マルコフ・システムの制御

大野　勝久

1. はじめに

ORにおける代表的な確率システムといえば，在庫管理，待ち行列，信頼性であり，これらはすべてマルコフあるいはその一般化としてのセミ・マルコフ過程として論じられてきた．経済，工学等の他の分野においても，確率的変動が無視できない動的な問題をあつかうさいには，マルコフ過程に帰着させて論ずるのが通例である．このような過程で記述されるマルコフ・システムの制御問題の例として，多品目在庫管理問題を考えよう．

例1．多品目在庫管理問題

問題は，期のはじめに各品目の在庫量をみて，発注するかしないか，発注するとすればどれだけかを与える最適発注政策を求めることである．費用としては，発注費用，品切れ費用，在庫費用を考え，発注時にかかる固定費用は注文する品目の組合せに依存するものとする．したがって，必ずしも(σ, S)政策が最適とはならず，混合発注政策等を考える必要がある[4, 7, 8]．さて，品目数をcで表わし，n期における品目$l(=1,\cdots,c)$の在庫量をx_l^n，需要量をz_l^n，発注量をa_l^nで表わそう．発注品は次期のはじめまでに納品されるものとすれば，$n+1$期のはじめにおける在庫量x_l^{n+1}は，

$$x_l^{n+1}=\max\{x_l^n-z_l^n,0\}+a_l^n \qquad (1)$$

で与えられる．需要量$(z_1^n,\cdots,z_c^n)$が各期独立に分布するものと仮定すれば，x_l^{n+1}の確率分布はx_l^n，a_l^nとz_l^nの周辺分布から計算できる．同様に，需要分布からn期にかかる平均総費用が計算できる．このように現在の状態$(x_1^n,\cdots,x_c^n)$と決定$(a_1^n,\cdots,a_c^n)$が与えられれば，その期にかかる平均総費用と次期の状態$(x_1^{n+1},\cdots,x_c^{n+1})$の確率分布が求められる．

今1つの例として，図1に示される$M/M/c$待ち行列システムの制御問題を考えよう[12]．

例2．$M/M/c$ 待ち行列システムの制御

客は到着率λのポアソン過程にしたがって到着し，各窓口でのサービス時間は指数分布にしたがい，待ち行列長はLに制限されている．問題はシステムの状態に応じて，各窓口$l(=1,\cdots,c)$のサービス率をM_l+1個のレベル$\{0,\mu_l,\cdots,M_l\mu_l\}$から選び，さらに複数の窓口が遊休中（サービス率が正で空）であればどの窓口に客を割り当てるかを与える最適制御政策を求めることである．客の待ち費用，窓口の稼動費用，遊休費用，サービス

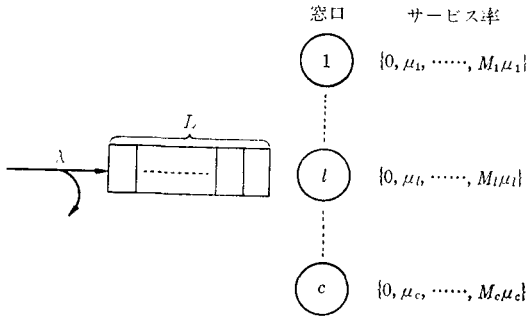

図1　$M/M/c$ 待ち行列システム

おおの　かつひさ　京都大学　工学部

率の切り換え費用を考え，状態として待ち行列長，各窓口の状態（空かどうか）と現在のサービス率[注1]をとる．決定は各窓口のサービス率および客の割り当てである．このとき例1と同様，現在の状態と決定が与えられれば，次に遷移する状態の分布とその遷移までにかかる平均総費用を計算することができる．しかし，この問題における決定はサービス率であり，(1)式の発注量のように直接的に状態を制御できず，サービス分布のパラメータとして確率的に制御できるにすぎない．このことが在庫管理問題との基本的な相違である．

例1，例2を含むマルコフ・システムの制御問題を統一的に論ずる理論が（セミ・）マルコフ決定過程である．以下では実際に最適制御政策を計算するアルゴリズムに焦点を絞り，簡単のためマルコフ決定過程について述べる．セミ・マルコフ決定過程への拡張は有限期間問題を除けば直接的である．また，例1，2とも費用最小化問題であったが，大勢にしたがい利得最大化問題を取りあつかうことにする．

2. 有限期間問題

システムは M 個の状態 $i\in I=\{1,\cdots,M\}$ をとり，各状態 i においてとりうる決定は集合 $A_i=\{1,\cdots,K_i\}$ で与えられている．このとき例1，2におけるように，状態 i で決定 $k\in A_i$ をとったとき次期の状態が j となる確率 $p_{ij}(k)$ およびその期における平均利得 $r_i(k)$ が与えられる．$p_{ij}(k)\geq 0$，$\sum_{j\in I}p_{ij}(k)=1$ である．単位期間当りの利子率 r から $\beta=1/(1+r)$ として決められる $\beta(0<\beta\leq 1)$ を割引率と呼ぶ．n 期のはじめにおける状態を x_n，その時の決定を a_n で表わせば，x_n，a_n が与えられたときの n 期における平均利得 y_n は，

$$y_n=r_x(a_n),\quad \text{ここで } x=x_n$$

となり，その現価換算値は $\beta^n y_n$ である．問題は初期状態 $x_0=i$ が与えられたとき，N期までに得られる総利得

$$\sum_{n=0}^{N}\beta^n y_n$$

の平均を最大にする決定列$(a_0,\cdots,a_N)$を定める最適政策を導くことである．

アルゴリズム[注2]

1：すべての $i\in I$ に対して，

$$v_i^N=\max_{k\in A_i} r_i(k) \qquad (2)$$

を計算し，最大を与える k を f_i^N とおく．$n=N-1$ とおく．

2：$v_i^{n+1}(i\in I)$ の値を用い，各 $i\in I$ に対して，

$$v_i^n=\max_{k\in A_i}\{r_i(k)+\beta\sum_{j\in I}p_{ij}(k)v_j^{n+1}\} \qquad (3)$$

を計算し，最大を与える k を f_i^n とおく．

3：$n=n-1$ とおき，$n<0$ となれば停止．さもなければステップ2へもどる．

v_i^0 が最大平均利得を与え，$f^n=(f_1^n,\cdots,f_M^n)$ とおけば，求める最適政策は$(f^0,f^1,\cdots,f^N)$で与えられる．すなわち，期間0では決定 f_i^0 をとり，n 期$(=1,\cdots,N)$のはじめに状態が j となれば決定 f_j^n をとるのが最適である．この政策は，システムの履歴 $(x_0,a_0,\cdots,x_n)$ を情報として用い，決定を確率的に定める非常に広い政策のなかで最適である．このアルゴリズムの計算量は，

乗算：$MN(M+1)$，加算：M^2N，

大小比較：$(N+1)\{\sum_{i=1}^{M}K_i-M\}$ (4)

であり，必要とするコア・メモリーは$\{v_i^n\},\{v_i^{n+1}\}$の記憶に要する $2M$ である．また最適政策を1つ求めればよいのであれば，最適政策の記憶に $(N+1)M$ のメモリーが必要である．したがって，たとえ連続状態をとるシステムであっても，最適政策の性質がくわしくわかっていないかぎり，それを計算する段階で必然的に有限状態で近似せざるを得ない．

上記アルゴリズムの計算量は N に比例して増えるだけであるが，得られた最適政策を実際用いるためには，$f^0,\cdots,f^N$ の表を必要とし，期間 n が変わるごとに対応する f^n の表を参照して決定を決めなければならない．したがって，計画期間長 N が長い問題では，あたかも永久にシステムをと

りまく現在の状況が維持されるものと考え，$N\to\infty$として得られる最適定常政策を採用するのが実際的であろう．

3. 割引利得問題

計画期間長が無限であり，割引率 $\beta<1$ のとき得られる総利得

$$\sum_{n=0}^{\infty}\beta^n y_n$$

の平均を最大にする最適定常政策$(f^*,f^*,f^*,\cdots)$を求める問題である．最適定常政策のもとでは，どの期間nにおいても$x_n=i$となれば決定f_i^*をとればよく，ただ1つの表f^*を必要とするだけである．この問題に対しては政策反復法，線形計画法，逐次近似法が古典的な結果として知られている[2]．

政策反復法

1：初期政策$f^0=(f_1^0,\cdots,f_M^0)$を与え，$n=0$とおく．

2：(値決定ルーチン)M元連立一次方程式

$$v_i-\beta\sum_{j\in I}p_{ij}(f_i^n)v_j=r_i(f_i^n)\quad(i\in I)\qquad(5)$$

を解き，$v_i(i\in I)$を求める．

3：(政策改良ルーチン)各$i\in I$に対して

$$\max_{k\in A_i}\{r_i(k)+\beta\sum_{j\in I}p_{ij}(k)v_j\}\qquad(6)$$

を与える決定の集合F_i^{n+1}を求める．もし$f_i^n\in F_i^{n+1}$ならば$f_i^{n+1}=f_i^n$とおき，さもなければf_i^{n+1}をF_i^{n+1}の任意の要素にとる．すべてのiで$f_i^{n+1}=f_i^n$となれば停止．f^nは最適定常政策f^*であり，$v_i(i\in I)$は状態iから出発したときの最大割引利得v_i^*を与える．あるiで$f_i^{n+1}\neq f_i^n$となれば，$n=n+1$としてステップ2へもどる．

政策反復法は有限回の反復でf^*を与えることが知られている．このアルゴリズムがN_1回の反復で停止したときの計算量は，ステップ2でガウスの消去法を用いるとすれば，[注3)]

$$\text{乗除算}：N_1\left\{\frac{1}{3}M^3+O(M\sum_{i=1}^{M}K_i)\right\}\qquad(7)$$

$$\text{加減算}：\text{同上，比較}：N_1\left\{\sum_{i=1}^{M}K_i-M\right\}$$

となり，必要とするコア・メモリーはM^2+4Mである．

線形計画法

$\sum_{i\in I}b_i=1$を満たす適当な正数b_iを与え，次のLP問題：

$$\max\ \sum_{i\in I}\sum_{k\in A_i}r_i(k)x_{ik}\qquad(8)$$

$$\text{subject to}\ \sum_{k\in A_i}x_{ik}-\beta\sum_{j\in I}\sum_{k\in A_j}p_{ji}(k)x_{jk}=b_i\quad(i\in I)$$

$$x_{ik}\geq 0\quad(i\in I,\ k\in A_i)$$

を解く．最適解x_{ik}^*は各iに対して唯一の$k\in A_i$で正となり，そのkをf_i^*ととれば最適定常政策f^*が得られる．

LP問題(8)は$\sum_{i=1}^{M}K_i$変数M制約式問題である．また，政策反復法の初期政策f_i^0に対応するx_{ik}を基底変数としてとれば，初期基底実行可能解が得られる．このとき改訂単体法を用いてN_2回で最適解が得られたものとすれば，計算量は[4)]，

$$\text{乗除算}：M^3+O(N_2M\sum_{i=1}^{M}K_i)\qquad(9)$$

$$\text{加減算}：\text{同上，比較}：N_2\{\sum_{i=1}^{M}K_i-M\}$$

となり，基底逆行列をコア・メモリーに保持すれば$M^2+O(M)$が必要となる．線形計画法では反復当り1つの状態の決定しか改良されず，N_2はN_1にくらべて相当大きく，Mのオーダになることを注意しなければならない．

ここで，例1において品目数$c=3$ととり，各品目の在庫の上限を9ととれば，$M=10^3$，$\sum_{i=1}^{M}K_i=10^5$となる．ゆえに政策反復法では千元連立一次方程式を解かねばならず，計算量は(7)式より$(N_1/3)\cdot 10^9$のオーダとなる．一方線形計画法では，10万変数千制約式のLP問題となり，計算量は(9)式より$10^9+O(N_2\cdot 10^8)$のオーダとなる．このように，いずれのアルゴリズムを用いるにせよ，状態数Mが数百以上となれば実際に最適政策を計算するのは容易ではない．

連立一次方程式の数値解法では，数百元以上の問題に対してヤコビ法，ガウス・ザイデル法等の反復解法がよく用いられている．したがって，こ

れら反復解法を値決定ルーチン（ステップ2）で用いる政策反復法が考えられる．しかし，ステップ3では改良された政策が得られる程度の(5)の近似解で十分であり，反復解法を有限回の反復mで打ち切ったほうが効率的であろう．すなわち，ステップ2で反復解法をm回用いる政策反復法が得られる．この方法を修正政策反復法と呼び，特に$m=0$のとき古典的な逐次近似法が得られる．以下のアルゴリズムでは，最大割引利得の上，下限を計算し，それらを用いて最適になり得ない決定をA_iから除去する非最適政策の除去法を併用している．

修正政策反復法［11］

1：初期値$v_i^0(i\in I)$，非負数ε，非負整数mを与え，$A_i^0=A_i$, $l_i=-\infty(i\in I)$, $\xi=0$, $n=0$とおく．

2：各iに対して，

$$w_i^{n+1}=\max_{k\in A_i^n}\{r_i(k)+\beta\sum_{j\in I}p_{ij}(k)v_j^n\} \tag{10}$$

を計算し，同時に最大を与える決定の集合F_i^{n+1}と

$$A_i^{n+1}=\{k\in A_i^n\,;\,r_i(k)+\beta\sum_{j\in I}p_{ij}(k)v_j^n \geq l_i-\beta\xi\}$$

を定める．もし$f_i^n\in F_i^{n+1}$ならば$f_i^{n+1}=f_i^n$とおき，さもなければf_i^{n+1}をF_i^{n+1}の任意の要素にとる．すべてのiでF_i^{n+1}が唯一の要素f_i^{n+1}になればステップ6へ．

3：$z_i^0=w_i^{n+1}$とおき，$l=0,1,\cdots,m-1$に対して

$$z_i^{l+1}=r_i(f_i^{n+1})+\beta\sum_{j\in I}p_{ij}(f_i^{n+1})z_j^l\quad(i\in I) \tag{11}$$

を計算し，$v_i^{n+1}=z_i^m(i\in I)$とおく．

4：$\Delta=\max_{i\in I}\{v_i^{n+1}-v_i^n\}$, $\nabla=\min_{i\in I}\{v_i^{n+1}-v_i^n\}$

$a=\max_{i\in I}\{v_i^{n+1}-w_i^{n+1}\}$, $b=\min_{i\in I}\{v_i^{n+1}-w_i^{n+1}\}$

$$\xi=((\beta\Delta-b)/1-\beta) \tag{12}$$

$$\eta=\max\{\beta\nabla-a)/(1-\beta),\ b\beta^m/(1-\beta^m)\} \tag{13}$$

を計算し，$\xi-\eta<2\varepsilon$となればステップ6へ．さもなければ，$l_i=v_i^{n+1}+\eta(i\in I)$，$n=n+1$とおいてステップ2へもどる．

5：最大割引利得v_i^*のε近似値が，

$$v_i=v_i^{n+1}+\frac{1}{2}(\xi+\eta)\quad(i\in I) \tag{14}$$

で与えられ，f^{n+1}はε'最適政策である．ここで，

$$\delta=\max_{i\in I}\{v_i-r_i(f_i^{n+1})-\beta\sum_{j\in I}p_{ij}(f_i^{n+1})v_j\} \tag{15}$$

$$\varepsilon'=\varepsilon+\{\delta/(1-\beta)\}. \tag{16}$$

6：f^{n+1}は唯一の最適定常政策である．

$c=\max_{i\in I}\{w_i^{n+1}-v_i^n\}$, $d=\min_{i\in I}\{w_i^{n+1}-v_i^n\}$

を計算し，$\varepsilon''=\beta(c-d)/(1-\beta)$とおけば，

$$v_i=w_i^{n+1}+\{\beta(c+d)/2(1-\beta)\}(i\in I) \tag{17}$$

は最大割引利得v_i^*のε''近似値を与える．

このアルゴリズムは任意の$v_i^0(i\in I)$と$\varepsilon>0$に対して有限回の反復で停止し，ε'最適政策（ステップ5）あるいは唯一の最適政策f^*（ステップ6）を与える．特に最適政策が唯一であることがあらかじめ知られていれば，$\varepsilon=0$ととることで有限回の反復でf^*を求めることができる．修正政策反復法がN_3回の反復で停止したときの計算量は，$|A_i^n|$でA_i^nの要素数を表わすことにすれば注5）

$$\begin{aligned}
&\text{乗算}\quad:(M+1)\{\sum_{n=1}^{N_3}\sum_{i=1}^{M}|A_i^{n-1}|+mN_3\}+o(M)\\
&\text{加減算}:M\{\sum_{n=1}^{N_3}\sum_{i=1}^{M}|A_i^{n-1}|+(m+3)N_3\}+o(M)\\
&\text{比較}\quad:2\sum_{n=1}^{N_3}\sum_{i=1}^{M}|A_i^{n-1}|+3MN_3+o(M)
\end{aligned} \tag{18}$$

となる．必要なコア・メモリーはv_i^n等に$4M$，A_i^nのために$\sum_{i=1}^{M}K_i$ビット必要である．たとえば例1で$M=10^3$，$\sum_{i=1}^{M}K_i=10^5$の場合，計算量は反復の初期では10^8であるが終期では10^3のオーダにまで低下し有力な手法であることが示される．

例2で$c=2$，$L=10$，$M_1=M_2=4$の場合$M=740$，$\sum_{i=1}^{M}K_i=4980$となるが，FACOM M−200による計算結果では，$m=6(12)$のとき25(15)回の反復で唯一の最適政策が16(13)秒で得られた．なお，修正政策反復法では(10)，(11)式でガウス・ザイデル法の反復を用いることもでき，$m=6(12)$のとき13(9)回の反復で9(8)秒に短縮された．mを反復ごとに自動的に調整する修正政策反復法が，例2と同様な直列型待ち行列システムの制御問題に応用されている［9］．

4. 時間平均利得問題

問題は無限期間における単位時間当り利得

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{n=0}^{N}y_n$$

の平均（ゲインと呼ばれる）を最大にする最適定常政策を求めることである．

政策反復法

1：初期政策 f^0 を与え，$n=0$ とおく．

2：(値決定ルーチン) g_i^n, $v_i^n(i\in I)$に関する次の $2M$ 元連立一次方程式を解く．

$$g_i^n-\sum_{j\in I}p_{ij}(f_i^n)g_j^n=0\quad(i\in I)\tag{19}$$

$$v_i^n-\sum_{j\in I}p_{ij}(f_i^n)v_j^n=-g_i^n+r_i(f_i^n)\ (i\in I)\tag{20}$$

3：(政策改良ルーチン)各 i に対して，

$$G_i^{n+1}=\{\sum_{j\in I}p_{ij}(k)g_j^n \text{ を最大化する } k\in A_i^n\}$$

$$F_i^{n+1}=\{r_i(k)+\sum_{j\in I}p_{ij}(k)v_j^n \text{ を最大化する } k\in G_i^{n+1}\}$$

を定める．$f_i^n\in F_i^{n+1}$ ならば $f_i^{n+1}=f_i^n$ とおき，さもなければ f_i^{n+1} を F_i^{n+1} の適当な要素にとる．すべての i で $f_i^{n+1}=f_i^n$ となれば停止し，そうでなければ $n=n+1$ とおいてステップ2へもどる．

$P(f)$で定常政策 f から定まる遷移確率 $p_{ij}(f_i)$ を i 行 j 列要素とする行列を表わせば，$P(f)$ はマルコフ連鎖を与える．この連鎖のエルゴード(再帰)集合の個数を $e(f)$とおくと，行列$(I-P(f))$の階数は $M-e(f)$となる[注6]．ゆえに(19),(20)で解 g_i^n は一意に決まるものの，v_i^n は $e(f^n)$ 個の自由度をもつ．Howard[5]は各エルゴード集合に属する1つの状態で $v_i^n=0$ とおき，得られた $2M-e(f^n)$元連立一次方程式を解けばよいとしている．しかし，そのためには反復ごとにマルコフ連鎖の状態分類をしなければならず，アルゴリズム[3]は知られているものの決して効率的ではない．この点を改善したものが次の値決定アルゴリズムであり，

$$\Delta g_i=g_i^n-g_i^{n-1},\ \Delta v_i=v_i^n-v_i^{n-1}\quad(i\in I)$$

に関する連立一次方程式を解いている．

値決定アルゴリズム[14]

1：行列 $(I-P(f^n))$ にガウスの消去法等を用い，適当に状態を並べ換えて，

$$L(I-P(f^n))=\begin{pmatrix}Q_1 & Q_2\\ \cdots\cdots & \\ 0 & \end{pmatrix}\tag{21}$$

なると $M\times M$下三角行列L, $(M-e(f^n))\times(M-e(f^n))$上三角行列 Q_1, $(M-e(f^n))\times e(f^n)$行列 Q_2 を求める．以下添字1で Q_1 に対応する $M-e(f^n)$ 次元ベクトルを表わし，添字2で残りの $e(f^n)$ 次元ベクトルを表わすことにする．

2：$$c_1=(P(f^n)g^{n-1})_1-g_1^{n-1}\tag{22}$$

$$d_l=r(f^n)_l+(P(f^n)v^{n-1})_l-g_l^{n-1}-v_l^{n-1}\quad(l=1,2)\tag{23}$$

を計算し，$c_1=d_2=0$ となればステップ4へ．

3：$$Q_1\Delta g_1+Q_2\Delta g_2=c_1,\ (L\Delta g)_2=d_2\tag{24}$$

を解き，その解 $\Delta g(M$次元ベクトル) を用いて，

$$Q_1\Delta v_1=d_1-(L\Delta g)_1\tag{25}$$

を解く．$\Delta v_2=0$ とおきステップ6へ．

4：(23)式の d_1 で要素が正となる状態の集合を S とおき，次の集合[注7]

$$T=\{i\in I;\ \text{ある } n\geq 0 \text{ で } \sum_{j\in S}p_{ij}^n(f^n)>0\}$$

を最短経路を求める Dijkstra のアルゴリズム(たとえば[1])を用いて求める．

5：$i\in T$ に対して $\Delta v_i=0$ とおいて，

$$Q_1\Delta v_1+Q_2\Delta v_2=d_1\tag{26}$$

を解き，$\Delta v_i(i\in T)$ を求める．$\Delta g_i=0(i\in I)$ とおく．

6：$$g_i^n=g_i^{n-1}+\Delta g_i,\ v_i^n=v_i^{n-1}+\Delta v_i(i\in I)\tag{27}$$

とおく．

このアルゴリズムを政策反復法の値決定ルーチンで用いれば有限回の反復で最適定常政策 $f^*=f^n$ および最大ゲイン $g^*=g^n$ が得られる．

線形計画法[6]

$\sum_{i\in I}b_i=1$ を満たす適当な正数 $b_i(i\in I)$ を与え，次のLP問題：

$$\max\quad\sum_{i\in I}\sum_{k\in A_i}r_i(k)x_{ik}\tag{28}$$

$$\text{subject to}\ \sum_{k\in A_i}x_{ik}-\sum_{j\in I}\sum_{k\in A_j}p_{ji}(k)x_{jk}=0\quad(i\in I)$$

$$\sum_{k\in A_i}(x_{ik}+y_{ik})-\sum_{j\in I}\sum_{k\in A_j}p_{ji}(k)y_{jk}=b_i\quad(i\in I)$$

$$x_{ik} \geq 0,\ y_{ik} \geq 0 \quad (i \in I,\ k \in A_i)$$

を解き最適解 x_{ik}^*, y_{ik}^* を求める．最適政策 f_i^* は，

(i) $\sum_{k \in A_i} x_{ik}^* > 0$ ならば $x_{ik}^* > 0$ となる任意の k

(ii) $\sum_{k \in A_i} x_{ik}^* = 0$ ならば $y_{ik}^* > 0$ となる任意の k

として得られる．

政策反復法では M 元連立一次方程式 (24) を解かねばならず，線形計画法では $\sum_{i \in I} K_i$ 変数 $2M$ 制約式のＬＰ問題を解かなければならない．したがって，M が大きな問題では逐次近似法によらざるを得ないものと思われる．

逐次近似法[13]

1：初期値 $v_i^0 (i \in I)$，正数 ε および

$$0 < \tau < \min\{1/(1-p_{ii}(k));\ p_{ii}(k) < 1,\ i \in I,\ k \in A_i\}$$

を満たす τ を適当に与える．$n=0$ とおく．

2：各 i に対して，

$$g_i^{n+1} = \max_{k \in A_i}\{r_i(k) + \sum_{j \in I} p_{ij}(k) v_j^n - v_i^n\} \quad (29)$$

を求め，$\Delta = \max_{i \in I} g_i^{n+1}$，$\nabla = \min_{i \in I} g_i^{n+1}$ を計算し，$\Delta - \nabla < \varepsilon$ となればステップ４へ．

3：$v_i^{n+1} = v_i^n + \tau(g_i^{n+1} - g_M^n)$ $(i \leq M-1)$ (30)

$$v_M^{n+1} = 0$$

とおき，$n = n+1$ としてステップ２へもどる．

4：(29) 式で最大を与える k を f_i とおけば，$f = (f_1, \cdots, f_M)$ は最大ゲイン g^* との差が ε 以下となる ε 最適政策を与える．

このアルゴリズムの収束を保証するには，残念ながら，最大ゲイン g^* が状態に依存しない定数であることを仮定しなければならない．この仮定は f^* のもとでのマルコフ連鎖がただ１つのエルゴード集合をもてば成立し，例１では常に，例２ではサービス率の切り換え費用が低ければ成立する．

時間平均利得問題に対する修正政策反復法は現在のところ公表されていないが，筆者は前節に述べた導出と同じ考え方で，$m=0$ で逐次近似法と一致する修正政策反復法を得ている．この反復法は $e(f^*) \geq 2$ となる一般的な問題へも拡張できる．

5. おわりに

有限状態，有限決定をもつマルコフ決定過程におけるアルゴリズムについて，現在までに得られた結果を紹介した．紙数の関係もあり，引用すべき文献を多数省略してしまったのは心残りである．最後に，この解説では状態が完全に観測できるものとしてきたが，実際の問題では観測が不完全な場合も多く，この場合については最近の解説[10]を参照されたい．

引用文献

[1] エイホ，A.V.他(野崎，野下訳)：アルゴリズムの設計と解析，サイエンス社，1977

[2] Derman, C.: *Finite State Markovian Decision Processes*. Academic Press, New York, 1970

[3] Fox, B.L. and Landi, D.M.: An Algorithm for Identifying the Ergodic Subchains and Transient States of a Stochastic Matrix. *Comm. ACM*, 11(1968), 619-621

[4] Goswick, T.E. and Sivazlian, B.D.: The Mixed Ordering Policy in Periodic Review Stochastic Multi-Commodity Inventory Systems. *Naval Res. Log. Quart.*, 21(1974), 389-410

[5] ハワード，R.A.(関根他訳)：ダイナミック・プログラミングとマルコフ過程，培風館，1971

[6] Hordijk, A. and Kallenberg, L.C.M.: Linear Programming and Markov Decision Chains. *Management Sci.*, 25(1979), 352-362

[7] Johnson, E.L.: Optimality and Computation of (σ, S) Policies, in the Multi-Item Infinite Horizon Inventory Problem. *Management Sci.*, 13(1967), 475-491

[8] Kalin, D.: On the Optimality of (σ, S) Policies, Math. *Operations Res.*, 5(1980), 293-307

[9] 三根，大野，市木：直列型待ち行列システムの最適制御に関する計算アルゴリズムについて，ＯＲ学

会春季アブストラクト集. 175-176, 1982年3月

[10] Monahan, G. E. : A Survey of Partially Observable Markov Decision Processes : Theory, Models, and Algorithms. *Management Sci.*, 28(1982), 1-16

[11] Ohno, K. : A Unified Approach to Algorithms with a Suboptimality Test in Discounted Semi-Markov Decision Processes. *J. Operations Res. Soc. Japan*, 24(1981), 296-324

[12] 大野勝久：待ち行列システムの最適制御に関する計算アルゴリズムについて，数理解析研究所講究録 452, 1-19, 1982年2月

[13] Platzman, L. : Improved Conditions for Convergence in Undiscounted Markov Renewal Programming. *Operations Res.*, 25(1977), 529-533

[14] Spreen, D. : A Further Anticycling Rule in Multichain Policy Iteration for Undiscounted Markov Renewal Programs, *Zeitschriftfür Operations Res.*, 25(1981), 225-233

注)

1) サービス率の切り換え費用の計算に必要である.

2) $p_{ij}(k)$, $r_i(k)$ が n に依存する非定常な場合にも適用できる.

3) $O(M)$ は $M\to\infty$ のとき M と同位の無限大を表わす.

4) 理論的には $O(M^{2.81})$ のアルゴリズムがあるが，実用的ではない[1].

5) $O(M)$ は $M\to\infty$ のとき M の低位の無限大を表わす.

6) I は単位行列を表わす.

7) $p_{ij}^n(f^n)$ は $P(f^n)^n$ の i 行 j 列要素を表わす.

特集に当って

大野 勝久

「確率システム？　知らないナァ」とページをめくり「ウーン」と頭をかかえられた読者もおありかと思いますが，現実に解決をせまられている問題の多くは，多かれ少なかれはっきりしない不確定部分を含んでおります．このような問題を状態あるいは決定変数を用い，不確定部分を確率として定式化したものが確率システムです（少々オーバーないいかたかも知れませんし，また定式化されたからといって現在の理論，計算機で必ず解けるわけでもありませんが……）．自動制御の分野では，「確率システム・シンポジウム」（日本自動制御協会主催）が毎秋，専門の異なった参加者が集まって開かれており，昨秋で13回と回を重ねております．そこで，このシンポジウムをはじめられた砂原先生に，その理論的側面を概説していただき，中溝，片山両先生にメイン・テーマである安定性と推定をお願いいたしました．

一方，ORをながめてみますと，待ち行列，信頼性，在庫管理等確率システムとよぶことのできる各々永い歴史をもった分野があります．研究の進展につれ，専門化，細分化するのは学問の常ですが，現在では相互の交流，批判もなく孤立化しているようにみうけられます．そこでこれら分野で横断的に用いられている（セミ・マルコフ過程をとりあげ，高橋先生に「マルコフ分析」をお願いいたしました．

なお，同様な趣旨でOR学会関西支部に「応用確率論研究部会」（主査，西田俊夫先生）が設けられ，今春から活動しております．ぜひご参加ください．

おわりに，ご多忙のところをご執筆いただいた上記諸先生に厚くお礼申しあげます．

おおの　かつひさ　京都大学