1. 待ち行列理論の概念とケンドールの記号

概要: 待ち行列理論はサービス業務を扱う。サービス業務では、顧客は系を訪れ、待ち、サービスを受け、立ち去るサービスには様々が形態があり、その形態に応じて理論も発展している。そのサービスの種類はケンドールの記号で一般的に表現されている。待ち行列理論では、この現象をどう扱うのかをこの章で説明する。

キーワード:待ち行列理論;ケンドールの記号; 来客密度; サービス率;オペレーション率;定常状態

1.1. 序

待ち行列理論は銀行やコールセンターのようなサービス業務の数理解析に用いられる。その中では、客は系を訪問し、待ち、サービスを受け、系を立ち去る。そのサービスの種類いろいろあり、その種類に応じて理論は発展している。そのサービスの種類はケンドールの記号で一般的に表現されている。この解析は一般性があり、重要であるため、多くの研究がなされている。 これらの本は基本的であるか、専門家のためのものであることが多い。ここでは、初等的な知識で最先端のモデルまで理解できるようになることを目指す。

待ち行列は長い列、または長い待ち時間になる場合に適用される、というイメージを持っている。しかし、このような場合は待ち行列理論は必要ではなく、もっと簡単に系を扱うことができる。長い待ち行列や待ち時間は系のもつバラツキを吸収し、平均値で現象を議論できるためである。

待ち行列理論は、比較的問題がない状態までもっていった場合の最後の仕上げに用いられる。この場合は、系のバラツキが最終的な結果に影響を及ぼす。

1.2. 待ち行列の発生

銀行の待合室を考えよう。 この待合室、対応する場所のことを系と呼ぶ。

顧客は銀行を毎分 5 人訪れると仮定する。1 人あたりの対応時間は 4 分とする。この状況を図 1 に示す。これらの状況が周期的に起こるならば、どの顧客も対応を待つことはない。この状況も図 1 に示されている。

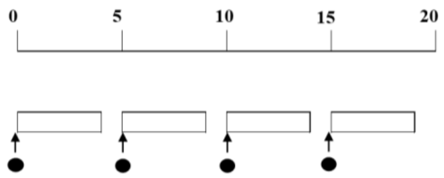

図 1　周期的な訪問、および一定のサービス時間

しかしながら、顧客は系を系統的に訪れるわけではない。顧客の訪問時間間隔とメンバーの対応時間はその時々で変動する。つまり、これらは確率変数である。訪問時間間隔、サービス時間の平均はそれぞれ 5 分、4 分であるにしても、実際はその周りの値に分布する。

より現実的な状況を考えよう。訪問時間間隔、サービス時間はばらついており、表 1　乱数発生による訪問時間間隔、サービス時間のリストに示すようなものであると仮定する。

図 2 に顧客が系を訪問した時刻を矢印で示し、待っている時間を灰色で示し、サービスを受けている時間を白で示す。図 3 に顧客の系内にいる人数を示す。系内の人数は 0 から 3 の間で大きく変動する。しかし、平均系内人数は観測時間を長くしていくと 1 より大きなある一定値に近づいていく。待ち行列理論は、この長い時間領域の平均の待ち人数を扱う。これを定常状態での値という。つまり、我々は累積の系内人数を $Q(t)$ とすると、長い時間での平均はある一定値に近づくとしている。これは、以下のように表現される。

$$\lim_{t\to\infty} = \frac{Q(t)}{t} = const \tag{1}$$

待ち行列理論はこの定常状態における値を扱う。すなわち、その系を見ていては値がばらついているので、その値は分からないが、長い時間の平均としてみればその状況がわかるのである。

表 1　乱数発生による訪問時間間隔、サービス時間のリスト

No	1	2	3	4	5	6	7	8	9	10
Time duration (min)	7	5	4	0	6	5	7	8	2	1
Service time (min)	9	2	3	6	2	3	4	7	2	5

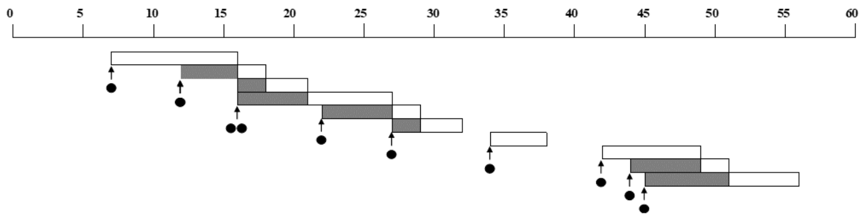

図 2　表 1 の訪問時間間隔、サービス時間を利用した場合の系内の状況

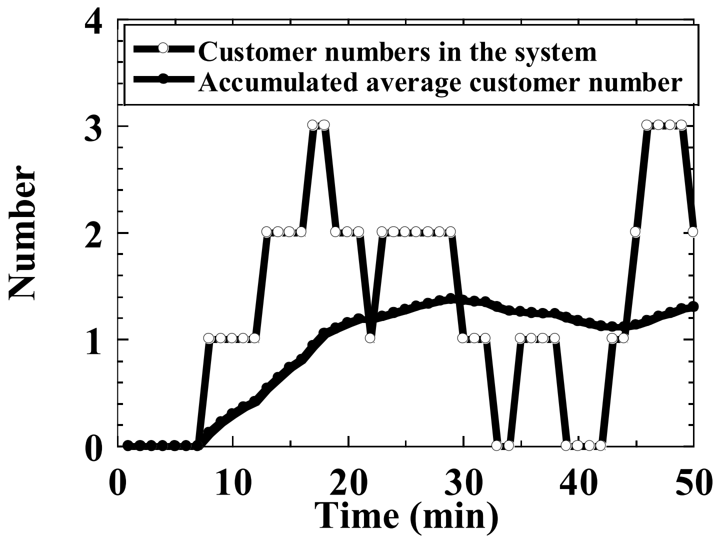

図 3　系内の人数と累積平均系内人数（図 2 の結果を表示）

1.3.　一般的な系の命名

ここでは、基本的に顧客は系を訪れ、サービスを受け、去っていく。その場合にも訪れる人や物、サービスの方法に様々な種類がある。ここでは、扱うシステムの種類の命名法を示す。

系を訪れるメンバーを顧客と呼ぶ。待ち行列理論はもともとコールセンターの解析のもとで発展した。したがって、この顧客のことをしばしば呼と呼ぶ。このように、訪問するのは人とは限らない。したがって、入力数はしばしば呼と呼ばれる。ここでは、入ってくるものを顧客、または呼と呼ぶ。

サービスをする場所をサービス機関と呼び、サービスをする人をサーバーまたはサービス提供者と呼ぶ。サービス提供者は銀行においては銀行業務のサービスをする人のことである。

1.4.　ケンドールの表記法

ケンドールの表記法は待ち行列の系を分類するのに一般的に利用される。 このでは、その表記方法を説明する。

例として銀行の待合室を考える。

顧客は銀行を訪れ、待ち、サービスを受け、去っていく。これらの要素を以下のように表記する。

1. 顧客の訪問する時間間隔に対する確率分布

我々は、顧客は系をランダムに訪問すると仮定する。これはポアソン分布で表現される。そして、対応する時間間隔は指数分布になる。指数分布は別名，マルコフ過程に対する分布である。したがって、これはMと表記される。 つまり、M は訪問時間間隔はマルコフ過程、つまり指数分布に従うことを意味する。もし、時間間隔がある分布に従うとすると、それに対応する記号で表現する。ここで、扱う分布は以下である

G: General distribution

M: Exponential distribution

D: Constant distribution

E_K: Erlang distribution with a phase K

2.サービス時間に対する確率分布

顧客が受けるサービスの時間に関する確率分布を考える。一般的には、これもMと表記する。つまり、サービス時間は指数分布に従うと仮定する。もし、サービス時間がある分布に従うとすると、それに対応する記号で表現する。

3.サービスをする人の人数。

これは、サービスをする人の人数であり、通常はsと表現される。

4. 系に入れる人の数

系にいることのできる人数の最大値を考慮する。一般にはこれは無限大とする。しかし、現実的には待合室の面積は有限であるから、入れる最大量は決まってくる。

5.サービスルール

サービスのルールはいろいろある。ここでは、一つのルールのみを扱う。つまり、入った順にサービスを受けるとする。したがって、これはケンドールの表記には表さない。最初に入った人がサービスを受けるとしている。

その外には、最後に入った人、ものからサービスを受ける場合、優先順位をつけて、その順に行う場合がある。そのような場合をここでは扱わない。

5. 訪問顧客の供給源

我々は、訪問顧客の供給は無限であることを暗に仮定していた。しかし、機械修理のような問題を考える際、その供給量、つまり故障する機械の量はもとの機械の量に制限を受ける。我々はこのような場合も解析する。

6. かたまりの供給

これまでは、顧客は短い間に 1 人、または 1 個だけ発生するとしてきた。しかし、トラックや船を利用した場合、供給量はかたまりとなる。これは、集団待ち行列と呼ばれる。このような場合もここでは扱う。

ケンドールの表記法では、最初の項は訪問、次の項はサービス、その次の項はサービス人数、最後の括弧の中身は系に存在できる人数の最大値を表す。

具体的には以下である。

M/M/s(∞)

最初の M は訪問間隔時間は指数分布に従うことを表している。二番めの M はサービス時間が指数分布に従うことを表している。三番目の s はサービスをする人の人数を表している。最後の無限大は、系にはいくらでも人が入れることを意味している。

もしも系に入ることのできる人数が有限であり、それが N であるとすると、その系は M/M/s(N)のように表現される。

入力の供給源の数が m に制限される場合は以下のように表現される。この場合は最初の M は訪問間隔時間は指数分布に従うことを表している。二番めの M はサービス時間が指数分布に従うことを表している。三番目の s はサービスをする人の人数を表している。M(m)/M/s(∞)

もしも供給がブロックで表現される場合は、関連する系は M[b]/M/1(∞)のように表現される。

この他にも様々な系が存在する。しかし、我々がここで扱うのはその中のいくつかである。ここで扱う形の全てを表 2 に示す。これは全てをカバーしてはいないが、多くの重要な場合を含んでいる。

最も一般的なシステムは以下である。

G/G/s(N)

これは解析的には扱わない。しかし、これは Monte Carlo では容易に扱うことができる。

M/G/1 は解析解を求めることができる。この結果を視察することで、以下の系に対する経験的な

モデルを提案することができる。

M/G/s(∞), M/G/s(N)

表 2　この本で扱う系のケンドールの記号

Kendall's notation
M/M/s(∞)
M/M/s(N)
M/M/s(s)
M/M/∞(∞)
M/M/∞(N)
M/G/1(∞)
$M/E_k/1(\infty)$
M(m)/M/s(∞)
M[b]/M/1(∞)

1.5.　理論で利用するパラメータ表示

パラメータの表記は任意である。しかし、いくつかのパラメータは頻繁に用いられる。したがって、この本では、あるパラメータはどのように表記されるか示しておくとわかり易い。ここでは、そのいくつかを示す。先に述べたように、表記には任意性があり、それは本に依存する。どう定義されているかを知ることが重要である。

1.5.1.　確率 P_n

この解析では現象を確率的に扱う。したがって、変数は確率的であり、確定したものではない。したがって、確率 P_n を導入する。P_n は系内にいるメンバーの数が n であることを表す。

1.5.2.　訪問客の容量 m

我々は、通常理論では、顧客の人数は無限と仮定している。したがって、顧客の訪問頻度は一定としている。しかし、それが仮定できない場合がある。機械修理の問題を考える。訪問顧客数はこの場合は、故障した機械となる。（この場合は訪問顧客は人ではなく機械となる。）もともとの機械の総数が m である場合、故障する機械の数は m 以下になる。このような特殊な場合もここでは扱う。その数を上に示しるように m とする。

1.5.3.　顧客の訪問頻度　λ

顧客の系への訪問頻度を λ .とする。これは、単位時間あたりの系への訪問人数である。これは

ポアソン分布に従うと仮定する。

1.5.4. サービス時間 t_s

顧客を系を訪れサービスを受ける。そのサービス時間を t_s とする。 t_s は確率変数である。つまりサービスごとに異なる。しかし、それが平均値を持つとする。その平均値を $\langle t_s \rangle$ とすると、それは以下で表現される。

$$\langle t_s \rangle = \frac{\sum t_s}{n} \tag{2}$$

1.5.5. 当たりのサービス頻度 μ

サービス利用率 μ は 1 サービス提供者当たりの単位時間あたりに対応できる数である。それは以下のように評価される。

$$\mu = \frac{1}{\langle t_s \rangle} \tag{3}$$

したがって, μ is は平均値である。別の本では、μ はトータルのメンバーによるサービス頻度として定義される場合がある。ここでは、1 サービス提供者あたりの利用率である。

1.5.6. サービス提供者数 s

顧客はサービスを受ける。サービスを提供する人数を s とする。

1.5.7. 1 サービス提供者あたりの利用率 ρ

1 サービス提供者当たりのサービス利用率は ρ で与えられる。 ρ は λ と μ と使い以下のように表現される。

$$\rho = \frac{\lambda}{\mu} \tag{4}$$

これは、各本により定義が異なる場合がある。このでは、以上の定義で常に利用する。

1.5.8. トータル利用率 η

η は全てのサービス提供者の利用率である。すなわち、以下のように与えられる。

$$\eta = \frac{\lambda}{s\mu} = \frac{\rho}{s} \tag{5}$$

これは、各本により定義が異なる場合がある。この本では、以上の定義で常に利用する。

1.5.9. 系にいる平均人数 L

系にいる顧客の平均人数を L と表記する。これは待っている人数とサービスを受けている人数

の和である。

系にいる顧客の人数は。図 3 に示すように 時間t によってばらついている。しかし、その人数は経過時間時間を長くとり、その平均値をとると、時間が経つにつれある一定値に近づく。つまり、

$$L = \lim_{t \to \infty} \frac{\int_0^t N(t)\,dt}{t} \tag{6}$$

である。ここで、$N(t)$ は時間tにおける系内の顧客の数である。したがって、L はP_n を用いて以下のようにも評価される。

$$L = \sum_{n=0}^{\infty} nP_n \tag{7}$$

1.5.10. 系内で待っている平均人数L_q

系内の待っている人数の平均値はL_qと表記される。

L_q も図 3 に示すように 時間t によってばらついている。しかし、その人数は経過時間を長くとり、その平均値をとると、時間が経つにつれある一定値に近づく。つまり、

$$L_q = \lim_{t \to \infty} \frac{\int_0^t N_q(t)\,dt}{t} \tag{8}$$

である。ここで、$N_q(t)$ は時間tにおける系内のまっている顧客の数である。したがって、L_q はP_n を用いて以下のようにも評価される。

$$L_q = \sum_{n=s+1}^{\infty} (n-s)P_n \tag{9}$$

系内のサービスを実際にうけている顧客の数s'は以下のようになる。

$$s' = L - L_q \tag{10}$$

1.5.11. 平均総滞在時間 W

顧客が系に滞在している平均総時間はWと表記される。これは、待っている時間とサービスを受けている時間の和である。

1.5.12. 平均待ち時間 W_q

平均待ち時間はW_qと表記される。

1.5.13. 系にいることができる制限値N or ∞

我々は通常は系にいることのできる最大人数に制限がないとする。この場合は∞と表記するか、表記を省く。実際の系では有限の広さの待合室であり、実質的にはその最大人数は有限である。それを明示する場合はNが用いられる。Nが扱う系内人数にくらべて十分大きい場合は、実質的にはNを∞としてよい。

1.6. 待ち行列の例

待ち行列は多くの例に適用できる。ここでは、その事例を表 3に示す。入ってくる顧客は、実際は顧客であったり、電話であったりする。したがって、解析する対象が何であるかもここでは示している。

表 3 待ち行列の例

Subject	Service	Queueing
Bank	Bank service	Waiting customers for the service
ATM	Money in and out	Waiting customers for finishing the treatment
Airplane	Landing or leaving the place	Waiting airplanes for departure or landing
Hospital	Medical check	Patents for the medical check
Call center	Telephone	Waiting customers for the telephone response
Barber	Hair cut	Waiting customers for the hair cut
Restaurant	Cooking	Waiting customers for the dish
Supermarket	Resister	Waiting customers for the resist
Supermarket	Selecting material to be bought	Customers who visit the supermarket
Harbor	Carrying with trucks	Block laggage
Museum	Display	Customers who visit the museum
Art gallerly	Display	Customers who visit the art gallaley
Library	Book lent	Customers who want to boroow a book
Rental video	Video lent	Customers who want to boroow a video
Machine repair	Machin repair person	Troubled machines for repair
Corona infection	Medical treatment	Infected customers

1.7. まとめ

この章のまとめを行う。

解析する対象の全ての領域を系と呼ぶ。

系にはいってくるものを顧客、または呼と呼ぶ。

サービスをする人のことをサービスをする人、またはサービス提供者と呼ぶ。

待ち行列理論はケンドールの記号で種別を表現される。 ケンドールの記号では 4 つの要素で系を表現する。訪問/サービス/サービス人数(系に入れる最大値)で表現される。ここで扱う系はケンドール で以下となる。

M/M/s(∞)

M/M/s(N)

M/M/s(s)

M/M/∞(∞)

M/M/∞(N)

M/G/1(∞)

M/E_k/1(∞)

M(m)/M/s(∞)

M[b]/M/1(∞)

ここで、最初の訪問を表す表記で(m)がついているものは、訪問のソースがmに制限を受けているものを表す。最初の訪問を表す表記で[b]がついているものは、訪問のソースがブロックであるものを表す。

もっとも一般的な系は以下である。

G/G/s(∞) and G/G/s(N)

これは解析的に扱うことはできない。しかし、これらは Monte Carlo simulation で表現できる。ここでは、以下を扱う。

M/G/s(∞) and M/G/s(N)

これらは、M/G/1 の解析モデルの結果を利用して経験的なモデルを導出する。

よく使われるパラメータの表記は以下である。

λ: 訪問顧客頻度

t_s: サービス時間

μ: サービス頻度

s: サービス提供者 (サービスをする人) の人数

ρ: λ のμに対する比率、これは 1 サービス提供者あたりのサービス率に相当する。

η: λ の$s\mu$に対する比率、これはトータルの サービス提供者のサービス率に相当する。

L: 系の平均顧客人数

L_q: 系の平均待ち人数

W: 系の平均総滞在時間

W_q: 平均待ち時間

系内にいることのできる最大人数は通常は∞とし、表記を省略す場合がある。これが、有限であればその数をNとする。

また、待ち行列に対応する事例も示した。