Ver. 1.2 平成 16 年 4 月 8 日

確率離散事象論講義資料

滝根 哲哉*

目 次

はじめに 2

1 待ち行列モデル 2

2 待ち行列理論のための基礎知識 3
2.1 用語と記号 . . . 3
2.2 リトルの公式 . . . 5
2.3 ポワソン過程と指数分布 . . . 8
2.3.1 ランダムな到着とポワソン過程 . . . 8
2.3.2 ポワソン到着における到着間隔と指数分布 . . . 9
2.3.3 一定到着率の仮定 . . . 10
2.3.4 一定到着率の仮定とポワソン過程 . . . 11
2.3.5 ポワソン過程の重畳と分岐 . . . 12
2.3.6 ポワソン到着する客が見るシステムの状態 (Cooper 1981) . . . 13

3 出生死滅過程と待ち行列 14
3.1 出生死滅過程 . . . 14
3.2 待ち行列モデルへの応用 . . . 16
3.2.1 $M/M/1$ . . . 16
3.2.2 $M/M/c$ . . . 18
3.2.3 $M/M/\infty$ . . . 20
3.2.4 $M/M/1/K$ . . . 20
3.2.5 $M/M/c/c$ . . . 21

4 離散時間マルコフ連鎖とその応用 22
4.1 離散時間マルコフ連鎖とその性質 . . . 23
4.1.1 離散時間マルコフ連鎖の遷移確率 . . . 23
4.1.2 再帰時間と状態の分類 . . . 24
4.1.3 既約な離散時間マルコフ連鎖の極限確率と定常状態分布 . . . 25
4.2 待ち行列モデルの系内客数分布 . . . 27
4.2.1 $M/G/1$ の系内客数分布 . . . 27
4.2.2 $M/G/1/K$ の系内客数分布 . . . 33
4.2.3 $GI/M/1$ の客数分布 . . . 36

5 待ち時間分布 40
5.1 指数サービスをもつ FCFS 待ち行列の待ち時間分布 . . . 40
5.1.1 $M/M/1$ の待ち時間分布 . . . 41
5.1.2 $GI/M/1$ の待ち時間分布 . . . 42
5.1.3 $M/M/c$ の待ち時間分布 . . . 43
5.1.4 $M/M/1/K$ の待ち時間分布 . . . 44
5.2 $M/G/1$ の待ち時間分布 . . . 44

6 その他の話題 45
6.1 残余寿命分布 . . . 45
6.2 $M/G/1$ の平均値公式 . . . 47
6.3 複数のポワソン流を収容する $M/G/1$ と非割込み優先規律 . . . 48
6.4 プロセッサシェアリング待ち行列と公平性 . . . 50
6.5 多呼種 $M/G/c/c$ . . . 51
6.6 待ち行列網と積形式解 . . . 52

*連絡先：
京都大学大学院情報学研究科数理工学専攻（〒 606-8501 京都市左京区吉田本町）
電話：(075)753-4758 FAX：(075)753-4756 電子メール：takine@kuamp.kyoto-u.ac.jp

はじめに

本講義では，離散的な状態をもち，かつ，確率的な振る舞いをするモデルの代表例として**待ち行列モデル**（queueing model）を取り上げる．待ち行列モデルとは，有限の資源が複数の利用者によって共有されている状況を表現した確率モデルであり，これを扱う理論は**待ち行列理論**（queueing theory）と呼ばれている．

待ち行列モデルで表現される状況では，ある利用者によって利用したい資源が占有されている場合，他の利用者はその資源が解放されるまで待つか，あるいは，その資源の利用をあきらめなければならない．この「待つ」という状況，あるいは「競合により資源利用を拒否される」という状況を抽象的に表すため，通常，資源利用要求を客，共有資源をサーバ，客が共有資源を占有する時間をサービス時間と呼ぶ．通信ネットワークへの応用においては，客はパケット，サーバは回線，サービス時間はパケットの送信時間に対応する．

以下では，待ち行列理論を理解する上で必要な基本的な知識として，ポワソン分布と指数分布ならびにリトルの公式を紹介した後，出生死滅過程，隠れマルコフ連鎖について解説すると共に，様々な待ち行列モデルへの応用を示す．

1　待ち行列モデル

待ち行列理論は共有資源に対する利用要求が確率的に発生するという仮定の下で，資源競合問題を抽象化した数学モデルの構築とその解析に関する理論である．待ち行列理論において扱われる確率モデルは待ち行列モデルと呼ばれる．図 1 に示されているように，待ち行列モデルとは共有の資源であるサーバと待合室からなるシステムに外部から客が到着し，これらの客はシステム内で暫く滞在した後，システムを去るというものである．通信ネットワークにおける応用では客はパケットに対応し，サーバと待合室はそれぞれ回線とルータ内のバッファに対応する．

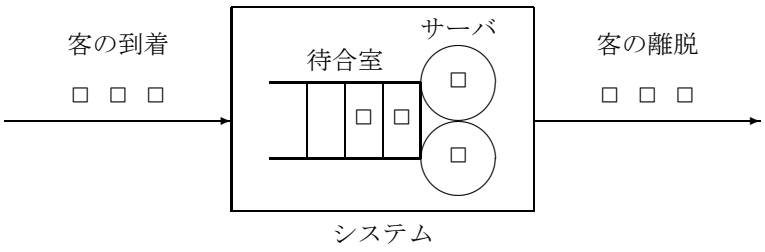

図 1: 待ち行列モデル

一般に待ち行列モデルは次の五つの要素からなる．

1. 到着過程（パケットの発生時点に関する統計的情報）
2. サービス時間分布（パケットの伝送時間に関する統計的情報）
3. サーバ数（回線の数）

4. システム容量（ルータ内で保持できるパケットの最大数）

5. サービス規律（ルータ内のパケットの送信順序を定める規則）

このような要素からなる待ち行列モデルを記述する方法として**ケンドールの記法**（Kendall's notation）が広く用いられている．これは，通常 $A/B/c/N$ の形をしており，それぞれ，到着間隔分布／サービス時間分布／サーバ数／システム容量を示している．A，B に関しては M（指数分布），D （一定分布），E_k（k 次のアーラン分布），H_k（k 次の超指数分布），G（一般分布：特定の分布を仮定しない），GI（独立同一分布：特定の分布は仮定しないが，到着間隔あるいはサービス時間が独立で同一の分布に従うことを強調したもの）などが用いられる．システム容量が無限大の場合は単に $A/B/c$ と書かれる．サービス規律は通常，先着順サービス（利用要求を到着順に処理するサービス規律）が仮定されており，**FCFS** (First-come, First-served) と書かれる．[1]サービス規律を明示する場合は，FCFS $A/B/c$ のように，この記号の前に書くことが多い．

待ち行列理論は通信ネットワークを含む様々な資源競合問題に対して抽象化されたモデルを通してシステムの定量的な評価を行い，問題解決の指針を与える数学的道具である．実際のシステムを評価するために待ち行列理論あるいは待ち行列モデルを利用する際には，まず，上記 (1)–(5) の要素を定める必要がある．ここで注意すべきことは，異なるシステムが同一のモデルで表現される場合や，同じシステムが（どの程度詳細にモデル化するかによって）異なるモデルで表現される場合があるということである．言い替えれば，待ち行列モデルは実際のシステムにおける資源競合を抽象的に表現した確率モデルであり，個々のモデルは特定のシステムのみを表現したものではない．以下では待ち行列理論の習慣に従い，個々の資源利用要求を客，個々の客が資源を占有する時間をサービス時間，個々の共有資源をサーバと呼ぶことにする．

2　待ち行列理論のための基礎知識

2.1　用語と記号

待ち行列モデルにおいて興味のある性能指標には，客の待ち時間やシステム内の客数（以下では系内客数と呼ぶ）などがある．客の到着あるいはサービス時間が確率的に定まる場合，これらの性能指標は確定的な値ではなく**確率変数**（random variable）となる．確率変数とは事象を数値で表現したものであり，例えば，時刻 t における系内客数を $L(t)$ としたとき，時刻 t に系内客数が j 人であるという事象は $\{L(t) = j\}$ で記述され，この事象が起こる確率を $\Pr(L(t) = j)$ と書く．

確率変数 X の定義域を $\mathcal{S}$ とする．すなわち $\Pr(X \in \mathcal{S}) = 1$ である．定義域 $\mathcal{S}$ が可算である場合 X は離散確率変数と呼ばれ，そうでない場合は連続確率変数と呼ばれる．

確率変数 X と Y に対して $\{X \le x\}$ という事象と $\{Y \le y\}$ という事象が同時に起こる確率 $\Pr(X \le x, Y \le y)$ を結合確率という．さらに事象 $\{X \le x\}$ が起こったという条件の下で事象 $\{Y \le y\}$ が起こる確率を条件付き確率といい $\Pr(Y \le y \mid X \le x)$ と書く．結合確率は条件付き確率を用いて以下のように表現できる．

$$\Pr(X \le x, Y \le y) = \Pr(X \le x) \Pr(Y \le y \mid X \le x)$$

もし，結合確率 $\Pr(X \le x, Y \le y)$ が $\Pr(X \le x) \Pr(Y \le y)$ に等しいならば，確率変数 X と Y は**独立**（independent）であるといわれ，$\Pr(Y \le y \mid X \le x) = \Pr(Y \le y)$ となる．

[1]FIFO (First-in, First-out) と呼ばれることもある．

離散確率変数 X の**平均**（mean）$E[X]$ は

$$E[X] = \sum_{x \in \mathcal{S}} x \Pr(X = x) \tag{1}$$

で与えられる．$E[X]$ は X の**期待値**（expectation）とも呼ばれる．また，実数 $\mathcal{R}$ を定義域にもつ連続確率変数 X は**分布関数**（distribution function）

$$F(x) = \Pr(X \leq x)$$

によって特徴付けられる．$F(x)$ は非減少関数で $F(\infty) = 1$ である．特に

$$\int_{-\infty}^{x} f(x)dx = F(x)$$

なる $f(x)$ が存在するとき，$f(x)$ は X の**密度関数**（density function）と呼ばれる．密度関数 $f(x)$ は $F(x)$ が微分可能である場合 $f(x) = dF(x)/dx$ である．また Δx を微小な正数としたとき，

$$f(x)\Delta x \approx \Pr(x < X \leq x + \Delta x)$$

という確率的意味を持つ．さらに定義から

$$\int_{-\infty}^{\infty} f(x)dx = F(\infty) = 1 \tag{2}$$

であり，これは確率の和が 1 であることと等価である．密度関数 $f(x)$ をもつ連続確率変数 X の期待値 $E[X]$ は

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

で与えられる．もし $\Pr(X < 0) = 0$ ならば，X の**補分布** (complementary distribution) $F^C(x) = \Pr(X > x) = 1 - F(x)$ を用いて

$$E[X] = \int_{0}^{\infty} F^C(x)dx$$

と書くことが出来る．

定数 c_1，c_2 と確率変数 X，Y に対して，期待値は次の性質を持つ．

$$E[c_1X + c_2Y] = c_1E[X] + c_2E[Y], \tag{3}$$

また，もし，確率変数 X と Y が独立ならば

$$E[XY] = E[X]E[Y]$$

が成立する．さらに，連続確率変数 X と関数 $u(x)$ に対して

$$E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx \tag{4}$$

である．離散確率変数 X の場合，式 (4) は，

$$E[u(X)] = \sum_{x \in \mathcal{S}} u(x) \Pr(X = x) \tag{5}$$

である．$\mathcal{S} \subset \mathcal{R}$ のとき，式 (4) と式 (5) をまとめて

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)dF(x)$$

という記法を用いる．[2]この記法を用いれば，X が連続か離散かという区別をせずに統一的に期待値を表すことができる．特に，$u(x) = x^n$ のとき $E[u(X)] = E[X^n]$ は $\boldsymbol{n}$ **次積率**（the nth moment）と呼ばれる．

[2]Riemann-Stieltjes 積分と呼ばれる．

最後に**分散** (variance) を定義する．確率変数 X の分散 $\mathrm{Var}(X)$ は

$$\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

で与えられる．分散は確率変数の取る値が平均からどの程度離れやすいかを示す指標であり，最初の等号から非負の値を取ることが分かる．また，$\sqrt{\mathrm{Var}[X]}$ は**標準偏差**（standard deviation）と呼ばれる．

2.2 リトルの公式

最初に述べたように，待ち行列モデルとは，システムの外部から客が到着し，システム内で一時的に滞在した後，システムを去る，という動作を表現したモデルである．よって，時刻 t におけるシステム内の客数，すなわち系内客数を $L(t)$ とし，$A(0,t]$ を区間 $(0,t]$ の間にシステムに到着した客数，$D(0,t]$ を区間 $(0,t]$ の間にシステムを離脱した客数とすると，

$$L(t) = L(0) + A(0,t] - D(0,t] \tag{6}$$

という関係を満たすモデルを対象としていることになる．最も興味ある量は平均系内客数 L と平均系内滞在時間 W である．W_n $(n = 1, 2, \ldots)$ を時刻 0 以降，n 番目に到着した客の系内滞在時間とすると，L と W はそれぞれ

$$L = \lim_{T\to\infty} \frac{1}{T} \int_0^T L(t)dt, \qquad W = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} W_n$$

で与えられる．ここで L は十分に長い間，システムを観測したときの累積客数の平均であり，時間平均と呼ばれる．一方，W は滞在時間の総和を客数で割った平均であり，客平均と呼ばれる．

一般に式 (6) を満たすモデルにおける平均系内客数 L と平均系内滞在時間 W の間には**リトルの公式**（Little's formula）と呼ばれる非常に単純な関係が成立する．まず，A_n を n 番目の客の到着時刻，$D_n = A_n + W_n$ を n 番目の客の離脱時刻とする．ここで指示関数 $I_n(t)$ を次式で定義する．

$$I_n(t) = \begin{cases} 1, & n \text{ 番目の客が時刻 } t \text{ に系内にいる場合} \\ & \quad (\text{すなわち } A_n \le t < D_n \text{ のとき}) \\ 0, & \text{その他} \end{cases}$$

定義より，W_n ならびに $L(t)$ はそれぞれ $I_n(t)$ を用いて

$$W_n = \int_0^\infty I_n(t)dt, \qquad L(t) = \sum_{n=1}^{\infty} I_n(t)$$

で与えられる．図 2 に W_n と $I_n(t)$ の関係を示す．

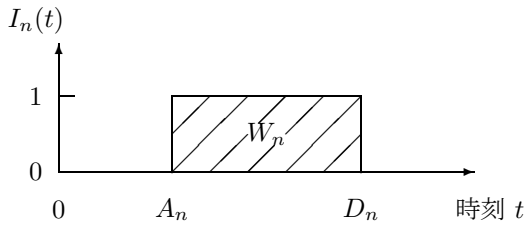

図 2: W_n と $I_n(t)$ の関係

ここで $N(t) = \max\{n;\, A_n \le t\} = A(0,t]$ を時刻 t までに到着した客の総数とすると，単位時間当たりに到着する平均客数，すなわち，平均到着率 λ は

$$\lambda = \lim_{t\to\infty} \frac{N(t)}{t}$$

で与えられる．

定理 1 (リトルの公式) L，λ，W が全て有限である場合，これらの間にはサービス順序によらず次式が成立する．

$$L = \lambda W \tag{7}$$

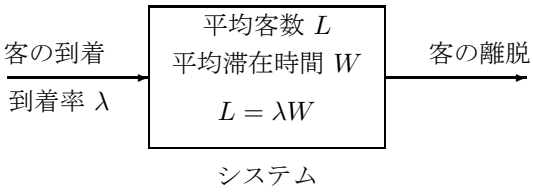

図 3: リトルの公式

リトルの公式がなぜ成立するかを見るために，$L(T) = 0$，すなわちシステムが空であるような時刻 T を考える．図 4 は時刻 T までに到着した 5 人の客が全て時刻 T までに離脱し，時刻 T における系内客数が 0 である場合を示している．

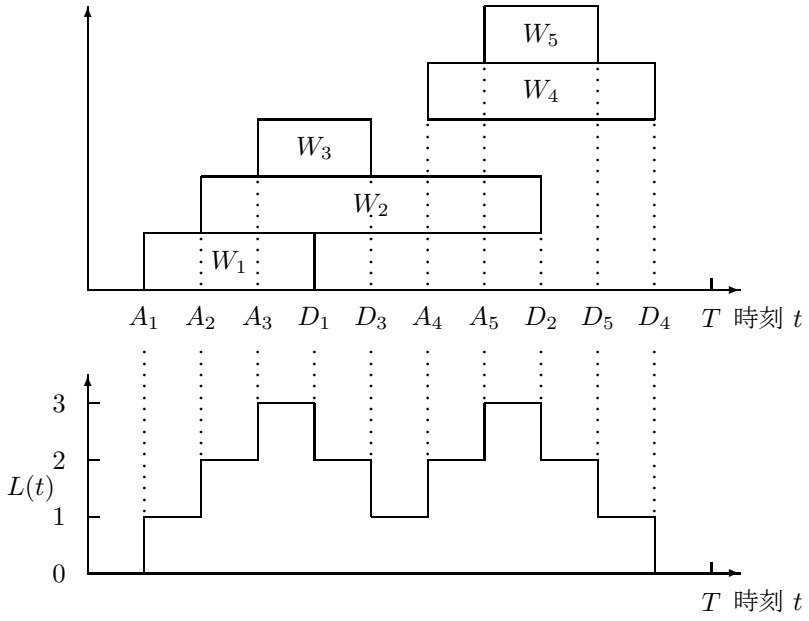

図 4: 客数の時間累積和と系内滞在時間の総和

時刻 T までに到着した全ての客はシステムを離脱しているので

$$\int_0^\infty I_n(t)dt = \int_0^T I_n(t)dt, \qquad n = 1, \ldots, N(T)$$

が成立する．さらに時刻 T 以降に到着する客，すなわち，$n > N(T)$ なる n に対しては，時間区間 $[0,T]$ において $I_n(t) = 0$ である．よって

$$\int_0^T L(t)dt = \int_0^T \sum_{n=1}^{N(T)} I_n(t)dt = \sum_{n=1}^{N(T)} \int_0^T I_n(t)dt = \sum_{n=1}^{N(T)} W_n$$

を得る．すなわち客数の時間累積和は系内滞在時間の総和に等しい（図 4 参照）．よって，時間区間 $[0,T]$ における平均系内客数と平均系内滞在時間の間には

$$\frac{1}{T}\int_0^T L(t)dt = \frac{N(T)}{T} \cdot \frac{1}{N(T)} \sum_{n=1}^{N(T)} W_n$$

が成立する．ここで $N(T)/T$ は時間区間 $[0,T]$ における平均到着率である．

一方，任意に選ばれた時刻 t においては

$$\int_0^t L(\tau)d\tau = \sum_{n=1}^{N(t)} W_n + X_E$$

である．ここで X_E は誤差の項を表現している．時刻 t における系内客数 $L(t)$ が正ならば，ある n（$\leq N(t)$）に対して $\int_0^t I_n(\tau)d\tau < W_n$ となるので誤差の項 X_E は負となる．t で両辺を割ると

$$\frac{1}{t}\int_0^t L(\tau)d\tau = \frac{N(t)}{t}\frac{1}{N(t)} \sum_{n=1}^{N(t)} W_n + \frac{X_E}{t}$$

を得る．もし

$$\lim_{t\to\infty} \frac{X_E}{t} = 0$$

ならば，$t \to \infty$ とすることにより式 (7) を得る．ここで仮定した $\lim_{t\to\infty} X_E/t = 0$ は 十分時間が経った後にも誤差の項 X_E が高々有限の値に押えられていることと等価である．すなわち，いかなる時刻においても系内に滞在している客が離脱するまでの時間の和が有限であれば，この条件が成立する．実際の安定なシステムへの応用では，無条件にリトルの公式が成立すると考えても支障はない．

ここで考えた「システム」は待ち行列モデル全体を意味する必要はない．例えば，待合室のみをシステムと見なせば，平均待ち客数 L_q と平均待ち時間 W_q の間には $L_q = \lambda W_q$ が成立する．また，サーバ部分のみをシステムと見なすこともできる．これを 平均到着率 λ，平均サービス時間 b をもつ安定な $G/G/c$ を対象に考えてみる．$G/G/c$ が安定ならば到着した客は全てサービスされるので，サーバへは単位時間当たり平均 λ 人の客が到着する．また，サーバでの平均滞在時間は平均サービス時間 b に等しい．よって，リトルの公式より，サーバにいる，すなわちサービス中の平均客数は λb で与えられる．

特に $c = 1$，すなわち $G/G/1$ の場合，サービス中の客は高々 1 人なので，この結果をサーバが稼働している確率を用いて表現すると，

$$\lambda b = 0 \text{ 人} \times \Pr(\text{サーバが休止}) + 1 \text{ 人} \times \Pr(\text{サーバが稼働})$$

となる．$\rho = \lambda b$ とすれば ρ は $\Pr$(サーバが稼働) という確率を与える．このため ρ は**利用率**（utilization factor）とも呼ばれる．一般に単一サーバをもつシステムが安定であるための条件は $\rho < 1$ である．[3]

[3]一定間隔到着，一定時間サービスをもつ $D/D/1$ の場合の安定条件は $\rho \leq 1$ である．

2.3 ポワソン過程と指数分布

待ち行列理論で用いられる到着過程の内，最も基本的なものはランダムな到着である．ランダムな到着を数学的に表現するため次のような状況を考える (高橋 1995)．時間区間 $(0,T]$ に K 人の客がでたらめに到着すると仮定する．すなわち，それぞれの客は他の客とは独立に時間区間 $(0,T]$ 内で一様分布に従って到着時点を選ぶと仮定する．この結果，幅 x をもつ任意に選ばれた区間に到着がある確率は区間の幅だけに依存する．すなわち，ある客の到着時刻を τ としたとき， τ が時間区間 $(y,y+x]$ に含まれる確率は $\Pr(\tau \in (y,y+x] \subset (0,T]) = x/T$ で与えられ，区間の位置を表す y の値とは独立である．

$A(y,y+x]$ を時間区間 $(y,y+x] \subset (0,T]$ に到着する客数を表す確率変数とする．それぞれの客は互いに独立に到着時点を選ぶので，

$$\Pr(A(y,y+x]=k) = \frac{K!}{k!(K-k)!}\left(\frac{x}{T}\right)^k\left(1-\frac{x}{T}\right)^{K-k} \tag{8}$$

で与えられる．$\lambda = K/T$ とし，期待値の公式に従って計算すると

$$E[A(y,y+x]] = \lambda x$$

を得る．定義より λ は単位時間当たりに到着する平均客数，すなわち，平均到着率を表しており，時間区間 $(y,y+x]$ に到着する平均客数は時間区間の長さ x と平均到着率の積で与えられ，区間の位置とは独立である．

2.3.1 ランダムな到着とポワソン過程

ここで平均到着率 $\lambda = K/T$ を一定の値に保ちながら $T\to\infty$，$K\to\infty$ の極限を考える．まず，式 (8) は

$$\Pr(A(y,y+x]=k) = \frac{x^k}{k!}\left(1-\frac{x}{T}\right)^K \frac{K}{T-x}\cdot\frac{K-1}{T-x}\cdot\cdots\cdot\frac{K-k+1}{T-x}$$

と変形できる．ここで $x/T = \lambda x/K$ に注意すると

$$\lim_{T\to\infty}\left(1-\frac{x}{T}\right)^K = \lim_{K\to\infty}\left(1-\frac{\lambda x}{K}\right)^K = e^{-\lambda x}$$

となり，また

$$\lim_{T\to\infty}\frac{K-n}{T-x} = \lim_{T\to\infty}\frac{\lambda T-n}{T-x} = \lambda \qquad (n=0,\ldots,k-1)$$

なので，平均到着率 $\lambda = K/T$ を一定に保ちながら $T\to\infty$，$K\to\infty$ の極限を取ると次式を得る．

$$\Pr(A(y,y+x]=k) = e^{-\lambda x}\frac{(\lambda x)^k}{k!} \qquad (k=0,1,\ldots) \tag{9}$$

定義 1 (ポワソン分布) 式 *(9)* の右辺で与えられる確率分布を，平均 λx をもつ**ポワソン分布**（*Poisson distribution*）という．

次に，互いに重なり合わない二つの時間区間に到着する客数の結合確率を考える．K 人の客が時間区間 $(0,T]$ の間に互いに独立に一様分布に従って到着するとき，時間区間 $(0,T]$ に含まれる重なり合わない二つの時間区間 $(y_1,y_1+x_1]$，$(y_2,y_2+x_2]$ にそれぞれ k_1 人，k_2 人の客が到着する結合確率は

$$\begin{aligned}
&\Pr(A(y_1,y_1+x_1]=k_1, A(y_2,y_2+x_2]=k_2)\\
&\quad= \Pr(A(y_1,y_1+x_1]=k_1)\Pr(A(y_2,y_2+x_2]=k_2 \mid A(y_1,y_1+x_1]=k_1)\\
&\quad= \frac{K!}{k_1!k_2!(K-k_1-k_2)!}\left(\frac{x_1}{T}\right)^{k_1}\left(\frac{x_2}{T}\right)^{k_2}\left(1-\frac{x_1+x_2}{T}\right)^{K-k_1-k_2}
\end{aligned} \tag{10}$$

で与えられる．さらに式 (10) は

$$\begin{aligned}\Pr(A(y_1, y_1+x_1] = k_1, A(y_2, y_2+x_2] = k_2) \\ = \quad \frac{x_1^{k_1}}{k_1!} \cdot \frac{x_2^{k_2}}{k_2!} \left(1 - \frac{x_1+x_2}{T}\right)^K \\ \cdot \frac{K}{T-x_1-x_2} \cdot \frac{K-1}{T-x_1-x_2} \cdot \cdots \cdot \frac{K-k_1-k_2+1}{T-x_1-x_2}\end{aligned}$$

と書き換えられるので，平均到着率 $\lambda = K/T$ を一定に保ちながら $T \to \infty$，$K \to \infty$ の極限を考えると

$$\begin{aligned}\Pr(A(y_1, y_1+x_1] = k_1, A(y_2, y_2+x_2] = k_2) \\ = \quad e^{-\lambda(x_1+x_2)} \frac{(\lambda x_1)^{k_1}}{k_1!} \cdot \frac{(\lambda x_2)^{k_2}}{k_2!} \\ = \quad e^{-\lambda x_1} \frac{(\lambda x_1)^{k_1}}{k_1!} \cdot e^{-\lambda x_2} \frac{(\lambda x_2)^{k_2}}{k_2!} \qquad (11)\end{aligned}$$

を得る．式 (9) に注意すると式 (11) は

$$\begin{aligned}\Pr(A(y_1, y_1+x_1] = k_1, A(y_2, y_2+x_2] = k_2) \\ = \quad \Pr(A(y_1, y_1+x_1] = k_1) \Pr(A(y_2, y_2+x_2] = k_2)\end{aligned}$$

と書くことができる．すなわち上記の極限で与えられる客の到着過程において，互いに重なり合わない時間区間に到着する客数は独立な確率変数となる．

定義 2 (ポワソン過程) 区間 $(y, y+x]$ の間に到着する客数分布が平均 λx のポワソン分布に従い，かつ，重なり合わない区間に到着する客数は互いに独立であるような到着過程を率 λ をもつ**ポワソン過程**（*Poisson process*），あるいは率 λ をもつ**ポワソン到着**（*Poisson arrivals*）という．

2.3.2 ポワソン到着における到着間隔と指数分布

次に率 λ でポワソン到着する客の到着間隔 X について考察する．時刻 t_0 に客の到着があったという事象を $Z(t_0)$ で表し，事象 $Z(t_0)$ が起こったという条件の下で，その次の客の到着までの間隔 X が t より大きい確率を考える．ポワソン到着の独立性より $A(t_0, t_0+t]$ は t_0 以前の到着とは独立なので，

$$\Pr(X > t \mid Z(t_0)) = \Pr(A(t_0, t_0+t] = 0 \mid Z(t_0)) = \Pr(A(t_0, t_0+t] = 0) = e^{-\lambda t}$$

を得る．よって，到着間隔の分布関数 $\Pr(X \le t \mid Z(t_0))$ は次式で与えられる．

$$\Pr(X \le t \mid Z(t_0)) = 1 - e^{-\lambda t} \qquad (12)$$

定義 3 (指数分布) 分布関数が式 *(12)* の右辺で与えられる確率分布をパラメタ λ をもつ**指数分布**（*exponential distribution*）という．

式 (12) より，パラメタ λ をもつ指数分布の密度関数は $\lambda \exp(-\lambda x)$ で与えられ，n 次積率（$n = 1, 2, \ldots$）は $n!/\lambda^n$ で与えられることが分かる．

次に，t_0 に客の到着があったという条件の下で，その後到着する 2 人の客の到着間隔 X_1, X_2 の結合分布を考える．$X_1 = y$ で条件付けを行うと，ポワソン到着の独立性より

$$\Pr(X_1 \le x_1, X_2 > x_2 \mid Z(t_0))$$

$$
\begin{aligned}
&= \int_0^{x_1} \lambda e^{-\lambda y} \Pr(A(t_0+y, t_0+y+x_2] = 0 \mid Z(t_0), Z(t_0+y)) dy \\
&= \int_0^{x_1} \lambda e^{-\lambda y} \Pr(A(t_0+y, t_0+y+x_2] = 0) dy \\
&= \int_0^{x_1} \lambda e^{-\lambda y} e^{-\lambda x_2} dy \\
&= (1-e^{-\lambda x_1}) e^{-\lambda x_2}
\end{aligned}
$$

となる．ここで $\Pr(X_1 \le x_1, X_2 \le x_2) + \Pr(X_1 \le x_1, X_2 > x_2) = \Pr(X_1 \le x_1)$ に注意すると

$$
\begin{aligned}
\Pr(X_1 \le x_1, X_2 \le x_2) &= \Pr(X_1 \le x_1) - \Pr(X_1 \le x_1, X_2 > x_2) \\
&= (1-e^{-\lambda x_1})(1-e^{-\lambda x_2}) \\
&= \Pr(X_1 \le x_1) \Pr(X_2 \le x_2)
\end{aligned}
$$

を得る．これは率 λ でポワソン到着する客の連続する到着間隔は互いに独立であり，それぞれ同じパラメタ λ をもつ指数分布に従うことを示している．

さて，時刻 0 に客が到着したと仮定し，次の客の到着までの間隔を X で表す．このとき，$(0, t_0]$ の間，次の客が到着しなかったという条件の下で時刻 t_0+t までに次の客が到着する条件付き確率 $\Pr(X \le t_0+t \mid X > t_0)$ を考える．定義に従って計算を進めると

$$
\begin{aligned}
\Pr(X \le t_0+t \mid X > t_0) &= \frac{\Pr(t_0 < X \le t_0+t)}{\Pr(X > t_0)} \\
&= \frac{\Pr(X \le t_0+t) - \Pr(X \le t_0)}{\Pr(X > t_0)} \\
&= \frac{(1-e^{-\lambda(t_0+t)}) - (1-e^{-\lambda t_0})}{e^{-\lambda t_0}} \\
&= 1-e^{-\lambda t} \qquad (13)
\end{aligned}
$$

を得る．式 (13) は条件付き確率 $\Pr(X \le t_0+t \mid X > t_0)$ が t_0 とは独立であり，時刻 t_0 から次の到着までの間隔は元の到着間隔 X と同じ確率分布に従うことを示している．この性質は指数分布の**無記憶性**（memoryless property）と呼ばれ，後で見るように待ち行列モデルの解析において極めて重要な役割を果たす．

2.3.3 一定到着率の仮定

客の到着間隔が独立同一なパラメタ λ をもつ指数分布に従うとき，その到着過程は以下で定義される一定到着率の仮定を満たす．

定義 4 (一定到着率の仮定) 次の 3 つの仮定を満たす客の到着過程は一定到着率の仮定を満たすと呼ばれる．

1. 独立増分：客の到着は互いに独立である．すなわち，交わらない二つの時間区間の間に到着する客の数は互いに独立な確率変数となる．
2. 定常増分：微小な時間区間 $(t, t+\Delta t]$ の間に *1* 人の客が到着する確率は $\lambda\Delta t + o(\Delta t)$ で与えられ，t とは独立である．
3. 順序性：客は 1 人ずつ到着する．すなわち，微小な時間区間 $(t, t+\Delta t]$ の間に *2* 人以上の客が到着する確率は $o(\Delta t)$ である．

ここで $o(\Delta t)$ は $\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$ となる項，すなわち Δt の高次の項を表す．仮定の (2)，(3) ならびに確率の和が 1 であることから，

4. 微小な時間区間 $(t, t+\Delta t]$ の間に客が到着しない確率は $1 - \lambda\Delta t + o(\Delta t)$ で与えられる．

が成立することに注意する．以下では客の到着間隔が独立同一な指数分布に従うとき客の到着過程は一定到着率の仮定を満たすことを示す．

まず初めに仮定 (1) を考える．２つの時間区間 $(y_1, y_1 + x_1]$ と $(y_2, y_2 + x_2]$ が重なり合わない（$y_1 + x_1 \leq y_2$）ならば，$\Pr(A(y_2, y_2 + x_2] = k_2 \mid A(y_1, y_1 + x_1] = k_1)$ は指数分布の無記憶性より $\Pr(A(y_2, y_2 + x_2] = k_2)$ と等しい．すなわち

$$\begin{aligned}
&\Pr(A(y_1, y_1 + x_1] = k_1, A(y_2, y_2 + x_2] = k_2) \\
&\quad = \Pr(A(y_1, y_1 + x_1] = k_1) \Pr(A(y_2, y_2 + x_2] = k_2 \mid A(y_1, y_1 + x_1] = k_1) \\
&\quad = \Pr(A(y_1, y_1 + x_1] = k_1) \Pr(A(y_2, y_2 + x_2] = k_2)
\end{aligned}$$

となり，仮定 (1) を満たすことが分かる．

次に時刻 t までに客が到着していないという条件の下で区間 $(t, t+\Delta t]$ の間に 1 人の客が到着する確率 $\Pr(X \leq t + \Delta t \mid X > t)$ を考える．指数分布の無記憶性より

$$\begin{aligned}
\Pr(X \leq t + \Delta t \mid X > t) &= 1 - e^{-\lambda\Delta t} \\
&= 1 - \left[1 - \lambda\Delta t + \frac{(\lambda\Delta t)^2}{2} - \cdots\right] \\
&= \lambda\Delta t + o(\Delta t)
\end{aligned}$$

となり，仮定 (2) を満たす．

最後に時刻 t までに客が到着しないという条件の下で区間 $(t, t+\Delta t]$ の間に 2 人以上の客が到着する確率を考える．

$$\begin{aligned}
&\Pr(A(t, t + \Delta t] \geq 2 \mid X > t) \\
&\qquad = 1 - \Pr(A(t, t + \Delta t] = 0) - \Pr(A(t, t + \Delta t] = 1)
\end{aligned}$$

ここで

$$\Pr(A(t, t + \Delta t] = 0) = e^{-\lambda\Delta t} = 1 - \lambda\Delta t + o(\Delta t)$$

に注意すると

$$\Pr(A(t, t + \Delta t] \geq 2 \mid X > t) = 1 - (1 - \lambda\Delta t) - \lambda\Delta t + o(\Delta t) = o(\Delta t)$$

となり仮定 (3) を満たす．よって，独立同一な指数分布間隔で到着する客は一定到着率の仮定を満たすことが示された．

2.3.4　一定到着率の仮定とポワソン過程

これまでに，ポワソン過程に従い到着する客の到着間隔が独立同一な指数分布に従う確率変数列となること，さらに客の到着間隔が独立同一な指数分布に従うとき，一定到着率の仮定を満たすことを見てきた．最後に，客の到着が一定到着率の仮定を満たすとき，客の到着がポワソン過程に従うことを示す．これにより，客の到着がポワ

ソン過程に従うこと，到着間隔が独立同一な指数分布に従うこと，ならびに客の到着が一定到着率の仮定に従うことが等価であることが示される．

まず，仮定 (1) より，ポワソン到着における独立性は明らかに満たされる．次に一定到着率の仮定の下で区間 $(y, y+x]$ に到着する客数 $A(y, y+x]$ を考える．$\Delta t = x/n$ とし，$o(\Delta t)$ の項を無視すると

$$\begin{aligned}
\Pr(A(y, y+x] = k) &= \lim_{\Delta t \to 0} \frac{n!}{k!(n-k)!} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{n-k} \\
&= \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda x}{n}\right)^k \left(1 - \frac{\lambda x}{n}\right)^{n-k} \\
&= \lim_{n \to \infty} \left(1 - \frac{\lambda x}{n}\right)^n \frac{(\lambda x)^k}{k!} \left(1 - \frac{\lambda x}{n}\right)^{-k} \\
&\qquad \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdot \cdots \cdot \frac{n-k+1}{n} \\
&= e^{-\lambda x} \frac{(\lambda x)^k}{k!}
\end{aligned}$$

となり，$\Pr(A(y, y+x] = k)$ は式 (9) を満たす．

定理 2 (ポワソン過程，指数分布到着間隔，一定到着率の仮定の等価性) 客の到着がポワソン過程に従うこと，到着間隔が独立同一な指数分布に従うこと，ならびに客の到着が一定到着率の仮定に従うことは等価である．

2.3.5 ポワソン過程の重畳と分岐

率 λ_i $(i = 1, \ldots, N)$ をもつ N 個の独立なポワソン過程を重ね合わせた到着過程を考える（図 5 参照）．

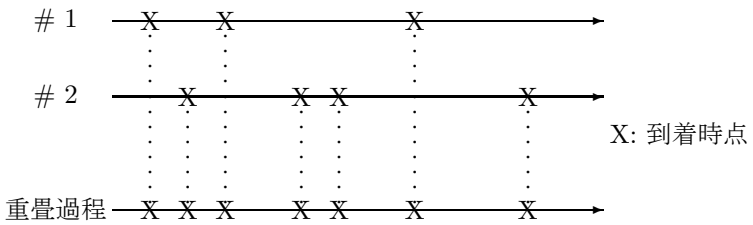

図 5: ポワソン過程の重畳（$N = 2$）

重ね合わせた到着流における到着間隔を X としたとき，時刻 t までに客が到着しない確率は，いずれの到着流からも客が到着しない確率に等しいため

$$\Pr(X > t) = \prod_{i=1}^{N} e^{-\lambda_i t} = \exp\left[-\left(\sum_{i=1}^{N} \lambda_i\right) t\right]$$

となり，到着間隔 X はパラメタ $\lambda = \lambda_1 + \cdots + \lambda_N$ の指数分布に従うことが分かる．さらに到着間隔 X が t であったという条件の下で，その到着が j 番目の到着流から到着した客である確率 d_j は

$$\begin{aligned}
d_j &= \lim_{\Delta t \to 0} \Pr(j \text{ 番目からの到着} \mid t < X \le t + \Delta t) \\
&= \lim_{\Delta t \to 0} \frac{\Pr(\text{時間区間 } (t, +\Delta t] \text{ に } j \text{ 番目からの到着})}{\Pr(t < X \le t + \Delta t)}
\end{aligned}$$

$$= \lim_{\Delta t \to 0} \frac{\lambda_j \Delta t + o(\Delta t)}{\lambda \Delta t + o(\Delta t)} = \frac{\lambda_j}{\lambda}$$

となり，到着間隔 X とは独立に平均到着率の割合で与えられる．

定理 3 (独立なポワソン過程の重畳) N 個の独立な率 λ_i $(i = 1, \ldots, N)$ をもつポワソン過程を重ね合わせた到着過程は率 $\lambda = \lambda_1 + \cdots + \lambda_N$ のポワソン過程となる．また，到着があったという条件の下で，その到着が j 番目の到着流からの客である確率は到着間隔とは独立に λ_j/λ で与えられる．

次に，率 λ でポワソン到着する客を，それぞれ独立に確率 p_i $(i = 1, \ldots, N)$ で i 番目の支流へ割り当てることを考える．以下では $N = 2$ の場合について結果を示すが，これは任意に選ばれた N について成立する．$A_i(0, t]$ を時間区間 $(0, t]$ の間に支流 i（$i = 1, 2$）に到着した客数とする．個々の客は独立に確率 p_i で支流 i に割り当てられるので $p_1 + p_2 = 1$ に注意すると，

$$\begin{aligned}
&\Pr(A_1(0,t] = n_1, A_2(0,t] = n_2) \\
&\quad = \Pr(A_1(0,t] = n_1 \mid A(0,t] = n_1 + n_2) \Pr(A(0,t] = n_1 + n_2) \\
&\quad = \frac{(n_1+n_2)!}{n_1! n_2!} p_1^{n_1} p_2^{n_2} \cdot e^{-\lambda t} \frac{(\lambda t)^{n_1+n_2}}{(n_1+n_2)!} \\
&\quad = e^{-\lambda p_1 t} \frac{(\lambda p_1 t)^{n_1}}{n_1!} \cdot e^{-\lambda p_2 t} \frac{(\lambda p_2 t)^{n_2}}{n_2!} \\
&\quad = \Pr(A_1(0,t] = n_1) \Pr(A_2(0,t] = n_2)
\end{aligned}$$

を得る．

定理 4 (ポワソン過程の分岐) 率 λ でポワソン到着する客をそれぞれ独立に確率 p_i $(i = 1, \ldots, N)$ で i 番目の支流へ割り当てたとき，i 番目の支流は他の支流とは独立な率 $p_i\lambda$ のポワソン過程となる．

2.3.6 ポワソン到着する客が見るシステムの状態 (Cooper 1981)

最後に，率 λ でポワソン到着する客の見るシステムの状態を考える．$Q(t)$ を時刻 t における系内客数とする．また，$P(t)$ を時刻 t の直後に客が到着したという条件の下での，時刻 t における系内客数とする．すなわち $P(t)$ は到着した客が見るシステムの状態である．ここで $C(x, y]$ を時間区間 $(x, y]$ に客が到着する事象とすると

$$\Pr(P(t) = k) = \lim_{\Delta t \to 0} \Pr(Q(t) = k \mid C(t, t + \Delta t])$$

である．よって

$$\begin{aligned}
\Pr(P(t) = k) &= \lim_{\Delta t \to 0} \frac{\Pr(Q(t) = k, C(t, t + \Delta t])}{\Pr(C(t, t + \Delta t])} \\
&= \lim_{\Delta t \to 0} \frac{\Pr(C(t, t + \Delta t] \mid Q(t) = k) \Pr(Q(t) = k)}{\Pr(C(t, t + \Delta t])} \qquad (14)
\end{aligned}$$

を得る．ここまでは到着に関して特に何も仮定していないことに注意する．

時刻 t におけるシステムの状態 $Q(t)$ は時刻 t 以前の到着のみによって定まる．一方，客の到着はポワソン過程に従うため，時刻 t 以降の到着はそれ以前の到着とは独立である．よって，ポワソン到着の場合，システムの状態とは独立に到着がおこるため $\Pr(C(t, t + \Delta t] \mid Q(t) = k) = \Pr(C(t, t + \Delta t]) = \lambda \Delta t + o(\Delta t)$ である．これらを式 (14) に代入すると次式を得る．

$$\Pr(P(t) = k) = \Pr(Q(t) = k)$$

定理 5 (ポワソン到着する客が見るシステムの状態) ポワソン到着する客が見るシステムの状態 $P(t)$ は外部観察者が見るシステムの状態 $Q(t)$ に等しい．特に，システムが**定常状態**（*steady state*）にある，すなわち $\Pr(Q(t)=k)$ が時刻 t に依存しない場合，ポワソン到着する客は定常状態を見る．

この結果は，ポワソン到着以外の場合は必ずしも成り立たないことに注意する．例として図 6 に到着間隔 2 秒，サービス時間 1 秒の $D/D/1$ （一定の到着間隔と一定のサービス時間を持つ単一サーバ待ち行列）に系内客数の変化を示す．最初の客がシステムに到着する時刻を 0 とすると，$(2t, 2t+1]$ （$t=0,1,2,\ldots$）の間はサービス中であり，$(2t+1, 2t+2]$ （$t=0,1,2,\ldots$）の間はシステムは空である．よって，図に示されているように外部観察者は全時間の $1/2$ の間，稼働中のシステムと見るが，到着する客は常に空のシステムを見る．このように，一般には，到着する客の見るシステムの状態は外部観察者が見るシステムの状態と異なる．

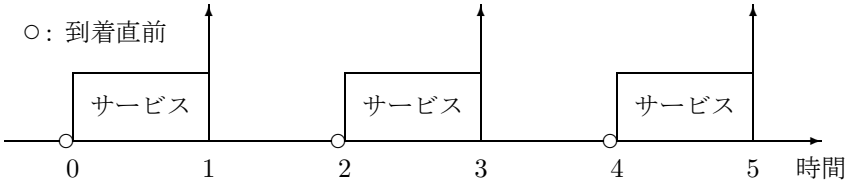

図 6: $D/D/1$ における客数の変化

3 出生死滅過程と待ち行列

この節では出生死滅過程と呼ばれる確率過程の定常状態確率分布の計算法ならびに待ち行列モデルへの応用を紹介する．

3.1 出生死滅過程

最初に出生死滅過程の定義を与える．

定義 5 (出生死滅過程) 時刻 t における系内客数 $L(t)$ の挙動が以下の仮定を満たすとき，$L(t)$ は**出生死滅過程** *(birth and death process)* と呼ばれる．

$$\Pr(L(t+\Delta t)=j \mid L(t)=i) = \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = i+1 \geq 1 \\ \mu_i \Delta t + o(\Delta t), & j = i-1 \geq 0 \\ o(\Delta t), & |j-i| \geq 2 \end{cases} \tag{15}$$

定義より，出生死滅過程 $L(t)$ は非負の整数値を取り，十分に小さな時間区間においてはその値は高々 1 しか増減しない．また，式 (15) ならびに確率の和が 1 であることから

$$\Pr(L(t+\Delta t)=i \mid L(t)=i) = \begin{cases} 1-\lambda_0 \Delta t + o(\Delta t), & i = 0 \\ 1-(\lambda_i+\mu_i)\Delta t + o(\Delta t), & i \geq 1 \end{cases}$$

が成立する．図 7 に出生死滅過程の状態遷移速度図を示す.

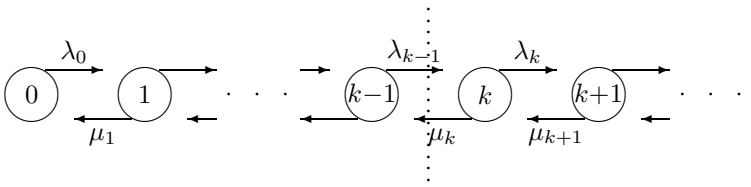

図 7: 出生死滅過程の状態遷移速度図

$r_i(t)$ を時刻 0 において状態 i にいるという条件の下で，時間間隔 $(0,t]$ の間状態 i から一度も他の状態に遷移しない確率とする．すなわち $r_i(t) = \Pr(L(\tau) = i, 0 \le \tau \le t \mid L(0) = i)$ である．このとき

$$\begin{aligned} r_i(t+\Delta t) &= r_i(t) \Pr(L(t+x) = i, 0 \le x \le \Delta t \mid L(\tau) = i, 0 \le \tau \le t) \\ &= r_i(t) \Pr(L(t+x) = i, 0 \le x \le \Delta t \mid L(t) = i) \end{aligned}$$

となるので，$i \ge 1$ の場合,

$$r_i(t+\Delta t) = r_i(t)(1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t))$$

を得る．さらに右辺の $r_i(t)$ を左辺に移項し，両辺を Δt で割り，$\Delta t \to 0$ の極限を考えると，$r_i(t)$ は以下の微分方程式を満たすことがわかる.

$$\frac{d}{dt} r_i(t) = -(\lambda_i + \mu_i) r_i(t)$$

$r_i(0) = 1$ に注意し，この微分方程式を解くと $r_i(t) = \exp(-(\lambda_i + \mu_i)t)$ を得る．$i = 0$ の場合は，同様の計算により $r_0(t) = \exp(-\lambda_0 t)$ となる．この結果より，出生死滅過程 $L(t)$ が各状態に留まる時間間隔は状態に依存したパラメタをもつ指数分布に従うことをが分かる.

時刻 t において出生死滅過程 $L(t)$ が k である確率を $p_k(t) = \Pr(L(t) = k)$ とする．時間区間 $(t, t+\Delta t]$ の間に起こる事象を考えると $p_k(t+\Delta t)$ は

$$\begin{aligned} p_0(t+\Delta t) &= p_0(t)(1 - \lambda_0 \Delta t) + p_1(t)\mu_1 \Delta t + o(\Delta t) \\ p_k(t+\Delta t) &= p_{k-1}(t)\lambda_{k-1}\Delta t + p_k(t)(1 - \lambda_k \Delta t - \mu_k \Delta t) \\ &\qquad + p_{k+1}(t)\mu_{k+1}\Delta t + o(\Delta t), \qquad k = 1, 2, \ldots \end{aligned}$$

を満たすことがわかる．これらの式に対して $r_i(t)$ の導出と同様の計算を行うと $p_k(t)$ が満たす微分差分方程式が得られる.

$$\begin{aligned} \frac{d}{dt} p_0(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ \frac{d}{dt} p_k(t) &= \lambda_{k-1} p_{k-1}(t) - (\lambda_k + \mu_k) p_k(t) + \mu_{k+1} p_{k+1}(t), \qquad k = 1, 2, \ldots \end{aligned}$$

以下では，システムが**定常**（stationary）であると仮定する．すなわち，$p_k(t)$ は時間に依存しないとする．このとき $p_k(t)$ の時間に関する微分値は 0 になるので，$p_k(t) = p_k$ とすると

$$\begin{aligned} 0 &= -\lambda_0 p_0 + \mu_1 p_1 \\ 0 &= \lambda_{k-1} p_{k-1} - (\lambda_k + \mu_k) p_k + \mu_{k+1} p_{k+1}, \qquad k = 1, 2, \ldots \end{aligned}$$

となり、これらを変形すると

$$\lambda_0 p_0 = \mu_1 p_1 \tag{16}$$

$$(\lambda_k + \mu_k) p_k = \lambda_{k-1} p_{k-1} + \mu_{k+1} p_{k+1}, \qquad k = 1, 2, \ldots \tag{17}$$

を得る．式 (16)，(17) の左辺は注目する状態から出ていく確率フローの量，右辺は注目する状態へ入る確率フローの量となっていることに注意する．一般に，定常状態においては，ある状態から出ていく確率フローの量はその状態へ入る確率フローの量に等しい．

さて，式 (16) ならびに (17) における $k = 1, \ldots, n-1$ を，辺々，足しあわせ，改めて $n = k$ とおくと

$$\lambda_{k-1} p_{k-1} = \mu_k p_k, \qquad n = 1, 2, \ldots \tag{18}$$

を得る．式 (18) の右辺は状態集合 $\{0, 1, \ldots, n-1\}$ から出ていく確率フローの総和であり，左辺は状態集合 $\{0, 1, \ldots, n-1\}$ へ入る確率フローの総和である．一般に，定常状態においては，ある状態集合から出ていく確率フローの総和はその状態へ入る確率フローの総和に等しい．特に，出生死滅過程の場合，状態遷移が隣り合う状態間でしか起こらないため，式 (18) の左辺 $\lambda_{k-1} p_{k-1}$ は客数が $k-1$ から k になる確率フローの量となり，右辺 $\mu_k p_k$ は客数が k から $k-1$ になる確率フローの量となる．すなわち，出生死滅過程の場合，$k-1$ と k の間を行きかう確率フローの量は定常状態では等しい（図 7 参照）．

さて，式 (18) より

$$p_k = \frac{\lambda_{k-1}}{\mu_k} p_{k-1} = \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k \mu_{k-1}\cdots\mu_1} p_0, \qquad k = 1, 2, \ldots \tag{19}$$

を得る．確率の総和は 1 なので，

$$\sum_{k=0}^{\infty} p_k = p_0 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k \mu_{k-1}\cdots\mu_1} p_0 = 1$$

を満たさなければならない．よって，未知の確率 p_0 は次式で与えられる．

$$p_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k \mu_{k-1}\cdots\mu_1} \right]^{-1} \tag{20}$$

定常状態が存在するための条件は式 (20) の右辺に現れる無限和が有限の値に収束することである．以上をまとめて次の定理を得る．

定理 6 (出生死滅過程の定常状態確率) 定義 *5* で与えられる出生死滅過程は式 *(20)* の右辺が有限の値に収束するとき定常状態確率をもち，それらは式 *(19)* ならびに式 *(20)* で与えられる．

3.2 待ち行列モデルへの応用

この節では，客の到着が率 λ のポワソン過程に従い，サービス時間がパラメタ μ の指数分布に従う様々な待ち行列モデルを考える．ただしサーバは系内に客がいる限り常にサービスを行うものとする．ポワソン到着をもつ待ち行列では，ポワソン到着の率 λ を到着率と呼ぶ．また，サービス時間が指数分布に従う場合，そのパラメタ μ をサービス率と呼ぶ．以下に見るように，このような待ち行列モデルにおける系内客数 $L(t)$ は出生死滅過程となるため，前節の結果を用いて定常状態における系内客数分布を得ることができる．なお，定常状態をもつ待ち行列モデルは**安定**（stable）であるといわれる．以下では

$$\rho = \lambda / \mu$$

とする．

3.2.1 $M/M/1$

まず初めに，単一サーバ待ち行列 $M/M/1$ を考える．

例 1 十分に大きなバッファをもつルータがあり，出力回線の容量は C bps であるとする．パケットの到着が率 λ のポワソン過程に従い，パケット長は平均 T バイトの指数分布に従うとする．このとき，ルータ内のパケット数の振舞いは，到着率 λ，サービス率 $\mu = C/8T$ をもつ $M/M/1$ でモデル化できる．

到着はポワソン過程に従うので，一定到着率の仮定より

$$\Pr(L(t+\Delta t) = 1 \mid L(t) = 0) = \lambda\Delta t + o(\Delta t)$$

である．また，サービス時間は指数分布に従うので $k = 1, 2, \ldots$ に対しても

$$\begin{aligned} \Pr(L(t+\Delta t) = k+1 \mid L(t) = k) &= (\lambda\Delta t + o(\Delta t))(1 - \mu\Delta t + o(\Delta t)) \\ &= \lambda\Delta t + o(\Delta t) \end{aligned}$$

となる．一方，客数が減少する場合は $k = 1, 2, \ldots$ に対して

$$\begin{aligned} \Pr(L(t+\Delta t) = k-1 \mid L(t) = k) &= (\mu\Delta t + o(\Delta t))(1 - \lambda\Delta t + o(\Delta t)) \\ &= \mu\Delta t + o(\Delta t) \end{aligned}$$

となる．$(t, t+\Delta t]$ の間に 2 人以上客が増加あるいは減少する確率は一定到着率の仮定より $o(\Delta t)$ である．以上の考察により $M/M/1$ の系内客数 $L(t)$ は $\lambda_k = \lambda$ （$k = 0, 1, \ldots$），$\mu_k = \mu$（$k = 1, 2, \ldots$）の出生死滅過程となることがわかる．図 8 に $M/M/1$ の状態遷移速度図を示す.

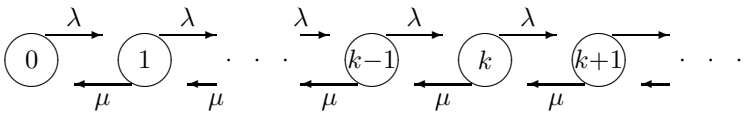

図 8: $M/M/1$ の状態遷移速度図

よって，定理 6 より，$\rho < 1$ のとき定常状態確率 p_k が存在し，

$$p_k = (1-\rho)\rho^k, \qquad k = 0, 1, \ldots \tag{21}$$

で与えられる．$E[L]$ を定常状態における平均系内客数とすると式 (21) より

$$E[L] = \sum_{k=1}^{\infty} k p_k = \frac{\rho}{1-\rho}$$

を得る．さらにリトルの公式を用いると平均系内滞在時間 $E[W]$ は

$$E[W] = E[L]/\lambda = \frac{1/\mu}{1-\rho}$$

となる．利用率 ρ に対する平均系内客数 $E[L]$ の変化を図 9 に示す．平均系内客数は利用率 ρ に対して非線形であり，ρ が 1 に近付くと急激に増加することに注意する．これは単一サーバ待ち行列に典型的に見られる性質である．

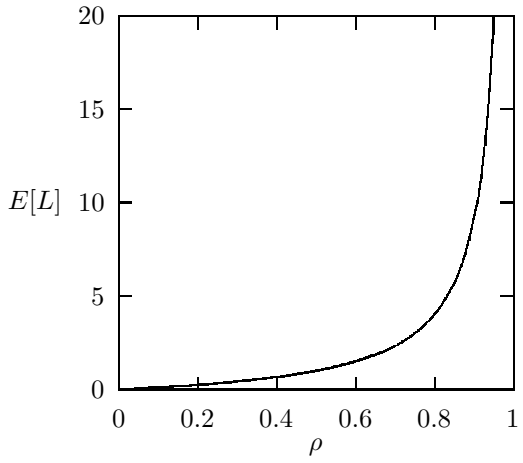

図 9: $M/M/1$ の平均系内客数 ($\mu = 1$)

3.2.2 $M/M/c$

c 個のサーバをもつ待ち行列 $M/M/c$ を考える．複数サーバ待ち行列は複数の CPU をもつコンピュータや通信ネットワークにおいて考察する送受信端末間に複数の経路が存在するような場合に適用できる．

$M/M/1$ の場合と同様に，ポワソン到着の仮定より

$$\Pr(L(t+\Delta t) = 1 \mid L(t) = 0) = \lambda \Delta t + o(\Delta t)$$

である．また，サービス時間は指数分布に従うので $k = 1, 2, \ldots$ に対しても

$$\begin{aligned} \Pr(L(t+\Delta t) = k+1 \mid L(t) = k) &= (\lambda \Delta t + o(\Delta t))(1 - \mu \Delta t + o(\Delta t))^{f(k)} \\ &= \lambda \Delta t + o(\Delta t) \end{aligned}$$

となる．ただし $f(k) = \min(k, c)$ はシステム内客数が k 人であるときのサービス中の人数を表す.

一方，客数が減少する場合は

$$\begin{aligned} &\Pr(L(t+\Delta t) = k-1 \mid L(t) = k) \\ &\quad = f(k)(\mu \Delta t + o(\Delta t))(1 - \mu \Delta t + o(\Delta t))^{f(k)-1}(1 - \lambda \Delta t + o(\Delta t)) \\ &\quad = f(k)\mu \Delta t + o(\Delta t), \quad k = 1, 2, \ldots \end{aligned}$$

を満たす．$(t, t+\Delta t]$ の間に 2 人以上客が増加あるいは減少する確率は一定到着率の仮定より $o(\Delta t)$ なので，$M/M/c$ の系内客数 $L(t)$ は $\lambda_k = \lambda$（$k = 0, 1, \ldots$），$\mu_k = f(k)\mu$（$k = 1, 2, \ldots$）の出生死滅過程となることがわかる．図 10 に $M/M/c$ の状態遷移速度図を示す.

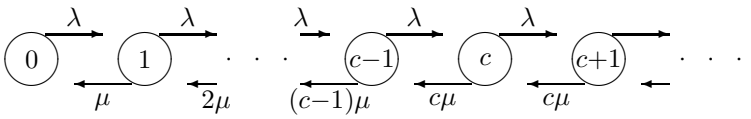

図 10: $M/M/c$ の状態遷移速度図

よって，定理 6 より，定常状態確率が存在するための条件は $\rho/c < 1$ であり，このとき定常状態における系内客数分布 p_k は次式で与えられる．

$$p_0 = \left[\sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{\rho^c}{(c-1)!(c-\rho)} \right]^{-1}$$

$$p_k = \begin{cases} \dfrac{\rho^k}{k!} p_0, & k = 0, \ldots, c-1 \\ \dfrac{\rho^k}{c! c^{k-c}} p_0, & k = c, c+1, \ldots \end{cases} \tag{22}$$

さらに平均系内客数 $E[L]$ は

$$E[L] = p_0 \left[\sum_{k=1}^{c-1} \frac{\rho^k}{(k-1)!} + \frac{\rho^c}{(c-1)!(c-\rho)^2} \{c^2 - (c-1)\rho\} \right]$$

となり，平均系内滞在時間 $E[W]$ はリトルの公式より $E[W] = E[L]/\lambda$ で与えられる．図 11 はシステムの能力を表すサーバ数 c で利用率 ρ を正規化したときの $M/M/c$ の平均系内客数 $E[L]$ を示したものである．c 個のサーバがある場合，系内客数が c 人以上であるときは全てのサーバが稼働するが，$c-1$ 人以下の場合，系内に客がいるにも拘らずサービスを行っていないサーバが存在する．結果として，システムの総能力を常時発揮することができないため性能が劣化する．言い替えれば c 個のサーバを用意するより，c 倍の能力をもつ一つのサーバを用意する方が能率的であることを示している．

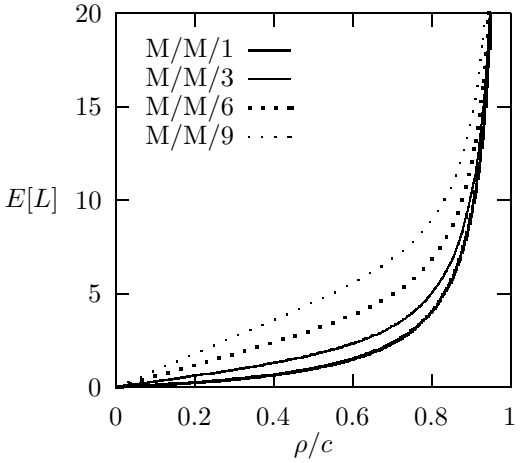

図 11: $M/M/c$ の平均系内客数

3.2.3 $M/M/\infty$

無限個のサーバをもつ待ち行列 $M/M/\infty$ を考える．サーバ数が無限であるので，到着した客は待たされることなく直ちにサービスを受けることができる．無限サーバ待ち行列はシステム内での客の振舞いが互いに独立と見なせる場合をモデル化したものであり，例えば，ある地域内でインターネットに接続中の人数をマクロな視点でモデル化する際に利用できる．

$M/M/c$ と同様の議論により $M/M/\infty$ の系内客数 $L(t)$ は $\lambda_k = \lambda$ $(k = 0, 1, \ldots)$，$\mu_k = k\mu$ $(k = 1, 2, \ldots)$ の出生死滅過程となる．図 12 に $M/M/\infty$ の状態遷移速度図を示す.

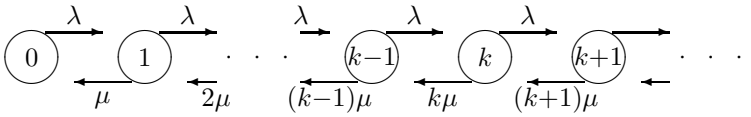

図 12: $M/M/\infty$ の状態遷移速度図

式 (20) の右辺は常に有限の値に収束するので，定理 6 より，$M/M/\infty$ は常に安定で定常状態が存在する．さらに定常状態確率 p_k は

$$p_k = e^{-\rho}\frac{\rho^k}{k!}, \qquad k = 0, 1, \ldots \tag{23}$$

で与えられ，平均 ρ をもつポワソン分布となる．

なお，式 (23) はサービス時間の平均が $1/\mu$ であるような全ての $M/G/\infty$ に対しても成立する．このように，サービス時間分布に関して，その平均値のみで系内客数の確率分布が定まる性質をサービス時間分布に関する**不感性**（insensitivity）という．

3.2.4 $M/M/1/K$

システム容量が K 人である単一サーバ待ち行列 $M/M/1/K$ を考える．すなわち，客が到着した時点で系内客数が K 人であれば，到着した客はシステムにはいることができず，棄却される．待ち行列理論ではこのように到着客が棄却されることを**呼損**（loss）[4]という．例 1 において，ルータのバッファが高々 K パケットしか保持できないと仮定すれば，そのようなルータの振舞いは $M/M/1/K$ でモデル化できる．また，この例のように客がパケットに対応している場合，呼損が起こる確率はパケット損確率あるいはパケット棄却率と呼ばれる．

この待ち行列の系内客数 $L(t)$ は $\lambda_k = \lambda$ $(k = 0, \ldots, K-1)$，$\lambda_k = 0$ $(k = K, K+1, \ldots)$，$\mu_k = \mu$ $(k = 1, 2 \ldots)$ の出生死滅過程となる．図 13 に $M/M/1/K$ の状態遷移速度図を示す.

[4]呼とは電話を接続する際の制御信号を指す call の訳語である．呼損という言葉を習慣的に用いているのは待ち行列理論が電話網の設計理論として発展してきた証でもある．

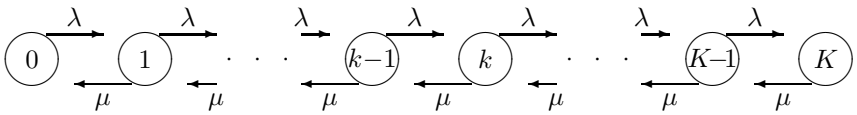

図 13: $M/M/1/K$ の状態遷移速度図

状態数が有限であるので，定理 6 より $M/M/1/K$ は常に安定であり，系内客数の定常状態確率 p_k は

$$p_k = \rho^k p_0, \qquad k = 0, \ldots, K$$

で与えられる．ただし p_0 は

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{K+1}}, & \rho \neq 1 \\ \dfrac{1}{K+1}, & \rho = 1 \end{cases}$$

である．

3.2.5 $M/M/c/c$

待合室を持たない c 個のサーバをもつ待ち行列 $M/M/c/c$ を考える．すなわち，客が到着した時点でサーバが全て稼働中であれば，到着した客はシステムにはいることが出来ず，呼損となる．

例 2 帯域が C *bps* の回線があり，各利用者は一定の帯域 r *bps* を要求すると仮定する．もし到着時に空き帯域が r *bps* 未満であるならば，この利用要求は拒否され，呼損となる．回線利用要求が率 λ のポワソン過程に従って発生し，回線接続時間が平均 μ^{-1} 秒の指数分布に従うならば，この回線の利用状況は，到着率 λ，サービス率 μ をもつ $M/M/c/c$ でモデル化できる．ただし $c = \lfloor C/r \rfloor$ は系内へ収容できる最大利用者数である．

$M/M/c/c$ の系内客数 $L(t)$ は $\lambda_k = \lambda$ $(k = 0, \ldots, c-1)$，$\lambda_k = 0$ $(k = c, c+1, \ldots)$，$\mu_k = k\mu$ $(k = 1, \ldots,$c$)$ の出生死滅過程となる．図 14 に $M/M/c/c$ の状態遷移速度図を示す．

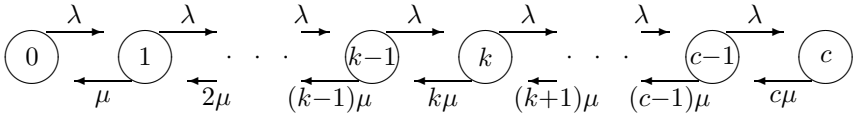

図 14: $M/M/c/c$ の状態遷移速度図

状態数が有限であるので，定理 6 より $M/M/c/c$ は常に安定であり，系内客数の定常状態確率 p_k は

$$p_k = \frac{\rho^k/k!}{\sum_{i=0}^{c} \rho^i/i!}, \qquad k = 0, \ldots, c$$

で与えられる．

この結果は $M/M/c/c$ の系内客数分布が $M/M/\infty$ において客数が c 人以下であるという条件の下での条件付き系内客数分布になっていることを示している．また $M/M/\infty$ と同様に，この結果もサービス時間分布に対して不感性をもっており，平均サービス時間が $1/\mu$ であるような全ての $M/G/c/c$ に対して成立する．特に**呼損率**（loss probability）を与える p_c に対する結果

$$p_c = \frac{\rho^c/c!}{\sum_{i=0}^{c} \rho^i/i!} \tag{24}$$

は **アーラン呼損式**（Erlang loss formula）と呼ばれている．利用率が ρ かつサーバ数が c のときの式 (24) で与えられる呼損率 p_c を $B(c,\rho)$ とすると $B(1,\rho) = \rho/(1+\rho)$ ならびに

$$B(c,\rho) = \frac{\rho B(c-1,\rho)}{c + \rho B(c-1,\rho)}, \qquad c = 2, 3, \ldots$$

により順次求めることができる．

図 15 は例 2 において一人当たりの要求帯域 r が 64kbps の場合の呼損率を示している．回線の帯域 C が 45Mbps の場合は $M/G/703/703$ に対応し，155Mbps の場合は $M/G/2421/2421$ に対応している．負荷の増大に伴い，呼損率は極めて急激に増加することが分かる．

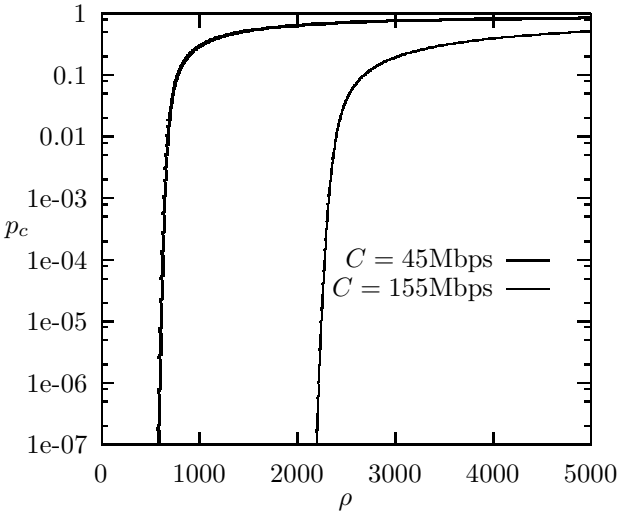

図 15: $M/G/c/c$ における呼損率 p_c

4 離散時間マルコフ連鎖とその応用

この節では離散時間マルコフ連鎖と呼ばれる確率過程と待ち行列モデルへの応用を考える．

4.1 離散時間マルコフ連鎖とその性質

X_n（$n = 0, 1, \ldots$）を対象となる確率変数列とする．これらの確率変数の取り得る値は非負の整数値とし，その集合を $\mathcal{S}$ で表す.

定義 6 (離散時間マルコフ連鎖) 確率変数列 X_n（$n = 0, 1, \ldots$）が

$$\Pr(X_{n+1} = j_{n+1} \mid X_k = j_k, k = 0, \ldots, n) = \Pr(X_{n+1} = j_{n+1} \mid X_n = j_n)$$

を満たすとき，確率変数列 X_n（$n = 0, 1, \ldots$）は**離散時間マルコフ連鎖**（*discrete-time Markov chain*）と呼ばれる．

離散時間マルコフ連鎖では将来の挙動 X_{n+1} が現在の状態 X_n のみで定まり，それ以前の状態 X_k $(k = 0, \ldots, n-1)$ とは独立である．この性質を **マルコフ性** (Markovian property) をいう．

4.1.1 離散時間マルコフ連鎖の遷移確率

離散時間マルコフ連鎖における $\Pr(X_{n+1} = j \mid X_n = i)$ は状態 i から状態 j への**遷移確率**（transition probability）と呼ばれる．以下では，遷移確率 $\Pr(X_{n+1} = j \mid X_n = i)$ が n に依存しないと仮定し，(i, j) 要素が

$$p_{i,j} = \Pr(X_{n+1} = j \mid X_n = i), \qquad i, j \in \mathcal{S}$$

で与えられる行列 $\boldsymbol{P} = (p_{i,j})$ を考える．$\boldsymbol{P}$ は**遷移確率行列**（transition probability matrix）と呼ばれる．行列 $\boldsymbol{P}$ の各行は，X_n の値が与えられたという条件の下での X_{n+1} の確率分布を表している．よって 行列 $\boldsymbol{P}$ は各要素が非負であり，かつ，各行の総和が 1 となる正方行列である．さらに (i, j) 要素が

$$p_{i,j}^{[n]} = \Pr(X_n = j \mid X_0 = i)$$

で与えられる正方行列 $\boldsymbol{P}^{[n]} = (p_{i,j}^{[n]})$ を n ステップ遷移確率行列と呼ぶ．特に $\boldsymbol{P}^{[1]} = \boldsymbol{P}$ であることに注意する．マルコフ性より，$n, m = 1, 2, \ldots$ に対して

$$\begin{aligned}
p_{i,j}^{[n+m]} &= \Pr(X_{n+m} = j \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} \Pr(X_{n+m} = j, X_n = k \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} \Pr(X_{n+m} = j \mid X_n = k, X_0 = i) \Pr(X_n = k \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} \Pr(X_{n+m} = j \mid X_n = k) \Pr(X_n = k \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} p_{i,k}^{[n]} p_{k,j}^{[m]}
\end{aligned}$$

となる．これを行列を用いて表現すると

$$\boldsymbol{P}^{[n+m]} = \boldsymbol{P}^{[n]} \boldsymbol{P}^{[m]}$$

である．特に $m = n = 1$ とすれば $\boldsymbol{P}^{[2]} = \boldsymbol{P}^2$ となるので，$m = 1$ として帰納法を用いると n ステップ遷移確率行列 $\boldsymbol{P}^{[n]}$ は

$$\boldsymbol{P}^{[n]} = \boldsymbol{P}^n$$

で与えられることが分かる．

ここで時点 n において状態が i である確率を $\pi_i^{[n]} = \Pr(X_n = i)$ $(i \in \mathcal{S})$ とし，X_n の確率分布 $\boldsymbol{\pi}^{[n]}$ を次式で定義する．

$$\boldsymbol{\pi}^{[n]} = (\pi_0^{[n]}, \pi_1^{[n]}, \ldots), \qquad n = 0, 1, \ldots$$

このとき

$$\begin{aligned} \Pr(X_n = j) &= \sum_{i=0}^{\infty} \Pr(X_n = j, X_0 = i) \\ &= \sum_{i=0}^{\infty} \Pr(X_n = j \mid X_0 = i) \Pr(X_0 = i) \end{aligned}$$

であるので $\boldsymbol{\pi}^{[n]} = \boldsymbol{\pi}^{[0]} \boldsymbol{P}^{[n]}$ となり，次の定理を得る．

定理 7 離散時間マルコフ連鎖の時点 n における確率分布 $\boldsymbol{\pi}^{[n]}$ は，初期状態確率分布 $\boldsymbol{\pi}^{[0]}$ ならびに遷移確率行列 $\boldsymbol{P}$ によって，以下のように一意に定まる．

$$\boldsymbol{\pi}^{[n]} = \boldsymbol{\pi}^{[0]} \boldsymbol{P}^n$$

4.1.2 再帰時間と状態の分類

ある時点で離散時間マルコフ連鎖が状態 i にあり，これ以降初めて状態 j に到達するまでの遷移回数 $T_{i,j}$ を状態 i から状態 j への**初到達時間**（first passage time）という．特に $i = j$ のとき，$T_{j,j}$ は状態 j の**再帰時間**（recurrence time）と呼ばれる．状態 i から状態 j への初到達時間が n である確率 $\Pr(T_{i,j} = n)$ を $f_{i,j}^{[n]}$ とする．すなわち

$$f_{i,j}^{[n]} = \Pr(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i), \quad n = 1, 2, \ldots$$

である．さらに，$f_{i,j}^{[n]}$ の総和を $f_{i,j}$ とする．

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{[n]}$$

$f_{i,j}$ は状態 i から出発し，いずれは状態 j に到達する確率である．もし $f_{j,j} = 1$ であるならば状態 j は**再帰的**（recurrent）と呼ばれ，$f_{j,j} < 1$ ならば状態 j は**過渡的**（transient）と呼ばれる．

まず，状態 i が再帰的であるか過渡的であるかを判定する条件を $p_{j,j}^{[n]}$ を用いて表すことを考える．ここで状態 i から出発するという条件の下で n ステップ目に状態 j にある確率を，最初に状態 j に訪れた時点を m として排反な事象に分けると

$$p_{i,j}^{[n]} = \sum_{m=1}^{n} f_{i,j}^{[m]} p_{j,j}^{[n-m]} \tag{25}$$

を得る．ただし $p_{j,j}^{[0]} = 1$ である．そこで，式 (25) において $i = j$ とし，両辺を n について和をとると

$$\begin{aligned} \sum_{n=1}^{\infty} p_{j,j}^{[n]} &= \sum_{n=1}^{\infty} \sum_{m=1}^{n} f_{j,j}^{[m]} p_{j,j}^{[n-m]} = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} f_{j,j}^{[m]} p_{j,j}^{[n-m]} = f_{j,j} \sum_{n=0}^{\infty} p_{j,j}^{[n]} \\ &= f_{j,j} + f_{j,j} \sum_{n=1}^{\infty} p_{j,j}^{[n]} \end{aligned}$$

となるので，形式的に

$$\sum_{n=1}^{\infty} p_{j,j}^{[n]} = \frac{f_{j,j}}{1 - f_{j,j}}$$

を得る．これより，状態 j は $\sum_{n=1}^{\infty} p_{j,j}^{[n]}$ が発散すれば再帰的であり，有限であれば過渡的であることがわかる．

次に再帰的な状態 j に対して平均再帰時間 $\nu_j = E[T_{j,j}]$ を考える．明らかに

$$\nu_j = \sum_{n=1}^{\infty} n f_{j,j}^{[n]}$$

である．もし，$f_{j,j} = 1$ かつ $\nu_j < \infty$ ならば状態 j は**正再帰的**（positive recurrent）と呼ばれ，$f_{j,j} = 1$ かつ $\nu_j = \infty$ ならば**零再帰的**（null recurrent）と呼ばれる．

次に**連結**（communicate）という考え方を導入する．まず，ある n （$n = 1, 2, \ldots$）に対して $p_{i,j}^{[n]} > 0$ ならば，状態 j は状態 i から到達可能という．さらに状態 i と j が互いに到達可能である時，状態 i と j は連結しているという．特に，状態 i はそれ自身と連結であるとする．このとき，マルコフ連鎖の全状態の集合 $\mathcal{S}$ の中から一つの状態を選び，それと連結している全ての状態を集めて集合 $\mathcal{S}$ の部分集合 $\mathcal{C}_1$ を作り（これを連結クラスという），さらに $\mathcal{C}_1$ に含まれていない状態を一つ選び，それと連結している全ての状態を集めて連結クラス $\mathcal{C}_2$ を作る，という手続きを繰り返すと，状態集合 $\mathcal{S}$ は

$$\mathcal{S} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots$$

のように互いに素な複数個の連結クラスの和集合で書くことができる．状態数が無限の場合は，連結クラスも無限個必要となる場合がある．

定義より，任意に選ばれた連結クラス $\mathcal{C}$ に含まれる二つの状態 i と j に対して，$p_{j,i}^{[n]} > 0$，$p_{i,j}^{[m]} > 0$ なる n，m が存在する．さらに

$$p_{j,j}^{[n+l+m]} \geq p_{j,i}^{[n]} p_{i,i}^{[l]} p_{i,j}^{[m]} \tag{26}$$

が成立するので，両辺を l について和をとれば

$$\sum_{l=1}^{\infty} p_{j,j}^{[l]} \geq \sum_{l=1}^{\infty} p_{j,j}^{[n+l+m]} \geq p_{j,i}^{[n]} \left(\sum_{l=1}^{\infty} p_{i,i}^{[l]} \right) p_{i,j}^{[m]}$$

を得る．これより次のことが分かる．もし，状態 i が再帰的ならば（すなわち $\sum_{l=1}^{\infty} p_{i,i}^{[l]} = \infty$），同じ連結クラスに属する他の状態 j もまた再帰的である（すなわち $\sum_{l=1}^{\infty} p_{j,j}^{[l]} = \infty$）．逆に状態 j が過渡的ならば（すなわち $\sum_{l=1}^{\infty} p_{j,j}^{[l]} < \infty$），同じ連結クラスに属する他の状態 i も過渡的である．

さらに，ある連結クラスに属する状態 i が正再帰的（零再帰的）であるならば，同じ連結クラスに属する他の状態 j もまた正再帰的（零再帰的）であることを示すことができる．特に再帰的な連結クラスに含まれる状態数が有限の場合，これらの状態は全て正再帰的である．

再帰的な連結クラス $\mathcal{C}$ の任意に選ばれた状態 i に対して，$p_{i,i}^{[l]} > 0$ となる l の最大公約数 d_i を考える．d_i は状態 i の周期と呼ばれる．同じ連結クラスに属する状態は同じ周期を持つことが知られており，$d = 1$ の時，連結クラスは**非周期**（aperiodic）であると呼ばれ，$d \geq 2$ のとき**周期的**（periodic）であると呼ばれる. 以上をまとめて次の定理を得る．

定理 8 連結クラス内の各状態は全て正再帰的か，全て零再帰的か，あるいは全て過渡的かのいずれかである．また，ある状態が周期的であるならば，その状態と同じ連結クラスに属する全ての状態は同じ周期をもつ．

4.1.3 既約な離散時間マルコフ連鎖の極限確率と定常状態分布

全ての状態 $i \in \mathcal{S}$ が一つの連結クラスに含まれるようなマルコフ連鎖は**既約**（irreducible）と呼ばれる．定義より，既約なマルコフ連鎖の任意に選ばれた状態 $i, j \in \mathcal{S}$ に対して，$p_{i,j}^{[n(i,j)]} > 0$ となるような自然数 $n(i,j)$ が存在する．

既約で正再帰的なマルコフ連鎖に対しては

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_i p_{i,j}, \qquad \sum_{j \in \mathcal{S}} \pi_j = 1 \tag{27}$$

を満たす正数 π_j（$j = 0, 1, \ldots$）が唯一定まる．$\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ としたとき，最初の式は $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ と等価であることに注意する．π_j を状態 j の**定常状態確率** (steady state probability) といい，$\boldsymbol{\pi}$ を**定常状態分布**（steady state distribution）という．定義より，もし $\boldsymbol{\pi}^{[0]} = \boldsymbol{\pi}$ ならば $\boldsymbol{\pi}^{[1]} = \boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$ であるので，帰納法により

$$\boldsymbol{\pi}^{[n]} = \boldsymbol{\pi}^{[n-1]}\boldsymbol{P} = \boldsymbol{\pi}, \qquad n = 1, 2, \ldots$$

を得る．すなわち初期状態分布 $\boldsymbol{\pi}^{[0]}$ が定常状態分布 $\boldsymbol{\pi}$ に等しければ，任意に選ばれた時点 n での状態分布 $\boldsymbol{\pi}^{[n]}$ は定常状態分布 $\boldsymbol{\pi}$ に等しい．

定常状態分布は時間平均分布と密接な関係がある．既約で正再帰的なマルコフ連鎖を時刻 0 から $N-1$ の間，観察したとき，状態 j にある時間平均を $g_j(N)$ とする．

$$g_j(N) = \frac{1}{N} \sum_{n=0}^{N-1} \pi_j^{[n]}$$

ここで $\boldsymbol{g}(N) = (g_0(N), g_1(N), \ldots)$ とすると

$$\boldsymbol{g}(N) = \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{\pi}^{[n]} = \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{\pi}^{[0]}\boldsymbol{P}^n$$

を得る．$\boldsymbol{I}$ を単位行列，$\boldsymbol{e}$ を全ての要素が 1 である列ベクトルとし，この両辺に $\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{e}\boldsymbol{\pi}$ を掛け，$\boldsymbol{P}\boldsymbol{e} = \boldsymbol{e}$，$\boldsymbol{\pi}^{[0]}\boldsymbol{e} = 1$ に注意すると

$$\boldsymbol{g}(N)[\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{e}\boldsymbol{\pi}] = \frac{1}{N}\boldsymbol{\pi}^{[0]}(\boldsymbol{I} - \boldsymbol{P}^N + N\boldsymbol{e}\boldsymbol{\pi}) = \boldsymbol{\pi}^{[0]}\frac{\boldsymbol{I} - \boldsymbol{P}^N}{N} + \boldsymbol{\pi}$$

となる．さらに $\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{e}\boldsymbol{\pi}$ が正則であり，$\boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{e}\boldsymbol{\pi})^{-1} = \boldsymbol{\pi}$ に注意すると

$$\boldsymbol{g}(N) = \boldsymbol{\pi}^{[0]}\frac{\boldsymbol{I} - \boldsymbol{P}^N}{N}(\boldsymbol{I} - \boldsymbol{P} + \boldsymbol{e}\boldsymbol{\pi})^{-1} + \boldsymbol{\pi}$$

を得る．ここで $N \to \infty$ の極限を考えると $\lim_{N\to\infty}(\boldsymbol{I} - \boldsymbol{P}^N)/N = \lim_{N\to\infty}(\boldsymbol{I} - \boldsymbol{P}^{[N]})/N = \boldsymbol{O}$ となるので，

$$\lim_{N\to\infty} \boldsymbol{g}(N) = \boldsymbol{\pi}$$

が得られる．

さらに，状態 j から出発し，平均 ν_j ステップで再び状態 j に戻って来るならば，平均して ν_j 回に一回，状態 j を訪れることになるので，時間平均分布 $\lim_{N\to\infty}\boldsymbol{g}(N)$，あるいは定常状態分布 $\boldsymbol{\pi}$ は平均再帰時間の逆数 $1/\nu_j$ に等しくなる．以上より，次の定理を得る．

定理 9 (マルコフ連鎖の定常状態確率と時間平均) 既約で正再帰的なマルコフ連鎖は式 *(27)* を満たす唯一の定常状態確率をもち，それは平均再帰時間の逆数に等しい．また，時間平均分布は定常状態分布と等しく，初期状態分布とは独立である．

一方，既約で非周期的なマルコフ連鎖に対して次式が成立する．

$$\lim_{n\to\infty} p_{i,j}^{[n]} = \lim_{n\to\infty} \pi_j^{[n]} = \pi_j^{[\infty]}$$

上式は，$n \to \infty$ の極限においてマルコフ連鎖がある状態 j にある確率が初期状態分布とは独立であることを示している．確率 $\pi_j^{[\infty]}$ は**極限確率** (limiting probability) と呼ばれる．極限確率に関しては次の定理が知られている．

定理 10 (マルコフ連鎖の極限確率) 既約なマルコフ連鎖が過渡的あるいは零再帰的である場合，全ての状態 $j \in \mathcal{S}$ に対して $\pi_j^{[\infty]} = 0$ である．また，正再帰的かつ非周期的である場合，極限確率 $\pi_j^{[\infty]}$ は定常状態確率 π_j に等しい．

4.2 待ち行列モデルの系内客数分布

以下で扱われる待ち行列モデルは，それぞれ，ある事象が起こった時点にのみ注目すると，系内客数に関する離散時間マルコフ連鎖を得ることができる．さらに得られたマルコフ連鎖の定常状態確率を用いて定常状態における系内客数分布を導出する手法について解説する．

4.2.1 $M/G/1$ の系内客数分布

到着率 λ，サービス時間分布関数 $H(x)$，平均サービス時間 b をもつ $M/G/1$ を考える．

例 3 十分に大きなバッファをもつルータがあり，出力回線の容量は C *bps* であるとする．パケットの到着が率 λ のポワソン過程に従い，パケット長（*byte*）の分布関数 は $G(x)$ であるとする．このとき，ルータ内のパケット数の振舞いは，到着率 λ，サービス時間分布関数 $H(x) = G(8x/C)$ をもつ $M/G/1$ でモデル化できる．

以下では利用率 $\rho = \lambda b$ が $\rho < 1$ であると仮定する．定常状態における系内客数分布を求めるため，客の離脱直後に注目する．n 番目の客の離脱直後の系内客数を X_n とし，A_n を n 番目の客のサービス時間の間に新たに到着する客数とする．

図 16 に系内客数の変化を示す．$X_n = 0$ ならば，この後，暫くの間サーバは休止しており，次の客が到着するとサービスを開始する．そして，この客のサービスが終了し離脱が起こった直後の系内客数は，この客のサービス中に新たに到着した客数に等しい．すなわち $X_n = 0$ ならば $X_{n+1} = A_{n+1}$ となる．

一方，$X_n = i$ $(i \geq 1)$ ならば，客の離脱の後，直ちに次のサービスが開始される．このサービスの終了直後における系内客数は，サービス開始時点で既に系内にいた $i-1$ 人の客とサービス中に新たに到着した客数の和で与えられる．すなわち $X_n \geq 1$ ならば $X_{n+1} = X_n - 1 + A_{n+1}$ である．よって

$$X_{n+1} = \max(X_n - 1, 0) + A_{n+1} \tag{28}$$

が成立する．

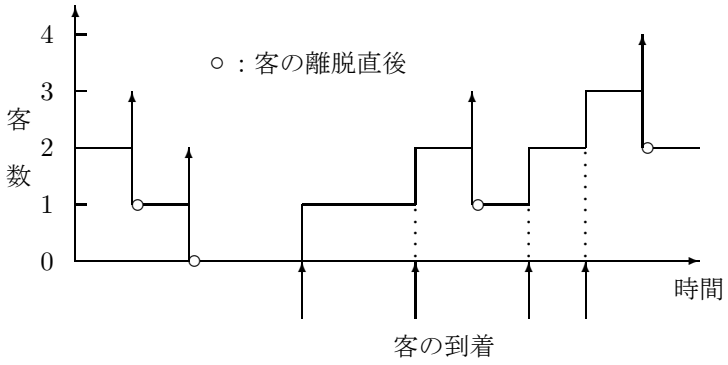

図 16: $M/G/1$ の系内客数の変化

さらに，一人の客のサービス時間の間に k 人の客が到着する確率を $a_k = \Pr(A_n = k)$ とすると，ポワソン到着の仮定より a_k は

$$a_k = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^k}{k!} dH(x), \qquad k = 0, 1, \ldots \tag{29}$$

で与えられ，n とは独立である．

以上の議論から，$n+1$ 番目の客の離脱直後の系内客数 X_{n+1} の確率分布は n 番目の客の離脱直後の系内客数 X_n のみに依存しており，それ以前の客の離脱時点での系内客数とは独立となる．すなわち，X_n はマルコフ性をもつことがわかる．このように系内客数が注目する時点でマルコフ性をもつとき，これらの時点を**隠れマルコフ点**（imbedded Markov point）といい，そのような時点のみに注目して構成されたマルコフ連鎖 X_n を**隠れマルコフ連鎖**（imbedded Markov chain）と言う．

定常状態における離脱直後の系内客数分布を求めるため，隠れマルコフ連鎖 X_n の遷移確率 $p_{i,j} = \Pr(X_{n+1} = j \mid X_n = i)$ を考える．式 (28) より $X_n = 0$ ならば $A_{n+1} = X_{n+1}$ なので

$$p_{0,j} = a_j, \qquad j = 0, 1, \ldots$$

を得る．また，$X_n = i \geq 1$ ならば $A_{n+1} = X_{n+1} - X_n + 1$ なので

$$p_{i,j} = \begin{cases} a_{j-i+1}, & j = i-1, i, \ldots \\ 0, & j = 0, \ldots, i-2 \end{cases}$$

を得る．以上より，客の離脱直後の系内客数の遷移確率行列 $\boldsymbol{P}$ は

$$\boldsymbol{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & \ldots \\ a_0 & a_1 & a_2 & a_3 & a_4 & \ldots \\ 0 & a_0 & a_1 & a_2 & a_3 & \ldots \\ 0 & 0 & a_0 & a_1 & a_2 & \ldots \\ 0 & 0 & 0 & a_0 & a_1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

で与えられる．状態 0 からは 1 ステップで任意の状態 j に到達可能であり，かつ，任意の状態 j ($j = 1, 2, \ldots$) からは j ステップで状態 0 に到達可能なので（そのような確率は a_0^j），マルコフ連鎖 X_n は既約であることが分かる．よって以下では X_n の定常状態確率が存在すると仮定する．

客の離脱直後における系内客数が k 人である定常状態確率を π_k とする．このとき定常状態分布 $\boldsymbol{\pi} = (\pi_j)$ は

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

を満たす．$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ を要素毎に書き下せば

$$\pi_j = \pi_0 a_j + \sum_{i=1}^{j+1} \pi_i a_{j+1-i}, \qquad j = 0, 1, \ldots \tag{30}$$

である．

ここで式 (30) の両辺を $j = 0$ から $k-1$ まで加えると

$$\begin{aligned}
\sum_{j=0}^{k-1} \pi_j &= \pi_0 \sum_{j=0}^{k-1} a_j + \sum_{j=0}^{k-1} \sum_{i=1}^{j+1} \pi_i a_{j+1-i} \\
&= \pi_0 \sum_{j=0}^{k-1} a_j + \sum_{i=1}^{k} \sum_{j=i-1}^{k-1} \pi_i a_{j+1-i} \\
&= \pi_0 \sum_{j=0}^{k-1} a_j + \sum_{i=1}^{k} \pi_i \sum_{j=0}^{k-i} a_j
\end{aligned}$$

を得る．これを π_k について解くと

$$\pi_k a_0 = \pi_0 \left(1 - \sum_{j=0}^{k-1} a_j \right) + \sum_{i=1}^{k-1} \pi_i \left(1 - \sum_{j=0}^{k-i} a_j \right) \tag{31}$$

となる．一回の遷移で高々一つしか系内客数が減らないことに注意すると，式 (31) の左辺は系内客数が k 以上から $k-1$ 以下に減少する確率フローの量と見なすことができ，一方，右辺は系内客数が $k-1$ 以下から k 以上になる確率フローの量と見なすことが出来る．一般に，定常状態では系内客数が k と $k-1$ の間を横切るフローの量は等しくなる．

定常状態において任意に選ばれた時点で系内客数が k 人である確率を p_k とする．π_k は客の離脱直後における系内客数分布であり，一般には p_k とは異なるが，$M/G/1$ においては以下の理由により両者が一致する (Gross and Harris 1998).

まず，系内客数を時間の関数として眺めると一回の変化で高々 1 しか動かない階段関数である（図 16 参照）．そこで，$A_k(t)$ を時間区間 $(0, t]$ の間に系内客数が k から $k+1$ に増加した時点の数とし，$D_k(t)$ を時間区間 $(0, t]$ の間に系内客数が $k+1$ から k に減少した時点の数とする．このとき $A(t) = \sum_{k=0}^{\infty} A_k(t)$ は時間区間 $(0, t]$ の間に到着する総客数を表し，$D(t) = \sum_{k=0}^{\infty} D_k(t)$ は時間区間 $(0, t]$ の間に離脱した総客数を表すことに注意する．定義より

$$\pi_k = \lim_{t \to \infty} \frac{D_k(t)}{D(t)}$$

である．ここで $L(t)$ を時刻 t における系内客数とすると $L(t) = L(0) + A(t) - D(t)$ が成立する．よって

$$\lim_{t \to \infty} \frac{D_k(t)}{D(t)} = \lim_{t \to \infty} \frac{A_k(t)}{A(t)} \cdot \frac{A(t)}{A(t) + L(0) - L(t)} + \lim_{t \to \infty} \frac{D_k(t) - A_k(t)}{D(t)}$$

を得る．さらに $L(0) < \infty$，かつ，$L(t)$ が極限分布をもつと仮定すると，$t \to \infty$ の極限では $A(t)/[A(t) + L(0) - L(t)] \to 1$ となり，一方，$|A_k(t) - D_k(t)| \leq 1$ であるので $[D_k(t) - A_k(t)]/D(t) \to 0$ となる．すなわち

$$\lim_{t \to \infty} \frac{D_k(t)}{D(t)} = \lim_{t \to \infty} \frac{A_k(t)}{A(t)} \tag{32}$$

が成立する．式 (32) は客の離脱直後の系内客数分布が客の到着直前の系内客数分布と等しいということを示している．また式 (32) を

$$\lim_{t\to\infty}\frac{D_k(t)}{t}\frac{t}{D(t)}=\lim_{t\to\infty}\frac{A_k(t)}{t}\frac{t}{A(t)}$$

のように変形し，$\lim_{t\to\infty}A(t)/t=\lim_{t\to\infty}D(t)/t$ に注意すると

$$\lim_{t\to\infty}\frac{D_k(t)}{t}=\lim_{t\to\infty}\frac{A_k(t)}{t} \tag{33}$$

を得る．式 (33) は単位時間当たりに状態 k を見る平均到着客数が離脱直後に状態 k を見る平均離脱客数に等しいことを示している．以上をまとめて次の補題を得る．

補題 1 客の到着ならびに離脱が順序性をもつ定常状態にある待ち行列モデルにおいて，客の到着直前の系内客数分布と客の離脱直後の系内客数分布は等しい．さらに，このとき，単位時間当たりに系内客数が k から $k+1$ へ変化する平均回数は単位時間当たりに系内客数が $k+1$ から k へ変化する平均回数に等しい．

一方，客の到着がポワソン過程に従うならば，定常状態において客の到着直前の系内客数分布は定常状態分布に等しい．それゆえ補題 1 より，離脱直後の系内客数分布は定常状態分布に等しくなる．特に $\pi_0=p_0$ であり，p_0 はサーバが稼働していない確率なので，$\pi_0=p_0=1-\rho$ を得る．

このように，式 (31) より π_k が順次求まり，離脱直後の系内客数分布 $\{\pi_k;k=0,1,\ldots\}$ は定常状態における系内客数分布 $\{p_k;k=0,1,\ldots\}$ に等しい．以上をまとめて次の定理を得る．

定理 11 $\rho<1$ のとき，$M/G/1$ の定常状態における系内客数分布 $\{p_k;k=0,1,\ldots\}$ は次式により計算される．

$$\begin{aligned}p_0 &= 1-\rho\\ p_k &= \frac{1}{a_0}\left[p_0\left(1-\sum_{j=0}^{k-1}a_j\right)+\sum_{i=1}^{k-1}p_i\left(1-\sum_{j=0}^{k-i}a_j\right)\right], \qquad k=1,2,\ldots\end{aligned}$$

図 17 はサービス時間が 1 である $M/D/1$ の系内客数 L の**裾野分布**（tail distribution） $\Pr(L>k)$ を対数軸で示している．図より，大きな k の値に対しては $\Pr(L>k)$ は一定の率で減少することがわかる．すなわち，十分大きな k に対して系内客数の裾野分布は $d>0$，$0<r<1$ なる定数 d，r を用いて $\Pr(L>k)\sim dr^k$ となっている．

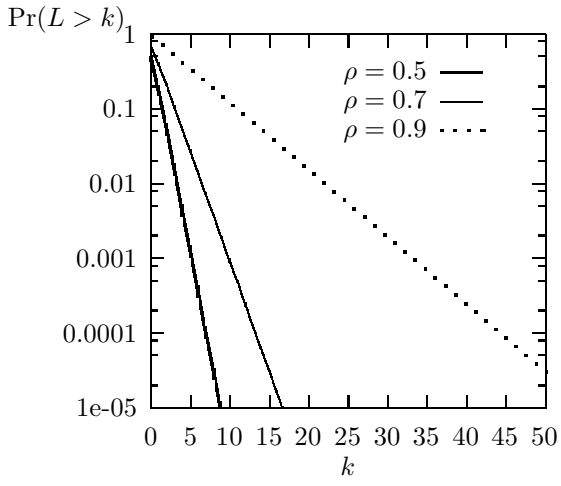

図 17: $M/D/1$ の系内客数の裾野分布（$b=1$）

次に系内客数分布の積率を計算する方法を考える．その準備として**確率母関数**（probability generating function）を導入する．[5]非負の整数値を取る確率変数 X に対して

$$G^*(z) = E\left[z^X\right] = \sum_{j=0}^{\infty} \Pr(X = j) z^j, \qquad |z| \leq 1$$

を X の確率母関数という．確率母関数 $G^*(z)$ は次の性質をもつ．

1. $G^*(z)$ は $|z| \leq 1$ で収束し，$|z| < 1$ で何回でも微分可能である．
2. $G^{(n)}(z)$ を $G^*(z)$ の n 階導関数とすると，$E[X^n] < \infty$ ならば $G^{(n)}(1) = E[X(X-1)\cdots(X-n+1)]$ であり，これは n 次の**階乗積率**（factorial moment）と呼ばれる．特に X の平均 $E[X]$ は $G^{(1)}(1)$ で与えられ，2 次積率 $E[X^2]$ は $G^{(2)}(1) + G^{(1)}(1)$ で与えられる．
3. 確率分布と確率母関数は 1 対 1 に対応している．
4. 独立な二つの確率変数 X，Y の確率母関数をそれぞれ $G_X^*(z)$，$G_Y^*(z)$ とすると，これらの和 $Z = X + Y$ の確率母関数 $G_Z^*(z)$ は $G_Z^*(z) = G_X^*(z) G_Y^*(z)$ で与えられる．

一方，非負の値を取る（連続）確率変数 X に対しては**ラプラス・スティルチェス変換**（Laplace-Stieltjes transform）がある．[6]非負の確率変数 X の分布関数を $G(x)$ としたとき，確率変数 X に対するラプラス・スティルチェス変換 $G^*(s)$ は

$$G^*(s) = E\left[e^{-sX}\right] = \int_0^{\infty} e^{-sx} dG(x), \qquad Re(s) > 0$$

で与えられる．X の密度関数 $g(x) = dG(x)/dx$ が存在する場合は，

$$G^*(s) = \int_0^{\infty} e^{-sx} g(x) dx$$

であり，密度関数に対する通常のラプラス変換と等価である．ラプラス・スティルチェス変換は密度関数が存在しないような場合に通常のラプラス変換を拡張したものと見ることができる．ラプラス・スティルチェス変換は次の性質をもつ．

1. $G^*(s)$ は $\mathrm{Re}(s) \geq 0$ で収束し，$\mathrm{Re}(s) > 0$ で何回でも微分可能である．
2. $G^{*(n)}(s)$ を $G(s)$ の n 階導関数とすると，$E[X^n] < \infty$ ならば $G^{*(n)}(0) = (-1)^n E[X^n]$ となる．特に X の平均 $E[X]$ は $-G^{*(1)}(0)$ で与えられ，2 次積率 $E[X^2]$ は $G^{*(2)}(0)$ で与えられる．
3. 確率分布とラプラス・スティルチェス変換は 1 対 1 に対応している．
4. 独立な確率変数 X，Y のラプラス・スティルチェス変換をそれぞれ $G_X^*(s)$，$G_Y^*(s)$ とすると，それらの和 $Z = X + Y$ のラプラス・スティルチェス変換 $G_Z^*(s)$ は $G_Z^*(s) = G_X^*(s) G_Y^*(s)$ で与えられる．

以上の準備のもとで $M/G/1$ の客の離脱時点における系内客数分布の確率母関数

$$L^*(z) = \sum_{j=0}^{\infty} \pi_j z^j$$

[5]確率母関数は Z 変換とも呼ばれる．
[6]ラプラス・スティルチェス変換は離散確率変数に対しても定義できるが，離散確率変数に対しては，通常，確率母関数を用いる．

を導く．$L^*(z)$ は定常状態における系内客数分布の確率母関数に等しいことに注意する．式 (30) の両辺に z^j を掛け，$j=0$ から無限大まで和を取ると

$$\begin{aligned} L^*(z) &= \pi_0 \sum_{j=0}^{\infty} a_j z^j + \sum_{j=0}^{\infty} \sum_{i=1}^{j+1} \pi_i a_{j+1-i} z^j \\ &= \pi_0 A^*(z) + \frac{1}{z} \sum_{i=1}^{\infty} \pi_i z^i \sum_{j=i-1}^{\infty} a_{j+1-i} z^{j+1-i} \\ &= \pi_0 A^*(z) + \frac{1}{z} \left(L^*(z) - \pi_0 \right) A^*(z) \end{aligned} \tag{34}$$

を得る．ただし $A^*(z)$ は任意に選ばれた客のサービス時間の間に到着する客数の確率分布 $\{a_k; k=0,1,\ldots\}$ の確率母関数であり

$$\begin{aligned} A^*(z) &= \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} \int_0^{\infty} e^{-\lambda x} \frac{(\lambda x)^k}{k!} dH(x) z^k \\ &= \int_0^{\infty} e^{-\lambda x} \sum_{k=0}^{\infty} \frac{(\lambda z x)^k}{k!} dH(x) = \int_0^{\infty} e^{-\lambda(1-z)x} dH(x) \end{aligned}$$

で与えられる．ここで，サービス時間分布のラプラス・スティルチェス変換

$$H^*(s) = \int_0^{\infty} e^{-sx} dH(x)$$

を用いれば

$$A^*(z) = H^*(\lambda - \lambda z)$$

となる．式 (34) を $L^*(z)$ について解くと

$$L^*(z) = \frac{\pi_0 (z-1) H^*(\lambda - \lambda z)}{z - H^*(\lambda - \lambda z)}$$

を得る．右辺に現れる未知数 π_0 は正規化条件 $L^*(1) = \sum_{j=0}^{\infty} \pi_j = 1$ を用いて定めることができる．すなわち

$$1 = \lim_{z \to 1-} L^*(z) = \frac{\pi_0}{1-\rho} \tag{35}$$

となるので，$\pi_0 = 1-\rho$ が得られ $L^*(z)$ が完全に決定される．

定理 12 ($M/G/1$ の系内客数分布の母関数) $\rho < 1$ のとき，$M/G/1$ の系内客数分布 $\{p_k; k=0,1,\ldots\}$ の確率母関数 $L^*(z)$ は

$$L^*(z) = \frac{(1-\rho)(z-1) H^*(\lambda - \lambda z)}{z - H^*(\lambda - \lambda z)} \tag{36}$$

で与えられる．

確率母関数の性質より，$L^*(z)$ を微分することで系内客数分布の階乗積率を求めることができる．定常状態における系内客数分布の n 次の階乗積率を $L^{(n)}$ とし，$b^{(n)}$ をサービス時間分布の n 次積率とする．式 (36) の両辺を n 回微分し，$z \to 1$ とすれば，系内客数分布の階乗積率は次の再帰式を満たすことを示すことができる．

$$L^{(n)} = \sum_{k=0}^{n-1} \binom{n+1}{k} \frac{\lambda^{n+1-k} b^{(n+1-k)} L^{(k)}}{(n+1)(1-\rho)} + \lambda^n b^{(n)}, \qquad n = 1, 2, \ldots$$

ただし $L^{(0)} = 1$ とした．特に平均系内客数 $E[L] = L^{(1)}$ は

$$E[L] = \frac{\lambda^2 b^{(2)}}{2(1-\rho)} + \rho$$

で与えられる．さらにリトルの公式より，平均滞在時間 $E[W]$ は

$$E[W] = \frac{\lambda b^{(2)}}{2(1-\rho)} + b$$

となる．図 18 は利用率 ρ に対して平均系内客数を示したものである．平均系内客数 は利用率 ρ のみならず，サービスの変動要因 $b^{(2)}$ にも依存することに注意する．特に ρ が比較的大きい場合，$b^{(2)}$ の影響が顕著になる．また，いずれの場合も ρ が 1 に近付くと平均系内客数が急激に増大することが分かる．

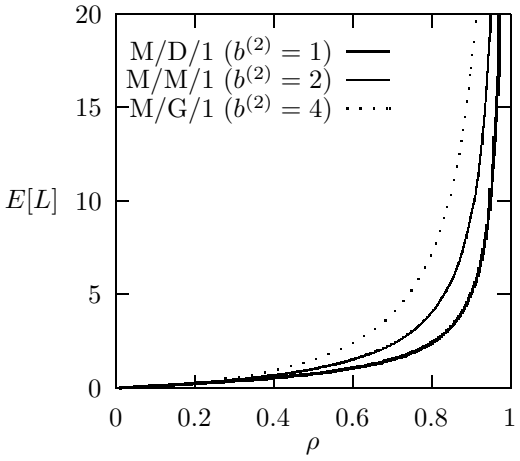

図 18: $M/G/1$ の平均系内客数（$b = 1$）

4.2.2 $M/G/1/K$ の系内客数分布

到着率 λ，サービス時間分布関数 $H(x)$，平均サービス時間 b をもつ $M/G/1/K$ を考える．システム容量が K なので，ある客の到着時に系内客数が K 人であれば，この客は棄却され呼損となる．例 3 において，ルータのバッファが高々 K パケットしか保持できないと仮定すれば，そのようなルータの振舞いは $M/G/1/K$ でモデル化できる．

定常状態における系内客数分布を導くため，$M/G/1$ の場合と同じく，客の離脱直後に注目する．X_n を n 番目の客の離脱直後の系内客数とし，A_n を n 番目の客のサービス中に新たに到着する客数とする．システムには高々 K 人の客しか存在し得ないので，離脱直後の系内客数は高々 $K-1$ であることに注意する．$M/G/1/K$ の系内客数の振舞いは，この点を除けば M/G/1 と同じであるので，$M/G/1$ において客数が $K-1$ 以上に遷移する場合は，$M/G/1/K$ では全て状態 $K-1$ に遷移することになる．すなわち

$$X_{n+1} = \min(\max(X_n - 1, 0) + A_{n+1}, K-1)$$

が成立する．この式より $M/G/1$ の場合と同様に X_n はマルコフ性をもち，客の離脱直後は隠れマルコフ点となることがわかる．

さらに $M/G/1$ の場合と同様の議論により，遷移確率 $p_{i,j}$（$i, j = 0, 1, \ldots, K-1$）は次式で与えられる．

$$p_{0,j} = \begin{cases} a_j, & j = 0, 1, \ldots, K-2, \\ \overline{a}_{K-1}, & j = K-1 \end{cases}$$

$$p_{i,j} = \begin{cases} 0, & i=1,\ldots,K-1, j=0,\ldots,i-2, \\ a_{j+1-i}, & i=1,\ldots,K-1, j=i-1,\ldots,K-2, \\ \overline{a}_{K-i}, & i=1,\ldots,K-1, j=K-1 \end{cases}$$

ただし

$$\overline{a}_j = \sum_{i=j}^{\infty} a_i = 1 - \sum_{i=0}^{j-1} a_i$$

である．よって客の離脱直後の系内客数の遷移確率行列 $\boldsymbol{P}^{(K)}$ は

$$\boldsymbol{P}^{(K)} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \ldots & a_{K-2} & \overline{a}_{K-1} \\ a_0 & a_1 & a_2 & a_3 & \ldots & a_{K-2} & \overline{a}_{K-1} \\ 0 & a_0 & a_1 & a_2 & \ldots & a_{K-3} & \overline{a}_{K-2} \\ 0 & 0 & a_0 & a_1 & \ldots & a_{K-4} & \overline{a}_{K-3} \\ 0 & 0 & 0 & a_0 & \ldots & a_{K-5} & \overline{a}_{K-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & a_0 & \overline{a}_1 \end{pmatrix}$$

で与えられ，定常状態分布 $\boldsymbol{\pi}^{(K)} = (\pi_j^{(K)})$ は

$$\boldsymbol{\pi}^{(K)} = \boldsymbol{\pi}^{(K)} \boldsymbol{P}^{(K)}, \qquad \sum_{j=0}^{K-1} \pi_j^{(K)} = 1$$

を満たす．$\boldsymbol{\pi}^{(K)} = \boldsymbol{\pi}^{(K)} \boldsymbol{P}^{(K)}$ を要素毎に書き下せば

$$\begin{aligned} \pi_j^{(K)} &= \pi_0^{(K)} a_j + \sum_{i=1}^{j+1} \pi_i^{(K)} a_{j+1-i}, \qquad j=0,\ldots,K-2 \\ \pi_{K-1}^{(K)} &= \pi_0^{(K)} \overline{a}_{K-1} + \sum_{i=1}^{K-1} \pi_i^{(K)} \overline{a}_{K-i} \end{aligned} \tag{37}$$

である．$\pi_j^{(K)}$ が満たすべき方程式は $\boldsymbol{\pi}^{(K)} = \boldsymbol{\pi}^{(K)} \boldsymbol{P}^{(K)}$ から K 個得られるが，行列 $\boldsymbol{P}^{(K)}$ のランクは $K-1$ であるので，その内一つは冗長である．よって，例えば，上に挙げた式 (37) と正規化条件 $\sum_{j=0}^{K-1} \pi_j^{(K)} = 1$ によって $\pi_j^{(K)}$ は一意に決定される．

ここで $j=0,\ldots,K-2$ の場合，$\pi_j^{(K)}$ は $M/G/1$ と同じ式に支配されていることに注意する．これより，$\pi_0^{(K)}$ に対して適当な非負の初期値を与え，順次 $\pi_j^{(K)}$（$j=1,\ldots,K-1$）を式 (31) より計算し，確率の総和が 1 となるように正規化してやれば客の離脱時点における定常状態確率を求めることができる．

次に，定常状態において任意に選ばれた時点における系内客数分布 $\{p_j^{(K)}; j=0.1.\ldots,K\}$ を求める．客の離脱直後の系内客数は高々 $K-1$ であるが，任意に選ばれた時点における系内客数は K である可能性があることに注意する．客の到着はポワソン過程に従っているため，呼損率は系内に K 人の客がいる定常状態確率 $p_K^{(K)}$ に等しい．さらに到着した客が系内に収容されるという条件の下で，その客が到着直前に見る客数分布は離脱直後の客数分布に等しいため，

$$\frac{p_j^{(K)}}{1-p_K^{(K)}} = \pi_j^{(K)}, \qquad j=0,\ldots,K-1 \tag{38}$$

を得る．特に $j=0$ を考えると

$$p_0^{(K)} = \left(1-p_K^{(K)}\right) \pi_0^{(K)} \tag{39}$$

が成立する．

一方，単位時間当たりにサービスを開始する平均客数は $\left(1-p_K^{(K)}\right)\lambda$ であり，サーバでの平均滞在時間は b なので，$\rho=\lambda b$ とするとリトルの公式よりサーバにいる平均客数，すなわちサーバが稼働している確率は $\left(1-p_K^{(K)}\right)\rho$ で与えられる．よって

$$p_0^{(K)} = 1-\left(1-p_K^{(K)}\right)\rho \tag{40}$$

が成立する．さらに，式 (39) 及び式 (40) より

$$\left(1-p_K^{(K)}\right)\pi_0^{(K)} = 1-\left(1-p_K^{(K)}\right)\rho \tag{41}$$

となり，これを $p_K^{(K)}$ について解けば $p_K^{(K)}$ が得られる．一方，式 (38) より $p_j^{(K)}=(1-p_K^{(K)})\pi_j^{(K)}$ $(j=0,\ldots,K-1)$ なので次の定理を得る．

定理 13 *$\pi_j^{(K)}$ $(j=0,\ldots,K-1)$ を式 (31) を満たし，かつ，和が 1 である正数とする．このとき $M/G/1/K$ の系内客数分布 $\{p_j^{(K)}; j=0,\ldots,K\}$ は*

$$\begin{aligned} p_j^{(K)} &= \frac{\pi_j^{(K)}}{\pi_0^{(K)}+\rho}, \qquad j=0,\ldots,K-1 \\ p_K^{(K)} &= 1-\frac{1}{\pi_0^{(K)}+\rho} \end{aligned}$$

で与えられる．

もし $\rho<1$ ならば，$M/G/1/K$ の離脱直後の系内客数分布は，$M/G/1$ において離脱直後の系内客数が $K-1$ 以下であるという条件の下での離脱直後の条件付き系内客数分布に等しいことがわかる．なぜならば，式 (31) は両方のモデルに対して成立し，この式によって π_j ならびに $\pi_j^{(K)}$ $(j=0,\ldots,K-1)$ の比が完全に決定されるからである．さらに，$\pi_j=p_j$ であるので次式を得る．

$$\pi_j^{(K)} = \frac{p_j}{\sum_{k=0}^{K-1} p_k}, \qquad j=0,\ldots,K-1$$

このように $\rho<1$ の場合，定常状態における $M/G/1/K$ の系内客数分布は，対応する $M/G/1$ の定常状態系内客数分布を用いて表すことができる．

さらに，$\rho<1$ の仮定の下で，対応する $M/G/1$ における客数の裾野分布 E_K を次式で定義する．

$$E_K = \sum_{j=K}^{\infty} p_j = 1-\sum_{j=0}^{K-1} p_j$$

E_K は $M/G/1$ において系内客数が K 以上である定常状態確率である．$p_0=1-\rho$ ならびに $\pi_0^{(K)}=p_0/(1-E_K)$ に注意して呼損率 $p_K^{(K)}$ を書き換えると

$$p_K^{(K)} = 1-\frac{1}{\dfrac{1-\rho}{1-E_K}+\rho} = 1-\frac{1-E_K}{1-\rho E_K} = \frac{(1-\rho)E_K}{1-\rho E_K} \tag{42}$$

を得る．このように $\rho<1$ の場合，$M/G/1/K$ の呼損率 $p_K^{(K)}$ は対応する $M/G/1$ の利用率 ρ と裾野分布 E_K を用いて表現できる．

図 19 はサービス時間が 1 である $M/D/1/K$ の呼損率 $p_K^{(K)}$ を対数軸で示している．図より，大きな K の値に対して呼損率はシステム容量に対して一定の率で減少することがわかる．この性質は次のように説明できる．まず，図 17 で説明したように多くの単一サーバ待ち行列モデルでは裾野分布 E_K は等比級数的に減少する．一方，

式 (42) より $p_K^{(K)} = (1-\rho)E_K + o(E_K)$ であるので，十分小さな呼損率に対しては $p_K^{(K)} \approx (1-\rho)E_K$ となる．よって K が大きくなると呼損率も等比級数的に減少する．この性質は，ルータが $M/G/1/K$ でモデル化可能であれば，ルータのバッファ容量を 2 倍にすることで呼損率をおよそ一桁小さくできることを示している．

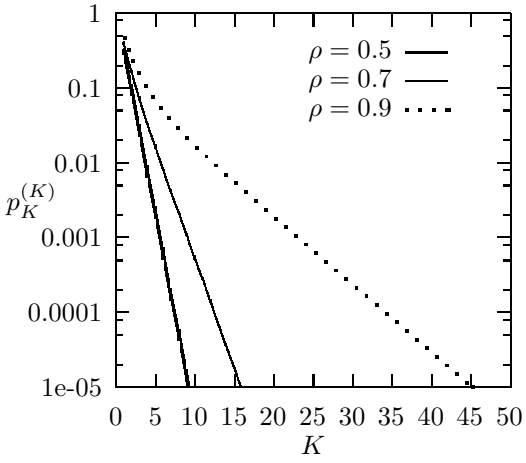

図 19: $M/D/1/K$ の呼損率（$b = 1$）

4.2.3 $GI/M/1$ の客数分布

到着間隔が平均 λ^{-1} をもつ独立同一な分布関数 $G(x)$ に従い，サービス時間がパラメタ μ の指数分布に従う $GI/M/1$ を考える．

例 4 十分に大きなバッファをもつルータがあり，出力回線の容量は C *bps* であるとする．パケットの到着間隔（秒）が独立同一な分布関数 $G(x)$ に従い，パケット長（*byte*）はパラメタ μ の指数分布に従うとする．このとき，ルータ内のパケット数の振舞いは，到着間隔分布関数 $G(x)$ とパラメタ $\mu C/8$ の指数サービスをもつ $GI/M/1$ でモデル化できる．

以下では利用率を $\rho = \lambda/\mu$ とし，$\rho < 1$ であると仮定する．定常状態における系内客数分布を導くため，客の到着直前の系内客数に注目する．Y_n を n 番目の客の到着直前の系内客数とし，B_n を $n-1$ 番目と n 番目の到着の間に離脱した客数とする．図 20 に $GI/M/1$ の客数の変化を示す．

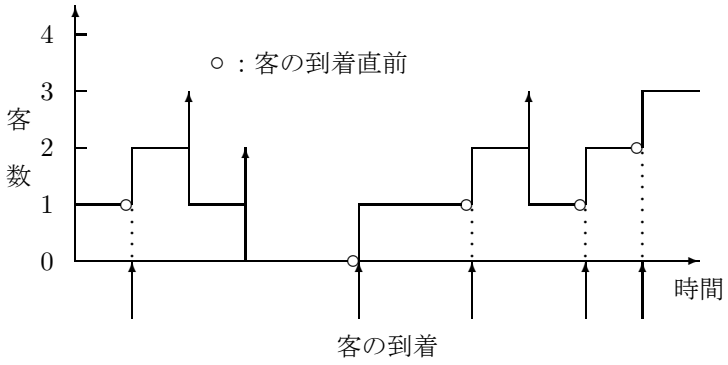

図 20: $GI/M/1$ の客数の変化

n 番目の客の到着直後の系内客数は Y_n+1 であり，この後，次の到着までに B_{n+1} 人の客が離脱するので，Y_n は次式を満たす.

$$Y_{n+1} = Y_n + 1 - B_{n+1}, \qquad n = 1, 2, \ldots$$

さらに $Y_n = i$ $(i = 0, 1, \ldots)$ という仮定の下で，ある客の到着直前から次の客の到着直前までの間に k 人 $(k = 0, \ldots, i)$ の客がサービスを受けて離脱する確率を b_k とする．もし到着間隔が x であれば，b_k は x の間に率 μ で k 人の客のサービスが終了する確率となり，これは平均 μx のポワソン分布に従う．よって $Y_n = i$ のとき b_k $(k = 0, \ldots, i)$ は

$$b_k = \int_0^\infty e^{-\mu x} \frac{(\mu x)^k}{k!} dG(x)$$

で与えられ，このとき $Y_{n+1} = i + 1 - k$ となる．一方，$i+1$ 人全てのサービスが終了する場合，$Y_{n+1} = 0$ となり，このような事象は確率 $1 - \sum_{k=0}^{i} b_k$ で起こる．以上の議論より，客の到着直前の系内客数に対する遷移確率は

$$p_{i,j} = \begin{cases} \overline{b}_{i+1} & j = 0 \\ b_{i+1-j}, & j = 1, \ldots, i+1 \\ 0, & j = i+2, i+3, \ldots \end{cases}$$

で与えられる．ただし

$$\overline{b}_k = \sum_{i=k}^{\infty} b_i$$

である．よって遷移確率行列 $\boldsymbol{P}$ は

$$\boldsymbol{P} = \begin{pmatrix} \overline{b}_1 & b_0 & 0 & 0 & 0 & 0 & \ldots \\ \overline{b}_2 & b_1 & b_0 & 0 & 0 & 0 & \ldots \\ \overline{b}_3 & b_2 & b_1 & b_0 & 0 & 0 & \ldots \\ \overline{b}_4 & b_3 & b_2 & b_1 & b_0 & 0 & \ldots \\ \overline{b}_5 & b_4 & b_3 & b_2 & b_1 & b_0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

となり，到着直前の客数の定常状態確率ベクトル $\boldsymbol{\pi} = (\pi_j)$ は

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

を満たす．$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ を要素毎に書き下せば

$$\pi_0 = \sum_{i=0}^{\infty} \pi_i \overline{b}_{i+1}, \qquad \pi_j = \sum_{i=j-1}^{\infty} \pi_i b_{i+1-j}, \qquad j = 1, 2, \ldots \tag{43}$$

である．

さて，式 (43) の 2 番目の式を γ 倍したものと 2 番目の式で π_{j+1} に対応するものの差を取ると

$$\gamma\pi_j - \pi_{j+1} = \sum_{i=j-1}^{\infty} (\gamma\pi_i - \pi_{i+1}) b_{i+1-j}, \qquad j = 1, 2, \ldots \tag{44}$$

が成立する．よって，もし $0 < \gamma < 1$ なる γ に対して

$$\gamma\pi_j = \pi_{j+1} \tag{45}$$

であるならば，式 (44) の両辺は等しくなるため，式 (43) の 2 番目の式が成立し，$\sum_{j=0}^{\infty} \pi_j = 1$ より

$$\pi_j = (1-\gamma)\gamma^j, \qquad j = 0, 1, \ldots \tag{46}$$

となる．さらに，もし，式 (46) で与えられる π_j が式 (43) の 1 番目の式を満たせば，定常状態確率を支配する全ての制約を満たすことになる．定常状態が存在すればそれは唯一の定常状態確率をもつため，このとき式 (46) で与えられる π_j は定常状態確率に他ならない．

よって，以下では式 (46) を仮定し，γ を求めることを考える．式 (43) の 1 番目の式に式 (46) を代入すると

$$1 - \gamma = \sum_{i=0}^{\infty} (1-\gamma)\gamma^i \overline{b}_{i+1}$$

となり，これより

$$1 = \sum_{i=0}^{\infty} \gamma^i \overline{b}_{i+1} = \frac{1 - G^*(\mu - \mu\gamma)}{1-\gamma} \tag{47}$$

を得る．ただし $G^*(s)$ は分布関数 $G(x)$ のラプラス・スティルチェス変換であり

$$G^*(s) = \int_0^{\infty} e^{-sx} dG(x)$$

で与えられる．それゆえ，もし

$$\gamma = G^*(\mu - \mu\gamma) \tag{48}$$

なる $0 < \gamma < 1$ が存在すれば，到着直前の定常状態確率 π_j は式 (48) を満たす γ を用いて式 (46) で与えられることになる．

$0 < \gamma < 1$ なる γ が存在するための条件は次のようにして得られる．$f(z) = G^*(\mu - \mu z)$ とおくと，$f(0) > 0$, $f(1) = 1$ であり，$df(z)/dz = -\mu dG^*(s)/ds > 0$ かつ $d^2 f(z)/dz^2 = \mu^2 d^2 G^*(s)/ds^2 > 0$ であるので，$z = f(z)$ が区間 (0,1) 内に解を持つための条件は $df(z)/dz|_{z=1-} > 1$ である（図 21 参照）．ここで

$$\frac{d}{dz} f(z)|_{z=1} = -\mu \frac{d}{ds} G^*(s)|_{s=0} = -\mu \times -\lambda^{-1} = \rho^{-1}$$

なので，$\rho < 1$ ならば $df/dz|_{z=1-} > 1$ であり，式 (48) を満たす γ は区間 (0,1) 内に唯一存在する．

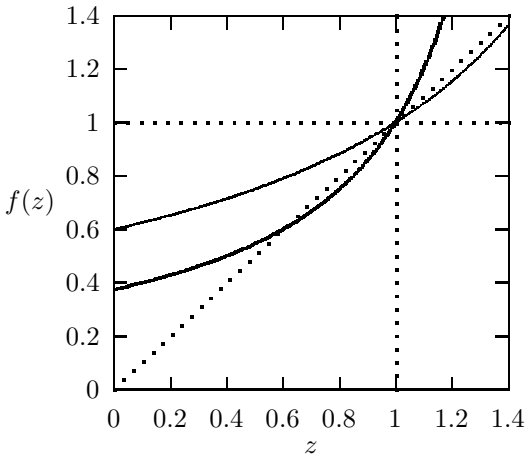

図 21: $f(z) = G^*(\mu - \mu z)$ の挙動

次に定常状態における任意に選ばれた時点で系内客数が j 人である確率 p_j を到着直前における系内客数の定常確率 π_j を用いて表すことを考える．単位時間当たりに到着する平均客数は λ で与えられ，到着客は確率 π_j で到着時に j 人の客を見る．よって，単位時間当たりに系内客数が j 人から $j+1$ 人へ変化する平均回数は $\lambda\pi_j$ で与えられる．一方，$j+1$ 人から j 人への変化は，ある時点で $j+1$ 人の客がおり，サービスが終了したとき起こる．よって，単位時間当たりに系内客数が $j+1$ 人から j 人へ変化する平均回数は μp_{j+1} で与えられる．一方，補題 1 より，定常状態では系内客数が j 人から $j+1$ 人へ変化する平均回数は $j+1$ 人から j 人へ変化する平均回数に等しい．よって

$$\lambda\pi_j = \mu p_{j+1}, \qquad j = 0, 1, \ldots$$

が成立し

$$p_j = \rho\pi_{j-1}, \qquad j = 1, 2, \ldots$$

となる．また $p_0 = 1 - \sum_{j=1}^{\infty} p_j$ なので $p_0 = 1 - \rho$ となる．[7]以上をまとめて次の定理を得る．

定理 14 $\rho < 1$ のとき，定常状態における $GI/M/1$ の系内客数分布 $\{p_j; j = 0, 1, \ldots\}$ は式 *(48)* を満たす唯一の $\gamma \in (0, 1)$ を用いて

$$\begin{aligned} p_0 &= 1 - \rho \\ p_j &= \rho(1-\gamma)\gamma^{j-1}, \qquad j = 1, 2, \ldots \end{aligned}$$

で与えられる．

また，定理 14 ならびにリトルの公式より，$GI/M/1$ の平均系内客数 $E[L]$ と平均系内滞在時間 $E[W]$ はそれぞれ

$$E[L] = \frac{\rho}{1-\gamma}, \qquad E[W] = \frac{1}{\mu(1-\gamma)}$$

で与えられる．$M/G/1$ の平均系内客数ならびに平均系内滞在時間がサービス時間分布の平均と 2 次積率で決まるのに対し，$GI/M/1$ では γ が到着間隔分布の関数になっているため，確率分布そのものに依存していることに注意する．$GI/M/1$ において到着間隔分布が性能に与える影響については次節で改めて論じる．

[7]システムが空である確率が $1 - \rho$ であることから，この結果は自明である．

5　待ち時間分布

この節では 3 3.2 節ならびに 4 4.2 節で考察した待ち行列モデルの待ち時間分布を導出する．

5.1　指数サービスをもつ FCFS 待ち行列の待ち時間分布

指数サービスをもつ定常な FCFS 単一サーバ待ち行列における待ち時間分布を考える．以下では，任意に選ばれた客の到着直前の系内客数分布 $\{q_k; k = 0, 1, \ldots\}$ が既知であるとする．ある客の到着直前に k 人の客がシステムにいた場合，これら k 人の客のサービスが終了すれば，到着した客のサービスが開始される．サービス時間がパラメタ μ の指数分布に従っているならば，指数分布の無記憶性より，到着時点でサービス中の客がサービスを終了するまでの時間もパラメタ μ の指数分布に従い，他の事象とは独立である．よって，任意に選ばれた客の待ち時間 W_q は，到着時点で系内に k 人の客がいれば，パラメタ μ をもつ k 個の独立な指数分布に従う確率変数の和で与えられる．

そこで，パラメタ μ の指数分布に従う独立な確率変数列 H_i （$i = 1, 2, \ldots$）に対して，$F_k = H_1 + \cdots + H_k$ と定義し，F_k の従う確率分布を見い出す．F_k の分布関数を $F_k(x)$ とすれば

$$F_2(x) = \int_0^x (1 - e^{-\mu(x-y)})\mu e^{-\mu y} dy = 1 - e^{-\mu x} - e^{-\mu x}\mu x$$

となる．同様に，

$$F_3(x) = \int_0^x F_2(x-y)\mu e^{-\mu y} dy = 1 - e^{-\mu x} - e^{-\mu x}\mu x - e^{-\mu x}\frac{(\mu x)^2}{2}$$

であり，帰納法によって

$$F_k(x) = 1 - \sum_{i=0}^{k-1} e^{-\mu x}\frac{(\mu x)^i}{i!}, \qquad k = 1, 2, \ldots \tag{49}$$

となることを示すことができる．さらに $F_k(x)$ を微分することにより，F_k の密度関数 $f_k(x)$ は

$$f_k(x) = \frac{(\mu x)^{k-1}}{(k-1)!} e^{-\mu x}\mu \tag{50}$$

で与えられることが分かる．

定義 7 k 個の独立なパラメタ μ をもつ指数分布に従う確率変数の和が従う確率分布を k 次の**アーラン分布**（*Erlang distribution*）といい，その分布関数ならびに密度関数はそれぞれ式 *(49)* ならびに式 *(50)* で与えられる．特に平均は k/μ で与えられ，分散は $(k/\mu)^2/k$ で与えられる．

図 22 に平均が 1 であるアーラン分布の密度関数を示す．この図から次数が高くなるに従って分散が減少する様子がわかる．また，アーラン分布の分散は常に同じ平均をもつ指数分布よりも小さいことに注意する．

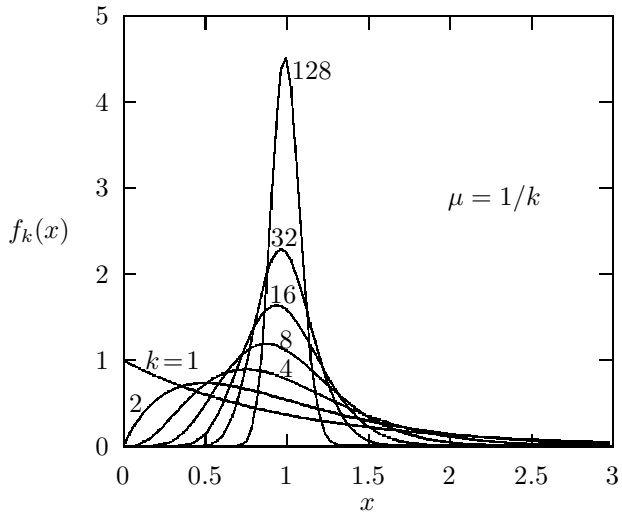

図 22: 平均 1 のアーラン分布の密度関数

5.1.1 $M/M/1$ の待ち時間分布

到着率 λ，サービス率 μ をもつ定常な FCFS $M/M/1$ の待ち時間分布を考える．到着直前の系内客数分布を $\{q_k; k = 0, 1, \ldots\}$ とすると，待ち時間 W_q の分布関数 $W_q(x) = \Pr(W_q \le x)$ は

$$W_q(x) = q_0 + \sum_{k=1}^{\infty} q_k F_k(x)$$

で与えられる．ポワソン到着の性質より，到着直前における系内客数の定常状態確率 q_k は任意に選ばれた時点における定常状態確率 $p_k = (1-\rho)\rho^k$ に等しいことに注意すると，

$$\begin{aligned}
W_q(x) &= 1 - \rho + \sum_{k=1}^{\infty}(1-\rho)\rho^k \left(1 - \sum_{n=0}^{k-1} e^{-\mu x}\frac{(\mu x)^n}{n!}\right) \\
&= 1 - \sum_{k=1}^{\infty}(1-\rho)\rho^k \sum_{n=0}^{k-1} e^{-\mu x}\frac{(\mu x)^n}{n!} \\
&= 1 - \rho e^{-\mu x}\sum_{n=0}^{\infty}\left(\frac{(\mu x)^n}{n!}\right)\sum_{k=n+1}^{\infty}(1-\rho)\rho^{k-1} \\
&= 1 - \rho e^{-\mu x}\sum_{n=0}^{\infty}\frac{(\mu x)^n}{n!}\rho^n
\end{aligned}$$

となる．さらに最後の等号の右辺に現れる無限和は $\exp(\mu\rho x)$ となり，次の定理を得る．

定理 15 (FCFS $M/M/1$ の待ち時間分布) $\rho < 1$ のとき，定常な *FCFS* $M/M/1$ における客の待ち時間の分布関数 $W_q(x)$ は

$$W_q(x) = 1 - \rho e^{-(1-\rho)\mu x}, \qquad x \ge 0 \tag{51}$$

で与えられる．

式 (51) より待ち時間の平均 $E[W_q]$ ならびに分散 $\mathrm{Var}[W_q]$ は，それぞれ

$$E[W_q] = \frac{\rho}{(1-\rho)\mu}, \qquad \mathrm{Var}[W_q] = \frac{\rho(2-\rho)}{((1-\rho)\mu)^2}$$

で与えられる．図 23 に FCFS $M/M/1$ の待ち時間の分布関数 $W(x)$ ならびに裾野分布 $\overline{W}(x) = 1 - W(x)$ を示す．利用率 ρ の値によって，待ち時間の特性が大きく異なることが分かる．

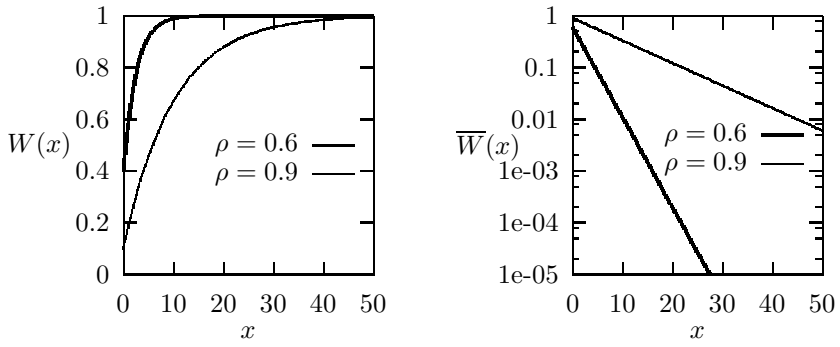

図 23: $M/M/1$ の待ち時間分布（$\mu = 1$）

5.1.2 $GI/M/1$ の待ち時間分布

到着間隔が平均 λ^{-1} をもつ分布関数 $G(x)$ に従い，サービス率 μ をもつ定常な FCFS $GI/M/1$ の待ち時間分布を考える．$\rho = \lambda/\mu < 1$ のとき，到着直前の系内客数が k 人である確率は $q_k = \pi_k = (1-\gamma)\gamma^k$ で与えられるので，$M/M/1$ と同様の議論により次の定理を得る．

定理 16 (FCFS $GI/M/1$ の待ち時間分布) $\rho < 1$ のとき，定常な *FCFS* $GI/M/1$ における待ち時間の分布関数 $W_q(x)$ は

$$W_q(x) = 1 - \gamma e^{-(1-\gamma)\mu x}, \qquad x \geq 0 \tag{52}$$

で与えられる．

式 (52) より待ち時間の平均 $E[W_q]$ ならびに分散 $\mathrm{Var}[W_q]$ は，それぞれ

$$E[W_q] = \frac{\gamma}{(1-\gamma)\mu}, \qquad \mathrm{Var}[W_q] = \frac{\gamma}{((1-\gamma)\mu)^2}$$

で与えられる．

到着間隔分布 $G(x)$ が性能に与える影響を見るために，同じ平均到着間隔 λ^{-1} ならびにサービス率 $\mu = 1$ をもつ $D/M/1$，$E_2/M/1$，$M/M/1$，$H_2/M/1$ における平均待ち時間を考える．ここで E_k は k 次のアーラン分布，H_k は k 次の**超指数分布**（hyperexponential distribution）を表すケンドールの記号である．k 次の超指数分布とは異なるパラメタ μ_i（$i = 1, \ldots, k$）をもつ指数分布を分岐確率 p_i（$i = 1, \ldots, k$）を用いて混合したものであり，その分布関数 $G_k(x)$ は

$$G_k(x) = \sum_{i=1}^{k} p_i(1 - e^{-\mu_i x})$$

で与えられ，平均 $\sum_{i=1}^{k} p_i/\mu_i$，2 次積率 $\sum_{i=1}^{k} 2p_i/\mu_i^2$ をもつ．超指数分布の 2 次積率は同じ平均をもつ指数分布の 2 次積率より常に大きいことが知られている．すなわち超指数分布の方が指数分布よりも変動が大きくな

る．図 24 にこれら 4 つの待ち行列モデルの平均待ち時間を示す．ただし 2 次の超指数分布は $G(x) = 0.25(1 - \exp(-\lambda x/2)) + 0.75(1 - \exp(-3\lambda x/2))$ である．図より到着時間間隔の変動が大きくなるに従い，性能が悪化することが分かる．

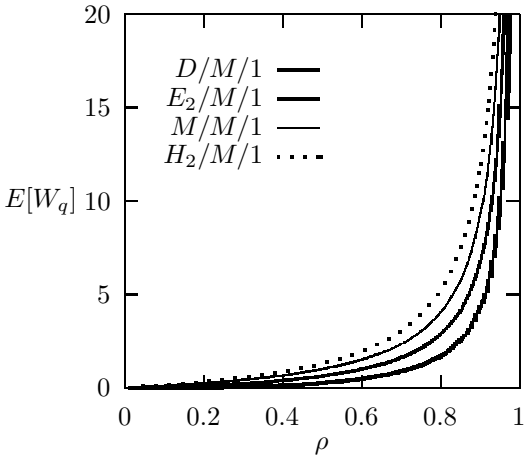

図 24: $GI/M/1$ における平均待ち時間（$\mu = 1$）

5.1.3 $M/M/c$ の待ち時間分布

到着率 λ，サービス率 μ をもつ定常な FCFS $M/M/c$ の待ち時間分布を考える．到着時点における系内客数が c 人以上のときのみ，到着した客は待たなければならない．さらに，客の到着直前の系内客数が $c-1+j$ $(j = 1, 2, \ldots)$ 人ならば，到着以降 j 人の客のサービスが終了した時点で，この客のサービスが開始される．指数分布の無記憶性より，到着時点でサービス中である個々の客がサービスを終了するまでの時間はそれぞれ独立なパラメタ μ の指数分布に従う．よって，$k \geq c$ 人の客が系内にいるとき，客の離脱間隔はパラメタ $c\mu$ の指数分布に従う．また，ポワソン到着の性質から，到着時点における系内客数分布 $\{q_k; k = 0, 1, \ldots\}$ は定常状態分布 $\{p_k; k = 0, 1, \ldots\}$ に等しい．

以上の考察により，サービス時間が指数分布に従う場合，複数のサーバをもつ待ち行列モデルの待ち時間分布は単一サーバの場合と同様にして求めることができることがわかる．特に $M/M/c$ の場合，$q_k = p_k = p_{c-1}(\rho/c)^{k-c+1}$ $(k = c, c+1, \ldots)$ なので，待ち時間 W_q の分布関数 $W_q(x) = \Pr(W_q \leq x)$ は

$$\begin{aligned} W_q(x) &= \sum_{k=0}^{c-1} p_k + \sum_{k=1}^{\infty} p_{c-1} \left(\frac{\rho}{c}\right)^k \left(1 - \sum_{i=0}^{k-1} e^{-c\mu x} \frac{(c\mu x)^i}{i!}\right) \\ &= \sum_{k=0}^{c-1} p_k + p_{c-1} \frac{\rho}{c-\rho} (1 - e^{-(c\mu-\lambda)x}) \end{aligned}$$

となる．さらに $\sum_{j=0}^{c-1} p_j = 1 - p_{c-1}\rho/(c-\rho)$ なので，次の定理を得る．

定理 17 (FCFS $M/M/c$ の待ち時間分布) $\rho < c$ のとき，定常な *FCFS* $M/M/c$ における待ち時間の分布関数 $W_q(x)$ は

$$W_q(x) = 1 - p_{c-1} \frac{\rho}{c-\rho} e^{-(c\mu-\lambda)x}, \qquad x \geq 0$$

で与えられる．

5.1.4 $M/M/1/K$ の待ち時間分布

到着率 λ，サービス率 μ をもつ定常な FCFS $M/M/1/K$ の待ち時間分布を考える．系内には高々 K 人の客しか収容できないため，呼損が起こる．このようなシステムにおける待ち時間は，通常，系内に収容された客に対してのみ定義される．ポワソン到着の性質より，到着直前の系内客数分布は定常状態分布と等しい．よって客が系内に収容されるという条件の下で，到着直前の系内客数が k 人である確率 q_k は

$$q_k = \frac{p_k}{1-p_K}, \qquad k = 0, \ldots, K-1$$

で与えられる．さらに，到着直前における系内客数が k 人であれば，この客の待ち時間は k 次のアーラン分布で与えられる．よって待ち時間 W_q の分布関数 $W_q(x) = \Pr(W_q \le x)$ は

$$\begin{aligned} W_q(x) &= q_0 + \sum_{k=1}^{K-1} q_k \left(1 - \sum_{i=0}^{k-1} e^{-\mu x} \frac{(\mu x)^i}{i!} \right) \\ &= 1 - \sum_{k=1}^{K-1} \frac{p_k}{1-p_K} \sum_{i=0}^{k-1} e^{-\mu x} \frac{(\mu x)^i}{i!} \end{aligned}$$

となり，和の順序を交換すると次の定理を得る．

定理 18 (FCFS $M/M/1/K$ の待ち時間分布) 定常な *FCFS* $M/M/1/K$ における待ち時間の分布関数 $W_q(x)$ は

$$W_q(x) = 1 - \sum_{i=0}^{K-2} \left(\sum_{k=i+1}^{K-1} \frac{p_k}{1-p_K} \right) e^{-\mu x} \frac{(\mu x)^i}{i!}, \qquad x \ge 0$$

で与えられる．

5.2 $M/G/1$ の待ち時間分布

到着率 λ，サービス時間が平均 b をもつ分布関数 $H(x)$ に従う定常な FCFS $M/G/1$ の待ち時間分布を考える．客の到着直前に k 人の客がいたとすると，この客の待ち時間は現在サービス中の客の残余サービス時間とそれに続く $k-1$ 人のサービス時間の和で与えられる．一般に，到着直前に系内にいる客は，現在行われているサービスの開始時点で既にいた客と現在のサービスの経過時間の間に到着する客に分類することができる．サービス時間が一般分布に従う場合，経過サービス時間と残余サービス時間の間には相関があるため，経過時間内に到着する客数と残余サービス時間の間に相関が生じる．[8]よって，指数サービスの場合と同様の議論を行うことができない．

そこで，以下では待ち時間分布と客の離脱直後の客数分布を関連づけることを考える．サービス時間分布 $H(x)$ のラプラス・スティルチェス変換を $H^*(s)$ とし，任意に選ばれた客の待ち時間分布のラプラス・スティルチェス変換を $W_q^*(s)$ とする．客の系内滞在時間は待ち時間とサービス時間の和で与えられ，両者は独立なので，系内滞在時間分布のラプラス・スティルチェス変換 $W^*(s)$ は

$$W^*(s) = W_q^*(s) H^*(s)$$

で与えられる．

FCFS $M/G/1$ では，客の離脱直後に系内にいる全ての客は離脱した客の系内滞在時間の間に到着している．また，任意に選ばれた客の離脱時点における客数の確率母関数は式 (36) で与えられた定常状態における系内客数分布の確率母関数 $L^*(z)$ と等しいことに注意する．

[8]指数分布の場合は，無記憶性により，経過サービス時間と残余サービス時間は独立となるが，例えば，サービス時間が一定の場合，経過サービス時間が比較的長ければ残余サービス時間は短くなる．結果として，残余サービス時間が短ければ経過サービス時間の間に到着する客数は比較的多くなることが期待される．

そこで系内滞在時間の間に到着する客数を考える．任意に選ばれた客の系内滞在時間が x であったという条件の下で，その間に率 λ でポワソン到着する客数の確率母関数は

$$\sum_{n=0}^{\infty} e^{-\lambda x} \frac{(\lambda x)^n}{n!} z^n = e^{-(\lambda - \lambda z)x}$$

で与えられる．よって $W(x)$ を任意に選ばれた客の系内滞在時間の分布関数とすると，この客の離脱直後の客数の確率母関数 $L^*(z)$ は

$$L^*(z) = \int_0^{\infty} e^{-(\lambda - \lambda z)x} dW(x)$$

を満たす．この式は，系内滞在時間分布のラプラス・スティルチェス変換の引数 s に $\lambda - \lambda z$ を代入した形になっている．よって

$$L^*(z) = W^*(\lambda - \lambda z) = W_q^*(\lambda - \lambda z) H^*(\lambda - \lambda z) \tag{53}$$

を得る．ここで $s = \lambda - \lambda z$ とおき，式 (53) を $W_q^*(s)$ について解くと

$$W_q^*(s) = L^*\left(\frac{\lambda - s}{\lambda}\right) / H^*(s)$$

を得る．さらに式 (36) を用いると次の定理を得る．

定理 19 (FCFS $M/G/1$ の待ち時間分布) $\rho < 1$ のとき，*FCFS* $M/G/1$ における待ち時間分布のラプラス・スティルチェス変換 $W_q^*(s)$ は

$$W_q^*(s) = \frac{(1-\rho)s}{s - \lambda + \lambda H^*(s)} \tag{54}$$

で与えられる．

$W_q^*(s)$ を微分することにより，待ち時間分布の積率を求めることができる．定常状態における待ち時間分布の n 次積率を $W_q^{(n)}$ とし，$b^{(n)}$ をサービス時間分布の n 次積率とする．式 (54) の両辺を微分し，$s \to 0$ とすることで，系内滞在時間分布の積率は一般に次の再帰式を満たすことを示すことができる．

$$W_q^{(n)} = \frac{\lambda}{(n+1)(1-\rho)} \sum_{k=0}^{n-1} \frac{(n+1)!}{k!(n+1-k)!} W_q^{(k)} b^{(n+1-k)}, \qquad n = 1, 2, \ldots$$

ただし $W_q^{(0)} = 1$ とした．特に，待ち時間の平均 $E[W_q] = W_q^{(1)}$ ならびに分散 $\mathrm{Var}[W_q] = W_q^{(2)} - E[W_q]^2$ は次式で与えられる．

$$E[W_q] = \frac{\lambda b^{(2)}}{2(1-\rho)}, \qquad \mathrm{Var}[W_q] = \frac{\lambda b^{(3)}}{3(1-\rho)} + E[W_q],$$

6 その他の話題

この節では，ここまでに触れることはできなかったが通信ネットワークの性能評価に有用な幾つかの話題を応用を交えて紹介する．

6.1 残余寿命分布

この節では，一般分布を要素に持つ待ち行列モデルを考察する上で非常に重要な残余寿命分布について解説する．**残余寿命**（residual life）とは，時間軸上に到着時点が与えられているとき，ランダムに選んだ時点から次の到着が起こる時点までの長さで定義される．

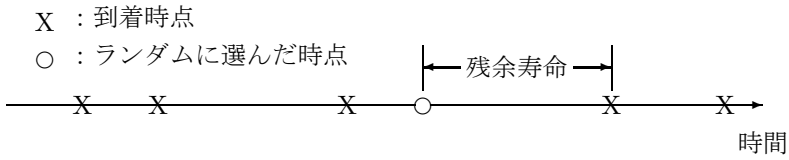

図 25: 残余寿命

これを次の例を用いて考える．

例 5 平均伝送時間が *10ms* であるパケット（以下，通常パケットと呼ぶ）が連続的に伝送されている回線を用いて特別なパケットを優先的に伝送することを考える．この特別なパケットは，発生時に伝送中の通常パケットが伝送を完了した後，直ちに伝送されるものとする．このとき，特別なパケットの伝送が開始されるまでの平均待ち時間を考える．特別なパケットの待ち時間は通常パケットの伝送時間の残余寿命に等しいことに注意する．以下では通常パケットの伝送時間 X *(ms)* の残余寿命を $\widetilde{X}$ *(ms)* で表す.

1. $\Pr(X = 10) = 1$ であるとする．この場合，特別なパケットの発生が通常パケットの伝送開始直後に起こる場合から伝送終了直前に起こる場合まで，すべて同様に確からしいので $E[\widetilde{X}] = 10/2 = 5ms$ となる．
2. $\Pr(X = 5) = 1/2$，$\Pr(X = 15) = 1/2$ であるとする．この場合，特別なパケットの発生時に伝送中の通常パケットの伝送時間は *5ms* と *15ms* の *2* 通りがあるが，伝送時間が *15ms* の通常パケットが伝送中である時間帯は時間軸全体の *3/4* を占める．すなわち，特別なパケットの発生時に伝送中の通常パケットの伝送時間は確率 *3/4* で *15ms* である．よって $E[\widetilde{X}] = 5/2 \times 1/4 + 15/2 \times 3/4 = 6.25ms$ となる．
3. $\Pr(X = 5) = 3/4$，$\Pr(X = 25) = 1/4$ であるとする．この場合，特別なパケットが発生時に伝送中の通常パケットの伝送時間は *5ms* と *25ms* の *2* 通りがあり，伝送時間には *5* 倍の差があるが，*5ms* の通常パケットの発生頻度は *25ms* の通常パケットに比べて *3* 倍である．よって，伝送時間が *25ms* の通常パケットが伝送中である時間帯は時間軸全体の
$$\frac{5 \times 1}{1 \times 3 + 5 \times 1} = \frac{5}{8}$$
を占める．すなわち，特別なパケットの発生時に伝送中の通常パケットの伝送時間は確率 *5/8* で *25ms* である．よって $E[\widetilde{X}] = 5/2 \times 3/8 + 25/2 \times 5/8 = 8.75ms$ となる．

この例からわかるように，特別なパケットが発生した時点でどのような伝送時間をもつ通常パケットが伝送中であるかという確率は，伝送時間とその発生頻度の両方に比例する．今，通常パケットの伝送時間 X の確率分布関数を $F(x) = \Pr(X \leq x)$ とし，X の密度関数 $f(x) = dF(x)/dx$ が存在するとする．このとき，ランダムに選ばれた時点において伝送中のパケットの伝送時間の密度関数は伝送時間の大きさ x と発生頻度 $f(x)$ の双方に比例する．すなわち選ばれる間隔の密度関数は，ある定数 α を用いて
$$\frac{xf(x)}{\alpha}$$
で与えられる．ここで，正規化定数 α は全区間に渡る積分が 1 となるように定められ（式 (2) 参照），$\alpha = E[X]$ であることがわかる．さらに，長さ x の間隔が選ばれた時，残余寿命はその中で一様に分布しているので
$$\Pr(y < \widetilde{X} \leq y + \Delta y) = \int_y^\infty \frac{xf(x)}{E[X]} \cdot \frac{\Delta y}{x} dx = \frac{1 - F(y)}{E[X]} \Delta y$$
を得る．

定理 20 確率変数 X が有限の平均 $E[X]$ ならびに分布関数 $F(x)$ をもつとき，X の残余寿命 $\widetilde{X}$ の密度関数は

$$\frac{1-F(x)}{E[X]} \tag{55}$$

で与えられる．また，残余寿命 $\widetilde{X}$ の n 次積率は

$$E[\widetilde{X}^n] = \frac{E[X^{n+1}]}{(n+1)E[X]} \tag{56}$$

で与えられる．

特に，平均残余寿命 $E[\widetilde{X}] = E[X^2]/(2E[X])$ が到着間隔の平均のみならず 2 次の積率に依存していることに注意する．これを X の分散 $\mathrm{Var}[X]$ を用いて書き換えると

$$E[\widetilde{X}] = \frac{E[X]}{2} + \frac{\mathrm{Var}[X]}{2E[X]}$$

となる．これより，分散 $\mathrm{Var}[X]$ が 0，すなわち X が確率 1 で一定の値を取るならば，残余寿命の平均は $E[X]/2$ で与えられ，最小となる．また，平均が同じであっても分散に上限はないため，平均残余寿命はいくらでも大きくなる可能性がある．

例 6 X がパラメタ λ の指数分布に従う場合，平均 $1/\lambda$，分布関数 $1-\exp(-\lambda x)$ をもつので，残余寿命の密度関数は

$$\frac{1-(1-e^{-\lambda x})}{\lambda^{-1}} = \lambda e^{-\lambda x}$$

となる．すなわち，指数分布の残余寿命分布は元の指数分布と等しい．[9]

6.2 $M/G/1$ の平均値公式

リトルの公式，ポワソン到着の性質ならびに残余寿命の平均を用いると定常な FCFS $M/G/1$ の平均待ち時間を求めることができる．以下では到着率を λ，平均サービス時間を b とし，利用率 $\rho = \lambda b < 1$ であると仮定する．$E[L_q]$ を平均待ち客数，$E[W_q]$ を平均待ち時間，$b^{(2)}$ をサービス時間の 2 次モーメントとする．

任意に選ばれた客の待ち時間は到着時点でサービス中ならばそのサービスの終了するまでの時間と到着時にサービスを待っていた客のサービス時間の和で与えられる．客の到着はポワソン分布に従うので，任意に選ばれた客の到着時に他の客がサービス中である確率は，定常状態におけるサービス中である時間割合に等しく，利用率 ρ で与えられる．また，到着時点で他の客がサービス中ならば，そのサービスが終了するまでにサービス時間の平均残余寿命 $b^{(2)}/2b$ だけ待たなければならない．さらに，客の到着はポワソン分布に従うので，到着時に待っている客の平均数は定常状態における平均待ち客数 $E[L_q]$ に等しく，到着した客はこれらの待っている客一人当たり平均 b のサービス時間分だけ待たなければならない．以上の考察より

$$E[W_q] = \rho\frac{b^{(2)}}{2b} + (1-\rho)\times 0 + bE[L_q] \tag{57}$$

を得る．一方，サーバを除いた部分を単独のシステムとして捉え，リトルの公式を用いると

$$E[L_q] = \lambda E[W_q] \tag{58}$$

を得る．よって，式 (57) ならびに式 (58) より平均待ち客数 $E[L_q]$ ならびに平均待ち時間 $E[W_q]$ はそれぞれ次式で与えられる．

$$E[L_q] = \frac{\lambda^2 b^{(2)}}{2(1-\rho)}, \qquad E[W_q] = \frac{\lambda b^{(2)}}{2(1-\rho)}$$

[9] 無記憶性よりこの結果は明らかである．

6.3 複数のポワソン流を収容する $M/G/1$ と非割込み優先規律

N 個の独立なポワソン到着流をもつ $M/G/1$ を考える．j 番目の到着流の到着率を λ_j，平均サービス時間を b_j，2 次積率 $b_j^{(2)}$，サービス時間の分布関数を $H_j(x)$ とする．このとき，定理 3 より，独立なポワソン流の重畳はポワソン流となり，到着した客が j 番目の到着流から来た客である確率は到着間隔とは独立に $\lambda_j/\sum_{i=1}^N \lambda_i$ で与えられるので，独立な複数のポワソン到着流をもつ待ち行列は以下のようなパラメタをもつ $M/G/1$ 待ち行列となる．

$$\lambda = \sum_{i=1}^N \lambda_i, \qquad b = \sum_{i=1}^N \frac{\lambda_i}{\lambda} b_i, \qquad b^{(2)} = \sum_{i=1}^N \frac{\lambda_i}{\lambda} b_i^{(2)}, \qquad H(x) = \sum_{i=1}^N \frac{\lambda_i}{\lambda} H_i(x)$$

よってサービスが FCFS で行われているならば，上記のパラメタを用いることで単一の到着流をもつ $M/G/1$ の結果が独立な複数の到着流をもつ $M/G/1$ にそのまま適用できる．

以下では FCFS ではなく，異なる到着流から来た客に異なる優先権を与えることを考える．今後，j 番目の到着流から来た客をクラス j の客と呼ぶ．クラス j の客はクラス $j+1,\ldots,N$ の客に対して**非割込み優先権**（nonpreemptive priority）を持つと仮定する．すなわち，サービス終了時に複数のクラスの客がサービスを待っている場合は，最も小さなクラスの客が次のサービスを受けることができる．一旦サービスが開始されると，その終了まで他の客がサービスされることはない．このようなサービス規律は非割込み型と呼ばれる．なお，システムが空のときに到着した客のサービスは直ちに行われる．以下ではクラス j の客の利用率を $\rho_j = \lambda_j b_j$ とし，システム全体での総利用率 $\rho = \sum_{i=1}^N \rho_i$ は $\rho < 1$ であると仮定し，定常状態における各クラスの客の平均待ち時間を導出する．

各クラスの客の到着はポワソン過程に従うため，任意に選ばれた客の到着時点においてクラス j の客がサービス中である確率は，任意に選ばれた時点においてサーバがクラス j の客をサービスしている確率 ρ_j に等しい．また，定常状態におけるクラス j の平均待ち客数を $E[L_{q,j}]$ としたとき，客の到着時点においてサービスを受けずに待っているクラス j の平均客数は定常状態における平均待ち客数 $E[L_{q,j}]$ で与えられる．さらに，客の到着時にクラス j の客がサービス中であったとき，客の到着時点からサービス終了時点までの平均長はクラス j のサービス時間の平均残余寿命 $b_j^{(2)}/(2b_j)$ で与えられる．

まず，クラス 1 の客の平均待ち時間 $E[W_{q,1}]$ を考える．$E[W_{q,1}]$ は到着時にサービスを待っているクラス 1 の客のサービス時間の総和の平均 $E[L_{q,1}]b_1$ と，到着時にサービス中ならば，サービス時間の平均残余寿命の和で与えられる．よって

$$E[W_{q,1}] = E[L_{q,1}]b_1 + \sum_{i=1}^N \rho_i \frac{b_i^{(2)}}{2b_i}$$

が成立する．一方，リトルの公式より $E[L_{q,1}] = \lambda_1 E[W_{q,1}]$ なので，$E[L_{q,1}]$ を消去することにより

$$E[W_{q,1}] = \frac{\sum_{i=1}^N \rho_i \frac{b_i^{(2)}}{2b_i}}{1-\rho_1}$$

を得る．

次にクラス 2 の平均待ち時間 $E[W_{q,2}]$ を考える．もし，注目するクラス 2 の客のサービスが開始されるまで，後続の客が誰も到着しなかったとすれば，この客の平均待ち時間 $E[T_2]$ は到着時に待っているクラス 1 および 2 の客のサービス時間と，もし到着時にサービス中ならば，サービス時間の平均残余寿命の和で与えられる．

$$E[T_2] = E[L_{q,1}]b_1 + E[L_{q,2}]b_2 + \sum_{i=1}^N \rho_i \frac{b_i^{(2)}}{2b_i}$$

しかし，注目するクラス 2 の客の待ち時間の間にクラス 1 の客が到着すれば，クラス 2 の客より先にサービス

されるため，注目するクラス 2 の客の待ち時間はこれらの客のサービス時間分だけ長くなる．そこで，注目するクラス 2 の客の平均待ち時間が，もし後続の到着がなければ $E[T_2]$ であるという条件の下で，実際にはどれだけ待たなければならないかを考える．

注目する客の到着後，単位時間当たり 平均 λ_1 人のクラス 1 の客が到着し，それぞれ平均 b_1 のサービスを受けるので，これらの到着によって平均 $\lambda E[T_2] \times b_1 = \rho_1 E[T_2]$ だけ待ち時間が長くなる．さらに，この増加した待ち時間 $\rho_1 E[T_2]$ の間に到着するクラス 1 の客によって，さらに待ち時間が長くなる．この増分の平均は，同様の議論により $\lambda\rho_1 E[T_2]b_1 = \rho^2 E[T_2]$ である．この議論を繰り返すと，結局，注目するクラス 2 の平均待ち時間は $E[T_2](1+\rho_1+\rho_1^2+\cdots) = E[T_2]/(1-\rho_1)$ となる．よって $E[L_{q,j}] = \lambda_j E[W_{q,j}]$（$j=1,2$）より

$$E[W_{q,2}] = \frac{E[T_2]}{1-\rho_1} = \frac{\displaystyle\sum_{i=1}^{N} \rho_i \frac{b_i^{(2)}}{2b_i}}{(1-\rho_1)(1-\rho_1-\rho_2)} \tag{59}$$

を得る．

最後にクラス j ($j=3,\ldots,N$）の平均待ち時間 $E[W_{q,j}]$ を考える．クラス j の客の視点に立つと，クラス 1 からクラス $j-1$ までの客は高い優先権を持っており，これらの客のサービス順序がどのようなものであろうとも，全てのサービスが終了するまで待たされることになり，その時間は高い優先権をもつ客のサービス順序に依らない．よって，クラス 1 からクラス $j-1$ の客をひとつのクラス H とみなし，クラス H, クラス $j, j+1,\ldots,N$ からなるシステムを考える．このとき最も高い優先権をもつクラス H の到着率 λ_H, 平均サービス時間 b_H，サービス時間の 2 次積率 $b_H^{(2)}$ はそれぞれ

$$\lambda_H = \sum_{i=1}^{j-1} \lambda_i, \qquad b_H = \sum_{i=1}^{j-1} \frac{\lambda_i}{\lambda_H} b_i, \qquad b_H^{(2)} = \sum_{i=1}^{j-1} \frac{\lambda_i}{\lambda_H} b_i^{(2)},$$

となり，利用率 ρ_H は $\rho_H = \rho_1 + \cdots + \rho_{j-1}$ で与えられる．これを式 (59) に適用すると

$$E[W_{q,j}] = \frac{\displaystyle\sum_{i=1}^{N} \rho_i \frac{b_i^{(2)}}{2b_i}}{(1-\rho_H)(1-\rho_H-\rho_j)} = \frac{\displaystyle\sum_{i=1}^{N} \rho_i \frac{b_i^{(2)}}{2b_i}}{\left(1-\displaystyle\sum_{i=1}^{j-1}\rho_i\right)\left(1-\displaystyle\sum_{i=1}^{j}\rho_i\right)}$$

を得る．非割込み優先規律を用いれば，異なる種類の客に異なる優先権を与えることでクラス間の性能を差別化することができる．

次に，クラス j ($j=1,\ldots,N$) の客の平均待ち時間に対して，単位時間当たり正のコスト C_j がかかるとし，適当な非割込み型サービス規律を用いて，システム全体でのコスト C_{total} を最小にするという問題を考える．

$$\text{minimize} \quad C_{\text{total}} = \sum_{j=1}^{R} C_j E[W_{q,j}]$$

このとき

$$\frac{C_{r_1}}{\rho_{r_1}} \geq \frac{C_{r_2}}{\rho_{r_2}} \geq \cdots \geq \frac{C_{r_N}}{\rho_{r_N}} \tag{60}$$

をみたすようなクラスの順序 $r_1, r_2, \ldots, r_N$ を定め，クラス r_j がクラス $r_{j+1},\ldots,r_N$ に対して優先権をもつ非割込み優先規律を用いると総コスト C_{total} が最小化されることが知られている (Gelenbe and Mitrani 1980).

例 7 特に $C_j = \lambda_j / \sum_{i=1}^{N} \lambda_i$ とすると，総コスト C_{total} は任意に選ばれた客の平均待ち時間，すなわち総平均待ち時間となる．このとき $C_j/\rho_j = 1/b_j \times 1/\sum_{i=1}^{N}\lambda_i$ であるので，式 *(60)* より平均サービス時間が短いクラスに対してより高い優先権を与えることで総平均待ち時間を最小化できることが分かる．

図 26 はクラス j（$j = 1, 2, 3$）の客がそれぞれ独立に同じ率 λ^* をもつポワソン過程に従い到着し，サービス時間がパラメタ $\mu_j = 2/j$ の指数分布に従う場合に，クラス 1, 2, 3 の順に高い優先権を与えたときの平均待ち時間を示したものである．参考のため，総平均待ち時間と FCFS の場合の平均待ち時間も示されている．この例では，$b_j = j/2$，$b_j^{(2)} = j^2/2$，$\rho_j = j\lambda^*/2$ である．また，FCFS の場合は $\lambda = 3\lambda^*$，$b = 1$，$b^{(2)} = 14/6$，$\rho = 3\lambda^*$，サービス時間の分布関数 $H(x)$ は 3 次の超指数分布 $H(x) = \frac{1}{3}(1 - e^{-2x}) + \frac{1}{3}(1 - e^{-x}) + \frac{1}{3}(1 - e^{-2x/3})$ で与えられる $M/G/1$ となることに注意する．図よりクラス間で性能に大きな違いがあることが分かる．特にシステムの負荷が 1 に近付くに従ってクラス 3 の平均待ち時間は発散していくが，クラス 1 と 2 の平均待ち時間は有限の値に留まっている．また，総平均待ち時間は FCFS の場合よりも小さくなっているが，これは最も低い優先権をもつクラス 3 の客の性能を犠牲にすることによって得られていることに注意する．

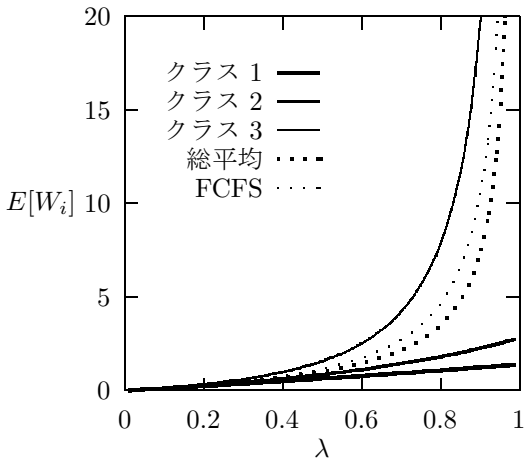

図 26: 非割込み優先規律における平均待ち時間（$\lambda = 3\lambda^*$）

6.4 プロセッサシェアリング待ち行列と公平性

以下のような単一サーバ待ち行列を考える．系内に n 人の客がいるとき，サーバはサービス能力を均等に n 分割し，これらの客をそれぞれ $1/n$ の能力で並行してサービスを行う．このようなサービス規律をプロセッサシェアリング（Processor-Sharing: PS）という (Kelly 1979).

例 8 ボトルネックとなっている回線を多くのフローが共有している状況を考える．回線速度を C *bps* としたとき，この回線を流れるフローが n 本であれば，個々のフローはあたかも C/n の回線速度でサービスされているように見なせる．そこで，通信ネットワーク内に流れる個々のフローを客とみなし，ボトルネックとなる回線をサーバとみなすと，プロセッサシェアリング規律に従う単一サーバ待ち行列モデルが得られる．

特に，客の到着が率 λ のポワソン過程に従い，客のサービスが独立なパラメタ μ の指数分布に従う場合，PS $M/M/1$ における系内客数は，出生率 λ，死滅率 μ をもつ出生死滅過程で表現される．すなわち，定常状態における系内客数分布は FCFS $M/M/1$ と同じになり，式 (21) で与えられる．

プロセッサシェアリング規律は幾つかの興味深い性質をもつことが知られている．まず，客の到着が率 λ のポワソン過程に従い，客のサービスが独立で平均 μ^{-1} をもつ PS $M/G/1$ において，$\rho = \lambda/\mu < 1$ ならば，定常状態における系内客数分布は，サービス時間分布の平均のみで定まり，サービス時間分布そのものには依存しない．

すなわち，$M/G/\infty$ や $M/G/c/c$ と同様にサービス時間分布に関して不感性をもっており，平均サービス時間が μ^{-1} で与えられるならば，定常状態における系内客数分布は式 (21) で与えられる．さらに，サービス時間 $H = x$ をもつ客の平均系内滞在時間 $E[W \mid H = x]$ はサービス時間分布には依存せず，

$$E[W \mid H = x] = \frac{x}{1-\rho} \tag{61}$$

で与えられる．

例 9 （例 *8* の続き）以下では一つのフローが伝送するデータ量の平均を h *byte* とする．特に，フローの発生が率 λ のポワソン過程に従うならば，この回線での遅延（全てのデータを送信し終るまでに必要な時間）は，利用率 $\rho = 8\lambda h/C < 1$ をもつ *PS* $M/G/1$ でモデル化でき，特に x *byte* のデータを伝送するのに必要な平均時間 $T[x]$ は式 *(61)* より

$$T[x] = \frac{8x/C}{1-\rho} \tag{62}$$

で与えられる．

式 (62) より，次のようなことがわかる．平均伝送時間は伝送するデータ量に線形である．例えば，2 倍のデータを送るためには平均 2 倍の時間がかかる．これを言い替えると，1 byte の伝送を行うために必要となる余分な時間は一定ということである．すなわち，データ量が x のフローが被る余分な時間は $(8x/C)/(1-\rho) - 8x/C = 8x/C \times \rho/(1-\rho)$ となるので，1 byte の伝送を行うために必要となる余分な時間は $8\rho/\{C(1-\rho)\}$ で与えられる．この余分な時間は 1 単位のデータの伝送に必要なペナルティーと見なすことができ，そのペナルティーが伝送するデータの総量と独立であるという意味で公平である．これ以外に，例 7 のように，短いデータの伝送に高い優先権を与えるという考え方もある．一方，長いデータの伝送を優先的に扱う場合，短いデータの伝送に対する平均遅延よりも長いデータの伝送に対する平均遅延の方が小さくなる可能性が生じる．もしこのようなことが起きれば，利用者は不必要に長いデータを通信ネットワークに流そうとするかも知れない．プロセッサシェアリングで達成される公平性は利用者のこのような振舞いを防ぐ働きがある．

6.5 多呼種 $M/G/c/c$

$M/M/c/c$ あるいは $M/G/c/c$ ではそれぞれの客が一つのサーバを占有すると仮定されているが，このモデルを拡張し，クラス 1 からクラス N まで N 種類の客があり，クラス j の客は c_j 個のサーバを同時に使用する場合を考える．クラス j の客は率 λ_j のポワソン過程に従い到着し，到着時に c 個のサーバのうち c_j 個以上のサーバが空であれば，その中から c_j 個のサーバを同時に占有し，サービスを受ける．空のサーバが c_j 個未満の場合は呼損となる．クラス j の客のサービス時間は一般分布に従い，その平均は b_j であるとする．サービスが終了すると占有していた c_j 個のサーバを同時に開放する．

例 10 帯域が C *bps* の回線があり，クラス j（$j = 1, \ldots, N$）の利用者は一定の帯域 c_j *bps* を要求する．もし到着時に空き帯域が c_j *bps* 未満であるならば，この利用要求は拒否され，呼損となる．クラス j の利用者の回線利用要求が率 λ_j のポワソン過程に従い発生し，回線接続時間が平均 b_j ならば，この回線の利用状況は上記の多呼種 $M/G/c/c$ でモデル化できる．ただし $c = C$ である．

このモデルで最も興味のある量はクラス j の客が呼損となる確率 B_j であり，これは次式で与えられる (Kelly 1979).

$$B_j = 1 - G(c - c_j)/G(c), \qquad j = 1, \ldots, R$$

ただし $\vec{n} = (n_1, \ldots, n_N)$ $(n_j \geq 0)$ と $\vec{c} = (c_1, \ldots, c_N)^T$ に対して $\Omega(x) = \{\vec{n}; 0 \leq \vec{n}\vec{c} \leq x\}$ とし，$\rho_j = \lambda_j b_j$ としたとき $G(x)$ は

$$G(x) = \sum_{\vec{n} \in \Omega(x)} \prod_{j=1}^{N} \frac{\rho_j^{n_j}}{n_j!}, \qquad 0 \leq x \leq c$$

で与えられる．$N = 1$ のとき，式 (24) のアーラン呼損式と等価である．

6.6 待ち行列網と積形式解

これまで議論してきた待ち行列モデルは全て一つの地点でサービスを要求するものであった．ここでは，ある地点でサービスを受けた客が他の地点に移動し，そこで新たなサービスを受ける待ち行列モデル，すなわち，**待ち行列網**（queueing network）について良く知られている結果を紹介する．図 27 は二つのノードからなる待ち行列網の例である．外部から網への到着は両方のノードにある．１段目のノードでサービスを終了した客の一部は網を離脱するが，残りは２段目のノードへ向かう．２段目のノードでは１段目のノードから来た客と外部から来た客が一つの待ち行列を作り，順次，サービスを受けた後，網を離脱する．

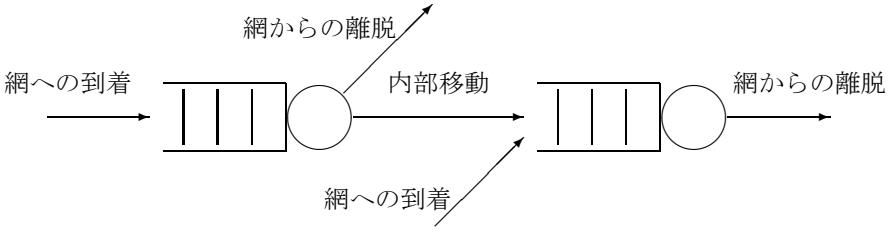

図 27: 待ち行列網

さて，J 個のノードからなる待ち行列網を考える．j 番目のノードには率 β_j のポワソン過程に従い外部から客が到着すると仮定する．j 番目のノードでのサービスは以下に示す３つのタイプのいずれか一つであるとする．

***M/M/c* 型** c_j 個のサーバがあり，サービス時間はパラメタ μ_j の指数分布に従う．

***M/G/∞* 型** 無限個のサーバがあり，平均サービス時間は $1/\mu_j$ である．便宜上 $c_j = \infty$ とする．

PS 型 １個のサーバがあり，客はプロセッサシェアリングサービス規律に従いサービスを受ける．平均サービス時間は $1/\mu_j$ である．便宜上 $c_j = 1$ とする．

さらに j 番目のノードでのサービスを終了した客は，確率 $r_{j,k}$ で k 番目のノードへ向かうと仮定する．よって，j 番目のノードでのサービスを終了後，網の外部へ離脱する確率 $r_{j,0}$ は

$$r_{j,0} = 1 - \sum_{k=1}^{J} r_{j,k} \geq 0$$

で与えられる．

ここで，単位時間当たりに j 番目のノードに到着する客数の平均を λ_j とすると，λ_j は次式を満たす．

$$\lambda_j = \beta_j + \sum_{i=1}^{N} \lambda_i r_{i,j} \tag{63}$$

さらに網に到着した客はいずれ必ず網を離脱すると仮定する．言い替えると，式 (63) を満たす正値 $\lambda_j < \infty$ が存在し，$\rho_j = \lambda_j/\mu_j$ としたとき，全ての j に対して $\rho_j < c_j$ が成立すると仮定する．このとき，この待ち行列網は安定であり，系内客数の結合確率に関して定常状態解が存在することが知られている．

L_j を定常状態における j 番目のノードの系内客数とする．系内客数の結合確率 $p(k_1, \ldots, k_J)$ を

$$p(k_1, \ldots, k_J) = \Pr(L_1 = k_1, \ldots, L_J = k_J)$$

で定義する．

定理 21 結合確率 $p(k_1, \ldots, k_J)$ は，以下のように各ノードにおける系内客数分布 $p_j(k_j) = \Pr(L_j = k_j)$ の積で与えられる *(Baskett et al. 1975)*.

$$p(k_1, \ldots, k_J) = p_1(k_1) \cdots p_J(k_J) \tag{64}$$

ただし $p_j(k_j)$ は $M/M/c$ 型ならば式 *(22)* で与えられ，$M/G/\infty$ 型ならば式 *(23)* で与えられ，*PS* 型ならば式 *(21)* で与えられる．

このように，系内客数に関する結合確率が個々のノードに関する確率の積で与えられるため，式 (64) は**積形式解**（product-form solution）と呼ばれている．