

待ち行列モデルにおける シミュレーション解法

逆瀬川 浩 孝

1. はじめに

世の中には待ち行列的な現象が数多く存在し、それらを理論的に解析した論文がこれまた多数存在しているが、いざ現実の問題を解決する段になると、その問題にちょうどあうモデルがないという理由で、GPSSのような計算機言語でモデルを組み立て、シミュレーションで解を求めるといった方法がとられやすい。シミュレーションを使えばとにかく答が求まるということから、モデルを作って計算し、平均値を出しておしまい、という例が多々見られるが、シミュレーションは確率現象の1つの実現値にすぎないから、確率の大きい事象が必ず出現するとはかぎらないわけで、得られた解がどの程度信用できるかを調べる必要がある。この小論では、待ち行列のシミュレーションでよくとりあげられる平衡状態での特性量に関するシミュレーション解の精度についての2, 3の話題を論ずる。 $\{X(t); t \geq 0\}$ を $t \rightarrow \infty$ の時、ある確率変数 X に弱収束するような待ち行列過程とし、この平衡状態での分布 X の、ある特性量

$$\theta = E[f(x)]$$

を乱数によって作り出された長さ有限のサンプルパスの値によって推定する問題を考える。たとえば利用率が1よりも小さい単一窓口モデルで時刻 t における系人数を $X(t)$ とし、 $f(x) = x^2$ とすれば、これは平衡状態での系人数分布の二次モーメ

ントを推定する問題になるし、 n 番目の客の待ち時間を $X(n)$ とし、 $f(x) = 1 (x=0), =0 (x \neq 0)$ とすればこれは平衡状態において到着客が待たないでサービスを受けられる確率の推定の問題になる。シミュレーションで観測する時点をと t_1, \dots, t_N とし $X_n = X(t_n)$ とおけば、結局、大きさ N の標本 (X_1, \dots, X_N) にもとづく θ の推定量

$$\hat{\theta}_N = \hat{E}[f(X)] = g(X_1, \dots, X_N)$$

で、分散、あるいは平均二乗誤差が最小となるような $g(\cdot)$ を定める、あるいはまた、偏りとか分散の大きさを計算する問題を考えることになる。

以下、データの観測開始時期の問題、再生過程を利用したシミュレーションの有効性の問題、組平均法の逐次決定方式、等について解説しよう。

2. 初期状態の影響

$\{X_n\}$ から $\theta = E[f(X)]$ を次式によって推定する問題を考える。

$$\hat{\theta}_N = \frac{1}{N} \sum_{n=0}^{N-1} f(X_n) \quad (2.1)$$

$\{X_n\}$ が独立同分布にしたがっている場合とか、 X_0 が X と同じ分布で t_1, t_2, \dots, t_N がうまく選ばれている場合には $\hat{\theta}_N$ は θ の不偏推定になるが、シミュレーションの場合の $\{X_n\}$ はこのような条件を満たしていないので、 $\hat{\theta}_N$ の偏りに対してどのように対処したらよいか、というのが本節の主題である。 $\{X_n\}$ は X に弱収束するから、 $\{t_n\}$ がうまく選ばれていれば、 $\hat{\theta}_N$ は θ に確率1で収

束することが知られているので、(2.1)式はそれなりの意味をもった推定量ではある。

シミュレーションの実例として、ある窓口の出力が別の窓口の入力になっており、各窓口の前には、決められた数の客しか待てない、いわゆるブロッキングのある直列型待ち行列モデルをとりあげよう。現実の待ち行列的な現象にはこうした例がしばしば見られるにもかかわらず、解析的にはほとんど何もわかっていないというのが現状である。まず問題になるのが、全体としてのサービス能力がどれぐらいか、ということ、すなわち第1段の窓口の前だけに無限大容量の待ち合い室を置いた時、待ち行列長が発散しないような到着率の上限(λ_{\max})がどうなるか、ということである。各段の窓口数が1の場合、この上限が存在することはわかっているが、サービス分布が与えられた特定のモデルについてその値を計算する方法は与えられていない。これをシミュレーションで求めるためには、第1段の前に常に客が待っているという状態での客の退去間隔を X_n とし、(2.1)式によって、平均退去間隔を推定し、その逆数を λ_{\max} の推定値とすればよい。実際に計算されたモデルは、平均サービス時間が1の指数分布にしたがう窓口が K 個直列に並び、中間待ちが許されない、というものである。このモデルはマルコフ過程となっているから、 K が小さい場合には平衡方程式をたてることによって λ_{\max} を数値的に求められるが、 K が大きくなった場合にはシミュレーションに頼らざるを得ない。このモデルは、計算は簡単であり、しかも、かなり離れた客相互の相関が無視できないという点で、 X_n の相関を調べるための格好の例になっている。時刻 $t=0$ で最初の客が第1段でサービスを開始したと考え、手近かな乱数を使って計算してみたところ、表2のような結果が得られた。表1から、この値は大体2.67に収束することが期待される。このモデルでは、先頭の客と次の客の退去間隔は K の増大につれて発散することが知られているから、 N が小さいと

表1 λ_{\max}

| K | 2 | 3 | 4 | 5 | 6 | 10 | 100 | 1000 |
|------------------|------|------|------|------|------|-------|-------|-------|
| λ_{\max} | .667 | .564 | .515 | .486 | .467 | .430* | .377* | .374* |

* はシミュレーションによる

表2 シミュレーション結果(その1)

| N | 50 | 100 | 500 | 1000 | 2000 | 5000 |
|------------------|------|------|------|------|------|------|
| $\hat{\theta}_N$ | 9.49 | 7.03 | 4.11 | 3.51 | 3.11 | 2.87 |

表3 シミュレーション結果(その2)

| N_0 | 50 | 100 | 500 | 1000 | 2000 |
|-------------------------|------|------|------|------|------|
| $\hat{\theta}_{N_0, N}$ | 2.80 | 2.79 | 2.73 | 2.71 | 2.71 |

ここで \bar{X}_N が大きな値になっているのは、最初のいくつかの長い退去間隔データがそのまま使われていることによるものと思われる。その影響がかなり長く尾を引いており、得体の知れない初期値を使う場合に単純な標本平均で推定することの危険性を示している。

この初期値の影響を消すためには最初のいくつかのデータをスキップし途中からのデータだけを使って標本平均を求めれば良いように思われる。

$$\hat{\theta}_{N_0, N} = \frac{1}{N - N_0} \sum_{n=N_0}^{N-1} f(X_n) \quad (2.2)$$

上の例で、 N_0 をいろいろに変えて計算すると、表3のようになる。このようにして N_0 を十分大きくとれば θ はうまく推定できるように思われるが、さてそれでは N_0 はどのように決めたらよいだろうか。最初のデータをスキップするという考え方は、 X_n が X に収束するのだから、どんな状態から出発しても X_n の「分布」は X のそれに近づくので、ある程度大きい n では確率変数は定常とみなせる、という考えにもとづいているが、サンプルパスを扱うシミュレーションではこのような論拠は成り立たないように思われる。実際、 t_{N_0} 以降に現われる確率過程の状態を、どんなものでもよいから1つとり、それを初期状態としてシミュレーションを開始させたと考えれば、もとの実験で、悪くはないとみなされたデータが最初から得られるので、データをスキップする必要が

表 4 シミュレーション結果 (その3)

| N または N_0 | 100 | 500 | 1000 | 2000 | 5000 |
|---------------------------|------|------|------|------|------|
| $\hat{\theta}_N$ | 2.53 | 2.59 | 2.64 | 2.66 | 2.68 |
| $\hat{\theta}_{N_0,5000}$ | 2.68 | 2.69 | 2.69 | 2.69 | — |

ない。つまり、そのような初期状態を選ぶことができれば最初の観測値からデータとして採用することができるはずである。さて、目的とする状態は、定常とみなされた確率過程がある時刻に訪問したものであるから、定常状態において、その状態をとる定常確率がある程度大きいものであることが必要である。このようにするためには、定常分布に対する知識が必要で、それがわかればシミュレーションをやる必要がなく、自家撞着に陥るようであるが、ある程度定常分布を予測することによって初期状態を定め、推定の偏りを小さくすることが可能である場合が少なくない。たとえば上の例では、初期状態はどの窓口も空の状態を選んだが、それに対して、最終段を除いてすべての窓口がブロックされている状態を初期状態として選び同じようなシミュレーションを行なうと表4のような結果が得られる。このように、初期状態を工夫することによって、推定の偏りをある程度減ずることはできるが、それでも最初のデータをスキップするという手順はやめるわけにはいかない。これに対しては次のような逐次決定手法が考えられる。まず、 $\{X_n\}$ を適当な大きさ m (たとえば10~50ぐらい)の偏りに分けて、各組で標本平均 $\hat{\theta}_j$ を計算し、

$$\hat{\theta}_j = \frac{1}{m} \sum_{i=0}^{m-1} f(X_{m(j-1)+i}) \quad (2.3)$$

十分先の組から計算された標本平均のいくつか、 $\{\hat{\theta}_j; j=l, l+1, \dots, k=\frac{N}{m}\}$ を使って t -分布をあてはめる。次に $j=l-1, l-2, \dots, 1$ について $\hat{\theta}_j$ がこの t -分布からの標本とみなし得るか否かを検定し、 α 回(たとえば2回)連続して棄却されるまで繰り返し、 α 回連続して棄却された組のデータおよびそれ以前の観測値をスキップする、というものである。この手順にはパラメータが4つ ($m, l,$

α および検定の有意水準) 入っており、どのような値を使うべきかについてはモデルごとに変わると思われるので詳しい検討が必要である。

このように最初のいくつかの観測値をスキップするという考え方はきわめて自然な発想だが、実用上の指針となるようなパラメータ選択の規則は確立していない。この問題を避けるためにある種の確率過程のシミュレーションに対して新しい方法が考えられているので、それを次節で述べる。

3. 再生過程のシミュレーション

前節で数値例として使った直列型待ち行列モデルで、最初の段を除いて窓口が空いている状態で最初の段のサービスが終わった時点を考えよう。 $\tau_0=0, \tau_1, \tau_2, \dots (\tau_i < \tau_{i+1})$ をそのような時点の列とすれば、 τ_i 以前におきた事象は、 τ_i から先の事象に対してなんら影響を及ぼさず、 $[\tau_i, \tau_{i+1})$ の間の種々な確率事象、たとえば第1段のサービスを終了した客の数というようなものは各 i ごとに独立な確率変数で共通の分布にしたがっていると言える。つまり各 τ_i ごとにシステムが再生していると考えられる。このような性質をもつ確率過程を再生過程 (regenerative process) といい、 $\{\tau_n\}$ を再生点 (regeneration point あるいは epoch) と呼ぶ(「再生過程」は renewal process の訳語として定着しているので混乱を招くが適当な訳語がないのでこの稿にかぎりこの用語を借用する)。待ち行列論で扱うモデルはほとんどがこのような意味で再生過程になっている。

シミュレーションモデルが再生過程になっていれば、前節で苦勞した問題は避けて通ることができるばかりでなく独立標本にもとづく推定論が適用できるため推定量の精度が計算できるという大きなメリットがある。これにつき説明しよう。

$$Y_n = \int_{\tau_{n-1}}^{\tau_n} f(X(t)) dt, Z_n = \tau_n - \tau_{n-1} \quad (3.1)$$

とおけば、 $\{\tau_n\}$ が再生点になっていることから、 $\{(Y_n, Z_n)\}$ は独立同分布にしたがう二次元の確率

変数列になっており、適当な条件の下で、

$$\frac{1}{\tau_N} \int_0^{\tau_N} f(X(t)) dt = (\sum_{n=1}^N Y_N) / (\sum_{n=1}^N Z_N) \quad (3.2)$$

は、 $N \rightarrow \infty$ の時確率 1 で θ に収束することが言える。大数の法則を使うと、

$$\theta = E(Y_1) / E(Z_1) \quad (3.3)$$

となるので問題は Y_1, Z_1 の期待値の比を有限の $\{(Y_n, Z_n)\}$ によってどのように推定するかということになる。1つの方法は(3.2)式をもとに、分母・分子を独立に推定してその比をとる方法

$$\hat{\theta}_N = \left(\frac{1}{N} \sum_{n=1}^N Y_n \right) / \left(\frac{1}{N} \sum_{n=1}^N Z_n \right) = \bar{Y} / \bar{Z} \quad (3.4)$$

である。 $\hat{\theta}_N$ は $N \rightarrow \infty$ で θ に収束し、独立な確率変数列 $\{U_n = Y_n - \theta Z_n\}$ に対して中心極限定理を適用することにより、

$$P \left(\frac{\sqrt{n} \bar{Z}}{\sigma} (\hat{\theta}_N - \theta) \leq x \right) \rightarrow \Phi(x) \quad (3.5)$$

(ただし、 $\sigma^2 = \text{Var}(U_1)$, $\Phi(x)$ は標準正規分布) が言える。したがって $\Phi(x)$ の上側 50% 点を Z_α と書けば、100(1- α)% 信頼区間は、

$$\left[\hat{\theta}_N - \frac{Z_\alpha}{\sqrt{N}} \cdot \frac{\sigma}{E(Z_1)}, \hat{\theta}_N + \frac{Z_\alpha}{\sqrt{N}} \cdot \frac{\sigma}{E(Z_1)} \right] \quad (3.6)$$

によって与えられる(たとえば[1])。

このように、再生過程のシミュレーションでは最初の観測値から、独立サンプルを得るためのデータとして用いられるので初期状態の影響を考慮する必要がなく、大標本にもとづいてはいるものの精度がきちんと計算できるという良さがある。また、あらかじめ $\sigma, E(Z_1)$ をテストランによって推定しておけば、(3.6)を使って必要な精度を得るための実験の長さを事前に知ることができる。いくつかのモデルを使って実際に計算し、この方法がうまく働いているという報告が多くなされているが、実用的に見て十分に有効なものと認められるためには、なおいくつかの問題点が解明されなければならない。第1点は、最も基本的な問題で、モデルごとにどのようにしたら再生点を見出すことができるのかという問題である。前述の直列型待ち行列モデルでは、この節の最初に定義された $\tau_1, \tau_2, \dots, \tau_N$ は確かに再生点にはなっているが、

段数が増えた時、もしサービス能力が各段でバランスしているとすれば、そのような状態が生ずる確率はきわめて低いものになる。再生点としては他にも、ある段から先の窓口が空いており、その段の前の各窓口がブロックされている状態でその段のサービスが終了した時点、とか、指数分布以外のサービス時間分布をもつ窓口のうち一つだけが稼動している状態で、その窓口のサービスが終了した時点、等、いろいろ考えられるが、それらのうち、いずれが上の意味で適当な再生点となるのか、については残念ながらあらかじめ知ることはむずかしい。再生点が見つければ、後の計算は単純だから信頼区間を求めるところまでは順調にいくが、ここにもいくつかの問題が残されている。1つは推定量の偏りの問題であり、また他の1つは信頼区間の信頼度の問題である。(3.4)式によって定義された $\hat{\theta}_N$ は $N \rightarrow \infty$ の時確率 1 で θ に収束することが保証されているが、不偏推定量ではない。シミュレーションは有限の数しか扱っていないし、複雑なモデルで再生点が多くとれないとすればあまり大きな N を期待することはむずかしいので、 $\hat{\theta}_N$ の偏りの大きさ $|E(\hat{\theta}_N) - \theta|$ をきちんと評価しておく。あるいは $\hat{\theta}_N$ より偏りの小さい別の推定量を考える、ということが必要になってくる。[3]では、いくつかの小さいモデルでいろいろな推定量の偏りと精度を調べるためのシミュレーション実験をした結果、 N が小さい場合には jackknife 型の推定量を使うと良い結果が得られる、という報告がなされている。一方、信頼区間の問題は推定量の偏りにも大いに関係してくるが、真の値のわかっているモデルについて同じような実験を 100 回繰り返してそのつど 100(1- α)% 信頼区間を計算すると、100 回のうち平均 100(1- α) 回は信頼区間が正しく真の値を含んでいることが期待されるにもかかわらず、期待通りにことが運ばないという数値例が少なからずある(たとえば[4])、という問題である。これは、推定量が偏っている、の他に、分散が正しく

推定されていない(過小評価されている)という点にも原因がある。これらの諸問題について解明されれば、実用上強力な道具になることが期待できるが、これは容易なことではなさそうである。

再生点を見つけることのむずかしさを回避する方法として、擬似再生過程法が提案されている[2]。これは、再生点間をいくつかの区間に分割し、それらの区分点も再生点とみなして上の推定法を適用しても、分割がうまくいっていれば同様な結論が得られる、というものである。ある再生過程は状態が a から b に変化した時点が再生点になっていると仮定する。これに対して、 $a(b)$ を含む任意の状態集合を $A(B)$ とし、 $X(t)$ が A に含まれる状態から B に含まれる状態に変化した点を再生点と考え、この擬似再生過程に上の方法を適用すると、(3.4)式の形の推定量は、再生過程の場合のそれと同じ性質をもつ。すなわち、標本数が多くなれば $\hat{\theta}_N$ は正規分布し、その期待値は θ に近づく、ということが証明されている。 A, B はそれぞれ a, b を含みさえすればどんな集合でもよいということから当然 $\hat{\theta}_N$ の θ への収束が遅くなることが予想される。そこでこの方法を実際の問題に適用するためには、再生過程による方法に比べてより一層 $\hat{\theta}_N$ の偏り、また N が小さい場合の正規分布からのずれに対する解析が必要となるが、これらの点についてはまだほとんど何も確かなことは言えていない。しかし、いくつかのやや複雑なモデルで A, B を適当に選んでこの方法を適用してみると、前節で述べた方法に比べてはるかに少ないデータで同じ程度の精度で推定できることが数値的に確かめられている[6]。

4. 組平均法

再生過程のシミュレーションが $\{X_n\}$ を再生点ごとに区切って独立サンプルをとり出したのに対して、組平均 batch means 法は、 $\{X_n\}$ を決められた数、たとえば m コずつの組に分け、各組ごとに m コのデータから標本平均を計算した時、こ

れらの標本平均が同一の正規分布からの独立サンプルであるとみなして t -分布による区間推定を行なう方法である。第 j 組の m コのデータから計算された標本平均を $\hat{\theta}_j (j=1, 2, \dots, k = \frac{N}{m})$, それらの k コのデータの平均を $\tilde{\theta}$ とすると、 $\hat{\theta}_j$ が正規分布するという仮定は良いとしても、 $\hat{\theta}_j$ と $\hat{\theta}_l (l \neq j)$ は独立であるとは言えないから、上述のように推定された信頼区間は正確なものではない。すなわち、 $\tilde{\theta}$ の分散 $\sigma^2 = \text{Var}(\tilde{\theta})$ は、データから、

$$\sigma^2 = \frac{1}{k(k-1)} \sum_{j=1}^k (\hat{\theta}_j - \tilde{\theta})^2$$

によって推定されるが、 $\hat{\theta}_j$ が独立でないために $\hat{\sigma}^2$ は σ^2 の不偏推定にならない。その偏りを計算することは一般にはむずかしいが、特に $\{X_n\}$ が二次定常過程になっていれば、自己相関係数 $\rho_j = c_j/c_0 (c_j = \text{Cov}(f(X_n), f(X_{n+j})))$ を使って次のように評価することができる。

$$E(\hat{\sigma}^2) = b(k, m)\sigma^2 \quad (4.1)$$

$$b(k, m) = \frac{1}{k-1} \left\{ k / (1 + 2 \sum_{j=1}^{k-1} (1 - \frac{j}{k}) \rho_j) - 1 \right\}$$

この式によれば $\rho_j > 0$ の時 $\hat{\sigma}^2$ は負の偏りをもつ、すなわち σ^2 を過小評価することになるが、もしそれが十分小さければその偏りは小さくなり上述のように $\hat{\theta}_j$ を独立サンプルと考えて区間推定してもかまわないことになる。この考えは ρ_j の次のような性質によって正当性を与えられる、すなわちもし $\sum c_j < \infty$ ならば、 m を大きくした時 ρ_j はいくらでも 0 に近づく。実際の計算では m をいくつにすれば良いかが問題になるが、[5] では多くの数値実験で得られた ρ_1 と m の大きさの関係をを利用して逐次決定法によって m を決める方法が提案されている。その数値実験の結果とは多くの待ち行列モデルでは ρ_1 に比べて ρ_2 以下は無視でき、 ρ_1 は正で、 m の単調減少関数になっており、ある m で $\rho_1 = \rho_1(m)$ が 0.4 より小さくなっているれば $\rho_1(10m)$ は 0.05 より小さくなっている、というものである。そこでまず m コずつに区切られたデータから標本平均 $\{\hat{\theta}_j\}$ を計算し、 $\{\hat{\theta}_j\}$ の遅れ 1 の自己相関係数 $\rho_1(m)$ を計算してその値が

0.4以下になるように m を定める. この場合 $\rho_1(m)$ の推定の偏りを少なくするため組の数 k は 100 以上の数にあらかじめ決めておく必要がある. そうすると $k \cdot m = N$ という関係から m を定めるとことは標本の大きさを決めることを意味していることがわかる. m が決まったところで改めて全部のデータを大きさ $10m$ の組に分けて標本平均を計算し, これらが独立サンプルであるかのようによく考えて信頼区間を求める. もし信頼区間を短くしたければデータ数を増やして同じ手順を繰り返す必要な精度が得られるまで判定を続ける.

この逐次決定法に含まれるパラメータは $\rho_1(m)$ の収束条件を表わす 0.4, 0.05, 10 という 3 種類の数字, $\rho_1(m)$ を推定するのに使うデータの数 $k (> 100)$, 信頼区間の長さの基準等であるが, 最も重要なのは $\rho_1(m)$ の形状で, 上で述べた実験結果が成立つようなモデルの範囲がどんなものであるかを知る必要があるし, 逆に与えられたモデルの $\rho_1(m)$ の形状を知る手順を与える必要がある. 実際の数値実験では幾つかの簡単なモデルに対して前節で述べたような信頼区間が真の値を含む比率を計算した例が報告されている. それによれば $\{X_n\}$ の相関が高くなったり分散が大きくなるということは単段モデルでいえば, たとえば ρ が 1 に近づくとサービス規律を LIFO にする事に対応しているが, その比率が期待値を大きく下まわるといふ結果になり (4.1) 式における ρ_j の影響をも少し考慮する必要があることを示唆している.

5. ま と め

数値的な実験によってある種のモデルに対して有効性が確かめられた推定方式を理論的に裏づけようとすると, ほとんどの場合, 得られるのは大数の法則と中心極限定理にもとづく, 推定量の一致性と, 漸近正規性ということになる. 実際のシミュレーションでは, 有限の, それもかなり小さい数の標本にもとづいて推定される場合が多いから, そのような性質はあまり役に立たない. 実用

上必要なのは推定量の偏りと正規分布からのズレを標本の大きさの関数として実際に与えることであるが, この問題を一般的状況のもとで解析することは容易なことではなさそうである. そこで推定量の有限の範囲のふるまいに関しても, いくつかのモデルを使った数値的な実験によって検証がなされているが, とりあげられるモデルは, 真の値と比較する必要があるというので, 規模の小さいものが多く, 正当性の検証が十分になされているとは言い難いので, もっとモデルの範囲を拡げてそれぞれの推定方式の守備範囲を明確にする努力が必要と思われる.

以上のように現在のところ解の精度を評価し得るシミュレーションの方法で決定的なものは存在しないのでシミュレーションで計算する場合にはいろいろな方法を試してみても同じような結果が得られればそれを解とする, というなんとも歯切れの悪いやり方が最上の策ではないかと思われる.

参 考 文 献

- [1] Crane, M. A. and Lemoine, A. J. (1977), *An Introduction to the Regenerative Method for Simulation Analysis*, Springer.
- [2] Gunther, F. L. and Wolff, R. W. (1980), The almost regenerative method for stochastic system simulations, *OR* 28, 2, 375-386.
- [3] Iglehart, D. L. (1975), Simulating stable stochastic systems, V: Comparison of ratio estimators, *Naval Res. Logist. Quart.* 22, 553-565.
- [4] Lavenberg, S. S. and Sauer, C. H. (1977), Sequential stopping rules for the regenerative method of simulation, *IBM J. Res. Develop.* 21, 545-558.
- [5] Law, A. M. and Carson, J. S. (1979), A sequential procedure for determining the length of a steady-state simulation, *OR* 27, 5, 1011-1025.
- [6] 逆瀬川浩孝 (1981) 「Regenerative Simulation について」京都大学数理解析研究録.