# $M/G/1$ 待ち行列とバンディット問題

西村　彰一

## 1. はじめに

待ち行列のサービス規律の中で FCFS (First-Come-First-Service) は最も多く研究されてきた．それ以外にも，バッチ処理，優先順位のある待ち行列，タイム・シェアの待ち行列など，CPUの高速化と記憶容量の増大化など，計算機の進歩にともなって，待ち行列理論で取り扱われるサービス規律も多様化してきている．ここでは，$M/G/1$ 待ち行列の平均待ち時間を最小にするには，どのようなサービス規準が良いのか，理論的限界を与える政策について考えてみることにする．

ベイズ流の学習理論から発展したバンディット問題(bandit problem)と呼ばれるマルコフ決定過程がある．ある種のジョブ・スケジューリングもこの bandit 問題で表現できるので，これらの関連について考えることにする．

## 2. bandit 問題

$N$個の独立な機械があると仮定する．時刻 $t$ ($t=0, 1, 2, \cdots$) における $i$ 番目の機械の状態を $x_i(t)$ とする．この機械を操作すると利益 $R_i(x_i(t))$ が得られ，機械の次の状態はマルコフ的に定められた確率で推移する．また操作しなかった機械からは何も得られず，また状態の変化はないと仮定する．この時，平均利益を最大にする最適政策を求める問題が multi-armed bandit 問題である．時刻 $t$ における利益を $R(t)=R_i(x_i(t))$ として $a$ ($0<a<1$) を一定の割引き率として，目的関数

$$\mathrm{E}\sum_{t=0}^{\infty} a^t R(t) \tag{1}$$

を最大にする無限期間計画である．

### (i) スロットル・マシン

スロットル・マシンが2台($i=1, 0$)あって，どちらの機械で賭けをしたら良いかを考える．簡単のために，賭けは勝つか負けるか，2種類の結果しかないとして，勝つか負けるかは，各試行が独立同分布，いわゆる，ベルヌイ試行列とする．ここで勝つ確率$\theta_i$の正確な値は知られておらず，$\theta_i$の先験分布がベータ分布として与えられているとする．勝った回数と負けた回数から，ベイズの定理を用いて事後確率を計算すると，先験分布と同様にして，$\theta_i$の事後確率がベータ分布となる．マシンの状態をこのベータ分布のパラメータとおくと，賭けの結果により，マシンの状態がマルコフ的に推移する([1],[3])．

### (ii) 病気の治療法

ある病気に対して，2つの治療法が開発されており，どちらの治療法が有効であるかを決定するとする．治癒する確率$\theta_i$を治療しながらベイズ的に推定する．目的関数を(1)式とすると(i)の問題として解くことができる．一方，治療という問題は治癒する確率が明確に決定しない時に，$\theta_i$の推定のために，ボランティアの患者に実験段階の治療法を多数回使用するのは良い方法ではない．また，治療回数が十分大きくなった時には，治癒する確率が高い治療法がほぼ全員に受けられるという相反する2つの要請が必要である．この観点に立って，と(i)は異なる形式で研究されている[2]．

### (iii) サービスの順序

$N$個のジョブがサービス窓口に待っており，どのジョブからサービスを行なったら良いかという問題を考える．外部からの新たな到着はないとして，一度に1つのジョブしかサービスできないとする．各ジョブについては，サービス時間が非負整数値をとり，サービス分布のみが知られているとする，各ジョブのサービスは，整数時に中断して他のジョブの処理を行なうことができるとし，どのジョブからサービスを行なうのが最適であろうか．

(1)式では，サービスを行なったジョブに対しては利益を計算しているが，待ち時間のようにサービスを行なわなかったジョブに対してコストを算入する構造にはなっていない．しかし，サービスが終了しなければ将来かか

にしむら　しょういち　筑波大学　社会工学系
〒305 つくば市天王台1－1－1

るであろう割引かれた待ち時間が，サービスの完了により免除される．言いかえると，サービス完了時に割引かれた利益を得るという形で，保持費用(待ち時間)の問題を，(1)式の形のサービス終了時の利益の問題にすることができる．特に割引き率$a\to 1$とすると(1)式を最大にする政策が平均待ち時間を最小にする政策と一致する．

bandit 問題をジョブ・スケジューリングの言葉で，統一的に扱うことにする．$N$個のジョブがサービス窓口で待っているとする．ジョブの処理時間は単位時間に区切ることができ，状態$x_i(t)$のジョブを処理することによって得られた利益は$R_i(x_i(t))$とする．サービスが終了するとマルコフ的に状態が推移する．Gittins [3] は目的関数が(1)式で与えられる割引かれた無限期計画に対して，非常に明確な方法で最適政策が達成されることを次のように示した．状態$x_i$のジョブ$i$を停止時間(stopping time) $\tau_i$サービスした後に，状態$x_j$のジョブ$j$を$\tau_j$サービスした時の割引かれた平均利益は，

$$V_{ij}=\mathrm{E}\{\sum_{t=0}^{\tau_i-1} a^tR_i(x_i(t))+\sum_{t=\tau_i}^{\tau_i+\tau_j-1} a^tR_j(x_j(t))\}$$
$$=\mathrm{E}\{\sum_{t=0}^{\tau_i-1} a^tR_i(x_i(t))\}+E\{a^{\tau_i}\}\mathrm{E}\{\sum_{t=0}^{\tau_j-1} a^tR_j(x_j(t))\}$$

同様にして，$i$と$j$の順序を交換した利益を$V_{ji}$とする．2つの政策を比較すると，

$$V_{ij}-V_{ji}$$
$$=\mathrm{E}\{\sum_{t=0}^{\tau_i-1} a^tR_i(x_i(t))\}\ (1-\mathrm{E}\{a^{\tau_j}\})$$
$$-\mathrm{E}\{\sum_{t=0}^{\tau_j-1} a^tR_j(x_j(t))\}(1-\mathrm{E}\{a^{\tau_i}\})$$

となる．$V_{ij}\geqq V_{ji}$を決定するための状態$x_i$の指標(Gittins index) は

$$\gamma_i(x_i)=\max_{\tau}[\mathrm{E}\{\sum_{t=0}^{\tau-1} a^tR_i(x_i(t))\}/(1-\mathrm{E}\{a^{\tau}\})] \quad (2)$$

となる．(2)式で停止時間$\tau$で最大化している理由は，ジョブ$i$のサービスしながら，時刻0における状態$x_i$の価値より下まわらない時刻までの停止時間を$\tau_i$として，指標の最大化を行なっている．$N$個のジョブの中で，この指標が最大となるジョブからサービスを行なうことが最適であることが証明されている ([3],[10],[11])．

Gittins の方法では，指標が他のジョブには影響されずに自分のサービス時間からのみ計算できる．そして，最適政策は指標の大小によって優先順位が決定される．

Whittle [12] は multi-armed bandit に新たに外部からジョブの到着を許し，指標から決定される優先サービスで，最適政策が達成されることを示した．この場合，サービス中に到着した客の効果を考慮しなければならないので，単一のジョブについてだけでは，指標が表現できなくなる．

## 3. Harrison の方法

Harrison [4],[5] は，いくつかのクラスに分けられたジョブがポアソン到着する時，どのジョブからサービスをするのが平均待ち時間を最小にするかを考えた．サービスは，1つのジョブのサービスが開始されると，他のジョブによって割込みのない (nonpreemptive) 方法と仮定する．

$i$番目($i=1,\cdots,K$)のクラスに属するジョブは，到着率$\lambda_i$の独立なポアソン到着し，サービス時間は分布関数$F_i(x)$にしたがうとする．また，$F_i(x)$のラプラス変換を$\psi_i(\beta)$とする．$i$番目のジョブのサービスが終了すると$r_i$の利益が得られる．これは前節iii)と同様に，割引かれた保持費用をサービス終了による利益とみなしたものである．

クラス$i$ ($i=1,\cdots,K$)のジョブが$n_i$サービスを受けるために窓口の前で待っている時，システムの状態を$s=(n_1,\cdots,n_K)$とする．最適政策が優先待ち行列によって達成されるとはかぎらないが，まず優先待ち行列について解析を行なう．便宜的にクラス1の最も優先順位が高く，$K$の優先順位が最も低いものとする．

初期状態を$s$として，クラス$k$までのジョブが窓口においてすべてサービスされ，$k$より優先順位が高いジョブがなくなるまでの稼働期間(busy period)を$B_k(s)$として，これのラプラス変換を$\pi_k(s)$とする（ラプラス変換のパラメータ$\beta$の文字は一定として省略する)．また，この間の$\beta$割引きの平均利益を$U_k(s)$とおく．$\pi_k(s)$は通常の$M/G/1$待ち行列システムの稼働期間のラプラス変換はよく知られている．$s$に対して予備的な状態として，

$$s'=(n_1,\cdots,n_{k-1},\infty,n_{k+1},\cdots,n_K)$$
$$\bar{s}=(0,\cdots,0,\infty,n_{k+1},\cdots,n_K) \quad (3)$$

と$k$番目の単位ベクトル$\delta_k=(0,\cdots,0,1,0,\cdots,0)$を導入する．まず，$\alpha_k\equiv\pi_k(\delta_k)$とすると，

$$c_k\equiv U_k(\bar{s})$$
$$=U_k(\delta_k)/(1-\alpha_k) \quad (4)$$

となる．$\alpha_k$は初期状態が$k$番目のジョブが1つあり，$k$より高いジョブのサービスを行なう時の稼働期間のラプラス変換であり，分子の$U_k(\delta_k)$はその間の割引かれた利益である．もし外部からのポアソン到着がない時はGittinsの指標(2)式に対応するもの（ただし，まだ max を取っていない)．ポアソン到着が存在する場合は，新たな到着も考慮して指標が計算される．(3)式を用いると，

$$U_k(s')=U_{k-1}(s)+\pi_{k-1}(s)c_k \quad (5)$$
$$=U_k(s)+\pi_k(s)c_k \quad (6)$$

なる関係が満たされる．$s'$において$n_k=\infty$なる特異な性質を用いる．まず，優先順位が$k-1$までサービスを行なったのち，サービスを行なうのは優先順位が$k$までのジョブであるから，(5)式が成立する．一方，初期状態を$s$として優先順位が$k$までのジョブに対するサービスを行なったのちの利益は$c_k$となるので(6)式が成立する．それ故に(5)，(6)式より，

$$U_k(s)=U_{k-1}(s)+[\pi_{k-1}(s)-\pi_k(s)]c_k$$
$$=\sum_{i=1}^{k}[\pi_{i-1}(s)-\pi_i(s)]c_i \quad (7)$$

となる．$\pi_i(s)$を用いて$U_k(s)$が表現できた．

これまで優先順位の政策と比較して，まず固定されたジョブ$j$を1つだけ処理し，次に1～$k$の優先順位でサービスを行なう政策を導入する．この政策に対する稼働期間のラプラス変換，平均利益に$j$を右肩につけることにする．これまでと同様にして，

$$U^j{}_k(s)=U^j{}_{k-1}(s)+[\pi^j{}_{k-1}(s)-\pi^j{}_k(s)]c_k$$
$$=\phi_j r_j+\sum_{i=1}^{k}[\pi^j{}_{i-1}(s)-\pi^j{}_i(s)]c_i \quad (8)$$

となる．$k=K$の時 $V(s)=U_K(s)$，$V_j(s)=U_K{}^j(s)$ とおいて，2つの政策を比較すると，

$$V(s)-V_j(s)$$
$$=(1-\phi_j)c_1-\phi_j r_j-\sum_{i=1}^{K-1}(\pi_i(s)-\pi^j{}_i(s))\,(c_i-c_{i+1})$$

となる．マルコフ決定過程の政策改良法を用いると，すべての$j$に対して，$V(s)\geqq V_j(s)$であれば，(7)式が最大の平均利益となる．結論として，Harrison は次のような結果を得ている．優先順位と指標の順位が一致する政策は最適である．言いかえると，$c_1\geqq\cdots\geqq c_K$であれば，1を最も優先順位が高いクラス，……，$K$を最も優先順位が低いクラスとする政策が最適である．また，このような最適政策は，次々に指標の高いジョブを優先順位につけ加えることによって構成され，そのアルゴリズムが示されている．

## 4. 割込みのある$M/G/1$待ち行列

$M/G/1$ 待ち行列においてサービス時間を量子時間で分割し，量子時間内は割込みのないサービスを行ない，量子時間のサービスが終了した時には，サービス経過時間の記録を状態として持ったジョブがフィード・バックすることにする．量子時間を一様に細かくすると，割込みのある待ち行列が表現できる．このような割込みのある（preemptive resume）待ち行列は，サービス時間の総和が政策に不変（保存則）になる．稼働期間とその間に到着する客数は不変であるから，稼働期間内の平均利益を最大化する政策を求めることにする．

サービス時間の密度関数を$f(x)$，分布関数を$F(x)$，故障率を$g(x)=f(x)/\{1-F(x)\}$とする．瞬時$x$を使用する政策の下での指標は，(2)式または(4)式に代入することにより，

$$\gamma(x)=g(x)/\beta$$

で与えられる．故障率の概形によって，最適政策を分類する．

(i) $g(x)$ が単調増加と仮定する．FCFS でサービスを行なうと，この時，サービス経過時間が$x$のジョブに対する指標 $c_1(x)$ は，図1のように単調増加になる．FCFSでは，一度サービスを開始するとサービスが完了するまで，1つのジョブに対してサービスを継続する．ここで指標が単調増加であれば，優先順位と指標の順序が一致するので，FCFSが最適であることが証明される．

(ii) $g(x)$ が単調減少としよう．Kleinrock で述べられているＦＢ(foreground-background)スケジューリングを考える．これはサービス経過時間が大きくなると優先順位が低くなる方式である．新しいジョブが到着した時には，優先順位が高いので，一時サービスを中断して，同じレベルになると同時にサービスを行なう．$g(x)$が単調減少の場合にＦＢを用いると，図2のように指標$c_2(x)$ も単調減少になり，前と同様にしてＦＢの最適性が示される．

(iii) $g(x)$が山型とする．$[0,u)$間でFCFSで行ない，$[u,\infty)$ 間をＦＢで行なう，レベルによって異なった2つのスケジューリングの複合を考える．このような複合したサービス方式は Kleinrock によってとり扱われている．この時の指標 $c_3(x)$ も一山型となり，この方式の最適性が示される．

(iv) $g(x)$が一谷型としよう．この場合，$[0,u)$間はＦＢで行ない，$[u,\infty)$はサービス経過時間が大きくなるにつれて優先順位が高くなる方式によって，優先順位と指標 $c_4(x)$ の順序が一致することが示される（図4）．

(i)～(iii)の方式については，稼働期間中の平均利益が求められるが，(iv)については，$[0,u)$ 間と $[u,\infty)$ 間の指標が一致する関数関係を陽の形で表現できない [9].

## 5. 有限呼源待ち行列

ＣＰＵと端末の動作解析のモデルとして，しばしば，有限呼源待ち行列が用いられる．ＣＰＵに$K$個の端末が接続されている．客は端末で思考時間を経過したのち，ＣＰＵでサービスを受ける．ＣＰＵでのサービスが終了

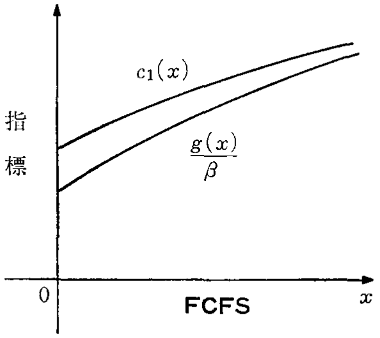


**図1**

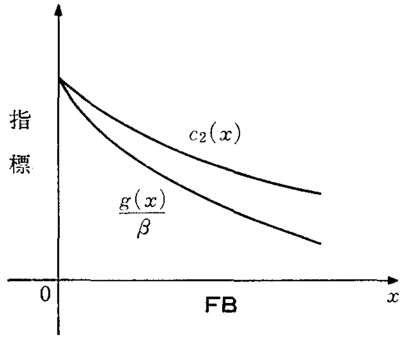


**図2**

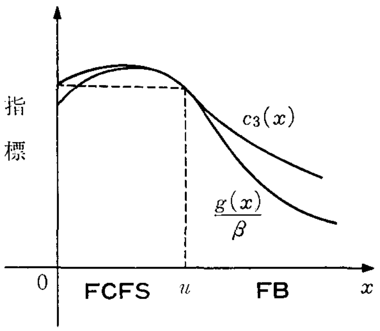


**図3**

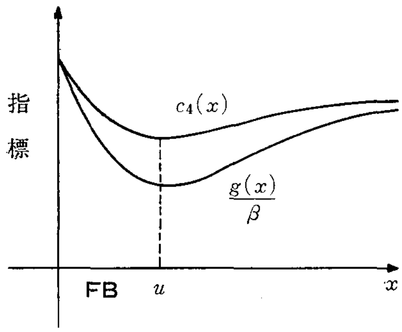


**図4**

すると再び端末にもどる．解析上の要請から各客の思考時間は独立同分布で，パラメータ$\mu$の指数分布にしたがうと仮定する．また，ＣＰＵにおけるサービス時間は，２つのクラスに分類されており，$F_1(x)$と$F_2(x)$にしたがっているとする(**図5**)．この時，サービス終了後に受ける平均利益を最大にする最適政策が求められている．ＣＰＵでのサービス時間が指数分布，または，ガンマ分布の場合，クラス１のジョブを優先順位の高いクラスとし，クラス２のジョブを低いクラスのジョブとする政策が最適であることが示される．$M/G/1$と異なり有限呼源待ち行列では，保存則が成立しない．ＣＰＵでのサービス順序によって，端末を通るフィード・バックの頻度

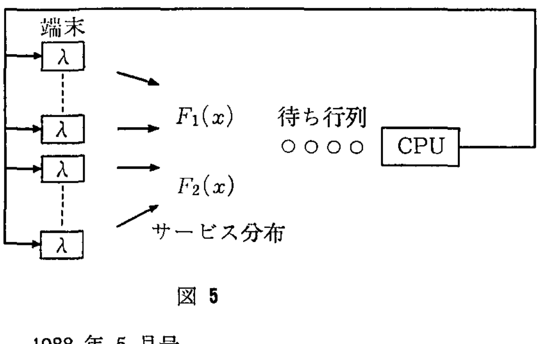


図 **5**

が異なる．したがって，稼働期間の分布が変化する．しかし，Harrisonの(5)および(6)式の関係が成立するので，最適性が証明できる（[6],[8]).

## 6. むすび

コンピュータの進歩に伴って，多種多様なサービスの方法やデータの通信の方法が開発され，実用化されている．このような問題に対しては，大量な仕事量の処理を伴うので，個々のサービス時間が明示されないが，分布は経験的に知られている．この時，どのようなサービス方法が良いかという，確率過程を用いた待ち行列のコントロールとしての解析が必要となる．理論的に解ける範囲はごく限られているので，実際問題への実用化にさいしては，理論上の長所と現実の複雑さを複合して，分析されている．

ここでは，平均利益の最大化，または，平均待ち時間の最小化と，bandit 問題との関係をまとめた．bandit 問題の考え方では，各ジョブに対して指標を計算し，指標の大きいものから優先順位をつける方法である．この考え方は，ある程度複雑なネット・ワーク待ち行列にも応用できる考え方ではないであろうか．

参 考 文 献

[1] Berry, D. A. and Fristedt, B. (1985) "BANDIT PROBLEMS", Chapman and Hall.

[2] Bather, J. A. (1981) "Randomized allocation of treatments in sequential experiments", *J. Roy. Statist. Soc.* **43**, 265-292.

[3] Gittins, J. C. (1979) "Bandit processes and dynamic allocation indices", *J. Roy. Statist. Soc.* **41**, 148-177.

[4] Harrison, J. (1975) "A priority queue with discounted linear costs", *Oper. Res.* **23**, 260-269.

[5] Harrison, J. (1975) "Dynamic scheduling of a multiclass queue : Discount optimality", *Opes. Res.* **23**, 270-282.

[6] Hirayama, T. (1987) "Nonpreemptive scheduling of a finite-source queue with two customer classes", *J. Oper. Res. Society Japan.* **30**, 200-217.

[7] Kleinrock, L. (1976) "QUEUEING SYSTEMS", Vol. 2, Wiley.

[8] Nishimura, S. "Priority job scheduling and finite-source queue" (投稿中)

[9] Nishimura, S. "*M/G/*1待ち行列におけるＦＣＦＳとＦＢスケジューリング" (投稿中)

[10] Varaiya, P., Walrand, J. and Buyukkoc, C. (1985) "Extensions of the multi-armed bandit problem : the discounted case", *IEEE Trans. Automat. Contr.* **30**, 426-439.

[11] Whittle, P. (1980) "Multi-armed bandits and the Gittins index", *J. Roy. Statist. Soc.* **42**, 143-149.

[12] Whittle, P. (1981) "Arm-acquiring bandits", *Ann. Prob.* **9**, 284 292.