行列解析法によるコールセンターのモデル化

河西　憲一

待ち行列理論 (Queueing Theory) には様々な解析手法が存在する．その一翼を担う行列解析法 (Matrix Analytic Methods) は今日ではシステム性能評価に使われる主要な解析手法である．また，行列解析法は数値計算を意識したアルゴリズム構築を志向している側面があり，その意味では電子計算機の性能が飛躍的に進歩した現代にあっては基本となる解析手法ともいえる．そこで本稿では行列解析法の典型的な応用であるシステムのモデル化と性能評価について述べる．具体的な待ち行列システムとしてコールセンターを取り上げる．行列解析法の中でも扱いやすい準出生死滅過程を中心にその手法の一端を紹介する．

キーワード：行列解析法，準出生死滅過程，コールセンター，リソース設計，性能評価

1. はじめに

日常生活をおくる中で，我々はしばしばサービスを受けるために待ちに遭遇する．待ちはサービスを提供する施設のリソース（資源）を超える負荷が発生した結果として生じる現象である．また，視点を変えると，待ちは施設に加わる負荷の度合いに応じて決まるサービス品質の程度を表しているともいえる．そのような文字通り列をなして並ぶ行列も含め，待ちについて解析する方法論の体系が待ち行列理論である．より細かく捉えればサービスを提供する施設における，(1) リソース，(2) 負荷，(3) サービス品質の3者を定量的に関係づける数理モデルに基づく体系ともいえる．

待ち行列理論の専門書を紐解くと，まず初等的なマルコフ連鎖に基づく方法を学ぶことから始まる．初等的なマルコフ連鎖の方法ではサービス施設内の客数に着目し，その推移の様子を客の到着と退去という事象に注視して観察する．そして「定常状態では入と出の客の流量が確率的にバランスする」という，いわば確率流に関する保存則をもとに系内客数を解析する．確率流の保存則に依拠する方法は定常状態を前提とした待ち行列理論の方法論に共通する考え方である．

待ち行列理論をさらに学び進めていくと，行列解析法と呼ばれる方法論に到達する．行列解析法は今日の待ち行列理論では欠くことのできない方法論としてその地位を確立しているといっても過言ではなく，かつ十分実用にも耐え得る方法論でもある．行列解析法の基本となるアイディアは「相 (phase)」と呼ばれる補助的な変数を導入することにある．主たる変数のみでは1次元空間上での状態の記述に留まるが，主たる変数に補佐する変数を組み込むことで多次元空間に拡張される．考える空間が多次元空間に様変わりしたものの，上手に空間を整理するといわば「1次元的な」空間上の問題に帰着させることができる．ただし，上手に整理した空間上での基本的な構成要素は行列となる．

行列解析法のもう1つの特徴はシステムの性能評価指標を算出する数値的に安定したアルゴリズムが構築できるという側面である．電子計算機が飛躍的な進歩を遂げている現代にあっては，その処理能力を最大限に活用するためにも数学的な性質に裏打ちされた数値的に安定した算術は歓迎すべきものであり，実用に供する方法論としても望ましい．

本稿ではコールセンターというサービスを提供する施設を舞台にして，待ち行列理論における主要な方法論である行列解析法がどのようにその役割を果たし得るのかを探る．はじめにコールセンターのリソース配分問題を1次元空間上での状態記述に制約されているともいえる初等的な待ち行列モデルを基軸に概観する．初等的なモデルから始めることで，1次元空間では表現しきれない課題を浮かび上がらせる．次にその課題を受けて行列解析法を用いた手法に移る．多次元空間に拡張された表現力をもつ行列解析法の特徴がどのように課題解決に活かされるかをみる．

2. コールセンターのリソース設計

コールセンターを運用するために必要なリソースを準備する際に，「どれくらいのリソースを用意すべきか」，という問題に直面する．ここでいうリソースとしては少なくとも2種類を対象とする．すなわち，1つは

かわにし　けんいち
群馬大学理工学研究院
〒376–8515 群馬県桐生市天神町 1–5–1

電話回線であり，もう 1 つはオペレータである．なお，ここでの電話回線とはコールセンターに外部から電話をかけてくる顧客とオペレータをつなぐために設置される外線を指すことにする．外線数は顧客がコールセンターに電話をかける場合のつながりやすさと関係する．一方，オペレータの人数は電話がつながった後の待ちに遭遇する可能性に関係する．当然，外線数が多ければつながりやすくなり，少ないとなかなかつながらない．また，オペレータが多ければ待ちに遭遇する可能性は低くなり，逆もしかりである．これら「つながりやすさ」や「待つ可能性」のようなサービス品質を定量化し，相応のリソースを算出するのがコールセンターのリソース設計である．

2.1 アーラン計算式によるリソース設計

このようなコールセンターにおけるリソース設計では，今日でもアーラン計算式が用いられている．ここでいうアーラン計算式とは待ち行列理論でいうところのアーラン B 式とアーラン C 式である．前者は必要な外線数を求めるために，後者はオペレータの人数を求めるために使われる．コールセンターのリソース設計は大方アーラン計算式で解決する．ただしこのような方法は，アーラン B 式とアーラン C 式を個別に使って外線数とオペレータ人数を決めるため，コールセンターへのつながりやすさと，つながった後に待ちに遭遇する可能性の相関が考慮されない方法でもある．実際よく考えてみると，つながりやすくするために外線をたくさん用意すれば，それに応じて多くの客がコールセンターにつながることになるものの，結果として待つ可能性も上がりそうである．また，アーラン C 式ではコールセンター内に待たせておくことができる顧客数に制限がないことが前提であり，この点は現実とは大きくかけ離れているといわざるを得ない．現実のコールセンターでは外線数を超えた客数を同時に抱えることはできないはずである．

2.2 アーラン計算式を超える

アーラン計算式を用いる標準的なコールセンターのリソース設計にはやや難点があり，改善すべき課題がある．具体的には，つながりやすさと待つ可能性との相関，および外線数の有限性を考慮することが望まれる．これらの視点を取り込むためにはアーラン計算式が基礎とする待ち行列モデルを見直すことが必要である．そのような待ち行列モデルとして M/M/c/K と表される待ち行列モデルが考えられる．この待ち行列モデルをコールセンターに対応づけると次のようになる．

1. 顧客からの問い合わせは率 λ のポアソン過程に従う．
2. オペレータから受けるサービス時間は平均 $1/\mu$ の指数分布に従う．
3. オペレータの人数は c 人である．
4. 外線の数は K 本である．

数理モデルとしての M/M/c/K モデルは状態空間が有限であるようなマルコフ連鎖として記述できる．ここでの状態とは待ち行列システムに滞在する客数で指定可能であり，その結果状態空間は有限な 1 次元空間 $\mathcal{S}_1 = \{0, 1, \ldots, K\}$ となる．確率流の保存則から定常状態で客数が i である確率 $\pi_i\,(i = 0, 1, \ldots, K)$ について漸化式 $\lambda\pi_i = \min(c, i+1)\mu\pi_{i+1}\,(i = 0, 1, \ldots, K-1)$ が導出される．この漸化式と正規化条件 $\pi_0+\pi_1+\cdots+\pi_K = 1$ から，例えば π_c は

$$\pi_c = \frac{a^c}{c!}\left[\sum_{n=0}^{c-1}\frac{a^n}{n!} + \frac{a^c}{c!}\sum_{n=0}^{K-c}\left(\frac{a}{c}\right)^n\right]^{-1},$$

で求まる．ここで，$a = \lambda/\mu$ であり，a は呼量と呼ばれる．さらに漸化式を用いると，π_K は π_c を使って

$$\pi_K = \left(\frac{a}{c}\right)^{K-c}\pi_c,$$

のように求まる．ここで求めた π_K は，PASTA (Poisson Arrivals See Time Averages) [1] という性質のため，コールセンターにつながらない確率でもある．

コールセンターにつながった顧客は即座にオペレータと通話を開始するか，さもなければしばらく待つ．待つ確率は待ち時間を W とすると $W > 0$ となる確率 $P(W > 0)$ で評価でき，次のように算出できる．

$$P(W > 0) = \frac{\pi_c}{1-\pi_K}\sum_{n=0}^{K-1-c}\left(\frac{a}{c}\right)^n.$$

ただし，$K = c$ の場合の和はゼロと解釈する．このように M/M/c/K モデルにおいては，つながりにくさを定量化した π_K と，待つ可能性を定量化した $P(W > 0)$ がお互いに関係づけられることが見てとれる．

2.3 M/M/c/K モデルの注意点

M/M/c/K モデルはアーラン B 式やアーラン C 式が基礎とするそれぞれの待ち行列モデルよりはコールセンターの特徴をよく捉えているものの，適用にあたって注意することがないわけではない．まず第 1 の注意点はコールセンター内の待ち室が最大で $K - c$ に制約されるという前提である．これは $K \geq c$ であることを暗に仮定することになり，したがって M/M/c/K モデルでは外線数よりも多くのオペレータを配置するコールセンターには適用できない．コールセンターにおいて，オペレータの人数が外線数よりも多くなる状

況は特に珍しいことではないだろう．例えば，できるだけ顧客の待ち時間を短くしようと意図するコールセンターを運営する場合，外線数以上にオペレータを配置することは十分考えられる．

さらにもう少し細かく観察すると M/M/c/K モデルではコールセンターにおけるリソースの占有の仕方がうまく表現できていないこともわかる．顧客がオペレータと通話中である状況を考える．このとき，顧客は2種類のリソースを占有している．すなわち，オペレータと外線である．ここで通話が終了したとする．通話が終了するとオペレータと外線が解放される．解放された外線は新たに電話をかけてくる顧客のために利用可能となり，解放されたオペレータは次の顧客と即座に通話を開始する．このような前提では M/M/c/K モデルは上手く表現できている．ところが，実際のコールセンターでは通常オペレータは顧客と通話を終えると，顧客情報の更新のような後処理と呼ばれる作業に取りかかる．オペレータが次の顧客と通話を開始するためには後処理が完了している必要がある．したがって，オペレータが占有される時間は通話時間と後処理時間を合計した時間と考えられる．

このように M/M/c/K モデルのような初等的な待ち行列モデルは細かい視点にまで立ち入ることができないこともあり，十分にシステムの挙動をモデル化できない可能性もある．コールセンターの例でいうならば，M/M/c/K モデルは状態を待ち行列システムに滞在する客数のみで記述するため，2種類のリソース，すなわち外線とオペレータの占有具合を表現するにはこと足りないことに起因している．

コールセンターの場合に限らず，何かしらのシステムのリソース設計においては各々のリソースがどのように配分され，どのように占有されるのかを見極めることが重要である．その上でモデルを構築する必要がある．どのようなモデルを選択するかは目的に応じて異なる．初等的な待ち行列モデルはシステムの状態を1つの変数で観察するので表現力に限界があるが簡易な手法でもあり，目的に適うのであれば躊躇することはない．

3. 行列解析法で解析する

コールセンターのリソース設計に対して M/M/c/K モデルにもその表現力に難点があることを指摘した．M/M/c/K モデルの表現力では不十分であるという立場から他のモデル化を探ることにする．以下，オペレータが c 人配置されているコールセンターに，K 本の外線が用意されているとする．オペレータの人数 c と外線数 K との間には大小関係の制約を設けないことにする．外部から電話をかけてくる顧客は空いている外線があればコールセンターに接続され，待ち行列に並び先着順にサービスを受けるものとする．ここでのサービスとはオペレータと通話することによって受けるサービスとする．サービス終了後，すなわち通話終了後に顧客は外線を解放しコールセンターを退去する．通話終了後にオペレータは後処理を開始する．後処理を終えたオペレータは次の顧客がいれば通話を開始する．

3.1 変数を追加する

時刻 t におけるコールセンターに滞在する客数を $N_1(t)$ と書くことにする．これは時刻 t で占有されている外線数に等しく，M/M/c/K モデルでも状態を記述するために使われる主たる変数である．主たる変数である $N_1(t)$ により外線の占有具合は表現できるが，オペレータの占有具合までは表現しきれない，というのが M/M/c/K モデルの難点であった．例えば，$N_1(t) = c$ である場合，外線は c 本占有されていることは確定できるが，オペレータがどのように占有されているか（通話中かあるいは後処理中なのか）までは確定できない．そこで，$N_1(t)$ を補佐する変数として，時刻 t での後処理中のオペレータ数を表す変数 $N_2(t)$ を導入する．この結果コールセンターの状態をより細かく記述することが可能になる．実際，補佐する変数 $N_2(t)$ と主たる変数 $N_1(t)$ の情報とを組み合わせることで，後処理中のオペレータ数はもちろんのこと，通話中のオペレータ数も確定し，さらに新たな顧客のために待機中のオペレータ数までも確定できる．

例えば $N_1(t) = K$ であり，かつ $N_2(t) = c$ であるならば，c 人のオペレータ全員が後処理中であることを意味し，待機中のオペレータはいない状況を示す．さらに，外線は K 本占有されているのだから，K 人の顧客が待ちの状態にある状況を表す．なお，オペレータの人数と外線数の間には大小関係がないので，M/M/c/K モデルのような $K \geq c$ という制約はなく，$K < c$ という状況もあり得ることに注意する．したがって，例えば $N_1(t) = K$ であり，かつ $N_2(t) = 0$ であるならば，時刻 t では外線が K 本占有されており，後処理中のオペレータはいない状態を表す．外線が K 本占有されているので，この場合は $\min(c, K)$ 人のオペレータが顧客と通話中である状態を意味する．さらに，$K \geq c$ ならばオペレータ全員が通話中であるため待機中のオペレータは存在せず，$K < c$ ならば $c - K$ 人のオペレータが待機中である状況を表すことになる．

3.2 構造的なマルコフ連鎖でモデル化

主たる変数のみではなく補佐する変数を付け加えることで表現力は豊かにはなった．ただし，結果として観察すべき変数が取り得る状態空間は 1 次元空間から多次元空間にまで広がることになる．記述の簡略化のため，主たる変数と補佐する変数の 2 つを組にして $X(t) = (N_1(t), N_2(t))$ と書くことにする．すると，オペレータの人数と外線数との間には大小関係のような制約はなく，互いに無関係に値をとり得るため，$X(t)$ の状態空間は $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ と表すことができる．ここで，$\mathcal{S}_1 = \{0, 1, \ldots, K\}$ は M/M/c/K モデルの場合と同様に主たる変数 $N_1(t)$ が，$\mathcal{S}_2 = \{0, 1, \ldots, c\}$ は補佐する変数 $N_2(t)$ が取り得る状態空間を表す．

仮に，客の到着間隔，オペレータの通話時間，およびオペレータの後処理時間がパラメータこそ違えどもそれぞれが指数分布であるとし，さらにそれらが互いに独立であるならば，$\{X(t); t \geq 0\}$ は連続時間のマルコフ連鎖を形成することは容易に理解できる．このマルコフ連鎖の状態の数は状態空間 $\mathcal{S}$ の要素の個数に等しく，全部で $(c+1)(K+1)$ 個の状態が存在する．よく知られているように，状態の数が有限個であるようなマルコフ連鎖の場合，定常分布を求めることにさほどの難しさはない．ひとたび定常分布が求められれば，M/M/c/K モデルの場合と同様に様々な性能評価指標を少なくとも数値的には手にすることが可能である．コールセンターのリソース設計に限らず，一般にシステムの性能評価に際して待ち行列理論の方法論を適用しようとするとき，マルコフ連鎖によるモデル化の手法は常套手段といってもよい．さらにマルコフ連鎖の状態の個数が有限に止められるのであれば，システム性能評価の問題は格段に易しくなるといえる．

実際に定常分布を求めるためには定常方程式を解けばよい．定常方程式は未知数が状態の数に等しい線形連立方程式で表すことができる．すなわち定常分布をその状態について適当に並べて行ベクトル $\boldsymbol{\pi}$ で表すと，定常方程式は $\boldsymbol{\pi Q} = \mathbf{0}$ と表現することができる．ここで，$\mathbf{0}$ は要素がすべて 0 である行ベクトルである．定常方程式を満足する定常分布は正規化条件 $\boldsymbol{\pi}\mathbf{1}^\top = 1$ のもとで一意に定める．ここで，$\mathbf{1}^\top$ は要素が全て 1 である列ベクトルである．係数行列である $\boldsymbol{Q}$ は推移速度行列と呼ばれ，状態の個数に等しい次元の正方行列で与えられる．この推移速度行列 $\boldsymbol{Q}$ は状態の並べ方を上手に選ぶと次のような行列を基本要素とするようなブロック 3 重対角構造をもつように整理できる．

$$
\boldsymbol{Q} = \begin{bmatrix}
\boldsymbol{A}_1^{(0)} & \boldsymbol{A}_0^{(0)} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\
\boldsymbol{A}_2^{(1)} & \boldsymbol{A}_1^{(1)} & \boldsymbol{A}_0^{(1)} & \ddots & \vdots \\
\boldsymbol{O} & \boldsymbol{A}_2^{(2)} & \ddots & \ddots & \boldsymbol{O} \\
\vdots & \ddots & \ddots & \ddots & \boldsymbol{A}_0^{(K-1)} \\
\boldsymbol{O} & \cdots & \boldsymbol{O} & \boldsymbol{A}_2^{(K)} & \boldsymbol{A}_1^{(K)}
\end{bmatrix}.
$$

ここで，$\boldsymbol{A}_0^{(j)}$ $(0 \leq j \leq K-1)$，$\boldsymbol{A}_1^{(j)}$ $(0 \leq j \leq K)$ および $\boldsymbol{A}_2^{(j)}$ $(1 \leq j \leq K)$ は $(c+1)$ 次元正方行列であり，$\boldsymbol{O}$ は要素がすべてゼロの適当な次元をもつ行列である．推移速度行列を構成するこれらの行列は主たる変数を中心に補佐する変数をまとめてブロック化した結果である．このようなブロックを基本単位として眺めると，状態空間が主たる変数について「1 次元的」ともいえる構造をもつことがわかる．

このような推移速度行列の特殊構造を十二分に活用することで，定常分布を算出する効率的なアルゴリズムが構築できる [2]. アルゴリズムという性格上，モデルを記述するパラメータを用いた定常分布の陽な姿は見えないが，アルゴリズムが示す手順に沿って計算を進めていけば定常分布が得られることになる．その詳細はともかく，例えば推移速度行列のブロック構造にならって定常分布 $\boldsymbol{\pi}$ もブロック単位に分解すると，元々解くべき定常方程式 $\boldsymbol{\pi Q} = \mathbf{0}$ がより規模の小さい，すなわちより小さいサイズの行列で記述される方程式系に還元されることは想像に難くないであろう．一般に行列解析法におけるアルゴリズムにおいて，行列を抜きにした議論は難しいといえる．行列解析法は定常分布なり性能評価指標を少なくとも数値的に評価することが目的の 1 つであるため，扱う行列のサイズを小さくする工夫は実用的にも意義がある．これは電子計算機の処理能力が日進月歩で向上しつつある現代にあっても変わらない．

4. 再呼も考慮する

行列解析法に頼るモデル化を採用するならば，構造化されたマルコフ連鎖に帰着できるように状態空間を上手く構築することが重要である．「主たる変数に対応させる量を何とし，それを補佐する変数をどのように見立てるか」はモデル化したいシステムごとに異なる．また，モデル化の仕方によってそれら変数が走る状態空間もその様相が変容し得る．そのような例として，再びコールセンターを題材に前節とは異なるモデル化を試みる．

待ち行列理論の専門書を読み進めると再試行，あるいは再呼と称する用語を目にする．客がサービス施設に到着したが他の客がサービス中であるため一旦施設を離れ，しばらくした後に再度サービス施設を訪れるという現象のことを指す．コールセンターの文脈でいうならば，コールセンターに電話をかけてみたがつながらず，一旦電話を切りしばらく待ってから再び電話をかけ直す顧客の振る舞いのことである．このような再呼の影響が行列解析法により，どのようにモデル化できるのかを前節のモデルを基礎に考える．

4.1　変数をさらに追加する

再呼はコールセンターに着信できなかったために生じる．着信できるかどうかはコールセンターの混雑具合に依存する．そこで混雑しているかどうかを判断するためにコールセンターの状態を観察する必要がある．前節のモデル化では，コールセンターを滞在する客数 $N_1(t)$ と後処理中のオペレータ数 $N_2(t)$ から構成される $X(t) = (N_1(t), N_2(t))$ で記述した．すなわち前節のモデルを基礎とするならば，$X(t)$ はある時刻 t におけるコールセンターの状態そのものを記述していると考えられる．よって $X(t)$ をまず状態を表す変数として採用する．再呼を考慮するならば明らかに $X(t)$ だけでは情報不足である．そこで再呼を考慮するモデルでは時刻 t で再呼のために待機中である顧客数 $N(t)$ を導入する．例えば $X(t) = (N_1(t), N_2(t)) = (K, c)$ であり，かつ $N(t) = 1$ ならば，K 人の顧客がコールセンターにつながっているが c 人のオペレータ全員が後処理中のため待ちの状態にあって，かつ再呼待機中の顧客が 1 人いる状況を表す．このように変数を追加することで，再呼待ちの顧客の状態と合わせたコールセンターの状態が観察できる．

観察する変数を主たる変数と補佐する変数に色分けし，構造化されたマルコフ連鎖として記述する，というのが行列解析法でのモデル化の考え方である．ここでは 2 つの変数（実際には 3 つの変数）$X(t)$ と $N(t)$ のうち，$N(t)$ を主たる変数，$X(t)$ を補佐する変数として再呼を考慮するモデルを構築する．補佐する変数 $X(t)$ の状態空間は $\mathcal{S}$ である．主たる変数の状態空間については $\mathcal{N} = \{0, 1, \ldots\}$ を選択する．

再呼待機中の顧客がお互いの状況を把握することは難しく，それぞれ独立に振る舞うと考えられる．仮に再呼待機中の顧客が再度電話をかけ直すまでの待機時間が指数分布に従うとすると，2 つ組 $(N(t), X(t))$ は状態空間が $\tilde{\mathcal{S}} = \mathcal{N} \times \mathcal{S}$ であるようなマルコフ連鎖として記述できる．そのマルコフ連鎖の推移速度行列 $\tilde{\boldsymbol{Q}}$ は次のような構造を有する．

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \boldsymbol{Q}_1^{(0)} & \boldsymbol{Q}_0^{(0)} & \boldsymbol{O} & \boldsymbol{O} & \cdots \\ \boldsymbol{Q}_2^{(1)} & \boldsymbol{Q}_1^{(1)} & \boldsymbol{Q}_0^{(1)} & \boldsymbol{O} & \cdots \\ \boldsymbol{O} & \boldsymbol{Q}_2^{(2)} & \boldsymbol{Q}_1^{(2)} & \boldsymbol{Q}_0^{(2)} & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{Q}_2^{(3)} & \boldsymbol{Q}_1^{(3)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

ここで，$\boldsymbol{Q}_0^{(j)}, \boldsymbol{Q}_1^{(j)}$ $(j \geq 0)$，および $\boldsymbol{Q}_2^{(j)}$ $(j \geq 1)$ の 3 つはいずれも $(c+1)(K+1)$ 次元正方行列である．このように再呼を考慮する場合でもブロック 3 重対角構造をもつマルコフ連鎖としてモデル化が可能である．ただし，$\tilde{\boldsymbol{Q}}$ は無限次元の行列であることに注意する．主たる変数の状態空間を $\mathcal{N} = \{0, 1, \ldots\}$ のように選択した結果である．再呼待ちにある潜在的な客数は非負の整数を取り得ると考えることは不自然ではない．そもそもコールセンターへの客の到着がポアソン過程に従うと考えるのなら，このような状態空間を想定することは理にかなっている．

4.2　数値計算の実際

以上より，補佐する変数の状態空間は有限次元ではあるが，主たる変数の状態空間は非負の整数の集合であるようなマルコフ連鎖を得た．これはすなわち状態の個数が有限ではないことを意味する．よって，正規化条件 $\tilde{\boldsymbol{\pi}}\mathbf{1}^\top = 1$ を満たすような $\tilde{\boldsymbol{\pi}}$ はもし存在すれば無限次元の線形連立方程式 $\tilde{\boldsymbol{\pi}}\tilde{\boldsymbol{Q}} = \mathbf{0}$ の解として表現される．つまり定常分布を求めたければ無限個の連立方程式を解く必要がある．

推移速度行列が $\tilde{\boldsymbol{Q}}$ のような構造をもつマルコフ連鎖は状態依存型の準出生死滅過程と呼ばれる．構造的なマルコフ連鎖ではあるので定常分布もその構造を活用した形式で表現は可能である．ただし，状態空間が無限次元であるため数値計算の観点から容易に計算可能な形式で定常分布を表現することは一般的には難しいとされる．したがって，主たる変数の状態空間が，ある $M > 0$ について $\mathcal{N}_M = \{0, 1, \ldots, M\}$ であるように切り詰めて，有限状態のマルコフ連鎖で近似する方法が実際的には用いられてきた．このような処方箋により前節での数値計算法が適用可能となるものの，どのように $M > 0$ を決定すべきかが課題となる．理想的には M を超える状態空間での定常分布を集めてきてその合計が無視できるほど小さくなるような M を選択できれば好ましいだろう．残念ながら $\tilde{\boldsymbol{Q}}$ のように主たる変数に応じて異なる $\boldsymbol{Q}_0^{(j)}, \boldsymbol{Q}_1^{(j)}$，および $\boldsymbol{Q}_2^{(j)}$ をもつ場合，一般的な枠組みでそのような M を決める

ことは難しく，現状ではモデルごとの性質を活用し決める方法に頼らざるを得ないようである．

主たる変数の状態空間は $\mathcal{N} = \{0, 1, \ldots\}$ を維持しながら $\boldsymbol{Q}_0^{(j)}, \boldsymbol{Q}_1^{(j)}$，および $\boldsymbol{Q}_2^{(j)}$ を修正し，元のマルコフ連鎖を近似する「解ける」モデルに帰着させる方法もある．例えば，ある $M > 0$ に対して，$j \geq M$ では $\boldsymbol{Q}_0^{(j)} = \boldsymbol{Q}_0^{(M)}, \boldsymbol{Q}_1^{(j)} = \boldsymbol{Q}_1^{(M)}, \boldsymbol{Q}_2^{(j)} = \boldsymbol{Q}_2^{(M)}$ のように修正すると，主たる変数が M 以上であるならば同じ推移速度をもつため，実質的に状態に依存しない準出生死滅過程が構成できる．状態に依存しない準出生死滅過程は行列幾何形式と呼ばれる形式で定常分布が表現できるため，少なくとも数値的には評価可能である [3].

4.3 計算コストの観点から

再呼のある待ち行列モデルについて行列幾何形式解で近似する方法は提案されており [4]，同じ考え方を再呼を考慮したコールセンターモデルに応用することは容易い．一方，主たる変数の状態空間を切り詰めるアプローチに基づく方法も提案されている [5]．両者とも行列を基本要素とする演算が不可避であるが，後者では行列による演算回数を減らす工夫が施されている．

後者の近似法での計算コストを例示する．再呼待機時間の平均値を 1 とし，平均到着間隔が 7.2，通話時間と後処理時間の平均値がそれぞれ 180 と 30 とし，その上でオペレータの人数を $c = 32$ に固定する．この条件で $K \geq c$ ならばトラヒック密度とみなせる値が 0.91 程度になる．切り詰めた状態空間 $\mathcal{N}_M = \{0, 1, \ldots, M\}$ の M はこのトラヒック密度と要求する定常分布の精度に応じて決定する．なお，ここでの精度とは主たる変数が M に等しい場合の定常分布と考えてよい．一般的にトラヒック密度が 1 に近いほど，また，高い精度を要求するほど M は大きくなる．

以上の条件で定常分布を計算し，計算コストとして CPU 時間を計測した結果の例を表 1 に示す．動作周波数が 2.2GHz である 8 コアの CPU とメモリ 16GB をもつ PC 上での結果である．演算の基本要素となる行列のサイズは $(c+1)(K+1)$ であるので，外線数 K を大きくするとより大規模な行列を扱うことになり，数値計算上も負荷が重くなる．行列解析法でのモデル化がどの程度のコストで実現可能かを見極めるための参考となれば幸いである．

表 1　精度ごとの M と CPU 時間（秒）

精度	M	$K = 35$	$K = 45$
10^{-6}	109	824	1,540
10^{-9}	175	851	1,585

5. おわりに

コールセンターを待ち行列システムとして眺め，リソース設計を目的としたモデル化の手法について述べた．初等的な待ち行列モデルから概観し，いくつかの難点を浮き上がらせた．次に待ち行列理論で整備されてきた方法論の中でも行列解析法を取り上げ，それらの難点を克服するモデル化の例を紹介した．

行列解析法に基づく方法ならばシステムの状態を細かく表現できるので，自由度の高いモデル化が可能となる．さらに数値的に安定したアルゴリズムが構築できるので十分実用に供する．一方で行列に関する演算が避けれられず，相応の計算コストは発生する．そのような側面を考えると，初等的な枠組みを維持しながらモデルを構築する方法も見直される．例えば後処理についていうならば，後処理中のオペレータ数の平均値を評価し，その平均値に等しい数だけ実効的に外線数を増やした M/M/c/K モデル [6] も検討に値する．

とはいえ行列解析法は現代の待ち行列理論においてはやはり必須の解析手法といえる．リソース設計や性能評価のための道具として使うという立場ならばもはや基本的ツールといえるかもしれない．洋書による解説 [7] があるが，日本語による解説 [8] も用意されているので是非とも学び，応用する機会を得てほしい．

参考文献

[1] R. W. Wolff, "Poisson arrivals see time averages," *Operations Research*, vol. 32, no. 2, pp. 223–231 (1982).

[2] D. Gaver, P. Jacobs and P. Latouche, "Finite birth-and-death models in randomly changing environments," *Advances in Applied Probability*, vol. 16, pp. 715–731 (1984).

[3] M. F. Neuts, *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*, Dover Publications, 1981.

[4] M. F. Neuts and B. M. Rao, "Numerical investigation of a multiserver retrial model," *Queueing Systems*, vol. 7, pp. 169–190 (1990).

[5] T. Phung-Duc and K. Kawanishi, "Multiserver retrial queues with after-call work," *Numerical Algebra, Control and Optimization*, vol. 1, no. 4, pp. 639–656 (2011).

[6] M. J. Fischer, D. A. Garbin and A. Gharakhanian, "Performance modeling of distributed automatic call distribution systems," *Telecommunication Systems*, vol. 9, pp. 133–152 (1998).

[7] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM, 1999.

[8] 牧本直樹，待ち行列アルゴリズム—行列解析アプローチ—，朝倉書店，2001.

