で定義された順序は半順序であることがわかる。また$\{\mu\}$上の関数$u(\mu)$を考え、この関数がμに関して増加であるとは、$u(\mu)$が実数値関数であって$\mu > \nu$であれば$u(\mu) \geqq u(\nu)$が成り立つことを言うものとする。$u(\mu)$の様な関数としては、例えば$u(\mu) = E_\mu X$の様な関数を考えればよい。

3. Preliminaries

前節で述べた仮定1－5及び定義のもとで、$\bar{T}(\mu, t|x)$及びそれらに関連した次のような性質が成り立つ。

Lemma 1.

(p_{ij}) に関する仮定１及び $(g_j(t))$ に関する仮定５のもとで次の性質が成り立つ。" 任意の $i, j(i<j), k, l(k<l)$ 及び $t(\geqq 0)$ に対して $q_{ik}(t)q_{jl}(t) \geqq q_{il}(t)q_{jk}(t)$ が成り立つ。"

Lemma 2.

$\bar{T}(\mu, t|x)$ は、μ 及び x に関して non-decreasing な関数である。

Lemma 1 は仮定１及び仮定５より簡単な計算で得られる。Lemma 2 は Nakai [3] において用いられたと同様の方法により得られる。また、同様にして

Lemma 3.

$\bar{T}(\mu, t|x)$ は、t に関して non-increasing な関数である。

Lemma 2 及び Lemma 3 で increasing または decreasing であるとは、ここで定義した順序に関して大小関係があるという意味で得られるものである。また Lemma 3 は $q_{ij}(t)$ に関する仮定を用いて得られる性質である。

Lemma 4.

$\{a_i\}$ が i に関して減少する数列であって $v(i, \mu)$ が i に関して non-increasing かつ μ に関して non-decreasing な関数であれば、

$$u(\mu) = \sum_{i=0}^{\infty} a_i v(i, \mu)$$

は μ に関して nondecreasing な関数である。

Lemma 5.

(p_{ij}) が TP_2 (total positivity of order two) であるとき

$$\sum_{j=k}^{\infty} p_{ij}$$

は i に関して non-increasing な関数である。

Lemma 4 及び Lemma 5 は (p_{ij}) が TP_2 である性質と $\{\mu\}$ の上に定義された順序の性質を利用すれば得られる性質である。

4. Formulation and Solution

残りの計画期間が T、その時点で直接観測することの出来ない semi-Markov process の state に関する事前の情報が μ であるとする。さらに、その時点より時間 t が経て未だに transiiton が起こっていず、まだ残された期間 $T-t$ の間に起こる transition のうち最初の N 回において決定を取ることが出来るとき、この問題の状態を (N, T, t, μ) で表わすことにする。この問題の状態が (N, T, t, μ) のとき最適に振舞って得ることの出来る値の期待値を $v^N(T, t, \mu)$ で表わすとすれば、この関数は次の最適方程式を満足する。

$$v^N(T, t, \mu) = \sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(t)\Delta t E\, max\{X, v^{N-1}(T-t-\Delta t, 0, \overline{T}(\mu, t+\Delta t\,|\,x))\}$$

$$+ (1 - \sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(t)\Delta t)v^N(T, t+\Delta t, \mu)$$

このとき移項・整理をして、$\Delta t \to 0$とすれば

$$\frac{\partial}{\partial t} v^N(T, t, \mu) = \sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(t)\{v^N(T, t, \mu) - \phi^N(T, t, \mu)\}$$

但し、$\phi^N(T, t, \mu) = E \max\{X, v^{N-1}(T-t, 0, \bar{T}(\mu, t\,|\,x))\}$

となる。このとき $v^N(T, t, \mu)$ 及び $\phi^N(T, t, \mu)$ に関して次のような性質が、

前節で仮定した条件のもとで得られる。以下で得られる性質はNに関する帰納法を用いて示すことが出来る。$N=1$の場合は明かであり、$N-1$のとき成り立つものと仮定する。

Proposition 1.

$\phi^N(T,t,\mu)$はμ及びTに関してnon-decreasingな関数であり、tに関してnon-increasingな関数である。

このPropositionは、帰納法の仮定と$\phi^N(T,t,\mu)$の定義より、Lemma 4を用いて示すことが出来る。また次のLemmaもこのPropositionとLemma 4から示される

Lemma 6.

$$\sum_{j}^{\infty} q_{ij}(t)\phi^N(T,t,\mu)$$

はμ及びiに関してnon-increasingな関数であり、またtに関してnon-decreasingな関数である。

Corollary 1.

$$\sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(t)\phi^N(T,t,\mu)$$

はμ及びTに関してnon-decreasingな関数であり、またtに関してnon-increasingな関数である。

次のLemmaは、Proposition 2を導くために必要な性質であり、この性質は$q_{ij}(t)$に関する仮定1より導かれる。

Lemma 7.

$dQ_{ij}(t)=q_{ij}(t)\,dt$とするとき

$$\sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} (Q_{ij}(t) - Q_{ij}(s))$$

はμに関して non - decreasing な関数である。

Proposition 2.

$v^N(T,t,\mu)$はT及びμに関して non - decresing な関数であり、tに関して non - inceasing な関数である。また、Nに関して non - decreasing な関数である。

略証） $v^N(T,t,\mu)$に関する方程式より

$$v^N(T,t,\mu) = \int_t^T \sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(s) \phi^N(T,t,\mu) e^{\sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(t)(Q_{ij}(t) - Q_{ij}(s))} ds$$

が求められ、

$$\sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(s) \phi^N(T,t,\mu)$$

及び

$$\sum_{i}^{\infty} \mu_i \sum_{j}^{\infty} q_{ij}(t)(Q_{ij}(t) - Q_{ij}(s))$$

に関する性質すなわち Corollary 1 及び Lemma 7 を用いて$v^N(T,t,\mu)$に関するこの Proposition を求めることが出来る。

また、最適戦略は、時刻tにおいて transition が起こり、state に depend した確率変数の実現値xを観測した時

$x \geqq v^{N-1}(T-t, 0, \bar{T}(\mu, t|x))$ ならば stop し

$x < v^{N-1}(T-t, 0, \bar{T}(\mu, t|x))$ ならば、次の値を observe することである。

References

1. Karlin, S. and Novikoff, A., " Generalized Convex Inequalities ", Pacific Journal of Mathematics, 1963, vol. 13, 1251 - 1279.
2. Nakai, T., " The Problem of Optimal Stopping in a Partially Observable Markov Chain ", Journal of Optimization Theory and Applications, 1985, vol. 45, 425 - 442.
3. Nakai, T., " A Sequential Stochastic Assignment Problem in a Partially Observable Markov Chain ", Mathematics of Operations Research, 1986, vol. 11, 230 - 240.
4. Ross, S. M., " Applied Probability with Optimization Applications " 1970, Holden Day, San Fransisco, California.
5. Ross, S. M., " Stochastic Processes ", 1983, John - Wiley and Sons, New York, New York.