数理解析研究所講究録
490巻 1983年 108-118

後着順サービスのGI/G/1の特性とその応用

神戸大 藤井進 (Susumu Fujii)
工学院大 山崎源治 (Genji Yamazaki)

1. はじめに

計算機の性能評価のための待ち行列ネットワークの研究が多数なされてきている。その非常に簡単なモデル—ノード数が小さく，各ノードでのジョブの処理時間は指数分布—については，既存の数値解法により解くことができるが，モデルの規模が少し大きくなると，最早それができなくなる（もちろん，原理的には可能であるが）。そのため，システムの各ノードのジョブ数の同時分布などが「積形式」となるモデルの研究に関心が集中し，それになるための各ノードにおけるサービス規律が幾つか求められてきた（例えば，Kelly [2]）。もちろん，この場合ジョブのシステムへの到着過程は，Poisson過程である。この積形式が成立するための規律と，その結果の特性を，あるノード—そこでのジョブの処理時間の分布は任意，1度に1ジョブしか処理できない—についてみると，

そのノードに到着したジョブの処理は，すぐに開始しなければならない（後着順・割込み優先サービス規律；PR－LCFSと略す），任意時点でのそのノードのジョブ数分布は幾何分布，存在する各ジョブの残り処理時間は i.i.d. r.v.'s となる，ということである。

本稿では，上述のノードを単独に取出し，それへのジョブの到着過程が任意の場合（すなわち，GI/G/1），PR-LCFSがどのような特性をもたらすか，を考えてみる。Fakinos(1981)[1]は，ジョブの到着時点のみでシステムを観察したとき，システム内ジョブ数の分布は，やはり幾何分布となること，存在する各ジョブの残り処理時間が i.i.d. r.v.'s となること，を証明した。その後，Yamazaki (1982)[5]は，そのシステムを，ジョブの到着時点と退去時点で観察することにより両時点で Fakinos の結果を証明し，さらに，あるジョブの退去時点から観測を始めて最初のジョブの到着時点までの時間の分布がシステムの状態とは独立で，その分布が先着順サービス規律のもとでの GI/G/1 の idle time の分布と一致することを明らかにした。

本稿では，[5]の結果を簡単に要約し（2節），その結果の応用（3節）について考えてみる。

2. 不変特性

ジョブのシステムへの到着時点を，$a = a_0, a_1, a_2, \cdots$ ($a_0=0$; $a_0 < a_1 < a_2 \cdots$)とし，$T_n \overset{d}{=} a_n - a_{n-1}$ とする。T_nは i.i.d.の r.v.'s で，その分布関数を $A(y)$，平均を λ^{-1}，n 番目に到着するジョブの処理時間を S_n で表わし，S_n も i.i.d. r.v.'s で，その分布関数を $B(x)$，平均を μ^{-1} とする。もちろん，$\{T_n\}$ と $\{S_n\}$ は独立で，$\rho \equiv \lambda/\mu < 1$ と仮定する。このシステムのサーバは1人で，1度に1ジョブのみを処理する。

このとき，$U_{n+1} = U_n + V_n$ ($n = 0, 1, 2, \cdots$)，$V_n = S_n - T_{n+1}$ ($n = 0, 1, 2, \cdots$)で定義される「ランダム・ウォーク」を考える。このランダム・ウォークのサンプルを図示すると，それは，上述の GI/G/1 システムの仮待ち時間のサンプルと類似している。ただ，前者は0に壁を持たないが，後者は0が壁となることのみ異なる。このランダム・ウォークで，その最後の最大値を最初にジャンプする点は "ascending ladder indices" と呼ばれているが，$\lambda/\mu < 1$ より $n \to \infty$ で $U_n \to -\infty$ となるため，その indices の数：K は確率1で有界な r.v. となる。そして，この K の分布が幾何分布となることはよく知られている（Kleinrock [3]）。すなわち，

(1) $\quad Pr(K = n) = (1-\sigma)\sigma^n \quad (n = 0, 1, 2, \cdots)$,

ここで，$\quad 1 - \sigma \equiv Pr(U_n \leq U_0;\ n = 1, 2, \cdots)$.

この $1-\alpha$ は，待ち行列論的には，ジョブがシステムに到着したとき，システムが空の状態である確率（平衡状態で），と解釈できる。さらに，$K=k$ であるとき（その相続く indices を $n_1, n_2, \cdots, n_k$ とする），明らかに $J_i \equiv U_{n_i} - U_{n_{i-1}}$ （$i=1,2,\cdots,k$；$n_0=0$）は，i.i.d. の r.v.'s であり，その分布関数を次のように定義する。

(2) $\qquad F(x) = \Pr(J_i \leq x) \qquad (x \geq 0).$

さらに，$1-I(x)$ を，上述のランダム・ウォークが $U_0=0$ で始まり，$x\ (\geq 0)$ に対して一度も $-x$ をこえない事象の確率として，次のように定義する。

(3) $\qquad 1-I(x) = \Pr(U_n \leq -x\,;\ n=1,2,\cdots \mid U_0=0,\ K=0).$

以上の準備のもとで，上述の GI/G/1 の特性を考える。そのシステムで，ジョブの待ちスペースを適当に分割して，サーバの前から順に番号づけをして，それらを positions 2,3,… と呼ぶ。便宜上，サーバを position 1 と呼ぶことにする。このとき，システムは次のように作動する。システム内に n ジョブ存在するとき，(i) 到着したジョブはすぐに position 1

に入る，そして positions $1, 2, \cdots, n$ にいたジョブは順に，positions $2, 3, \cdots, n+1$ へ移る，(ii) position 1 のジョブの処理が終了し，そのジョブがシステムを去ったとき，positions $2, 3, \cdots, n$ にいたジョブは順に，positions $1, 2, \cdots, n-1$ へ移る。他のジョブの到着のため，処理を中断されたジョブの残り処理時間は，もちろん次の処理まで不変であるとする。

$Q(t)$, $X_i(t)$ $(i=1, 2, \cdots, Q(t))$, $Y(t)$ を順に，このシステムの時刻 t におけるジョブ数，position i のジョブの残り処理時間，時刻 t から観測を始めて最初のジョブの到着時点までの時間とし，$\alpha = \alpha_1, \alpha_2, \cdots$ をジョブの退去時点としよう。

このとき，[5] で得られた結果は，次のように要約できる（ただし，以下の結果は平衡状態で）．

(4) $$\Pr(Q(a-0)=n;\ X_1(a-0)\le x_1,\ X_2(a-0)\le x_2, \cdots, X_n(a-0)\le x_n;\ Y(a-0)\le y) = r_n A(y)\prod_{j=1}^{n} F(x_j),$$

(5) $$\Pr(Q(\alpha+0)=n;\ X_1(\alpha+0)\le x_1,\ X_2(\alpha+0)\le x_2, \cdots, X_n(\alpha+0)\le x_n;\ Y(\alpha+0)\le y) = g_n I(y)\prod_{j=1}^{n} F(x_j),$$

(6) $$r_n = g_n = (1-\rho)\rho^n \qquad (n=0, 1, 2, \cdots).$$

3. 応 用

本節では，FCFSのもとでのGI/G/1のジョブの待ち時間（平衡状態での）：W_q の期待値 $E(W_q)$，分散 $Var(W_q)$ をシミュレーションで推定する際の，前節の結果の応用を試る。

FCFSのもとでの $E(W_q)$，$Var(W_q)$ の一つの表現は，次のようになる（Marshall [4]）。

$$(7) \qquad E(W_q) = \frac{C_a^2 + \rho^2 C_s^2}{2\lambda(1-\rho)} + \frac{1-\rho}{2\lambda} - \frac{E(I^2)}{2E(I)}$$

$$(8) \qquad Var(W_q) = \frac{\lambda\{E(S^3) - E(T^3)\}}{3(1-\rho)} + \frac{\rho}{\lambda^2} + \frac{\rho(C_a^2 - \rho C_s^2)}{\lambda^2(1-\rho)} + \frac{1}{4\lambda^2}\left\{\frac{C_a^2 + \rho^2 C_s^2}{1-\rho} + 1 - \rho\right\}^2 + \frac{E(I^3)}{3E(I)} - \left\{\frac{E(I^2)}{2E(I)}\right\}^2,$$

ここで，

C_a：ジョブの到着間隔の変動係数，　C_s：ジョブの処理時間の変動係数，　S：処理時間の分布に従うr.v.，

T：到着間隔の分布に従うr.v.，　I：idle time.

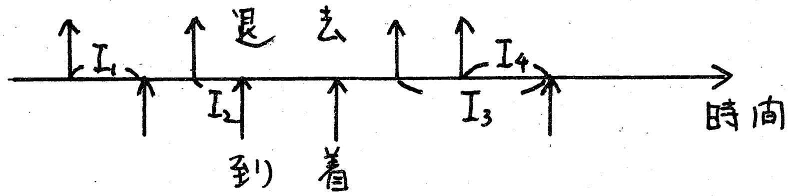
シミュレーションで $E(W_q)$ を推定する際，(7)式の右

辺のρ, C_a, C_s, λは必然的に定まるため，未知の項は$E(I^2)/2E(I)$のみであり，ρが1に近いところではそれは他の項と比べて微少となることは，よく知られている。従って，その項のよい推定法が見つかれば，各ジョブの待ち時間から直接$E(W_q)$を推定するよりも，$E(I^2)/2E(I)$を推定し，(7)式より$E(W_q)$を求める方が有効となることが期待できる。$Var(W_q)$についても同様である。それゆえ，FCFSのもとで，シミュレーションによる$E(W_q)$, $Var(W_q)$の推定法としては次の2つが考えられる。

ケース1：各ジョブの待ち時間から直接的に推定する.

ケース2：idle timeから$E(I)$, $E(I^2)$, $E(I^3)$を推定し，(7), (8)式を用いる。

ケース2を用いる場合，ρが大きいところではデータ数



図・1：PR－LCFCの退去時点と次の到着の間隔

が少なくなり，かなりのジョブ数をランさせなければならない，という心配を伴う。一方，前節の(5)式に着目すると，PR-LCFSのもとでのGI/G/1のある退去時点から次の到着時点までの間隔の分布は，システム内ジョブ数とは独立となる。また，FCFSのもとでも，PR-LCFSのもとでも同じサンプルでみた場合，システムが空になる時点は一致する。それゆえ，(5)式の右辺の$I(y)$は，FCFSのもとでのIの分布と一致する。

この特性を利用して，(7)，(8)式の$E(I)$, $E(I^2)$, $E(I^3)$に関する次の推定法が考えられる。

ケース3：PR-LCFSのもとでランさせて，ジョブの退去時点から最初の到着時点までの間隔のデータから$E(I)$, $E(I^2)$, $E(I^3)$を推定し，(7)，(8)式を用いる（図・1）。

もちろん，ケース3の相続くデータは互いに独立とはならないが，データ数はケース1と同程度になる。

一例として，M/M/1の$E(W_q)$, $Var(W_q)$の推定を，ケース1，2，3で行った結果をプロットしたのが，図・2，図・3である。これらの図は，一回のシミュレーションのランで，

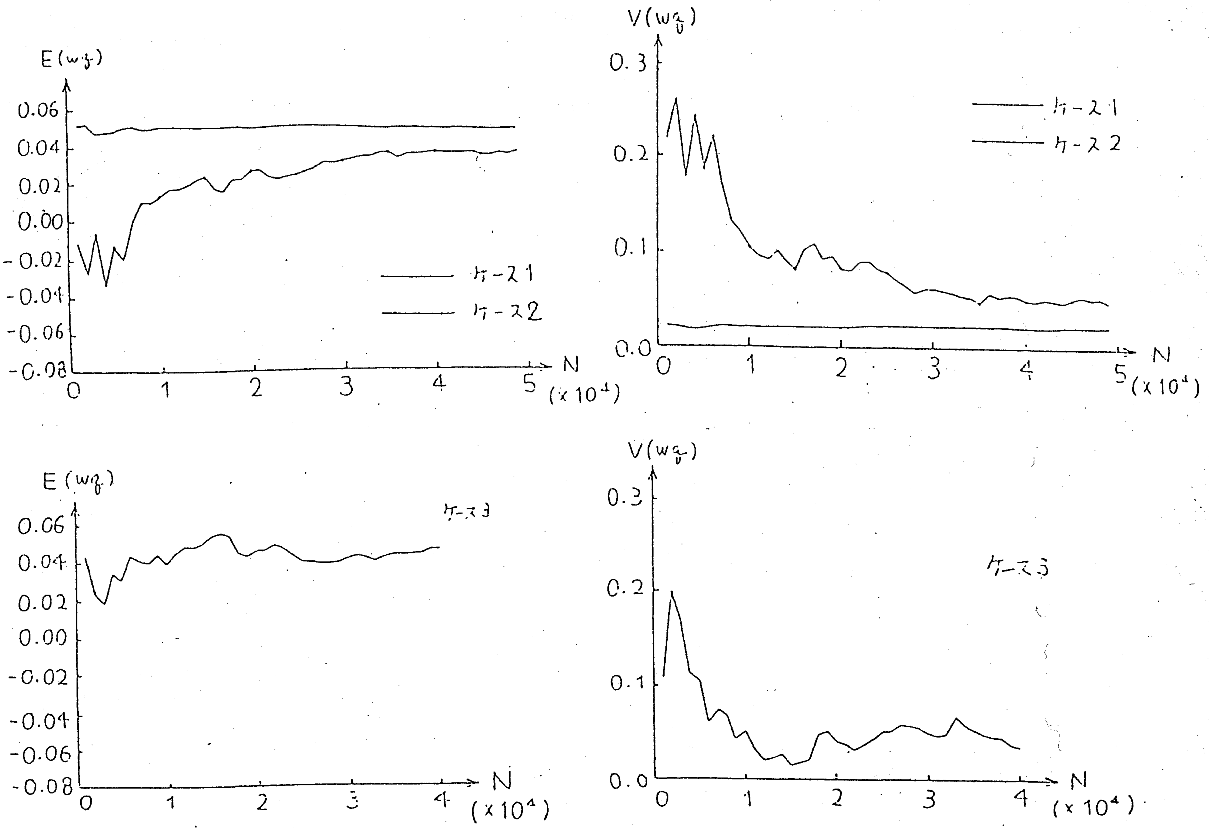
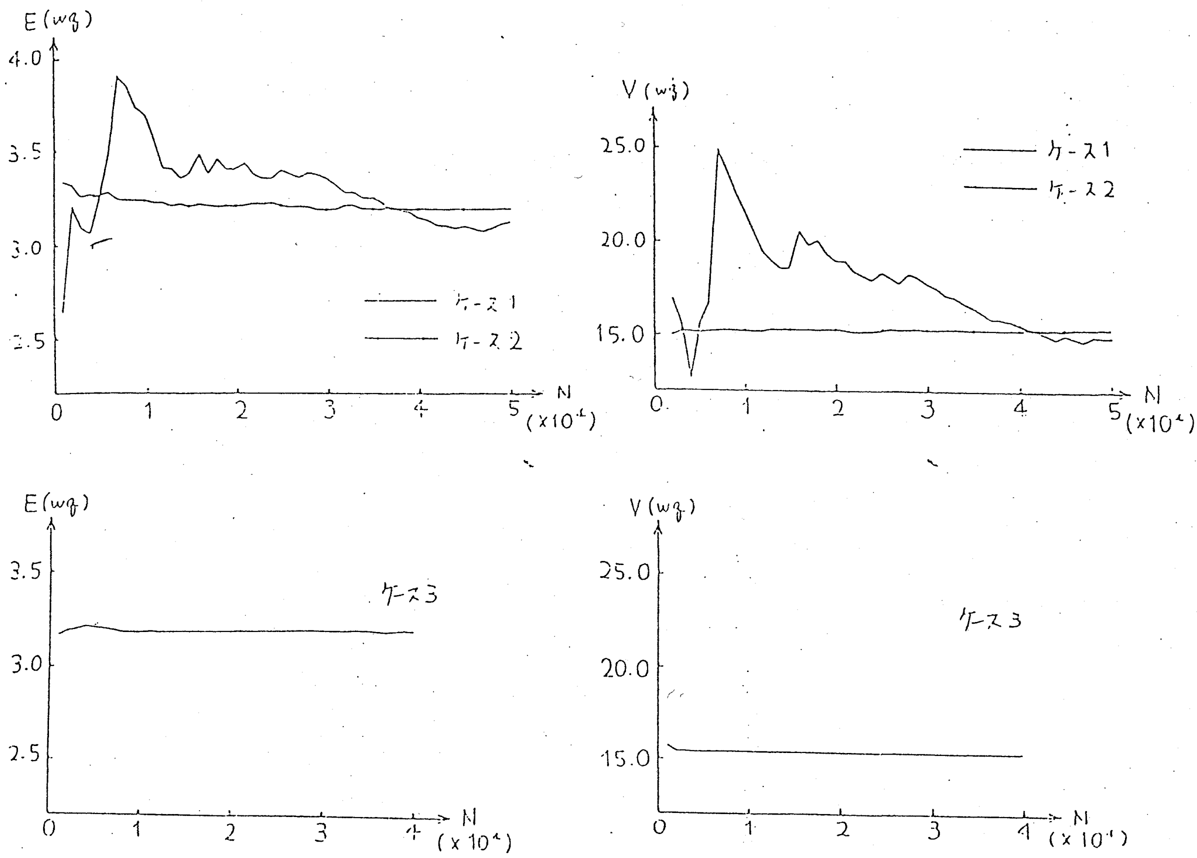


図2: M/M/1の $E(W_q)$, $Var(W_q)$の推定, $\rho=0.2$

$N=4\times10^4$のジョブを流し, 10^3ごとに各ケースで, $E(W_q)$ $Var(W_q)$ の推定値をプロットしたものである。

これらの図から明らかなように, ρが小さいところではケース3の推定法はあまりよくないが, ρが大きいとき, ケース3の推定法がかなり有効である。

まだ, 数値実験の結果が少なく, 結論を出す段階ではない。しかし, ケース3を用いる場合, 互いに独立なデータではないため, そのとり方に若干の工夫が必要であると思われるが, ρの大きいところでは, 一回のランのジョブ数をかな



図·3：M/M/1 の $E(W_q)$, $Var(W_q)$ の推定, $\rho = 0.8$

り少なくしても，有効な推定が可能であると期待できる。今後，さらに検討し，次の機会で報告したい。

Refernces

1. Fakinos, D.(1981). The GI/G/1 queueing system with a particular queue discipline, J. R. Statist Soc., B. 43, 190-196.

2. Kelly, F. P. (1978). Reversibility and Stochastic Networks, Wiley.

3. Kleinrock, L.(1975). Queueing Systems, Vol. 1, Wiley.

4. Marshall, K. T.(1968). Some inequalities in queueing, Opns. Res.,

Vol.16, 651-665.

5. Yamazaki, G.(1982). The GI/G/1 queue with last-come-first-served, Ann. Inst. Statist. Math., Vol. 34, 599-604.