

5 群 (通信・放送) - 1 編 (待ち行列理論とシミュレーション)

1 章 モデルと基礎概念

(執筆者: 塩田茂雄) [2008 年 10 月受領]

概要

待ち行列理論は、電話網の設計のために、コペンハーゲンの電話会社に勤める A.K. Erlang が確率論を応用した論文を 1908 年に発表したことに始まるとされる。それ以降も、待ち行列理論は情報通信技術の進歩とともに歩んできた。

インターネットの普及に代表されるように、近年、通信ネットワークは大きな変化を遂げた。通信ネットワークのブロードバンド化に伴いネットワーク内のトラフィック量は 1990 年代後半より指数関数的に増大している。また、通信ネットワークのマルチメディア化によりネットワークの利用形態も多様化し、電話網の時代に比べて、トラフィックは統計的に複雑な挙動を示すようになった。ネットワークの設計、制御の基礎理論となるべき待ち行列理論にも、異速度混在型ネットワークへの対応、相関をもつ到着過程のモデル化、長時間相関性を有する分布関数の考慮、待ち時間など各種特徴量の分布の裾の評価など解決すべき様々な課題が生じ、これら課題を解決するために待ち行列理論は、近年、一層の発展を見た。

待ち行列モデルは通信ネットワークの混雑現象を解析するためのモデルとして使われることが多い。「混雑」は本質的にランダムな現象であり、そのモデル化には確率の考え方が必要となる。従って、待ち行列モデルの記述にも確率変数や確率過程といった確率論の言葉を用いる。解析可能な待ち行列モデルはマルコフモデルであることが多い。しかし、マルコフモデルでなくとも成立する普遍的な関係式 (リトルの公式など) が存在し、これら関係式を見通しよく証明するために、点過程論によるアプローチが使われる。

【本章の構成】

本章では、待ち行列モデルの必要性と記述法 (1-1 節)、待ち行列を確率モデルで記述するための基本概念である到着過程とサービス時間分布 (1-2 節)、また待ち時間など待ち行列モデルの様々な評価尺度 (1-3 節) について説明する。さらに、定常な待ち行列モデルを解析するための基本的な道具となる点過程論について概説する (1-4 節)。なお本章では以下の記号を用いる。

- $P[A]$: 事象 A が生ずる確率
- $E[X]$: 確率変数 X の期待値
- $P_N[A]$: 事象 A が生ずる Palm 測度における確率 (Palm 測度については本章 1-4 を参照)
- $E_N[A]$: 確率変数 X の Palm 測度における期待値
- $\mathbf{1}(A)$: A が正のとき 1, 偽のとき 0 をとる関数 (identity function)

また \mathbb{N} , \mathbb{Z} , \mathbb{R} はそれぞれ自然数, 整数, 実数の集合を表す。

5 群 - 1 編 - 1 章

1-1 モデルの記述

(執筆者：塩田茂雄)[2008年10月受領]

1-1-1 モデルの必要性

モデルという言葉が初めて学術用語として用いられたのは、理論物理学の分野らしい¹⁾。例えば、物質の様々な物理現象は、物質を構成する原子・分子がシュレディンガー方程式のような基礎方程式に従って運動する結果、現れるものである。しかし、アボガドロ数オーダーの分子から構成される物質のマクロな現象を、個々の粒子のミクロな運動を厳密に積み上げることで説明するのは困難であろう。通常は、物理現象に最も本質的に効く要素に焦点をあて、その他の要素を大胆に無視する「モデル」を構築し、物理現象を説明する手法が使われる。当然、そこで用いられるモデルは現実の物質を近似したものにすぎないが、そのモデルが物理現象を適格に捉えたものであれば、モデル化誤差を超えて我々に有用な知見を与えてくれる。

このようなアプローチは物理現象のみならず、人工的なシステムの解析を行う上でも有効である。例えば、コンピュータネットワークの輻輳（ふくそう：混雑の意味）は、通信回線の帯域不足により、ルータなどの中継装置の内部にデータの待ち行列が生ずることなどで引き起こされる。しかし、ルータの構成や処理機能は大変複雑であり、これらをすべて考慮して厳密に輻輳現象を解析することは難しい。通常は、通信回線の帯域に比べてルータ内部の処理が充分高速であるとして、ルータの出力ポートに待ち行列が発生するというモデルを用いる（図 1-1 参照）。このモデルはルータの処理モデルとしては簡易すぎるが、輻輳現象の解析には充分有効である。現象の本質をついたシンプルなモデルを構築することは、システムの性能評価上、最も重要な貢献であると言っても差し支えない。

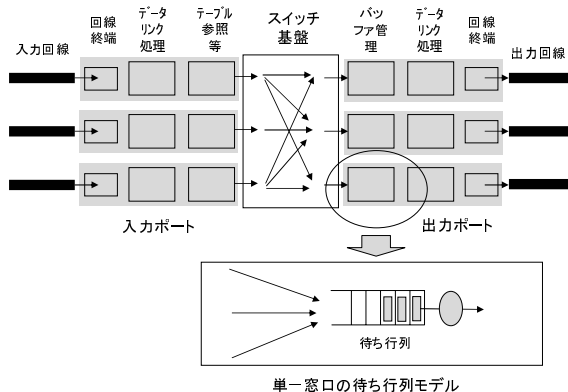


図 1-1 ルータのモデル化例

1-1-2 通信ネットワークと待ち行列モデル

インターネットのようなパケット交換型ネットワークではパケットと呼ばれるデータのかたまりを中継装置がリレー方式で通信相手のコンピュータに転送する。中継装置はパケット

を受け取ると、次の中継装置に通信回線の帯域幅で決まる速度でパケットを転送するが、たまたま、複数のコンピュータからパケットを大量に受け取ると、次の中継装置への転送が間に合わず、一時的にパケットの「待ち行列」が形成されてしまう（図 1・1 参照）。

一般に、通信ネットワークでは、このように一時輻輳による待ち行列がシステム内に頻繁に形成され、ネットワークの性能に大きな影響を及ぼすことが知られている。このために、通信ネットワークの性能評価においては古くから待ち行列モデルが使われてきた。

「待ち」は本質的に確率現象であるため、待ち行列モデルの解析に確率論は欠かせない²⁾。例えば、ATM（自動現金引き出し機）の前にできる待ち行列を考えよう。ATM には平均して 2 分間に 1 人の割合で人が立ち寄り、現金引き出しに平均 1 分かかるとする。さて、この情報だけで、ATM の前にできる平均待ち行列長はわかるだろうか。残念ながら答えは NO である。人が正確に 2 分おきに立ち寄り、常に 1 分で現金を引き出すならば、待ち行列は発生しない。しかし、人の到着間隔や現金引き出し時間にバラツキがあると、待ち行列が形成される。つまり、待ち行列長は、到着間隔やサービス時間（ATM の例では現金引き出し時間）の期待値のみならず、2 次以上のモーメントにも強く依存する。先に述べたパケットの待ち行列の場合も同様であって、パケットの到着間隔とパケット長（パケットの待ち行列の場合、窓口でのサービス時間はパケット長を通信回線の帯域幅で割ったものに相当する）の期待値だけでなく、その高次モーメントがパケットの待ち時間に大きく影響する。

一般にシステムのランダム性（バラツキの大きさ）が高いほど、待ち行列が形成されやすい。このランダム性を評価するために、確率論という道具が必要不可欠である。

なお、電話網のような回線交換型ネットワークのモデル化には交換機間の回線が窓口となる「複数窓口待ち行列モデル」が使われる（図 1・2 参照）。回線交換網では回線がすべて使われている状態（回線全塞状態）で新たに到着した呼は直ちに呼損となる。従って、回線交換網用に用いられる待ち行列モデルには、いわゆる「待ち室」は存在しない。

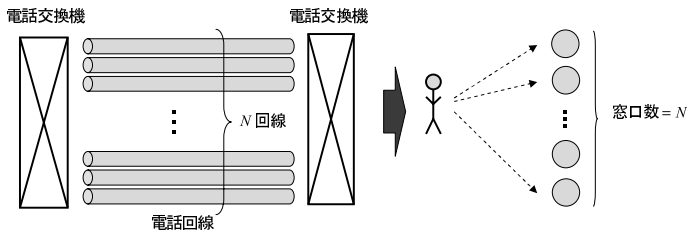


図 1・2 電話網のモデル化例

1-1-3 ケンドールの記法

以下では、待ち行列の最も基本的なモデルとして、客が一つの待ち行列を訪れ、サービス終了後にシステムを退去する「単一ノード待ち行列モデル」を説明する。単一ノード待ち行列モデルは、(1) 到着過程、(2) サービス時間分布、(3) (サービスを行う) 窓口の数、(4) 最大収容人数、(5) サービス規律（どのような順序で来た客をサービスするか）により特徴づけられる。待ち行列理論では、これら要素による特徴づけを「ケンドールの記法（Kendall's

notation)」を用いて表すのが一般的である。ケンドールの記法では、待ち行列システムは

$$A/B/N/K/Z$$

と表される。ここで、 A には到着過程の種類、 B にはサービス時間分布の種類を表す記号が入る。また、 N はサービスを行う窓口の数を、 K は最大収容人数を表す（待ち行列内の収容人数が最大値 K に達しているときに来た客は、サービスを受けずに立ち去る）。 Z は処理規律を表し、例えば来た客が先着順にサービスされる場合は Z は FIFO (First-In-First-Out) と書かれる。なお、収容人数に制限が無く、また処理規律が先着順の場合、 $A/B/N/\infty/\text{FIFO}$ と書かず、単に

$$A/B/N$$

と書くことが多い。

到着過程とは、客の到着時刻（到着間隔）を支配する確率法則のことを指す（本章 1-2 で詳述する）。到着間隔が互いに独立で同一の分布に従う到着過程を「再生過程」と呼ぶ。再生過程はケンドールの記法では GI で表される（すなわち到着過程が再生過程の場合、 A には「 GI 」という記号が入る）。到着間隔が指数分布に従う再生過程を「ポアソン過程」と呼ぶ。ポアソン過程はケンドールの記法では M で表される。到着過程の確率法則に何の制限も設けないときは、到着過程を G という記号で表す。

サービス時間分布とは、客が窓口で受けるサービス時間を支配する確率法則（確率分布）を指す（本章 1-2 で詳述する）。到着過程同様、各々の客のサービス時間が互いに独立で同一の分布に従う場合、ケンドールの記法では B に「 GI 」という記号が入り、サービス時間が互いに独立な指数分布に従う場合、 B には「 M 」という記号が入る。サービス時間の確率法則に何の制限も設けないときは、サービス時間は G という記号で表される。例えば、 $M/M/1$ は客がポアソン過程に従って到着し、客の要求するサービス時間が指数分布に従い、窓口が一つの待ち行列システムを表す。

参考文献

- 1) 木村英紀, モデル学は可能か, 日本オペレーションズリサーチ学会誌, vol.50, no.8, 2005.
- 2) 高橋幸雄, 森村英典, 混雑と待ち, 朝倉書店, 2001.

5 群 - 1 編 - 1 章

1-2 到着過程とサービス時間分布

(執筆: 塩田茂雄) [2008 年 10 月受領]

待ち行列モデルにおける基本的なデータは客の到着時刻と窓口で受けるサービス時間である。待ち行列システムが時刻 0 から稼働を始めたとし、システムに到着した客に順に 1 から番号をふり、 n 番目の客の到着時刻を T_n 、サービス時間を σ_n により表すこととする。待ち行列理論では、通常、 $\{T_n, \sigma_n\}_{n \in \mathbb{N}}$ に何らかの確率法則を仮定して解析を行う。到着時刻列 $\{T_n\}_{n \in \mathbb{N}}$ を支配する確率法則を到着過程と呼ぶ。一方、各々の客のサービス時間の分布（正確には、サービス時間列 $\{\sigma_n\}_{n \in \mathbb{N}}$ の周辺分布）をサービス時間分布と呼ぶ。本節では代表的な到着過程とサービス時間分布の幾つかについて概説する。

1-2-1 ポアソン過程

本章 1-1 で述べたように、到着間隔が指数分布に従う再生過程をポアソン過程と呼ぶ。ポアソン過程は最も基本的な到着過程であり、待ち行列理論においても重要な役割を果たす。以下では、 $(0, t]$ 間の到着客数を $N(t)$ で表す。ポアソン過程は次の性質を持つことが知られる。

- (1) 任意の自然数 n と任意の $0 < t_1 < t_2 < \dots < t_n < \infty$ に対して、

$$N(t_1), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$$

は独立な確率変数列となる。

- (2) 各 $t > 0$ において、 $N(t)$ の増分 $(N(t) - N(t-))$ は 0 または 1 に等しい。
 (3) 各 $s, t > 0$ に対して、 $N(t+s) - N(t)$ の分布は t とは独立で、 s にのみ依存する。

ポアソン過程では、「ある時刻以降の到着客数や到着時刻は、その時刻以前の客の到着履歴とは無関係である」という意味での「無記憶性」が成り立つ。条件 (1) は、この無記憶性を数学的に表現したものである。ケンドールの記法において、ポアソン過程を表す「M」という記号は、無記憶性 (Memoryless) の頭文字をとったものと言われている。条件 (2) は、同時に 2 人以上の到着が生じないことを示す。この条件を満たす到着過程を「単純」な到着過程と呼ぶ（単純な到着過程はポアソン過程以外にも存在する）。条件 (3) は、客の到着率が一定であることを示す。なお、条件 (1), (2), (3) を満たす到着過程はポアソン過程以外には存在しない。すなわち

定理 1. 条件 (1), (2), (3) は到着過程がポアソン過程であるための必要十分条件である。

条件 (1) と (2) のみを満たす到着過程は「非正常ポアソン過程」と呼ばれる。非正常ポアソン過程は無記憶性と単純性を満たし、客の到着率が時刻とともに変動する到着過程である。

さらに、ポアソン過程の幾つかの重要な性質を述べておく。次の補題は「ポアソン過程」の名前の由来となる事実を示している。

補題 2. ポアソン過程において、 $N(t)$ の増分はポアソン分布に従う。すなわち

$$P[N(t+s) - N(t)] = \frac{(\lambda s)^n}{n!} e^{-\lambda s}$$

ポアソン過程は一定の到着率のもとで客がランダムに到着する到着過程と考えるとイメージしやすい。次の補題はこの事実を示したものである³⁾。

補題 3. ポアソン過程において、時刻 0 から時刻 t までの到着客数が与えられた条件の下で、それぞれの客の到着時刻は $(0, t]$ 間に一樣かつ独立に分布する。すなわち

$$P[T_1 \leq t_1, T_2 \leq t_2 \leq \dots T_n \leq t_n | N(t) = n] = \frac{n! \prod_{k=1}^n t_k}{t^n}$$

1-2-2 指数分布

指数分布は待ち行列理論において頻繁に登場する、最も基本的なサービス時間分布である。確率変数 X が以下の分布関数に従うとき、 X は (パラメータ μ の) 指数分布に従うという。

$$P[X \leq t] = \begin{cases} 1 - e^{-\mu t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

サービス時間 σ が (パラメータ μ の) 指数分布に従うならば、その期待値は $1/\mu$ に等しい。

$$E[\sigma] = \int_0^{\infty} P[\sigma > t] dt = \int_0^{\infty} e^{-\mu t} dt = 1/\mu$$

実際、今、ある客がサービスを受け始めてから、ちょうど時間 t が経過したとしよう。客の残りのサービス時間を「残余サービス時間」と呼ぶ。一般に残余サービス時間の分布は t に依存するが、サービス時間が指数分布に従うならば、残余サービス時間分布は t に非依存であるという、次の驚くべき結果が得られる。

補題 4. サービス時間が (パラメータ μ の) 指数分布に従うならば、残余サービス時間も (パラメータ μ の) 指数分布に従う。すなわち

$$P[\sigma \leq s + t | \sigma \geq t] = 1 - e^{-\mu s}$$

補題 4 は、残余サービス時間がこれまでにその客が受けたサービス時間と無関係に決まるという意味での「無記憶性」を表している。ケンドールの記法において、指数分布も「M」という記号を用いて表されるのは、この理由による。

1-2-3 位相型分布

到着間隔もしくはサービス時間分布を指数分布以外の分布におきかえることで、待ち行列の様々なモデル化が可能になるが、多くの場合、解析可能性は失われる。例えば、到着間隔の独立性とサービス時間相互の独立性を残しつつ、到着間隔とサービス時間の分布を指数分布以外に置き換えたモデル (GI/GI/1 モデル) を厳密に解析する手法は知られていない。しかし、到着間隔もしくはサービス時間の分布にマルコフ性 (無記憶性を拡張した概念) を持

たせれば、解析可能性を維持しながら指数分布の拡張を行うことが可能である。

マルコフ性を有する確率分布の一つに「位相型分布」¹⁾がある。位相型分布は「位相 (phase)」と呼ばれる内部変数 $J(t)$ を持つ。 $J(t)$ の時間変化は有限状態マルコフ過程に従う。 $J(t)$ は決まった初期状態 α から出発し、推移率行列 Q により推移し、最終的に「吸収状態」($J(t)$ が取りうる状態の一つ) に吸収される。位相型分布とは、内部変数 $J(t)$ が初期状態から出発して吸収状態に吸収されるまでの時間 X の確率分布を表すものであり、具体的には次の分布関数を有する。

$$P[X \leq t] = \begin{cases} 1 - \alpha e^{tU} \mathbf{1} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

ここで U は非吸収状態間の遷移を表す推移率行列、 $\mathbf{1}$ は要素がすべて 1 の縦ベクトルである。指数分布の分布関数と比較すると、位相型分布が指数分布のある種の拡張になっていることが理解できよう。事実、位相型分布は指数分布はもちろん α -アーラン分布、超指数分布を含む、非常に広いクラスの確率分布である。ケンドールの記号では、到着間隔が位相型分布に従う再生過程を「PH」という記号で表す。サービス時間が互いに独立な位相型分布に従う場合も、サービス時間分布は「PH」という記号で表される。

1-2-4 相関を持つ到着過程

すべての再生過程は、到着間隔を位相型分布でモデル化した再生過程により任意の精度で近似できる。しかし、現実には再生過程でモデル化できない到着過程が数多く存在する。特に、音声や動画を可変ビットレートでコーデックした際に発生するバケットの到着過程は正の相関を持つことが知られる。ここで「到着過程が相関を持つ」とは、客の到着間隔が互いに独立でなく、相関を持つ (互いに従属である) ことを指す。到着過程が正の相関を持つ場合、一旦、到着間隔が短くなりだすと、しばらく到着間隔が短い状態が継続し、逆に、一旦、到着間隔が長くなりだすと、しばらく到着間隔が長い状態が継続する。このような到着過程は再生過程では記述できない。

このような事情から、これまでに相関を持つ到着過程のモデルが幾つか提案されており、特にマルコフ性を持たせて解析可能性を残した到着過程のモデルが頻りに利用されてきた。Markov Modulated Poisson Process (MMPP) や Markovian Arrival Process (MAP) ¹⁾ がマルコフ性を有する到着過程として有名である。例えば MAP は、位相型分布同様、内部変数 $J(t)$ を持つ。 $J(t)$ の時間変化は有限状態マルコフ過程に従う。 $J(t)$ の推移率行列は $C + D$ の形を有し、 D による推移が生じた際には、客の到着が生ずる (位相型分布と異なり、吸収状態はもたない)。MAP の詳細については、本編 3 章を参照いただきたい。

参考文献

- 1) 宮沢政清, 待ち行列の数理とその応用, 牧野書店, 2006.
- 2) 滝根哲哉, 伊藤大雄, 西尾章治郎, ネットワーク設計理論 (岩波講座インターネットシリーズ第 5 巻), 岩波書店, 2001.
- 3) S. Ross, Stochastic Processes, Wiley, 1996.

5 群 - 1 編 - 1 章

1-3 評価尺度

(執筆: 塩田茂雄) [2008 年 10 月受領]

1-3-1 客の経験値に基づく評価尺度

待ち行列モデルを解析する目的の一つは、例えば通信ネットワークの性能を調べて、その設計や制御手法に反映させることである。通信ネットワークの性能は、End-to-End での遅延時間、パケットロス率、スループットなどの指標で評価され、これらはいずれも、ルータ内部で各パケットが経験する遅延時間やパケットロス率に依存する。ルータの処理を待ち行列でモデル化する場合、遅延時間は待ち行列での客の待ち時間、滞在時間と、またパケットロス率は客の損失率（待ち行列到着時にシステム内容数がシステムの最大収容人数に達しており、到着客がシステムに入れない確率）と深く関わる。多くの場合、これら客の経験値に基づく指標が、待ち行列モデルの評価尺度として使われる。

まず、いくつかの定義を述べておく。滞在時間は客がシステムに到着してから、サービスを受け終わってシステムから退去するまでの時間で定義される。待ち時間は滞在時間からサービス時間を引いたものである。いったんサービスを受け始めた客が他の客から割り込まれることがない（割り込み中断がない: non-preemptive）待ち行列モデルの場合は、待ち時間は到着時点からサービス開始時点までの時間に等しい。損失率は、客到着時点において、システム内容数が最大収容人数に等しい確率である。今、 n 番目の客の滞在時間、待ち時間を S_n 及び W_n で表すこととする。もし、充分時間が経過した後にシステムが定常状態に落ち着くならば、定常状態における評価尺度が解析対象になる。つまり、システムが定常状態に収束し、以下の極限

$$S \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} S_n, \quad V \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} V_n$$

が（法則収束の意味で）存在するならば、収束先の確率変数 S や V の期待値を用いて、客が経験する平均滞在時間や平均待ち時間を評価する。また S や V の分布が分かれば、滞在時間や待ち時間の $x\%$ 値など、より精密な評価が可能になる。

1-3-2 待ち行列システムの特性に基づく評価尺度

客の経験値に基づく評価尺度以外に、システム内容数、待ち行列長、workload など待ち行列システム自身の特性も解析の対象となる。システム内容数はサービス中の客と待ち室にいる客の総和で定義される。待ち行列長は待ち室にいる客の総数である。workload（「仮待ち時間」と呼ばれることもある）はシステム内の客が退去するまでに受けるべき残余サービス時間の和で定義される。時刻 t のシステム内容数、待ち行列長、workload をそれぞれ $L(t)$ 、 $Q(t)$ 、 $W(t)$ と記す。充分時間が経過した後にシステムが定常状態に落ち着くならば、以下の極限

$$L \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} L(t), \quad Q \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} Q(t), \quad W \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} W(t)$$

が存在し、これら定常状態での評価尺度を評価することで、待ち行列システム自身の特性の評価ができる。

なお、workload については次の漸化式が存在する。

$$W(T_1) = 0, \quad W(T_{n+1}) = \max\{W(T_n) + \sigma_n - \tau_n, 0\}$$

上式は Lindley の方程式と呼ばれ、システムの安定性などを調べる際などに利用される。

1-3-3 客の経験値と待ち行列システムの特性間の関係

待ち時間や滞在時間など客の経験値と、システム内客数や待ち行列長など待ち行列システムの特性の間にはある種の関係が存在する。例えば $L(t)$ は、 T_n 、 S_n と以下の関係にある。

$$L(t) = \sum_n \mathbf{1}(T_n \leq t < T_n + S_n) \quad (1\cdot1)$$

また、割り込み中断がない待ち行列モデルの場合

$$Q(t) = \sum_n \mathbf{1}(T_n \leq t < T_n + V_n) \quad (1\cdot2)$$

さらに待ち室が無限大の場合

$$W(t) = \sum_n g(t - T_n; V_n, \sigma_n) \mathbf{1}(T_n \leq t) \quad (1\cdot3)$$

が成り立つ。ここで $g(t, x, y)$ は以下で定義される関数である（図 1・3 参照）。

$$g(t; x, y) \stackrel{\text{def}}{=} \begin{cases} y & 0 \leq t \leq x \\ y - t + x & x < t \leq x + y \\ 0 & t < 0, t > x + y \end{cases}$$

定常状態における客の経験値 (S や V) と待ち行列システムの特性値 (L や Q) の間には、リトルの公式として知られる普遍的な関係式が成立する。本章 1-4 において、式 (1・1) や式 (1・2) を用いて、この関係式を導出する。

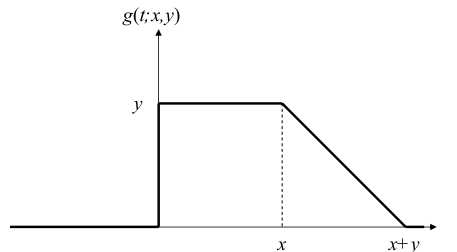


図 1・3 $g(t; x, y)$

参考文献

- 1) F. Baccelli and P. Bremaud, Elements of Queueing Theory, 2nd ed., Springer, 2003.

5 群 - 1 編 - 1 章

1-4 点過程論

(執筆者: 塩田茂雄) [2008 年 10 月受領]

本章 1-3 で述べたように待ち行列モデルの解析は、一般に、モデルの定常状態を調べることに帰着される。定常状態を数学的に解析できるモデルはそれほど多くはない。厳密な解析ができるモデルは、待ち行列の動きをマルコフ過程としてモデル化するマルコフモデルにほぼ限られる(マルコフモデルについては本編 2 章を参照のこと)。

マルコフモデルは非常に強力な解析ツールであるが、反面、到着過程もしくはサービス時間分布がマルコフ性を満たさなければならないという制約が生ずる。しかし、待ち行列モデルにはマルコフモデルでなくとも成立する幾つかの普遍的な関係式が存在する。マルコフモデルに基づいて、これら普遍的な関係式を証明するのは必ずしも容易でなく、マルコフモデルを超えた一般化は難しい。

本章では、到着過程やサービス時間について定常性が仮定できる場合に、待ち行列モデルに広く成立する普遍的な関係式を証明するための強力な道具となる点過程論について紹介する。

1-4-1 定常性

定常性とは、確率過程の確率法則が時間の並進に対して不変であることを指す概念である。例えば、ある到着過程の実現値(サンプルパス)において、客の到着が

$$\dots, -8s, -2s, -1s, 3s, 9s, 12s, 13s, \dots$$

であったとする(s は秒を表す)。このとき、例えば時間を 2 秒早めにずらしたサンプルパス

$$\dots, -10s, -4s, -3s, 1s, 7s, 10s, 11s, \dots$$

も同じ確率で出現するというのが、直感的な定常性の意味である。

定常性を正確に理解するためには、確率空間の知識が必要になる。ここでは、確率空間の初歩的な知識を仮定して話を進める。到着過程 $\{T_n\}_{n \in \mathbb{Z}}$ 、サービス時間列 $\{\sigma_n\}_{n \in \mathbb{Z}}$ が確率空間 (Ω, \mathcal{F}, P) 上で定義されているとする。客の到着は無限の過去からあるとし、客の番号は到着時刻列が以下を満たすように定める。

$$\dots T_{-2} \leq T_{-1} \leq T_0 \leq 0 < T_1 \leq T_2 \leq \dots$$

次に Ω 上の時間ずらしの作用素 θ_s を定義する。 $\theta_s \omega$ ($\omega \in \Omega$) は、 ω による実現値(サンプルパス)を s だけ時間を進ませた時計で見ることに対応する。例えば、 $(s, t]$ 間の到着客数を $N(s, t)$ とすると、 $N(s, t)$ は (Ω, \mathcal{F}, P) 上の確率過程であり、 Ω 上の任意の要素 ω に対する $N(s, t)$ の実現値を $N(s, t)(\omega)$ と記すならば、

$$N(0, t)(\theta_s \omega) = N(s, s+t)(\omega) \quad (1.4)$$

が成立するとする(これを $N(0, t) \circ \theta_s = N(s, s+t)$ と記す)。同様に

$$T_n \circ \theta_s = T_{N(s)+n} - s \quad (1.5)$$

が成立する .

さて、確率測度 P が定常であるとは、任意の $A \in \mathcal{F}$ 及び $s \in \mathbb{R}$ に対して

$$P(A) = P(\theta_s A)$$

が成立することをいう . 確率測度 P が定常であるとき、式 (1.4) (もしくは式 (1.5)) を満たす到着過程を「定常点過程」という . サービス時間列 $\{\sigma_n\}_{n \in \mathbb{Z}}$ が、やはり時間ずらし作用素 θ_s の影響を受けて変化するならば、 $\{T_n, \sigma_n\}_{n \in \mathbb{Z}}$ は「マーク付き定常点過程」になる .

連続時間型確率過程 $X(t)$ が時間ずらし作用素 θ_s の影響を受けて時間がずれる、すなわち $X(t) \circ \theta_s = X(s+t)$ が成立するならば、「 $X(t)$ は θ_s に連動する」という . θ_s に連動する確率過程は定常である . システム内容数、待ち行列長、workload はいずれも θ_s に連動するので (式 (1.1) などを利用して証明できる)、これら確率過程は定常である . 従って、時刻 0 のシステム内容数 $L(0)$ は定常状態でのシステム内容数に等しく、その期待値 $E[L(0)]$ も定常状態での平均システム内容数に等しい .

1-4-2 Palm 測度

$L(t)$ が定常であっても、到着間隔列が定常とは限らない . n 番目の客と $n+1$ 番目の客との到着間隔を $\tau_n (= T_{n+1} - T_n)$ で表すと、平均的に τ_0 は τ_1 に比べて大きくなる (到着間隔が大きいほど時間原点はトラップされやすい) . つまり $\{\dots, \tau_{-1}, \tau_0, \tau_1, \dots\}$ は同分布に従うとは限らない . この結果、 $E[\tau_0]$ は真の平均到着間隔に必ずしも一致しない .

しかし、 $N(s, t)$ が定常であれば、 $\{\dots, \tau_{-1}, \tau_0, \tau_1, \dots\}$ が同分布に従う、つまり定常となる確率測度を構築することができ、それは事実上、客が到着時点で見える定常的なシステムの様子を表す確率測度となる . Palm 測度とは、このような確率測度であり、以下で定義される .

$$P_N(A) (= E_N[\mathbf{1}(A)]) \stackrel{\text{def}}{=} \frac{1}{\lambda} E \left[\int_0^1 (\mathbf{1}(\theta_s^{-1} A) N(ds) \right], \quad \lambda \stackrel{\text{def}}{=} E[N(1)]$$

Palm 測度のもとでの期待値である $E_N[\tau_0]$ や $E_N[\sigma_0]$ は、真の平均到着間隔や平均サービス時間に一致する . 一方、例えば $E_N[L(0-)]$ は、定常状態における任意時点での平均システム内容数に一致しない . $E_N[L(0-)]$ は客到着時の平均システム内容数になる .

Palm 測度においては、次の Campbell の公式が重要である .

定理 5. [Campbell の公式] $\{T_n, Z_n\}$ をマーク付き定常点過程とする . 任意の可測関数 f に対して、次が成立する .

$$E \left[\sum_n f(T_n, Z_n) \right] = \lambda E_N \left[\int f(t, Z_0) dt \right]$$

1-4-3 Little の公式

Campbell の公式から、幾つかの普遍的な関係式が導かれる . 式 (1.1) に Campbell の公式を適用し (マークは滞在時間 $\{S_n\}$ とする)、 $f(T_n, S_n) = \mathbf{1}(T_n \leq 0 < T_n + S_n)$ とおくことにより

$$\begin{aligned}
 E[L(0)] &= E\left[\sum_n \mathbf{1}(T_n \leq 0 < T_n + S_n)\right] \\
 &= \lambda E_N \int \mathbf{1}(t \leq 0 < t + S_0) dt \\
 &= \lambda E_N[S_0]
 \end{aligned}$$

つまり「平均システム内容数は到着率と平均滞在時間の積に等しい」。これをリトルの公式と呼ぶ。同様に式 (1・2) に Campbell の公式を適用し

$$E[Q(0)] = \lambda E_N[V_0]$$

を得る。つまり「平均待ち行列長は到着率と平均待ち時間の積に等しい」。更にサービス中の客数について同様の関係式を考えることにより

$$P[L(0) > 0] = \lambda E_N[\sigma_0]$$

が得られる。窓口が一つの待ち行列の場合、 $\rho \stackrel{\text{def}}{=} \lambda E_N[\sigma_0]$ は使用率と呼ばれる。

1-4-4 Brumelle の公式

リトルの公式以外にも、Campbell の公式から幾つかの普遍的な関係式を導くことができる。例えば、式 (1・3) に Campbell の公式を適用すると

$$\begin{aligned}
 E[W(0)] &= E\left[\sum_n g(-T_n; V_n, \sigma_n) \mathbf{1}(T_n \leq 0)\right] \\
 &= \lambda E_N \int g(-t; V_0, \sigma_0) \mathbf{1}(t \leq 0) dt \\
 &= \lambda E_N \int_0^\infty g(t; V_0, \sigma_0) dt \\
 &= \lambda E_N[\sigma_0 V_0] + \frac{\lambda}{2} E_N[\sigma_0^2]
 \end{aligned}$$

特に σ_0 と V_0 が独立の場合 (GI/GI/1/ FIFO など)

$$E[W(0)] = \lambda E_N[\sigma_0] E_N[V_0] + \frac{\lambda}{2} E_N[\sigma_0^2]$$

これは Brumelle の公式と呼ばれる。到着過程がポアソン過程で先着順サービスの場合 (M/G/1 FIFO queue), $E[W(0)] = E_N[V_0]$ であるから

$$E_N[V_0] = \frac{\lambda E_N[\sigma_0^2]}{2(1-\rho)} = \frac{\rho E_N[\sigma_0]}{2(1-\rho)} \left(1 + \frac{\text{Var}[\sigma_0]}{E_N[\sigma_0]^2}\right)$$

上式はボラチェック・ヒンチンの平均値公式と呼ばれている。

参考文献

- 1) 宮沢政清, 待ち行列の数理とその応用, 牧野書店, 2006.
- 2) F. Baccelli and P. Bremaud, Elements of Queueing Theory, 2nd ed., Springer, 2003.