数理解析研究所講究録
680巻 1989年 137-145

Optimal control of an M/G/1 queue with imperfectly observed queue length

長岡高専　涌田和芳（Kazuyoshi Wakuta）

§1. Introduction

待ち行列の制御問題で，arrival rateの未知な適応制御問題が最近研究されている（cf. Lam Yeh and Thomas [2], Schäl [3]）．これは，情報の欠如した待ち行列の制御問題といえる．ここでは，もう一つの情報の欠如した待ち行列の制御問題，すなわち，queue length（システムの客の数）が不完全にしか観測できない問題を考える．ほとんどの待ち行列の制御問題では，システムの状態はqueue lengthで，決定はその観測にもとづいて行なわれる．しかし，我々の制御問題では，queue lengthはサービスを受けた客がシステムを離れる時点で，観測システムにより生成される観測値を通して不完全にしか観測されない．そして，その時点でのシステムの観測可能な厂史にもとづいて，新しいservice rateが選択される．コストは，通常のservice costとholding cost

の2つとする．問題は，期待合計割引コストを最小にする決定規則を見つけることである．

§2. The control problem

ここで考える待ち行列は $M/G/1$ タイプとする．arrival rate λ は一定で，service time は確率分布 $G_a(t)$, $a \in A$, をもつ．$G_a(t)$ は $A \times R_+$ 上でボレル可測な密度関数 $g_a(t)$ をもつとする．ここで，A はボレル空間で，$R_+ = [0, \infty)$．システムの状態，すなわち，queue length は，サービスを受けた客がシステムを離れる時点で観測システム q により観測される．q は S が与えられたときの M 上の stochastic kernel である．ここで，$S = M = \{0, 1, \cdots\}$ で，S は状態の集合，M は観測値の集合を表わす．q による観測誤差には限界があり，$q(\{i, i\pm1, \cdots, i\pm L\} \mid i) = 1$, $i \in S$ なる正の整数 L が存在すると仮定する．i_n は n 番目の客がシステムを離れたときの queue length とする．このとき，$q(\cdot \mid i_n)$ により観測値 m_n が生成され，制御パラメーター a_n が選択される．システムの初期状態 i_0 に対しては，事前分布 $p \in P(S)$ が与えられているとする．$P(S)$ は S 上の確率測度の集合である．制御パラメーターは，初期時刻とサービスを受けた客がシステムを離れる時刻で選択される．これらを decision epoch と呼ぶ

ことにする．もし，decision epoch に客がいなければ，制御パラメーターを選択する意味はないが，このときも制御パラメーターは選択され，しかし，その選択は何も引き起こさないとする．そのような状況では，新しい客の到着時刻を次の decision epoch とみなし，それまでの時間を便宜上 service time と呼ぶことにする．そこで，n 番目の客の service time を t_n で表わし，$T_0 = 0$，$T_n = t_1 + \cdots + t_n$ とすると，T_n は n 番目の decision epoch を表わす．コストは，service cost b_a，$a \in A$ と1人につき h の割合で holding cost があるとし，b_a は A 上有界なボレル可測関数とする．

ある decision epoch で，システムの状態は i ($i \geq 0$) にあり，制御パラメーター $a \in A$ が選択され，t 時間後にサービスが終了するまでに $(j-i+1)$ 人の新しい客が到着するとする．このとき，期待割引コストは次のように表わされる．

$$l(i,a,j,t) := \begin{cases} b_a + ih\int_0^t e^{-\alpha s}ds + h\sum_{k=1}^{j-i+1}\left[\int_0^t\int_\tau^t e^{-\alpha s}ds\,dF^k(\tau)\right], & i \geq 1 \\ 0, & i = 0 \end{cases}$$

ここで，α は正数で，割引率を表わし，F^k は指数分布 $F(t) = 1 - e^{-\lambda t}$ の k-fold convolution を表わす．policy は $\omega =$

$\{\omega_0, \omega_1, \cdots\}$ で，各 ω_n は $H_n := P(S) \times M \times (A \times R_+ \times M)^n$ が与えられたときの A 上のボレル可測な stochastic kernel である．問題は，期待合計割引コストを最小にする policy を見つけることであり，そのような policy を最適な policy という．

§3. Formulation as an imperfect state information semi-Markov decision process

§2 で述べた制御問題は，状態情報の不完全な semi-Markov 決定過程（ISISMDP）$(S, M, A, q, g, c, \alpha)$ として定式化される．ここで，q は $S \times A$ が与えられたときの $S \times R_+$ 上のボレル可測な stochastic kernel で，状態の推移法則を表わし，次のようなボレル可測な密度関数 k をもつ．$k(j,t|i,a) = g_a(t) e^{-\lambda t} (\lambda t)^{j-i+1} / (j-i+1)!$ $(i \geq 1, j \geq i-1)$；$k(j,t|i,a) = 0$ $(i \geq 1, j < i-1)$；$k(j,t|i,a) = \lambda e^{-\lambda t}$ $(i=0, j=1)$；$k(j,t|i,a) = 0$ $(i=0, j \neq 1)$．c は次のように定義される $S \times A$ 上のボレル可測関数で，期待コスト関数を表わす．

$$c(i,a) := \begin{cases} \int_0^\infty \sum_{j \geq i-1} l(i,a,j,t) \left[e^{-\lambda t} (\lambda t)^{j-i+1} / (j-i+1)! \right] dG_a(t), & i \geq 1 \\ 0, & i = 0. \end{cases}$$

policy ω に対する期待合計割引コストは,

$$J_\omega(p) := E_{\omega,p}\left[\sum_{n=0}^{\infty} e^{-\alpha T_n} c(i_n, a_n)\right], \quad p \in P(S)$$

と定義される. ここで, $E_{\omega,p}$ は, ω, p, q, g により定義される $S \times M \times (A \times S \times R_+ \times M)^\infty$ 上の確率測度による期待値を表わす. すべての policy ω とすべての事前分布 $p \in P(S)$ に対して $J_{\omega^*}(p) \leq J_\omega(p)$ が成り立つとき, ω^* は最適であるという.

Lemma 3.1. 次のような正数 β と γ ($\gamma \geq 1$) が存在する.

$$|c(i,a)| \leq \beta i + \gamma, \quad i \in S, \ a \in A.$$

コスト関数の非有界性を処理するために, 次の仮定をおく.

Assumption (I). 次のような正数 ρ ($0 < \rho < 1$) が存在する.

$$\sum_{j \in S} \int_0^\infty e^{-\alpha t} w(j+2L)\, k(j, t \mid i, a)\, dt \leq \rho\, w(i), \quad i \in S, \ a \in A.$$

ただし, $w(i) = \beta i + \gamma \geq 1$, $i \in S$.

Assumption (II). 事前分布 $p \in P(S)$ は有限な期待値をもつ.

§4. Reduction to a perfect state information semi-Markov decision process

§3で定式化された ISISMDP は，状態の事後分布を新しいシステムの状態とする状態情報の完全な semi-Markov 決定過程 (SMDP) $(P, A, \overline{q}, \overline{c}, \alpha)$ に変換される (cf. Wakuta [4]). ただし，$P := \{ p \in P(S) : p(\{i, i\pm1, \cdots, i\pm L\} \cap S) = 1 \text{ for some } i \in S \}$. この SMDP での policy を I-policy と呼び，その期待合計割引コストを $I(\pi)(p_0)$, $p_0 \in P$ と表わす.

Lemma 4.1. 任意の $p \in P$ と $a \in A$ に対して次のことが成り立つ.

(i) $|\overline{c}(p,a)| \leq \overline{w}(p)$

(ii) $\iint_{P\times R_+} e^{-\alpha t} \overline{w}(p')\, d\overline{q}((p',t)|p,a) \leq \rho\, \overline{w}(p)$

ただし，$\overline{w}(p) := \sum_{i\in S} w(i)p(i)$, $p \in P$.

Proposition 4.1. 任意の I-policy π に対して，$\|I_\pi\| \leq 1/(1-\rho)$. ただし，$\|v\| := \sup_{p\in P} |v(p)|\, \overline{w}(p)^{-1}$.

Proposition 4.2 (cf. Bertsekas and Shreve [1]). I-policy π^* が SMDP $(P, A, \overline{q}, \overline{c}, \alpha)$ に対して最適ならば，それは ISISMDP $(S, M, A, q, g, c, \alpha)$ に対しても最適である.

§5. Optimal stationary I-policies

我々の制御問題に対する最適方程式は次のようになる.

$$v(p) = \inf_{a \in A} \left\{ \bar{c}(p,a) + \iint_{P \times R_+} e^{-\alpha t} v(p') \, d\bar{q}((p',t) \mid p,a) \right\}, \quad p \in P$$

これは次のようにもかくことができる.

$$v(p) = \inf_{a \in A} \left\{ \bar{c}(p,a) + \sum_{m \in M} \int_0^\infty e^{-\alpha t} v(u(p,a,t,m)) z(p,a,t,m) \, dt \right\}, \quad p \in P.$$

ここで, $z(p,a,t,m) := \sum_{i \in S} \sum_{j \in S} g(m \mid j) k(j,t \mid i,a) p(i)$. u は $P \times A \times R_+ \times M$ から P へのボレル可測写像である (cf. Wakuta [4]). 上式の右辺のカッコ内を $v(p,a)$ で表わし, 次のようにオペレーターを定義する. ただし, f は P から A へのボレル可測写像である.

$$T_a v(p) := v(p,a), \quad a \in A, \ p \in P.$$

$$T v(p) := \inf_{a \in A} T_a v(p), \quad p \in P.$$

$$T_f v(p) := v(p, f(p)), \quad p \in P.$$

Assumption (III). (i) A はコンパクト距離空間である.
(ii) $g_a(t)$ は $A \times R_+$ 上で有界で, 各 $t \in R_+$ に対して A 上連続である.
(iii) b_a は A 上有界で, l.s.c. である.

Lemma 5.1. 距離空間 $(S(P), \|\cdot\|)$ は完備である. ただし, $S(P)$ は $\|v\| < \infty$ である P 上の l.s.c. funs. の集合である.

Lemma 5.2. $v \in S(P)$ とすると, $v_n(p) \uparrow v(p)$, $p \in P$ で, $\|v_n\| \leq \|v\|$ であるような P 上の有界連続関数列 $\{v_n\}$ が存在する. 逆に, v が $\|v\| < \infty$ なる P 上の関数で, $v_n(p) \uparrow v(p)$, $p \in P$ なる連続関数列 $\{v_n\}$ が存在すれば, $v \in S(P)$ である.

Lemma 5.3 (cf. Schäl [3]). X は距離空間, (Y, β) は可測空間, μ は (Y, β) 上の測度, b と d は $|b| \leq d$ で, $\int_Y d(x,y)\, d\mu(y) < \infty$, $x \in X$ なる $X \times Y$ 上の関数で, さらに各 $y \in Y$ に対して, b と d は X 上連続で, $\int_Y d(x,y)\, d\mu(y)$ は X 上連続であるとする. このとき, $\int_Y b(x,y)\, d\mu(y)$ も X 上連続である.

Lemma 5.4. Assumption (III) を仮定する. このとき, T は $S(P)$ 上の縮小写像で, 一意の不動点をもつ.

Theorem 5.1. (cf. Wakuta [5]) $I_\pi(p_0)$, $p_0 \in P$ が最適方程式を満たすときに限り I-policy π は $SMDP(P, A, \bar{g}, \bar{c}, \alpha)$ に対して最適である.

Theorem 5.2. Assumption (Ⅲ)を仮定する．このとき，Tの不動点，$v^* \in S(P)$でvをおきかえた最適方程式の右辺を最小にする行動を選択するボレル可測写像f^*が存在し，定常な policy $(f^*)^\infty$が SMDP $(P, A, \bar{g}, \bar{\tau}, \alpha)$に対して最適となる．

References

[1] Bertsekas and Shreve (1978) Optimal Control : The Discrete Time Case. Academic Press.

[2] Lam Yeh and Thomas (1983) Adaptive control of M/M/1 queues — continuous-time Markov decision process approach. J. Appl. Prob. 20. 368–379.

[3] Schäl (1981) Estimation and control in stochastic dynamic programming with applications to queueing models. Preprint.

[4] Wakuta (1987) コンパクトな行動空間をもつ部分的観測セミマルコフ決定過程 数理解析研究所講究録 611. 42-53.

[5] _____ (1987) Arbitrary state semi-Markov decision processes with unbounded rewards. Optimization 18. 447–454.