

2016年度待ち行列シンポジウム「確率モデルとその応用」

2017年1月19日(木)～21日(土)
東京理科大学 森戸記念館(東京都新宿区)

質疑応答含め一般 1人あたり30分, 学生 1人あたり25分
*=発表者(複数著者の場合)

1月19日(木)

13:20 開会にあたって 塩田 茂雄(千葉大学)

Session T1 (13:30-15:00) 企画セッション 座長: 小林 正弘(東海大学)
「確率・統計及びその応用」

13:30 正方分割表における非対称性のモデルについて
田畑 耕治(東京理科大学)

14:15 1次元線形セルオートマトンの極限挙動
竹居 正登(横浜国立大学)

Session T2 (15:15-16:30) 学生セッション 座長: 井家 敦(神奈川工科大学)

15:15 Binomial catastrophe の発生を伴うマルコフ変調 $M^X/M/\infty$ 待ち行列に対する中心極限定理
*矢島 萌子(東京工業大学), フンドックトゥアン(筑波大学)

15:40 Uplink coverage analysis for heterogeneous cellular networks
*小林 拓矢, 三好 直人(東京工業大学)

16:05 セルラネットワークにおける移動車両の通信精度解析手法
徳山 喜一(東京工業大学)

Session T3 (16:45-17:45) 一般セッション 座長: フンドックトゥアン(筑波大学)

16:45 M/G/1+PH 待ち行列における系内客数分布の数値計算法
井上 文彰(大阪大学)

17:15 Distribution of the times to service completion and abandonment in the M/M/m preemptive LCFS queue with impatient customers
高木 英明(筑波大学)

懇親会 (18:00-20:00) 司会: 井上 文彰(大阪大学)

1月20日(金)

Session F1 (9:30-11:10) 学生セッション 座長: 藤本 衡(東京電機大学)

9:30 ビットコインのマイニング・モデルとトランザクション承認時間解析
*河瀬 良亮, 笠原 正治(奈良先端科学技術大学院大学)

9:55 BGPネットワークにおける障害規模推定方法の検討
*後谷 浩輔, 小林 香(富山県立大学)

10:20 BGPメッセージを用いた迂回路検出手法の検討
*山谷 周平, 小林 香(富山県立大学)

10:45 ランダム多スタート局所探索法に対する極値統計を用いた計算停止基準
*仲川 諒馬, 増山 博之(京都大学), 檀 寛成(関西大学), 高橋 豊(京都大学)

11:10-12:40 昼休み

Session F2 (12:40–14:20) 学生セッション 座長: 木村 達明 (NTT)

- 12:40 既約でLumpableなマルコフ連鎖の上界の改善
*石川 真也, 河西 憲一 (群馬大学)
- 13:05 レベル依存するM/G/1型マルコフ連鎖における条件付き定常分布の数値計算アルゴリズム
*木村 雅俊, 滝根 哲哉 (大阪大学)
- 13:30 Geometric convergence of finite-level M/G/1-type Markov chains
*勝又 洋介, 増山 博之, 高橋 豊 (京都大学)
- 13:55 セルラネットワークの空間確率モデルに対するポアソン点過程近似について
*高橋 雄太, 三好 直人 (東京工業大学)

Session F3 (14:35–16:35) 企画セッション 座長: 岡村 寛之 (広島大学)

「マルコフ決定過程の信頼性評価への応用」

- 14:35 Maintenance optimization for multiple-unit systems based on Markov decision process
金 路 (電気通信大学)
- 15:15 Monotone properties of an optimal maintenance policy for a Markovian deteriorating system with imperfect repair and incomplete information
田村 信幸 (法政大学)
- 15:55 契約電力超過確率最小化に関する研究
*小柳 淳二, 古林 雄貴, 河野 大地 (鳥取大学)

特別講演 (16:50–17:50) 座長: 塩田 茂雄 (千葉大学)

情報システムと待ち行列
高橋 豊 (京都大学)

1月21日 (土)

Session S1 (9:30–10:20) 学生セッション 座長: 岸 康人 (東京交通短期大学)

- 9:30 Revisiting the heavy-traffic asymptotics of the GI/G/1-type Markov chain: a complete proof of the asymptotic formula for the stationary distribution
*木村 達明 (NTT), 増山 博之, 高橋 豊 (京都大学)
- 9:55 マルチエージェントモデルを用いたセルフレジ導入によるレジサービスへの影響と解析
*平田 直也, 中桐 齊之 (兵庫県立大学)

Session S2 (10:35–12:05) 一般セッション 座長: 河西 憲一 (群馬大学)

- 10:35 An arrival-time distribution for the equilibrium mean waiting time of a discrete-time single-server queue with acceptance period and Poissonian population of customers
*佐久間 大 (防衛大学校), 小林 正弘 (東海大学), 増山 博之 (京都大学)
- 11:05 A finite bottleneck game with homogeneous commuters
河崎 亮 (東京工業大学), 小西 秀男 (ボストンカレッジ), *湯川 隼貴 (東京工業大学)
- 11:35 上界と平均の上界が与えられた確率過程の特徴づけ
高田 寛之 (長崎大学)
- 12:15 待ち行列研究部会 研究奨励賞 受賞者発表 塩田 茂雄 (千葉大学)
閉会にあたって 塩田 茂雄 (千葉大学)

目 次

正方分割表における非対称性のモデルについて	1
田畑 耕治 (東京理科大学)	
1 次元線形セルオートマトンの極限挙動	11
竹居 正登 (横浜国立大学)	
Binomial catastrophe の発生を伴うマルコフ変調 $M^x/M/\infty$ 待ち行列に対する中心極限定理	19
矢島 萌子 (東京工業大学), フンドック トゥアン (筑波大学)	
Uplink coverage analysis for heterogeneous cellular networks	29
小林 拓矢, 三好 直人 (東京工業大学)	
セルラネットワークにおける移動車両の通信精度解析手法	38
徳山 喜一 (東京工業大学)	
$M/G/1+PH$ 待ち行列における系内客数分布の数値計算法	44
井上 文彰 (大阪大学)	
Distribution of the times to service completion and abandonment in the $M/M/m$ preemptive LCFS queue with impatient customers	54
高木 英明 (筑波大学)	
ビットコインのマイニング・モデルとトランザクション承認時間解析	64
河瀬 良亮, 笠原 正治 (奈良先端科学技術大学院大学)	
BGP ネットワークにおける障害規模推定方法の検討	74
後谷 浩輔, 小林 香 (富山県立大学)	
BGP メッセージを用いた迂回路検出手法の検討	81
山谷 周平, 小林 香 (富山県立大学)	
ランダム多スタート局所探索法に対する極値統計を用いた計算停止基準	84
仲川 諒馬, 増山 博之 (京都大学), 檀 寛成 (関西大学), 高橋 豊 (京都大学)	
既約で Lumpable なマルコフ連鎖の上界の改善	89
石川 真也, 河西 憲一 (群馬大学)	
レベル依存する $M/G/1$ 型マルコフ連鎖における条件付き定常分布の数値計算アルゴリズム	99
木村 雅俊, 滝根 哲哉 (大阪大学)	
Geometric convergence of finite-level $M/G/1$ -type Markov chains	109
勝又 洋介, 増山 博之, 高橋 豊 (京都大学)	
セルラネットワークの空間確率モデルに対するポアソン点過程近似について	119
高橋 雄太, 三好 直人 (東京工業大学)	
Maintenance optimization for multiple-unit systems based on Markov decision process	128
金 路 (電気通信大学)	
Monotone properties of an optimal maintenance policy for a Markovian deteriorating system with imperfect repair and incomplete information	134
田村 信幸 (法政大学)	
契約電力超過確率最小化に関する研究	140
小柳 淳二, 古林 雄貴, 河野 大地 (鳥取大学)	

Revisiting the heavy-traffic asymptotics of the GI/G/1-type Markov chain: a complete proof of the asymptotic formula for the stationary distribution	143
木村 達明 (NTT), 増山 博之, 高橋 豊 (京都大学)	
マルチエージェントモデルを用いたセルフレジ導入によるレジサービスへの影響と解析	149
平田 直也, 中桐 齊之 (兵庫県立大学)	
An arrival-time distribution for the equilibrium mean waiting time of a discrete-time single-server queue with acceptance period and Poissonian population of customers	157
佐久間 大 (防衛大学校), 小林 正弘 (東海大学), 増山 博之 (京都大学)	
A finite bottleneck game with homogeneous commuters	162
河崎 亮 (東京工業大学), 小西 秀男 (ボストンカレッジ), 湯川 隼貴 (東京工業大学)	
上界と平均の上界が与えられた確率過程の特徴づけ	168
高田 寛之 (長崎大学)	

正方分割表における 非対称性のモデルについて

東京理科大学 田畑耕治

2016年度待ち行列シンポジウム「確率モデルとその応用」@東京理科大学
2017年1月19日

1

はじめに

表1：英国人女性の左右裸眼視力データ（Stuart, 1955）

Right Eye Grade	Left Eye Grade			
	Best	Second	Third	Worst
Best	1520	266	124	66
Second	234	1512	432	78
Third	117	362	1772	205
Worst	36	82	179	492

1

・ 順序カテゴリ $r \times r$ 正方分割表

・ (i, j) セル確率: p_{ij} ($i = 1, \dots, r; j = 1, \dots, r$)

・ 対称 (S) モデル (Bowker, 1948)

$$p_{ij} = \psi_{ij} \quad (i = 1, \dots, r; j = 1, \dots, r)$$

ただし $\psi_{ij} = \psi_{ji}$

$\begin{smallmatrix} Y \\ X \end{smallmatrix}$	1	2	3	4
1	p_{11}	p_{12}	p_{13}	p_{14}
2	p_{21}	p_{22}	p_{23}	p_{24}
3	p_{31}	p_{32}	p_{33}	p_{34}
4	p_{41}	p_{42}	p_{43}	p_{44}

1

・ 周辺同等 (MH) モデル (Stuart, 1955)

$$p_{i\cdot} = p_{\cdot i} \quad (i = 1, \dots, r)$$

ただし

$$p_{i\cdot} = \sum_{t=1}^r p_{it}, \quad p_{\cdot i} = \sum_{s=1}^r p_{si}$$

$\begin{smallmatrix} Y \\ X \end{smallmatrix}$	1	2	3	4	計
1	p_{11}	p_{12}	p_{13}	p_{14}	$p_{1\cdot}$
2	p_{21}	p_{22}	p_{23}	p_{24}	$p_{2\cdot}$
3	p_{31}	p_{32}	p_{33}	p_{34}	$p_{3\cdot}$
4	p_{41}	p_{42}	p_{43}	p_{44}	$p_{4\cdot}$
計	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	$p_{\cdot 4}$	1

1

- ・ 準対称 (QS) モデル (Caussinus, 1965)

$$p_{ij} = \mu \alpha_i \beta_j \psi_{ij} \quad (i = 1, \dots, r; j = 1, \dots, r)$$

ただし $\psi_{ij} = \psi_{ji}$

- ・ 特に, $\{\alpha_i = \beta_i\}$ のとき S モデル
- ・ i 行と j 行, s 列と t 列におけるオッズ比 :

$$\theta_{(i < j; s < t)} = \frac{p_{is} p_{jt}}{p_{js} p_{it}}$$

1

- ・ QSモデルの別表現 :

$$\theta_{(i < j; s < t)} = \theta_{(s < t; i < j)} \quad (i < j; s < t)$$

$\begin{smallmatrix} Y \\ X \end{smallmatrix}$	1	2	3	4
1	p_{11}	p_{12}	p_{13}	p_{14}
2	p_{21}	p_{22}	p_{23}	p_{24}
3	p_{31}	p_{32}	p_{33}	p_{34}
4	p_{41}	p_{42}	p_{43}	p_{44}

$\theta_{(1 < 2; 3 < 4)}$

$\theta_{(3 < 4; 1 < 2)}$

Sモデルが成り立つための必要十分条件はQSモデルとMHモデルの両方が成り立つことである

(Caussinus, 1965)

1

- ・ 条件付き対称 (CS) モデル (McCullagh, 1978)

$$p_{ij} = \begin{cases} \gamma \psi_{ij} & (i < j) \\ \psi_{ij} & (i \geq j) \end{cases}$$

ただし $\psi_{ij} = \psi_{ji}$

- ・ 特に, $\gamma = 1$ のときSモデル

- ・ グローバル対称 (GS) モデル (Read, 1977)

$$\delta_U = \delta_L$$

ただし

$$\delta_U = \sum_{i=1}^{r-1} \sum_{j=i+1}^r p_{ij}, \quad \delta_L = \sum_{i=1}^{r-1} \sum_{j=i+1}^r p_{ji}$$

1

- ・ CSモデルとGSモデル

$\begin{smallmatrix} Y \\ \backslash X \end{smallmatrix}$	1	2	3	4	$\begin{smallmatrix} Y \\ \backslash X \end{smallmatrix}$	1	2	3	4
1	p_{11}	p_{12}	p_{13}	p_{14}	1	p_{11}	p_{12}	p_{13}	p_{14}
2	p_{21}	p_{22}	p_{23}	p_{24}	2	p_{21}	p_{22}	p_{23}	p_{24}
3	p_{31}	p_{32}	p_{33}	p_{34}	3	p_{31}	p_{32}	p_{33}	p_{34}
4	p_{41}	p_{42}	p_{43}	p_{44}	4	p_{41}	p_{42}	p_{43}	p_{44}

CSモデル

GSモデル

Sモデルが成り立つための必要十分条件は**CSモデル**と**GSモデル**の両方が成り立つことである

(Read, 1977)

- ・ 2比パラメータ対称 (TRPS) モデル

(Tomizawa, 1987)

$$p_{ij} = \gamma \theta^{j-i} p_{ji} \quad (i < j)$$

- ・ $\gamma = 1$ のとき, 線形対角パラメータ対称 (LDPS) モデル (Agresti, 1983)

- ・ $\theta = 1$ のとき, CSモデル

- ・ $\gamma = \theta = 1$ のとき, Sモデル

- ・ TRPSモデルの拡張モデルを提案し, 提案モデルを用いたSモデルの分解および適合度検定統計量の性質を調べる

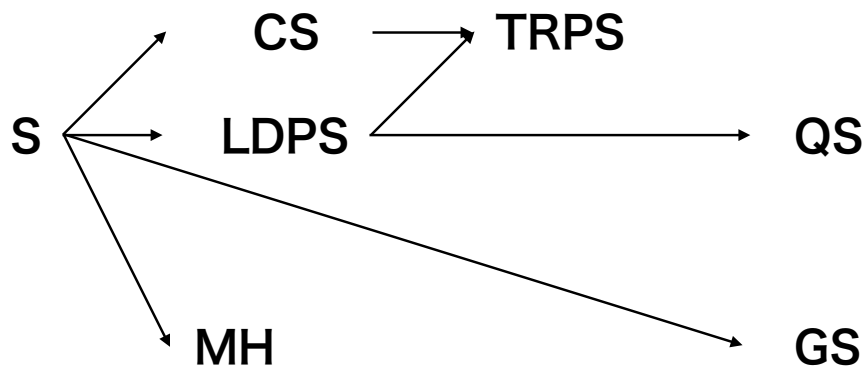


図1：モデルの包含関係

(注：A → B はモデルAはモデルBを示す)

3

提案モデル

・ 拡張k次線形非対称 (ELS_k) モデル

任意に与えられた k ($k = 1, \dots, r - 1$) に対して

$$p_{ij} = \mu \left(\prod_{l=1}^k \alpha_l^{i^l} \beta_l^{j^l} \right) \psi_{ij} \quad (i = 1, \dots, r; j = 1, \dots, r)$$

ただし $\psi_{ij} = \gamma \psi_{ji} \quad (i < j)$

・ $\gamma = 1$ のとき, k次線形非対称 (LS_k) モデル

(Tahata and Tomizawa, 2011)

3

・ LS_kモデルの特別な場合について

◆ LS₁モデルはLDPSモデル (Agresti, 1983)

◆ LS₂モデルは拡張線形対角パラメータ対称
(ELDPS) モデル (Tomizawa, 1991)

◆ LS_{r-1}モデルはQSモデル (Caussinus, 1965)

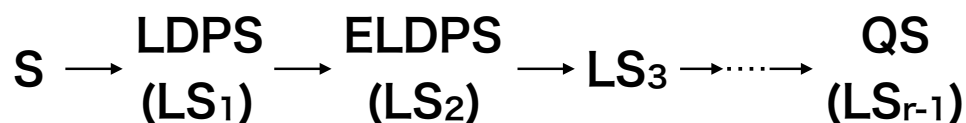


図2：モデルの包含関係

(注：A → B はモデルAはモデルBを示す)

3

- ・ **拡張準対称 (EQS) モデル** (Tomizawa, 1984)

$$p_{ij} = \mu \alpha_i \beta_j \psi_{ij} \quad (i = 1, \dots, r; j = 1, \dots, r)$$

ただし $\psi_{ij} = \gamma \psi_{ji} \quad (i < j)$

- ・ EQSモデルとELS_{r-1}モデルの関連について：

$$\frac{p_{ij}}{p_{ji}} = \gamma \frac{\lambda_j}{\lambda_i} \quad (i < j)$$

$$\frac{p_{ij}}{p_{ji}} = \gamma \left(\prod_{l=0}^{r-1} \frac{\theta_l^{j^l}}{\theta_l^{i^l}} \right) \quad (i < j)$$

EQSモデルとELS_{r-1}モデルは同値である

4

対称性の分解

- ・ 行変数を X ，列変数を Y とする

- ・ **周辺 k 次積率一致 (ME_k) モデル**

任意に与えられた k ($k = 1, \dots, r - 1$) に対して

$$E(X^l) = E(Y^l) \quad (l = 1, \dots, k)$$

ただし

$$E(X^l) = \sum_{i=1}^r i^l p_{i\cdot}, \quad E(Y^l) = \sum_{j=1}^r j^l p_{\cdot j}$$

- ・ ME_{r-1}モデルはMHモデル

4

定理1.

任意に与えられた k ($k = 1, \dots, r - 1$) に対して,
Sモデルが成り立つための必要十分条件は
ELS_kモデル, ME_kモデル, GSモデル
のすべてが成り立つことである

- ・ 定理1 は
Caussinus (1965)
Read (1977)
Tahata and Tomizawa (2011)
などの結果を含む

4

- ・ ME_kモデルとGSモデルの両方の構造をもつモデル
をGM_kモデルと記す. 特に, MHモデルとGSモデル
の両方の構造をもつモデルをGMHモデルと記す

系1.

任意に与えられた k ($k = 1, \dots, r - 1$) に対して,
Sモデルが成り立つための必要十分条件は**ELS_kモ
デルとGM_kモデル**の両方が成り立つことである

ELS_{r-1}モデルはEQSモデルであり, GM_{r-1}モデルは
GMHモデルであることに注意する

5

モデルの包含関係

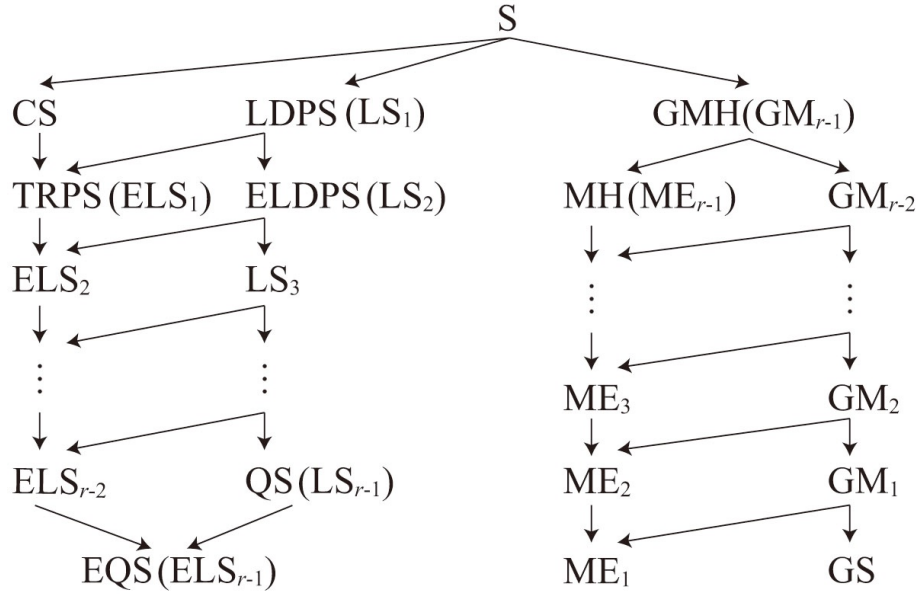


図3：モデルの包含関係

(注：A → B はモデルAはモデルBを示す)

6

検定統計量の性質

- ・ 尤度比カイ二乗統計量：

$$G^2(M) = 2 \sum_{i=1}^r \sum_{j=1}^r n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right)$$

n_{ij} ：(i, j)セル観測度数

$\hat{m}_{ij}$ ：モデルMの下での期待度数 m_{ij} の最尤推定量

定理 2.

任意に与えられた k ($k = 1, \dots, r - 1$) に対して,
 $G^2(S)$ は漸近的に $G^2(ELS_k)$ と $G^2(GM_k)$ の和に同
 等である

1. Agresti, A. (1983). A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics and Probability Letters*, 1, 313-316.
2. Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43, 572-574.
3. Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse*, Série 4, 29, 77-182.
4. McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65, 413-418.
5. Read, C. B. (1977). Partitioning chi-square in contingency tables: A teaching approach. *Communications in Statistics-Theory and Methods*, 6, 553-562.

6. Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42, 412-416.
7. Tahata, K. and Tomizawa, S. (2011). Generalized linear asymmetry model and decomposition of symmetry for multiway contingency tables. *Biometrics and Biostatistics*, 2, 1-6.
8. Tomizawa, S. (1984). Three kinds of decompositions for the conditional symmetry model in a square contingency table. *Journal of the Japan Statistical Society*, 14, 35-42.
9. Tomizawa, S. (1987). Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories. *Biometrical Journal*, 29, 45-55.
10. Tomizawa, S. (1991). An extended linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *Metron*, 49, 401-409.

1 次元線形セルオートマトンの極限挙動

竹居 正登
(横浜国立大学大学院工学研究院)

概要： Rule 90 に代表される 1 次元線形セルオートマトンの極限挙動に関する話題を紹介する.

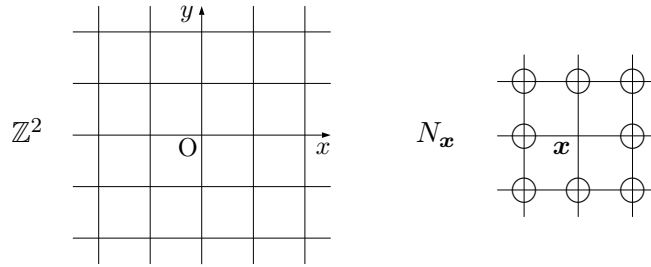
1. はじめに

1.1. Conway のライフゲーム

Conway による “Game of Life” は次のようなものである ([4] 参照) : 2 次元正方格子の格子点全体を $\mathbb{Z}^2 = \{\mathbf{x} = (x^1, x^2) : x^1, x^2 \text{ は整数}\}$ で表し, $\mathbb{Z}^2$ 上に 0 と 1 を配置したもの全体を

$$\Omega = \{0, 1\}^{\mathbb{Z}^2} = \{\omega = (\omega(\mathbf{x}) : \mathbf{x} \in \mathbb{Z}^2) : \omega(\mathbf{x}) \in \{0, 1\}\}$$

で表す. 点 $\mathbf{x} \in \mathbb{Z}^2$ の近傍を $N_{\mathbf{x}} = \{\mathbf{y} \in \mathbb{Z}^2 : \max\{|x^1 - y^1|, |x^2 - y^2|\} = 1\}$ とし, 配置 $\omega \in \{0, 1\}^{\mathbb{Z}^2}$ における点 $\mathbf{x} \in \mathbb{Z}^2$ の周囲の「人口」を $J(\omega, \mathbf{x}) = \sum_{\mathbf{y} \in N_{\mathbf{x}}} \omega(\mathbf{y})$ とする.



$\mathbb{Z}^2$ 上の 0, 1 の配置 $\omega \in \Omega$ に, 新しい配置 $\mathfrak{L}\omega \in \Omega$ を対応させる写像 $\mathfrak{L}$ を次のように定める :

$$(\mathfrak{L}\omega)(\mathbf{x}) = \begin{cases} 0 & (J(\omega, \mathbf{x}) \geq 4 \text{ のとき} \dots \text{「過密による死」}), \\ 1 & (J(\omega, \mathbf{x}) = 3 \text{ のとき} \dots \text{「誕生」}), \\ \omega(\mathbf{x}) & (J(\omega, \mathbf{x}) = 2 \text{ のとき} \dots \text{「生存」}), \\ 0 & (J(\omega, \mathbf{x}) \leq 1 \text{ のとき} \dots \text{「過疎による死」}). \end{cases}$$

初期配置 $\omega \in \Omega$ を適当に与えることによって, 時間発展 $\mathfrak{L}\omega, \mathfrak{L}^2\omega, \dots, \mathfrak{L}^n\omega, \dots$ の様子に様々な興味深い特徴が観察される. 初期配置として複雑なものを与えたとき (特に, ランダムに 0, 1 を配置するとき) はどうなるだろうか?

1.2. “Rule 90” — 1 次元ライフゲーム

初期配置 ω の分布を μ とするとき, $\mathfrak{L}\omega$ の分布を $\mathfrak{L}\mu$ で表す. しかし, ライフゲームの写像 $\mathfrak{L}$ は複雑なので, $\mathfrak{L}\mu = \mu$ を満たす定常分布 μ をみつけるといった問題を調べるのは難しいように思われる. そこで, $\mathfrak{L}$ の特徴をとらえて 1 次元に単純化した写像 Λ を考える ([10]; [15], Exercise (2.6) も参照) :

$$\Omega = \{0, 1\}^{\mathbb{Z}} := \{\omega = (\omega(x) : x \in \mathbb{Z}) : \omega(x) \in \{0, 1\}\}$$

とし, 配置空間 Ω 上の変換 Λ を

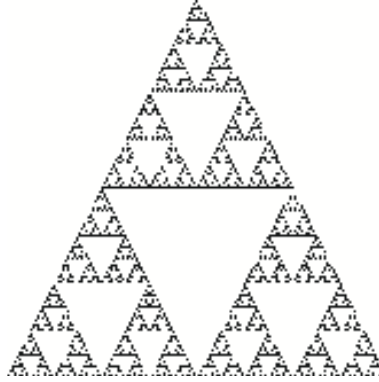
$$(\Lambda\omega)(x) = \omega(x-1) + \omega(x+1) \pmod{2} \quad [\omega \in \Omega, x \in \mathbb{Z}]$$

によって定める．後に Wolfram による 1 次元基本セルオートマトンの分類 [22] によって広く知られるようになり， Λ には rule 90 という番号が与えられ，数値計算により観察された結果のいくつかには数学的な証明がなされた ([6] など)．また，このような変換は 1960 年代からソ連のグループにより確率セルオートマトンの特別な場合としても研究されていた ([19, 20, 21] など；[17] も参照)．

任意の $m = 0, 1, 2, \dots$ に対して，

$$(\Lambda^{2^m} \omega)(x) = \omega(x - 2^m) + \omega(x + 2^m) \pmod{2} \quad [\omega \in \Omega, x \in \mathbb{Z}]$$

が成り立つことから， Λ^{2^m} は $2^m \mathbb{Z}$ の上の rule 90 と同型になる．原点だけが 1 で，他の点は 0 であるという初期配置から出発した時空図は，Pascal の三角形に現れる奇数だけを取り出したものになる：



上の図の n 段目で 1 になっている点の集合を $\Lambda^n\{0\}$ で表し，非負の整数 n の 2 進展開に現れる 1 の個数を $B(n)$ で表すとき， $\#(\Lambda^n\{0\}) = 2^{B(n)}$ であることが分かる．この時空図に適切なスケールを施すことで Sierpiński gasket と呼ばれる自己相似集合が得られる．そのため，時空図のパターンは pre-Sierpiński gasket と呼ばれる． Λ の線形性から，一般の初期配置から出発した場合 pre-Sierpiński gasket の重ね合わせが現れる．

初期配置 ω の分布を μ とするとき， $\Lambda\omega$ の分布を $\Lambda\mu$ で表す．どのような μ が Λ で不変となるだろうか？ また， $n \rightarrow \infty$ のとき $\Lambda^n\mu$ はどのように振る舞うだろうか？

初期配置の確率分布の中で最も基本的と考えられるのが密度 ρ の **Bernoulli 測度** β_ρ である： β_ρ のもとで $(\omega(x) : x \in \mathbb{Z})$ は各点独立で，任意の $x \in \mathbb{Z}$ に対して，

$$\omega(x) = 1 \text{ となる確率} = \rho, \quad \omega(x) = 0 \text{ となる確率} = 1 - \rho$$

とする． β_0 は「全ての点の状態が 0」である配置に集中したデルタ測度であるが，これは Λ で不変である．また，公平なコイン投げの確率分布に相当し「最もランダムな状態」を表すものと考えられる一様測度 $\beta_{1/2}$ も Λ で不変であることが分かる．

任意の $\omega \in \Omega$ と任意の非負の整数 n に対して， $(\Lambda^n \omega)(0) = \sum_{x \in \Lambda^n\{0\}} \omega(x)$ および $\#(\Lambda^n\{0\}) = 2^{B(n)}$ が成り立つことに注意すると， $0 < \rho < 1$ のとき

$$\beta_\rho((\Lambda^{2^m} \omega)(x) = 1) = 2\rho(1 - \rho), \quad \lim_{m \rightarrow \infty} \beta_\rho((\Lambda^{2^m-1} \omega)(x) = 1) = \frac{1}{2}$$

となることが分かる ([3], 5d 節を参照)．Miyamoto [10] と Lind [6] は次のことを証明した：

- $\lim_{n \rightarrow \infty} \Lambda^n \beta_\rho$ が存在する $\iff \rho \in \{0, 1/2, 1\}$ ．従って， β_ρ が Λ -不変 $\iff \rho \in \{0, 1/2\}$ ．
- $0 < \rho < 1$ ならば，Cesàro 平均 $\frac{1}{N} \sum_{n=0}^{N-1} \Lambda^n \beta_\rho$ は $N \rightarrow \infty$ のとき $\beta_{1/2}$ に収束する．

この結果は次のことを示している：

- 自明でない Bernoulli 測度の中では，一様測度 $\beta_{1/2}$ が唯一つの Λ -不変測度である．
- 自明でない Bernoulli 測度から出発したとき，そのままでは収束しなくても Cesàro 平均は一様測度に収束する．

前者の性質は“rigidity”と呼ばれる： Λ に対する周期点から生成されるような「退化した」確率測度を除外する何らかの条件を課すと、一様測度は唯一つの Λ -不変測度である（この種の性質に関する議論は既に Vasil'ev and Pyatetskii-Shapiro [21] に見られる）。最近のサーヴェイとして Maass [7] がある。後者の性質は“asymptotic randomization”と呼ばれる。初期測度のクラスを広げる研究は種々なされてきたが、Pivato and Yassawi [13] は“harmonically mixing measures”の概念を導入し、このクラスの測度から出発したとき、非自明な線形セルオートマトンによって Cesàro 平均が一様測度に収束することを極めて一般の設定で証明している。

本稿では、Bernoulli 測度を含むより広いクラスの初期分布から出発した場合の極限分布を調べる（2 節）。また、 Λ を含むより広いクラスの 1 次元線形セルオートマトンに加法的なエラーを付加した場合の極限挙動についても調べる（3 節）。

2. Addition modulo p の極限分布

本節では [16] で得られた結果をさらに改良したものを紹介する。

Rule 90 を次のように一般化する： p を素数とし、 $\mathcal{A} = \{0, 1, \dots, p-1\}$ とおく。配置空間

$$\Omega = \mathcal{A}^{\mathbb{Z}} := \{\omega = (\omega(x) : x \in \mathbb{Z}) : \omega(x) \in \mathcal{A}\}$$

の上の変換 Λ を

$$(\Lambda\omega)(x) = \omega(x-1) + \omega(x+1) \pmod{p} \quad [\omega \in \Omega, x \in \mathbb{Z}]$$

で定め、addition modulo p と呼ぶ。

$\theta = (\theta_0, \theta_1, \dots, \theta_{p-1})$ を $\mathcal{A}$ 上の確率分布とする：すなわち、 $\theta_k \geq 0$ [$k \in \mathcal{A}$], $\sum_{k \in \mathcal{A}} \theta_k = 1$ 。確率変数列 $(\omega(x) : x \in \mathbb{Z})$ が各点独立で、任意の $x \in \mathbb{Z}$ に対して $\omega(x) = k$ となる確率 $= \theta_k$ [$k \in \mathcal{A}$] を満たすとき、その分布を μ_θ で表し、密度 θ の直積測度と呼ぶ。特に、全ての $k \in \mathcal{A}$ で $\theta_k = \frac{1}{p}$ であるときの μ_θ を $\mu_{1/p}$ で表し、一様測度と呼ぶ。また、 $a \in \mathcal{A}$ に対して、「全ての点の状態が a 」である配置を $\mathbf{a}$ で表し、 $\mathbf{a}$ に集中したデルタ測度を $\delta_{\mathbf{a}}$ で表す。

Cai and Luo [1] は Miyamoto [10], Lind [6] の結果を奇素数 p の場合に拡張した：

- $\lim_{n \rightarrow \infty} \Lambda^n \mu_\theta$ が存在する $\iff \mu_\theta = \delta_0$ 又は $\mu_\theta = \mu_{1/p}$ 。従って、 Λ -不変な直積測度はこれら 2 つに限られる。
- 全ての $k \in \mathcal{A}$ に対して $\theta_k < 1$ ならば、Cesàro 平均 $\frac{1}{N} \sum_{n=0}^{N-1} \Lambda^n \mu_\theta$ は $N \rightarrow \infty$ のとき一様測度 $\mu_{1/p}$ に収束する。

本節では、より広いクラスの μ に対する $\lim_{n \rightarrow \infty} \Lambda^n \mu$ の存在について調べる。

2.1. Shift-不変な確率測度

自然数 L に対して、 $\Omega_L = \mathcal{A}^L := \{\sigma = (\sigma_1, \sigma_2, \dots, \sigma_L) : \sigma_1, \sigma_2, \dots, \sigma_L \in \mathcal{A}\}$ と定める。 $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_L) \in \Omega_L$ と $a \in \mathbb{Z}$ に対して

$$[\sigma]_{a+1}^{a+L} = [\sigma_1 \sigma_2 \cdots \sigma_L]_{a+1}^{a+L} := \{\omega \in \Omega : \omega(a+x) = \sigma_x (x = 1, 2, \dots, L)\}.$$

とおく。この形の Ω の部分集合をシリンダー集合という。シリンダー集合から生成される σ -加法族を $\mathcal{B}$ で表す。 Ω に離散位相の直積位相を入れて考えたときの Borel 集合族は $\mathcal{B}$ に一致する。今後は $(\Omega, \mathcal{B})$ の上の確率測度について考える：これらは Ω の上の Borel 確率測度と呼ばれ、シリンダー集合の確率を決めると測度が一意的に決まる。例えば、密度 $\theta = (\theta_0, \theta_1, \dots, \theta_{p-1})$ の直積測度 μ_θ は

$$\mu_\theta([\sigma]_{a+1}^{a+L}) = \prod_{x=1}^L \theta_{\sigma_x} \quad [a \in \mathbb{Z}, L = 1, 2, \dots; \sigma \in \Omega_L] \quad (*)$$

によって特徴付けられる。確率測度 μ_1, μ_2 について、任意のシリンダー集合 A に対して $\mu_1(A) = \mu_2(A)$ が成り立つとき $\mu_1 = \mu_2$ と表す。確率測度の列 $\{\mu_n\}$ と確率測度 μ について、任意のシリンダー集合 A に対して $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ が成り立つとき $\lim_{n \rightarrow \infty} \mu_n = \mu$ と表す。

(*) により, 確率 $\mu_\theta([\sigma]_{a+1}^{a+L})$ は a によらない: この性質は μ_θ の shift-不変性と呼ばれる. 一般の定義を述べよう. Ω の元の左への shift 変換 θ を

$$(\theta\omega)(x) = \omega(x+1) \quad [\omega \in \Omega, x \in \mathbb{Z}]$$

によって定める. θ^{-1} は右への shift 変換である. n 歩の shift 変換を θ^n, θ^{-n} で表す. 事象 $A \in \mathcal{B}$ に対して,

$$\theta^{-n}A := \{\omega \in \Omega : \theta^n\omega \in A\} = \{\theta^{-n}\omega : \omega \in A\}$$

とする. 例えば, $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_L) \in \Omega_L$ に対して,

$$\begin{aligned} \theta^{-n}[\sigma]_1^L &= \{\omega \in \Omega : (\theta^n\omega)(x) = \sigma_x (x = 1, 2, \dots, L)\} \\ &= \{\omega \in \Omega : \omega(x+n) = \sigma_x (x = 1, 2, \dots, L)\} = [\sigma]_{1+n}^{L+n}. \end{aligned}$$

Ω 上の確率測度 μ が shift-不変であるとは, 任意の $A \in \mathcal{B}$ に対して $\mu(\theta^{-1}A) = \mu(A)$ が成り立つときにいう. これは次のことと同値である: 任意の L, n と $\sigma \in \Omega_L$ に対して $\mu([\sigma]_{1+n}^{L+n}) = \mu([\sigma]_1^L)$. この値を $\mu(\sigma)$ と書くことがある.

2.2. Shift-不変な確率測度の混合性

$a, b \in \mathbb{Z}$ とし, L, L' を自然数とすると, $[a+1, a+L] \cap [b+1, b+L'] = \emptyset$ ならば, 任意の $\sigma \in \Omega_L$ と $\sigma' \in \Omega_{L'}$ に対して

$$\mu_\theta([\sigma]_{a+1}^{a+L} \cap [\sigma']_{b+1}^{b+L'}) = \mu_\theta([\sigma]_{a+1}^{a+L})\mu_\theta([\sigma']_{b+1}^{b+L'})$$

が成り立つ: これを μ_θ におけるペアごとの独立性という. 次のさらに強い性質を独立性という: 依存する座標に重なりのない有限個のシリンダー集合の共通部分の μ_θ -確率は, 各々のシリンダー集合の μ_θ -確率の積で与えられる.

μ を shift-不変な確率測度とすると, shift 写像に関する漸近的な独立性は混合性と呼ばれている (例えば Smorodinsky [14] や 十時 [18] を参照). Shift-不変な μ が混合的 ((strong) mixing) であるとは, 任意の $A, B \in \mathcal{B}$ に対して

$$\lim_{n \rightarrow \infty} \mu(A \cap \theta^{-n}B) = \mu(A)\mu(B)$$

が成り立つことをいう. これは次のことと同値である: 任意の自然数 L, L' と $\sigma \in \Omega_L, \sigma' \in \Omega_{L'}$ に対して, $n \rightarrow \infty$ のとき

$$\mu([\sigma]_1^L \cap [\sigma']_{1+n}^{L'+n}) \rightarrow \mu([\sigma]_1^L) \mu([\sigma']_1^{L'}).$$

Shift-不変な μ がエルゴード的 (ergodic) であるとは, $\theta^{-1}A = A$ を満たす事象 $A \in \mathcal{B}$ に対しては $\mu(A) = 0$ 又は 1 が成り立つときにいう. 混合的ならばエルゴード的である.

r を自然数とする. Ω 上の shift-不変な確率測度 μ が r 次の混合性をもつ (r -fold mixing) とは, 任意の $A, B_1, \dots, B_r \in \mathcal{B}$ に対して

$$n_1 \rightarrow \infty, \quad n_2 - n_1 \rightarrow \infty, \quad \dots, \quad n_r - n_{r-1} \rightarrow \infty$$

のとき

$$\mu(A \cap \theta^{-n_1}B_1 \cap \dots \cap \theta^{-n_r}B_r) \rightarrow \mu(A)\mu(B_1) \dots \mu(B_r)$$

が成り立つことをいう. 特に, 1 次の混合性をもつときが strong mixing である. r 次の混合性をもつ Ω 上の確率測度の全体を $\mathcal{M}_r$ で表すとき,

$$\mathcal{M}_1 \supset \mathcal{M}_2 \supset \dots \supset \mathcal{M}_r \supset \dots$$

が成り立つ.

$\mathcal{G}_n$ を, $\{\omega(i) : i \geq n\}$ によって生成される σ -加法族とする. Shift-不変な確率測度 μ が K-mixing であるとは, 任意の $A \in \mathcal{B}$ に対して,

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{G}_n} |\mu(A \cap B) - \mu(A)\mu(B)| = 0$$

が成り立つことをいう.

$$\mathcal{G}_\infty := \bigcap_{n=1}^{\infty} \mathcal{G}_n$$

と置くとき, μ が K-mixing であることと $\mathcal{G}_\infty$ が μ に関して trivial であることは同値であることが知られている. また, K-mixing である Ω 上の確率測度の全体を $\mathcal{M}$ で表すとき,

$$\mathcal{M} \subset \bigcap_{r=1}^{\infty} \mathcal{M}_r$$

が成り立つことが分かっている. Kolmogorov の 0-1 法則により直積測度 μ_θ は K-mixing であり, 上記の混合性を全てもつことになる.

2.3. 主結果

Ω 上の確率測度の族 $\mathcal{P}$ に対して, $\mathcal{P}$ の凸包 (convex hull) を

$$\text{Conv}(\mathcal{P}) := \left\{ \int_{\mathcal{P}} \mu d\pi(\mu) : \pi \text{ は } \mathcal{P} \text{ 上の確率測度} \right\}$$

で定義する.

定理 2.1. $p = 2$ とする. $P \in \text{Conv}(\mathcal{M}_1)$ とする. $\lim_{n \rightarrow \infty} \Lambda^n P$ が存在するための必要十分条件は,

$$P = \alpha\beta_0 + \alpha'\beta_{1/2} + \alpha''\beta_1 \quad [\alpha, \alpha', \alpha'' \geq 0, \alpha + \alpha' + \alpha'' = 1]$$

と表されることである. 従って, $P \in \text{Conv}(\mathcal{M}_1)$ が定常分布であるための必要十分条件は,

$$P = \alpha\beta_0 + \alpha'\beta_{1/2} \quad [\alpha, \alpha' \geq 0, \alpha + \alpha' = 1]$$

と表されることである.

この方向の結果として, Miyamoto の論文 ([10], Theorem 3 及び [11], Theorem 1) では, 定理 2.1 で mixing な確率測度の集合 $\mathcal{M}_1$ を K-mixing な確率測度の集合 $\mathcal{M}$ で置き換えた主張がなされている ([10, 11] の証明法をよく見ると, $\mathcal{M}_1$ を $\mathcal{M}_3$ で置き換えた形で証明できている).

定理 2.2. p を奇素数とする. $\mu \in \mathcal{M}_1$ のとき, $\lim_{n \rightarrow \infty} \Lambda^n \mu$ が存在するための必要十分条件は $\mu = \delta_0$ 又は $\mu = \mu_{1/p}$ となることである. 特に, $\mathcal{M}_1$ に属する Λ -不変測度はこの 2 つしかない.

定理 2.2 は Marcovici による結果 ([9], Proposition 5.5 及び [2], Proposition 3.2.2) の改良である. また, Pivato [12] もある種の混合性条件のもとでの不変測度について, 極めて一般的な設定のもとで論じている. p が奇素数で $P \in \text{Conv}(\mathcal{M}_1)$ の場合に $\lim_{n \rightarrow \infty} \Lambda^n P$ が存在するかという問題については,

$p = 2$ に比べて状況が複雑であることが次の例から示唆される: $p = 3$ のとき, $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2 \in \text{Conv}(\mathcal{M}_1)$ も Λ -不変である.

3. 加法的なエラーのある 1 次元線形基本セルオートマトンのエルゴード性

本節の結果は上妻遼太氏との共同研究 ([5]) に基づく.

Rule 90 の写像 Λ に「加法的なエラー」を付加する: $n = 1, 2, \dots$ と $x \in \mathbb{Z}$ に対し, $Y^{(n)}(x)$ は互いに独立で

$$P(Y^{(n)}(x) = 1) = \alpha^{(n)}(x), \quad P(Y^{(n)}(x) = 0) = 1 - \alpha^{(n)}(x)$$

を満たすとする. $\alpha^{(n)}(x)$ は時刻 n に点 $x \in \mathbb{Z}$ でエラーの生じる確率である.

Ω に値をとる確率変数の列 $\{X^{(n)} : n = 0, 1, 2, \dots\}$ を次のように定義する：初期配置 $X^{(0)}$ はある確率測度 μ_0 に従って選ばれるものとする。 $n = 1, 2, \dots$ に対して、 $Y^{(n)} = (Y^{(n)}(x) : x \in \mathbb{Z})$ を Ω に値をとる確率変数とみなし、

$$X^{(n)} = \Lambda X^{(n-1)} + Y^{(n)}$$

とする：例えば、

$$\begin{array}{r} X^{(n-1)} = \dots 0010100100 \dots \\ \Lambda X^{(n-1)} = \dots 0100011010 \dots \\ +) \quad Y^{(n)} = \dots \mathbf{1001000010} \dots \\ \hline X^{(n)} = \dots \mathbf{1101011000} \dots \end{array}$$

$X^{(n)}$ の分布を μ_n で表す。 $\{X^{(n)}\}$ が ergodic であるとは、ある Ω 上の確率測度 μ_∞ が存在し、

$$\text{任意の初期分布 } \mu_0 \text{ に対して、 } \lim_{n \rightarrow \infty} \mu_n = \mu_\infty$$

が成り立つときにいう。 $\mu_0 = \beta_{1/2}$ のとき、任意の $n \geq 0$ に対して $\mu_n = \beta_{1/2}$ となるから、 $\{X^{(n)}\}$ が ergodic であるとき $\mu_\infty = \beta_{1/2}$ となる。

エラーの確率が一樣な場合、すなわち全ての n と $x \in \mathbb{Z}$ で $\alpha^{(n)}(x) \equiv \alpha$ であるとき、 $\{X^{(n)}\}$ が ergodic であるための必要十分条件は $0 < \alpha < 1$ である。(この結果は、例えば Vasershtein [20] の Example 3 に見られる。) エラーの確率が時間的に非一樣な場合、すなわち全ての $x \in \mathbb{Z}$ で $\alpha^{(n)}(x) = \alpha^{(n)}$ であるときについて、Miyamoto [10] は $\{X^{(n)}\}$ が ergodic となるための必要十分条件を与えており、 Λ を取り除いた $\widetilde{X^{(n)}} = \widetilde{X^{(n-1)}} + Y^{(n)}$ が ergodic となるための必要十分条件よりもずっと緩やかであることを示した(3.2節を参照)。これは Λ が「コイン投げに近づけるはたらきを有する」ことを示唆している。

確率セルオートマトンの最近のサーヴェイに Mairesse and Marcovici [8] がある。

本節では、変換 Λ を他の 1 次元線形セルオートマトンに置き換えたときの ergodicity の必要十分条件を調べよう。

3.1. 加法的なエラーのある 1 次元線形基本セルオートマトン

$\Omega = \{0, 1\}^{\mathbb{Z}}$ の変換 L について、ある $f_L : \{0, 1\}^3 \rightarrow \{0, 1\}$ が存在して

$$(L\omega)(x) = f_L(\omega(x-1), \omega(x), \omega(x+1)) \quad [x \in \mathbb{Z}]$$

と表されるとき、 L を 1 次元基本セルオートマトン (elementary cellular automaton: ECA) と呼ぶ。Wolfram [22] による命名法がある： $(s, t, u) \in \{0, 1\}^3$ を 2 進数とみて $4s + 2t + u$ に対応させ、 $f_L(s, t, u) = a_{4s+2t+u}$ とおく。

$$\begin{array}{cccccccccc} stu & 111 & 110 & 101 & 100 & 011 & 010 & 001 & 000 \\ f_L(s, t, u) & a_7 & a_6 & a_5 & a_4 & a_3 & a_2 & a_1 & a_0 \end{array}$$

さらに、 $(a_7, a_6, \dots, a_1, a_0) \in \{0, 1\}^8$ を 2 進数とみて $N = \sum_{k=0}^7 2^k a_k$ に対応させ、1 次元基本セルオートマトン L を rule N と呼ぶ。

L が 1 次元線形基本セルオートマトン (linear ECA) であるとは、

$$f_L(s, t, u) = C_s s + C_t t + C_u u \quad [C_s, C_t, C_u \in \{0, 1\}]$$

と表されるときにいう (今後、加法はつねに mod 2 で考えるものとする)。Linear ECA L に対して、

$$f_{L^*}(s, t, u) = C_u s + C_t t + C_s u$$

で定められる L^* を dual linear ECA と呼ぶ。

係数 C_s, C_t, C_u のうち 2 つ以上が 1 である linear ECA を non-trivial であるという：

- rule 102 (Ledrappier CA と呼ばれる) $\mathcal{L}$: $f_{\mathcal{L}}(s, t, u) = t + u$. $\mathcal{L}^*$ は rule 60: $f_{\mathcal{L}^*}(s, t, u) = s + t$.
- rule 90 (1次元ライフゲームとも見られる) Λ : $f_{\Lambda}(s, t, u) = s + u$. この場合 $\Lambda^* = \Lambda$ である.
- rule 150 $\tilde{\Lambda}$: $f_{\tilde{\Lambda}}(s, t, u) = s + t + u$. やはり $(\tilde{\Lambda})^* = \tilde{\Lambda}$ である.

その他の linear ECA を trivial という (完全に trivial な rule 0: $f_L(s, t, u) \equiv 0$ は除外しよう) :

- 恒等写像 I は rule 204 に相当する : $f_I(s, t, u) = t$.
- rule 170 (左への shift 写像): $f_{\theta}(s, t, u) = u$. θ^* は rule 240 (右への shift 写像): $f_{\theta^*}(s, t, u) = s$. 互いに独立で

$$P(Y^{(n)}(x) = 1) = \alpha^{(n)}(x), \quad P(Y^{(n)}(x) = 0) = 1 - \alpha^{(n)}(x)$$

である $\{Y^{(n)}(x) : n = 1, 2, \dots; x \in \mathbb{Z}\}$ を考える. ECA L と初期配置 $X^{(0)}$ が与えられたとき, Ω -値確率変数列 $\{X^{(n)} : n = 1, 2, \dots\}$ を

$$X^{(n)} := LX^{(n-1)} + Y^{(n)} \quad [n = 1, 2, \dots]. \quad (1)$$

によって定める. この $\{X^{(n)}\}$ を, エラー確率 $\alpha = \{\alpha^{(n)}(x) : n = 1, 2, \dots; x \in \mathbb{Z}\}$ の ECA L と呼ぶ. 初期配置 $X^{(0)}$ はある確率測度 μ_0 に従って選ばれるとし, $X^{(n)}$ の分布を μ_n で表す. L が linear ECA であるとき, $\mu_0 = \beta_{1/2}$ ならば任意の $n \geq 0$ に対して $\mu_n = \beta_{1/2}$ となるから, $\{X^{(n)}\}$ が ergodic であるとき $\mu_{\infty} = \beta_{1/2}$ でなければならない.

3.2. 主結果

実数 $\alpha \in [0, 1]$ に対して, $\hat{\alpha} := \min\{\alpha, 1 - \alpha\}$ とおく : $\hat{\alpha} > 0$ と $0 < \alpha < 1$ は同値である.

まず, 空間的に一様で時間的に非一様なエラーのある場合を考える :

$$\alpha^{(n)}(x) = \alpha^{(n)} \quad [n = 0, 1, 2, \dots; x \in \mathbb{Z}].$$

定理 3.1. 時間的に非一様なエラーのある trivial な linear ECA が ergodic となるための必要十分条件は,

$$\text{ある } n \in \{1, 2, \dots\} \text{ で } \alpha^{(n)} = 1/2, \quad \text{又は} \quad \sum_{n=1}^{\infty} \widehat{\alpha^{(n)}} = +\infty.$$

Rule 204 に対して定理 3.1 が成り立つことは [10], p.526 において (証明無しで) 指摘されている.

定理 3.2. 時間的に非一様なエラーのある non-trivial な linear ECA が ergodic となるための必要十分条件は,

$$\text{ある } n \in \{1, 2, \dots\} \text{ で } \alpha^{(n)} = 1/2, \quad \text{又は} \quad \text{無限個の } n \text{ で } \widehat{\alpha^{(n)}} > 0.$$

Rule 90 に対して定理 3.2 が成り立つことは [10] で示された.

次に, 時間的に一様で空間的に非一様なエラーのある場合を考える :

$$\alpha^{(n)}(x) = \alpha(x) \quad [n = 0, 1, 2, \dots; x \in \mathbb{Z}].$$

定理 3.3. 空間的に非一様なエラーのある linear ECAs が ergodic となるための必要十分条件は次の通りである.

- rule 204 ($L = I$): 全ての $x \in \mathbb{Z}$ で $\widehat{\alpha(x)} > 0$.
- rule 170 ($L = \theta$): $\sum_{x=0}^{\infty} \widehat{\alpha(x)} = +\infty$.
- rule 102 ($L = \mathcal{L}$): 無限個の $x \geq 0$ で $\widehat{\alpha(x)} > 0$.
- rule 90 or 150 ($L = \Lambda$ or $\tilde{\Lambda}$): $x_0 < x_1$ を満たすある $x_0, x_1 \in \mathbb{Z}$ に対して $\widehat{\alpha(x_0)}, \widehat{\alpha(x_1)} > 0$.

Rule 170 に対しては, 定理 3.3 よりも一般の結果 ($0 \rightarrow 1$ というエラーの確率と $1 \rightarrow 0$ というエラーの確率が異なることを許す) が Vasershtein [20], Example 1 で述べられている. θ^{-1} , $\mathcal{L}^*$ については, それぞれ (ii) と (iii) の左右を入れ替えた条件が ergodicity の必要十分条件となる.

参考文献

- [1] Cai, H. and Luo, X., Laws of large numbers for a cellular automaton, *Ann. Probab.*, **21**, (1993), 1413–1426.
- [2] de Menibus, B. H., Asymptotic behaviour of cellular automata: Computation and randomness, PhD thesis, (2014), Aix-Marseille University.
- [3] Durrett, R., Lecture notes on particle systems and percolation, (1988), Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- [4] Gardner, M., Mathematical games — The fantastic combinations of John Conway’s new solitaire game “life”, *Scientific American*, October 1, (1970), 120–123.
- [5] Kouduma, R. and Takei, M., On the ergodicity of one-dimensional linear cellular automata with additive error, *Proceedings of 4th International Workshop on Applications and Fundamentals of Cellular Automata (AFCA’16), held in conjunction with CANDAR’16*, (2016), 222–228.
- [6] Lind, D. A., Applications of ergodic theory and sofic systems to cellular automata, *Phys. D*, **10**, (1984), 36–44.
- [7] Maass, A., Rigidity results in cellular automata theory: Probabilistic and ergodic theory approach, (2015), available at <https://cantorsalta2015.sciencesconf.org/>
- [8] Mairesse, J. and Marcovici, I., Around probabilistic cellular automata, *Theoret. Comput. Sci.*, **559**, (2014), 42–72.
- [9] Marcovici, I., Automates cellulaires probabilistes et mesures spécifiques sur des espaces symboliques, PhD thesis, (2013), Université Paris 7.
- [10] Miyamoto, M., An equilibrium state for a one-dimensional life game, *J. Math. Kyoto Univ.*, **19**, (1979), 525–540.
- [11] Miyamoto, M., Stationary measures for automaton rules 90 and 150, *J. Math. Kyoto Univ.*, **34**, (1994), 531–538.
- [12] Pivato, M., Module shifts and measure rigidity in linear cellular automata, *Ergodic Theory Dynam. Systems*, **28**, (2008), 1945–1958..
- [13] Pivato, M. and Yassawi, R., Limit measures for affine cellular automata, *Ergodic Theory Dynam. Systems*, **22**, (2002), 1269–1287.
- [14] Smorodinsky, M., Ergodic theory, entropy, *Lecture Notes in Math.*, **214**, (1971), Springer.
- [15] Spitzer, F., Random fields and interacting particle systems, (1971), Math. Assoc. America.
- [16] Takei, M., On limiting measures for a class of one-dimensional linear cellular automata, *Proceedings of 4th International Workshop on Applications and Fundamentals of Cellular Automata (AFCA’16), held in conjunction with CANDAR’16*, (2016), 236–242.
- [17] Toom, A. L., Vasilyev, N. B., Stavskaya, O. N., Mityushin, L. G., Kurdyumov, G. L., and Pirogov, S. A., Discrete local Markov systems, *Stochastic cellular systems: ergodicity, memory, morphogenesis (Dobrushin, R. L., Kryukov, V. I. and Toom, A. L. Eds.)*, (1990), 1–182, Manchester University Press.
- [18] 十時東生, エルゴード理論入門, 共立講座現代の数学, **30**, (1971), 共立出版.
- [19] Vallander, S. S., One system of automata with local interactions, *Multicomponent random systems, Adv. Probab. Related Topics*, **6**, (1980), 577–587, Dekker.
- [20] Vasershtein, L. N., Markov processes over denumerable products of spaces describing large system of automata, *Probl. Inf. Transm.*, **5**, (1969), 47–52.
- [21] Vasil’ev, N. B. and Pyatetskii-Shapiro, I. I., The classification of uniform homogeneous networks, *Probl. Inf. Transm.*, **7**, (1971), 340–346.
- [22] Wolfram, S., Statistical mechanics of cellular automata, *Rev. Modern Phys.*, **55**, (1983), 601–644.

Binomial catastrophe の発生を伴うマルコフ変調 $M^X/M/\infty$ 待ち行列 に対する中心極限定理

矢島 萌子	フンドック トゥアン
東京工業大学	筑波大学
大学院情報理工学研究科	システム情報系
数理・計算科学専攻	社会工学域

概要： 集団到着のあるマルコフ変調無限サーバ待ち行列のうち、binomial catastrophe の発生を伴うモデルを考察する。無限個のサーバを有するこの待ち行列システムでは、到着率、集団の大きさの分布、サービス率および binomial catastrophe に関する各パラメータが背後過程と呼ばれる有限状態空間上の連続時間マルコフ連鎖に支配される。マルコフ変調待ち行列の定常解析は非常に困難であるため、本稿では元のモデルをスケールリングしたモデルに着目する。そしてスケールリングモデルの定常系内容数に対し、中心極限定理が成立することを示す。

1 はじめに

無限サーバ待ち行列は、非常に大きいサイズの有限サーバ待ち行列の近似に用いられる他、生産システムや通信システムを始めとする多くの応用領域をもつため、研究が非常に盛んである。さらに近年では、catastrophe の発生を伴う無限サーバ待ち行列への拡張が注目されている [6, 8, 9]。Catastrophe はそのメカニズムの違いによりいくつか種類に分けられ、total catastrophe はその中で最も単純なメカニズムを持つ。

Total catastrophe を含む catastrophe のメカニズムはもともと、人口の振る舞いを出生死滅過程を用いてモデル化する際に取り入れられた [5, 11]。Total catastrophe とは人口がゼロになるイベントであり、それがある確率過程に従って発生する。人口は突如として発生する災害や食糧不足により急激に減少し得るため、人口の振る舞いのモデル化に、total catastrophe の発生を伴う出生死滅過程はよく利用される。Total catastrophe が発生すると人口はゼロになってしまうが、現実には人口の一部が死滅する災害なども多く存在する。そのような状況を適切にモデル化するため、しばしば random catastrophe と呼ばれるものが用いられる。Random catastrophe とは、人口が確率的な数だけ減少するイベントであり、本稿で扱う **binomial catastrophe** は random catastrophe の一種である [1, 4, 7]。人口モデルにおいて binomial catastrophe がある確率過程に従って発生すると、そのとき存在する人は互いに独立に確率 $p \in (0, 1)$ で生き残る。つまり人口の総数が n のとき binomial catastrophe が発生したならば、その直後の人口の総数の分布はパラメータ (n, p) の二項分布に従う。

我々が待ち行列を用いてシステムの振る舞いを表現する際、何かしらか深刻な技術的障害が発生して客を全てまたは一部失うといった状況を表現したいことがある。このような技術的障害、例えばパソコンのウイルス感染の発生などは人口モデルにおける catastrophe と見なせるため、近年では人口モデルの分野で研究された手法を用い、catastrophe の発生を伴う待ち行列が研究されている [6, 8, 9]。

本稿では binomial catastrophe の発生を伴う無限サーバ待ち行列のうち、各パラメータが有限状態空間 $\mathbb{D}$ 上の連続時間マルコフ連鎖 $\{J(t); t \geq 0\}$ に支配されるものを考える。 $\{J(t)\}$ の無限小生成作用素を $\mathbf{Q} := (q_{\ell,j})_{\ell,j \in \mathbb{D}}$ と記し、これ以降 $\{J(t)\}$ を背後過程と呼ぶ。 $J(t) = j$ のとき、大きさが確率変数 X_j に従う集団が率 λ_j のポアソン過程に従って到着し、そのときシステム内に滞在している客の残余サービス時間は互いに独立なパラメータ μ_j の指数分布に従う。また $J(t) = j$ のとき率 γ_j のポアソン過程に従って binomial catastrophe が発生し、そのときシステムに滞在している客は互いに独立に確率 p_j で系内に滞在したままである。本稿ではこのような待ち行列を、**binomial catastrophe の発生を伴うマルコフ変調 $M^X/M/\infty$ 待ち行列**と呼ぶことにする。

一般的に待ち行列の到着率やサービス率などのパラメータは常に一定であると仮定されるが、この仮定は現実的とはいえない。一方、本稿で扱うようなマルコフ変調待ち行列は、背後過程に依存する可変なパラメータを持つことのできる待ち行列であるため、実用上とても自然であり、多くの応用例が考えられる。実際にマルコフ変調待ち行列の研究は、近年活発に行なわれている [2, 10, 12]。

マルコフ変調待ち行列は、たとえ集団到着や catastrophe の発生を伴わないモデルであっても、待ち行列長の定常分布を陽な形で得ることは困難である [10]。そして本稿が扱う集団到着と binomial catastrophe の発生が伴うマルコフ変調無限サーバ待ち行列では、後に述べる補題 3.1 から分かる通り、定常系内容数の各次モーメントの連立漸化式が定常系内容数分布について直接的に得られる情報の限度である。さらに本稿が扱うモデルでは、集団到着と binomial catastrophe の発生の両方を伴うため、系内容数を表すマルコフ連鎖は自然数上

で任意の幅だけ上下にジャンプできる．そのため本稿が扱うモデルの定常系内客数分布は，数値的に計算することさえ困難である．

そのため本稿では，以下のようなスケーリングを行なったスケーリングモデルを解析することで，定常系内客数分布に関する洞察を得る．あるスケーリングファクター N とスケーリング係数 $\alpha > 0$ に対して，次のようなスケーリングを行なったスケーリングモデルを考える．各 $j, \ell \in \mathbb{D}$ に対して，到着率 λ_j を N 倍，背後過程の推移率 $q_{\ell,j}$ を N^α 倍，catastrophe の発生率 γ_j を N 倍，そして死滅確率 $(1-p_j)$ を N^{-1} 倍とする．このスケーリングにおいて，catastrophe の発生率と死滅確率の積 $\gamma_j(1-p_j)$ は常に一定である．すなわち，ある客が catastrophe の発生が原因で死滅する率は，スケーリングを行っても変化しないということである．このスケーリングではサービス率は変化させないため，スケーリングモデルにおける客の退去率（サービス完了および catastrophe による死滅）は元のモデルと変わらないといえる．

本稿の主な目標は，スケーリングモデルの定常系内客数に対して中心極限定理を導くことである．すなわち， $M^{(N)}$ をスケーリングモデルの定常系内客数を表す確率変数として，次の確率変数がある分散をもつ平均ゼロの正規分布に弱収束することを示す．

$$N^{\beta/2} \left(\frac{M^{(N)}}{N} - \rho \right). \quad (1.1)$$

ただし， β は α の値のみで決定する定数である．

マルコフ変調待ち行列の解析は複雑であるため，このようなスケーリングモデルを用いた解析が時として行なわれる．例えば Blom et al.[3] ではスケーリングモデルを考えることで，マルコフ変調 $M/M/\infty$ 待ち行列の中心極限定理を導いている．つまり，本稿の結果は，Blom et al.[3] の結果の拡張といえる．

本稿では以下のような手順で，中心極限定理を導出する．まずは確率変数 $M^{(N)}N^{-1}$ は定数 ρ に確率収束することを示す．その後，確率変数 (1.1) の特性関数を $N \rightarrow \infty$ としたときに，平均ゼロの正規分布の特性関数に収束することを示す．そして，Lévy の連続定理を用いることで，確率変数 (1.1) が平均ゼロの正規分布に弱収束することを示す．ちなみに [3] においても，これと似たような手順で中心極限定理が示されている．しかし，本稿では特性関数の収束を示しているところを，[3] では (1.1) に該当する確率変数の積率母関数およびその微分がある区間で存在することを暗に仮定し，積率母関数の収束を示す方法をとっている．それに対し，本稿では積率母関数などの存在を暗に仮定することはせず，直接的に特性関数の収束の議論を行なう．

本稿の構成は以下のとおりである．2 節では本稿で扱うモデルを説明する．3 節では中心極限定理を示すために必要ないくつか補題を示す．4 節では本稿の主な結果である中心極限定理の証明を行なう．最後に 5 節では， N を大きくするにつれて，シミュレーションによる計算結果が正規分布に近づいていく様子を観察する．

2 モデルの設定

本節では，本稿が扱う binomial catastrophe の発生を伴うマルコフ変調 $M^X/M/\infty$ 待ち行列について説明する．無限個のサーバを有するこの待ち行列システムは，有限状態空間 $\mathbb{D} := \{1, 2, \dots, d\}$ 上の既約な連続時間マルコフ連鎖 $\{J(t); t \geq 0\}$ に支配される．背後過程 $\{J(t)\}$ の無限小生成作用素は $Q := (q_{\ell,j})_{\ell,j \in \mathbb{D}}$ で与えられるものとする．また各 $j \in \mathbb{D}$ に対し， $\{J(t)\}$ の状態 j における定常分布を τ_j として，定常分布のベクトルを $\tau := (\tau_j)_{j \in \mathbb{D}}$ と定める．

$J(t) = j$ のとき，客の集団は率 λ_j のポアソン過程に従って到着し，そのときに到着する集団の大きさ，すなわち集団に属する客数は $N := \{1, 2, \dots\}$ 上の確率変数 X_j に従う． X_j の分布は $x_{j,n} = P(X_j = n)$ で与えられるものとして，集団の大きさの分布の積率母関数を $\hat{X}_j(z) := \sum_{n=1}^{\infty} x_{j,n} z^n$ と記す．また本稿を通し，各 $j \in \mathbb{D}$ に対して $E[X_j^3] < \infty$ が成立すると仮定する．

客はシステムに到着後すぐに空きサーバを 1 つ占有してサービスを受け始める．そしてサービスが完了後すぐにシステムから退去する．客のサービス要求量は独立な平均 1 の指数分布に従うとして， $J(t) = j$ のときのサービスレート，すなわち 1 つのサーバが単位時間あたりに処理できるサービス要求量を $\mu_j \in (0, \infty)$ とする．したがって $J(t) = j$ のとき，システムに滞在している客の残余サービス時間は互いに独立な平均 $1/\mu_j$ の指数分布に従う．

$J(t) = j$ のとき，率 γ_j のポアソン過程に従って，binomial catastrophe が発生すると仮定する．そして，そのとき客は確率 $p_j \in (0, 1)$ でそのまま系内に滞在し，確率 $1-p_j$ で消滅，すなわちシステムから退去するものとする．客がそれぞれシステム内に滞在したままか，または消滅するかは互いに独立であると仮定する．

本稿では，上述の待ち行列に対してスケーリングを行なったスケーリングモデルを考える．具体的には，あるスケーリングファクター N とスケーリング係数 $\alpha > 0$ に対して，以下のようにスケーリングを行なう．各 $j, \ell \in \mathbb{D}$ に対して， $\lambda_j \mapsto N\lambda_j$ ， $q_{\ell,j} \mapsto N^\alpha q_{\ell,j}$ ， $\gamma_j \mapsto N\gamma_j$ ， $(1-p_j) \mapsto N^{-1}(1-p_j)$ とする．ただし， α は 1 よ

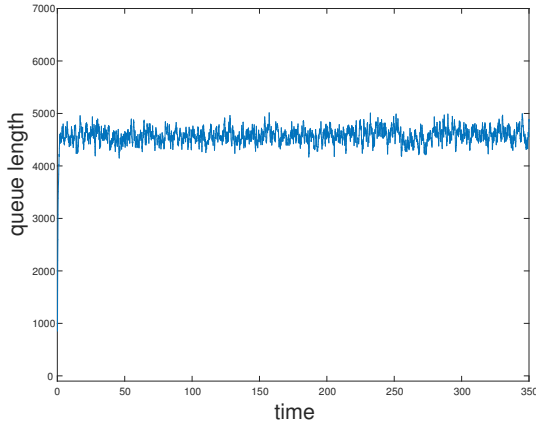


図 1: $\alpha = 2$

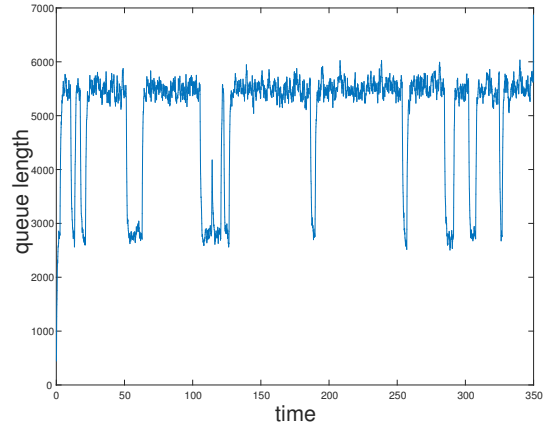


図 2: $\alpha = 0.1$

り大きい場合, 1 より小さい場合, そして 1 の場合, 全てを考えていることに注意する. α と 1 の大小関係によって, スケーリングされたシステムの挙動は大きく異なる (注意 2.1 参照).

スケーリングファクターが N であるスケーリングモデルの時刻 t での系内客数を $M^{(N)}(t)$, 背後過程の状態を $J^{(N)}(t)$ と記すと, スケーリングモデルの結合系内客数過程 $\{(J^{(N)}(t), M^{(N)}(t)); t \geq 0\}$ は状態空間 $\mathbb{D} \times \mathbb{Z}_+$ 上の連続時間マルコフ連鎖となる. ただし, $\mathbb{Z}_+ := \{0, 1, \dots\}$ である. また $M^{(N)}(0) = 0$ を仮定する. 本稿を通して, このマルコフ連鎖が唯一つの定常分布をもつことを仮定し, 各 $(j, n) \in \mathbb{D} \times \mathbb{Z}_+$ に対して状態 (j, n) の定常分布を $\pi_{j,n}^{(N)}$ と記す. そして, 定常状態における系内客数と背後過程の状態を表す確率変数をそれぞれ, $M^{(N)}, J^{(N)}$ とする. スケーリングモデルにおいて $N = 1$ とすれば, 元のモデルと一致することに注意する.

注意 2.1 図 1, 2 は binomial catastrophe の発生を伴うマルコフ変調 $M^X/M/\infty$ 待ち行列を $N = 1000$ でスケーリングしたモデルの待ち行列長のサンプルパスである. 図 1 は $\alpha = 2 (> 1)$, 図 2 は $\alpha = 0.1 (< 1)$ の場合のサンプルパスであり, 両者において N と α 以外のパラメータは 5 節で提示されているものと同じであり, 特に背後過程の次元は $d = 2$ である. 図 1 では, パラメータが一定である待ち行列長のサンプルパスのようなものが観察できる. これは α が 1 より大きく, スケーリングモデルでは到着間隔より背後過程の推移率のほうが極端に大きくなるため, N を大きくしていくとある一定のパラメータをもつ $M^X/M/\infty$ 待ち行列にシステムが収束していくことが原因であると推測できる. また図 2 では異なるパラメータを持つ 2 つの定常なサンプルパスが交互に出現しているように観察できる. これは α が 1 より小さく, スケーリングモデルでは到着間隔より背後過程の推移率のほうが極端に小さくなるため, 背後過程の一方の状態で十分時間が経過してシステムが定常になってから, 次に背後過程が推移するイベントが発生することが原因であると推測できる.

3 中心極限定理を導く準備

本節では, 中心極限定理を導く際に使用するいくつかの補題を示す. $|z| \leq 1$ を満たす z に対して, 行ベクトル $\hat{\pi}^{(N)}(z) = (\hat{\pi}_j^{(N)}(z))_{j \in \mathbb{D}}$ と $\hat{\pi}_p^{(N)}(z) = (\hat{\pi}_{p,j}^{(N)}(z))_{j \in \mathbb{D}}$ を次のように定義する.

$$\begin{aligned}\hat{\pi}_j^{(N)}(z) &= \mathbb{E}[z^{M^{(N)}} \mathbf{1}\{J^{(N)} = j\}], \quad j \in \mathbb{D}, \\ \hat{\pi}_{p,j}^{(N)}(z) &= \hat{\pi}_j^{(N)}(z + N^{-1}(1-z)(1-p_j)), \quad j \in \mathbb{D}.\end{aligned}$$

ただし, $\mathbf{1}\{A\}$ は事象 A の指示関数である.

補題 3.1 は系内客数の母関数が満たす微分方程式を示している. 証明は紙面の都合上省略する. 本稿ではこれ以降, (3.1) 式を利用して特性関数などの収束性を議論していくが, (3.1) 式の未知関数は単純に系内客数の母関数 $\hat{\pi}_j^{(N)}(z)$ のみでないことに注意しなければならない. (3.1) 式の未知関数には $\hat{\pi}_j^{(N)}(z)$ に $z = z + N^{-1}(1-z)(1-p_i)$ を代入した関数 $\hat{\pi}_{p,j}^{(N)}(z)$ が含まれており, これ以降の極限操作の難しさは, 主にこの項が原因となっている. ちなみに [3] でも, マルコフ変調 $M/M/\infty$ 待ち行列の中心極限定理を導出するために, このようにスケーリングされた系内客数の母関数が満たす微分方程式を導出するところから始めているが, (3.1) 式における $\hat{\pi}_{p,j}^{(N)}(z)$ のような項は出現しない. すなわち系内客数の母関数のみが未知関数である連立微分方程式を導出することができており, それ以降の極限操作の難しさは本稿のモデルと比較すると, かなり軽減されているといえる.

補題 3.1 各 $j \in \mathbb{D}$ に対し、母関数 $\hat{\pi}_j^{(N)}(z)$ は、 $\hat{\pi}_j^{(N)}(1) = \tau_j$ および以下の微分方程式を満たす。

$$\mu_j(1-z) \frac{d}{dz} \hat{\pi}_j^{(N)}(z) = N\lambda_j \{1 - \hat{X}_j(z)\} \hat{\pi}_j^{(N)}(z) - N^\alpha \sum_{\ell \in \mathbb{D}} q_{\ell,j} \hat{\pi}_\ell^{(N)}(z) - N\gamma_j \{\hat{\pi}_{p,j}^{(N)}(z) - \hat{\pi}_j^{(N)}(z)\}. \quad (3.1)$$

次に確率変数 $M^{(N)}N^{-1}$ に関する補題 3.2 と補題 3.3 を示す。詳しい証明は紙面の都合上省略するが、証明では集団の大きさの 3 次モーメントが有限である仮定を利用している。

補題 3.2 以下の式が成立する。

$$\mathbb{E}[M^{(N)}N^{-1}] = O(1), \quad \mathbb{E}[\{M^{(N)}N^{-1}\}^2] = O(1), \quad \mathbb{E}[\{M^{(N)}N^{-1}\}^3] = O(1).$$

ただし、 $f(N) = O(g(N))$ は $\lim_{N \rightarrow \infty} |f(N)/g(N)|$ が存在する場合、その値が有限であることを意味する。

補題 3.3 任意の $\theta \in (-\infty, \infty)$ に対して、以下の式が成立する。ただし、 $i = \sqrt{-1}$ 。

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E}[M^{(N)}N^{-1} \exp(i\theta M^{(N)}N^{\beta/2-1})] &= \rho \lim_{N \rightarrow \infty} \mathbb{E}[\exp(i\theta M^{(N)}N^{\beta/2-1})], \\ \lim_{N \rightarrow \infty} \mathbb{E}[\{M^{(N)}N^{-1}\}^2 \exp(i\theta M^{(N)}N^{\beta/2-1})] &= \rho^2 \lim_{N \rightarrow \infty} \mathbb{E}[\exp(i\theta M^{(N)}N^{\beta/2-1})]. \end{aligned}$$

最後に定常系内客数に関する大数の法則を示す。これ以降、式が煩雑になるのを避けるため、次の対角行列を導入する。 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$, $\widehat{\mathbf{X}}(z) = \text{diag}(\widehat{X}_1(z), \widehat{X}_2(z), \dots, \widehat{X}_d(z))$, $\overline{\mathbf{X}} = \text{diag}(\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_d])$, $\underline{\mathbf{X}} = \text{diag}(\mathbb{E}[X_1^2], \mathbb{E}[X_2^2], \dots, \mathbb{E}[X_d^2])$, $\mathbf{M} = \text{diag}(\mu_1, \mu_2, \dots, \mu_d)$, $\mathbf{\Gamma} = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_d)$, $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_d)$ 。

また、補題 3.4 で示した確率収束先は非常に単純で分かり易い形になっている。 ρ の分子はスケーリング前のモデルにおける単位時間あたり平均の到着客数であり、分母の逆数はスケーリング前のモデルにおいてある客がシステムから離れるまでにかかる時間の平均である。

補題 3.4 確率変数 $M^{(N)}N^{-1}$ を $N \rightarrow \infty$ とすると、以下で定められる定数 ρ に確率収束する。

$$\rho = \frac{\sum_{j \in \mathbb{D}} \tau_j \lambda_j \mathbb{E}[X_j]}{\sum_{j \in \mathbb{D}} \tau_j \{\mu_j + (1-p_j)\gamma_j\}}.$$

証明. (3.1) 式は次のような形に書き換えられる。

$$\hat{\pi}^{(N)}(z) \mathbf{Q} = N^{-\alpha}(z-1) \frac{d}{dz} \hat{\pi}^{(N)}(z) \mathbf{M} - N^{1-\alpha} \hat{\pi}^{(N)}(z) \mathbf{\Lambda} (\widehat{\mathbf{X}}(z) - \mathbf{I}) - N^{1-\alpha} \{\hat{\pi}_p^{(N)}(z) - \hat{\pi}^{(N)}(z)\} \mathbf{\Gamma}. \quad (3.2)$$

$\theta \in (-\infty, 0]$ に対して $z(\theta) := \exp(\theta N^{-1})$ と定義する。特に明記はしないが、 $z(\theta)$ は N に依存していることに注意する。さらに、ベクトル $\tilde{\pi}^{(N)}(\theta) = \hat{\pi}^{(N)}(z(\theta))$ および $\tilde{\pi}_p^{(N)}(\theta) = \hat{\pi}_p^{(N)}(z(\theta))$ を定義する。ベクトル $\tilde{\pi}^{(N)}(\theta)$ の総和は確率変数 $M^{(N)}N^{-1}$ の積率母関数となる。そして、 $M^{(N)}N^{-1}$ が非負実数値確率変数であるため、 $\tilde{\pi}^{(N)}(\theta)$ は任意の $\theta \in (-\infty, 0]$ に対して全ての要素が有限である。(3.2) 式に $z = z(\theta)$ を代入すると、

$$\tilde{\pi}^{(N)}(\theta) \mathbf{Q} = N^{-\alpha}(z(\theta)-1) \frac{d}{dz} \hat{\pi}^{(N)}(z(\theta)) \mathbf{M} - N^{1-\alpha} \tilde{\pi}^{(N)}(\theta) \mathbf{\Lambda} (\widehat{\mathbf{X}}(z(\theta)) - \mathbf{I}) - N^{1-\alpha} \{\tilde{\pi}_p^{(N)}(\theta) - \tilde{\pi}^{(N)}(\theta)\} \mathbf{\Gamma}.$$

ここで、 $\frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) = N^{-1} z(\theta) \frac{d}{dz} \hat{\pi}^{(N)}(z(\theta))$ より、上式は次のような微分方程式に変形できる。

$$\begin{aligned} \tilde{\pi}^{(N)}(\theta) \mathbf{Q} &= N^{1-\alpha} (1 - z(\theta)^{-1}) \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \mathbf{M} - N^{1-\alpha} \tilde{\pi}^{(N)}(\theta) \mathbf{\Lambda} (\widehat{\mathbf{X}}(z(\theta)) - \mathbf{I}) \\ &\quad - N^{1-\alpha} \{\tilde{\pi}_p^{(N)}(\theta) - \tilde{\pi}^{(N)}(\theta)\} \mathbf{\Gamma}. \end{aligned} \quad (3.3)$$

以降しばらく、 $N \rightarrow \infty$ としたときの (3.3) 式の収束性について議論していく。まず $z(\theta)^{-1}$ のテーラー展開を考えると、次の式が成立する。

$$1 - z(\theta)^{-1} = \theta N^{-1} + O(N^{-2}). \quad (3.4)$$

また、各 $j \in \mathbb{D}$ に対して $|N(z(\theta)^{X_j} - 1)| \leq -\theta X_j$ が任意の $\theta \in (-\infty, 0]$ で成立し、さらに $\mathbb{E}[X_j] < \infty$ を仮定していたので、優収束定理より次の式を得る。

$$\lim_{N \rightarrow \infty} N(\widehat{X}_j(z(\theta)) - 1) = \mathbb{E} \left[\lim_{N \rightarrow \infty} N(z(\theta)^{X_j} - 1) \right] = \theta \mathbb{E}[X_j].$$

よって、次の式が成立することが分かる。

$$\widehat{\mathbf{X}}(z(\theta)) - \mathbf{I} = N^{-1}\theta\overline{\mathbf{X}} + o(N^{-1}). \quad (3.5)$$

ただし、 $f(N) = o(g(N))$ は $\lim_{N \rightarrow \infty} |f(N)/g(N)| = 0$ を意味する。そして補題 A.1(付録参照) より、(3.3) 式の最後の項は次のように変形できる。

$$N^{1-\alpha} \{ \tilde{\pi}_p^{(N)}(\theta) - \tilde{\pi}^{(N)}(\theta) \} \mathbf{\Gamma} = -\theta N^{-\alpha} \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} + o(N^{-\alpha}). \quad (3.6)$$

(3.4), (3.5), (3.6) 式を (3.3) 式に適用すると、 $\theta \in (-\infty, 0]$ に対して次の式を得る。

$$\tilde{\pi}^{(N)}(\theta) \mathbf{Q} = \theta N^{-\alpha} \left\{ \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \mathbf{M} - \tilde{\pi}^{(N)}(\theta) \mathbf{\Lambda} \overline{\mathbf{X}} + \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} \right\} + o(N^{-\alpha}). \quad (3.7)$$

$\mathbf{T} := \mathbf{1} \cdot \tau$ と定義すると、 $\mathbf{T}^2 = \mathbf{T}$ 、 $\mathbf{T}\mathbf{1} = \mathbf{1}$ および $\mathbf{Q}\mathbf{T} = \mathbf{O}$ が成立する。さらに、これらから以下の式が導かれる。

$$\mathbf{Q}(\mathbf{T} - \mathbf{Q})^{-1} = -(\mathbf{T} - \mathbf{Q})(\mathbf{T} - \mathbf{Q})^{-1} + \mathbf{T}(\mathbf{T} - \mathbf{Q})^{-1} = -\mathbf{I} + \mathbf{T}, \quad (3.8)$$

$$(\mathbf{T} - \mathbf{Q})^{-1} \mathbf{1} = (\mathbf{T} - \mathbf{Q})^{-1} \mathbf{T}\mathbf{1} = \mathbf{1} + (\mathbf{T} - \mathbf{Q})^{-1} \mathbf{Q}\mathbf{1} = \mathbf{1}. \quad (3.9)$$

ただし、 $\mathbf{O}$ は全ての要素が 0 である正方行列であり、 $\mathbf{1}$ は全ての要素が 1 である列ベクトルである。ここで $\mathbf{T} - \mathbf{Q}$ は、 $\mathbf{Q}$ の既約性から狭義優対角行列であることが示せるため、逆行列の存在が保証される。(3.7) 式の右側から $(\mathbf{T} - \mathbf{Q})^{-1}$ を掛けると、(3.8) 式より、

$$\tilde{\pi}^{(N)}(\theta) \mathbf{T} - \tilde{\pi}^{(N)}(\theta) = N^{-\alpha} \theta \left[\frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \{ \mathbf{M} + (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} \} - \tilde{\pi}^{(N)}(\theta) \mathbf{\Lambda} \overline{\mathbf{X}} \right] (\mathbf{T} - \mathbf{Q})^{-1} + o(N^{-\alpha}). \quad (3.10)$$

さらに上式の右側から $N^\alpha \mathbf{1}$ を掛けると、 $\mathbf{T}\mathbf{1} = \mathbf{1}$ と (3.9) 式より、次の式が得られる。

$$0 = \left[\frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \{ \mathbf{M} + (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} \} - \tilde{\pi}^{(N)}(\theta) \mathbf{\Lambda} \overline{\mathbf{X}} \right] \mathbf{1} + o(1). \quad (3.11)$$

ここで (3.10) 式より、以下の関係式が成立することが分かる。

$$\tilde{\pi}^{(N)}(\theta) \mathbf{T} - \tilde{\pi}^{(N)}(\theta) = o(1), \quad \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \mathbf{T} - \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) = o(1).$$

これらを (3.11) 式に代入すると、次の式が得られる。

$$0 = \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \mathbf{T} \{ \mathbf{M} + (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} \} \mathbf{1} - \tilde{\pi}^{(N)}(\theta) \mathbf{T} \mathbf{\Lambda} \overline{\mathbf{X}} \mathbf{1} + o(1).$$

$\tilde{\pi}^{(N)}(\theta) = \tilde{\pi}^{(N)}(\theta) \mathbf{1}$ とすると、上式は次のように書き換えられる。

$$0 = \frac{d}{d\theta} \tilde{\pi}^{(N)}(\theta) \sum_{j \in \mathbb{D}} \tau_j \{ \mu_j + (1 - p_j) \gamma_j \} - \tilde{\pi}^{(N)}(\theta) \sum_{j \in \mathbb{D}} \tau_j \lambda_j + o(1).$$

上式において $N \rightarrow \infty$ として極限をとると、次の微分方程式を得る。

$$0 = \rho \tilde{\pi}^{(\infty)}(\theta) - \frac{d}{d\theta} \tilde{\pi}^{(\infty)}(\theta).$$

ただし、 $\tilde{\pi}^{(\infty)}(\theta) = \lim_{N \rightarrow \infty} \tilde{\pi}^{(N)}(\theta)$ 。 $\tilde{\pi}^{(\infty)}(0) = 1$ より、この微分方程式の解は次の形で得られる。

$$\tilde{\pi}^{(\infty)}(\theta) = \exp(\rho\theta).$$

これは定数 ρ の積率母関数に他ならない。よって、Lévy の連続定理より、確率変数 $M^{(N)} N^{-1}$ は定数 ρ へ弱収束する。さらに、確率変数が定数へ弱収束することと確率収束することは同値であるため、 $M^{(N)} N^{-1}$ は定数 ρ へ確率収束することが示された。 $\square$

4 中心極限定理の導出

本節では、本稿の主結果である中心極限定理の導出を行なう。定理 4.1 の結果は、各 $j \in \mathbb{D}$ に対して、 X_j を定数 1, $\gamma_j = 0$ とすれば、[3] で導出されているマルコフ変調 $M/M/\infty$ 待ち行列に対する中心極限定理と一致する。

また、定理 4.1 で示されている収束先の分散は α の値によって変化する。 α が 1 より小さい場合は σ_1^2 , 1 より大きい場合は σ_2^2 , 1 の場合にはそれらの和となる。つまりスケーリングモデルの系内客数分布は α と 1 の大小関係に大きく依存するといえる。これは注意 2.1 のサンプルパスの挙動および直感的な予想と一致している。

定理 4.1 確率変数

$$N^{\beta/2} \left(\frac{M^{(N)}}{N} - \rho \right), \quad (4.1)$$

は $N \rightarrow \infty$ としたとき、分散 σ^2 をもつ平均ゼロの正規分布に弱収束する。すなわち、

$$N^{\beta/2} \left(\frac{M^{(N)}}{N} - \rho \right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

ただし、 $\beta = \min(\alpha, 1)$ であり、分散 σ^2 は以下のとおり。

$$\begin{aligned} \sigma^2 &= \sigma_1^2 1\{\alpha \leq 1\} + \sigma_2^2 1\{\alpha \geq 1\}, \\ \sigma_1^2 &= \frac{1}{\mu_\infty} \tau [\Lambda \bar{X} - \rho(M + (I - P)\Gamma)](T - Q)^{-1} [\Lambda \bar{X} - \rho(M + (I - P)\Gamma)] 1, \\ \sigma_2^2 &= \frac{1}{2\mu_\infty} \sum_{j \in \mathbb{D}} \tau_j \{ \lambda_j E[X_j^2] + \rho \mu_j + \rho \gamma_j (1 - p_j) + \rho^2 \gamma_j (1 - p_j)^2 \}, \\ \mu_\infty &= \sum_{j \in \mathbb{D}} \tau_j \{ \mu_j + \gamma_j (1 - p_j) \}. \end{aligned}$$

ただし、 $T = \mathbf{1} \cdot \tau$ であり、 $1\{A\}$ は事象 A の指示関数。

証明. 確率変数 (4.1) を $\psi^{(N)}$ と記す。さらに、ベクトル $\psi_*^{(N)}(\theta) := \exp(-\rho i \theta N^{\beta/2}) \hat{\pi}^{(N)}(\exp(i \theta N^{-1+\beta/2}))$ を定義する。 $\psi_*^{(N)}(\theta)$ を $\psi^{(N)}$ の特性関数とすると、 $\psi_*^{(N)}(\theta) = \psi_*^{(N)}(\theta) \mathbf{1}$ が成立することが分かる。

(3.2) 式の右側から $(T - Q)^{-1}$ を掛けたものに (3.8) 式を代入すると

$$\begin{aligned} \hat{\pi}^{(N)}(z) &= \hat{\pi}^{(N)}(z) T - N^{-\alpha} (z - 1) \frac{d}{dz} \hat{\pi}^{(N)}(z) M (T - Q)^{-1} \\ &\quad + N^{1-\alpha} \hat{\pi}^{(N)}(z) \Lambda(\widehat{X}(z) - I) (T - Q)^{-1} + N^{1-\alpha} \{ \hat{\pi}_p^{(N)}(z) - \hat{\pi}^{(N)}(z) \} \Gamma (T - Q)^{-1}. \end{aligned}$$

$\theta \in (-\infty, \infty)$ に対して $z(\theta) := \exp(i \theta N^{-1+\beta/2})$ と定義する。ここで、特に明記はしないが $z(\theta)$ は N に依存していることに注意する。上式に $z = z(\theta)$ を代入した式に $\exp(-\rho i \theta N^{\beta/2})$ を掛けると、次の式を得る。

$$\begin{aligned} \psi_*^{(N)}(\theta) &= \psi_*^{(N)}(\theta) T - N^{-\alpha} \exp(-\rho i \theta N^{\beta/2}) (z(\theta) - 1) \frac{d}{dz} \hat{\pi}^{(N)}(z(\theta)) M (T - Q)^{-1} \\ &\quad + N^{1-\alpha} \psi_*^{(N)}(\theta) \Lambda(\widehat{X}(z(\theta)) - I) (T - Q)^{-1} \\ &\quad + N^{1-\alpha} \{ \exp(-\rho i \theta N^{\beta/2}) \hat{\pi}_p^{(N)}(z(\theta)) - \psi_*^{(N)}(\theta) \} \Gamma (T - Q)^{-1}. \end{aligned} \quad (4.2)$$

ここで $\frac{d}{d\theta} \psi_*^{(N)}(\theta) = -i \rho N^{\beta/2} \psi_*^{(N)}(\theta) + i N^{-1+\beta/2} e^{-\rho i \theta N^{\beta/2}} z(\theta) \frac{d}{dz} \hat{\pi}^{(N)}(z(\theta))$ を (4.2) 式に代入すると、次の微分方程式を得る。

$$\begin{aligned} \psi_*^{(N)}(\theta) &= \psi_*^{(N)}(\theta) T + i N^{1-\alpha-\beta/2} (1 - z(\theta)^{-1}) \frac{d}{d\theta} \psi_*^{(N)}(\theta) M (T - Q)^{-1} \\ &\quad - \rho N^{1-\alpha} (1 - z(\theta)^{-1}) \psi_*^{(N)}(\theta) M (T - Q)^{-1} + N^{1-\alpha} \psi_*^{(N)}(\theta) \Lambda(\widehat{X}(z(\theta)) - I) (T - Q)^{-1} \\ &\quad - N^{1-\alpha} \{ \psi_*^{(N)}(\theta) - \exp(-\rho i \theta N^{\beta/2}) \hat{\pi}_p^{(N)}(z(\theta)) \} \Gamma (T - Q)^{-1}. \end{aligned} \quad (4.3)$$

以降しばらく、 $N \rightarrow \infty$ としたときの (4.3) 式の収束性を議論していく。 $z(\theta)^{-1}$ のテーラー展開を考えると、

$$1 - z(\theta)^{-1} = i \theta N^{-1+\beta/2} + \frac{1}{2} \theta^2 N^{-2+\beta} + O(N^{-3+3\beta/2}). \quad (4.4)$$

また、部分積分を行なうことで、任意の $x \in (-\infty, \infty)$ に対して次の式が成立することが分かる。

$$z(\theta)^x = 1 + ix\theta N^{\beta/2-1} - \frac{1}{2}x^2\theta^2 N^{\beta-2} - \frac{i}{2} \int_0^{x\theta N^{\beta/2-1}} (x\theta N^{\beta/2-1} - u)^2 e^{iu} du. \quad (4.5)$$

(4.5) 式を利用すると、各 $j \in \mathbb{D}$ に対して、次の関係式が成立することが分かる。

$$\lim_{N \rightarrow \infty} \left| N(\widehat{X}_j(z(\theta)) - 1) - i\theta N^{\beta/2} \mathbb{E}[X_j] + \frac{\theta^2}{2} N^{\beta-1} \mathbb{E}[X_j^2] \right| \leq \lim_{N \rightarrow \infty} \frac{1}{6} N^{3/2\beta-2} |\theta^3| \mathbb{E}[X_j^3] = 0.$$

上式は次のように書き換えられる。

$$N(\widehat{\mathbf{X}}(z(\theta)) - 1) = i\theta N^{\beta/2} \overline{\mathbf{X}} - \frac{\theta^2}{2} N^{\beta-1} \underline{\mathbf{X}} + o(1). \quad (4.6)$$

また、補題 A.2(付録参照) より、次の式が成立することが分かる。

$$\begin{aligned} & N\{\psi_*^{(N)}(\theta) - \exp(-\rho i\theta N^{\beta/2}) \widehat{\pi}_p^{(N)}(z(\theta))\} \\ &= \theta \frac{d}{d\theta} \psi_*^{(N)}(\theta) (\mathbf{I} - \mathbf{P}) + iN^{\beta/2} \theta \psi_*^{(N)}(\theta) (\mathbf{I} - \mathbf{P}) \rho + \frac{1}{2} N^{\beta-1} \theta^2 \psi_*^{(N)}(\theta) \{(\mathbf{I} - \mathbf{P}) \rho + (\mathbf{I} - \mathbf{P})^2 \rho^2\} + o(1). \end{aligned} \quad (4.7)$$

(4.4), (4.6), (4.7) 式を (4.3) 式に代入すると、次の式が得られる。

$$\begin{aligned} & \psi_*^{(N)}(\theta) - \psi_*^{(N)}(\theta) \mathbf{T} \\ &= -N^{-\alpha} \theta \frac{d}{d\theta} \psi_*^{(N)}(\theta) [\mathbf{M} + (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] (\mathbf{T} - \mathbf{Q})^{-1} + i\theta N^{-\alpha+\beta/2} \psi_*^{(N)}(\theta) [\mathbf{\Lambda} \overline{\mathbf{X}} - \rho \mathbf{M} - \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] (\mathbf{T} - \mathbf{Q})^{-1} \\ & \quad - \frac{1}{2} N^{-\alpha+\beta-1} \theta^2 \psi_*^{(N)}(\theta) [\mathbf{\Lambda} \underline{\mathbf{X}} + \rho \mathbf{M} + \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} + \rho^2(\mathbf{I} - \mathbf{P})^2 \mathbf{\Gamma}] (\mathbf{T} - \mathbf{Q})^{-1} + o(N^{-\alpha}). \end{aligned} \quad (4.8)$$

(4.8) 式の右側から $\theta^{-1} N^\alpha \mathbf{1}$ を掛けると、 $\mathbf{T} \mathbf{1} = \mathbf{1}$ より、次の式が導かれる。

$$\begin{aligned} 0 &= -\frac{d}{d\theta} \psi_*^{(N)}(\theta) [\mathbf{M} + (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] \mathbf{1} + iN^{\beta/2} \psi_*^{(N)}(\theta) [\mathbf{\Lambda} \overline{\mathbf{X}} - \rho \mathbf{M} - \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] \mathbf{1} \\ & \quad - \frac{1}{2} N^{\beta-1} \theta \psi_*^{(N)}(\theta) [\mathbf{\Lambda} \underline{\mathbf{X}} + \rho \mathbf{M} + \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} + \rho^2(\mathbf{I} - \mathbf{P})^2 \mathbf{\Gamma}] \mathbf{1} + o(1). \end{aligned} \quad (4.9)$$

さらに、 $\beta \leq 1$ かつ $\beta \leq \alpha$ であることを考慮すると、(4.8) 式より以下の式が成立することが示される。

$$\begin{aligned} & \psi_*^{(N)}(\theta) = \psi_*^{(N)}(\theta) \mathbf{T} + o(1), \\ & \frac{d}{d\theta} \psi_*^{(N)}(\theta) = \frac{d}{d\theta} \psi_*^{(N)}(\theta) \mathbf{T} + o(1), \\ & N^{\beta/2} \psi_*^{(N)}(\theta) = N^{\beta/2} \psi_*^{(N)}(\theta) \mathbf{T} + iN^{\beta-\alpha} \theta \psi_*^{(N)}(\theta) [\mathbf{\Lambda} \overline{\mathbf{X}} - \rho \mathbf{M} - \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] (\mathbf{T} - \mathbf{Q})^{-1} + o(1). \end{aligned}$$

これらの式と $\tau[\mathbf{\Lambda} \overline{\mathbf{X}} - \rho \mathbf{M} - \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] \mathbf{1} = 0$ を (4.9) 式に適用すると、次の式が得られる。

$$\begin{aligned} 0 &= -\frac{d}{d\theta} \widetilde{\psi}^{(N)}(\theta) \tau[\mathbf{M} + (\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] \mathbf{1} \\ & \quad - N^{\beta-\alpha} \theta \widetilde{\psi}^{(N)}(\theta) \tau[\mathbf{\Lambda} - \rho \mathbf{M} - \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma}] (\mathbf{T} - \mathbf{Q})^{-1} [\mathbf{\Lambda} \overline{\mathbf{X}} - \rho \mathbf{M} - \rho \mathbf{\Gamma} (\mathbf{I} - \mathbf{P})] \mathbf{1} \\ & \quad - \frac{1}{2} N^{\beta-1} \theta \widetilde{\psi}^{(N)}(\theta) \tau[\mathbf{\Lambda} \underline{\mathbf{X}} + \rho \mathbf{M} + \rho(\mathbf{I} - \mathbf{P}) \mathbf{\Gamma} + \rho^2(\mathbf{I} - \mathbf{P})^2 \mathbf{\Gamma}] \mathbf{1} + o(1). \end{aligned}$$

σ_1^2 と σ_2^2 の定義より、上式は次のように書き換えられる。

$$0 = -\frac{d}{d\theta} \widetilde{\psi}^{(N)}(\theta) - N^{\beta-\alpha} \sigma_1^2 \theta \widetilde{\psi}^{(N)}(\theta) - N^{\beta-1} \sigma_2^2 \theta \widetilde{\psi}^{(N)}(\theta) + o(1).$$

上式を両辺 $N \rightarrow \infty$ とすると、次の微分方程式を得る。

$$\frac{d}{d\theta} \widetilde{\psi}^{(\infty)}(\theta) = -\sigma^2 \theta \widetilde{\psi}^{(\infty)}(\theta).$$

ただし、 $\widetilde{\psi}^{(\infty)}(\theta) = \lim_{N \rightarrow \infty} \widetilde{\psi}^{(N)}(\theta)$ 、 $\widetilde{\psi}^{(N)}(0) = 1$ より、この微分方程式の解は次の形で得られる。

$$\widetilde{\psi}^{(\infty)}(\theta) = \exp\left(-\frac{\sigma^2}{2} \theta^2\right).$$

これは、分散 σ^2 をもつ平均ゼロの正規分布の特性関数に他ならない。よって、Lévy の連続定理より、確率変数 $\psi^{(N)}$ は分散 σ^2 をもつ平均ゼロの正規分布に弱収束する。 $\square$

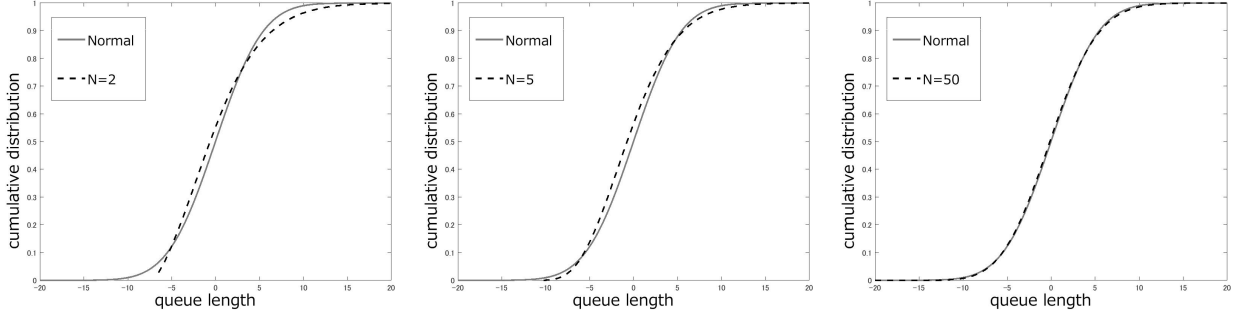


図 3: $\alpha = 2$

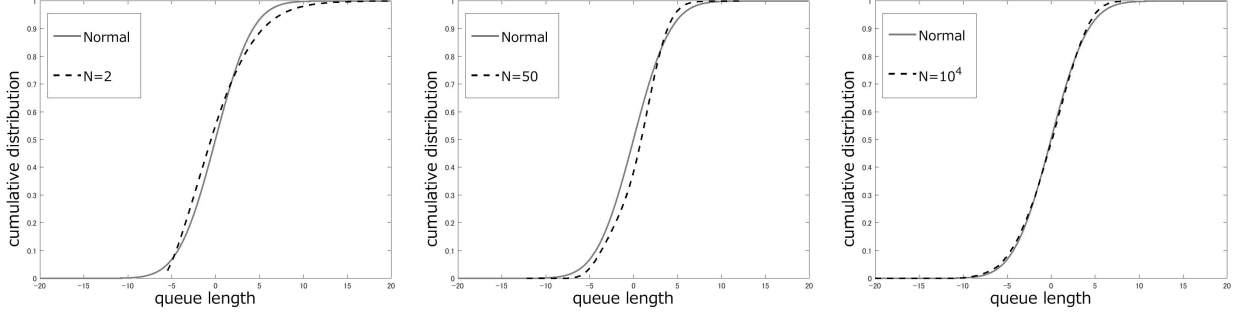


図 4: $\alpha = 0.5$

5 正規近似とシミュレーション結果の比較

本節では、シミュレーションにより得た確率変数 (4.1) の分布と定理 4.1 で示されている弱収束先の正規分布を比較し、 N が大きくなるにつれて、前者が後者に収束していく様子を確認する。 α 以外のパラメータは共通して以下のとおりとする。

$$d = 2, \quad Q = \begin{pmatrix} -0.2 & 0.2 \\ 0.1 & -0.1 \end{pmatrix}, \quad (\lambda_1, \lambda_2) = (1, 2), \quad (\gamma_1, \gamma_2) = (10^{-4}, 2 \times 10^{-3}), \\ (\mu_1, \mu_2) = (2, 2), \quad (p_1, p_2) = (0.2, 0.1).$$

図 3, 4 では、横軸は確率変数 (4.1) の値を、縦軸はその累積分布関数を表している。図 3 は $\alpha = 2 (> 1)$, N を左から順に 2, 5, 50 としてシミュレーションを行なった結果である。これらと正規分布の比較をすると、 N がさほど大きくなくてもシミュレーション結果と正規分布はかなり近いことが分かる。よって α が 1 より大きい場合、定理 4.1 で示されている収束の収束スピードはかなり速いと推測できる。図 4 は $\alpha = 0.5 (< 1)$, N を左から順に 2, 50, 10^4 としてシミュレーションを行なった結果である。これらと正規分布の比較をすると、 N をかなり大きくしないとシミュレーション結果は正規分布に近づいていないことが分かる。よって α が 1 より小さい場合、定理 4.1 で示されている収束の収束スピードは、 α が 1 より大きい場合と比較するとかなり遅いと推測できる。

参考文献

- [1] R. Bartoszyński, W. J. Buehler, W. Chan, and D. K. Pearl. Population processes under the influence of disasters occurring independently of population size. *Journal of Mathematical Biology*, 27(2):167–178, 1989.
- [2] M. Baykal-Gursoy and W. Xiao. Stochastic decomposition in M/M/ ∞ queues with markov modulated service rates. *Queueing Systems*, 48(1):75–88, 2004.
- [3] J. Blom, K. De Turck, and M. Mandjes. A central limit theorem for markov-modulated infinite-server queues. In *Proceedings of ASMTA 2013, Lecture Notes in Computer Science*, 7984:81–95, 2013.
- [4] P. J. Brockwell, J. Gani, and S. I. Resnick. Birth, immigration and catastrophe processes. *Advances in Applied Probability*, 14(4):709–731, 1982.
- [5] X. Chao and Y. Zheng. Transient analysis of immigration birth–death processes with total catastrophes. *Probability in the Engineering and Informational Sciences*, 17(1):83–106, 2003.

- [6] A. Di Crescenzo, V. Giorno, A. G. Nobile, and L. M. Ricciardi. On the M/M/1 queue with catastrophes and its continuous approximation. *Queueing Systems*, 43(4):329–347, 2003.
- [7] S. Kapodistria, T. Phung-Duc, and J. Resing. Linear birth/immigration-death process with binomial catastrophes. *Probability in the Engineering and Informational Sciences*, 30(1):79–111, 2016.
- [8] B. Krishna Kumar and D. Arivudainambi. Transient solution of an M/M/1 queue with catastrophes. *Computers & Mathematics with applications*, 40(10-11):1233–1240, 2000.
- [9] B. Krishna Kumar, A. Vijayakumar, and S. Sophia. Transient analysis for state-dependent queues with catastrophes. *Stochastic Analysis and Applications*, 26(6):1201–1217, 2008.
- [10] C. A. O’cinneide and P. Purdue. The M/M/ ∞ queue in a random environment. *Journal of Applied Probability*, 23(1):175–184, 1986.
- [11] R. J. Swift. Transient probabilities for a simple birth-death-immigration process under the influence of total catastrophes. *International Journal of Mathematics and Mathematical Sciences*, 25(10):689–692, 2001.
- [12] M. Yajima, T. Phung-Duc, and H. Masuyama. The stability condition of BMAP/M/ ∞ queues (in press). In *Proceedings of 11th International Conference on Queueing Theory and Network Applications, QTNA 2016*, 2016.

A 証明内で用いた補題

補題 A.1 $z(\theta) = \exp(M^{(N)}N^{-1}\theta)$ とすると, 任意の $j \in \mathbb{D}$ と $\theta \in (-\infty, 0]$ に対して以下の式が成立する.

$$\begin{aligned} & NE[\{N^{-1}(1-p_j) + z(\theta)(1-N^{-1}(1-p_j))\}^{M^{(N)}} 1\{J^{(N)} = j\}] - NE[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\ &= -\theta(1-p_j) \frac{d}{d\theta} E[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] + o(1). \end{aligned} \quad (A.1)$$

証明. (A.1) 式の左辺は次のように変形できる.

$$\begin{aligned} & NE[\{N^{-1}(1-p_j) + z(\theta)(1-N^{-1}(1-p_j))\}^{M^{(N)}} 1\{J^{(N)} = j\}] - NE[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\ &= N(1-p_j) E \left[M^{(N)} N^{-1} z(\theta)^{M^{(N)}} \sum_{k=1}^{M^{(N)}} \binom{M^{(N)}-1}{k-1} \frac{1}{k} (z(\theta)^{-k} - 1) \right. \\ & \quad \left. \times \{N^{-1}(1-p_j)\}^{k-1} \{1-N^{-1}(1-p_j)\}^{M^{(N)}-k} 1\{J^{(N)} = j\} \right]. \end{aligned} \quad (A.2)$$

ここで, 任意の $k \geq 1$ と $\theta \in (-\infty, 0]$ に対し, 次の不等式が成立する.

$$z(\theta)^{-k} - 1 = (z(\theta)^{-1} - 1)k + (z(\theta)^{-1} - 1)^2 \sum_{\ell=1}^{k-1} \sum_{h=0}^{\ell-1} z(\theta)^{-h} \leq (z(\theta)^{-1} - 1)k + (z(\theta)^{-1} - 1)^2 \frac{1}{2} k(k-1) z(\theta)^{-k+1}.$$

(A.2) 式にこれを代入した式に対して, 補題 3.2 と (3.4) 式を適用すると, 次の不等式を得る.

$$\begin{aligned} & NE[\{N^{-1}(1-p_j) + z(\theta)(1-N^{-1}(1-p_j))\}^{M^{(N)}} 1\{J^{(N)} = j\}] - NE[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\ & \leq -\theta(1-p_j) \frac{d}{d\theta} E[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] + f(N). \end{aligned} \quad (A.3)$$

ただし, $f(N)$ は $\lim_{N \rightarrow \infty} f(N) = 0$ を満たすある関数とする. 一方で, 任意の $k \geq 1$ と $\theta \in (-\infty, 0]$ に対して, $z(\theta)^{-k} \geq 1 - k\theta N^{-1}$ が成立するため, (A.2) 式より次の不等式が導かれる.

$$\begin{aligned} & NE[\{N^{-1}(1-p_j) + z(\theta)(1-N^{-1}(1-p_j))\}^{M^{(N)}} 1\{J^{(N)} = j\}] - NE[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\ & \geq -\theta(1-p_j) \frac{d}{d\theta} E[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}]. \end{aligned}$$

上式を (A.3) 式と組みあわせると, (A.1) 式が成立することが分かる. □

補題 A.2 $z(\theta) = \exp(i\theta N^{\beta/2} \{M^{(N)} N^{-1} - \rho\})$ とすると, 任意の $j \in \mathbb{D}$ と $\theta \in (-\infty, \infty)$ に対して以下の式が成立する.

$$\begin{aligned}
& N \exp(-\rho i N^{\beta/2} \theta) \left\{ \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] - \mathbb{E}[\{z(\theta) + (1 - z(\theta))N^{-1}(1 - p_j)\}^{M^{(N)}} 1\{J^{(N)} = j\}] \right\} \\
&= \theta(1 - p_j) \frac{d}{d\theta} \exp(-\rho i N^{\beta/2} \theta) \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\
&\quad + i\theta(1 - p_j) \rho N^{\beta/2} \exp(-\rho i N^{\beta/2} \theta) \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\
&\quad + \frac{1}{2} (1 - p_j) \rho + (1 - p_j)^2 \rho^2 \theta^2 N^{\beta-1} \exp(-\rho i N^{\beta/2} \theta) \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] + o(1). \tag{A.4}
\end{aligned}$$

証明. 補題 3.3 より, 次の式が成立すれば (A.4) 式が成立することがいえる.

$$\begin{aligned}
& N \left\{ \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] - \mathbb{E}[\{z(\theta) + (1 - z(\theta))N^{-1}(1 - p_j)\}^{M^{(N)}} 1\{J^{(N)} = j\}] \right\} \\
&= i\theta(1 - p_j) \mathbb{E}[M^{(N)} N^{\beta/2-1} z(\theta) 1\{J^{(N)} = j\}] + \frac{1}{2} (1 - p_j) \theta^2 N^{\beta-1} \mathbb{E}[M^{(N)} N^{-1} z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \\
&\quad + \frac{1}{2} (1 - p_j)^2 \theta^2 N^{\beta-1} \mathbb{E}[\{M^{(N)} N^{-1}\}^2 z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] + o(1).
\end{aligned}$$

すなわち, 次の式の絶対値を $N \rightarrow \infty$ としたときに, 0 に収束することを示せばよい.

$$\begin{aligned}
& N \left\{ \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] - \mathbb{E}[\{z(\theta) + (1 - z(\theta))N^{-1}(1 - p_j)\}^{M^{(N)}} 1\{J^{(N)} = j\}] \right\} \\
&\quad - (1 - p_j) \theta \mathbb{E} \left[\left\{ i M^{(N)} N^{\beta/2-1} + \frac{1}{2} \theta M^{(N)} N^{\beta-2} + \frac{1}{2} (1 - p_j) \theta \{M^{(N)}\}^2 N^{\beta-3} \right\} z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\} \right]. \tag{A.5}
\end{aligned}$$

ここで各 $j \in \mathbb{D}$ に対して, 次の式が成立する.

$$\begin{aligned}
& N \left\{ \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] - \mathbb{E}[\{z(\theta) + (1 - z(\theta))N^{-1}(1 - p_j)\}^{M^{(N)}} 1\{J^{(N)} = j\}] \right\} \\
&= N \mathbb{E} \left[z(\theta)^{M^{(N)}} \sum_{k=1}^{M^{(N)}} \binom{M^{(N)}}{k} (1 - z(\theta))^{-k} \{N^{-1}(1 - p_j)\}^k \{1 - N^{-1}(1 - p_j)\}^{M^{(N)}-k} 1\{J^{(N)} = j\} \right]. \tag{A.6}
\end{aligned}$$

よって, 次の不等式が得られる.

$$\begin{aligned}
& \left| N \left\{ \mathbb{E}[z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] - \mathbb{E}[\{z(\theta) + (1 - z(\theta))N^{-1}(1 - p_j)\}^{M^{(N)}} 1\{J^{(N)} = j\}] \right\} \right. \\
&\quad \left. - (1 - p_j) \theta \mathbb{E} \left[\left\{ i M^{(N)} N^{\beta/2-1} + \frac{1}{2} \theta M^{(N)} N^{\beta-2} + \frac{1}{2} (1 - p_j) \theta \{M^{(N)}\}^2 N^{\beta-3} \right\} z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\} \right] \right| \\
&= \left| - (1 - p_j)^2 N^{\beta-2} \mathbb{E}[M^{(N)} N^{-1} z(\theta)^{M^{(N)}} 1\{J^{(N)} = j\}] \right. \\
&\quad \left. + \frac{iN}{2} \mathbb{E} \left[z(\theta)^{M^{(N)}} \sum_{k=1}^{M^{(N)}} \binom{M^{(N)}}{k} \int_0^{-k\theta N^{\beta/2-1}} (k\theta N^{\beta/2-1} + u)^2 e^{iu} du \{N^{-1}(1 - p_j)\}^k \{1 - N^{-1}(1 - p_j)\}^{M^{(N)}-k} 1\{J^{(N)} = j\} \right] \right| \\
&\leq \frac{1}{2} (1 - p_j)^2 N^{\beta-2} \mathbb{E}[M^{(N)} N^{-1} 1\{J^{(N)} = j\}] \\
&\quad + (1 - p_j) \mathbb{E} \left[M^{(N)} \sum_{k=1}^{M^{(N)}} \binom{M^{(N)} - 1}{k-1} \frac{k^2 |\theta^3| N^{3/2\beta-3}}{6} \{N^{-1}(1 - p_j)\}^{k-1} \{1 - N^{-1}(1 - p_j)\}^{M^{(N)}-k} \right] \\
&\leq \frac{1}{2} (1 - p_j)^2 N^{\beta-2} \mathbb{E}[M^{(N)} N^{-1}] + (1 - p_j)^2 \frac{|\theta^3|}{2} N^{3/2\beta-3} \mathbb{E}[M^{(N)} (M^{(N)} - 1) N^{-2}] \\
&\quad + (1 - p_j)^3 \frac{|\theta^3|}{6} N^{3/2\beta-2} \mathbb{E}[M^{(N)} (M^{(N)} - 1) (M^{(N)} - 2) N^{-3}] + (1 - p_j) \frac{|\theta^3|}{6} N^{3/2\beta-2} \mathbb{E}[M^{(N)} N^{-1}].
\end{aligned}$$

ここで1つめの等号では, (A.6) 式の変形を利用した後, $z(\theta)^{-k}$ に対して (4.5) 式の変形を利用した. また, 1つめの不等号では, 三角不等式および $|e^{iu}| \leq 1$ を利用した. 上式の右辺は $N \rightarrow \infty$ とすると 0 に収束する. よって, (A.5) 式の絶対値は $N \rightarrow \infty$ としたときに, 0 に収束することがいえた. $\square$

Uplink coverage analysis for heterogeneous cellular networks

Takuya Kobayashi

Naoto Miyoshi

Department of Mathematical and Computing Sciences

Tokyo Institute of Technology

Abstract The analysis of multi-tier cellular networks (heterogeneous networks, HetNets) have been attracting attention because of the increasing variety, popularity of mobile applications, and network traffic. Most prior studies on such models assume that the base stations (BSs) are placed according to homogeneous Poisson Point Processes (PPPs); that is, their spatial correlation is ignored. For the downlink communication, a stochastic geometry model of downlink cellular networks was proposed in which the wireless BSs are deployed according to the Ginibre point process (GPP). The GPP can express repulsion between BSs. We should consider the configuration of BSs depending on the BS class since the strength of the received signal and the interference depends on the distance between nodes.

On the other hand, in the case of uplink communication, the transmit power of user equipment (UE) depends on its location because of the transmit power control. Furthermore, the interfering UEs construct a Voronoi perturbed lattice according to the configuration of BSs. Therefore the uplink analysis for HetNets is more complicated than the downlink one.

In this study, we analyze the uplink coverage probability for K -tier HetNet model, which is composed of the PPP, Hexagonal lattice, and GPP. In this model, we approximate the configuration of interfering UEs with inhomogeneous PPP based on the configuration of BSs. For this model, we investigate the impact of the power control parameters and the configuration of BSs on the coverage probability from the results of some numerical experiments.

1. Introduction

The multi-tier cellular networks (heterogeneous networks, HetNets) have been attracting attention because of the increasing variety and popularity of mobile applications. A multi-tier cellular network consists of multi-tiers of base stations (BSs), in which the BSs of each tier have a particular transmit power, intensity, path-loss exponent and configuration policy.

Recently the studies for downlink HetNets have been made a lot, e.g., Poisson multi-input multi-output (MIMO) HetNet models (see, [2, 3, 4]). However, in the case of Poisson point processes (PPPs), the wireless nodes are positioned independently of each other and their spatial correlation is ignored. For example, the configuration of the macrocells tends to be more regular than that of the femtocells. We should consider the configuration of BSs depending on the BS class since the strength of the received signal and the interference depends on the distance between nodes. By taking the correlation of BS deployments into account, we expect that the models will enable more accurate analysis to be conducted. The wireless network models considering the negative correlation between BSs is analyzed, in which the configuration of BSs follows Ginibre point process (GPP) or α -GPP (see e.g., [5, 6, 7, 8]). The GPP and α -GPP are a kind of Determinantal point processes (DPPs) that can be used to express repulsion between the points (see e.g., [10, 11, 12]). In the α -GPP, the parameter α can be used to interpolate smoothly between the PPP and the GPP. For example, the downlink heterogeneous cellular networks with multi-tiers was analyzed, in which the BSs are deployed according to α -GPP in [6]. Also the downlink cellular network model was proposed in [7] where the BSs are deployed according to the GPP over Nakagami- m fading. In the case of HetNet model, since the macro BSs are strategically placed, it is appropriate to express the configuration of the macro BSs using the α -GPP. On the other hand, since the femto BSs are arranged opportunistically, the PPP is sufficient to represent the configuration of the femto BSs.

On the other hand, in the case of uplink communication, the transmit power of user equipment (UE) depends on its location because of the transmit power control. Furthermore, the interfering UEs construct a Voronoi perturbed lattice according to the configuration of BSs. Therefore the uplink analysis for HetNets is more complicated than the downlink one and requires a different

approach. For the single-tier case, Novlan *et al.* [13] proposed a model wherein the locations of the UEs follow a PPP and the communication distances of the UEs from their associated BSs follow an independent identical distribution. Furthermore, the model proposed in [14] places the UEs in annular areas around a typical BS. The model of [13] is regarded as a UE-centric scenario, and the model of [14] is regarded as a BS-centric one.

In this study, we analyze the uplink coverage probability for K -tier HetNet model, which is composed of the PPP, Hexagonal lattice, and GPP. In this model, we approximate the configuration of interfering UEs with inhomogeneous PPP based on the configuration of BSs in a manner similar to [1]. For this model, we investigate the impact of the power control parameters and the configuration of BSs on the coverage probability from the results of some numerical experiments.

The rest of this paper is organized as follows. In Section 2, we describe our stochastic model of general K -tier HetNets and introduce the cell selection rule and the proposed approximation of the configuration of interfering UEs. The numerically computable representations of the coverage probability is described in Section 3. The results of numerical experiments are described in Section 4, and conclusion is given in Section 5.

2. K -tier cellular network models

Now we consider K -tier HetNets in which each BSs is assumed to belong to one of K distinct classes. Let Φ_k denote a point process with intensity λ_k on $\mathbb{R}^2$, $k \in \mathcal{K} = \{1, 2, \dots, K\}$. The point process Φ_k represents the configuration of the BSs of k th tier. We assume that the point process Φ_k , $k \in \mathcal{K}$, are simple (the probability that there are more than one point in the same place is 0), locally finite a.s. and also stationary. Therefore, the superposition $\Phi_B = \bigcup_{k \in \mathcal{K}} \Phi_k$ is locally finite a.s. and stationary. In the same way, let $\Phi_U = \{Y\}$ denote a point process on $\mathbb{R}^2$. the point process Φ_U represents the configuration of active UEs. Assuming that there is only one active UE in each cell, Φ_B and Φ_U have the same intensity. Therefore Φ_U constructs a Voronoi perturbed lattice based on Φ_B and cell selection rule. Since Φ_B is locally finite and stationary, Φ_U is also locally finite and stationary. The path-loss function ℓ representing the attenuation of the signal of the signal with distance r is given by $\ell(r) = r^{-2\beta}$, $r > 0$, where $\beta > 1$ is the path-loss exponent. The random effects of fading $\{H_{X,Y}\}$ are assumed to be Rayleigh distributed; that is, the $\{H_{X,Y}\}$, $X \in \Phi_B$, $Y \in \Phi_U$, are mutually independent and exponentially distributed with unit mean and are independent of the point processes Φ_U and Φ_B . Furthermore, let P_k denote the constant transmit power emitted by BS (k, i) , $i \in \mathbb{N}$.

The received power at the BS located at X from a typical UE located at Y is

$$P_{X,Y} = P_Y H_{X,Y} \ell(\|X - Y\|),$$

where P_Y denotes the transmit power emitted by UE located at Y . We assume the fractional power control for uplink transmission; that is, the transmit power P_Y is given by $P_Y = P_u (\ell(R_Y))^{-\epsilon}$, where $\epsilon \in [0, 1]$ denotes the power control fraction, P_u denotes the target received signal power at each BS and $R_Y = \|Y - X_Y\|$ denotes the communication distance between UE located at Y and BS connected with the UE. We can also regard R_Y as the distance between a typical UE and nearest BS of desired tier.

The signal-to-interference-ratio (SIR) of the typical user at the origin is given by

$$\text{SIR}_o = \frac{P_{X_o} H_{X_o,o} \ell(\|X_o\|)}{I_o(X_o)},$$

where X_o denotes the location of BS associated with typical UE at the origin and $I_o(X_o)$ denotes the cumulative interference received by BS X_o from all the UEs except typical UE at the origin; that is,

$$I_o(X_o) = \sum_{Y \in \Phi_U \setminus \{o\}} P_Y H_{X_o,Y} \ell(\|X_o - Y\|).$$

2.1. Flexible cell selection in Hetnets

In this paper, we assume that every UE use flexible cell selection rule for both uplink and downlink association; that is, a user selects to be served by the BS that provides the maximum biased average received power. Let $T_k = P_k b_k$, $k \in \mathcal{K}$, denote the uplink association weight for BSs of the k th tier, where b_k is the corresponding bias. Hence, noting the $\ell(r) = r^{-2\beta}$, $r > 0$, the event that a typical UE located at Y communicates with BS of the k th tier located at X_Y is given by

$$\begin{aligned} & \{T_k \ell(\|Y - X_Y\|) \geq T_m \ell(\|Y - X\|), X \in \Phi_m, m \in \mathcal{K}\} \\ & = \{\|Y - X\| \geq \left(\frac{T_m}{T_k}\right)^{\frac{\delta}{2}} \|Y - X_m\|, X \in \Phi_m, m \in \mathcal{K}\}, \end{aligned} \quad (1)$$

where $\delta = 1/\beta$.

2.2. Approximated interfering UE point process

It is difficult to express the interfering UE point process $\Phi_U \setminus \{o\}$ since Φ_U is a Voronoi perturbed lattice based on the superposition of distinct point processes. Therefore we approximate the point process with inhomogeneous Poisson point process in a manner similar to [1]. Conditioned on the BS being located at X and of tier k , the point process of interfering UEs from tier m to BS X , $\Phi_{U,m}$ is assumed to be Poisson point process with intensity measure

$$\Lambda_{U,m}(dx) = \lambda_m \left(1 - \mathbb{P}(\|X\| \geq \left(\frac{T_m}{T_k}\right)^{\frac{\delta}{2}} \|x\|, X \in \Phi_m, m \in \mathcal{K})\right) dx.$$

This approximation is based on thinning according to distance x between a transmit UE and BS receiving signal and their tiers.

3. Uplink coverage probability

The uplink coverage probability is defined as the probability with which the uplink SIR of the typical user located at the origin achieves a target threshold associated with the desirable BS; that is, $\mathbb{P}(\text{SIR}_o > \theta)$ for $\theta > 0$. Although $\{R_Y\}$, $Y \in \Phi_U$ are not independent because of their dependence on the BS configuration Φ_B , we assume that they are mutually independent. The independence assumption is appropriate enough [15].

3.1. General stationary BS deployments

Lemma 1. *Consider the K -tier downlink HetNet model, in which the tiers are mutually independent and BSs of tier k are deployed according to the stationary point process Φ_k on $\mathbb{R}^2$ with intensity λ_k , and we assume the channel model and the flexible cell selection described above. Then, the coverage probability for a typical user is given by*

$$\mathbb{P}(\text{SIR}_o > \theta) = \mathbb{E} \left[\prod_{Y \in \Phi_U \setminus \{o\}} \left(1 + \theta \left(\frac{R_Y^\epsilon R_o^{1-\epsilon}}{\|Y - X_o\|}\right)^{2\beta}\right)^{-1} \right]. \quad (2)$$

Proof. The coverage probability is given by

$$\begin{aligned} \mathbb{P}(\text{SIR}(X_o) > \theta) &= \mathbb{P}\left(\frac{P_u H_{X_o,o} \ell(R_o)^{1-\epsilon}}{I_o(X_o)} > \theta\right) \\ &= \mathbb{P}\left(H_{X_o,o} > \theta \frac{I_o(X_o)}{P_u \ell(R_o)^{1-\epsilon}}\right). \end{aligned} \quad (3)$$

Since $H_{X_o,o}$ is exponentially distributed,

$$\begin{aligned} & \mathbb{P}\left(H_{X_o,o} > \theta \frac{I_o(X_o)}{P_u \ell(R_o)^{1-\epsilon}} \mid \Phi_U, \Phi_B, \{H_{X,Y}\}, (X, Y) \neq (X_o, o)\right) \\ &= \exp\left(-\theta \frac{I_o(X_o)}{P_u \ell(R_o)^{1-\epsilon}}\right) \\ &= \prod_{Y \in \Phi_U \setminus \{o\}} \exp\left(-H_{X_o,Y} \theta \frac{\ell(\|Y - X_o\|)}{\ell(R_Y)^\epsilon \ell(R_o)^{1-\epsilon}}\right). \end{aligned} \quad (4)$$

Since $H_{X_o, Y}$, $Y \in \Phi_U \setminus \{o\}$, are mutually independent and the Laplace transform of $H_{X_o, Y}$ is given by $\mathcal{L}_H(s) = \mathbb{E}[\exp(-sH)] = (1+s)^{-1}$. Then,

$$\begin{aligned} & \mathbb{E} \left[\prod_{Y \in \Phi_U \setminus \{o\}} \exp \left(-H_{X_o, Y} \theta \frac{\ell(\|Y - X_o\|)}{\ell(R_Y)^\epsilon \ell(R_o)^{1-\epsilon}} \right) \middle| \Phi_U, \Phi_B \right] \\ &= \prod_{Y \in \Phi_U \setminus \{o\}} \mathbb{E} \left[\exp \left(-H_{X_o, Y} \theta \frac{\ell(\|Y - X_o\|)}{\ell(R_Y)^\epsilon \ell(R_o)^{1-\epsilon}} \right) \middle| \Phi_U, \Phi_B \right] \\ &= \prod_{Y \in \Phi_U \setminus \{o\}} \left(1 + \theta \frac{\ell(\|Y - X_o\|)}{\ell(R_Y)^\epsilon \ell(R_o)^{1-\epsilon}} \right)^{-1}. \end{aligned} \quad (5)$$

Hence, we obtain (2) by applying (4) and (5) to (3). $\square$

3.2. Single-tier cases

Theorem 1 (Poisson deployment). *Consider single-tier uplink cellular network model with path-loss function $\ell(r) = r^{-\beta}$, $r > 0$, in which BSs are deployed according to the homogeneous PPP with intensity λ_P , and we assume the channel model and the flexible cell selection described above. The coverage probability for the typical UE is given by*

$$\mathbb{P}(\text{SIR}_o > \theta) = \int_0^\infty \exp \left(-t - \int_0^\infty \int_0^\infty e^{-u} (1 - e^{-s}) (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right) dt. \quad (6)$$

Proof. The probability density function of R_Y is given by $f_R(r) = 2\pi\lambda r e^{-\pi\lambda r^2}$, $r > 0$. Since the intensity measure of interfering UEs is given by $\Lambda_U(dx) = \lambda(1 - e^{-\pi\lambda\|x\|^2})dx$, from lemma 1, the coverage probability is given by

$$\begin{aligned} & \mathbb{P}(\text{SIR}_o > \theta) \\ &= \int_0^\infty 2\pi\lambda r e^{-\pi\lambda r^2} \exp \left(- \int_{x \in \mathbb{R}^2} \left(1 - \int_0^\infty 2\pi\lambda y e^{-\pi\lambda y^2} (1 + \theta (\frac{y^\epsilon r^{1-\epsilon}}{\|x\|})^{2\beta})^{-1} dy \right) \Lambda_U(dx) \right) dr \\ &= \int_0^\infty 2\pi\lambda r e^{-\pi\lambda r^2} \exp \left(- 2\pi\lambda \int_0^\infty \left(1 - \int_0^\infty 2\pi\lambda y e^{-\pi\lambda y^2} (1 + \theta (\frac{y^\epsilon r^{1-\epsilon}}{z})^{2\beta})^{-1} dy \right) (1 - e^{-\pi\lambda z^2}) z dz \right) dr \\ &= \int_0^\infty e^{-t} \exp \left(- \int_0^\infty \left(1 - \int_0^\infty e^{-u} (1 + \theta (\frac{u^\epsilon t^{1-\epsilon}}{s})^\beta)^{-1} du \right) (1 - e^{-s}) ds \right) dt. \end{aligned} \quad (7)$$

$\square$

Theorem 2 (Hexagonal lattice deployment). *Consider single-tier uplink cellular network model with path-loss function $\ell(r) = r^{-\beta}$, $r > 0$, in which BSs are deployed according to the Hexagonal lattice with intensity λ_H , and we assume the channel model and the flexible cell selection described above. The coverage probability for the typical UE is given by*

$$\mathbb{P}(\text{SIR}_o > \theta) = \int_0^1 \exp \left(- \int_{s \geq 1} \int_{0 \leq u \leq 1} (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right) dt. \quad (8)$$

Theorem 3 (Ginibre deployment). *Consider single-tier uplink cellular network model with path-loss function $\ell(r) = r^{-\beta}$, $r > 0$, in which BSs are deployed according to the GPP with intensity λ_G , and we assume the channel model and the flexible cell selection described above. The coverage probability for the typical UE is given by*

$$\mathbb{P}(\text{SIR}_o > \theta) = \int_0^\infty f_G(t) \exp \left(- \int_0^\infty \int_0^\infty f_G(u) (1 - f_G(s)) (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right) dt, \quad (9)$$

where $f_G(t) = \sum_{i \in \mathbb{N}} \frac{t^{i-1} e^{-t}}{(i-1)!} \prod_{j \neq i} Q(j, t)$, and $Q(j, t) = \int_t^\infty \frac{u^{i-1} e^{-u}}{(i-1)!} du$ denotes the regularized incomplete gamma function.

Remark 1. The coverage probability for the α -Ginibre deployment case is given by replacing $f_G(t)$ in Theorem 3 with

$$f_G(t) = \alpha \sum_{i \in \mathbb{N}} \frac{t^{i-1} e^{-t}}{(i-1)!} \prod_{j \neq i} (1 - \alpha + \alpha Q(j, t)).$$

3.3. Multi-tier cases

Theorem 4 (Hexagonal lattice-Poisson overlaid BS deployments). Consider 2-tier uplink cellular network model with path-loss function $\ell(r) = r^{-\beta}$, $r > 0$, in which BSs of tier $H \in \mathcal{K}$ (tier $P \in \mathcal{K}$) are deployed according to the Hexagonal lattice (Poisson point process) with intensity λ_H (λ_P), and we assume the channel model and the flexible cell selection described above. The coverage probability for the typical UE is given by

$$\begin{aligned} \mathbb{P}(\text{SIR}_o > \theta) = & \mathcal{A}_H \int_0^1 \exp \left(- \int_{s>0} \int_{0 \leq u \leq 1} (1 - \mathbf{1}_{s \leq 1} e^{-s \frac{c_P}{c_H}}) (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right. \\ & \left. - \int_{s>0} \int_{u>0} e^{-u} (1 - \mathbf{1}_{s \leq \frac{c_P}{c_H}} (1 - s \frac{c_H}{c_P}) e^{-s}) (1 + \theta^{-1} (\frac{s}{u^\epsilon (\frac{\lambda_P}{\lambda_H} t)^{1-\epsilon}})^\beta)^{-1} du ds \right) dt \\ & + \mathcal{A}_P \int_0^\infty \exp \left(- \int_{s>0} \int_{0 \leq u \leq 1} (1 - \mathbf{1}_{s \leq 1} e^{-s \frac{c_P}{c_H}}) (1 + \theta^{-1} (\frac{s}{u^\epsilon (\frac{\lambda_P}{\lambda_H} t)^{1-\epsilon}})^\beta)^{-1} du ds \right. \\ & \left. - \int_{s>0} \int_{u>0} e^{-u} (1 - \mathbf{1}_{s \leq \frac{c_P}{c_H}} (1 - s \frac{c_H}{c_P}) e^{-s}) (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right) dt, \quad (10) \end{aligned}$$

where $c_H = \lambda_H T_H^\delta$, $c_P = \lambda_P T_P^\delta$, $\mathcal{A}_H = \frac{c_H}{c_P} (1 - e^{-\frac{c_P}{c_H}})$, $\mathcal{A}_P = 1 - \mathcal{A}_H$.

Theorem 5 (Ginibre-Poisson overlaid BS deployments). Consider 2-tier uplink cellular network model with path-loss function $\ell(r) = r^{-\beta}$, $r > 0$, in which BSs of tier $G \in \mathcal{K}$ (tier $P \in \mathcal{K}$) are deployed according to the Ginibre point process (Poisson point process) with intensity λ_G (λ_P), and we assume the channel model and the flexible cell selection described above. The coverage probability for the typical UE is given by

$$\begin{aligned} \mathbb{P}(\text{SIR}_o > \theta) = & \mathcal{A}_G \int_0^\infty \exp \left(- \int_{s>0} \int_{u>0} f_G(u) (1 - f_G(s) e^{-s \frac{c_P}{c_G}}) (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right. \\ & \left. - \int_{s>0} \int_{u>0} e^{-u} (1 - e^{-s} \bar{F}_G(s \frac{c_G}{c_P})) (1 + \theta^{-1} (\frac{s}{u^\epsilon (\frac{\lambda_P}{\lambda_G} t)^{1-\epsilon}})^\beta)^{-1} du ds \right) dt \\ & + \mathcal{A}_P \int_0^\infty \exp \left(- t - \int_{s>0} \int_{u>0} e^{-u} (1 - f_G(s) e^{-s \frac{c_P}{c_G}}) (1 + \theta^{-1} (\frac{s}{u^\epsilon (\frac{\lambda_P}{\lambda_G} t)^{1-\epsilon}})^\beta)^{-1} du ds \right. \\ & \left. - \int_{s>0} \int_{u>0} e^{-u} (1 - e^{-s} \bar{F}_G(s \frac{c_G}{c_P})) (1 + \theta^{-1} (\frac{s}{u^\epsilon t^{1-\epsilon}})^\beta)^{-1} du ds \right) dt, \quad (11) \end{aligned}$$

where $c_G = \lambda_G T_G^\delta$, $\mathcal{A}_G = \int_0^\infty f_G(t) \exp(-t \frac{c_P}{c_G}) dt$, $\mathcal{A}_P = 1 - \mathcal{A}_G$.

4. Numerical experiments

Here, we show the results of numerical experiments on computing the coverage probabilities. Fig. 1 and Fig. 2 show the coverage probabilities for given values of the SINR threshold θ in 3 models: Hexagonal lattice, PPP and Ginibre with the pair of path-loss exponent β and compensation factor ϵ given as $(\beta, \epsilon) = (2, 1)$ and $(1.625, 0.75)$, respectively. From these figures, the uplink coverage probability is achieved when the repulsion degree of the configuration of BSs increases.

5. Conclusion

We proposed the multi-tier uplink cellular network models. For this model, we analyze the coverage probabilities and investigate the impact of the fractional power control and varying the repulsion by using PPP, Hexagonal lattice, and GPP. We approximate the configuration of interfering UEs

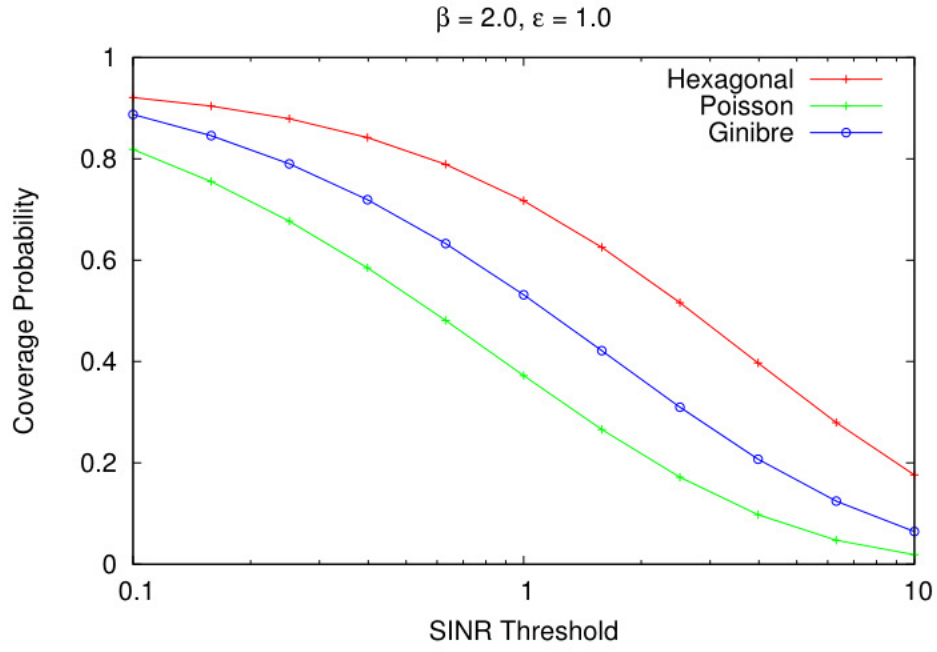


Figure 1: Comparison of uplink coverage probability of a single-tier cellular networks between the Hexagonal lattice, PPP, and GPP, where $\beta = 2$, $\epsilon = 1$.

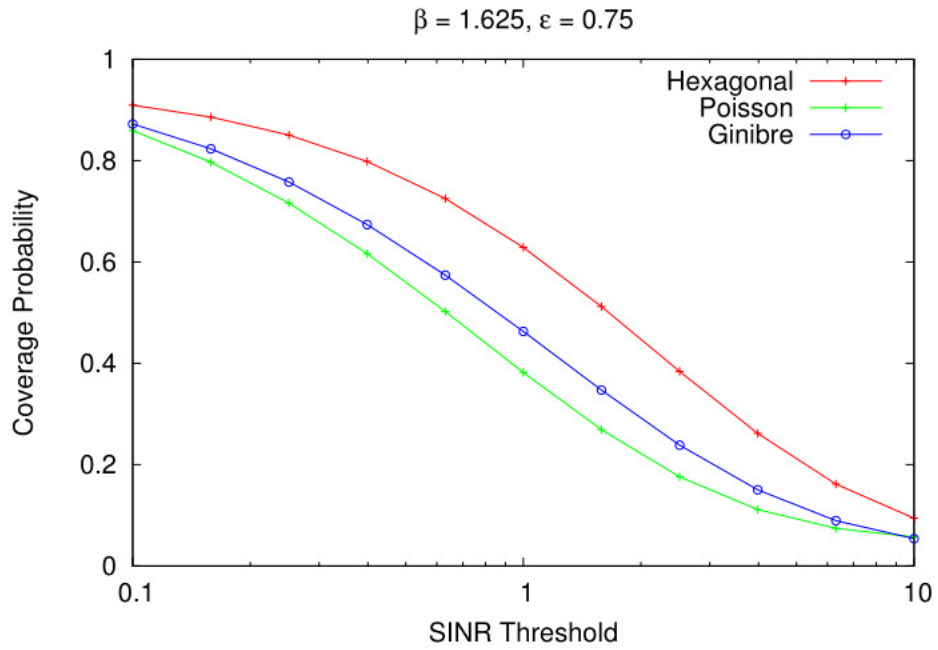


Figure 2: Comparison of uplink coverage probability of a single-tier cellular networks between the Hexagonal lattice, PPP, and GPP, where $\beta = 1.625$, $\epsilon = 0.75$.

with inhomogeneous Poisson point process based on the configuration of BSs. Furthermore, the uplink coverage probability is achieved when the repulsion degree of the BSs increases.

However, from the complexity of the GPP models, the numerical calculation of the performance indices of the GPP models takes a longer time than that of PPP and Hexagonal lattice models. In [8], the approximative approaches for downlink HetNets are proposed (see, [8, 9]).

References

- [1] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in HetNets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, 2015.
- [2] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 5208–5222, 2013.
- [3] A. K. Gupta, H. S. Dhillon, S. Vishwanath, and J. G. Andrews, "Downlink multi-antenna heterogeneous cellular network with load balancing," *IEEE Trans. Commun.*, vol. 62, pp. 4052–4067, 2014.
- [4] —, "Downlink coverage probability in MIMO HetNets with flexible cell selection," *2014 IEEE GLOBECOM*, pp. 1534–1539, 2014.
- [5] N. Miyoshi and T. Shirai, "A cellular network model with Ginibre configured base stations," *Adv. Appl. Probab.*, vol. 46, pp. 832–845, 2014.
- [6] I. Nakata and N. Miyoshi, "Spatial stochastic models for analysis of heterogeneous cellular networks with repulsively deployed base stations," *Perform. Eval.*, vol. 78, pp. 7–17, 2014.
- [7] N. Miyoshi and T. Shirai, "Downlink coverage probability in a cellular network with Ginibre deployed base stations and Nakagami- m fading channels," *WiOpt 2015*, pp. 483–489, 2015.
- [8] H. Wei, N. Deng, W. Zhou, and M. Haenggi, "A simple approximative approach to the SIR analysis in general heterogeneous cellular networks," *2015 IEEE GLOBECOM*, 2015.
- [9] R. K. Ganti and M. Haenggi, "Asymptotics and approximation of the SIR distribution in general cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 2130–2143, 2015.
- [10] T. Shirai and Y. Takahashi, "Random point fields associated with certain Fredholm determinants I: Fermion, Poisson and Boson processes," *J. Funct. Analysis*, vol. 205, pp. 414–463, 2003.
- [11] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. American Math. Soc., 2009.
- [12] A. Goldman, "The Palm measure and the Voronoi tessellation for the Ginibre process," *Annals Appl. Probab.*, vol. 20, pp. 90–128, 2010.
- [13] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Trans. on Wireless Comm.*, vol. 12, pp. 2669–2679, 2013.
- [14] M. Coupechoux and J. Kelif, "How to set the fractional power control compensation factor in LTE?," *IEEE Sarnoff Symp.*, pp. 1–5, 2011.
- [15] T. Kobayashi, and N. Miyoshi, "Uplink cellular network models with Ginibre deployed base stations," *26th International Teletraffic Congress (ITC26)*, pp. 1–7, 2015.

A. Ginibre and α -Ginibre point processes

In this section, we define the GPP and α -GPP, then describe their tractable properties. Figure. 3 show the voronoi tessellations of samples of a PPP and GPP. From Fig. 3, we can see that the configuration of the GPP is more regular than that of PPP.

A.1. Ginibre point process

The GPP is a DPP on the complex plane $\mathbb{C}$, which can express repulsion between the points of each other. Let Φ denote a simple point process on $\mathbb{C}$ and $\rho_n: \mathbb{C}^n \rightarrow \mathbb{R}_+$, $n \in \mathbb{N}$, denote its joint

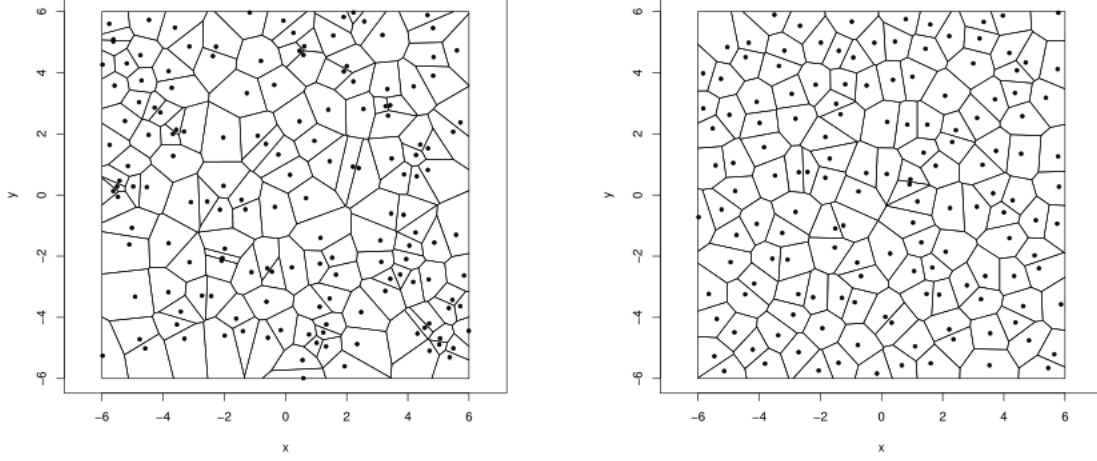


Figure 3: Voronoi tessellations of samples of a PPP (left) and GPP (right).

intensities with respect to some locally finite measure ν on $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$. That is, for any continuous symmetric function f on $\mathbb{C}^n$ with compact support, we have

$$\mathbb{E} \left[\sum_{\substack{X_1, \dots, X_n \in \Phi \\ \text{distinct}}} f(X_1, X_2, \dots, X_n) \right] = \int_{\mathbb{C}^n} f(z_1, \dots, z_n) \rho_n(z_1, \dots, z_n) \nu(dz_1) \cdots \nu(dz_n).$$

The point process Φ is said to be a determinantal point process with kernel $K: \mathbb{C}^2 \rightarrow \mathbb{C}$ with respect to the reference measure ν if the ρ_n , $n \in \mathbb{N}$, satisfy

$$\rho_n(z_1, \dots, z_n) = \det(K(z_i, z_j))_{1 \leq i, j \leq n},$$

for $z_1, \dots, z_n \in \mathbb{C}$, where $\det$ denotes the determinant. Furthermore, Φ is said to be the GPP if the kernel K is given as following definition.

Definition 1 (Ginibre point process). *Φ is said to be the GPP if the kernel K is given by $K(z, w) = e^{z\bar{w}}$, $z, w \in \mathbb{C}$, with respect to the Gaussian measure $\nu(dz) = \pi^{-1} e^{-|z|^2} m(dz)$, where $\bar{w}$ denotes the complex conjugate of $w \in \mathbb{C}$ and m denotes the Lebesgue measure on $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$.*

This choice of pair of K and ν is not unique. Indeed, the determinantal point process associated with the kernel $\tilde{K}(z, w) = \pi^{-1} e^{-(|z|^2 + |w|^2)/2} e^{z\bar{w}}$ with respect to $\tilde{\nu}(dz) = m(dz)$ coincides with the GPP. From this expression, it is easy to see that $\tilde{\rho}_n(z_1, \dots, z_n) = \det(\tilde{K}(z_i, z_j))_{1 \leq i, j \leq n}$ is motion-invariant, or equivalently, the GPP is motion-invariant (stationary and isotropic). From its definition, we see that $\mathbb{E}[\Phi(C)] = m(C)/\pi$ for $C \in \mathcal{B}(\mathbb{C})$; i.e. the (first order) intensity is equal to π^{-1} with respect to the Lebesgue measure. In order to adjust the intensity, we consider a scaled GPP Φ_c with a scaling parameter $c > 0$. Φ_c is given by the kernel $K_c(z, w) = c e^{cz\bar{w}}$ with respect to the reference measure $\nu_c(dz) = \pi^{-1} e^{-c|z|^2} m(dz)$, or equivalently, $\tilde{K}_c(z, w) = (c/\pi) e^{-c(|z|^2 + |w|^2)/2} e^{cz\bar{w}}$ with respect to the Lebesgue measure. Since $\mathbb{E}[\Phi_c(C)] = (c/\pi) m(C)$, Φ_c has the intensity c/π .

One of the useful properties of the GPP comes from its radial symmetry, and it is described as follows.

Proposition 1 (Kostlan [?]). *Let X_i , $i \in \mathbb{N}$, denote the points of GPP Φ . Then, the set $\{\|X_i\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{Z_i}\}_{i \in \mathbb{N}}$, where the Z_i , $i \in \mathbb{N}$, are mutually independent and each Z_i follows the i th Erlang distribution with a unit rate parameter, denoted by $Z_i \sim \text{Gam}(i, 1)$, $i \in \mathbb{N}$.*

Proposition 1 can be extended to scaled GPP Φ_c ; i.e., for the points X_i , $i \in \mathbb{N}$, of Φ_c , the set $\{\|X_i\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{Z_i}\}_{i \in \mathbb{N}}$, where Z_i , $i \in \mathbb{N}$, are mutually independent and each Z_i follows the i th Erlang distribution with rate parameter c , denoted by $Z_i \sim \text{Gam}(i, c)$, $i \in \mathbb{N}$.

A.2. α -Ginibre point process

As is the usual GPP, the α -GPPs are determinantal point process on the complex plane $\mathbb{C}$ defined as follows.

Definition 2 (α -Ginibre point process). Φ^α is said to be an α -GPP with $\alpha \in (0, 1]$ if the kernel K^α is given by $K^\alpha(z, w) = e^{z\bar{w}/\alpha}$, $z, w \in \mathbb{C}$, with respect to the scaled Gaussian measure $\nu^\alpha(dz) = \pi^{-1} e^{-|z|^2/\alpha} m(dz)$.

This choice of pair of K^α and ν^α is not unique and the determinantal point process associated with the kernel $\tilde{K}^\alpha(z, w) = \pi^{-1} e^{-(|z|^2+|w|^2)/2} e^{z\bar{w}}$ with respect to $\tilde{\nu}^\alpha(dz) = m(dz)$ coincides with the α -GPP. The usual GPP is just the one with $\alpha = 1$ and the α -GPP converges weakly to the PPP of intensity $1/\pi$ as $\alpha \rightarrow 0$ (see [12]). Therefore, the α -GPP constitute an intermediate class between the PPP and GPP by regulating the value of $\alpha \in (0, 1]$. As is usual GPP, it can be seen that the GPP is motion-invariant and the intensity is equal to π^{-1} . In order to adjust the intensity, we consider a scaled α -GPP Φ_c^α with a scaling parameter $c > 0$. Φ_c^α is given by the kernel $K_c^\alpha(z, w) = c e^{cz\bar{w}/\alpha}$ with respect to the reference measure $\nu_c^\alpha(dz) = \pi^{-1} e^{-c|z|^2/\alpha} m(dz)$, or equivalently, $\tilde{K}_c^\alpha(z, w) = (c/\pi) e^{-c(|z|^2+|w|^2)/2\alpha} e^{cz\bar{w}/\alpha}$ with respect to the Lebesgue measure. Since $\mathbb{E}[\Phi_c^\alpha(C)] = (c/\pi) m(C)$, Φ_c^α has the intensity c/π .

Due to the radial symmetry of α -GPP, we can extend Proposition 1 to the scaled α -GPP.

Proposition 2. Let X_i , $i \in \mathbb{N}$, denote the points of α -GPP Φ_c^α with intensity c/π . Then, the set $\{\|X_i\|\}_{i \in \mathbb{N}}$ has the same distribution as $\{\sqrt{\tilde{Z}_i}\}_{i \in \mathbb{N}}$, which is constructed from $\{Z_i\}_{i \in \mathbb{N}}$ such that Z_i , $i \in \mathbb{N}$, are mutually independent and each Z_i follows the i th Erlang distribution with a rate parameter c/π , denoted by $Z_i \sim \text{Gam}(i, c/\alpha)$, $i \in \mathbb{N}$ and it is included in $\{\tilde{Z}_i\}_{i \in \mathbb{N}}$ with probability α independently of others.

From Proposition 2, we can construct the α -GPP Φ_c^α with intensity c/π from the usual GPP $\Phi_{c/\alpha} = \{\bar{X}_i\}_{i \in \mathbb{N}}$ with intensity $c/\alpha\pi$ by independent α -thinning; that is, by deleting each point of GPP with probability $1 - \alpha$ independently. Let $\{\xi_i\}_{i \in \mathbb{N}}$ denote the set of marks of $\Phi_{c/\alpha}$ such that ξ_i , $i \in \mathbb{N}$, are mutually independent and identically distributed as $\mathbb{P}(\xi_i = 1) = \alpha$ and $\mathbb{P}(\xi_i = 0) = 1 - \alpha$. Then, the α -GPP Φ_c^α is obtained by

$$\Phi_c^\alpha(C) = \sum_{i \in \mathbb{N}} \xi_i \mathbf{1}(\bar{X}_i \in C), \quad C \in \mathcal{B}(\mathbb{C}).$$

セルラネットワークにおける移動車両の通信精度解析手法

徳山喜一

東京工業大学大学院 情報理工学研究科

概要： 本論文では、無線通信のセルラネットワークにおいて電波を受信する端末が、直線上を一定ではない速度で移動する場合にも対応可能な通信の性能判定手法を提案する。

移動物体の無線通信の受信性能判定で考慮される主要素として、通信のデータ容量変化とハンドオフに係るコストの2つが挙げられるが、物体の移動速度の変化に対応し、これら2つの量を導出することで、軌道上を不規則な速度で移動する物体に対しても受信性能判定を行うことができる。特に、移動する物体の受信データ容量の変化をとらえることが従来研究で困難な課題の一つであるが、歩行者の移動よりも比較的平均速度が高い車両の移動を想定した場合に、直線軌道上を不規則な速度で移動する物体の受信データ容量の変化を近似可能な手法を提案した。本手法の評価では、同一の平均速度を保ちながら、軌道上を終始等速で移動する場合とそうでない場合とで、性能判定結果に起こる影響を観察した。

1. はじめに

本論文では、点過程理論を応用したセルラ方式の無線通信ネットワーク（セルラネットワーク）のモデリングを取り扱う。

近年、セルラネットワーク上で無線電波を受信する際の性能解析に点過程理論を応用する研究が盛んに行われている。無線通信を受信する端末が任意のある一点に配置されたとき、無線通信の基地局からその端末に発信される電波の受信容量を計算する手法については、解析的に計算する手法が様々に確立されている [1], [2]。一方で、通信を受信する端末が任意の軌道に沿って移動する際は、移動により発生する受信容量の変化についての解析が困難であり、これを計算する手法の確立が課題となっている。また、セルラネットワークの基地局は互いにセルと呼ばれる通信区画を生成し、各セルに含まれる受信端末を対象として無線通信を行うが、受信端末が移動する際はセル同士の境界線を跨ぐことが想定されるため、それに伴い通信先の基地局を変更する処理（ハンドオフ）を考慮する必要がある。

これらの問題に取り組んだ研究として代表的なものが Wei らによる結果である [3]。彼らは、複数の規格の無線通信が混在するセルラネットワーク（ヘテロジニアスネットワーク）上で、通信端末が任意の軌道を描いて移動する場合に、通信端末の移動に伴い発生するハンドオフの平均回数を解析的に求める手法を提案した。さらに、それを毎秒ごとの平均ハンドオフレートへと換算し、平均電波受信容量と平均ハンドオフレートの双方を考慮した通信の評価関数を与え、評価関数が最大値をとるハンドオフ方法を導出した。一方 Wei らは、通信端末における毎秒ごとの平均電波受信容量は通信端末の位置によらず一定としていた。毎秒ごとのハンドオフレートは物体の移動する速度 v に依存するが、毎秒ごとの平均電波受信容量を一定としている場合は、それに対し比較されるハンドオフレートを v に関して不変とする必要があるため、物体は軌道の上を一定の速度で動くことが求められた。

そこで本研究では、物体の移動に伴う変化を踏まえた平均電波受信容量を導出することで、軌道上を一定でない速度で移動する通信端末に対しても、平均電波受信容量と平均ハンドオフレートによる評価を行えるよう、既存手法を改良した。これにより、通信端末の移動における速度の変化が及ぼす影響の観察を試みる。特に本論文では、ある平均速度で軌道上を通信端末が移動する際に、常に同じ速度で移動する場合と、軌道上で速度を変化させながら移動する場合との間で生じる変化についての観察を試みる。今回、直線軌道上を指数分布に従って速度を変化させながら移動する場合と、一定速度の場合との間で、通信の評価関数に一定の差がみられることを観察することができた。本結果は、軌道上で物体が速度を変化させながら移動するという、実際の移動物体の動きにより類似した状況下での通信の評価を可能とした。

本研究で考察する通信端末の移動は、直線上を一方方向へ移動するものである。また、移動物体の速度と移動距離は、車両の移動に伴うスケールのもを想定している。将来、自動運転車の導入等も相まって、無線通信端末を搭載した自動車の移動を想定した性能判定の手法提案が求められていると考えられ、本研究は、以上の背景を動機として行われている。

2. システムモデル

2.1. ネットワークモデル

セルラネットワークを構成する無線通信規格は 1 種類のみホモジニアスネットワークを想定し、基地局は intensity を λ に持つ定常ポワソン点過程 Φ に従って配置されるとする。セルラネットワーク上の通信端末は最も近い位置に存在する基地局と通信することを仮定した場合、基地局は、それらの配置がなすポワソンボロノイセルの内部に含まれる通信端末に対し無線通信を行うこととなる。ここで、intensity が λ のポワソン点過

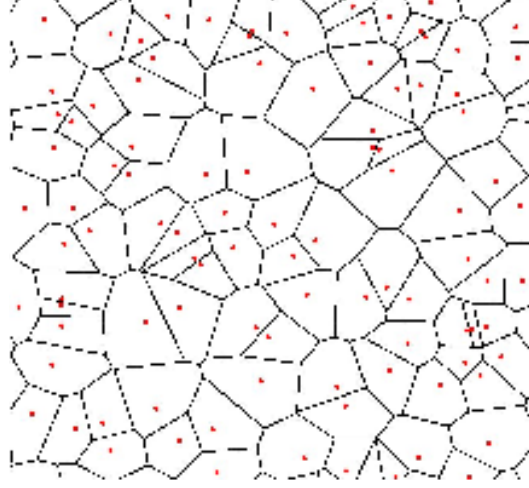


図 1: ポワソン点過程がなすポワソンボロノイセルの例

程がなすポワソンボロノイセルにおいて、セルを構成する辺の長さは平均 $2/(3\sqrt{\lambda})$ のガンマ分布に従い、またそれらの辺の中心点の分布は intensity が 3λ のポワソン点過程に従うとみなせることが知られている [4], [5]。

通信端末が基地局から電波を受信する際、その電波の強度は $\alpha \geq 2$ の指数オーダーで減衰するとする。また、受信電波の強度はフェーディング効果によっても左右される。そのフェーディング効果は平均 1 の Rayleigh Fading を仮定し、また、基地局が発信する電波強度は一律に $1/\mu$ であるとする。このとき、通信端末が、距離が r だけ離れた基地局から電波を受信する際の受信強度は $hr^{-\alpha}$ と表せられる。ここで、 h は平均 μ の指数分布に従う確率変数である。

また、通信先以外の基地局から発せられる障害電波（インターフィアレンス）も同様に平均 μ の指数分布に従うとする。通信端末に対するインターフィアレンス強度 I_r と SINR は、それぞれ以下のように表される。

$$I_r = \sum_{i \in \Phi \setminus \{b_0\}} g_i R_i \quad (1)$$

$$\text{SINR} = \frac{hr^{-\alpha}}{\sigma^2 + I_r} \quad (2)$$

ここで、 $b_0 \in \Phi$ は通信端末の通信先である基地局の配置を表す。また、通信端末へインターフィアレンスを発信する基地局 $i \in \Phi \setminus \{b_0\}$ は、通信端末からの距離が R_i であり、電波へ影響するフェーディング効果は指数分布 $g_i \sim \exp(\mu)$ に従っている。

通信端末とその通信先である基地局との間の距離 r が従う分布関数 $F(r)$ は

$$F(r) = P\{r \leq R\} = 1 - e^{-\lambda\pi r^2} \quad (3)$$

である。よって、その密度関数 $f(r)$ は

$$f(r) = \frac{dF(r)}{dr} = e^{-\lambda\pi r^2} 2\pi\lambda r \quad (4)$$

と表せられる。

2.2. 通信端末の移動

直線軌道上にある規則に従って速度を変化させながら通信端末が移動するモデルを考える。通信端末が移動する直線軌道は、原点を軌道の出発点として直交座標における横軸の正方向へ移動するものとしてよい。以後、通信端末が移動する長さ L の直線軌道を $TR(L)$ と呼称する。

$TR(L)$ 上に k 個の点 $p_1, \dots, p_k$ を、一次元点過程 Ψ に従って配置する。通信端末は、点 p_i, p_{i+1} 間の距離 $l_i (i = 1, \dots, k-1)$ をそれぞれ s 秒で移動するように $TR(L)$ 上を移動する。なお、 p_i, p_{i+1} の間の区間を通信端末はそれぞれ等速で移動する。 Ψ が一様点過程の場合は、通信端末は $TR(L)$ 上を等速で移動する。また、 $TR(L)$ 上で速度を変化させながら移動する例として、 Ψ が intensity を ν とした一次元ポワソン点過程に従う場合が考えられる。なお、 s は $s \geq 10$ をみたとすることを前提とする。

3. ハンドオフレートと受信容量、通信の評価関数

ここでは、 $TR(L)$ 上の区間 $p_i, p_{i+1} (i = 1, \dots, k-1)$ を物体が移動する際に発生するハンドオフレート、また、受信データ容量の変化量を解析的に導出する手法を述べる。ここで、ハンドオフレートとは、区間 $p_i, p_{i+1} (i = 1, \dots, k-1)$ の間を通信端末が移動した際に発生するハンドオフの平均回数を指すものとする。また、受信データ容量の変化量とは、区間 $p_i, p_{i+1} (i = 1, \dots, k-1)$ を通信端末が移動する際、ハンドオフを行った場合と行わなかった場合との間で発生する受信データ容量の差分値のことである。これら 2 つの量を用いて、通信の評価関数を定義する。

以後、区間 p_i, p_{i+1} の距離 l_i の添字を一般化し、 l と表すこととする。

3.1. ハンドオフレートとその解析

proposition 1 基地局の配置が intensity を λ とする定常ポワソン点過程 Φ に従うとき、距離 l の直線上を移動する通信端末が通過するポワソンボロノイセルの境界面の数の平均、すなわちハンドオフレート $\mathcal{N}(\lambda, l)$ は、以下のように求まる。

$$\mathcal{N}(\lambda, l) = \frac{\sqrt{\lambda}}{4\pi} l \quad (5)$$

証明 Φ によるポワソンボロノイセルを構成する辺の中心点の集合を Θ 、辺の長さを X 、また辺が座標横軸の正方向となす角度を ϕ とする ($0 \leq \phi \leq \pi$)。 Θ は intensity が 3λ の定常ポワソン点過程であるため [4], [5]、長さ l の直線の一方の端点は原点に、もう一方の端点は座標 $(l, 0)$ に配置してよい。このとき、ハンドオフレート $\mathcal{N}(\lambda, l)$ は、 Θ の元を中心点に持つポワソンボロノイセルの辺と長さ l の直線との間での交点の数に等しい。ゆえに、 $\mathcal{N}(\lambda, l)$ は、縦長を $l + \frac{X}{2} \cos \phi$ 、横長を $X \sin \phi$ とする長方形の領域 B の内部に存在する、定常ポワソン点過程 Θ の元の個数の平均である。 X の従う分布の密度関数を $g(x)$ とすれば、領域 B の平均面積 $|B|$ は、

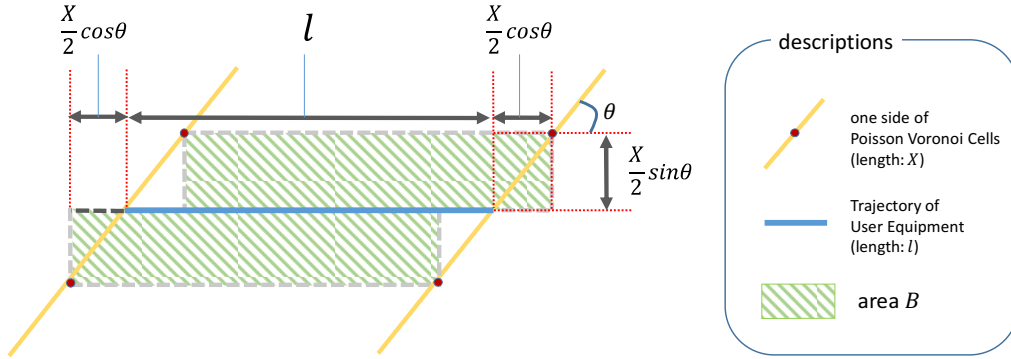


図 2: 領域 B の概念図

$$\begin{aligned} |B| &= \frac{1}{\pi} \int_0^\pi \int_0^\infty g(x) \left(l + \frac{x}{2} \cos \phi \right) x \sin \phi \, dx d\phi \\ &= \frac{1}{\pi} \int_0^\infty g(x) \left(\int_0^\pi l x \sin \phi d\phi + \int_0^\pi \frac{x}{4} \sin 2\phi d\phi \right) dx \\ &= \frac{1}{\pi} \int_0^\infty g(x) (2lx + 0) dx \\ &= \frac{1}{\pi} \bar{X} l \\ &= \frac{4l}{3\pi\sqrt{\lambda}} \end{aligned} \quad (6)$$

となる。ここで、 $\bar{X} = \frac{2}{3\sqrt{\lambda}}$ 、 $\phi \sim U(0, \pi)$ であることを用いた [4], [5]。ゆえに、

$$\begin{aligned}\mathcal{N}(\lambda, l) &= \sum_{n=0}^{\infty} nP(\Theta(|B|) = n) \\ &= 3\lambda |B| \\ &= \frac{4\sqrt{\lambda}}{\pi} l\end{aligned}\quad (7)$$

となる。 $\square$

なお、式 (5) の結果は、Wei らの結果 [3] をホモジニアスネットワークの場合に置き換えたものと合致する。

3.2. 平均受信データ容量の変化量の解析

平均受信データ容量の変化量 $\Delta D(s, l, \lambda, \alpha)$ は、区間 $p_i, p_{i+1} (i = 1, \dots, k-1)$ を通信端末が移動する際、ハンドオフを行わなかった場合の平均受信データ容量 $D_1(l, \lambda, \alpha)$ を、ハンドオフを行った場合の平均受信データ容量 $D_2(s, \lambda, \alpha)$ から差し引いた量、すなわち

$$\Delta D(s, l, \lambda, \alpha) = D_2(s, \lambda, \alpha) - D_1(l, \lambda, \alpha) \quad (8)$$

をみたす量である。 $D_1(l, \lambda, \alpha)$ 、 $D_2(s, \lambda, \alpha)$ はそれぞれ以下のように求まる。

proposition 2 基地局の配置が *intensity* を λ とする定常ポワソン点過程 Φ に従うとき、距離 l の直線上でハンドオフを一度も行わずに通信端末が移動を行ったときの平均受信データ容量 $D_1(l, \lambda, \alpha)$ は、近似的に以下のように求まる。

$$D_1(l, \lambda, \alpha) = \frac{1}{l} \int_0^l \frac{1}{2\pi} \int_0^{2\pi} \tau(\lambda, \alpha, \theta, u) d\theta du \quad (9)$$

ここで、

$$\tau(\lambda, \alpha, \theta, u) = \int_{r>0} r e^{-\pi\lambda r^2} \int_{t>0} e^{-\sigma^2 \mu r'^{\alpha} (e^t - 1)} \mathcal{L}_{I_{r'}}(\mu r'^{\alpha} (e^t - 1)) dt dr \quad (10)$$

$$\mathcal{L}_{I_{r'}}(\mu r'^{\alpha} (e^t - 1)) = \frac{\exp\left(-2\pi\lambda \int_0^{\infty} \frac{v}{1+v^{\alpha}\{r'^{\alpha}(e^t-1)\}^{-1}} dv\right)}{\int_0^{\infty} \frac{w e^{-\lambda\pi w^2}}{1+w'^{\alpha}\{r'^{\alpha}(e^t-1)\}} dw} \quad (11)$$

$$r' = \sqrt{r^2 + u^2 - 2ru\cos\theta} \quad (12)$$

$$w' = \sqrt{w^2 + u^2 - 2wu\cos\theta} \quad (13)$$

である。

証明 *proposition 1* と同様、長さ l の直線上を通信端末が移動をする出発点を原点、到達点を $(l, 0)$ としてよい。点 $H(u)$ を、座標 $(u, 0)$ 上の点とする ($0 \leq u \leq l$)。通信端末が原点に存在するときの通信先の基地局 b_0 と原点との距離、 $H(u)$ との距離をそれぞれ r 、 r' とする。原点と基地局 b_0 とを結ぶ直線が横軸正方向となす角を θ とする。

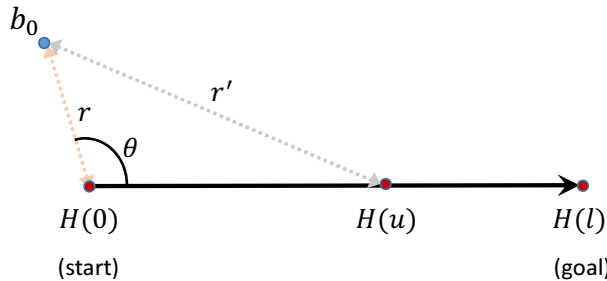


図 3: 移動端末と基地局との位置関係

$H(u)$ が基地局 b_0 と通信するときのインターフィアレンスの強度 $I_{r'}$ を

$$I_{r'} = \sum_{i \in \Phi/b_0} g_i R_i^{-\alpha} \quad (14)$$

とおく。ここで、 R_i は $H(u)$ と $i \in \Phi/b_0$ との距離、 $g_i \sim \exp(\mu)$ は $i \in \Phi/b_0$ からの障害電波におけるフェーディングである。

$I_{r'}$ の Laplace 変換 $\mathcal{L}_{I_{r'}}(s)$ は以下のように計算される。

$$\begin{aligned}
\mathcal{L}_{I_{r'}}(s) &= \mathbf{E}_\Phi \left[\prod_{i \in \Phi/b_0} \mathbf{E}_g [\exp(-sgR_i^{-\alpha})] \right] \\
&= \frac{\mathbf{E}_\Phi [\prod_{i \in \Phi} \mathbf{E}_g [\exp(-sgR_i^{-\alpha})]]}{\mathbf{E}_\Phi [\prod_{i \in \{b_0\}} \mathbf{E}_g [\exp(-sgR_i^{-\alpha})]]} \\
&= \frac{\exp \left(-2\pi\lambda \int_0^\infty (1 - \mathbf{E}_g [\exp(-sgv^{-\alpha})]) v dv \right)}{\int_0^\infty \mathbf{E}_g [\exp(-sgr'^{-\frac{\alpha}{2}})] 2\lambda\pi r e^{-\lambda\pi r^2} dr} \\
&= \frac{\exp \left(-2\pi\lambda \int_0^\infty \left(1 - \frac{\mu}{\mu+sv^{-\alpha}}\right) v dv \right)}{2\pi\lambda \int_0^\infty r e^{-\pi\lambda r^2} \frac{\mu}{\mu+sr'^{-\alpha}} dr} \tag{15}
\end{aligned}$$

なお、以上の式の第 2 行目は、 Φ と $\{b_0\}$ とが互いに独立とみなしている点で近似的な式変形となっていることに注意する。なお、同様の近似計算の手法が [6] においても試みられている。

$H(u)$ における、毎秒あたりの基地局 b_0 からの平均受信データ容量を $\tau(u, \lambda, \alpha)$ と表す。このとき、距離が l の区間を通信端末が s 秒かけて移動する間に受信する総受信データ量の平均 $D_1(l, \lambda, \alpha)$ は

$$D_1(l, \lambda, \alpha) \simeq \frac{1}{l} \int_0^l \tau(u, \lambda, \alpha) du \tag{16}$$

となる。ここで、 $D_1(l, \lambda, \alpha)$ は、 $\tau(u, \lambda, \alpha)$ の区間 0 から l までの推移を s 回に分割した区分求積として求められている。

以下、 $\tau(u, \lambda, \alpha)$ を求める。

$$\begin{aligned}
\tau(u, \lambda, \alpha) &= \mathbf{E}[\log(1 + SINR)] \\
&= \frac{1}{2\pi} \int_0^{2\pi} \tau(\lambda, \alpha, \theta, u) d\theta \tag{17}
\end{aligned}$$

ここで、

$$\begin{aligned}
\tau(\lambda, \alpha, \theta, u) &= \int_{r>0} f(r) \mathbf{E}[\log(1 + \frac{hr'^{-\alpha}}{\sigma^2 + I_{r'}})] dr \\
&= 2\pi\lambda \int_{r>0} r e^{-\pi\lambda r^2} \int_{t>0} \mathbf{P}[\log(1 + \frac{hr'^{-\alpha}}{\sigma^2 + I_{r'}}) > t] dt dr \\
&= 2\pi\lambda \int_{r>0} r e^{-\pi\lambda r^2} \int_{t>0} e^{-\sigma^2 \mu r'^{\alpha}(e^t-1)} \mathcal{L}_{I_{r'}}(\mu r'^{\alpha}(e^t-1)) dt dr \tag{18}
\end{aligned}$$

であり、これは式 (10) に合致する。なお、以上の式における $\mathcal{L}_{I_{r'}}(\mu r'^{\alpha}(e^t-1))$ は式 (15) で求めた Laplace 変換を代入することで求めることができ、その結果は式 (11) に合致する。□

proposition 3 ([2], Theorem 3) 基地局の配置が *intensity* を λ とする定常ポワソン点過程 Φ に従うとき、距離 l の直線上でハンドオフを行いながら通信端末が移動を行ったときの平均受信データ容量 $D_2(s, \lambda, \alpha)$ は以下のように求まる。

$$D_2(s, \lambda, \alpha) = s \cdot \tau(\lambda, \alpha) \tag{19}$$

ここで、

$$\tau(\lambda, \alpha) = \int_{r>0} r e^{-\pi\lambda r^2} \int_{t>0} e^{-\sigma^2 \mu r^{\alpha}(e^t-1)} \mathcal{L}_{I_r}(\mu r^{\alpha}(e^t-1)) dt 2\pi\lambda r dr \tag{20}$$

$$\mathcal{L}_{I_r}(\mu r^{\alpha}(e^t-1)) = \exp \left(-\pi\lambda r^2 (e^t-1)^{2/\alpha} \int_{(e^t-1)^{-2/\alpha}}^\infty \frac{1}{1+x^{\alpha/2}} \right) \tag{21}$$

である。

証明 基地局のなすセルの境界を超えるたびに通信端末がハンドオフを行う場合、その平均受信データ容量は、最短距離に存在する基地局から通信を受信するときの毎秒あたりの平均受信データ容量 $\tau(\lambda, \alpha)$ と、距離 l の区間の移動に要する秒数 s との積に等しい。ここで、 $\tau(\lambda, \alpha)$ が式 (20) で与えられることは、[2], Theorem 3 から従う。□

3.3. 通信の評価関数

これまでの議論により、一次元点過程 Ψ により分布する k 個の点で直線軌道 $TR(L)$ 上を分割した各区間の距離を l としたとき、その区間を通信端末が s 秒かけて移動するときのハンドオフレート $\mathcal{N}(\lambda, l)$ と、平均受信データ容量の変化量 $\Delta D(s, l, \lambda, \alpha)$ をそれぞれ求めた。これらを用いて、通信の評価関数 $QoS(s, l, \lambda, \alpha)$ を

$$QoS(s, l, \lambda, \alpha) = \mathcal{U}\Delta D(s, l, \lambda, \alpha) - \mathcal{C}\mathcal{N}(\lambda, l) \quad (22)$$

と定義する。ここで、 $\mathcal{U}$ はデータ 1 ビットに対応するユーティリティを表し、 $\mathcal{C}$ はハンドオフ 1 回あたりのコスト換算量を表す。なお、 $\mathcal{C}$ の単位はユーティリティ $\mathcal{U}$ と同じである。式 (22) における $\mathcal{U}\Delta D(s, l, \lambda, \alpha)$ と $\mathcal{C}\mathcal{N}(\lambda, l)$ はそれぞれ、距離 l の区間上の移動において、セルの境界を越えてもハンドオフを行わない場合のユーティリティ損失量と、セルの境界を超えるたびにハンドオフを行う場合に発生するコストのユーティリティ換算量とを表す。

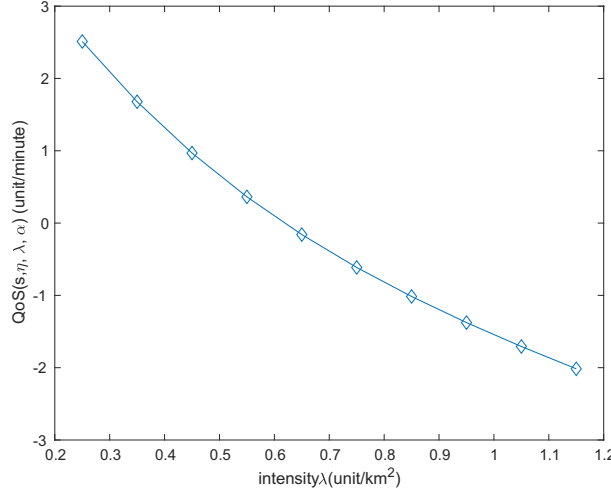


図 4: $QoS(s, l, \lambda, \alpha)$ の数値例

図 4 は、 λ に関する $QoS(s, l, \lambda, \alpha)$ の推移である。なお、他のパラメータは $s=100(s)$, $l=1(km)$, $\alpha=3$ と定めた。以上の結果より、 $QoS(s, l, \lambda, \alpha)$ には λ に関する減少傾向が認められる。これは、式 (23) の第二項におけるハンドオフレート $\mathcal{N}(\lambda, l)$ が、第一項における $\Delta D(s, l, \lambda, \alpha)$ に対し基地局の配置の $\text{intensity } \lambda$ に関してより高いオーダーで増加していることを示している。すなわち、 λ が十分小さいときはハンドオフを行わない場合のユーティリティ損失量がハンドオフを行う場合に対し大きくなり、逆に λ が十分大きい場合は、ハンドオフを行う場合のコストがハンドオフを行わない場合に対し大きくなる事が分かる。

以上の $QoS(s, l, \lambda, \alpha)$ を用いて直線軌道 $TR(L)$ 全体での評価関数値の平均を求める。直線軌道 $TR(L)$ を通信端末が移動する際の速度変化は、一次元点過程 Ψ により $TR(L)$ に配置された k 個の点の分布により変化する。 k 個の点により分割された $TR(L)$ 上の区間の長さが任意の分布 η に従うとき、 $QoS(s, l, \lambda, \alpha)$ の l に関する平均 $QoS_k(s, \eta, \lambda, \alpha)$ は

$$QoS_k(s, \eta, \lambda, \alpha) = \mathbf{E}_l[QoS(s, l, \lambda, \alpha)] \quad (23)$$

と求まる。

図 5 は、分布 η が平均 ν の指数分布に従う場合 ($\eta = \eta_{exp}$) に対する、一様分布に従う場合 ($\eta = \eta_{uni}$) の $QoS_k(s, \eta, \lambda, \alpha)$ の差分値 $\Delta QoS_k(s, \eta, \lambda, \alpha) = QoS_k(s, \eta_{exp}, \lambda, \alpha) - QoS_k(s, \eta_{uni}, \lambda, \alpha)$ を、 k の値に関して推移させた結果である。なお、実際の車両の移動を想定し、移動軌道の長さ L を $L = 10(km)$ 、移動軌道を k 個の点により分割した各区間の移動にかかる秒数 s を $s = 100(秒)$ 、基地局の配置密度を $\lambda = 0.25(個/km^2)$ 電波強度の減衰係数 α を $\alpha = 3$ と設定した。以上の結果より、 $\Delta QoS_k(s, \eta, \lambda, \alpha)$ の値は k の増加に伴い増加することが分かる。すなわち、直線上を移動する物体の通信において、ハンドオフの有無により発生する通信のユーティリティの変化量は、軌道上を一定速度で移動する場合に対し一定ではない速度で移動する場合の方が大きくなり、平均移動速度の増加に伴いその傾向は大きくなることを示している。

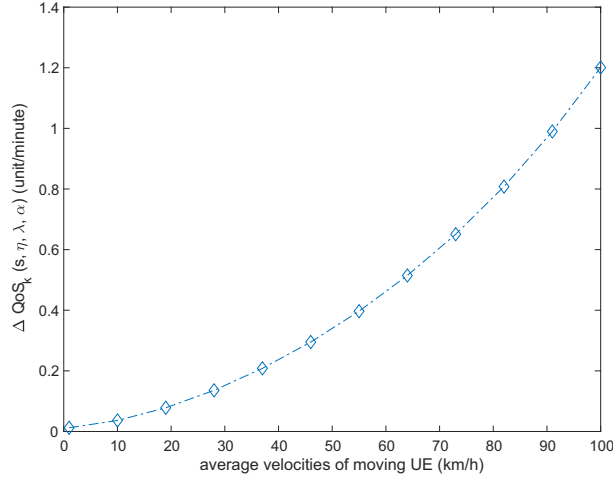


図 5: $\Delta QoS_k(s, \eta, \lambda, \alpha)$ の数値例

4. おわりに

本論文では、物体の移動速度の変化に対応し、受信データ容量の変化とハンドオフレートを導出することで、軌道上を不規則な速度で移動する物体に対して受信性能判定を行う手法を提案した。さらに手法の評価では、同一の平均速度を保ちながら、軌道上を終始等速で移動する場合とそうでない場合との間で、性能判定結果に起こる影響を観察した。なお、本論文で導入した無線通信の性能評価指標を用いれば、無線通信を行う物体の移動パターン（平均移動速度、移動速度の変化が従う分布）に対応した最適な基地局の配置強度が求められると考えられる。よって今後の展望としては、実際の車両の移動速度の変化を分布関数により近似した上での、最適な基地局の配置強度の導出等がある。また、無線通信の規格が複数混在するヘテロジニアスネットワークへと本論文のモデルを拡張することも、より現実に即した無線ネットワークのモデリングのためには求められる。

参考文献

- [1] H. Dhillon, and R. Ganti, Modeling and Analysis of K-tier Downlink Heterogeneous Cellular Networks, IEEE Journal. selected Areas in Communications, **30** No.3(2012), 550–560.
- [2] J. Andrews, F. Baccelli, and R. Ganti, A Tractable Approach to Coverage and Rate in Cellular Networks, IEEE Transactions and Communications, **59** No.11 (2011), 3122–3134.
- [3] W. Bao, B. Liang, Stochastic Geometric Analysis of User Mobility in Heterogeneous Wireless Networks, IEEE Journal. selected Areas in Communications, **33** No.10(2015), 2212–2225.
- [4] M. P. Quine and D. F. Watson, Radical Generarion of n-Dimensional Poisson Process, Journal of Applied Probability, **21**(1984), 548–557.
- [5] A. L. Hinde, R. E. Miles, Monte Carlo Estimates of the Distributions of the Voronoi Tessallation with Respect to a Poisson Process, **10**(1980), 205–223.
- [6] H.Q. Nguyen, F. Baccelli, and D. Kofman, A Stochastic Geometric Analysis of Dense IEEE 802.11 networks, in Proc, 26th IEEE International Conf. Comut. Commun. (INFOCOM’07), 2007, 1199–1207.

M/G/1+PH 待ち行列における系内客数分布の数値計算法

井上 文彰
大阪大学工学研究科
電気電子情報工学専攻

概要：本稿では、客の待ち時間に制約があり、制約の長さが相型分布に従う M/G/1 待ち行列における定常系内客数分布を考察し、その確率関数を計算するアルゴリズムを構築する。さらに、このアルゴリズムの誤差上界を導出し、いくつかの数値例を示す。

1. はじめに

本稿では、客の待ち時間に制約のある定常な M/G/1 待ち行列における系内客数分布を考察する。客の到着は率 λ ($0 < \lambda < \infty$) のポアソン過程に従い、到着客は先着順でサービスを受けるとする。サービス時間は一般の分布関数 $H(x)$ ($x \geq 0$) に従って独立かつ同一に分布すると仮定する。それぞれの客の待ち時間には制約があり、経過待ち時間がその制約に達すると、客はサービスを受けずに系を途中退去する。待ち時間制約の長さは、相型分布 (α, T) に従って独立かつ同一に分布すると仮定する。ただし、 α は確率ベクトル、 T は劣無限小生成作用素を表す。すなわち、待ち時間制約長の分布関数 $G(x)$ は次式で与えられる。

$$G(x) = 1 - \alpha \exp[Tx]e, \quad x \geq 0 \quad (1)$$

ただし、 e は要素が全て 1 の列ベクトルを表す。このモデルは通常、M/G/1+PH 待ち行列と表記され、 $G(x)$ が一般の分布関数と仮定される M/G/1+G 待ち行列の特別な場合に相当する。

本稿では、サービス時間分布は有限の平均 $E[H] < \infty$ を有すると仮定する。このときトラヒック強度 $\rho = \lambda E[H]$ は有限の値を取り、また $\lim_{x \rightarrow \infty} G(x) = 1$ であるため、システムは安定であることが保証される [1]。無用な煩雑さを避けるため、本稿では $T - Te\alpha$ が既約であること、ならびに、 $H(0) = 0$ を仮定する。

本稿の目的は、M/G/1+PH 待ち行列における系内客数分布の数値計算法を構築することである。このモデルは M/PH/1+PH 待ち行列を特別な場合として含むことに注意せよ。M/PH/1+PH 待ち行列の系内客数過程は連続時間のマルコフ連鎖を成し、レベル依存準出生死滅過程として定式化される。待ち客数が n であるとき、この準出生死滅過程の背後状態数は待ち時間制約を表す相型分布が有する状態数の n 乗に比例するため、その定常分布を直接計算することは極めて難しい。本稿では、M/G/1+PH 待ち行列における仮待ち時間分布の解析結果 [6] をもとに系内客数分布を求めるというアプローチを取ることで、M/G/1+PH 待ち行列の系内客数分布に対する数値計算アルゴリズムを構築する。さらに、この計算アルゴリズムの誤差上界を導出し、いくつかの数値例を示す。

以下の本稿は次のように構成される。第 2 節では、M/G/1+PH 待ち行列に対する既存の解析結果をいくつか紹介する。第 3 節では、M/G/1+PH 待ち行列の数値計算アルゴリズムを構築し、その誤差上界を与える。第 4 節において数値例を示し、最後に第 5 節で本稿を締めくくる。

2. M/G/1+PH 待ち行列に関する既存の結果

$\bar{H}(x)$ ならびに $\bar{G}(x)$ ($x \geq 0$) を、それぞれサービス時間ならびに待ち時間制約長の補分布関数とする。

$$\begin{aligned} \bar{H}(x) &= 1 - H(x) \\ \bar{G}(x) &= 1 - G(x) \\ &= \alpha \exp[Tx]e \end{aligned} \quad (2)$$

また、 $h_e(x)$ をサービス時間の平衡分布の密度関数とする。

$$h_e(x) = \frac{\bar{H}(x)}{E[H]}, \quad x \geq 0$$

L ならびに V を定常状態における系内客数ならびに仮待ち時間とする。ただし、仮待ち時間とは、現時点において系内に存在する客が全て離脱するまでに要する総時間を表す。すなわち本稿のモデルにおける仮待ち時間は、系内に滞在中である客のうち、最終的にサービスを受けることになる客の残余サービス時間の総和に等しい。一方で、 L は最終的にサービスを受ける客と途中退去する客の両方を含んだ数を表す。 π_ℓ ($\ell = 0, 1, \dots$) を L の確率関数とし、 $v(x)$ ($x > 0$) を V の密度関数とする。

$$\pi_\ell = \Pr(L = \ell), \quad \ell = 0, 1, \dots, \quad v(x) = \frac{d\Pr(V \leq x)}{dx}, \quad x > 0$$

定義より $\Pr(V = 0) = \pi_0$ であり、次式が成立する。

$$\pi_0 + \int_{0+}^{\infty} v(x) dx = 1 \quad (3)$$

P_{loss} をランダムに選ばれた客が途中退去する確率とする。サーバのみに注目したリトルの公式 $1 - \pi_0 = \lambda(1 - P_{\text{loss}})E[H]$ より、 P_{loss} は π_0 を用いて次式で与えられることがわかる [2]。

$$P_{\text{loss}} = \frac{\rho - (1 - \pi_0)}{\rho} \quad (4)$$

補題 1 ([1, 7]). $u(n; x)$ ($x > 0, n = 1, 2, \dots$) を次の漸化式で定められる関数列とする。

$$u(1; x) = \rho h_e(x), \quad u(n; x) = \rho \int_{0+}^x u(n-1; y) \bar{G}(y) h_e(x-y) dy, \quad n = 2, 3, \dots$$

このとき、 π_0 ならびに $v(x)$ は次式で与えられる。

$$\pi_0 = \left[1 + \sum_{n=1}^{\infty} \int_{0+}^{\infty} u(n; y) dy \right]^{-1} \quad (5)$$

$$v(x) = \pi_0 \sum_{n=1}^{\infty} u(n; x), \quad x > 0$$

注意 1. [3] では、次のように $u(n; x)$ に対する確率的解釈が得られている。以下の (i)–(iv) で記述される、割り込み再開型後着順サービス (LCFS-PR; Last-Come First-Served Preemptive-Resume) M/G/1 待ち行列を考える。

- (i) 客は率 λ のポアソン過程に従って到着する
- (ii) 客のサービス時間は、分布関数 $H(x)$ ($x \geq 0$) に従って独立かつ同一に分布する
- (iii) 客は、到着直前における系内仕事量が x ($x \geq 0$) であるとき、確率 $\bar{G}(x)$ で呼損する
- (iv) 到着時に呼損しなかった客は、その後途中退去することなく必ずサービスを受け、そのサービス順序は LCFS-PR に従う

この LCFS-PR M/G/1 待ち行列の系内仕事量過程が、M/G/1+G 待ち行列における仮待ち時間過程と等価であることは容易に確かめられる。この等価な待ち行列モデルにおける定常系内客数を L_{LCFS} とし、このモデルの定常系内仕事量、すなわち、元の M/G/1+G 待ち行列における定常仮待ち時間を V とすると、次式が成立する。

$$\frac{d}{dx} [\Pr(L_{\text{LCFS}} = n, V \leq x)] = \pi_0 u(n; x), \quad x > 0, n = 1, 2, \dots$$

また、上式より、 L_{LCFS} の確率関数は次式で与えられる。

$$\Pr(L_{\text{LCFS}} = 0) = \left[1 + \sum_{n=1}^{\infty} \int_{0+}^{\infty} u(n; x) dx \right]^{-1}$$

$$\Pr(L_{\text{LCFS}} = n) = \pi_0 \int_{0+}^{\infty} u(n; x) dx, \quad n = 1, 2, \dots$$

よって、これらの関係式を用いると、補題 1 の結果が確率的解釈を有することがわかる。

以上の結果は、安定な M/G/1+G 待ち行列において一般に成立する。本稿で考察する M/G/1+PH 待ち行列では、待ち時間制約長が相型分布に従うという仮定、すなわち (2) によって、補題 1 の離散版に相当する関係式が得られる [6].

θ を $\mathbf{T}$ の対角要素の最大の絶対値とする。 θ を用いた一様化により、(2) は次式で書き換えられる。

$$\bar{G}(x) = \sum_{k=0}^{\infty} \frac{\exp[-\theta x](\theta x)^k}{k!} \cdot \bar{g}^{[k]}, \quad x \geq 0 \quad (6)$$

ただし、 $\mathbf{I}$ を単位行列として、 $\bar{g}^{[k]}$ は次式で与えられる。

$$\bar{g}^{[k]} = \alpha[\mathbf{I} + \theta^{-1}\mathbf{T}]^k \mathbf{e}, \quad k = 0, 1, \dots$$

(6) の両辺に θ を掛け、 $[0, \infty)$ に関して積分することで次式が得られる。

$$\sum_{k=0}^{\infty} \bar{g}^{[k]} = \theta E[G] \quad (7)$$

ただし、 $E[G]$ は平均待ち時間制約長を表す。

$v^{[k]}(j\theta)$ ($j = 1, 2, \dots$) を、定常仮待ち時間 V の間に率 $j\theta$ のポアソン到着を発生させたとき、総到着数が k である確率とする。

$$v^{[k]}(j\theta) = \int_{0+}^{\infty} \frac{\exp[-j\theta x](j\theta x)^k}{k!} \cdot v(x) dx, \quad k = 0, 1, \dots$$

同様に、 $u^{[k]}(n; j\theta)$ ($n = 1, 2, \dots, j = 1, 2, \dots$) ならびに $h_e^{[k]}(j\theta)$ ($j = 1, 2, \dots$) を次式で定義する。

$$\begin{aligned} u^{[k]}(n; j\theta) &= \int_{0+}^{\infty} \frac{\exp[-j\theta x](j\theta x)^k}{k!} \cdot u(n; x) dx, \quad k = 0, 1, \dots \\ h_e^{[k]}(j\theta) &= \int_{0+}^{\infty} \frac{\exp[-j\theta x](j\theta x)^k}{k!} \cdot h_e(x) dx, \quad k = 0, 1, \dots \end{aligned} \quad (8)$$

記法を簡単にするため、 $\mathbf{v}^*(j\theta)$ ($j = 1, 2, \dots$)、 $\mathbf{u}^*(n; j\theta)$ ($n = 1, 2, \dots, j = 1, 2, \dots$)、ならびに $\mathbf{h}_e^*(j\theta)$ ($j = 1, 2, \dots$) をそれぞれ、第 k 要素 ($k = 0, 1, \dots$) が $v^{[k]}(j\theta)$ 、 $u^{[k]}(n; j\theta)$ 、ならびに $h_e^{[k]}(j\theta)$ で与えられる $1 \times \infty$ ベクトルと定義する。また、同様に、 $\bar{\mathbf{g}}^*$ を第 k 要素 ($k = 0, 1, \dots$) が $\bar{g}^{[k]}$ で与えられる $\infty \times 1$ ベクトルと定義する。

$\mathbf{H}_{e,n}$ ($n = 1, 2, \dots$) ならびに $\bar{\mathbf{B}}_n$ ($n = 1, 2, \dots$) を、次式で定義される $\infty \times \infty$ 確率行列ならびに $\infty \times \infty$ 劣確率行列と定義する。

$$\mathbf{H}_{e,n} = \begin{pmatrix} h_e^{[0]}(n\theta) & h_e^{[1]}(n\theta) & h_e^{[2]}(n\theta) & \cdots \\ 0 & h_e^{[0]}(n\theta) & h_e^{[1]}(n\theta) & \cdots \\ 0 & 0 & h_e^{[0]}(n\theta) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (9)$$

$$\bar{\mathbf{B}}_n = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ b_n(1, 0)\bar{g}^{[1]} & b_n(1, 1) & 0 & \cdots \\ b_n(2, 0)\bar{g}^{[2]} & b_n(2, 1)\bar{g}^{[1]} & b_n(2, 2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (10)$$

ただし $b_n(m, k)$ ($n = 1, 2, \dots, m = 1, 2, \dots$) はパラメタ $(m, n/(n+1))$ の二項分布の確率関数を表す。

$$b_n(m, k) = \binom{m}{k} \left(\frac{n}{n+1} \right)^k \left(\frac{1}{n+1} \right)^{m-k}, \quad k = 0, 1, \dots, m$$

補題 2 ([6]). M/G/1+PH 待ち行列において, 次式が成立する.

$$\mathbf{v}^*(j\theta) = \pi_0 \sum_{n=1}^{\infty} \mathbf{u}^*(n; j\theta) \quad (11)$$

$$\mathbf{u}^*(1; j\theta) = \rho \mathbf{h}_e^*(j\theta), \quad \mathbf{u}^*(n; j\theta) = \mathbf{u}^*(n-1, (j+1)\theta) \cdot \rho \bar{\mathbf{B}}_j \mathbf{H}_{e,j}, \quad n = 2, 3, \dots \quad (12)$$

したがって, $\mathbf{u}^*(n; j\theta)$ は次式で与えられる.

$$\mathbf{u}^*(n; j\theta) = \rho \mathbf{h}_e((j+n-1)\theta) \cdot \rho \bar{\mathbf{B}}_{j+n-2} \mathbf{H}_{e,j+n-2} \cdot \rho \bar{\mathbf{B}}_{j+n-3} \mathbf{H}_{e,j+n-3} \cdots \rho \bar{\mathbf{B}}_j \mathbf{H}_{e,j}$$

系 1 ([5]). $\infty \times \infty$ 行列 $\mathbf{A}_n(j\theta)$ ($n = 1, 2, \dots, j = 1, 2, \dots$) を次式で定義する.

$$\mathbf{A}_n(j\theta) = \rho \bar{\mathbf{B}}_{j+n-1} \mathbf{H}_{e,j+n-1} \cdot \rho \bar{\mathbf{B}}_{j+n-2} \mathbf{H}_{e,j+n-2} \cdots \rho \bar{\mathbf{B}}_j \mathbf{H}_{e,j} \quad (13)$$

このとき, $\mathbf{A}_n(j\theta)$ の第 0 行は $\mathbf{u}(n; j\theta)$ に等しい.

注意 2. [6] では, 以上の結果をもとに, π_0 ならびに P_{loss} の数値計算法が考察されている.

次に, 系内容数分布 π_ℓ ($\ell = 1, 2, \dots$) を考える. $G_e(x)$ ($x \geq 0$) ならびに $\bar{G}_e(x)$ ($x \geq 0$) を, 待ち時間制約長の平衡分布の分布関数ならびに補分布関数とする.

$$G_e(x) = \int_0^x \frac{\bar{G}(y)}{\mathbf{E}[G]} dy, \quad \bar{G}_e(x) = \int_x^\infty \frac{\bar{G}(y)}{\mathbf{E}[G]} dy$$

(6) と同様に, これらの関数は次式で書き換えられる.

$$G_e(x) = \sum_{k=0}^{\infty} \frac{\exp[-\theta x](\theta x)^k}{k!} \cdot g_e^{[k]}, \quad \bar{G}_e(x) = \sum_{k=0}^{\infty} \frac{\exp[-\theta x](\theta x)^k}{k!} \cdot \bar{g}_e^{[k]}$$

ただし, $g_e^{[k]}$ ならびに $\bar{g}_e^{[k]}$ ($k = 0, 1, \dots$) は次式で与えられる.

$$g_e^{[0]} = 0, \quad g_e^{[k]} = \sum_{i=0}^{k-1} \frac{\bar{g}_e^{[i]}}{\theta \mathbf{E}[G]}, \quad k = 1, 2, \dots, \quad \bar{g}_e^{[k]} = \sum_{i=k}^{\infty} \frac{\bar{g}_e^{[i]}}{\theta \mathbf{E}[G]}, \quad k = 0, 1, \dots$$

g_e^* および $\bar{g}_e^*$ をそれぞれ, $g_e^{[k]}$ および $\bar{g}_e^{[k]}$ を第 k 要素とする $\infty \times 1$ ベクトルと定義する. また, $\infty \times \infty$ 劣確率行列 $\mathbf{B}_{e,n}$ ならびに $\bar{\mathbf{B}}_{e,n}$ を次式で定義する.

$$\mathbf{B}_{e,n} = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ b_n(1,0)g_e^{[1]} & 0 & 0 & \cdots \\ b_n(2,0)g_e^{[2]} & b_n(2,1)g_e^{[1]} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (14)$$

$$\bar{\mathbf{B}}_{e,n} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ b_n(1,0)\bar{g}_e^{[1]} & b_n(1,1) & 0 & \cdots \\ b_n(2,0)\bar{g}_e^{[2]} & b_n(2,1)\bar{g}_e^{[1]} & b_n(2,2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (15)$$

補題 3 ([5]). π_ℓ ($\ell = 0, 1, \dots$) は次式で与えられる.

$$\pi_0 = \left[1 + \sum_{\ell=1}^{\infty} r_\ell \right]^{-1}, \quad \pi_\ell = \pi_0 r_\ell, \quad \ell = 1, 2, \dots \quad (16)$$

$$r_\ell = \sum_{m=\ell-1}^{\infty} \frac{\exp[-\lambda \mathbf{E}[G]](\lambda \mathbf{E}[G])^m}{m!} \cdot \hat{r}_{m,\ell-1}, \quad \ell = 1, 2, \dots \quad (17)$$

ただし, $\hat{r}_{m,\ell}$ ($\ell = 0, 1, \dots, m, m = 0, 1, \dots$) は次式で定義される.

$$\hat{r}_{m,\ell} = \sum_{n=1}^{\infty} \int_{0+}^{\infty} K_{m,\ell}(x) u(n; x) dx$$

$$K_{0,0}(x) = 1, \quad K_{m,\ell}(x) = \binom{m}{\ell} \{G_e(x)\}^{\ell} \{\bar{G}_e(x)\}^{m-\ell}, \quad \ell = 0, 1, \dots, m, m = 1, 2, \dots$$

補題 4 ([5]). $\kappa_{m,\ell}^*$ ($\ell = 0, 1, \dots, m, m = 1, 2, \dots$) を, $\kappa_{1,0}^* = \bar{g}_e^*$, $\kappa_{1,1}^* = g_e^*$ ならびに次の漸化式で定められる $\infty \times 1$ ベクトルとする.

$$\kappa_{m,0}^* = \bar{B}_{e,m} \kappa_{m-1,0}^* \quad (18)$$

$$\kappa_{m,\ell}^* = \bar{B}_{e,m} \kappa_{m-1,\ell}^* + B_{e,m} \kappa_{m-1,\ell-1}^*, \quad \ell = 1, 2, \dots, m-1 \quad (19)$$

$$\kappa_{m,m}^* = B_{e,m} \kappa_{m-1,m-1}^* \quad (20)$$

このとき, $\hat{r}_{m,\ell}$ ($m = 0, 1, \dots$) は次式で与えられる.

$$\hat{r}_{0,0} = \sum_{n=1}^{\infty} \mathbf{u}^*(n; \theta) e \quad (21)$$

$$\hat{r}_{m,\ell} = \sum_{n=1}^{\infty} \mathbf{u}^*(n; m\theta) \kappa_{m,\ell}^*, \quad \ell = 0, 1, \dots, m, m = 1, 2, \dots \quad (22)$$

注意 3. 定義より, $m = 1, 2, \dots$ に対し次式が成立する.

$$\sum_{\ell=0}^m \kappa_{m,\ell}^* = e \quad (23)$$

したがって, 任意の k ($k = 0, 1, \dots$) について, $\infty \times 1$ ベクトル $\kappa_{m,0}^*, \kappa_{m,1}^*, \dots, \kappa_{m,m}^*$ のうち少なくとも一つは, 無視できないほど大きい第 k 要素を有する.

3. 系内客数分布の数値計算法

$m = 1, 2, \dots$ に対し, $\hat{r}_{m,\ell}(n)$ ($\ell = 0, 1, \dots, m, n = 1, 2, \dots$) を次式で定義する.

$$\hat{r}_{m,\ell}(n) = A_n(m\theta) \kappa_{m,\ell}^*$$

$A_n(m\theta)$ の定義 (13) より, $\hat{r}_{m,\ell}(n)$ は次の漸化式で与えられる.

$$\hat{r}_{m,\ell}(1) = \rho \bar{B}_m H_{e,m} \kappa_{m,\ell}^*, \quad \hat{r}_{m,\ell}(n) = \rho \bar{B}_{m+n-1} H_{e,m+n-1} \cdot \hat{r}_{m,\ell}(n-1), \quad n = 2, 3, \dots \quad (24)$$

系 1 より $\hat{r}_{m,\ell}(n)$ の第 0 要素は

$$[\hat{r}_{m,\ell}(n)]_0 = \mathbf{u}^*(n; m\theta) \kappa_{m,\ell}^*$$

を満たす. したがって, (22) より, $m = 1, 2, \dots$ に対し $\hat{r}_{m,\ell}$ ($\ell = 0, 1, \dots, m$) は次式で与えられる.

$$\hat{r}_{m,\ell} = \sum_{n=1}^{\infty} [\hat{r}_{m,\ell}(n)]_0 \quad (25)$$

また, (21) ならびに (23) より, $\hat{r}_{0,0}$ は次式で与えられる.

$$\begin{aligned} \hat{r}_{0,0} &= \sum_{n=1}^{\infty} \mathbf{u}^*(n; \theta) (\kappa_{1,0}^* + \kappa_{1,1}^*) \\ &= \hat{r}_{1,0} + \hat{r}_{1,1} \end{aligned} \quad (26)$$

以上より, M/G/1+PH 待ち行列における系内客数分布 π_{ℓ} ($\ell = 0, 1, \dots$) は, 原理的には次の手順で定められることがわかる.

- (i) $\kappa_{m,\ell}^*$ ($\ell = 0, 1, \dots, m, m = 1, 2, \dots$) を (18)–(20) より求める.
- (ii) $\hat{\mathbf{r}}_{m,\ell}(n)$ ($n = 1, 2, \dots, \ell = 0, 1, \dots, m, m = 1, 2, \dots$) を (24) より求める.
- (iii) $\hat{r}_{m,\ell}$ ($\ell = 0, 1, \dots, m, m = 0, 1, \dots$) を (25) および (26) より求める.
- (iv) r_ℓ ($\ell = 1, 2, \dots$) を (17) より求め、それを用いて π_ℓ ($\ell = 0, 1, \dots$) を (16) より求める.

ただし、この手順は無限次元のベクトルや行列を複数含んでおり、そのままでは計算機上で実装することはできない。以下では、 π_ℓ ($\ell = 1, 2, \dots$) を実際に計算するためのアルゴリズムを考察する。

3.1. 計算アルゴリズム

まず、与えられた $\epsilon_g > 0$ に対し、 $\bar{\mathbf{g}}^{*,\text{comp}}$ の切断点 k_g^* を次式で定める。

$$k_g^* = \min \left\{ i \in \{1, 2, \dots\}; \sum_{k=0}^{i-1} \alpha[\mathbf{I} + \theta^{-1}\mathbf{T}]^k \mathbf{e} \geq \theta \mathbf{E}[G] - \epsilon_g \right\} \quad (27)$$

ここで、(7) に注意せよ。このとき、 $\bar{\mathbf{g}}$ は次式で表される $\infty \times 1$ ベクトル $\bar{\mathbf{g}}^{*,\text{comp}}$ で近似される。

$$\bar{\mathbf{g}}^{*,\text{comp}} = \begin{pmatrix} \bar{\mathbf{g}}^{*,\text{comp}} \\ \mathbf{0} \end{pmatrix}$$

ただし、 $\bar{\mathbf{g}}^{*,\text{comp}}$ は第 k 要素が次式で与えられる $k_g^* \times 1$ ベクトルを表す。

$$\bar{g}^{[k],\text{comp}} = \bar{g}^{[k]} = \alpha[\mathbf{I} + \theta^{-1}\mathbf{T}]^k \mathbf{e}, \quad k = 0, 1, \dots, k_g^* - 1 \quad (28)$$

以降では同様に、実際に計算を行なう有限次元のベクトルや行列は上添字 comp を付けて表し、その有限次元の量によって暗黙に表現される無限次元の量を $\overline{\text{comp}}$ を付けて表す。

$\bar{\mathbf{g}}$ に対する上記の切断は (10) に波及し、 $\bar{\mathbf{B}}_n$ に対する近似行列は次式で表される [6]。

$$\bar{\mathbf{B}}_n^{\overline{\text{comp}}} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ b_n(1,0)\bar{g}^{[1]} & b_n(1,1) & 0 & \cdots \\ b_n(2,0)\bar{g}^{[2]} & b_n(2,1)\bar{g}^{[1]} & b_n(2,2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ b_n(k_g^* - 1, 0)\bar{g}^{[k_g^* - 1]} & b_n(k_g^* - 1, 1)\bar{g}^{[k_g^* - 2]} & b_n(k_g^* - 1, 2)\bar{g}^{[k_g^* - 3]} & \cdots \\ 0 & b_n(k_g^*, 1)\bar{g}^{[k_g^* - 1]} & b_n(k_g^*, 2)\bar{g}^{[k_g^* - 2]} & \cdots \\ 0 & 0 & b_n(k_g^* + 1, 2)\bar{g}^{[k_g^* - 1]} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

さらに、 k^* を $k^* > k_g^*$ なる切断点とし、 $\mathbf{B}_{e,n}$ ならびに $\bar{\mathbf{B}}_{e,n}$ を $k^* \times k^*$ 北西角行列により切断近似する。まず、 $\mathbf{g}_e^*$ ならびに $\bar{\mathbf{g}}_e^*$ に対する近似ベクトルを次式で表す。

$$\mathbf{g}_e^{*,\text{comp}} = \begin{pmatrix} \mathbf{g}_e^{*,\text{comp}} \\ \mathbf{e} \end{pmatrix}, \quad \bar{\mathbf{g}}_e^{*,\text{comp}} = \begin{pmatrix} \bar{\mathbf{g}}_e^{*,\text{comp}} \\ \mathbf{0} \end{pmatrix}$$

ただし、 $\mathbf{g}_e^{*,\text{comp}}$ および $\bar{\mathbf{g}}_e^{*,\text{comp}}$ はいずれも $k^* \times 1$ ベクトルを表し、それらの第 k 要素 ($k = 1, 2, \dots, k^* - 1$) は次式で与えられる。

$$g_e^{[k],\text{comp}} = \begin{cases} 0, & k = 0 \\ \frac{1}{\theta \mathbf{E}[G]} \sum_{i=0}^{k-1} \bar{g}^{[i],\text{comp}}, & k = 1, 2, \dots, k_g^* \\ 1, & k = k_g^* + 1, k_g^* + 2, \dots, k^* - 1 \end{cases} \quad (29)$$

$$\bar{g}_e^{[k],\text{comp}} = \begin{cases} \frac{1}{\theta \mathbf{E}[G]} \sum_{i=k}^{k_g^* - 1} \bar{g}^{[i],\text{comp}}, & k = 1, 2, \dots, k_g^* - 1 \\ 0, & k = k_g^*, k_g^* + 1, \dots, k^* - 1 \end{cases} \quad (30)$$

そして, (14) ならびに (15) において, $g_e^{[k]}$ を $g_e^{[k],\text{comp}}$ に, $\bar{g}_e^{[k]}$ を $\bar{g}_e^{[k],\text{comp}}$ にそれぞれ置き換えた行列と同じ (i, j) 要素 ($i, j \in \{0, 1, \dots, k^* - 1\}$) を有する $k^* \times k^*$ 行列 $\mathbf{B}_{e,n}^{\text{comp}}$ ならびに $\bar{\mathbf{B}}_{e,n}^{\text{comp}}$ を用いて, 行列 $\mathbf{B}_{e,n}$ ならびに $\bar{\mathbf{B}}_{e,n}$ を近似する.

$$\mathbf{B}_{e,n}^{\text{comp}} = \begin{pmatrix} \mathbf{B}_{e,n}^{\text{comp}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \bar{\mathbf{B}}_{e,n}^{\text{comp}} = \begin{pmatrix} \bar{\mathbf{B}}_{e,n}^{\text{comp}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

このとき, (18)–(20) より, $\boldsymbol{\kappa}_{m,\ell}^*$ は次式で近似される.

$$\boldsymbol{\kappa}_{m,\ell}^{*,\text{comp}} = \begin{pmatrix} \boldsymbol{\kappa}_{m,\ell}^{*,\text{comp}} \\ \mathbf{0} \end{pmatrix}, \quad \ell = 0, 1, \dots, m, \quad m = 1, 2, \dots$$

ただし $\boldsymbol{\kappa}_{m,\ell}^{*,\text{comp}}$ は, $\boldsymbol{\kappa}_{1,0}^{*,\text{comp}} = \bar{\mathbf{g}}_e^{*,\text{comp}}$, $\boldsymbol{\kappa}_{1,1}^{*,\text{comp}} = \mathbf{g}_e^{*,\text{comp}}$ ならびに

$$\boldsymbol{\kappa}_{m,0}^{*,\text{comp}} = \bar{\mathbf{B}}_{e,m}^{\text{comp}} \boldsymbol{\kappa}_{m-1,0}^{*,\text{comp}} \quad (31)$$

$$\boldsymbol{\kappa}_{m,\ell}^{*,\text{comp}} = \bar{\mathbf{B}}_{e,m}^{\text{comp}} \boldsymbol{\kappa}_{m-1,\ell}^{*,\text{comp}} + \mathbf{B}_{e,m}^{\text{comp}} \boldsymbol{\kappa}_{m-1,\ell-1}^{*,\text{comp}}, \quad \ell = 1, 2, \dots, m-1 \quad (32)$$

$$\boldsymbol{\kappa}_{m,m}^{*,\text{comp}} = \mathbf{B}_{e,m}^{\text{comp}} \boldsymbol{\kappa}_{m-1,m-1}^{*,\text{comp}} \quad (33)$$

により定められる $k^* \times 1$ ベクトルを表す. よって, $\bar{\mathbf{B}}_n^{\text{comp}}$ ($n = 1, 2, \dots$) を $\bar{\mathbf{B}}_n^{\text{comp}}$ の $\{n(k_g^* - 1) + k^*\} \times \{(n-1)(k_g^* - 1) + k^*\}$ 北西角行列, $\mathbf{H}_{e,n}^{\text{comp}}$ ($n = 1, 2, \dots$) を $\mathbf{H}_{e,n}$ の $\{(n-1)(k_g^* - 1) + k^*\} \times \{(n-1)(k_g^* - 1) + k^*\}$ 北西角行列とすると, (24) より,

$$\hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(1) = \rho \bar{\mathbf{B}}_m^{\text{comp}} \mathbf{H}_{e,m}^{\text{comp}} \boldsymbol{\kappa}_{m,\ell}^{*,\text{comp}}, \quad (34)$$

$$\hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(n) = \rho \bar{\mathbf{B}}_{m+n-1}^{\text{comp}} \mathbf{H}_{e,m+n-1}^{\text{comp}} \hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(n-1), \quad n = 2, 3, \dots \quad (35)$$

と定めることで, $\hat{\mathbf{r}}_{m,\ell}(n)$ は次式で近似されることが確かめられる.

$$\hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(n) = \begin{pmatrix} \hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(n) \\ \mathbf{0} \end{pmatrix}, \quad \ell = 0, 1, \dots, m, \quad m = 1, 2, \dots$$

さらに, 与えられた $\epsilon_r > 0$ に対し,

$$n_m^* = \min \left\{ i \in \{1, 2, \dots\}; \hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(i) < \epsilon_r \mathbf{e}, \forall \ell \in \{0, 1, \dots, m\} \right\}$$

で定義される n_m^* を用いて, (25) で定義される $\hat{\mathbf{r}}_{m,\ell}$ を次式で切断近似する.

$$\hat{\mathbf{r}}_{m,\ell}^{\text{comp}} = \sum_{n=1}^{n_m^*} [\hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(n)]_0, \quad \ell = 0, 1, \dots, m, \quad m = 1, 2, \dots \quad (36)$$

また, (26) より $\hat{\mathbf{r}}_{0,0}^{\text{comp}}$ を次式で定める.

$$\hat{\mathbf{r}}_{0,0}^{\text{comp}} = \hat{\mathbf{r}}_{1,0}^{\text{comp}} + \hat{\mathbf{r}}_{1,1}^{\text{comp}} \quad (37)$$

最後に, 与えられた $\epsilon_M > 0$ に対し, (17) における無限和の切断点 m^* を

$$m^* = \min \left\{ i \in \{1, 2, \dots\}; \sum_{m=0}^{i-1} \frac{\exp[-\lambda E[G]] (\lambda E[G])^m}{m!} \geq 1 - \epsilon_M \right\} \quad (38)$$

で定め,

$$r_\ell^{\text{comp}} = \sum_{m=\ell-1}^{m^*-1} \frac{\exp[-\lambda E[G]] (\lambda E[G])^m}{m!} \cdot \hat{\mathbf{r}}_{m,\ell-1}^{\text{comp}}, \quad \ell = 1, 2, \dots, m^* \quad (39)$$

とする. すなわち, $\boldsymbol{\kappa}_{m,\ell}^{*,\text{comp}}$, $\hat{\mathbf{r}}_{m,\ell}^{\text{comp}}(n)$, および $\hat{\mathbf{r}}_{m,\ell}^{\text{comp}}$ は $m = 1, 2, \dots, m^* - 1$ についてのみ計算すれば良い. 以上より, 系内容数分布 π_ℓ ($\ell = 0, 1, \dots$) に対する数値近似解は次式で与えられる.

$$\pi_0^{\text{comp}} = \left[1 + \sum_{\ell=1}^{m^*} r_\ell^{\text{comp}} \right]^{-1}, \quad \pi_\ell^{\text{comp}} = \pi_0^{\text{comp}} r_\ell^{\text{comp}}, \quad \ell = 1, 2, \dots, m^* \quad (40)$$

ここで, $\pi_{m^*+1}, \pi_{m^*+2}, \dots$ は暗黙に 0 で近似されていることに注意せよ.

3.2. 誤差上界

上記のアルゴリズムにおいて生じる切断誤差を精密に見積もることは容易ではない。一方で, [6] において計算法が示されている π_0 の下界 π_0^{low} を用いることで, 切断誤差の上界が得られる。 r_ℓ は非負の数の和と積で表されており, 切断近似の結果, 明らかに $r_\ell^{\text{comp}} \leq r_\ell$ が成立する。そこで, Δ_ℓ を

$$\Delta_\ell = \begin{cases} r_\ell - r_\ell^{\text{comp}}, & \ell = 1, 2, \dots, m^* \\ r_\ell, & \ell = m^* + 1, m^* + 2, \dots \end{cases}$$

とすると $\Delta_\ell \geq 0$ ($\ell = 1, 2, \dots$) であり, さらに $\Delta = \sum_{\ell=1}^{\infty} \Delta_\ell$ とおくと,

$$\Delta = \sum_{\ell=1}^{\infty} r_\ell - \sum_{\ell=1}^{m^*} r_\ell^{\text{comp}} = \frac{1}{\pi_0} - \frac{1}{\pi_0^{\text{comp}}}$$

が成立する。単純な計算により,

$$\begin{aligned} |\pi_\ell^{\text{comp}} - \pi_\ell| &= \left| \frac{r_\ell^{\text{comp}}}{1 + \sum_{\ell=1}^{m^*} r_\ell^{\text{comp}}} - \frac{r_\ell^{\text{comp}} + \Delta_\ell}{1 + \sum_{\ell=1}^{m^*} r_\ell^{\text{comp}} + \Delta} \right| = \pi_0 \Delta \left| \pi_\ell^{\text{comp}} - \frac{\Delta_\ell}{\Delta} \right| \\ &\leq \pi_0 \Delta \cdot \max(\pi_\ell^{\text{comp}}, \frac{\Delta_\ell}{\Delta}) \\ &\leq \pi_0 \Delta \\ &= 1 - \frac{\pi_0}{\pi_0^{\text{comp}}}, \quad \ell = 1, 2, \dots \end{aligned}$$

が確かめられるため, δ^{up} を

$$\delta^{\text{up}} = 1 - \frac{\pi_0^{\text{low}}}{\pi_0^{\text{comp}}} \quad (41)$$

と定義すると, 任意の $\ell = 1, 2, \dots$ に対して $|\pi_\ell^{\text{comp}} - \pi_\ell| \leq \delta^{\text{up}}$ が成立することがわかる。

本稿の数値計算アルゴリズムを図 1 にまとめる。手順 (e-I-1) において, (8) で定義される $h_e^{[m]}$ ($m = 0, 1, \dots$) を計算する必要があるが, これについては, サービス時間分布が一定分布, 相型分布, パレート分布に従う場合の計算法が [6] に示されている。手順 (h) と (i) は, 誤差上界 δ^{up} を求めない場合には不要であるが, δ^{up} を求める場合, [6] に従って π_0 の下界 π_0^{low} を計算する必要がある。

4. 数値例

本稿のアルゴリズムを用いた数値例を示す。ここでは, サービス時間および待ち時間制約長ともに 5 次のアーラン分布を用いる。また, 常に $E[H] = 1$ となるように時間の単位を定める。全ての数値例について, 切断点を決めるパラメータは $k^* = 1000$, $\epsilon_g = 10^{-11}$, $\epsilon_r = 10^{-9}$, $\epsilon_M = 10^{-9}$ と選んだ。

図 2 に, $E[G]$ を 20 に固定して ρ を変化させたときの, 系内客数分布の数値結果を示す。この図より, π_ℓ^{comp} ($\ell = 0, 1, \dots$) は ρ が小さいとき単調減少であり, ρ が大きくなると単峰型となることがわかる。図 2 に, この確率分布 π_ℓ^{comp} ($\ell = 0, 1, \dots$) を標準化した, すなわち平均を減じ, 標準偏差で除した確率変数の密度関数を示す。参考のため, 図 2 には標準正規分布 $\mathcal{N}(0, 1)$ の密度関数をあわせてプロットしている。この図から, ρ が大きいとき, 標準化された系内客数分布は標準正規分布に漸近することがわかる。表 1 に, この数値結果に対応する誤差上界 δ^{up} を示す。表 1 より, ρ が増加するにつれて δ^{up} は大きく増加し, 計算精度を保証するのが難しくなることがわかる。

次に, $\rho = 1.1$ に固定して $E[G]$ を変化させたときの, 系内客数分布の数値結果を図 4 に示す。図より, π_ℓ^{comp} ($\ell = 0, 1, \dots$) の概形は, $E[G]$ が小さいとき急尖的であり, $E[G]$ が大きくなるにつれて緩尖的となることがわかる。図 5 に, 図 3 と同様にして標準化した π_ℓ^{comp} の密度関数を示す。図 5 より, $E[G]$ が大きくなるにつれて標準化された系内客数分布は標準正規分布に漸近することがわかる。表 2 に, この数値結果に対応する誤差上界 δ^{up} を示す。表 2 より, δ^{up} は $E[G]$ が増加するにつれて大きくなるが, その度合いは ρ を増加させた場合 (表 1) に比べると緩やかであることがわかる。

入力: $\lambda, H(x) (x \geq 0), (\alpha, \mathbf{T}), k^*, \epsilon_g, \epsilon_r, \epsilon_M$.

出力: $m^*, \pi_\ell (\ell = 1, 2, \dots, m^*), \delta^{\text{up}}$.

- (a) (28) を用いて $\mathbf{g}^{*,\text{comp}}$ を計算し, (27) に基づいて k_g^* を決定する.
- (b) (38) より m^* を定める.
- (c) (29) および (30) を用いて, $\mathbf{g}_e^{*,\text{comp}}$ および $\bar{\mathbf{g}}_e^{*,\text{comp}}$ を計算する.
- (d) (31)–(33) を用いて, $\kappa_{m,\ell}^{*,\text{comp}} (\ell = 0, 1, \dots, m, m = 1, 2, \dots, m^* - 1)$ を計算する.
- (e) $m := 1$ とおく.
 - (e-I) $n := 1$ とおき, (34) より $\hat{r}_{m,\ell}^{\text{comp}}(1) (\ell = 0, 1, \dots, m)$ を計算する.
 - (e-I-1) (35) より $\hat{r}_{m,\ell}^{\text{comp}}(n+1) (\ell = 0, 1, \dots, m)$ を計算する.
 - (e-I-2) 全ての $\ell = 0, 1, \dots, m$ に対し $\hat{r}_{m,\ell}^{\text{comp}}(n+1) < \epsilon_r e$ ならば (e-II) へ進む. そうでなければ, $n := n+1$ として (e-I-1) へ戻る.
 - (e-II) (36) より $\hat{r}_{m,\ell}^{\text{comp}} (\ell = 0, 1, \dots, m)$ を計算し, $m = m^* - 1$ ならば (f) へ進む. そうでなければ $m := m+1$ として (e-I) へ戻る.
- (f) (37) より $\hat{r}_{0,0}^{\text{comp}}$ を計算する.
- (g) (39) より $r_\ell^{\text{comp}} (\ell = 1, 2, \dots, m^*)$ を計算し, (40) より $\pi_\ell (\ell = 0, 1, \dots, m^*)$ を計算する.
- (h) [6] のアルゴリズムを用いて, π_0 の下界 π_0^{low} を計算する.
- (i) (41) より δ^{up} を計算する.

図 1: 系内客数分布の数値計算アルゴリズム

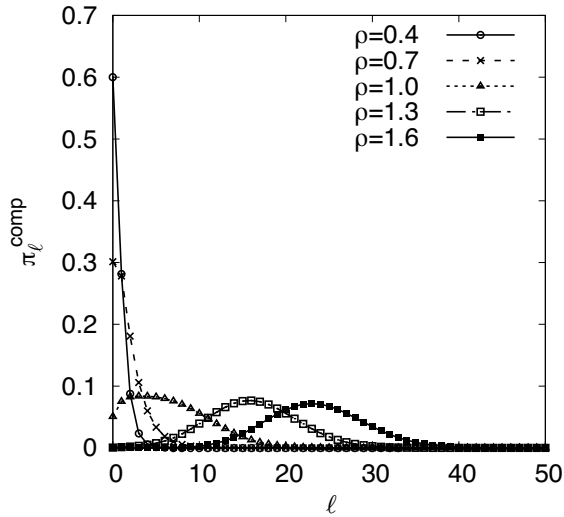


図 2: $E[G] = 20$ の M/Er₅/Er₅ 待ち行列における系内客数分布 π_ℓ^{comp}

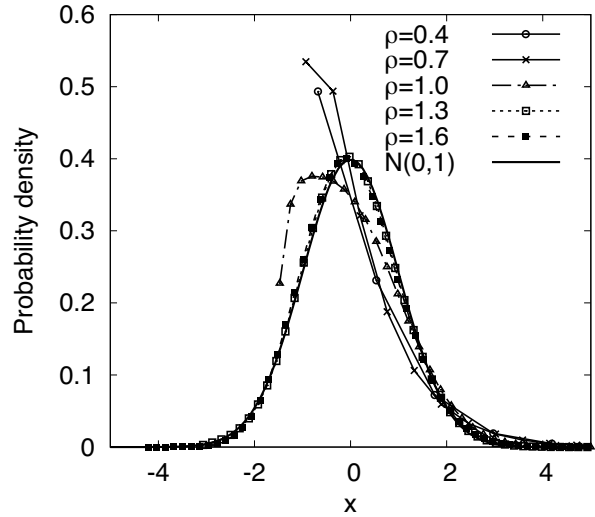


図 3: $E[G] = 20$ の M/Er₅/Er₅ 待ち行列における標準化された系内客数の確率密度

表 1: 図 2 の数値結果に対する誤差上界

ρ	0.4	0.7	1.0	1.3	1.6
δ^{up}	1.8×10^{-9}	1.8×10^{-9}	1.9×10^{-9}	1.9×10^{-6}	4.5×10^{-4}

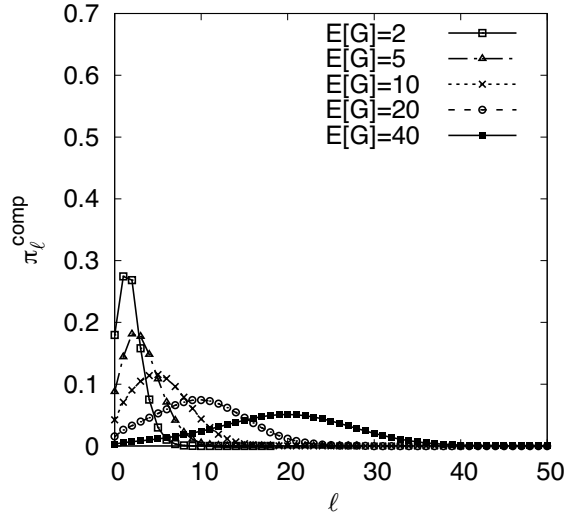


図 4: $\rho = 1.1$ の $M/Er_5/Er_5$ 待ち行列における系内客数分布 π_ℓ^{comp}

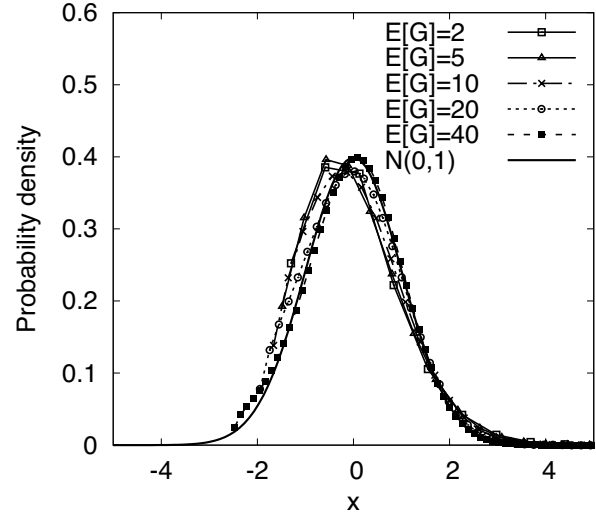


図 5: $\rho = 1.1$ の $M/Er_5/Er_5$ 待ち行列における標準化された系内客数の確率密度

表 2: 図 4 の数値結果に対する誤差上界

$E[G]$	2	5	10	20	40
δ^{up}	10^{-9} 未満	1.2×10^{-9}	3.3×10^{-9}	1.4×10^{-8}	9.5×10^{-8}

5. まとめ

本稿では, $M/G/1+PH$ 待ち行列における系内客数分布の数値計算法を考案した. 構築した計算アルゴリズムは図 1 にまとめている. また, 数値例を通じて, トラヒック強度 ρ と待ち時間制約の長さが系内客数分布に与える影響を考察した.

謝辞

本研究の一部は JSPS 科研費 JP16H06914 の支援を受けている.

参考文献

- [1] F. Baccelli, P. Boyer, and G. Hebuterne, Single-server queues with impatient customers. Adv. Appl. Prob., 16, 887–905, 1984.
- [2] D. J. Daley, General customer impatience in the queue $GI/G/1$. J. Appl. Prob., 2, 186–205, 1965.
- [3] Y. Inoue and T. Takine, Analysis of the loss probability in the $M/G/1+G$ queue. Queueing Syst., 80, 363–386, 2015.
- [4] Y. Inoue and T. Takine, The $M/D/1+D$ queue has the minimum loss probability among $M/G/1+G$ queues. Oper. Res. Lett., 43, 629–632, 2015.
- [5] 井上文彰, $M/G/1+PH$ 待ち行列における系内客数分布について. 日本オペレーションズ・リサーチ学会 2016 年秋季研究発表会アブストラクト集, 254–255, 2016.
- [6] Y. Inoue and T. Takine, A computational algorithm for the loss probability in the $M/G/1+PH$ queue. To appear in Stoch. Models.
- [7] I. N. Kovalenko, Some queueing problems with restrictions. Theor. Prob. Appl., 6, 205–208, 1961.

Distribution of the Times to Service Completion and Abandonment in the M/M/ m Preemptive LCFS Queue with Impatient Customers

Hideaki Takagi

Professor Emeritus, University of Tsukuba

takagi@sk.tsukuba.ac.jp

Abstract We study an M/M/ m preemptive last-come, first-served queue with impatient customers without priority classes. We focus on the time interval from the arrival to either service completion or to abandonment of an arbitrary customer in the steady state. The problem is formulated as a combination of two one-dimensional birth-and-death processes each with two absorbing states. We give explicit expressions in terms of Laplace-Stieltjes transforms of the distribution functions for the time to service completion or abandonment which is decomposed into the waiting time (time in limbo) and the service time. A numerical example is presented in order to demonstrate the computation of theoretical formulas.

1. Introduction

We consider an M/M/ m queueing system with impatient customers without exogenous priority classes. Customers arrive in a Poisson process with rate λ . The service time of each customer is exponentially distributed with mean $1/\mu$. There are m servers and an infinite capacity of the waiting room. Let us define a parameter $\rho := \lambda/(m\mu)$. Each customer present in the system is either being served or in limbo in the waiting room at any time. A customer only in limbo leaves the system (abandons waiting) with probability $\theta\Delta t$ during a short interval $(t, t + \Delta t)$. In other words, the patience time of each customer is exponentially distributed with mean $1/\theta$. We use another parameter $\tau := \theta/\mu$, which is the ratio of the mean service time to the mean patience time.

It is assumed that the service to each customer is started immediately upon arrival. If all servers are busy, the arriving customer preempts the ongoing service to the customer who arrived first among those who are being served. The customer whose service is preempted is placed at the head of the queue in the waiting room. When one of the servers becomes available, a customer at the head of the queue, if any, is called in for service to be resumed. This discipline is equivalent to the one called “preemptive *last-in, first-out* (LIFO)” for an M/G/1 queue by Wolff [4, p. 456].

We study the time interval from the arrival to either service completion or to abandonment of an arbitrary customer in the steady state. In our previous work [3], we formulated the problem as a one-dimensional birth-and-death process with two absorbing states and considered the first passage times in this process. We gave explicit expressions for the probabilities of service completion and abandonment. We also showed the sets of computational formulas for calculating the mean and the second moment of the times to service completion and abandonment.

In the present paper, we turn our attention to the distribution of the times to service completion and abandonment. We formulate the problem as a combination of two one-dimensional birth-and-death processes each with two absorbing states and consider the first passage times in these processes. We give explicit expressions in terms of Laplace-Stieltjes transforms (LST) of the distribution functions (DF) for the time to service completion or abandonment which is decomposed into the waiting time (time in limbo) and the service time. We present a numerical example for the mean and the covariance of conditional waiting and service times.

An analytic technique shown in this paper should be interesting on its own. It can also be extended and applied to the analysis of M/M/ m preemptive-resume priority queues with impatient customers in which customers of the same class are served in either first-come, first-served (FCFS) or last-come, first-served (LCFS) fashion.

2. Time to preemption or service completion from state k , $0 \leq k \leq m-1$

We begin analysis with focusing on a tagged customer in state k , signifying that there are k other customers who compete against him for service at any given time in the steady state. They are the customers who arrived after the tagged one and have been staying in the system until that time. Therefore, arriving customer starts with state 0.

We first consider a birth-and-death process for the time to preemption or service completion starting with state k , $0 \leq k \leq m-1$, in which the tagged customer is being served.

2.1. Birth-and-death process for the time to preemption or service completion

The state transition rate diagram for the one-dimensional birth-and-death process modeling the time to preemption or service completion is shown in Fig. 1. The process has m transient states and two absorbing states. The transition rates and the LST of the DF for the time spent by the tagged customer in state k are given by

$$\alpha_k = \frac{k\mu}{\lambda + (k+1)\mu} \quad ; \quad \beta_k = \frac{\mu}{\lambda + (k+1)\mu} \quad ; \quad B_k^*(s) = \frac{\lambda + (k+1)\mu}{s + \lambda + (k+1)\mu}.$$

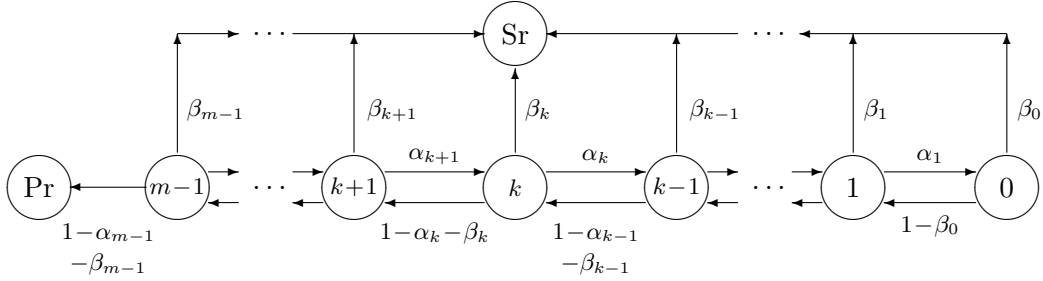


Figure 1: State transitions for a customer until preemption or service completion.

2.2. LST of the DF for the time to preemption or service completion

Let $H_k^*(s, \text{Pr})$ be the joint probability of service preemption and the LST of the DF for the time to move from state k to state m (“Pr”) without service completion (“Sr”). Let $H_k^*(s, \text{Sr})$ be the joint probability of service completion and the LST of the DF for the time to move from state k to state “Sr” without reaching state “Pr”.

Finite sets of equations for $\{H_k^*(s, \text{Pr}); 0 \leq k \leq m-1\}$ and $\{H_k^*(s, \text{Sr}); 0 \leq k \leq m-1\}$ are given by

$$(s + \lambda + \mu)H_0^*(s, \text{Pr}) = \lambda H_1^*(s, \text{Pr}),$$

$$[s + \lambda + (k+1)\mu]H_k^*(s, \text{Pr}) = k\mu H_{k-1}^*(s, \text{Pr}) + \lambda H_{k+1}^*(s, \text{Pr}) \quad 1 \leq k \leq m-2,$$

$$(s + \lambda + m\mu)H_{m-1}^*(s, \text{Pr}) = (m-1)\mu H_{m-2}^*(s, \text{Pr}) + \lambda$$

and

$$(s + \lambda + \mu)H_0^*(s, \text{Sr}) = \mu + \lambda H_1^*(s, \text{Sr}),$$

$$[s + \lambda + (k+1)\mu]H_k^*(s, \text{Sr}) = k\mu H_{k-1}^*(s, \text{Sr}) + \mu + \lambda H_{k+1}^*(s, \text{Sr}) \quad 1 \leq k \leq m-2,$$

$$(s + \lambda + m\mu)H_{m-1}^*(s, \text{Sr}) = (m-1)\mu H_{m-2}^*(s, \text{Sr}) + \mu,$$

respectively. The solution can be obtained in terms of $\{h_k^*(s); 0 \leq k \leq m\}$ in the form

$$H_k^*(s, \text{Pr}) = \frac{h_k^*(s)}{h_m^*(s)} \quad ; \quad H_k^*(s, \text{Sr}) = \frac{\mu}{s + \mu} \left[1 - \frac{h_k^*(s)}{h_m^*(s)} \right] \quad 0 \leq k \leq m.$$

In addition, let $H_m^*(s, \text{Pr}) \equiv 1$ and $H_m^*(s, \text{Sr}) \equiv 0$.

2.3. Solution for $\{h_k^*(s); 0 \leq k \leq m\}$

A finite set of equations for $\{h_k^*(s); 0 \leq k \leq m\}$ is given by

$$\begin{aligned} h_0^*(s) &= 1 \quad ; \quad s + \lambda + \mu = \lambda h_1^*(s), \\ [s + \lambda + (k+1)\mu]h_k^*(s) &= k\mu h_{k-1}^*(s) + \lambda h_{k+1}^*(s) \quad 1 \leq k \leq m-1, \end{aligned}$$

which can be written as the following set of recurrence relations:

$$h_k^*(s) = \frac{s + \lambda + k\mu}{\lambda} h_{k-1}^*(s) - \frac{(k-1)\mu}{\lambda} h_{k-2}^*(s) \quad 2 \leq k \leq m.$$

The solution is given by *Cramer's formula* as the determinant of the $k \times k$ tridiagonal matrix

$$h_k^*(s) = (-1)^k \begin{vmatrix} -\frac{s+\lambda+\mu}{\lambda} & 1 & 0 & 0 & 0 & \cdots & 0 \\ \frac{\mu}{\lambda} & -\frac{s+\lambda+2\mu}{\lambda} & 1 & 0 & 0 & \cdots & 0 \\ 0 & \frac{2\mu}{\lambda} & -\frac{s+\lambda+3\mu}{\lambda} & 1 & 0 & \cdots & 0 \\ 0 & 0 & \frac{3\mu}{\lambda} & -\frac{s+\lambda+4\mu}{\lambda} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\frac{s+\lambda+(k-1)\mu}{\lambda} & 1 \\ 0 & 0 & 0 & 0 & \cdots & \frac{(k-1)\mu}{\lambda} & -\frac{s+\lambda+k\mu}{\lambda} \end{vmatrix}$$

for $1 \leq k \leq m$. Note that $h_k^*(s)$ is a k th-order polynomial in s , the coefficient of s^k being $(1/\lambda)^k$.

At this moment, we do not have a simple expression for $h_k^*(s)$. For $s = 0$, we get

$$h_k^*(0) = \sum_{j=0}^k \frac{(m\rho)^j}{j!} \bigg/ \frac{(m\rho)^k}{k!} = \frac{1}{B(k, m\rho)} \quad 0 \leq k \leq m$$

with Erlang's B formula:

$$B(0, a) = 1 \quad ; \quad B(m, a) = \frac{a^m}{m!} \bigg/ \sum_{j=0}^m \frac{a^j}{j!} \quad m = 1, 2, \dots$$

Thus we have the probabilities of preemption and service completion

$$p_k\{\text{Pr}\} := H_k^*(0, \text{Pr}) = \frac{B(m, m\rho)}{B(k, m\rho)} \quad ; \quad p_k\{\text{Sr}\} := H_k^*(0, \text{Sr}) = 1 - p_k\{\text{Pr}\} \quad 0 \leq k \leq m.$$

In particular,

$$p_0\{\text{Pr}\} = B(m, m\rho) \quad ; \quad p_m\{\text{Pr}\} = 1 \quad ; \quad p_{m-1}\{\text{Pr}\} = \rho[1 - B(m, m\rho)].$$

The derivatives of $\{h_k^*(s); 0 \leq k \leq m\}$ at $s = 0$ are given by

$$h'_0(0) = 0 \quad ; \quad h'_k(0) := \frac{dh_k^*(s)}{ds} \bigg|_{s=0} = \frac{1}{\lambda} \sum_{j=1}^k \frac{(m\rho)^j}{j!} \sum_{l=0}^{j-1} h_l^*(0) \bigg/ \frac{(m\rho)^k}{k!} \quad 1 \leq k \leq m.$$

The ℓ th derivatives of $\{h_k^*(s); 0 \leq k \leq m\}$ at $s = 0$, $\ell = 2, 3, \dots$, are recursively given by

$$\begin{aligned} h_0^{(\ell)}(0) &= h_1^{(\ell)}(0) = \cdots = h_{\ell-1}^{(\ell)}(0) = 0, \\ h_k^{(\ell)}(0) &:= \frac{d^\ell h_k^*(s)}{ds^\ell} \bigg|_{s=0} = \frac{\ell}{\lambda} \sum_{j=1}^k \frac{(m\rho)^j}{j!} \sum_{l=0}^{j-1} h_l^{(\ell-1)}(0) \bigg/ \frac{(m\rho)^k}{k!} \quad \ell \leq k \leq m. \end{aligned}$$

3. Time to service resumption or abandonment from state k , $k \geq m$

We next consider another birth-and-death process for the time to service resumption or abandonment from state k , $k \geq m$, in which the tagged customer is waiting.

3.1. Birth-and-death process for the time to service resumption or abandonment

The state transition rate diagram for the one-dimensional birth-and-death process modeling the time to service resumption or abandonment is shown in Fig. 2. The process has an infinite number of transient states and two absorbing states. The transition rates and the LST of the DF for the time spent by the tagged customer in state k are given by

$$\alpha'_k = \frac{m\mu + (k-m)\theta}{\lambda + m\mu + (k+1-m)\theta} \quad ; \quad \beta'_k = \frac{\theta}{\lambda + m\mu + (k+1-m)\theta},$$

$$B_k^*(s) = \frac{\lambda + m\mu + (k+1-m)\theta}{s + \lambda + m\mu + (k+1-m)\theta}.$$

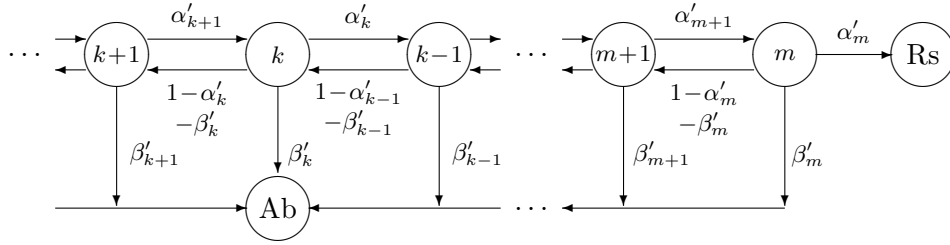


Figure 2: State transitions for a customer until service resumption or abandonment.

3.2. LST of the DF for the time to service resumption or abandonment

Let $W_k^*(s, \text{Rs})$ be the joint probability of service resumption and the LST of the DF for the time to move from state k to state $m-1$ (“Rs”) without reaching abandonment (“Ab”). Let $W_k^*(s, \text{Ab})$ be the joint probability of abandonment and the LST of the DF for the time to move state k to “Ab” without reaching state “Rs”.

Infinite sets of equations for $\{W_k^*(s, \text{Rs}); k \geq m\}$ and $\{W_k^*(s, \text{Ab}); k \geq m\}$ are given by

$$(s + \lambda + m\mu + \theta)W_m^*(s, \text{Rs}) = \lambda W_{m+1}^*(s, \text{Rs}),$$

$$[s + \lambda + m\mu + (k+1-m)\theta]W_k^*(s, \text{Rs})$$

$$= [m\mu + (k-m)\theta]W_{k-1}^*(s, \text{Rs}) + \lambda W_{k+1}^*(s, \text{Rs}) \quad k \geq m+1$$

and

$$(s + \lambda + m\mu + \theta)W_m^*(s, \text{Ab}) = \theta + \lambda W_{m+1}^*(s, \text{Ab}),$$

$$[s + \lambda + m\mu + (k+1-m)\theta]W_k^*(s, \text{Ab})$$

$$= [m\mu + (k-m)\theta]W_{k-1}^*(s, \text{Ab}) + \theta + \lambda W_{k+1}^*(s, \text{Ab}) \quad k \geq m+1,$$

respectively. The solution can be obtained in terms of $\{G_k^*(s); k \geq m\}$ in the form

$$W_k^*(s, \text{Rs}) = G_k^*(s + \theta) \quad ; \quad W_k^*(s, \text{Ab}) = \frac{\theta}{s + \theta}[1 - G_k^*(s + \theta)] \quad k \geq m.$$

Thus the probabilities of service preemption and abandonment are given by

$$p_k\{\text{Rs}\} := W_k^*(0, \text{Rs}) = G_k^*(\theta) = 1 - p_k\{\text{Ab}\} \quad k \geq m.$$

3.3. Busy period

A *busy period* started with k ($\geq m$) customers in an M/M/ m queue is the time interval from the instant at which there are k customers in the system (all servers are busy and $k - m$ customers are waiting) to the first instant at which any one of the servers becomes available. The LST of the DF for such a period, denoted $G_k^*(s)$, satisfies the following set of equations:

$$(s + \lambda + m\mu)G_m^*(s) = \lambda G_{m+1}^*(s) + m\mu,$$

$$[s + \lambda + m\mu + (k - m)\theta]G_k^*(s) = [m\mu + (k - m)\theta]G_{k-1}^*(s) + \lambda G_{k+1}^*(s) \quad k \geq m + 1.$$

Subba Rao [2] derived the LST of the DF for the duration of a busy period in an M/G/1 queue with impatient customers, which is also shown in [1]. We can obtain the LST of the DF for the duration of the busy period in the M/M/ m queue by using the exponentially distributed service times with mean $1/(m\mu)$ in their result. Thus we have

$$G_k^*(s) = \frac{\frac{m\mu}{s + m\mu} + \sum_{i=1}^{\infty} (-1)^i \psi_{i,k-m}(\lambda/\theta) \left[\prod_{j=0}^{i-1} \left(1 - \frac{m\mu}{s + m\mu + j\theta} \right) \right] \frac{m\mu}{s + m\mu + i\theta}}{1 + \sum_{i=1}^{\infty} \frac{(\lambda/\theta)^i}{i!} \prod_{j=0}^{i-1} \left(1 - \frac{m\mu}{s + m\mu + j\theta} \right)} \quad k \geq m,$$

where we have defined

$$\psi_{i,k}(x) := \sum_{j=\max\{0, i-k\}}^i \frac{(-x)^j}{j!} \binom{k}{i-j} \quad i \geq 1, \quad k \geq 0.$$

Thus

$$G_k^*(\theta) = \sum_{i=0}^{\infty} \frac{i!(-\tau/m)^i \psi_{i,k-m}(\lambda/\theta)}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \bigg/ \sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=0}^i (1 + j\tau/m)}$$

and

$$\begin{aligned} m\mu G_k'(\theta) &:= m\mu [dG_k^*(s)/ds]_{s=\theta} \\ &= \left\{ \left[\sum_{i=1}^{\infty} \frac{i!(-\tau/m)^i \psi_{i,k-m}(\lambda/\theta)}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \sum_{j=1}^i \frac{1}{(1 + j\tau/m)j\tau/m} \right] - \left[\sum_{i=0}^{\infty} \frac{i!(-\tau/m)^i \psi_{i,k-m}(\lambda/\theta)}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \cdot \frac{1}{1 + (i+1)\tau/m} \right] \right\} \\ &\quad \bigg/ \left[\sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=0}^i (1 + j\tau/m)} \right] - \left[\sum_{i=0}^{\infty} \frac{i!(-\tau/m)^i \psi_{i,k-m}(\lambda/\theta)}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \right] \left[\sum_{i=1}^{\infty} \frac{\rho^i}{\prod_{j=1}^i (1 + j\tau/m)} \sum_{j=1}^i \frac{1}{(1 + j\tau/m)j\tau/m} \right] \\ &\quad \bigg/ \left[\sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=0}^i (1 + j\tau/m)} \right]^2. \end{aligned}$$

In particular, since $\psi_{i,0}(x) = (-x)^i/i!$, we get

$$G_m^*(s) = \frac{\frac{m\mu}{s + m\mu} + \sum_{i=1}^{\infty} \frac{(\lambda/\theta)^i}{i!} \left[\prod_{j=0}^{i-1} \left(1 - \frac{m\mu}{s + m\mu + j\theta} \right) \right] \frac{m\mu}{s + m\mu + i\theta}}{1 + \sum_{i=1}^{\infty} \frac{(\lambda/\theta)^i}{i!} \prod_{j=0}^{i-1} \left(1 - \frac{m\mu}{s + m\mu + j\theta} \right)}$$

so that

$$G_m^*(\theta) = 1 - \frac{\theta}{\lambda} \sum_{i=1}^{\infty} \frac{i\rho^i}{\prod_{j=1}^i (1 + j\tau/m)} \bigg/ \sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=0}^i (1 + j\tau/m)}$$

and

$$m\mu G'_m(\theta) = \left[\sum_{i=1}^{\infty} \frac{\rho^i}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \sum_{j=1}^i \frac{1}{(1 + j\tau/m)j\tau/m} - \sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \cdot \frac{1}{1 + (i+1)\tau/m} \right] \\ \left/ \left[\sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=0}^i (1 + j\tau/m)} \right] \right. \\ \left. - \left[\sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=1}^{i+1} (1 + j\tau/m)} \right] \left[\sum_{i=1}^{\infty} \frac{\rho^i}{\prod_{j=1}^i (1 + j\tau/m)} \sum_{j=1}^i \frac{1}{(1 + j\tau/m)j\tau/m} \right] \left/ \left[\sum_{i=0}^{\infty} \frac{\rho^i}{\prod_{j=0}^i (1 + j\tau/m)} \right] \right|^2.$$

4. Joint Distribution of the Time in Limbo and the Service Time

We are now in a position to consider the distribution of the time until the departure (either abandonment or service completion) for the tagged customer in state k (≥ 0). Such a time consists of the waiting time (the time in limbo) and the time being served. Therefore, we derive the joint LST of the DF for the waiting time and the service time for the customer who abandons waiting, denoted $\hat{T}_k^*(s, s', \text{Ab})$, and that for the customer who gets served until completion, denoted $\hat{T}_k^*(s, s', \text{Sr})$. Then we get the marginal LSTs of the DFs for the waiting time, the service time, and the total time spent in the system as follows.

$$\begin{aligned} \hat{W}_k^*(s, \text{Ab}) &:= \hat{T}_k^*(s, 0, \text{Ab}) & \hat{H}_k^*(s, \text{Ab}) &:= \hat{T}_k^*(0, s, \text{Ab}) & \hat{T}_k^*(s, \text{Ab}) &:= \hat{T}_k^*(s, s, \text{Ab}), \\ \hat{W}_k^*(s, \text{Sr}) &:= \hat{T}_k^*(s, 0, \text{Sr}) & \hat{H}_k^*(s, \text{Sr}) &:= \hat{T}_k^*(0, s, \text{Sr}) & \hat{T}_k^*(s, \text{Sr}) &:= \hat{T}_k^*(s, s, \text{Sr}). \end{aligned}$$

4.1. Time to abandonment

We first consider the time to abandonment for a customer who abandons waiting before service completion. For the customer in state k , $0 \leq k \leq m-1$, we have

$$\hat{T}_k^*(s, s', \text{Ab}) = \frac{H_k^*(s', \text{Pr})W_m^*(s, \text{Ab})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s', \text{Pr})} = \frac{\theta}{s + \theta} \cdot \frac{h_k^*(s')[1 - G_m^*(s + \theta)]}{h_m^*(s') - h_{m-1}^*(s')G_m^*(s + \theta)},$$

which leads to the marginal distributions

$$\begin{aligned} \hat{W}_k^*(s, \text{Ab}) &= \frac{p_k \{\text{Pr}\} W_m^*(s, \text{Ab})}{1 - p_{m-1} \{\text{Pr}\} W_m^*(s, \text{Rs})} = \frac{\theta}{s + \theta} \cdot \frac{h_k^*(0)[1 - G_m^*(s + \theta)]}{h_m^*(0) - h_{m-1}^*(0)G_m^*(s + \theta)}, \\ \hat{H}_k^*(s, \text{Ab}) &= \frac{p_m \{\text{Ab}\} H_k^*(s, \text{Pr})}{1 - p_m \{\text{Rs}\} H_{m-1}^*(s, \text{Pr})} = \frac{h_k^*(s)[1 - G_m^*(\theta)]}{h_m^*(s) - h_{m-1}^*(s)G_m^*(\theta)}, \\ \hat{T}_k^*(s, \text{Ab}) &= \frac{H_k^*(s, \text{Pr})W_m^*(s, \text{Ab})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s, \text{Pr})} = \frac{\theta}{s + \theta} \cdot \frac{h_k^*(s)[1 - G_m^*(s + \theta)]}{h_m^*(s) - h_{m-1}^*(s)G_m^*(s + \theta)}. \end{aligned}$$

Then we get the probability of abandonment

$$P_k \{\text{Ab}\} = \hat{W}_k^*(0, \text{Ab}) = \hat{H}_k^*(0, \text{Ab}) = \hat{T}_k^*(0, \text{Ab}) = \frac{h_k^*(0)[1 - G_m^*(\theta)]}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)},$$

the mean waiting and service times

$$\begin{aligned} E[\hat{W}_k, \text{Ab}] &= \frac{1}{\theta} P_k \{\text{Ab}\} - \frac{h_k^*(0)[h_{m-1}^*(0) - h_m^*(0)]G'_m(\theta)}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2}, \\ E[\hat{H}_k, \text{Ab}] &= [1 - G_m^*(\theta)] \left\{ \frac{h_k^*(0)[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2} - \frac{h'_k(0)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)} \right\}, \end{aligned}$$

and the covariance $\text{Cov}[\hat{W}_k, \hat{H}_k; \text{Ab}] := E[\hat{W}_k \hat{H}_k, \text{Ab}] - E[\hat{W}_k, \text{Ab}]E[\hat{H}_k, \text{Ab}]$ with

$$\begin{aligned} E[\hat{W}_k \hat{H}_k, \text{Ab}] &= \frac{1}{\theta} E[\hat{H}_k, \text{Ab}] + G'_m(\theta) \left\{ \frac{2h_k^*(0)[h_m^*(0) - h_{m-1}^*(0)][h'_m(0) - h'_{m-1}(0)G_m^*(\theta)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^3} \right. \\ &\quad \left. - \frac{h_k^*(0)[h'_m(0) - h'_{m-1}(0)] + h'_k(0)[h_m^*(0) - h_{m-1}^*(0)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2} \right\}. \end{aligned}$$

For the customer in state k , $k \geq m$, we have

$$\begin{aligned}\hat{T}_k^*(s, s', \text{Ab}) &= W_k^*(s, \text{Ab}) + \frac{W_k^*(s, \text{Rs})H_{m-1}^*(s', \text{Pr})W_m^*(s, \text{Ab})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s', \text{Pr})} \\ &= \frac{\theta}{s + \theta} \left\{ 1 - \frac{[h_m^*(s') - h_{m-1}^*(s')]G_k^*(s + \theta)}{h_m^*(s') - h_{m-1}^*(s')G_m^*(s + \theta)} \right\},\end{aligned}$$

which leads to the marginal distributions

$$\begin{aligned}\hat{W}_k^*(s, \text{Ab}) &= W_k^*(s, \text{Ab}) + \frac{p_{m-1}\{\text{Pr}\}W_k^*(s, \text{Rs})W_m^*(s, \text{Ab})}{1 - p_{m-1}\{\text{Pr}\}W_m^*(s, \text{Rs})} = \frac{\theta}{s + \theta} \left\{ 1 - \frac{[h_m^*(0) - h_{m-1}^*(0)]G_k^*(s + \theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(s + \theta)} \right\}, \\ \hat{H}_k^*(s, \text{Ab}) &= p_k\{\text{Ab}\} + \frac{p_k\{\text{Rs}\}p_m\{\text{Ab}\}H_{m-1}^*(s, \text{Pr})}{1 - p_m\{\text{Rs}\}H_{m-1}^*(s, \text{Pr})} = 1 - \frac{[h_m^*(s) - h_{m-1}^*(s)]G_k^*(\theta)}{h_m^*(s) - h_{m-1}^*(s)G_m^*(\theta)}, \\ \hat{T}_k^*(s, \text{Ab}) &= W_k^*(s, \text{Ab}) + \frac{W_k^*(s, \text{Rs})H_{m-1}^*(s, \text{Pr})W_m^*(s, \text{Ab})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s, \text{Pr})} = \frac{\theta}{s + \theta} \left\{ 1 - \frac{[h_m^*(s) - h_{m-1}^*(s)]G_k^*(s + \theta)}{h_m^*(s) - h_{m-1}^*(s)G_m^*(s + \theta)} \right\}.\end{aligned}$$

Then we get the probability of abandonment

$$P_k\{\text{Ab}\} = 1 - \frac{[h_m^*(0) - h_{m-1}^*(0)]G_k^*(\theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)},$$

the mean waiting and service times

$$\begin{aligned}E[\hat{W}_k, \text{Ab}] &= \frac{1}{\theta} P_k\{\text{Ab}\} - [h_{m-1}^*(0) - h_m^*(0)] \left\{ \frac{G_k'(\theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)} + \frac{h_{m-1}^*(0)G_k^*(\theta)G_m'(\theta)}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2} \right\}, \\ E[\hat{H}_k, \text{Ab}] &= \frac{G_k^*(\theta)[h_m'(0)h_{m-1}^*(0) - h_{m-1}'(0)h_m^*(0)][1 - G_m^*(\theta)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2},\end{aligned}$$

and

$$\begin{aligned}E[\hat{W}_k \hat{H}_k, \text{Ab}] &= \frac{1}{\theta} E[\hat{H}_k, \text{Ab}] + \frac{h_m'(0)h_{m-1}^*(0) - h_{m-1}'(0)h_m^*(0)}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^3} \\ &\times \{G_k^*(\theta)G_m'(\theta)[h_m^*(0) - 2h_{m-1}^*(0) + G_m^*(\theta)h_{m-1}^*(0)] - G_k'(\theta)[1 - G_m^*(\theta)][h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]\}.\end{aligned}$$

4.2. Time to service completion

We next consider the time to service completion for a customer who gets served. For the customer in state k , $0 \leq k \leq m - 1$, we have

$$\begin{aligned}\hat{T}_k^*(s, s', \text{Sr}) &= H_k^*(s', \text{Sr}) + \frac{H_k^*(s', \text{Pr})W_m^*(s, \text{Rs})H_{m-1}^*(s', \text{Sr})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s', \text{Pr})} \\ &= \frac{\mu}{s' + \mu} \left\{ 1 - \frac{h_k^*(s')[1 - G_m^*(s + \theta)]}{h_m^*(s') - h_{m-1}^*(s')G_m^*(s + \theta)} \right\},\end{aligned}$$

which leads to the marginal distributions

$$\begin{aligned}\hat{W}_k^*(s, \text{Sr}) &= p_k\{\text{Sr}\} + \frac{p_k\{\text{Pr}\}p_{m-1}\{\text{Sr}\}W_m^*(s, \text{Rs})}{1 - p_{m-1}\{\text{Pr}\}W_m^*(s, \text{Rs})} = 1 - \frac{h_k^*(0)[1 - G_m^*(s + \theta)]}{h_m^*(0) - h_{m-1}^*(0)G_m^*(s + \theta)}, \\ \hat{H}_k^*(s, \text{Sr}) &= H_k^*(s, \text{Sr}) + \frac{p_m\{\text{Rs}\}H_k^*(s, \text{Pr})H_{m-1}^*(s, \text{Sr})}{1 - p_m\{\text{Rs}\}H_{m-1}^*(s, \text{Pr})} = \frac{\mu}{s + \mu} \left\{ 1 - \frac{h_k^*(s)[1 - G_m^*(\theta)]}{h_m^*(s) - h_{m-1}^*(s)G_m^*(\theta)} \right\}, \\ \hat{T}_k^*(s, \text{Sr}) &= H_k^*(s, \text{Sr}) + \frac{H_k^*(s, \text{Pr})W_m^*(s, \text{Rs})H_{m-1}^*(s, \text{Sr})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s, \text{Pr})} = \frac{\mu}{s + \mu} \left\{ 1 - \frac{h_k^*(s)[1 - G_m^*(s + \theta)]}{h_m^*(s) - h_{m-1}^*(s)G_m^*(s + \theta)} \right\}.\end{aligned}$$

Then we get the probability of service completion

$$P_k\{\text{Sr}\} = \hat{W}_k^*(0, \text{Sr}) = \hat{H}_k^*(0, \text{Sr}) = \hat{T}_k^*(0, \text{Sr}) = 1 - \frac{h_k^*(0)[1 - G_m^*(\theta)]}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)} = 1 - P_k\{\text{Ab}\},$$

the mean waiting and service times

$$E[\hat{W}_k, \text{Sr}] = \frac{h_k^*(0)[h_{m-1}^*(0) - h_m^*(0)]G'_m(\theta)}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2},$$

$$E[\hat{H}_k, \text{Sr}] = \frac{1}{\mu}P_k\{\text{Sr}\} - [1 - G_m^*(\theta)] \left\{ \frac{h_k^*(0)[h'_m(0) - h'_{m-1}(0)G_m^*(\theta)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2} - \frac{h'_k(0)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)} \right\},$$

and the covariance $\text{Cov}[\hat{W}_k, \hat{H}_k; \text{Sr}] := E[\hat{W}_k \hat{H}_k, \text{Sr}] - E[\hat{W}_k, \text{Sr}]E[\hat{H}_k, \text{Sr}]$ with

$$E[\hat{W}_k \hat{H}_k, \text{Sr}] = \frac{1}{\mu}E[\hat{W}_k, \text{Sr}] - G'_m(\theta) \left\{ \frac{2h_k^*(0)[h_m^*(0) - h_{m-1}^*(0)][h'_m(0) - h'_{m-1}(0)G_m^*(\theta)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^3} \right. \\ \left. - \frac{h_k^*(0)[h'_m(0) - h'_{m-1}(0)] + h'_k(0)[h_m^*(0) - h_{m-1}^*(0)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2} \right\}.$$

For the customer in state k , $k \geq m$, we have

$$\hat{T}_k^*(s, s', \text{Sr}) = \frac{W_k^*(s, \text{Rs})H_{m-1}^*(s', \text{Sr})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s', \text{Pr})} = \frac{\mu}{s' + \mu} \cdot \frac{[h_m^*(s') - h_{m-1}^*(s')]G_k^*(s + \theta)}{h_m^*(s') - h_{m-1}^*(s')G_m^*(s + \theta)}.$$

which leads to the marginal distributions

$$\hat{W}_k^*(s, \text{Sr}) = \frac{p_{m-1}\{\text{Sr}\}W_k^*(s, \text{Rs})}{1 - p_{m-1}\{\text{Pr}\}W_m^*(s, \text{Rs})} = \frac{[h_m^*(0) - h_{m-1}^*(0)]G_k^*(s + \theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(s + \theta)},$$

$$\hat{H}_k^*(s, \text{Sr}) = \frac{p_k\{\text{Rs}\}H_{m-1}^*(s, \text{Pr})}{1 - p_m\{\text{Rs}\}H_{m-1}^*(s, \text{Pr})} = \frac{\mu}{s + \mu} \cdot \frac{[h_m^*(s) - h_{m-1}^*(s)]G_k^*(\theta)}{h_m^*(s) - h_{m-1}^*(s)G_m^*(\theta)},$$

$$\hat{T}_k^*(s, \text{Sr}) = \frac{W_k^*(s, \text{Rs})H_{m-1}^*(s, \text{Pr})}{1 - W_m^*(s, \text{Rs})H_{m-1}^*(s, \text{Pr})} = \frac{\mu}{s + \mu} \cdot \frac{[h_m^*(s) - h_{m-1}^*(s)]G_k^*(s + \theta)}{h_m^*(s) - h_{m-1}^*(s)G_m^*(s + \theta)}.$$

Then we get the probability of service completion

$$P_k\{\text{Sr}\} = \hat{W}_k^*(0, \text{Sr}) = \hat{H}_k^*(0, \text{Sr}) = \hat{T}_k^*(0, \text{Sr}) = \frac{[h_m^*(0) - h_{m-1}^*(0)]G_k^*(\theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)} = 1 - P_k\{\text{Ab}\},$$

the mean waiting and service times

$$E[\hat{W}_k, \text{Sr}] = [h_{m-1}^*(0) - h_m^*(0)] \left\{ \frac{G'_k(\theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)} + \frac{h_{m-1}^*(0)G_k^*(\theta)G'_m(\theta)}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2} \right\},$$

$$E[\hat{H}_k, \text{Sr}] = \frac{1}{\mu}P_k\{\text{Sr}\} - \frac{G_k^*(\theta)[h'_m(0)h_{m-1}^*(0) - h'_{m-1}(0)h_m^*(0)][1 - G_m^*(\theta)]}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^2},$$

and

$$E[\hat{W}_k \hat{H}_k, \text{Sr}] = \frac{1}{\mu}E[\hat{W}_k, \text{Sr}] - \frac{h'_m(0)h_{m-1}^*(0) - h'_{m-1}(0)h_m^*(0)}{[h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]^3} \\ \times \{G_k^*(\theta)G'_m(\theta)[h_m^*(0) - 2h_{m-1}^*(0) + G_m^*(\theta)h_{m-1}^*(0)] - G'_k(\theta)[1 - G_m^*(\theta)][h_m^*(0) - h_{m-1}^*(0)G_m^*(\theta)]\}.$$

4.3. Time to departure (either abandonment or service completion)

We finally consider the time to departure (either abandonment or service completion) for an arbitrary customer. For the customer in state k , $0 \leq k \leq m-1$, we have

$$\begin{aligned}\hat{T}_k^*(s, s') &= \hat{T}_k^*(s, s', \text{Ab}) + \hat{T}_k^*(s, s', \text{Sr}) \\ &= \frac{\mu}{s' + \mu} + \left(\frac{\theta}{s + \theta} - \frac{\mu}{s' + \mu} \right) \frac{h_k^*(s')[1 - G_m^*(s + \theta)]}{h_m^*(s') - h_{m-1}^*(s')G_m^*(s + \theta)},\end{aligned}$$

which leads to the marginal distributions

$$\begin{aligned}\hat{W}_k^*(s) &= \hat{W}_k^*(s, \text{Ab}) + \hat{W}_k^*(s, \text{Sr}) = 1 - \frac{s}{s + \theta} \cdot \frac{h_k^*(0)[1 - G_m^*(s + \theta)]}{h_m^*(0) - h_{m-1}^*(0)G_m^*(s + \theta)}, \\ \hat{H}_k^*(s) &= \hat{H}_k^*(s, \text{Ab}) + \hat{H}_k^*(s, \text{Sr}) = \frac{\mu}{s + \mu} + \frac{s}{s + \mu} \cdot \frac{h_k^*(s)[1 - G_m^*(\theta)]}{h_m^*(s) - h_{m-1}^*(s)G_m^*(\theta)}, \\ \hat{T}_k^*(s) &= \hat{T}_k^*(s, \text{Ab}) + \hat{T}_k^*(s, \text{Sr}) = \frac{\mu}{s + \mu} + \left(\frac{\theta}{s + \theta} - \frac{\mu}{s + \mu} \right) \frac{h_k^*(s)[1 - G_m^*(s + \theta)]}{h_m^*(s) - h_{m-1}^*(s)G_m^*(s + \theta)}.\end{aligned}$$

The mean waiting time, service time, and total time are given by

$$\begin{aligned}E[\hat{W}_k] &= E[\hat{W}_k, \text{Ab}] + E[\hat{W}_k, \text{Sr}] = \frac{1}{\theta}P_k\{\text{Ab}\}, \\ E[\hat{H}_k] &= E[\hat{H}_k, \text{Ab}] + E[\hat{H}_k, \text{Sr}] = \frac{1}{\mu}P_k\{\text{Sr}\}, \\ E[\hat{T}_k] &= E[\hat{T}_k, \text{Ab}] + E[\hat{T}_k, \text{Sr}] = E[\hat{W}_k] + E[\hat{H}_k] = \frac{1}{\theta}P_k\{\text{Ab}\} + \frac{1}{\mu}P_k\{\text{Sr}\}.\end{aligned}$$

In particular, for an arriving customer ($k = 0$), we have the relation

$$E[\hat{W}_0] = \frac{1}{\theta}P_0\{\text{Ab}\} = \frac{E[L]}{\lambda} \quad ; \quad E[\hat{H}_0] = \frac{1}{\mu}P_0\{\text{Sr}\} = \frac{E[S]}{\lambda}$$

and Little's theorem

$$E[\hat{T}_0] = \frac{1}{\theta}P_0\{\text{Ab}\} + \frac{1}{\mu}P_0\{\text{Sr}\} = \frac{E[N]}{\lambda},$$

where $E[L]$, $E[S]$, and $E[N]$ are the mean number of customers present in the waiting room, in the service facility, and in the whole system, respectively, which are given in [3]. We also have

$$E[\hat{W}_k \hat{H}_k] = E[\hat{W}_k \hat{H}_k, \text{Ab}] + E[\hat{W}_k \hat{H}_k, \text{Sr}] = \frac{1}{\theta}E[\hat{H}_k, \text{Ab}] + \frac{1}{\mu}E[\hat{W}_k, \text{Sr}]$$

to be used in the covariance $\text{Cov}[\hat{W}_k, \hat{H}_k] := E[\hat{W}_k \hat{H}_k] - E[\hat{W}_k]E[\hat{H}_k]$.

For the customer in state k , $k \geq m$, we have

$$\begin{aligned}\hat{T}_k^*(s, s') &= \hat{T}_k^*(s, s', \text{Sr}) + \hat{T}_k^*(s, s', \text{Ab}) \\ &= \frac{\theta}{s + \theta} + \left(\frac{\mu}{s' + \mu} - \frac{\theta}{s + \theta} \right) \frac{[h_m^*(s') - h_{m-1}^*(s')]G_k^*(s + \theta)}{h_m^*(s') - h_{m-1}^*(s')G_m^*(s + \theta)},\end{aligned}$$

which leads to the marginal distributions

$$\begin{aligned}\hat{W}_k^*(s) &= \frac{\theta}{s + \theta} + \frac{s}{s + \theta} \cdot \frac{[h_m^*(0) - h_{m-1}^*(0)]G_k^*(s + \theta)}{h_m^*(0) - h_{m-1}^*(0)G_m^*(s + \theta)}, \\ \hat{H}_k^*(s) &= 1 - \frac{s}{s + \mu} \cdot \frac{[h_m^*(s) - h_{m-1}^*(s)]G_k^*(\theta)}{h_m^*(s) - h_{m-1}^*(s)G_m^*(\theta)}, \\ \hat{T}_k^*(s) &= \frac{\theta}{s + \theta} + \left(\frac{\mu}{s + \mu} - \frac{\theta}{s + \theta} \right) \frac{[h_m^*(s) - h_{m-1}^*(s)]G_k^*(s + \theta)}{h_m^*(s) - h_{m-1}^*(s)G_m^*(s + \theta)}.\end{aligned}$$

We get the same expressions for $E[\hat{W}_k]$, $E[\hat{H}_k]$, $E[\hat{T}_k]$, and $E[\hat{W}_k \hat{H}_k]$ as shown above, respectively.

Table 1: Numerical example.

k	$P_k\{\text{Ab}\}$	$P_k\{\text{Sr}\}$	$E[\hat{W}_k, \text{Ab}]$	$E[\hat{H}_k, \text{Ab}]$	$E[\hat{T}_k, \text{Ab}]$	$E[\hat{W}_k, \text{Sr}]$	$E[\hat{H}_k, \text{Sr}]$	$E[\hat{T}_k, \text{Sr}]$
0	0.51270	0.48730	0.24299	0.30992	0.55291	0.01336	0.17738	0.19074
1	0.56396	0.43604	0.26729	0.28965	0.55693	0.01470	0.14639	0.16108
2	0.62549	0.37451	0.29645	0.26019	0.55663	0.01630	0.11432	0.13062
3	0.70034	0.29966	0.33192	0.21777	0.54969	0.01825	0.08189	0.10014
4	0.79283	0.20717	0.37576	0.15678	0.53254	0.02066	0.05039	0.07105
5	0.90911	0.09089	0.43087	0.06878	0.49965	0.02369	0.02211	0.04579
6	0.94907	0.05093	0.45368	0.03854	0.49222	0.02085	0.01239	0.03324
7	0.96686	0.03314	0.46548	0.02508	0.49056	0.01795	0.00806	0.02601
8	0.97624	0.02376	0.47252	0.01798	0.49050	0.01560	0.00578	0.02138
9	0.98181	0.01819	0.47713	0.01377	0.49090	0.01378	0.00442	0.01820
10	0.98541	0.01459	0.48038	0.01104	0.49142	0.01233	0.00355	0.01588
15	0.99293	0.00707	0.48827	0.00535	0.49362	0.00819	0.00172	0.00991
20	0.99541	0.00459	0.49146	0.00347	0.49493	0.00624	0.00112	0.00736
30	0.99732	0.00268	0.49434	0.00203	0.49637	0.00432	0.00065	0.00497

k	$E[\hat{W}_k]$	$E[\hat{H}_k]$	$E[\hat{T}_k]$	$\text{Cov}[\hat{W}_k, \hat{H}_k; \text{Ab}]$	$\text{Cov}[\hat{W}_k, \hat{H}_k; \text{Sr}]$	$\text{Cov}[\hat{W}_k, \hat{H}_k]$
0	0.25635	0.48730	0.74365	0.08169	0.00896	0.16832
1	0.28198	0.43604	0.71802	0.07098	0.00897	0.15952
2	0.31274	0.37451	0.68726	0.05852	0.00888	0.14639
3	0.35017	0.29966	0.64983	0.04474	0.00862	0.12713
4	0.39642	0.20717	0.60358	0.03103	0.00807	0.09905
5	0.45456	0.09089	0.54544	0.02089	0.00703	0.05808
6	0.47454	0.05093	0.52546	0.01656	0.00582	0.04012
7	0.48343	0.03314	0.51567	0.01379	0.00487	0.03049
8	0.48812	0.02376	0.51188	0.01183	0.00417	0.02459
9	0.49090	0.01819	0.50910	0.01038	0.00365	0.02066
10	0.49270	0.01459	0.50730	0.00926	0.00324	0.01785
15	0.49647	0.00707	0.50353	0.00612	0.00212	0.01087
20	0.49771	0.00459	0.50229	0.00466	0.00160	0.00798
30	0.49866	0.00268	0.50134	0.00323	0.00110	0.00533

5. Numerical Example

Numerical values for $P_k\{\text{Ab}\}$, $P_k\{\text{Sr}\}$, $E[\hat{W}_k, \text{Ab}]$, $E[\hat{H}_k, \text{Ab}]$, $E[\hat{T}_k, \text{Ab}]$, $E[\hat{W}_k, \text{Sr}]$, $E[\hat{H}_k, \text{Sr}]$, $E[\hat{T}_k, \text{Sr}]$, $E[\hat{W}_k]$, $E[\hat{H}_k]$, $E[\hat{T}_k]$, $\text{Cov}[\hat{W}_k, \hat{H}_k; \text{Ab}]$, $\text{Cov}[\hat{W}_k, \hat{H}_k; \text{Sr}]$, and $\text{Cov}[\hat{W}_k, \hat{H}_k]$ are shown for $k = 0, 1, 2, \dots, 10, 15, 20$, and 30 in Table 1. In this case, we have assumed $m = 5$, $\mu = 1$, $\theta = 2$, and $\lambda = 10$ so that $\rho = 2$ and $\tau = 2$.

References

- [1] F. Iravani and B. Balcioglu, On priority queues with impatient customers, *Queueing Systems*, Vol.58, No.4, pp.239–260, May 2008.
- [2] S. Subba Rao, Queuing with balking and reneging in M/G/1 systems, *Metrika*, Vol.12, No.1, pp.173–188, December 1967.
- [3] H. Takagi, Times to service completion and abandonment in the M/M/m preemptive LCFS queue with impatient customers, *QTNA' 2016*, December 13–15, 2016, Wellington, New Zealand, DOI: <http://dx.doi.org/10.1145/3016032.3016036>
- [4] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice Hall, 1989.

Acknowledgment

This work is supported by the Grant-in-Aid for Scientific Research (C) No. 26330354 from the Japan Society for the Promotion of Science (JSPS) in 2016.

ビットコインのマイニング・モデルとトランザクション承認時間解析

河瀬 良亮 笠原 正治
奈良先端科学技術大学院大学 情報科学研究科

概要：ビットコインでは、ハッシュ関数の出力結果がある条件を満足するような文字列を探索するマイニングにより、取引台帳であるブロック・チェーンを更新する。最近の研究により、新たに発生したトランザクションは現在のマイニング対象ブロックには含まれず、次回以降のブロックに含まれるということが明らかになった。本稿では、この状況を集団サービス型待ち行列を用いてモデル化し、トランザクション単位の承認時間の解析を行う。具体的には、集団サービスを行う $M/G^B/1$ を考え、新規トランザクションの到着時点でサービス中の集団サイズが制限未満であってもそのトランザクションは待ち行列で待機するモデルを考える。このモデルにおいて、トランザクションの系内滞在時間はそのトランザクションの承認にかかる時間に相当する。このモデルに対してサービス中のトランザクション数と待ち行列内トランザクション数の結合分布を考え、それを基にトランザクション承認時間の平均を導出する。数値例において、集団サービスのサイズがトランザクション承認時間に与える影響を定量的に分析する。

1. はじめに

ビットコインは中央サーバや管理者が存在しない自律分散型の仮想通貨であり、暗号技術とピア・ツー・ピア (Peer-to-Peer) ネットワーク技術により、多重支払いやなりすまし等の不正を防止することに成功している [1]。ビットコインは高速かつ安全に国際取引が可能であり、その手数料の安さからインターネット上の投げ銭やコンテンツ一部分への課金といった少額取引 (マイクロペイメント) において利用拡大が期待されている [2]。ビットコインは年々使用される機会が増えており、2015 年の一日当たりの平均取引件数は 125,000 件に上り、その前年の 69,000 件と比較して約 2 倍に増加している。日本国内においては、財務省と金融庁が、ビットコインやイーサリアムなどの仮想通貨取得時にかかる消費税をなくす方向で来年度の税制改正に向けて議論を進めており [3]、仮想通貨の普及に拍車がかかることが考えられる。

ビットコインの取引は、トランザクションを通じて行われる。トランザクションの中には、送信者から受信者へある額面のコインを移動させるという情報が記載されている。ビットコインでは、トランザクションを複数集めてブロックを形成し、パズルの問題を解く作業によりブロックを承認する。この作業をマイニングと呼び、マイニングは平均 10 分で終了するように、パズルの難易度は自動で調整されている [4]。マイニングを行うノードをマイナーと呼び、マイナーはマイニングに成功すると報酬として 12.5BTC とブロックに含まれるトランザクションの送金手数料を受け取る。

ビットコインの技術的課題の一つに、最大ブロックサイズに起因する低いトランザクション処理速度がある [5]。ビットコインではブロックに複数のトランザクションを収容して承認処理を行うが、ブロックサイズが 1Mbyte に制限されており、平均 10 分のマイニング時間であることから、単位時間当たりのトランザクション処理数に限界がある。また、トランザクションは送金金額、前回の承認からの経過時間、データサイズによって優先権の値が決められ、その値と手数料を基にブロックに含まれるかどうか決定される [4]。そのため、高負荷時では低額送金のトランザクションや低い手数料のトランザクションは低い優先権のためにブロックに中々組み込まれず、承認に時間がかかってしまう。ブロック・チェーンの取引情報を提供している [6] によると、10 分間当たりのトランザクション発生量はブロックサイズの上限近くに迫ってきており、スケーラビリティを如何に拡張するかが喫緊の課題となっている。この問題に対し、トランザクションの電子署名部分を分離するアプローチ [7] など数多くの解決法が提案されているが、2016 年 11 月時点においてビットコイン・コミュニティで合意された方法は存在しない。

ビットコインのトランザクション承認処理は、集団サービスによる待ち行列としてモデル化することができる。集団サービスを持つ待ち行列の解析については、文献 [9] において $M/G^B/1$ に対し、経過サービス時間と待ち行列内容数の結合分布関数について解析が行われている。文献 [8] では、優先権付き $M/G^B/1$ 待ち行列を考え、高優先と低優先の二種類のトランザクションに対する平均承認時間の解析が行われている。文献 [8] の待ち行列モデルでは、トランザクションが新規に到着した時点で集団サービスが制限未満であれば、そのトランザクションは直ちにサービスに含まれるというモデルを検

討している. 文献 [8] ではブロック・チェーンの統計的分析を行い, マイニングにかかる時間 (ブロック生成時間) が指数分布に従っていること, トランザクションが到着してから承認されるまでの時間がサービス時間の約二倍近くかかっていること, の二点を報告し, これらの結果から, 新規到着トランザクションは直ちにサービスされるのではなく, 早くても次のブロック生成に回されることを指摘している.

そこで本研究では, $M/G^b/1$ 待ち行列において, 集団サービスが行われているときに到着したトランザクションはそのサービスに含められず, 待ち行列で待機させられるモデルを考える. 具体的には, サービス中のトランザクション数と待ち行列内のトランザクション数の結合確率母関数を考え, 平均系内トランザクション数からリトルの公式より平均トランザクション承認時間を導出する. 数値例では, ブロックサイズがトランザクション承認時間に与える影響について, 定量的な評価を行う.

本論文の構成は以下の通りである. 2 章ではトランザクション処理の待ち行列モデルを述べ, 3 章でサービス中のトランザクション数と待ち行列内トランザクション数の結合分布を導出する. 4 章で数値例を示し, 5 章で結論を述べる.

2. マイニングの待ち行列モデル

トランザクションは率 λ のポアソン過程に従ってシステムに到着する. トランザクションは複数個まとめられて一つのブロックに収容される.

マイニングによってブロックが生成されるため, 以下ではブロック生成間隔をサービス時間と呼ぶ. S_i ($i = 1, 2, \dots$) を i 番目のブロック生成時間間隔とする. $\{S_i\}$ は独立同一な分布関数 $G(x) = \Pr\{S \leq x\}$ に従い, $G(x)$ の確率密度関数を $g(x)$ とする. 平均ブロック生成時間 $E[S]$ は

$$E[S] = \int_0^\infty x dG(x) = \int_0^\infty xg(x) dx,$$

と表せる.

サーバは複数のトランザクションを集団サービスする. 集団サービスの最大サイズは b である. トランザクションが到着した時点でシステムが稼働中のとき, そのトランザクションは待ち行列で待機し, 次以降のブロックに到着順で含められる. システム内にトランザクションが存在しない場合でも, マイニングによるサービスは継続的に行われる. そのため, 新たに到着したトランザクションは次のサービス開始時点まで待機させられるものと仮定する.

3. 解析

$N_s(t)$ を時刻 t におけるサーバ内トランザクション数, $N_q(t)$ を待ち行列内トランザクション数とする. また $X(t)$ を時刻 t における経過サービス時間とし, $P_{m,n}$ を次式で定義する.

$$P_{m,n}(x, t)dx = \Pr\{N_s(t) = m, N_q(t) = n, x < X(t) \leq x + dx\}.$$

またサービス時間 S のハザードレート $\xi(x)$ を次式で定義する.

$$\xi(x) = \frac{g(x)}{1 - G(x)}.$$

$t \rightarrow \infty$ として定常状態を考える. 経過サービス時間 $x(t)$ の微小変化量に注意することで次式を得る.

$$\frac{d}{dx} P_{m,n}(x) = -\{\lambda + \xi(x)\}P_{m,n}(x) + \lambda P_{m,n-1}(x), \quad 0 \leq m \leq b, \quad n \geq 1, \quad (1)$$

$$\frac{d}{dx} P_{m,0}(x) = -\{\lambda + \xi(x)\}P_{m,0}(x), \quad 0 \leq m \leq b. \quad (2)$$

$x = 0$ における境界条件は次式で与えられる.

$$\begin{aligned} P_{b,n}(0) &= \sum_{m=0}^b \int_0^\infty P_{m,n+b}(x) \xi(x) dx, \quad n \geq 0, \\ P_{m,n}(0) &= 0, \quad m = 0, 1, \dots, b-1, \quad n \geq 1, \\ P_{k,0}(0) &= \sum_{m=0}^b \int_0^\infty P_{m,k}(x) \xi(x) dx, \quad k = 0, 1, \dots, b. \end{aligned}$$

正規化条件は

$$\sum_{n=0}^\infty \sum_{m=0}^b \int_0^\infty P_{m,n}(x) dx = 1,$$

と表せる. ここで, 結合母関数 $P(z_1, z_2; x)$ と $P(z_1, z_2)$ を以下のように定義する.

$$P(z_1, z_2; x) = \sum_{n=0}^\infty \sum_{m=0}^b P_{m,n}(x) z_1^m z_2^n, \quad (3)$$

$$P(z_1, z_2) = \int_0^\infty P(z_1, z_2; x) dx. \quad (4)$$

(1) 式と (2) 式より,

$$\begin{aligned} \sum_{n=0}^\infty \sum_{m=0}^b \frac{d}{dx} P_{m,n}(x) z_1^m z_2^n &= \sum_{n=0}^\infty \sum_{m=0}^b -\{\lambda + \xi(x)\} P_{m,n}(x) z_1^m z_2^n + \sum_{n=1}^\infty \sum_{m=0}^b \lambda P_{m,n-1}(x) z_1^m z_2^n \\ &= \sum_{n=0}^\infty \sum_{m=0}^b -\{\lambda + \xi(x)\} P_{m,n}(x) z_1^m z_2^n + \lambda z_2 \sum_{n=0}^\infty \sum_{m=0}^b P_{m,n}(x) z_1^m z_2^n, \end{aligned}$$

となり, 母関数を用いて書き換えると

$$\begin{aligned} \frac{d}{dx} P(z_1, z_2; x) &= -\{\lambda + \xi(x)\} P(z_1, z_2; x) + \lambda z_2 P(z_1, z_2; x) \\ &= -\{\lambda(1 - z_2) + \xi(x)\} P(z_1, z_2; x). \end{aligned}$$

これより $P(z_1, z_2; x)$ は次式で与えられる.

$$P(z_1, z_2; x) = P(z_1, z_2; 0) e^{-\lambda(1-z_2)x} \{1 - G(x)\}, \quad (5)$$

この両辺に $\xi(x)$ をかけて x で積分すると

$$\begin{aligned} \int_0^\infty P(z_1, z_2; x) \xi(x) dx &= \int_0^\infty P(z_1, z_2; 0) e^{-\lambda(1-z_2)x} \{1 - G(x)\} \frac{dG(x)}{1 - G(x)} \\ &= P(z_1, z_2; 0) \int_0^\infty e^{-\lambda(1-z_2)x} dG(x) \\ &= P(z_1, z_2; 0) G^*(\lambda - \lambda z_2). \end{aligned} \quad (6)$$

ここで

$$G^*(\lambda - \lambda z_2) = \int_0^\infty e^{-\lambda(1-z_2)x} dG(x),$$

である. (3), (6) 式より,

$$P(z_1, z_2; 0)G^*(\lambda - \lambda z_2) = \int_0^\infty \sum_{n=0}^\infty \sum_{m=0}^b P_{m,n}(x)\xi(x) dx z_1^m z_2^n. \quad (7)$$

また (3) 式で $x = 0$ とすると

$$\begin{aligned} P(z_1, z_2; 0) &= \sum_{n=0}^\infty \sum_{m=0}^b P_{m,n}(0) z_1^m z_2^n \\ &= \sum_{n=0}^\infty P_{b,n}(0) z_1^b z_2^n + \sum_{m=0}^{b-1} P_{m,0}(0) z_1^m \\ &= \sum_{n=0}^\infty \sum_{m=0}^b \int_0^\infty P_{m,n+b}(x)\xi(x) dx z_1^b z_2^n + \sum_{n=0}^{b-1} \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx z_1^n. \end{aligned} \quad (8)$$

ここで (8) 式の第一項目に (7) 式を用いることで

$$\begin{aligned} \sum_{n=0}^\infty \sum_{m=0}^b \int_0^\infty P_{m,n+b}(x)\xi(x) dx z_1^b z_2^n &= \left(\frac{z_1}{z_2}\right)^b \sum_{n=b}^\infty \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx z_2^n \\ &= \left(\frac{z_1}{z_2}\right)^b \left\{ P(1, z_2; 0)G^*(\lambda - \lambda z_2) - \sum_{n=0}^{b-1} \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx z_2^n \right\}, \end{aligned}$$

となる. よって, (8) 式は次のように書ける.

$$\begin{aligned} P(z_1, z_2; 0) &= \left(\frac{z_1}{z_2}\right)^b \left\{ P(1, z_2; 0)G^*(\lambda - \lambda z_2) - \sum_{n=0}^{b-1} \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx z_2^n \right\} \\ &\quad + \sum_{n=0}^{b-1} \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx z_1^n. \end{aligned} \quad (9)$$

(9) 式に $z_1 = 1$ を代入すると

$$\begin{aligned} P(1, z_2; 0) &= \left(\frac{1}{z_2}\right)^b \left\{ P(1, z_2; 0)G^*(\lambda - \lambda z_2) - \sum_{n=0}^{b-1} \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx z_2^n \right\} \\ &\quad + \sum_{n=0}^{b-1} \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx. \end{aligned}$$

上式の両辺に z_2^b をかけて整理すると

$$\{z_2^b - G^*(\lambda - \lambda z_2)\}P(1, z_2; 0) = \sum_{n=0}^{b-1} (z_2^b - z_2^n) \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx.$$

よって

$$P(1, z_2; 0) = \frac{\sum_{n=0}^{b-1} (z_2^b - z_2^n) \alpha_n}{z_2^b - G^*(\lambda - \lambda z_2)}. \quad (10)$$

ここで

$$\alpha_n = \sum_{m=0}^b \int_0^\infty P_{m,n}(x)\xi(x) dx,$$

である.

(10) 式の分母に対して Rouché の定理より [10],

$$z_2^b - G^*(\lambda - \lambda z_2) = 0,$$

の z_2 についての解は $|z_2| = 1 + \epsilon (\epsilon > 0)$ の範囲に b 個含まれている. その解の 1 つは $z_2 = 1$ であり, その他の解を $z_{2,k}^* (k = 1, 2, \dots, b-1)$ とすると式 (10) より次の $b-1$ 個の方程式を得る.

$$\sum_{n=0}^{b-1} \left\{ (z_{2,k}^*)^b - (z_{2,k}^*)^n \right\} \alpha_n = 0, \quad k = 1, 2, \dots, b-1. \quad (11)$$

(9) 式と (10) 式より

$$P(z_1, z_2; 0) = \left(\frac{z_1}{z_2} \right)^b \left\{ \frac{\sum_{n=0}^{b-1} (z_2^b - z_2^n) \alpha_n}{z_2^b - G^*(\lambda - \lambda z_2)} G^*(\lambda - \lambda z_2) - \sum_{n=0}^{b-1} \alpha_n z_2^n \right\} + \sum_{n=0}^{b-1} \alpha_n z_1^n. \quad (12)$$

また (4) 式と (5) 式より

$$\begin{aligned} P(z_1, z_2) &= P(z_1, z_2; 0) \int_0^\infty e^{-\lambda(1-z_2)x} \{1 - G(x)\} dx \\ &= P(z_1, z_2; 0) \frac{1 - G^*(\lambda - \lambda z_2)}{\lambda(1 - z_2)}. \end{aligned} \quad (13)$$

ここで両辺に $\lambda(1 - z_2)$ をかけて z_2 で偏微分すると

$$\frac{\partial P(z_1, z_2)}{\partial z_2} \lambda(1 - z_2) - P(z_1, z_2) \lambda = \frac{\partial P(z_1, z_2; 0)}{\partial z_2} \{1 - G^*(\lambda - \lambda z_2)\} - P(z_1, z_2; 0) \frac{\partial G^*(\lambda - \lambda z_2)}{\partial z_2}.$$

$z_1 = z_2 = 1$ を代入し $P(1, 1) = 1$ を用いると

$$P(1, 1) = P(1, 1; 0) E[S] = 1.$$

$P(1, 1; 0)$ を求めるために, (13) 式の両辺に $z_2^b(z_2^b - G^*(\lambda - \lambda z_2))$ をかけると

$$\begin{aligned} P(z_1, z_2; 0) z_2^b \{z_2^b - G^*(\lambda - \lambda z_2)\} &= z_1^b \sum_{n=0}^{b-1} (z_2^b - z_2^n) \alpha_n G^*(\lambda - \lambda z_2) - z_1^b \{z_2^b - G^*(\lambda - \lambda z_2)\} \sum_{n=0}^{b-1} \alpha_n z_2^n \\ &\quad + z_2^b \{z_2^b - G^*(\lambda - \lambda z_2)\} \left(\sum_{n=1}^{b-1} \alpha_n z_1^n + \alpha_0 z_1 \right). \end{aligned}$$

z_2 で偏微分して $z_1 = z_2 = 1$ とし, システムの安定条件 $b > \lambda E[S]$ の成立を仮定すると, 次式を得る.

$$P(1, 1; 0) = \frac{\sum_{n=0}^{b-1} (b-n) \alpha_n}{b - \lambda E[S]}.$$

よって正規化条件は次式で与えられる.

$$\frac{\sum_{n=0}^{b-1} (b-n) \alpha_n}{b - \lambda E[S]} E[S] = 1. \quad (14)$$

(11) 式と (14) 式を α_n について解くことにより, b 個の未知変数 α_n の値が決定される.

(12) 式と (13) 式より

$$P(z_1, z_2) = \left\{ \left(\frac{z_1}{z_2} \right)^b \frac{\sum_{n=0}^{b-1} (z_2^b - z_2^n) \alpha_n}{z_2^b - G^*(\lambda - \lambda z_2)} G^*(\lambda - \lambda z_2) - \left(\frac{z_1}{z_2} \right)^b \sum_{n=0}^{b-1} \alpha_n z_2^n \right. \\ \left. + \sum_{n=0}^{b-1} \alpha_n z_1^n \right\} \frac{1 - G^*(\lambda - \lambda z_2)}{\lambda(1 - z_2)}.$$

上式を z_1 で 1 回偏微分し z_2 で 2 回偏微分してから (14) 式に注意して $z_1 = z_2 = 1$ を代入すると

$$\left(\frac{\partial P(z_1, z_2)}{\partial z_1} \right)_{z_1=1, z_2=1} = \left\{ \frac{\sum_{n=0}^{b-1} (b-n) \alpha_n}{b - \lambda E[S]} E[S] \right\} \lambda E[S] \\ = \lambda E[S].$$

また z_2 で 3 回微分して $z_1 = z_2 = 1$ を代入し (14) 式を用いて整理すると

$$\left(\frac{\partial}{\partial z_2} P(z_1, z_2) \right)_{z_1=1, z_2=1} = \frac{1}{2(b - \lambda E[S])} \left(\lambda^2 E[S^2] - 2b(b - \lambda E[S]) - b(b-1) \right) \\ + \sum_{n=0}^{b-1} \left\{ \lambda E[S^2](b-n) + E[S]\{b(b-1) - n(n-1)\} + 2bE[S](b-n) \right\} \alpha_n.$$

よって系内トランザクション数の平均は次式で与えられる.

$$E[N] = \frac{1}{2(b - \lambda E[S])} \left(\lambda^2 E[S^2] - b(b-1) - 2(b - \lambda E[S])^2 \right) \\ + \sum_{n=0}^{b-1} \left\{ \lambda E[S^2](b-n) + E[S]\{b(b-1) - n(n-1)\} + 2bE[S](b-n) \right\} \alpha_n.$$

トランザクションの系内滞在時間を T とすると, 平均系内滞在時間 $E[T]$ は, リトルの公式より次式で与えられる.

$$E[T] = \frac{E[N]}{\lambda} \\ = \frac{1}{2\lambda(b - \lambda E[S])} \left(\lambda^2 E[S^2] - b(b-1) - 2(b - \lambda E[S])^2 \right) \\ + \sum_{n=0}^{b-1} \left\{ \lambda E[S^2](b-n) + E[S]\{b(b-1) - n(n-1)\} + 2bE[S](b-n) \right\} \alpha_n. \quad (15)$$

4. 数値解析

4.1. ブロック生成時間の分布

文献 [8] より, ブロック生成時間の分布 $G(x)$ は以下のパラメータを持つ指数分布で近似できることが報告されている.

$$G(x) = 1 - e^{-\mu x}, \quad \text{ただし } \mu = 0.0018378995.$$

これより $E[S]$ と $E[S^2]$ は,

$$E[S] = \frac{1}{\mu} = 544.0993884, \quad E[S^2] = \frac{2}{\mu^2} = 592088.2889,$$

と求められる. また, $G(x)$ のラプラス・スティルチェス変換は,

$$G^*(s) = \frac{\mu}{s + \mu},$$

となる. これらと式 (15) を用いて数値実験を行った.

4.2. 解析とシミュレーションの比較

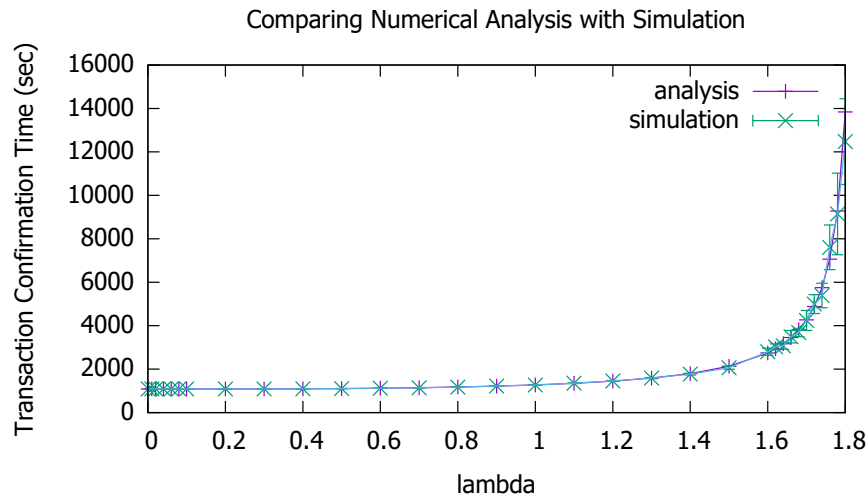


図 1: 解析とシミュレーションの比較

解析結果の妥当性を検証するため, 同一の待ち行列モデルに対するモンテカルロシミュレーションを行った. 図 1 は解析結果とシミュレーション結果を比較した図である. この図の横軸は到着率 λ を, 縦軸は平均トランザクション承認時間 $E[T]$ を表している. ここではブロックサイズを $b = 1000$ で固定して数値計算を行っている. 図 1 より, 解析結果とシミュレーション結果が一致しており, 解析の妥当性を確認することができた.

また図 1 より, λ が小さい領域ではトランザクション承認時間が 1088 秒で一定の値を示していることが観察される. 平均ブロック生成時間間隔 $1/\mu = 544$ 秒より, 到着率 λ がこの領域ではトランザクションは直後のブロックではなくそのもう 1 つ後のブロックに取り込まれている状況が推察される. さらに λ が b/μ に近づくとき急激に承認時間が長くなり発散することも観察される.

4.3. 解析と実測値の比較

解析と実測値との比較を行う. 文献 [8] によると, トランザクションの平均サイズが 571.34 バイトであることから, 最大ブロックサイズは $b = 1750$ と見積もることができる. 表 1 はこのブロックサイズの基で実データより得られた平均到着率 λ の値を基に計算した平均トランザクション承認時間と実測値を比較したものである. この表より

表 1: 解析と実測値との比較

トランザクション	到着率	実測値	解析
優先度なし	0.9709120529	1127.238651	1112.035745

実測値が 1127 秒に対して解析では 1112 秒と近い値を示している. 解析結果の方が約 15 秒ほど小さい結果となっているが, 平均値は 1000 のオーダーとなっており, 概ね妥当な精度の結果が得られていると考えられる.

4.4. 実データを用いたシミュレーションと解析の比較

文献 [6] にあるトランザクションの発行時間とブロックの生成時間のデータを用いてシミュレーションを行い, 解析結果と比較する. まず 2013 年 10 月から 2015 年 9 月までの 2 年間のデータに対して,

式 (15) で到着率 $\lambda = 0.9709120$ として数値実験を行った結果とシミュレーションの結果を図 2 に示す. 同様に, 2013 年 10 月から 2014 年 9 月の 1 年間と 2014 年 10 月から 2015 年 9 月までの 1 年間に對して, それぞれ到着率 $\lambda = 0.733693, 1.208131$ で解析を行った結果とシミュレーション結果を図 3, 4 に示す.

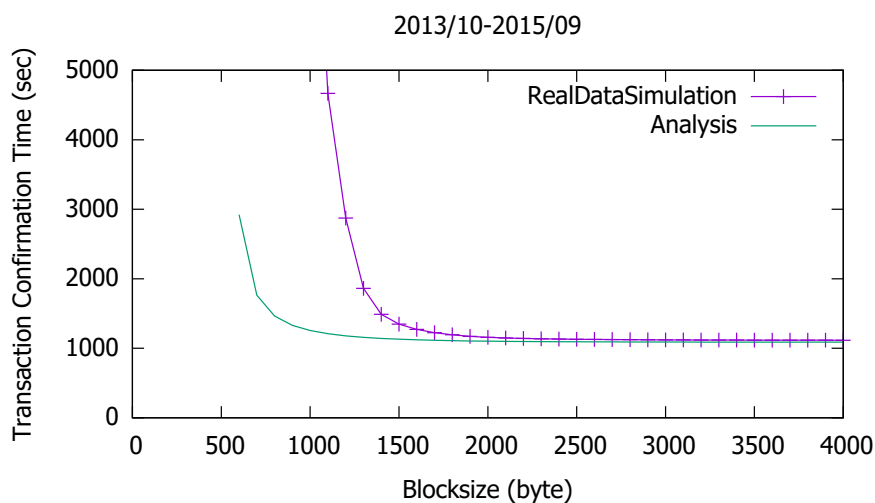


図 2: 2013/10-2015/09 における承認時間に対するブロックサイズの影響

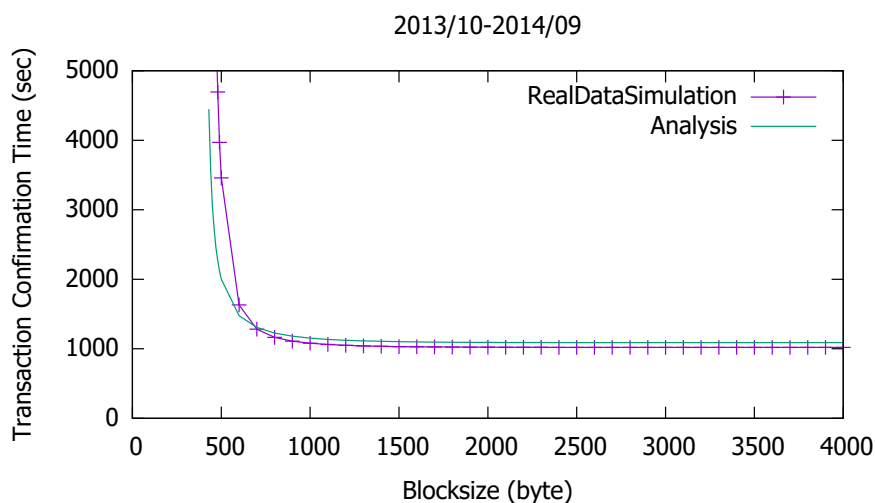


図 3: 2013/10-2014/09 における承認時間に対するブロックサイズの影響

1 日ごとのトランザクションの到着率を図 5 に示す. これより 1 日に到着するトランザクションの数は時間とともに大きく変動していることが分かる. 2013 年 10 月から 2014 年 9 月の間では到着率のばらつきが少ないため, 解析とシミュレーションでよく一致している. しかしながら 2014 年 10 月から 2015 年 9 月の間では到着率のばらつきが大きいため, 結果が解析とシミュレーションの間で差異が生じている. 同様の理由で 2013 年 10 月から 2015 年 9 月までの 2 年間に對してシミュレーションを行った際も合致しなかった.

4.5. トランザクション承認時間に対するブロックサイズの影響

本節では, ブロックサイズがトランザクション承認時間に与える影響について検討を行う.

ブロックサイズを $b = 1000, 2000, 4000, 8000$ と変えた場合の結果を図 6 に示す. 現在の最大ブロックサイズ 1Mbyte に含まれるトランザクションの最大数は文献 [8] より $b = 1750$ で近似的に与えられ

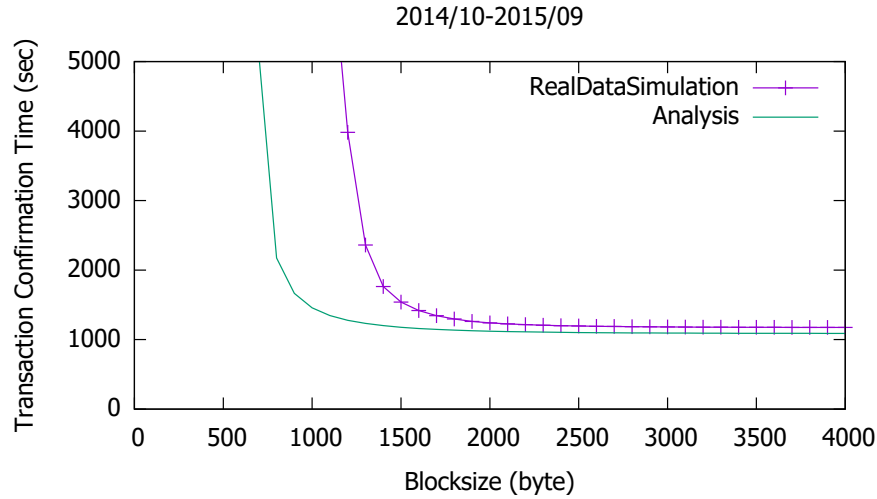


図 4: 2014/10-2015/09 における承認時間に対するブロックサイズの影響

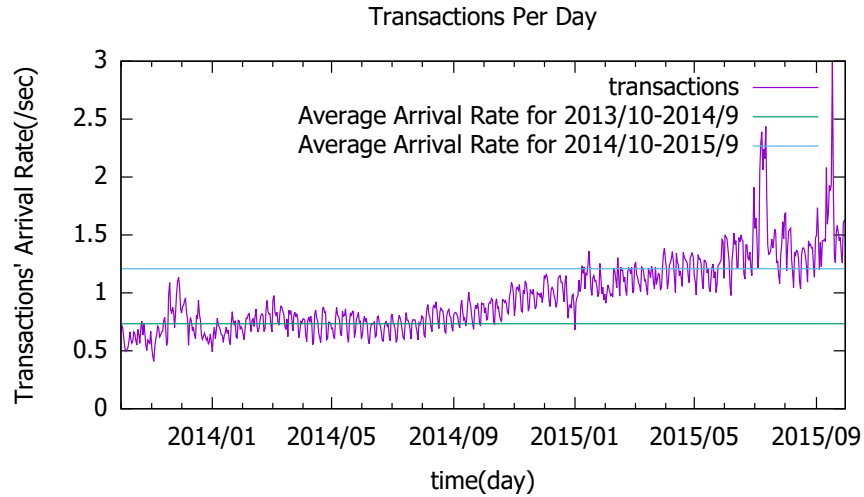


図 5: Arrival Rate Of Transactions Per Day

ている. この値に近いのが $b = 2000$ の場合で, $\lambda = 3.6$ 付近で発散している. これはブロックサイズの上限に関する問題で, ブロックサイズを増やしてもトランザクション処理速度が大幅に向上しないことを意味している.

それに対して Segregated Witness (Segwit) と呼ばれる改良を施すことが予定されている [11]. Segwit は, トランザクションから電子署名を分離し, ブロックチェーンに取り込まれるデータ量を削減し, セキュリティも強化するものである. Segwit に対応したウォレットを利用すると, 電子署名部分の容量を 75% 削減できる. Segwit とブロックサイズの増加を組み合わせることにより, トランザクション処理量を大幅に改善できることが期待できる.

5. 結論・今後の課題

本論文ではビットコインのトランザクション承認時間について, 集団サービスを行う待ち行列モデル $M/G^B/1$ を用いて解析を行った. この待ち行列モデルでは, 到着したトランザクションは一旦メモリ上のプールに蓄積され, 到着時点で生成中のブロックには含められない状況を模擬している. このモデルを用いてトランザクションの承認時間を定式化し, その正当性をシミュレーションとの比較によって確認した. さらにトランザクション承認時間に対するブロックサイズや到着率の影響について定量

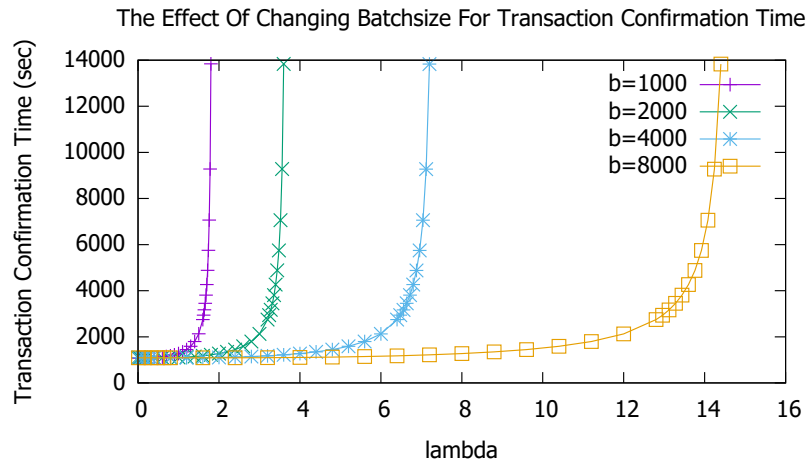


図 6: バッチサイズによるトランザクション承認時間の変化

的に評価を行った. ブロックサイズの増加により, トランザクション処理量をある程度増大させることができるが, その処理増加量は限定的であることが判明した. しかしながら Segwit によるトランザクションデータ量の削減と最大ブロックサイズを増加させることにより, トランザクション処理量を現状よりも大幅に改善できる可能性がある.

今後の課題として, 実際のビットコインのシステムではトランザクションに優先度が与えられ, 優先度の高いトランザクションほど先にブロックに格納される仕組みになっているため, 優先度を考慮したモデルの解析が必要である. とくにマイクロペイメントでの利用が見込まれているため, 優先度が低い場合のトランザクション承認時間の解析をする必要がある [2]. またマイクロペイメントによる利用が増えた場合に, ライトニングネットワークと呼ばれるマイクロペイメント専用のチャネルを実装することが検討されている [12]. その効果を数理モデルで定量的に解析することも重要である.

参考文献

- [1] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” <https://bitcoin.org/bitcoin.pdf>, 2008.
- [2] http://www.meti.go.jp/committee/kenkyukai/sansei/fintech_kadai/pdf/003_02_00.pdf
- [3] <http://www3.nhk.or.jp/news/html/20161016/k10010731691000.html>
- [4] A. M. Antonopoulos, *Mastering Bitcoin*, O’Reilly, 2014.
- [5] <http://www.coindesk.com/1mb-block-size-today-bitcoin/>
- [6] <https://blockchain.info/>
- [7] <http://www.coindesk.com/segregated-witness-bitcoin-block-size-debate/>
- [8] S. Kasahara, J. Kawahara, “Priority Mechanism of Bitcoin and Its Effect On Transaction-Confirmation Process,” Submitted for publication.
- [9] M. L. Chaudhry and J. G. C. Templeton, “The Queing System $M/G^B/1$ and its Ramifications,” *European Journal of Operational Research*, vol. 6, pp. 56–60, 1981.
- [10] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation Vol. 1, Vacation and Priority Systems, Part I*, North-Holland, 1991.
- [11] <https://github.com/bitcoin/bips/blob/master/bip-0141.mediawiki>
- [12] J. Poon, T. Dryja, “The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments,” <https://lightning.network/lightning-network-paper.pdf>, 2016.

BGP ネットワークにおける障害規模推定方法の検討

後谷 浩輔 小林 香
富山県立大学 工学部 情報システム工学科

概要： 本研究では，BGP ネットワークに障害が起こった時，その規模を推定する手法を検討する．BGP のメッセージの一つで経路更新情報を通知する UPDATE メッセージを用いて，障害規模を推定するアルゴリズムを検討する．

1. はじめに

インターネットは，何らかの影響で通信障害を発生させることが多々ある．そこでそのような障害に迅速に対応することが求められている．障害に迅速に対応するには，主に，位置と規模の大きさについて知りたい．しかし，BGP(Border Gateway Protocol) ネットワークにおいて，そのようなものを知るツールは少ない．そこで，BGP ネットワークにおける障害規模の推定手法を検討する．BGP は，いくつかのネットワーク (プレフィックス) をひとまとめた単位でルーティングを行う通信プロトコルである．現在，AS 間をつなぐルーティングプロトコルの多くは BGP を用いていて，BGP は TCP 上でピアを確立し，BGP メッセージをいくつかのピアとやりとりしている．メッセージの種類は，

- 通信を開通する際の OPEN メッセージ
- 生存確認をする KEEPALIVE メッセージ
- 経路更新情報を伝える UPDATE メッセージ

等がある．

2. 推定手法

BGP は，いくつかのネットワーク (プレフィックス) をひとまとめた単位でルーティングを行う通信プロトコルである．現在，AS 間をつなぐルーティングプロトコルの多くは BGP を用いていて，BGP は TCP 上でピアを確立し，BGP メッセージをいくつかのピアとやりとりしている．メッセージの種類には，

- 通信を開通する際の OPEN メッセージ
- 生存確認をする KEEPALIVE メッセージ
- 経路更新情報を伝える UPDATE メッセージ

等がある．その中で，本研究では経路更新情報を伝える UPDATE メッセージに着目した．UPDATE メッセージは，ピアに経路の差分情報を送るものであり，それには切断された (以後，withdrawn) プレフィックスと，新しく通信可能になったプレフィックス (以後，Announce されたプレフィックス) が含まれる．通常，新しい経路情報を受け取れば，即時 UPDATE メッセージによって差分の経路情報がピアに送出されることになっている．また，1 つのメッセージで，withdrawn プレフィックスや Announce されたプレフィックスをそれぞれ複数通知することが出来る．藤原 [5] は，1 つの UPDATE メッセージに含まれる withdrawn プレフィックスの数に注目して，一定の時間の UPDATE メッセージの中央値をとることで障害警告レベルを設定した．本研究の最終目標としては，この UPDATE メッセージを用いてリアルタイムに障害規模が推定出来ることである．本研究では，障害規模を推定するための幾つかのパラメータを UPDATE メッセージから抽出し，それらを 0~1 に正規化し，原点との距離を測定することで障害規模を推定する．

3. 検討

3.1. 用いるデータ

Update メッセージの取得先は Route Viwes Project[1] である．幾つかの AS の UPDATE メッセージを MRT format[2] で保存されており，それを RIPE NCC による libBGPDump[3] によりテキスト

データに変換する．表 1 にテキストデータに変換した UPDATE メッセージの内容サンプルを示す．表 1 のようにメッセージの届いた時刻，プロトコルタイプ，メッセージを送ってきた AS 番号，パス属性，AS パス，WITHDRAW，ANNOUNCE 等が含まれている．表 1 に，得られる情報の例を示す．

表 1: UPDATE メッセージから得られる情報例

TIME	11/11/14 00:00:08
TYPE	BGP4MP/MESSAGE/Update
FROM	メッセージを送ってきた IP アドレス AS 番号
TO	メッセージを受け取った IP アドレス AS 番号
ASPATH	通ってきた AS 番号のリスト
WITHDRAW	切断されたプレフィックスリスト
ANNOUNCE	新しく広告されたプレフィックスリスト

また，障害情報は，米 Dyn Research 社 [4](以後，Dyn 社) の情報を用いる．Dyn 社が提供するネットワーク障害情報 (internet event bulletin) の例として表 2 に 2015 年 4 月 30 日の障害 (outage) 情報を示す．

表 2: Dyn 社から得られる情報例

日時	障害のタイプ	影響を受けたネットワーク数	障害が起こった国
2015/04/30 04:08	outage	7	Paraguay
2015/04/30 05:13	outage	23	Paraguay
2015/04/30 06:36	outage	71	the United.States
2015/04/30 06:53	momentary_outage	23	Paraguay
2015/04/30 06:56	outage	57	the United.States
2015/04/30 07:02	outage	183	the United.States
2015/04/30 07:19	outage	15	Paraguay
2015/04/30 07:49	momentary_outage	23	Paraguay
2015/04/30 07:58	outage	15	Paraguay
2015/04/30 10:22	outage	57	Hong.Kong
2015/04/30 10:57	momentary_outage	88	Sweden
2015/04/30 10:59	momentary_outage	72	Sweden
2015/04/30 11:01	momentary_outage	71	Sweden
2015/04/30 11:07	outage	88	Sweden
2015/04/30 11:09	momentary_outage	42	Sweden
2015/04/30 15:57	outage	48	Philippines
2015/04/30 17:53	outage	68	Ukraine
2015/04/30 20:32	outage	111	Iran
2015/04/30 23:53	outage	101	Bulgaria

表 2 のように障害や復旧が発生した日時 (UTC)，その障害が起こった国，切断 (outage) または復旧 (restoration) したネットワーク数，障害の発生した経路の AS 番号などが掲載されている．しかし，Dyn 社の障害情報が，どの範囲までの UPDATE メッセージに影響があるかはわからない．しかし，経路の変更や消失情報を受け取れば，通常すべてのピアに通知するので，障害による UPDATE メッセージの影響は広範囲に渡ると考えた．従って，本研究では，障害が起こった国の違いを考慮しない．また，今回は，1 つの観測 AS the United.States にある routeviews_2 のみで検討する．

3.2. outagelevel の設定

Dyn 社の outage 情報から，障害のレベルを表 3 のように設定した．但し，同時刻に複数の国で障害が起こっている場合，それらを合計する．

表 3: outagelevel の設定

outagelevel	outage 計
1	0
2	1 ~ 100
3	101 ~ 200
4	201 ~

3.3. 正規化方法の設定

距離としては、最も一般的に使用されているユークリッド距離を用いるが、パラメータの値の幅に差があり、そのまま扱うと、値の幅が大きいパラメータの影響が非常に大きくなるため、正規化を行う必要がある。正規化には何種類がある。

1. 小数点をずらして合わせる方法
2. 最小値を 0, 最大値を 1 に合わせる方法
3. $Z=(X-\mu)/\sigma$ で変換する方法 (X が元のデータ, μ が平均, σ が分散)

等がある。本研究では、0 を原点にとれる 2 の方法で正規化した。変換する式は

$$\frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

である。

3.4. 用いるパラメータの検討

まず、データ点を時刻ごとまたはメッセージごとにとることが出来るが、時刻ごと（秒単位）のほうが設定できるパラメータが多いので時刻ごとにする。表 1 から得られる情報からいくつかのパラメータ候補を挙げた。

1. 前のメッセージ到着時刻との間隔 (interval)
2. Announce を含むメッセージ数 (Announce message)
3. 同時刻の Withdraw 計 (Withdraw length)
4. 同時刻の Announce (Announce length)
5. 同時刻の Withdraw を含むメッセージの割合 (Withdraw メッセージ率) (Withdraw message rate)
6. Withdraw を含むメッセージ数 (Withdraw message)

この 6 つのパラメータの正規化した値の分布から、outagelevel が上がればパラメータも上がる傾向があるものを選ぶ。つまり、outagelevel が上がれば中央値か平均値が上がるパラメータを選ぶ。正規化した 6 つのパラメータうち、例として同時刻の Announce 計の度数分布候補の度数分布を図 1, その中央値と平均値を図 2 に示す。但し、用いる UPDATE メッセージのデータは 2015 年 4 月の 1 ヶ月分で、度数分布の階級幅は 0.001 である。

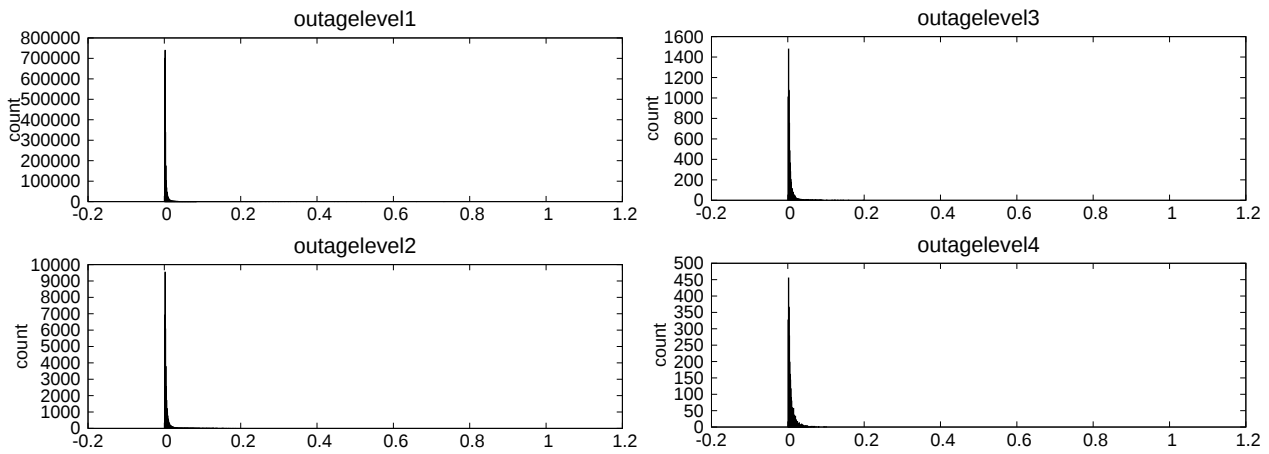


図 1: 2015 年 4 月の同時刻 Announce 計の outagelevel 別の度数分布

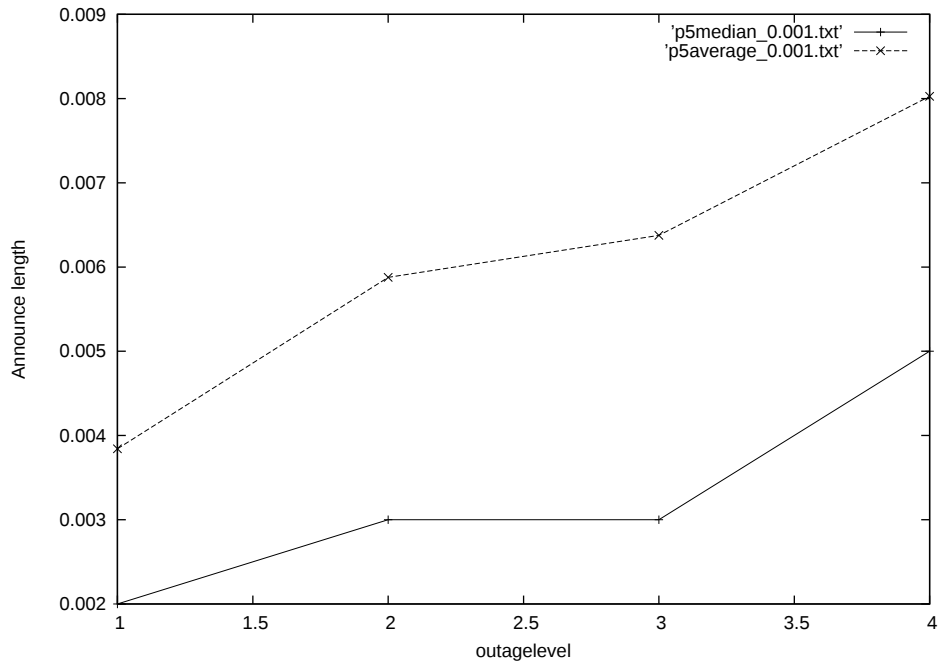


図 2: 2015 年 4 月の同時刻 Announce 計の中央値と平均値

図 2 より, outagelevel が上がれば同時刻の Announce 計は中央値も平均値も上がっていることから, パラメータとして選ぶ. このようにして, 障害規模推定に用いるパラメータを次の 4 つに絞った.

1. Announce を含むメッセージ数 (Announce message)
2. 同時刻の Withdraw 計 (Withdraw length)
3. 同時刻の Announce (Announce length)
4. Withdraw を含むメッセージ数 (Withdraw message)

3.5. 障害規模推定の分類方法

データ点の原点との最大の距離は2である．従って，原点との距離を d として，推定規模レベルを表4のように定める．

表 4: 推定規模レベルの設定

レベル0	$d \leq 1$
レベル1	$1 < d \leq 2$
レベル2	$2 < d \leq 3$
レベル3	$3 < d \leq 4$

4. 障害規模推定結果

3章の設定を踏まえ，3.4で選ばれたパラメータを用いて2015年4月30日，routeviews_2の障害規模を推定した結果を，

- Announce を含むメッセージ数 (Announce message)
- 同時刻の Announce (Announce length)
- Withdraw を含むメッセージ数 (Withdraw message)

の3軸で図3に示す．但し2015年4月30日は，障害規模推定レベル4はなかった．

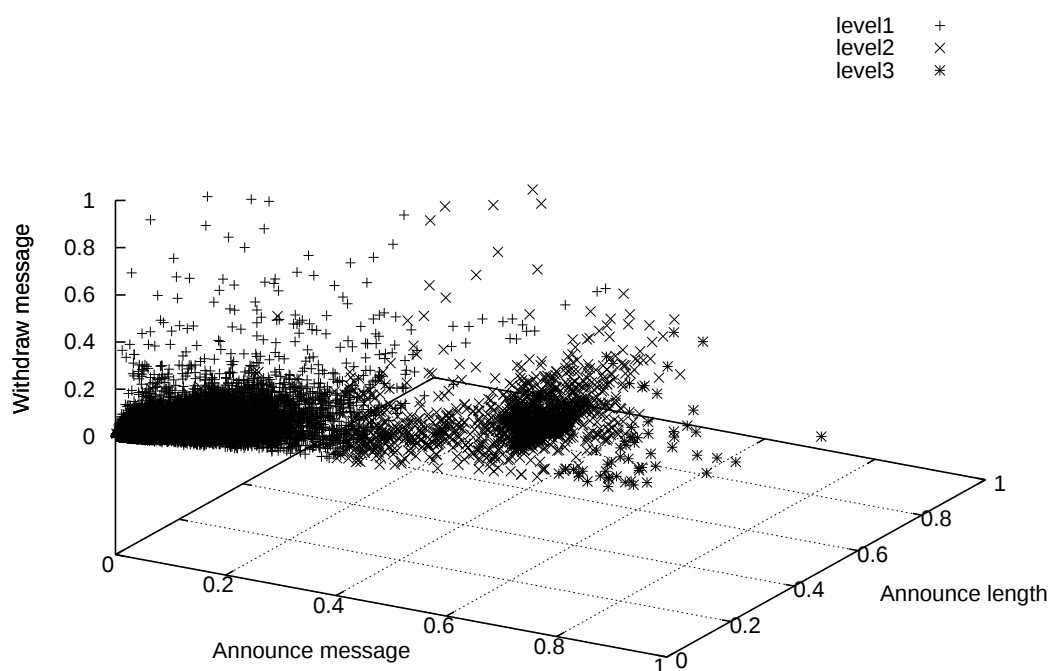


図 3: 2015 年 4 月 30 日の Announce message, Announce length, Withdraw message を軸にとった障害規模推定結果

5. 評価

障害規模推定をした結果を Dyn 社の障害報告情報で評価する．図4に2015年4月30日の障害規模推定レベルされた同時刻 Announce 計の時系列データを示す．

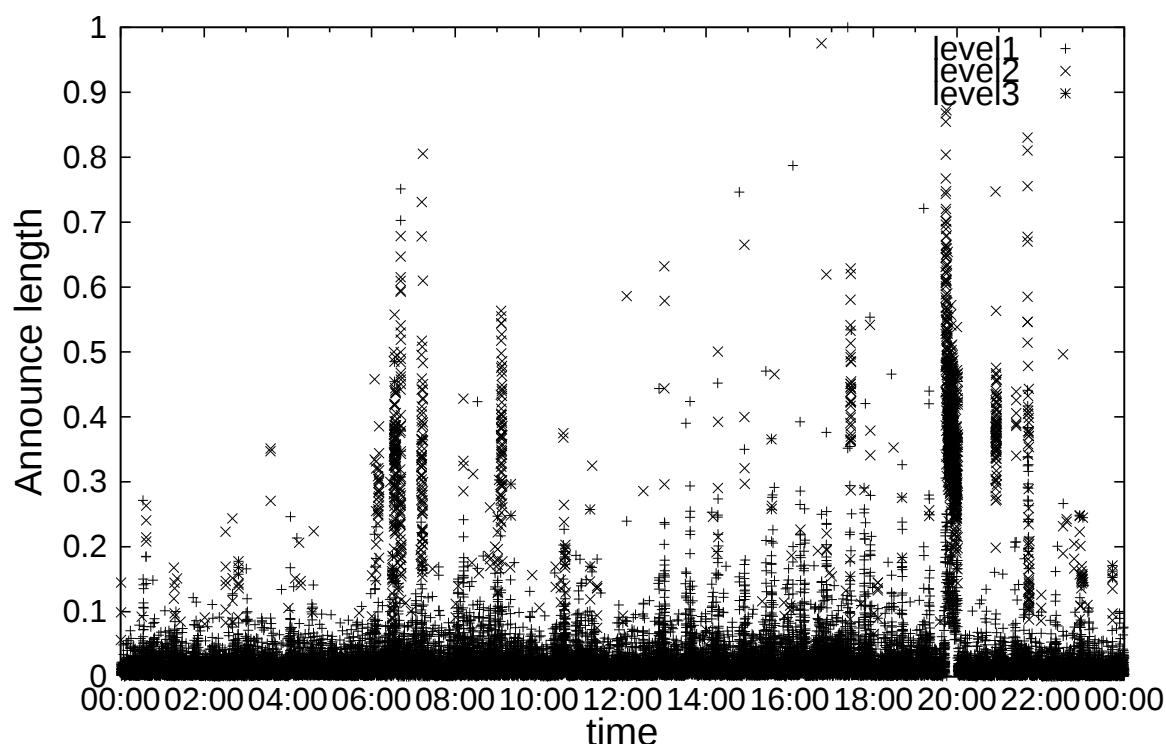


図 4: 2015 年 4 月 30 日の同時刻 Announce 計の時系列データ

表 2 に示した 2015 年 4 月 30 日の Dyn 社の障害情報と図 4 を見比べると, the United.States で障害が起きた 06:36, 06:56, 07:02, で障害規模推定レベルが 2 や 3 のデータが多数見られる．従って, この障害規模推定は United.States の障害において, 障害を検知出来たと言える．さらに, Dyn 社の報告にはないが, 20:00 あたりにも障害があったと予測できる．

6. まとめ

今回は, 1 つの AS についてのみで, その AS がある国の障害についてのみの検討を行ったが, 他の AS, 他の国の障害についてもどのような影響が起きるのか等を検討しなければならない．また今回は, 規模推定のレベル設定を単純に距離にし, 同心球上にレベルの境界を設けたが, 機械学習等を用いてより障害規模推定の精度が高まるような境界の設定方法が出来ると考える．従って今後は機械学習を用いた検討も行うことにする．

7. 謝辞

本研究を進めるにあたり, 多くのご指導を下さった小林 香講師に心から感謝いたします．また, BGPView のデータ提供元である GRI(Graceful Resilient network Investigation project), Update メッセージの提供元である米オレゴン大学の Route Views Project の参加者の方々に深く感謝いたします．

参考文献

- [1] University of Oregon Route Views Project, <http://www.routeviews.org/>
- [2] RFC6396, <https://tools.ietf.org/html/rfc6396>
- [3] libBGPdump, <https://bitbucket.org/ripence/bgpdump/wiki/Home>

- [4] Dyn Research , <http://research.dyn.com/>
- [5] 藤原一貴, 小林香, BGP メッセージの到着仮定を考慮した障害警告レベルの設定, 2015 年度待ち行列シンポジウム, 68-76
- [6] BGPView , <http://www.bgpview.org/>
- [7] RFC-4271 , <https://tools.ietf.org/html/rfc4271>
- [8] 林慶太, 小林香, インターネット接続障害時における迂回路形成の分析, 2014 年度待ち行列シンポジウム

BGP メッセージを用いた迂回経路検出手法の検討

山谷 周平 小林 香
富山県立大学 工学部 情報システム工学科

概要： インターネットの需要は年々増加しており、それに伴い安定した通信がもとめられている。本研究では迂回路の検出とマッピングを行うことで、障害発生時に重要度が高いと思われる AS を検出し、その経路変化を 1 日にわたり解析する。

1. はじめに

近年、ベトナムと香港間を結ぶ海底ケーブル「Asia-America Gateway(AAG)」の相次ぐ切断事故を受けて、NEC は 2016 年に日本からシンガポール間のアジア 11ヶ所を結ぶ大容量光海底ケーブル「Asia Pacific Gateway (APG)」を建設した。APG により不安定な AAG に依存せずに迂回ルートとしての働きに期待されている。海底ケーブルの建設はこれ以外にも、日米を結ぶ「FASTER」などがあり、世界の安定した通信需要が高まっていることが分かる。しかし、海底ケーブルの事故は地震やサメによる切断、人的な要因など様々でネットワーク障害を排除することは難しい。

このように、事故による障害などが発生した際、ネットワークは迂回路を形成して通信を保とうとする。この時、BGP(Border Gateway Protocol) という経路制御プロトコルが、AS(Autonomous System) と呼ばれる一意のポリシーの下で管理運用されるネットワークを相互に接続することで通信が行える。BGP は BGP メッセージという経路情報を AS 間で送受信させることで、経路の消失や更新を記録する。BGP は経路を自動で選択するが、経路の通信帯域が十分であるかは分からない。通信帯域が不十分だと、ルータの記憶する経路情報の数が大きな増減を繰り返すこととなり、パケットの転送に支障をきたすことがある。BGP オペレータは安定した通信を保つため、BGP ルータの設定を行い通信を維持する必要がある。しかし、BGP オペレータは経路の選択を BGP に任せているため、障害が発生した際、通信が行われていた経路がどのような迂回路を選択しているのか把握していない。そこで、迂回路を可視化し、障害が発生した際の迂回路形成パターンを把握できれば、それに応じた BGP ルータの設定が可能であると考ええる。

先行研究で林 [1], 田端 [2] は、AS 毎の prefix 数に着目して、1 日を通して一定以上の変化をしている AS の prefix 数を見ることで迂回路を検出する手法を提案した。ネットワーク障害が発生した時間帯に、先行研究の検出方法を用いることで重要な AS を特定し、その AS を含んだ経路変化を検出することができる。

本研究では、実際に障害が発生した日の通信経路を解析し、地図上にマッピングすることで経路の変化を可視化する。また、障害が発生した際の迂回路データを蓄積することで重要な AS を発見し、その AS の通信帯域増強などにつなげる。

2. BGP

2.1. BGP(Border Gateway Protocol)

BGP は経路制御プロトコルの一種で、AS(Autonomous System) と呼ばれる一意のポリシーの下で管理運用されるネットワークを相互に接続する際に用いられる。各 AS には AS 番号が割り当てられ、BGP はそれを下に関々の経路を識別する。AS 番号は原則として 2 バイト (16 ビット) の大きさを持ち、0 から 65535 までの整数で表現されている。しかし、AS 番号の需要増加に伴い枯渇が予想され、2007 年 3 月から 4 バイト (32 ビット) で表記する 4 バイト AS の割り当てを行っている。BGP は隣接ルータ間の通信に、ポート 179 で TCP を使用する。隣接ルータがお互いに TCP セッションを確立すると「メッセージ」という形で BGP 情報を交換する。経路が確立した後は、経路情報を伝達するためにメッセージを定期的に交換し、お互いの持つ経路情報を保持、更新し続ける。

2.2. BGP による経路形成

BGP を実行するほとんどのネットワークは、複数のネットワークの接続している。これは、複数の BGP 隣接ルータから複数の宛先に到達できることを意味し、BGP では様々な経路の中からベストルートを選択する。ベストルートに障害が発生した場合、BGP ルータは到達できる宛先から最適な経路を選択し、新たなベストルートを選択し、これが迂回路となる。

2.3. フルルート

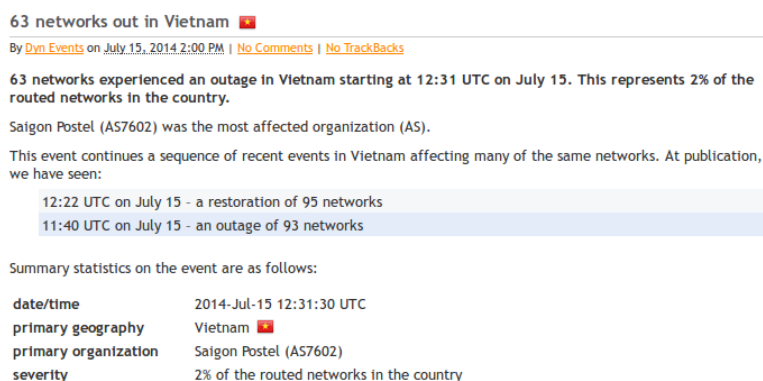
本研究室では BGPView というプログラムを使用して、フルルートと呼ばれるデータを取得している。フルルートは観測 AS から観測可能なすべての宛先への経路情報 (Prefix, Next Hop, MED 値, 継続時間, ASpath など) をまとめたものであり、30 分毎にダンプされる。

3. 迂回路検出手法

3.1. 分析対象とする障害事例

迂回路検出手法を評価するために、分析対象として Dyn 社にイベント報告がある日のフルルートを用いる。今回は、2014/7/15 20:36(JST) に発生した AAG 光海底ケーブルの切断事故と、2015/4/23 5:45(JST) に発生した AAG 光海底ケーブルの切断事故を対象とする。

図 1 に 2014/7/15 の Dyn 社に報告があったイベントを示す。



The screenshot shows a report from Dyn Events dated July 15, 2014, at 2:00 PM. The title is '63 networks out in Vietnam' with a Vietnam flag icon. The report states that 63 networks experienced an outage in Vietnam starting at 12:31 UTC on July 15, representing 2% of the routed networks in the country. It identifies Saigon Postel (AS7602) as the most affected organization. A timeline shows a restoration of 95 networks at 12:22 UTC and an outage of 93 networks at 11:40 UTC. Summary statistics include the date/time (2014-Jul-15 12:31:30 UTC), primary geography (Vietnam), primary organization (Saigon Postel (AS7602)), and severity (2% of the routed networks in the country).

63 networks out in Vietnam 🇻🇳	
By Dyn Events on July 15, 2014 2:00 PM No Comments No TrackBacks	
63 networks experienced an outage in Vietnam starting at 12:31 UTC on July 15. This represents 2% of the routed networks in the country.	
Saigon Postel (AS7602) was the most affected organization (AS).	
This event continues a sequence of recent events in Vietnam affecting many of the same networks. At publication, we have seen:	
12:22 UTC on July 15 - a restoration of 95 networks	
11:40 UTC on July 15 - an outage of 93 networks	
Summary statistics on the event are as follows:	
date/time	2014-Jul-15 12:31:30 UTC
primary geography	Vietnam 🇻🇳
primary organization	Saigon Postel (AS7602)
severity	2% of the routed networks in the country

図 1: 20140715 における Dyn 社のレポート

3.2. 迂回路の検出

フルルートが 30 分毎にダンプされるため、処理する時刻とその 30 分後のフルルートで同じ prefix かつ、ASpath が違うものをペアで抽出した。この時、迂回路が確立されたことを検出するため、抽出する経路は継続時間が 30 分以内のものとする。

4. 評価

4.1. 2014/7/15 の事故についての分析

実際に機能していた迂回路に多く含まれていた AS-A とその迂回路ペアである AS-A' を含む prefix 数の時間毎の変化を図 2 に示す。

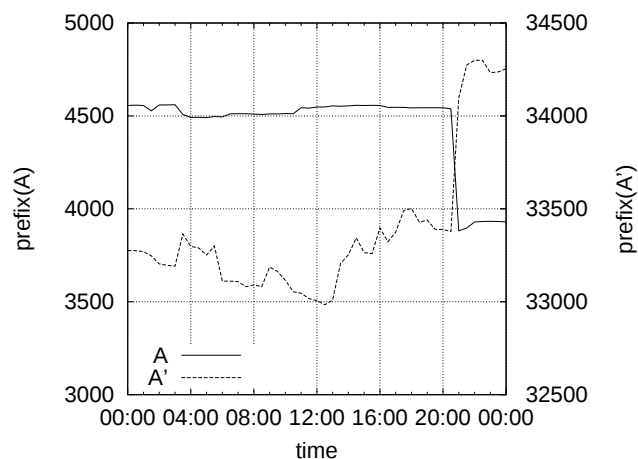


図 2: AS-A,A' を含む prefix 数の変化

Dyn 社では障害によって 20:40(JST) に 93 のネットワークが消失し, 21:22(JST) に 95 のネットワークが復元したと報告されている. 20:30 頃に AS-A を含む prefix 数は減少し 21:00 頃, AS-A' が増加していることから, 迂回路として経路が確立したことが分かる.

参考文献

- [1] 林 慶太, 小林 香, インターネット接続障害時における迂回路形成の分析, 2014 年度待ち行列シンポジウム報文集.
- [2] 田端 浩志, BGP ネットワークにおける迂回路検出アルゴリズムの提案, 2015 年度富山県立大学卒業論文.

ランダム多スタート局所探索法に対する極値統計を用いた計算停止基準

仲川 諒馬 a 増山博之 a 檀寛成 b 高橋豊 a

a: 京都大学大学院 情報学研究科

b: 関西大学

概要： 本論文では、組合せ最適化問題に対する代表的な近似解法の一つであるランダム多スタート局所探索法の計算停止基準として、暫定最適値の単位時間期待改善率を提案する。なお、単位時間期待改善率の推定には、極値統計におけるブロック最大法と閾値法の2つの手法を用いる。巡回セールスマン問題を例として数値実験を行い、2つの手法でそれぞれ推定された単位時間期待改善率の比較や、計算停止基準としての有効性を検討する。

1. 序論

大規模な組合せ最適化問題に対しては、膨大な計算時間を必要とする厳密解法ではなく、現実的な時間で「質の良い」解を求める近似解法がよく用いられる。本研究では代表的な近似解法の一つであるランダム多スタート局所探索法について考える。ランダム多スタート局所探索法では、ランダムに選んだ複数の初期解から局所探索によって得られた局所的最適解を最適解候補とし、その中から最良のものを「最適解」として採用することが一般的である。

ランダム多スタート局所探索法において、計算の停止判断は非常に難しい問題である。計算時間を長く取れば取るほど、より良い解が求まる可能性は高くなるが、求めた解が大域的最適解である保証はない。したがって、現実的にランダム多スタート局所探索法を用いる際には、あらかじめ設定した計算時間内で得られた最善の局所的最適解を「最適解」として採用するのが一般的である。しかし、計算期間の設定は経験則に拠るところが大きく、その判断が適当であるかを定量的に測ることは容易ではない。

本研究では、上で述べたような計算停止判断を、極値統計を用いて行うことを検討する。極値統計を組合せ最適化問題に応用した研究は過去にも数多く行われているが、それらは最適値の推定を目的としている(サーベイ論文[1]参照)。一方、本研究は、ランダム多スタート局所探索法の実行過程における暫定解の単位時間期待改善率を新たな指標として提案する。暫定解の単位時間期待改善率とは、計算を継続した場合に見込まれる暫定解の単位時間あたりの平均改善率である。この提案指標の有用性を議論するため、代表的な組合せ最適化問題である巡回セールスマン問題(TSP: Traveling Salesman Problem)を対象とし、2-opt 近傍探索法を実装したランダム多スタート局所探索法による数値実験を行う。

2. 極値統計

本章では、極値統計の理論について簡単に述べる。一般の統計学が、母集団分布の平均まわりの特性について推定を行うのに対し、極値統計では母集団分布の裾に関する特性について推定を行う。極値統計の推定に用いられる確率分布は大きく分けて2つ存在する。一つは一般極値分布と呼ばれ、これはブロック化された観測データ内の最大値が従う分布である。もう一つは一般パレート分布と呼ばれ、観測データの中で特定の閾値を超えるデータが従う分布である。一般極値分布を用いた極値モデルをGEV (Generalized Extreme Value) モデルと呼び、一般パレート分布を用いたモデルをGP (Generalized Pareto) モデルとよぶ。GEV モデルではデータ数が足りないときにはGP モデルがよく用いられる。他にも、推定の信頼性を向上させるために、上位 r (> 1) 個のデータを用いる r GEV モデルが存在する。なお、ここでモデルはいずれも観測データの独立性を仮定しているが、観測対象によってはこの仮定が適切でない場合も考えられる。そのようなデータに対しては、独立性の条件を緩和した非定常極値モデルが用いられる。

以下の節では、GEV モデルおよびGP モデルについて詳述する。

2.1. GEV モデル

2.1.1. 一般極値分布

GEV モデルに用いられる一般極値分布を導入する．まず，実数値データ (確率変数) の列 $\{X_n \in \mathbb{R} := (-\infty, \infty); n = 1, 2, \dots\}$ が独立かつ同一に分布 F に従うとする．さらに，確率変数 M_n ($n = 1, 2, \dots$) を以下のように定義する．

$$M_n = \max\{X_i; i = 1, 2, \dots, n\}. \quad (1)$$

このとき， n をブロックサイズ， M_n をブロック最大値データとよぶ．ブロック最大値データ M_n の分布に関して，以下の結果が知られている．

定理 1 (Fisher-Tippett の定理) ある定数 $c_n > 0$, $d_n \in \mathbb{R}$ と，非退化な分布 G が存在して，

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{M_n - d_n}{c_n} \leq z \right) = G(z), \quad z \in \mathbb{R},$$

となるならば，分布 G は次のように表現できる．

$$G(z) = e^{-(1+\xi z)^{-1/\xi}} =: G_\xi(z), \quad z \in \mathbb{R}.$$

ただし， $1 + \xi z > 0$ である．このとき，分布 F は一般極値分布の吸引領域に属するといい， $F \in \mathcal{D}(G_\xi)$ と表記する．

上の定理から，ブロックサイズ n が十分大きいとき，

$$\Pr(M_n \leq z) \approx G_\xi \left(\frac{z - d_n}{c_n} \right), \quad z \in \mathbb{R}, \quad (2)$$

となる．ここで，式 (2) の右辺に現れる定数 c_n, d_n をそれぞれ σ, μ と置き換えて得られる $G_\xi((z - \mu)/\sigma)$ を一般化極値分布と呼び， $\text{GEV}(\mu, \sigma, \xi)$ と表す．

$$\text{GEV}(\mu, \sigma, \xi) = \exp \left\{ - \left(1 + \xi \frac{z - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right\}, \quad z \in \mathbb{R}. \quad (3)$$

2.1.2. ブロックサイズの決定手法

一般極値分布 G_ξ を推定する際に問題となるのは，式 (1) におけるブロックサイズ n の設定である．一般極値分布は，ブロックサイズ n を無限大に向かって増大させたときのブロック最大値データ M_n の漸近分布である．したがって，ブロックサイズ n はできるだけ大きい方が望ましいと考えられる．一方で，ブロックサイズ n を大きくしすぎると，一般極値分布 G_ξ のパラメータ (μ, σ, ξ) の推定に利用できるデータ数が少なくなり，パラメータの推定精度が低下する．このトレードオフは Bias-Variance のジレンマと呼ばれる．

Bias-Variance のジレンマにより，ブロックサイズ n を決めるのは容易ではないが，本研究では，次式で定義される期待平均対数尤度 $L(n)$ を最大にするようなブロックサイズ n を採用する．

$$L(n) = \frac{1}{m} \sum_{i=1}^m \log g_{\xi,n}(z_i). \quad (4)$$

ただし， m はブロック最大値データ数， $g_{\xi,n}$ はブロックサイズを n として推定される一般極値分布 $G_{\xi,n}$ の確率密度関数である．ここで，

$$g_{\xi,n}(z) = \frac{\partial}{\partial z} G_{\xi,n}(z).$$

であることに注意する．

さて、期待平均対数尤度 $L(n)$ を最大化する意味について述べる。ブロックサイズを n としたときの真の分布を G_ξ とし、その密度関数を g_ξ とする。推定された分布 $G_{\xi,n}$ が G_ξ にどれだけ近いかを表す尺度として、以下の Kullback-Leibler 情報量 (以下 K-L 情報量) がある。

$$\begin{aligned} I(g_\xi(z); g_{\xi,n}(z)) &= E_{G_\xi} \left[\log \frac{g_\xi(Z)}{g_{\xi,n}(Z)} \right] \\ &= E_{G_\xi} [\log g_\xi(Z)] - E_{G_\xi} [\log g_{\xi,n}(Z)]. \end{aligned} \quad (5)$$

ただし、 $E_{G_\xi}[\cdot]$ は分布 G_ξ に関する期待値を表す。式 (5) の第二項は

$$E_{G_\xi} [\log g_{\xi,n}(Z)] = \int \log g_{\xi,n}(z) dG_\xi(z), \quad (6)$$

と書くことができ、平均対数尤度と呼ばれる。

式 (5) から、分布 $G_{\xi,n}$ が G_ξ に近いほど、 $I(g_n(z); g_{\xi,n}(z))$ の値は小さくなることがわかる。式 (5) の第一項は定数であるから、第二項の平均対数尤度の値が大きいほど、真の分布に近いといえる。しかし、真の分布 G_ξ は未知であるから、式 (6) で与えられる平均対数尤度を陽に求めることはできない。そのため、平均対数尤度を推定する必要がある、その自然な推定量として、式 (4) で表される期待平均対数尤度がある。したがって、 $L(n)$ を最大化するブロックサイズ n を選択するということは、推定される分布 $G_{\xi,n}$ が、K-L 情報量を尺度としたとき、真の分布 G_ξ に最も近くなるよう意図したものと言える。

2.2. GP モデル

本小節では GP モデルについて述べる。GEV モデルではブロック最大値データを用いたが、データ数が十分でない場合も多い。そのような場合、推定の信頼性は失われてしまう。そこで、GP モデルではまず、適当に閾値を定め、その閾値を超えたデータのみを採用する。次に、採用されたデータの値から閾値を引いたものを閾値超過データと呼び、これを用いて母集団の裾の推定を行う。

以下では、GP モデルに用いられる一般パレート分布について述べる。

2.2.1. 一般パレート分布

第 2.1.1 節の GEV モデルの説明と同様に、データの列 $\{X_n; n = 1, 2, \dots\}$ が独立かつ同一に分布 F に従うとする。このとき、閾値 u に関する閾値超過データの分布 (以下、閾値超過分布とよぶ) を F_u とすると、

$$F_u(y) = \Pr(X - u \leq y \mid X > u), \quad y \geq 0, \quad (7)$$

となる。この分布 F_u について、以下の定理が知られている。

定理 2 ([4, 定理 2.4.1]) 以下の (i), (ii) は等価である。

- (i) 分布 F は一般極値分布の吸引領域に属する。すなわち、 $F \in \mathcal{D}(G_\xi)$ である。
- (ii) ある関数 $a: \mathbb{R} \rightarrow (0, \infty)$ が存在して、次式が成り立つ。

$$\lim_{n \uparrow x_F} \sup_{0 < y < x_F - u} |F_u(y) - H_\xi(y/a(u))| = 0.$$

ただし、 $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ であり、 $H_\xi(x)$ は標準一般パレート分布とよばれ、以下の式で表される。

$$H_\xi(x) = 1 - (1 + \xi x)_+^{-1/\xi}.$$

ただし、 $(z)_+ := \max\{z, 0\}$ である。

上の定理から、分布 F が一般極値分布 G_ξ の吸引領域に属しているならば、十分大きな閾値 u に関する閾値超過データは分布 $H_\xi(y/a(u))$ に従うと考えられる。ここで、関数 $a(u)$ を σ に置き換えて得られる分布 $H_\xi(y/\sigma)$ を一般パレート分布と呼び、 $GP(\sigma, \xi)$ と表す。

$$GP(\sigma, \xi) = H_\xi(y/\sigma) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)_+^{-1/\xi}. \quad (8)$$

2.2.2. 一般パレート分布における閾値の決定

GEV モデルにおけるブロックサイズの選択と同様に, GP モデルにおいては閾値 u の設定が重要な問題となる. 以下では, 本研究で採用した閾値の決定について簡単に述べる. 詳細は高橋ら [4] を参照されたい.

まず, 値 u を超えるデータを $y_{[1]}, y_{[2]}, \dots, y_{[n_u]}$ とする. このとき, 値 u に関する閾値超過データ $\{y_{[i]} - u\}_{i=1}^{n_u}$ から, 最尤法を用いて推定される一般パレート分布 $GP(\sigma, \xi)$ のパラメータをそれぞれ $\hat{\sigma}$ と $\hat{\xi}$ とする. 値 u を色々と変化させて, $(u, \hat{\sigma}^*)$ と $(u, \hat{\xi})$ をプロットする. ただし, $\hat{\sigma}^* = \hat{\sigma} - \hat{\xi}u$ は修正尺度の推定値である. この二つの図で, ある値より右で推定値 $\hat{\sigma}^*$ と $\hat{\xi}$ がともに一定であると見なせるとき, それらの最小の値を閾値と決める.

こうして求められた推定分布 $\hat{H}_{\hat{\xi}}$ とデータが適合しているかは, 確率プロットおよび確率点プロットが用いて確かめられる. 確率プロットは

$$\left\{ \left(\hat{H}_{\hat{\xi}}(y_{(i)}), \frac{i}{n_u + 1} \right) : i = 1, \dots, n_u \right\},$$

で与えられ, 確率点プロットは

$$\left\{ \left(\hat{H}_{\hat{\xi}}^{-1} \left(\frac{i}{n_u + 1} \right) + u, y_{(i)} + u \right) : i = 1, \dots, n_u \right\},$$

で与えられる. ただし,

$$\hat{H}_{\hat{\xi}}^{-1} \left(\frac{i}{n_u + 1} \right) = \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(\frac{n_u + 1 - i}{n_u + 1} \right)^{-\hat{\xi}} - 1 \right],$$

である. この2つのプロットが対角線から大きく外れるような場合には, GP モデルを当てはめるのは適当ではないと判断する.

3. 極値統計による計算停止指標の提案

本節では, まず, 極値統計を応用して, 組合せ最適化におけるランダム多スタート局所探索法の計算停止指標を提案する. 具体的には, 局所最適値のデータから, 一般極値分布あるいは一般パレート分布を推定し, 単位時間計算後における解の改善率の期待値を計算停止指標として定義する. その後, 定義した単位時間期待改善率を TSP の近似解法に適用することを想定し, その具体的な手順および注意点を述べる.

3.1. 単位時間期待改善率

本小節では, 組合せ最適化における計算停止指標として単位時間期待改善率を定義する. なお, 以下では目的関数の最大化問題について述べる. これは, 極値統計では最大値データの方が扱いやすいためである. 最小化問題においても, 目的関数に -1 を乗ずることで最大化問題に読みかえられるため, 一般性は失われない.

さて, 提案する計算停止指標の概要について述べる. まず, ランダム多スタート局所探索法により得られた目的関数の局所最適値データに対して, GEV モデルおよび GP モデルを適用し, 一般極値分布および一般パレート分布をそれぞれ推定する. さらに, 推定された一般極値分布および一般パレート分布を用いて, 計算停止基準となる単位時間計算後の最適値の改善率, すなわち単位時間期待改善率を求める. 以下では局所最適値データに対して GEV モデルを適用した場合についてのみ, 計算停止指標の述べているが, GP モデルを用いる場合でも同様に議論できる.

ランダム多スタート局所探索法によって, n ($n = 1, 2, \dots$) 番目に得られた最適解候補の目的関数値を Y_n とし, $\{Y_n; n = 1, 2, \dots\}$ は独立かつ同一に分布すると仮定する. また, $\{Y_n\}$ を局所最適値データと呼び, これらのデータを得られた順に N 個ずつまとめたものをブロック, そして N をブロックサイズとよぶ. さらに, k ($k = 1, 2, \dots$) 番目のブロック $\{Y_{(k-1)N+1}, Y_{(k-1)N+2}, \dots, Y_{kN}\}$ の最大値を Z_k とし, これをブロック最大値とよぶ. 仮定より, $\{Z_k; k = 1, 2, \dots\}$ は独立かつ同一に分布し, その共通の分布を G とする.

いま, いくつかのブロック最大値が得られているとして, それらの中で最大の値を $\hat{z}$ とする. また, 1 ブロック分の試行にかかる計算時間の平均を T_B とする. ここで, T と R を次のように定義する.

$$T = \frac{T_B}{1 - G(\hat{z})}, \quad (9)$$

$$\begin{aligned} R &= \int_{\hat{z}}^{\infty} \frac{z - \hat{z}}{|\hat{z}|} \frac{dG(z)}{1 - G(\hat{z})} \\ &= \frac{1}{|\hat{z}|} \frac{\int_{\hat{z}}^{\infty} \{1 - G(z)\} dz}{1 - G(\hat{z})}. \end{aligned} \quad (10)$$

定義から, T は追加実験を行った場合に暫定最適値 $\hat{z}$ が改善されるまでの平均時間である. また, R は 1 ブロック分の追加実験によって暫定最適値が更新されるという条件の下での期待改善率である. したがって, T と R をそれぞれ平均更新期間と条件付きブロック単位期待改善率とよぶ. さらに, 次式で定義される r を, 暫定最適値の単位時間期待改善率とよぶ.

$$r = \frac{R}{T} = \frac{\int_{\hat{z}}^{\infty} \{1 - G_{\xi}(z)\} dz}{|\hat{z}| T}. \quad (11)$$

暫定最適値 $\hat{z}$ は計算を継続していく過程において非減少であるので, 単位時間期待改善率 r は非増加となる. よって, r が十分小さくなった時点で計算を停止すれば, 暫定最適値が更新される見込みのない無駄な計算を回避できる.

以上の議論では GEV モデルを用いているが, GP モデルを用いる場合には, 局所最適値データ $\{Y_n\}$ と適当な閾値 u を用いて, 一般パレート分布 H_u のパラメータを最尤推定すればよい. ただし, G_{ξ} が 1 ブロックにおける最大の局所最適値の分布を表すのに対し, H_u は局所最適値が閾値 u を超えたとき, その超過分の値であることを注意が必要である. したがって, 式 (11) から単位時間期待改善率を求める際には, $G_{\xi}(z)$ を $H_u(z - u)$ で置き換える. また, n 個の局所最適値データのうち n_u 個が閾値を超えたとすれば, 閾値を超過するデータは平均して n/n_u 回に一度現れると考えられるので, T_B は 1 つの局所最適値データを得る平均計算時間に n/n_u を乗じた値とする.

3.2. TSP への適用

本研究では, 単位時間期待改善率が, ランダム多スタート局所探索法の計算停止基準として妥当なものか検証するため, TSP を例に挙げて数値実験を行う. 以下では, 数値実験を行う際の注意点だけを述べ, 数値実験結果については, 口頭発表時に考察を添えて紹介する.

本研究の数値例では, 2-opt を実装したランダム多スタート局所探索法を用い, 初期点はメルセンヌ・ツイスタを用いて選択する. 極値統計を TSP に応用するにあたり, 巡回路の総距離に -1 を乗じたものを目的関数とし, 最大化問題に読み換える. また, 局所最適値データ $\{Y_n; n = 1, 2, \dots\}$ から一般化極値分布 $GEV(\mu, \sigma, \xi)$ のパラメータ μ, σ, ξ , または一般パレート分布 $GP(\sigma, \xi)$ のパラメータ σ, ξ を推定する際には, $\{Y_n\}$ が独立であることが前提となる. そこで, Swed ら [3] が提案した手法を用い, $\{Y_n\}$ が独立であるとみなすことの妥当性を確認した. なお, パラメータの推定には最尤法を用いる.

参考文献

- [1] A. P. Giddings, R. L. Rardin, and R. Uzsoy, “Statistical optimum estimation techniques for combinatorial optimization problems: a review and critique,” *Journal of Heuristics*. vol. 20, No. 3, pp. 329–358, 2014.
- [2] B. L. Golden, and F. B. Alt, “Interval estimation of a global optimum for large combinatorial problems,” *Naval Reserach Logistics Quarterly*. vol. 29, pp.69–77, 1979.
- [3] F. S. Swed, and C. Eisenhart, “Tables for testing randomness of grouping in a sequence of alternatives,” *The Annals of Mathematical Statistics*. vol. 14, No. 1, pp. 66–87, 1943.
- [4] 高橋倫也, 志村隆彰, 極値統計学, 近代科学社, 2016.
- [5] 柳浦睦憲, 茨木俊秀, 組合せ最適化—メタ戦略を中心として—, 朝倉書店, 2001.

既約で Lumpable なマルコフ連鎖の上界の改善

群馬大学大学院 電子情報数理教育プログラム 石川 真也
群馬大学大学院 理工学府電子情報部門 河西 憲一

Abstract 本論文では、大きな状態空間上の離散時間マルコフ連鎖において、増大した状態空間を縮小するアルゴリズムについて考え、より緊密かつ既約な上界を与えるマルコフ連鎖を構成する方法について提案する。文献 [1] では、性能指標の上界を求める方法として、マルコフ連鎖に確率的な順序関係を定めて、さらに増大した状態空間のサイズを減少させるアルゴリズムが提案されている。同アルゴリズムは、元のマルコフ連鎖が既約であれば、上界を与えるマルコフ連鎖も既約になるように設計されている。しかしながら、このアルゴリズムが導入する既約性を保証する手法では、必ずしも緊密な上界を与えるとは限らない。本論文ではこの課題を指摘し、その原因を考察していくとともに、解決策を提案する。待ち行列モデルなどのマルコフ連鎖に対して数値実験をし、本論文で提案する既約性を保ちつつより緊密な上界を得るための手法について検証していく。

1. はじめに

サービスシステムの性能を解析する目的で、マルコフ連鎖による待ち行列モデルが用いられる。現実的なモデル化のためにパラメータを細かく設定するとマルコフ連鎖の状態空間が増大し、計算にかなりのコストを費やすことになってしまう。マルコフ連鎖の定常分布からシステムの性能評価指標を計算するが、実際には厳密な計算結果に必ずしも拘泥する必要はなく、性能評価指標の上界でも十分に有用な情報となり得る。性能指標の上界を求める方法として、マルコフ連鎖に順序関係を定めて評価する方法がある。文献 [1] での上界の求め方は、マルコフ連鎖の確率的な順序関係に基づいたアルゴリズム的な手法である。このアルゴリズムは Vincent のアルゴリズム [2] を拡張させたものであり、アルゴリズムに入力として与える元のマルコフ連鎖が既約であるならば、出力として得られるマルコフ連鎖も既約であることが保証される。よって、得られたマルコフ連鎖の既約性を検証する必要がないため非常に有用である。

Vincent のアルゴリズムを拡張したアルゴリズムが既約性を保証する状態空間は、元のマルコフ連鎖の状態空間を意味する。文献 [1] ではこの評価方法に加えて、さらに元の状態空間のサイズを減少させるアルゴリズムが提案されている。このアルゴリズムは元のマルコフ連鎖が既約でありかつある条件を満たすならば、既約性を保証したサイズの小さいマルコフ連鎖を与える点に特徴があり、大きな状態空間をもつマルコフ連鎖の場合に有効である。

既約性を保証するために、これらのアルゴリズムではある操作を行うことで既約性を保証している。しかしその操作を行うことで、過大な上界を与えてしまう可能性があることが否定できない。本論文では文献 [1] で提案されているアルゴリズムの既約性を保証するための手法に焦点を当て、より緊密な上界を求めるための方法を検討していく。2 章では、文献 [1] のアルゴリズムについて述べ、3 章ではその課題を指摘していく。その課題の解決方法を 4 章で述べ、5 章では実際のモデルを使うことで、数値計算を行い、本論文で提案するアルゴリズムの有用性について述べていく。

2. 確率的上界のアルゴリズム

まず、本論文を進めるにあたって離散時間マルコフ連鎖における確率的な順序付けについての必要な定義、関連する記号を記載する。

2.1. マルコフ連鎖の比較

以下の定義式にしたがって、マルコフ連鎖の順序関係を定める。有限状態空間 S 上の離散時間マルコフ連鎖を考える。状態空間 S 上には全順序が定義されているとする。以下、特に断らないかぎり、 N を正の整数として、 $S = \{1, 2, \dots, N\}$ とする。

定義 1 状態空間 S 上に値をとる確率変数 X と Y について、任意の増加関数 $f: S \rightarrow \mathbb{R}$ に対して、 $E[f(X)] \leq E[f(Y)]$ であるならば、またそのときに限り、 Y は X より確率順序の意味で大きいといい、 $X \preceq_{\text{st}} Y$ と書く。

注意 1 確率変数 X と Y の確率関数をそれぞれ $p: S \rightarrow \mathbb{R}$ と $q: S \rightarrow \mathbb{R}$ とする。 $X \preceq_{\text{st}} Y$ であることと、任意の

$i \in \mathcal{S}$ について $\sum_{j=i}^N p[j] \leq \sum_{j=i}^N q[j]$ であることは同値である。ただし、 $p[j] = \Pr\{X = j\}$, $q[j] = \Pr\{Y = j\}$ である。確率関数 p は、 j 番目の成分が $p[j]$ である確率ベクトル p と同一視してもよい。よって、 $X \preceq_{\text{st}} Y$ であることを $p \preceq_{\text{st}} q$ と記述することもある。

定義 2 P と Q を状態空間 $\mathcal{S}$ 上の離散時間マルコフ連鎖の推移確率行列とする。 P の i 行目を $P[i, *]$, Q の i 行目を $Q[i, *]$ とする。任意の $i \in \mathcal{S}$ について、 $P[i, *] \preceq_{\text{st}} Q[i, *]$ が成立するならば、 Q は確率順序の意味で P よりも大きいといい、 $P \preceq_{\text{st}} Q$ と書く。(比較可能性)

マルコフ連鎖をサンプルパス毎に比較するために、推移を進めても確率順序が保存される性質を導入する。

定義 3 P を確率行列とする。全ての $u \preceq_{\text{st}} v$ となる確率ベクトル u, v に対して、

$$uP \preceq_{\text{st}} vP,$$

が成立するとき、 P は確率順序の意味で単調であるという。

注意 2 確率行列 P が単調であるならば、任意の $i < N$ に対して、 $P[i, *] \preceq_{\text{st}} P[i+1, *]$ が成立することが知られている。

これらの定義はマルコフ連鎖の順序付けを可能とするためのものである。これらの不等式を使いマルコフ連鎖の上界を考えていく。

2.2. 確率的な上界を求めるアルゴリズム

文献 [1] では離散時間マルコフ連鎖の推移確率行列の上界を求めるアルゴリズムが記述されている。本節ではそのアルゴリズムについて述べる。初めに、記述を簡単にするために以下の関数を導入する。

定義 4 正方行列の集合を $\mathcal{M}$ とする。状態空間 $\mathcal{S}$ 上のマルコフ連鎖が定める推移確率行列 $P \in \mathcal{M}$ に対して、 $\mathcal{M}$ から $\mathcal{M}$ への関数 $c: \mathcal{M} \rightarrow \mathcal{M}$ を、

$$c(P)[i, j] = \sum_{k=j}^N P[i, k], \quad \forall i, j \in \mathcal{S}$$

で定義する。

関数 c の逆関数 c^{-1} を次のように定義する。

定義 5 $M \in \mathcal{M}$ に対して、

$$c^{-1}(M)[m, n] = \begin{cases} M[m, N], & n = N, \\ M[m, n] - M[m, n+1], & n = 1, 2, \dots, N-1. \end{cases}$$

注意 3 確率行列の各行の確率ベクトルに対して、関数 c は確率ベクトルの累積補分布を求める演算となる。同様に、累積補分布を行の要素とする行列に対して、 c^{-1} は確率行列を与える。

$A, B \in \mathcal{M}$ に対して、 $B - A$ が非負行列、すなわち全ての要素が非負である行列であるとき、 $A \leq B$ と書くことにすれば、 P と Q を確率行列とすると、 $P \preceq_{\text{st}} Q$ であることと、 $c(P) \leq c(Q)$ であることは同値である。次の定理が成立する [4]。

定理 1 P と R を推移確率行列とする離散時間マルコフ連鎖を $\{X_n; n \geq 0\}$ と $\{Y_n; n \geq 0\}$ とする。次の条件

- 1) $X_0 \preceq_{\text{st}} Y_0$
- 2) P か R のうち少なくとも一つは単調

$$3) \quad c(P) \leq c(R)$$

が成立するならば、すべての $n \geq 0$ について、 $X_n \preceq_{\text{st}} Y_n$ である。

定理 1 により、必ずしも単調とはいえない確率行列 P に対して、その上界を与える確率行列 R は次の不等式を満たす必要があることがわかる。

$$\begin{aligned} c(R_{1,j}) &\geq c(P_{1,j}) & \forall j \in \mathcal{S}, \\ c(R_{i+1,j}) &\geq c(R_{i,j}) \vee c(P_{i+1,j}) & \forall i, j \in \mathcal{S}. \end{aligned} \quad (1)$$

ただし、 $a \vee b = \max(a, b)$ とする。ここで得られた $c(R)$ に定義 5 の逆関数 c^{-1} を使用することで、確率行列の P に対する確率的な上界 R を設計することができる。最も緊密な上界を得るためには、(1) 式で等式が成り立つ R を求めれば良い。これは Vincent のアルゴリズムとして知られている [2]。しかし、Vincent のアルゴリズムでは P から R を計算した時に P では正であった行列の要素が R では 0 になってしまう可能性があることが指摘されている。このため、Vincent のアルゴリズムは既約な P に対して、必ずしも既約な R を出力するとは限らないことになる。そのため文献 [1] では既約性という部分に着目し、確率行列 P の条件に既約性を加え、導出される確率行列 R でも既約性を保持するアルゴリズムを提案している。そのアルゴリズムを IMSUB とし、Algorithm 1 に示す。

Algorithm 1 IMSUB

```

1: for  $l = 1, 2, \dots, n$  do
2:    $r_{1,l} = p_{1,l}$ 
3: end for
4: for  $i = 2, 3, \dots, n$  do
5:    $r_{i,n} = r_{i-1,n} \vee p_{i,n}$ 
6: end for
7: for  $l = n-1, n-2, \dots, 1$  do
8:   for  $i = 2, 3, \dots, n$  do
9:      $c(r_{i,l}) = c(r_{i-1,l}) \vee c(p_{i,l})$ 
10:    if  $(c(r_{i,l}) = c(r_{i,l+1}))$  and  $(c(r_{i,l+1}) < 1)$  and  $((c(p_{i,l}) > c(p_{i,l+1}))$  or  $(i = l+1))$  then
11:       $c(r_{i,l}) = \epsilon \times (1 - c(r_{i,l+1})) + c(r_{i,l+1})$ 
12:    end if
13:   end for
14: end for

```

アルゴリズム IMSUB では式 (1) に加えて 10 行目に条件式を追加している。この条件が満たされる場合、既約性を保証するため、11 行目に示したように任意の正数 ϵ を導入して修正する。アルゴリズム IMSUB では次の性質が成立することが文献 [1] で示されている。

命題 1 (文献 [1] の Theorem 2) P を既約な確率行列とする。 $P[1, 1] > 0$ で、行列 P の下三角行列の各行に少なくとも 1 つの正の要素が含まれているならば、またその時に限りアルゴリズム IMSUB によって計算された行列 R は既約である。

2.3. Lumpable な上界を求めるアルゴリズム

この節ではアルゴリズム IMSUB が出力する上界を与える確率行列が Lumpable と呼ばれる性質をも有する場合について触れる。以下では、確率行列が Lumpable であるという性質を定義する。

定義 6 $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \dots \cup \mathcal{S}_K$ かつ $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset, i \neq j$ とする。このとき、 $A_K = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ を $\mathcal{S}$ の分割と呼ぶ。さらに、 $\mathcal{S}_i$ をマクロ状態 i と呼ぶ。

定義 7 P を確率行列とする。任意の $\mathcal{S}_i, \mathcal{S}_j \in A_K$ に対して、

$$\sum_{k \in \mathcal{S}_j} P[m, k] = \sum_{k \in \mathcal{S}_j} P[n, k], \quad \forall m, n \in \mathcal{S}_i$$

が成立するとき、 P は分割 A_K について Lumpable であるという.

既約であり、かつ定義 7 で示した Lumpable な確率行列で、確率的な上界を与えるアルゴリズムを Algorithm 2 に示す. そのアルゴリズムを LIMSUB と呼ぶことにする [1].

Algorithm 2 LIMSUB

```

1: for  $k = K, K-1, \dots, 1$  do
2:   for  $l = e(k), \dots, b(k)$  do
3:     for  $i = 1, 2, \dots, n$  do
4:        $c(r_{i,l}) = c(r_{i-1,l}) \vee c(p_{i,l})$ 
5:       if  $(c(r_{i,l}) = c(r_{i,l+1}))$  and  $(c(r_{i,l+1}) < 1)$  and  $((c(p_{i,l}) > c(p_{i,l+1})$  or  $(i = l+1))$  then
6:          $c(r_{i,l}) = \epsilon \times (1 - c(r_{i,l+1})) + c(r_{i,l+1})$ 
7:       end if
8:     end for
9:   for  $y = 1, 2, \dots, K$  do
10:     $tmp = \sum_{j=b(k)}^{e(k)} r_{e(y),j}$ 
11:    for  $i = b(y), \dots, e(y) - 1$  do
12:       $c(r_{i,b(k)}) = tmp - \sum_{j=b(k)+1}^{e(k)} r_{i,j} + c(r_{i,b(k)+1})$ 
13:    end for
14:  end for
15: end for
16: end for

```

ただし、 $b(k) = \min\{s \mid s \in \mathcal{S}\}$ であり、 $e(k) = \max\{s \mid s \in \mathcal{S}\}$ である. アルゴリズム LIMSUB では次の性質が成立することが文献 [1] で示されている.

命題 2 (文献 [1] の Theorem 3) P を既約な確率行列とし、分割 A_K を $A_K = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ とする. $P[1, 1] \neq 0$ であり、部分空間 $\mathcal{S}_1$ に対応するブロック行列 $P_{1,1}$ の下三角行列の各行に少なくとも 1 つの正の要素が含まれ、且つ全ての状態 i に対して、 $i \in \mathcal{S}_x$ と $j \in \mathcal{S}_y$ と $y < x$ となるような状態 j への遷移が存在する場合、アルゴリズム LIMSUB は分割 A_K に応じた Lumpable な既約確率行列の上界を計算する.

例 1 以下の 2 つの行列は P が元の確率行列、 R がアルゴリズム LIMSUB を使用した結果得られる確率行列である. ただし、分割を $A_2 = \{\{1, 2\}, \{3, 4, 5\}\}$ とする.

$$P = \left(\begin{array}{cc|ccc} 0.1 & 0.4 & 0.1 & 0.2 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.4 & 0.2 \\ \hline 0.1 & 0.2 & 0.1 & 0.4 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.5 \\ 0 & 0.1 & 0.2 & 0.2 & 0.5 \end{array} \right), \quad R = \left(\begin{array}{cc|ccc} 0.1 & 0.2 & 0.3 & 0.2 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.4 & 0.2 \\ \hline 0.1 & 0 & 0.3 & 0.4 & 0.2 \\ 0.1 & 0 & 0.2 & 0.2 & 0.5 \\ 0 & 0.1 & 0.2 & 0.2 & 0.5 \end{array} \right).$$

行列 R は分割 $A_2 = \{\{1, 2\}, \{3, 4, 5\}\}$ に対して定義 7 を満たすので Lumpable である. さらに、 R は既約であり、確率順序の意味で P の上界を与えることがわかる.

3. アルゴリズム IMSUB と LIMSUB の関数定義とその課題

本章ではアルゴリズム IMSUB と LIMSUB を表現する関数を定義し、そのアルゴリズムの課題について検討する.

3.1. アルゴリズムの関数定義

ここで、2 章のアルゴリズムを定義するため、2 つの関数を定義する. ここでは確率行列 P に対して累積確率行列 Q を $Q = c(P)$ とする. また、分割 A_K を $A_K = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ とする.

定義 8 関数 ℓ_{S_x} を,

1. $m \in S_i, i = 1, \dots, K, n = b(x)$ の場合

$$\ell_{S_x}(Q)[m, n] = \max\{Q[k, b(x)] \mid k \in S_i\},$$

2. その他の場合

$$\ell_{S_x}(Q)[m, n] = Q[m, n].$$

定義 9 関数 v_{S_p} を

(I) $p = K$ のとき,

$$v_{S_p}(Q)[i, j] = v_{S_p}(Q)[i - 1, j] \vee Q[i, j], \quad i = 1, \dots, n, \quad j = e(K), \dots, b(K),$$

(II) $p = K - 1, \dots, 1$ のとき,

1. 条件 1 の場合

$$v_{S_p}(Q)[i, j] = \epsilon(1 - (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1]) + (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1], \quad i = 1, \dots, n,$$

2. 条件 2 の場合

$$v_{S_p}(Q)[i, j] = \epsilon(1 - v_{S_p}(Q)[i, j + 1]) + v_{S_p}(Q)[i, j + 1], \quad i = 1, \dots, n, \quad j = e(p) - 1, \dots, b(p),$$

3. その他の場合

$$v_{S_p}(Q)[i, j] = v_{S_p}(Q)[i - 1, j] \vee Q[i, j], \quad i = 1, \dots, n, \quad j = e(p), \dots, b(p),$$

で定める。ただし, $v_{S_p}(S)[0, j] = 0$ であり, ϵ は任意の正数とする。条件 1 と条件 2 は

条件 1: $(\{v_{S_p}(Q)[i - 1, e(p)] \vee Q[i, e(p)]\} \leq (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1] < 1)$ and

$((P[i, e(p)] > 0) \text{ or } (i = j + 1)), \quad i = 1, \dots, n,$

条件 2: $((v_{S_p}(Q)[i - 1, j] \vee Q[i, j]) \leq v_{S_p}(Q)[i, j + 1] < 1)$ and $((P[i, j] > 0) \text{ or } (i = j + 1)),$

$i = 1, \dots, n, \quad j = e(p) - 1, \dots, b(p),$

である。

3.2. アルゴリズム LIMSUB の課題

分割 A_K と命題 2 の条件を満たす既約な確率行列 P に対し, $Q_f = \ell_{S_1} \circ v_{S_1} \circ \dots \circ \ell_{S_K} \circ v_{S_K}(Q)$ によって累積確率行列 Q_f を求める。このとき, Q_f に対応する確率行列 $P_f = c^{-1}(Q_f)$ が, アルゴリズム LIMSUB が出力する確率順序の意味で上界を与える Lumpable な既約確率行列である。関数 v_{S_p} での条件 1 と条件 2 がアルゴリズム LIMSUB の 5 行目に書かれている条件であり, アルゴリズム IMSUB の 10 行目の条件と同じである。このアルゴリズムは, 上界を保証するための制約式 (1) を維持し, かつ Lumpability を保証するため分割 A_K で定まる各ブロックの行和が等しくなるように調整しつつ, 既約性を保証するために,

操作 1: 元々正の確率を有する要素は非零とする

操作 2: 下側の副対角要素を非零とする

方針のもとに設計されている。操作 1 と操作 2 を実現するために, このアルゴリズムでは条件式内で任意の正数 ϵ を導入している。しかしながら, 文献 [1] では具体的な ϵ の値については何も言及してなく, 大きな ϵ を与えてしまうと, より大きな上界が出てしまい適切でないことがある。以下の例を見てみる。

例 2 P を次の確率行列とする。

$$P = \left(\begin{array}{cc|cc|cc} 0.1 & 0.3 & 0.1 & 0.1 & 0.2 & 0.2 \\ 0.1 & 0.2 & 0.0 & 0.2 & 0.2 & 0.3 \\ \hline 0.2 & 0.5 & 0.1 & 0.1 & 0.1 & 0.0 \\ 0.1 & 0.0 & 0.1 & 0.2 & 0.4 & 0.2 \\ \hline 0.0 & 0.1 & 0.1 & 0.4 & 0.1 & 0.3 \\ 0.1 & 0.0 & 0.1 & 0.1 & 0.2 & 0.5 \end{array} \right).$$

ただし、分割を $A_3 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ とする. L_{1f} と L_{2f} をそれぞれ ϵ の値が $\epsilon = 0.2$ と $\epsilon = 0.6$ であるときに、アルゴリズム LIMSUB が出力する確率行列を縮約した行列とする. このとき L_{1f} と L_{2f} は以下のようになる.

$$L_{1f} = \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.3 & 0.6 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}, \quad L_{2f} = \begin{pmatrix} 0.08 & 0.42 & 0.5 \\ 0.08 & 0.32 & 0.6 \\ 0.08 & 0.22 & 0.7 \end{pmatrix}.$$

これを見ると $L_{1f} \preceq_{\text{st}} L_{2f}$ が確認できる. 以下ではどのような ϵ の値を選択すべきかについて考察する.

4. ϵ の上限について

4.1. 過度な上界が出力される原因

任意の正数 ϵ によって、得られる確率行列の上界が過度になる原因は 2 つある.

原因 1. アルゴリズム LIMSUB では前述の通り v_{S_i} , ℓ_{S_i} と分割 S_i 列群部を求めた後 $v_{S_{i-1}} \dots$ という順序で求めている. これにより元の行列よりも, S_i 列群の要素が大きく変わることがある. すると, $v_{S_{i-1}}$ の列 $e(i-1)$ の計算で条件 1 を満たす可能性が高くなる.

原因 2. 与えられた正数 ϵ の値が大きいくほど, 条件の不等式内にある $\max$ 関数で採用される値がより大きくなるため確率順序の意味で大きな行列が出力される.

正数 ϵ は確率行列の既約性を保つために用いられているが, それは任意の値である. 仮に ϵ を大きい値にしてしまうと, 過度な上界がでてしまう可能性がある. そこで, ϵ をどの値にすれば最も適切か, 原因 2 を適用しないための条件をより詳しく考えてみる.

4.2. ϵ の上限

関数 v_{S_p} において, ϵ を使用するのは条件 1 あるいは条件 2 が成立するときである. アルゴリズム LIMSUB に従い, マクロ状態 $p+1$ (すなわち S_{p+1}) まで処理されているものと仮定し, 行列 R を次のように定める.

$$R = \ell_{S_{p+1}} \circ v_{S_{p+1}} \circ \dots \circ \ell_{S_K} \circ v_{S_K}(Q). \quad (2)$$

補題 1 $i = N$ の時に条件 2 に当てはまることはない.

証明 まず, $i = N$ のとき $v_{S_p}(R)[N, j+1]$ の値は

$$v_{S_p}(Q)[N, j+1] = \max\{Q[i, j+1] | 1 \leq i \leq N\}$$

仮に $\max$ 値をとる行を m と置くと, 累積確率行列であることから以下のことが言える.

$$Q[m, j+1] \leq Q[m, j]$$

また, 関数 v の定義により, 以下のことが言える.

$$v_{S_p}(Q)[1, j] \leq v_{S_p}(Q)[2, j] \leq \dots \leq v_{S_p}(Q)[N, j].$$

ここから以下のことが言える.

$$v_{S_p}(Q)[m, j] \leq v_{S_p}(Q)[N-1, j].$$

故に

$$v_{S_p}(Q)[N, j+1] \leq v_{S_p}(Q)[N-1, j].$$

よって, $i = N$ の時に ϵ を使用する条件 2 に当てはまることはないと言える. $\square$

以下では $i \neq N$ と $j \in S_p, p \neq K$ に対して, 条件 2 を $v_{S_p}(Q)[i, j]$ が満たしているとする. $v_{S_p}(Q)[i, j]$ が条件 2 を満たすので, $v_{S_p}(R)[i, j]$ は

$$v_{S_p}(R)[i, j] = v_{S_p}(R)[i, j+1] + \epsilon(1 - v_{S_p}(R)[i, j+1]). \quad (3)$$

ここで, $v_{S_p}(R)[i+1, j]$ を考える. 仮定より, 条件 1 と条件 2 に当てはまることはないため,

$$v_{S_p}(R)[i+1, j] = v_{S_p}(R)[i, j] \vee R[i+1, j].$$

以下, 2 つの場合に分けて考える.

- ケース 1: $v_{S_p}(R)[i, j+1] > R[i+1, j]$ の場合

このとき, $v_{S_p}(R)[i, j] > R[i+1, j]$ であるので, $v_{S_p}(R)[i+1, j] = v_{S_p}(R)[i, j]$ である. よって, 式 (4) の右辺第 2 項の分だけ最小となる上界よりも大きくなり, かつ $v_{S_p}(R)[i+1, j]$ にも伝搬する.

- ケース 2: $v_{S_p}(R)[i, j+1] \leq R[i+1, j]$ の場合

このとき, $v_{S_p}(R)[i+1, j] = R[i+1, j]$ が成立する, すなわち i 行目で選択した ϵ が $i+1$ 行目に影響を与えないためには, $v_{S_p}(R)[i, j] \leq R[i+1, j]$ を満たす ϵ を選ぶ必要がある. $j \neq e(p), i \neq N$ としたとき, この不等式から $\epsilon \times (1 - v_{S_p}(R)[i, j+1]) + v_{S_p}(R)[i, j+1] < R[i+1, j]$ であるので

$$\epsilon < \frac{R[i+1, j] - v_{S_p}(R)[i, j+1]}{1 - v_{S_p}(R)[i, j+1]} = \frac{Q[i+1, j] - v_{S_p}(R)[i, j+1]}{1 - v_{S_p}(R)[i, j+1]} = 1 - \frac{1 - Q[i+1, j]}{1 - v_{S_p}(R)[i, j+1]}, \quad (4)$$

と評価できる. ただし, 関数処理が施されていないため $R[i+1, j] = Q[i+1, j]$ であることに注意する.

また, 条件 1 が成り立つ時も, 同様に考えることが可能である. ここで, R' を次のように定める.

$$R' = \ell_{S_{p+2}} \circ v_{S_{p+2}} \circ \cdots \circ \ell_{S_K} \circ v_{S_K}(Q).$$

このとき, $j = e(p), i \neq N$ に対して, $(\ell_{S_{p+1}} \circ v_{S_{p+1}})(R')[i, j+1] \geq R'[i+1, j]$ となっていれば上記ケース 1 に該当し, $(\ell_{S_{p+1}} \circ v_{S_{p+1}})(R')[i, j+1] \leq R'[i+1, j]$ となっていればケース 2 に該当する. ケース 2 の時は, 上と同様の考え方で ϵ の上限を定めることができ, その上限は

$$\epsilon \leq 1 - \frac{1 - Q[i+1, j]}{1 - (\ell_{S_{p+1}} \circ v_{S_{p+1}})(R')[i, j+1]},$$

と計算できる.

文献 [1] のアルゴリズムにおける ϵ は任意の正数であるが, 行列の要素すべてに対して一様に定める必要はなく, 要素ごとに異なっても良い. そこで, i 行 j 列における ϵ を $\epsilon_{i,j}$ とすると, $\epsilon_{i,j}$ の上限を次のようにまとめることができる.

- (i) $j \neq e(p), v_{S_p}(R)[i, j+1] \leq Q[i+1, j]$ の場合

$$\epsilon_{i,j} \leq 1 - \frac{1 - Q[i+1, j]}{1 - v_{S_p}(R)[i, j+1]}. \quad (5)$$

- (ii) $j = e(p), (\ell_{S_{p+1}} \circ v_{S_{p+1}})(R')[i, j+1] \leq Q[i+1, j]$ の場合

$$\epsilon_{i,j} \leq 1 - \frac{1 - Q[i+1, j]}{1 - (\ell_{S_{p+1}} \circ v_{S_{p+1}})(R')[i, j+1]}. \quad (6)$$

4.3. 既約性の条件の緩和

アルゴリズム LIMSUB において, ケース 2 が必ず成立するならば, 行列の既約性を保証するために修正が必要な要素に導入する ϵ の値が, 他の要素に伝搬することなく, Lumpable なマルコフ連鎖を得ることができる. しかし, ケース 1 があり得るため, 既約性を保つために導入する ϵ の値が他の要素に影響を及ぼす可能性がある. アルゴリズム LIMSUB が出力する行列は既約でありかつ Lumpable である. さらに, Lumpable な行列の状態空間を縮約して得られる行列も既約である. ここで Lumpable な行列を得る目的が状態空間の縮約にあるとするならば, 縮約しない元の状態空間での既約性は必ずしも求められるものではなく, 縮約した状態空間での既約性を保証することで十分である. そこで, 「状態空間を縮約した行列に対してのみ既約性を保証する」ことを求める性質としアルゴリズム LIMSUB における操作 2 を次のように変更する.

操作 2': 縮約した行列の下側の副対角要素を非零とする.

このように変更することで, アルゴリズム LIMSUB が要求する条件が緩和される. 操作 2' を実現する方法を, 以下では例に基づいて考察する.

例 3 分割 $A_3 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ に対して確率行列 P を次のようにする.

$$P = \left(\begin{array}{cc|cc|cc} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,5} & p_{1,6} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,5} & p_{2,6} \\ \hline p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} & p_{3,5} & p_{3,6} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} & p_{4,5} & p_{4,6} \\ \hline p_{5,1} & p_{5,2} & p_{5,3} & p_{5,4} & p_{5,5} & p_{5,6} \\ p_{6,1} & p_{6,2} & p_{6,3} & p_{6,4} & p_{6,5} & p_{6,6} \end{array} \right).$$

ここで、行列内の線は分割を示しているものとする. このとき、縮約した行列の下側の副対角要素を決めるブロック行列は、次の 2 つの行列 $P_{2,1}$ と $P_{3,2}$ である.

$$P_{2,1} = \begin{pmatrix} p_{3,1} & p_{3,2} \\ p_{4,1} & p_{4,2} \end{pmatrix}, \quad P_{3,2} = \begin{pmatrix} p_{5,3} & p_{5,4} \\ p_{6,3} & p_{6,4} \end{pmatrix}.$$

アルゴリズム LIMSUB では、Lumpable な行列を作成するため、定義 7 を満たさなければならない. したがって、例えば $p_{5,3} + p_{5,4}$ の値は $p_{6,3} + p_{6,4}$ に等しくなる. 縮約した状態空間での既約性を保証するためには、元の状態空間の副対角要素を非零とする条件の代わりに、これらのどちらかの値を非零とする条件が考えられる. よって、縮約した状態空間での既約性を保つためには、 $p_{6,3} + p_{6,4} = 0$ となる場合に正の数 ϵ を導入して修正すれば良い. 以上の考察を一般化し、定義 9 を次のように修正し、定義 10 とする.

定義 10 関数 v_{S_p} を以下のように定める.

(I) $p = K$ のとき,

$$v_{S_p}(Q)[i, j] = v_{S_p}(Q)[i - 1, j] \vee Q[i, j], \quad i = 1, \dots, n, \quad j = e(K), \dots, b(K),$$

(II) $p = K - 1, \dots, 1$ のとき,

1. 条件 1' の場合

$$v_{S_p}(Q)[i, j] = (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1], \quad i = 1, \dots, n,$$

2. 条件 2' の場合

$$v_{S_p}(Q)[i, j] = v_{S_p}(Q)[i, j + 1], \quad i = 1, \dots, n, \quad j = e(p - 1), \dots, b(p),$$

3. 条件 3 の場合

$$v_{S_p}(Q)[i, j] = \epsilon(1 - (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1]) + (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1], \quad i = 1, \dots, n,$$

4. 条件 4 の場合

$$v_{S_p}(Q)[i, j] = \epsilon(1 - v_{S_p}(Q)[i, j + 1]) + v_{S_p}(Q)[i, j + 1], \quad i = 1, \dots, n, \quad j = e(p - 1), \dots, b(p),$$

5. その他の場合

$$v_{S_p}(Q)[i, j] = v_{S_p}(Q)[i - 1, j] \vee Q[i, j], \quad i = 1, \dots, n, \quad j = e(p), \dots, b(p),$$

ただし、条件 1', 条件 2', 条件 3, 条件 4 は以下のようにする.

条件 1': $(\{v_{S_p}(Q)[i - 1, e(p)] \vee Q[i, e(p)]\} \leq (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1] < 1) \text{ and } (P[i, e(p)] > 0), \quad i = 1, \dots, n,$

条件 2': $((v_{S_p}(Q)[i - 1, j] \vee Q[i, j]) \leq v_{S_p}(Q)[i, j + 1] < 1) \text{ and } (P[i, j] > 0), \quad i = 1, \dots, n, \quad j = e(p) - 1, \dots, b(p),$

条件 3: $(\{v_{S_p}(Q)[i - 1, e(p)] \vee Q[i, e(p)]\} \leq (\ell_{S_{p+1}} \circ v_{S_{p+1}})(Q)[i, e(p) + 1] < 1) \text{ and } (P[i, e(p)] > 0) \text{ and } \left(1 - \frac{1 - Q[i + 1, j]}{1 - (\ell_{S_{p+1}} \circ v_{S_{p+1}})(R')[i, j + 1]} < 0\right) \text{ and } i = e(p + 1), \quad i = 1, \dots, n,$

条件 4: $((v_{S_p}(Q)[i-1, j] \vee Q[i, j]) \leq v_{S_p}(Q)[i, j+1] < 1)$ and $(P[i, j] > 0)$ and $\left(1 - \frac{1 - Q[i+1, j]}{1 - v_{S_p}(R)[i, j+1]} < 0\right)$ and $i = e(p+1), \quad i = 1, \dots, n, \quad j = e(p-1), \dots, b(p).$

条件 3 と条件 4 の場合, ϵ に正数を与える. また, 条件 1' と条件 2' の場合は $\epsilon = 0$ として計算をする. 以上の考察に基づいて, 修正したアルゴリズムを LIMSUB 改とし, Algorithm 3 に示す.

Algorithm 3 LIMSUB 改

```

1: for  $k = K, K-1, \dots, 1$  do
2:   for  $l = e(k), \dots, b(k)$  do
3:     for  $i = 1, 2, \dots, n$  do
4:        $c(r_{i,l}) = c(r_{i-1,l}) \vee c(p_{i,l})$ 
5:       if  $(c(r_{i,l}) = c(r_{i,l+1}))$  and  $(c(r_{i,l+1}) < 1)$  and  $(c(p_{i,l}) > c(p_{i,l+1}))$  and  $\left(\frac{c(r_{i+1,l}) - c(r_{i,l+1})}{1 - c(r_{i,l+1})}\right)$  and  $(l < b(K))$  and  $(m = e(k+1))$  then
6:          $c(r_{i,l}) = \epsilon \times (1 - c(r_{i,l+1})) + c(r_{i,l+1})$ 
7:       else if  $(c(r_{i,l}) = c(r_{i,l+1}))$  and  $(c(r_{i,l+1}) < 1)$  and  $(c(p_{i,l}) > c(p_{i,l+1}))$  then
8:          $c(r_{i,l}) = c(r_{i,l+1})$ 
9:       end if
10:    end for
11:    for  $y = 1, 2, \dots, K$  do
12:       $tmp = \sum_{j=b(k)}^{e(k)} r_{e(y),j}$ 
13:      for  $i = b(y), \dots, e(y) - 1$  do
14:         $c(r_{i,b(k)}) = tmp - \sum_{j=b(k)+1}^{e(k)} r_{i,j} + c(r_{i,b(k)+1})$ 
15:      end for
16:    end for
17:  end for
18: end for

```

元の LIMSUB に比べて LIMSUB 改の方が, 副対角要素の数が減り, 既約性を保証するために導入する ϵ を使う回数が減る. そのため, より緊密な上界を得られることが期待できる.

5. 実際のモデルを使った数値例

5.1. 待ち行列モデルの例 [3]

提案したアルゴリズムを検証するために, 次の待ち行列システムをモデル化したマルコフ連鎖を考える. 単一のサーバと有限容量 B をもつバッファからなる待ち行列システムを考える. 優先度の異なる 2 種類のパケットが到着すると仮定する. 簡単のため, パケットのサイズは種類に関係なく全て同一であり一定であるとする. バッファが全て占有されているときに優先度の低いパケットが到着したら損失される. 優先度の高いパケットが到着した場合, もしバッファ内に優先度の低いパケットがあるならば, 優先度の高いパケットは優先度の低いパケットを押し出してバッファを占有するが (PushOut), 優先度の低いパケットがなければ, 優先度の高いパケットは損失される. サービスを受ける順番は, 常に優先度の低いパケットが優先度の高いパケットよりも先に受けるようにスケジューリングする (Head Of Line).

1 つのパケットがサービスを受けている間に, パケットは集団で到着すると仮定する. 簡単のため, 集団のサイズが 0 の確率を p とし, 1 の確率を q とし, 2 の確率を r とし, $p + q + r = 1$ とする. 到着したパケットが優先度の低いパケットである確率を a とし, 優先度の高いパケットである確率を $b = 1 - a$ とする. n 番目のサービスを完了した直後のバッファに存在するパケット数を T_n とし, 同じように優先度の高いパケット数を H_n とすると, $\{(T_n, H_n); n \geq 0\}$ は, 状態空間 $\mathcal{S}$ が $\mathcal{S} = \{(0,0), (1,0), (1,1), \dots, (B,0), \dots, (B,B)\}$ となるマルコフ連鎖である. バッファのサイズが B であるならば, 状態の数は $(B+1)(B+2)/2$ となる. 状態空間 $\mathcal{S}$ を $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \dots \cup \mathcal{S}_B$ のように分割する. ここで, $\mathcal{S}_k = \{(t,h) \in \mathcal{S} \mid t = k\}$, $k = 0, 1, \dots, B$ である.

例 4 $B = 4$ の場合で $p = 0.3, q = 0.5, r = 0.2$ とし, $a = 0.3, b = 0.7$ とする. 分割を $A_5 = \{\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_4\}$ としてアルゴリズムを実行する. アルゴリズム LIMSUB の結果を L_{1m} とし, アルゴリズム LIMSUB に, 式 (5) と

式 (6) の条件を加えた結果 [5] を L_{2m} とし、アルゴリズム LIMSUB 改の結果を L_{3m} とする。ただし、いずれのアルゴリズムにおいても ϵ を使用する場合は $\epsilon = 0.1$ とする。その結果を以下に示す。

$$L_{1m} = \begin{pmatrix} 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0.27 & 0.53 & 0.2 & 0 \\ 0 & 0 & 0.27 & 0.53 & 0.2 \\ 0 & 0 & 0 & 0.27 & 0.73 \end{pmatrix}, \quad L_{2m} = \begin{pmatrix} 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0.3 & 0.5 & 0.2 & 0 \\ 0 & 0 & 0.3 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0.3 & 0.7 \end{pmatrix},$$

$$L_{3m} = \begin{pmatrix} 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0.3 & 0.5 & 0.2 & 0 \\ 0 & 0 & 0.3 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0.3 & 0.7 \end{pmatrix}.$$

ここでは $L_{2m} \preceq_{\text{st}} L_{1m}, L_{3m} \preceq_{\text{st}} L_{1m}$ となっていることが確認できる。しかし、この例では L_{2m}, L_{3m} が同じである。

5.2. 縮約した行列のみ既約性を保証した例

この節では 4.3 節で示した「状態空間を縮約した行列に対してのみ既約性を保証」した場合の例を示す。

P を確率行列とし、分割 $A_3 = \{\{1, 2\}, \{3\}, \{4, 5\}\}$ とする。行列内に書かれている線はその分割を示したものとする。

$$P = \left(\begin{array}{cc|c|cc} 0.3 & 0.1 & 0.1 & 0.0 & 0.5 \\ 0.2 & 0.0 & 0.0 & 0.1 & 0.7 \\ \hline 0.3 & 0.3 & 0.1 & 0.0 & 0.3 \\ 0.2 & 0.7 & 0.0 & 0.1 & 0.0 \\ 0.4 & 0.4 & 0.1 & 0.0 & 0.1 \end{array} \right). \quad (7)$$

式 (7) にアルゴリズム LIMSUB を適用する。アルゴリズム LIMSUB の結果を L_1 とし、アルゴリズム LIMSUB に、式 (5) と式 (6) の条件を加えた結果 [5] を L_2 とし、アルゴリズム LIMSUB 改の結果を L_3 とする。ただし、いずれのアルゴリズムにおいても ϵ を使用する場合は $\epsilon = 0.1$ とする。その結果を以下に示す。

$$L_1 = \begin{pmatrix} 0.18 & 0.02 & 0.8 \\ 0.18 & 0.02 & 0.8 \\ 0.18 & 0.02 & 0.8 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 0.2 & 0.0 & 0.8 \\ 0.18 & 0.02 & 0.8 \\ 0.18 & 0.02 & 0.8 \end{pmatrix}, \quad L_3 = \begin{pmatrix} 0.2 & 0.0 & 0.8 \\ 0.2 & 0.0 & 0.8 \\ 0.18 & 0.02 & 0.8 \end{pmatrix}.$$

これを見ると、 $L_2 \preceq_{\text{st}} L_1$ であり、さらに $L_3 \preceq_{\text{st}} L_2$ となっている。4.3 節の条件を適用したときの方がより緊密な上界を得ることができていることが確認できる。

6. まとめ

これまでの考察により、条件 1, 条件 2 が成立する場合、 $j \neq e(p), i \neq N$ としたとき、 $v_{S_p}(R)[i, j+1] \leq Q[i+1, j]$ ならば、 ϵ の上限を定めることが可能であることを示した。また、より緊密な上界を得るために、縮約した行列のみを既約にするように条件を緩めたアルゴリズムを提案した。このアルゴリズムは既約を保ち且つ、なるべく緊密な上界を定めることができると期待され、サービスシステムのモデリングなどの研究において有用であると考えられる。

参考文献

- [1] J.M. Fourneau, M. Lecoq and F. Quessette, Algorithms for an irreducible and lumpable strong stochastic bound, *Linear Algebra and its Applications*, **386** (2004), 167–185.
- [2] O. Abu-Amsha, J.-M. Vincent, An algorithm to bound functionals of Markov chains with large state space, 4th INFORMS Conference on Telecommunications, Boca Raton, Florida, (1998).
- [3] J.M. Fourneau and N. Pekergin, An Algorithmic Approach to Stochastic Bounds, M.C. Calzarossa and S. Tucci (Eds.): *Performance 2002*, LNCS 2459 (2002), 64–88.
- [4] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, Wiley, 1983.
- [5] 石川真也, 河西憲一, 上界を与えるマルコフ連鎖を求めるアルゴリズムの考察, 日本オペレーションズ・リサーチ学会 2016 年秋季研究発表会アブストラクト集, (2016), 250–251.

レベル依存する M/G/1 型マルコフ連鎖における条件付き定常分布の数値計算アルゴリズム

木村 雅俊 滝根 哲哉
大阪大学 工学研究科 電気電子情報工学専攻

概要： 本稿では、任意のレベル依存する M/G/1 型連続時間マルコフ連鎖における、レベルが N 以下であるという条件付きの定常分布に対する数値計算アルゴリズムを提案する。このアルゴリズムは、先行研究において、レベルの降下方向への遷移を表す行列が正則であるという仮定の下で提案されたものとほぼ同じであり、計算結果が含む誤差の評価が可能なものとなっている。 N が十分大きい場合、この条件付き定常分布は、定常分布の近似として捉えることもできるため、任意のレベル依存する M/G/1 型マルコフ連鎖における定常分布の数値計算アルゴリズムとしても利用可能である。

1. はじめに

本稿では、既約で正再帰的な連続時間 2 変数マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ を考察する。ただし、レベルと呼ばれる変数 $L(t)$ は非負整数値を取り、相と呼ばれる変数 $J(t)$ は、 $L(t) = k$ ($k = 0, 1, 2, \dots$) が与えられたとき、有限集合 $\mathcal{M}_k = \{1, 2, \dots, M_k\}$ の要素から値を取る。以下では、連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の状態空間 $\{(k, j); k = 0, 1, \dots, j \in \mathcal{M}_k\}$ を互いに素な部分集合 $\{(k, j); j \in \mathcal{M}_k\}$ に分割し、各部分集合 $\{(k, j); j \in \mathcal{M}_k\}$ ($k = 0, 1, \dots$) をレベル k と呼ぶ。よって、辞書式順序に従って状態を並べると、連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の無限小生成作用素 Q は以下の形をとる。

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & Q_{0,3} & \cdots \\ Q_{1,-1} & Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,-2} & Q_{2,-1} & Q_{2,0} & Q_{2,1} & \cdots \\ Q_{3,-3} & Q_{3,-2} & Q_{3,-1} & Q_{3,0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (1)$$

ただし、 $Q_{k,l}$ ($k = 0, 1, \dots, l = -k, -k+1, \dots$) は $M_k \times M_{k+l}$ 行列であり、レベル k からレベル $k+l$ への遷移を表している。無限小生成作用素が式 (1) で示したブロック構造をもつ連続時間 2 変数マルコフ連鎖は G/G/1 型と呼ばれる。特に、レベルの降下方向への遷移に飛び越しが無い、すなわち

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & Q_{0,3} & \cdots \\ Q_{1,-1} & Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots \\ O & Q_{2,-1} & Q_{2,0} & Q_{2,1} & \cdots \\ O & O & Q_{3,-1} & Q_{3,0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

である場合、連続時間 2 変数マルコフ連鎖はレベル依存する M/G/1 型連続時間マルコフ連鎖と呼ばれる。以下では、連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ は定常であると仮定する。

レベル依存する G/G/1 型連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の定常分布を $\{q_k; k = 0, 1, \dots\}$ とする。ただし、 q_k ($k = 0, 1, \dots$) は $1 \times M_k$ ベクトルであり、 j 番目 ($j \in \mathcal{M}_k$) の要素は定常確率 $\Pr(L(0) = k, J(0) = j)$ を表す。なお、遷移確率行列 P をもつ G/G/1 型離散時間マルコフ連鎖の定常分布 q に興味がある場合は、 $q = qP$ が成立するので、無限小生成作用素 $Q := P - I$ をもつ G/G/1 型連続時間マルコフ連鎖の定常分布を考察すればよい。

レベル依存する G/G/1 型連続時間マルコフ連鎖において、レベル n ($n = 0, 1, \dots$) 内の各状態からレベル k ($k > n$) 以上の状態への初到達時間内に、レベル $k-1$ 以下の各状態に滞在する平均延べ時間の比は、 $k \rightarrow \infty$ の極限において（適当な条件の下で）定常分布に収束することが知られている [4]。

$$\lim_{k \rightarrow \infty} \text{diag}^{-1}((-k_+ Q)^{-1} e)(-k_+ Q)^{-1} = e(q_0 \ q_1 \ \dots) \quad (2)$$

ただし、 $k_+ Q$ ($k = 1, 2, \dots$) は無限小生成作用素 Q の $(\sum_{l=0}^{k-1} M_l) \times (\sum_{l=0}^{k-1} M_l)$ の北西角を表す。また、任意のベクトル x に対して $\text{diag}(x)$ は $[\text{diag}(x)]_{i,i} = [x]_i$ なる対角行列であり、 e は適当な次元をもつ全ての要素

が 1 である列ベクトルである。文献 [4] では、この結果を基に、レベル依存する M/G/1 型連続時間マルコフ連鎖の定常分布に対する数値計算アルゴリズムが提案されている。

本稿では、レベル k ($k = 1, 2, \dots$) への再帰時間内における、特定の下位レベル n ($n = 0, 1, \dots, k-1$) の各状態に滞在する平均延べ時間に注目する。

定義 1. レベル依存する G/G/1 型連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ に対して、 $M_k \times M_n$ 行列 $U_{k,n}$ ($k = 1, 2, \dots, n = 0, 1, \dots, k-1$) を導入する。ただし、 $U_{k,n}$ の (i, j) 番目 ($i \in \mathcal{M}_k, j \in \mathcal{M}_n$) の要素は、初期状態 (k, i) から開始したレベル k への再帰時間内に状態 (n, j) に滞在する平均延べ時間を、初期状態 (k, i) の平均滞在時間で割った値で与えられる。

$U_{k,n}$ の定義から次式が成立する。

$$q_n = q_k U_{k,n}, \quad k = 1, 2, \dots, n = 0, 1, 2, \dots, k-1 \quad (3)$$

ここで、 $u_{k,n}(i)$ ($i \in \mathcal{M}_k$) を $U_{k,n}$ の i 番目の行ベクトルとする。一般に、既約で正再帰的なレベル依存する G/G/1 型連続時間マルコフ連鎖において、全ての $i \in \mathcal{M}_k$ に対して $u_{k,n}(i)e > 0$ となることは保証されない。一方、もし、全ての i ($i \in \mathcal{M}_k$) に対して $u_{k,n}(i)e = 0$ であるような k, n の組が存在すると仮定すると、これはレベル k からレベル n に到達できないことを意味し、既約であるという仮定に矛盾する。よって、既約であるならば、任意の k, n の組 ($k = 1, 2, \dots, n = 0, 1, \dots, k-1$) に対して $u_{k,n}(i)e > 0$ となる i ($i \in \mathcal{M}_k$) が少なくとも一つ存在する。一方、各レベル内の相のラベルを任意に付け替えても、G/G/1 型、M/G/1 型といったレベル間の遷移に関する構造は保存される。そこで本節ならびに次節では、一般性を失うことなく、非負整数 n ($n = 0, 1, 2, \dots$) が与えられたとき、全ての k ($k = n+1, n+2, \dots$) に対して $u_{k,n}(1)e > 0$ が成立すると仮定する。

$u_{k,n}(i)e > 0$ なる $i \in \mathcal{M}_k$ に対して、正規化されたベクトル $\bar{u}_{k,n}(i)$ を次式で定義する。

$$\bar{u}_{k,n}(i) = \frac{u_{k,n}(i)}{u_{k,n}(i)e} \quad \text{if } u_{k,n}(i)e > 0$$

次節では次の定理を証明する。

定理 2. レベル依存する G/G/1 型連続時間マルコフ連鎖において

$$\lim_{k \rightarrow \infty} \bar{u}_{k,n}(1) = \bar{q}_n \quad (4)$$

が成立する。ただし $\bar{q}_n$ はレベルが n であるという条件下での相の条件付き定常確率ベクトルを表す。

$$\bar{q}_n = q_n / (q_n e)$$

上で述べたように、各レベル内の相のラベルを任意に張り替えても、G/G/1 型という構造は保存される。また、各レベルの相は有限集合の要素から値を取る。そこで、 $u_{k,n}(i)e > 0$ ($i \in \mathcal{M}_k$) なる任意の i のラベルを 1 に張り直し、定理 2 を適用すると以下の系を得る。

系 3. レベル依存する G/G/1 型連続時間マルコフ連鎖において、任意の非負整数 n と任意の正数 ϵ に対して、ある自然数 $K := K(n, \epsilon)$ が存在し、全ての自然数 $k = K, K+1, \dots$ に対して

$$\|\bar{u}_{k,n}(i) - \bar{q}_n\| < \epsilon \quad \text{if } u_{k,n}(i)e > 0, \quad i \in \mathcal{M}_k$$

が成立する。ただし $\|\cdot\|$ はベクトルノルムを表す。

さらに、ある非負整数 N に対して $\bar{q}_N$ が求まれば、式 (3) を用いて、レベルが N 以下であるという条件付きの定常分布 $\{q_k(N); k = 0, 1, \dots, N\}$ を得ることができる。

$$q_k(N) = q_k \left/ \left(\sum_{\ell=0}^N q_\ell e \right) \right. = \bar{q}_N U_{N,k} \left/ \left(\sum_{\ell=0}^N \bar{q}_N U_{N,\ell} e \right) \right., \quad k = 0, 1, \dots, N \quad (5)$$

すなわち、条件付き定常分布 $\{q_k(N); k = 0, 1, \dots, N\}$ の計算は $U_{k,n}$ ($k = 1, 2, \dots, n = 0, 1, \dots, k-1$) の計算に帰着させることができる。

残念ながら、レベル依存する G/G/1 型マルコフ連鎖における $U_{k,n}$ の計算は、全レベル間の遷移構造に関する情報が必要となるため、一般には困難である。しかし、G/G/1 型の特別な場合である、レベル依存する M/G/1 型連続時間マルコフ連鎖では、式 (6) で示すように、 $U_{k,n}$ を効率的に計算することができる [6]。

$$U_{k,n} = Z_{k-1} Z_{k-2} \cdots Z_n, \quad n = 0, 1, 2, \dots, k-1 \quad (6)$$

ただし,

$$\mathbf{Z}_k = \mathbf{Q}_{k+1,-1}(-\mathbf{T}_k)^{-1}, \quad k = 0, 1, \dots \quad (7)$$

であり, $\mathbf{T}_k$ ($k = 0, 1, \dots$) は次式に従い順次計算される.

$$\mathbf{T}_k = \mathbf{Q}_{k,0} + \sum_{n=0}^{k-1} \mathbf{Z}_{k-1} \mathbf{Z}_{k-2} \cdots \mathbf{Z}_n \mathbf{Q}_{n,k-n}, \quad k = 0, 1, \dots \quad (8)$$

文献 [6] では, レベル依存する M/G/1 型連続時間マルコフ連鎖を対象に,

仮定 4. あるレベル N より上位の全てのレベルにおいて, レベルの降下方向への遷移を表す行列 $\mathbf{Q}_{k,-1}$ ($k > N$) が正則な $M \times M$ 正方行列である,

という仮定の下で, $k > n \geq N$ なる任意の非負整数 k, n の組に対して, $\mathcal{M} = \{1, 2, \dots, M\}$ としたとき

(i) $\mathbf{u}_{k,n}(i)\mathbf{e} > 0$ ($i \in \mathcal{M}$)

(ii) $\bar{\mathbf{u}}_{k,n}(i)$ ($i \in \mathcal{M}$) は一次独立,

(iii) $\bar{\mathbf{u}}_{k,n}(i)$ ($i \in \mathcal{M}$) が張る $(M-1)$ -単体 $\mathcal{P}_{k,n}$ の相対的内部を $\text{ri } \mathcal{P}_{k,n}$ としたとき, $\mathcal{P}_{k+1,n} \subseteq \mathcal{P}_{k,n}$ かつ $\bar{\mathbf{q}}_n \in \bigcap_{k=n+1}^{\infty} \text{ri } \mathcal{P}_{k,n}$

となることを導いている. よって, 仮定 4 の下では, 式 (4) が $\bar{\mathbf{u}}_{k,n}(1)$ のみならず, 全ての $\bar{\mathbf{u}}_{k,n}(i)$ ($i \in \mathcal{M}$) に対して成立する. 文献 [6] では, この結果を利用して, 条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ の数値計算アルゴリズムを提案している.

本稿の 3 節では, 系 3 に基づいて, 任意のレベル依存する M/G/1 型連続時間マルコフ連鎖における条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ の数値計算アルゴリズムを提案する. これは, 仮定 4 の下で, 文献 [6] で構築された数値計算アルゴリズムとほぼ同じであるが, 停止条件に関しては新たな考察が必要となる. 仮定 4 の下では, (iii) より $(M-1)$ -単体 $\mathcal{P}_{k,n}$ が包含関係を保ったまま縮退していくため, 全ての $\bar{\mathbf{u}}_{k,n}(i)$ ($i = 1, 2, \dots, M$) が近似的に等しいと判断された時点で $\mathbf{U}_{k,n}$ の計算を停止すればよい. 本稿では割愛するが, 仮定 4 が成立しない場合でも, $\bar{\mathbf{u}}_{k,n}(i)$ ($i \in \mathcal{M}_k$) が定義可能な複数の $\bar{\mathbf{u}}_{k,n}(i)$ が存在した場合, それらで張られる凸多面体を $\mathcal{P}_{k,n}$ と記すと, $\mathcal{P}_{k+1,n} \subseteq \mathcal{P}_{k,n}$ を示すことができる. さらに, $\bar{\mathbf{q}}_n$ は $\mathcal{P}_{k,n}$ を張る頂点の凸結合で表現されるため, 仮定 4 が成立しない場合でも, 同様の停止条件が利用可能である.

これまで, レベル依存する M/G/1 型マルコフ連鎖における定常分布の数値計算アルゴリズムは, レベル依存する基本行列 (G -行列) を用いるものが主であった [2, 3]. 文献 [3] では, 十分大きな K に対して, レベル K 以上の構造がレベルに依存しないモデルに置き換え, 近似解を計算する手法が提案されている. また, 文献 [2] では, レベルが十分大きい場合, 遷移構造が近似的にレベルに依存しないと見なせるモデルを対象とした手法が提案されている. 一方, 本稿で提案する数値計算アルゴリズムは, 仮定 4 の下で文献 [6] において提案されたものとほぼ同一であり, レベルに関する降下方向の率行列 (R -行列) を用いる. なお, 本稿の結果を用いれば, 任意のレベル依存する M/G/1 型マルコフ連鎖における条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ の近似解の精度を評価することが可能である. また, 系 10 で示すように, ある自然数 K ($K > N$) に対して, 降下方向の遷移を表す行列 $\mathbf{Q}_{K,-1}$ のランクが 1 である場合, 条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ の厳密解を計算することができる.

本稿で議論する条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ は N を十分大きく取ることで, 定常分布の良い近似を与える [6]. しかし, レベル依存する M/G/1 型連続時間マルコフ連鎖において, どの程度の N であれば良い近似となるか, という点について明確な基準は得られていない. 文献 [1] では, レベル依存する M/G/1 型マルコフ連鎖の特別な場合であるレベル依存する準出生死滅過程を対象に, N の選択について議論されている. また, 文献 [4] で提案されている, 式 (2) に基づくレベル依存する M/G/1 型マルコフ連鎖に対する数値計算アルゴリズムは, N を事前に選択すること無く定常分布の数値計算が可能な手法である.

なお, レベル依存する M/G/1 型連続時間マルコフ連鎖に限定した場合, 文献 [4] で提案されている数値計算アルゴリズムの基礎となっている式 (2) と本稿で提案する数値計算アルゴリズムの基礎となっている定理 2 の間には密接な関係がある. まず, 式 (2) で現れる $(-_{k+}\mathbf{Q})^{-1}$ の $(k-1, n)$ ブロック行列 ($n = 0, 1, \dots, k-1$) における (i, j) 番目 ($i \in \mathcal{M}_{k-1}, j \in \mathcal{M}_n$) の要素は, 状態 $(k-1, i)$ から開始されるレベル k 以上の状態への初到達時間内において, 状態 (n, j) に滞在する平均延べ時間を与えることに注意する. よって, $(-_{k+}\mathbf{Q})^{-1}$ の $(k-1, n)$ ブロック行列 ($n = 0, 1, \dots, k-1$) を $_{k+}\mathbf{B}_{k-1,n}$ とすると, レベル依存する M/G/1 型連続時間マルコフ連鎖の場合には

$$\mathbf{U}_{k,n} = \mathbf{Q}_{k,-1} \cdot _{k+}\mathbf{B}_{k-1,n}$$

が成立する. この結果, M/G/1 型に限定すれば, 定理 2 は文献 [4] で示された式 (2) を用いて導くことができる. レベル依存する M/G/1 型マルコフ連鎖に対する文献 [4] の数値計算アルゴリズムは $(-_{k+}\mathbf{Q})^{-1}$ を扱っ

ており、本稿で考察する数値計算アルゴリズムでは $U_{k,n}$ を扱っているため、前者における停止条件に関しても、後者と同様の停止条件が利用可能と思われる。

以下、2 節で定理 2 を証明した後、一般のレベル依存する M/G/1 型連続時間マルコフ連鎖における条件付き定常分布の数値計算アルゴリズムを 3 節で提案する。最後に 4 節で本稿で得られた結果をまとめる。

2. 定理 2 の証明

証明の見通しを良くするため、G/G/1 型連続時間マルコフ連鎖 $\{(L(t), J(t)); t \geq 0\}$ の状態遷移直後に着目することで得られる隠れマルコフ連鎖 $\{(L_\tau, J_\tau); \tau = 0, 1, \dots\}$ を考える。このとき、隠れマルコフ連鎖 $\{(L_\tau, J_\tau); \tau = 0, 1, \dots\}$ は G/G/1 型となり、その遷移確率行列 P は次式で与えられる。

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} & P_{0,2} & P_{0,3} & P_{0,4} & \cdots \\ P_{1,-1} & P_{1,0} & P_{1,1} & P_{1,2} & P_{1,3} & \cdots \\ P_{2,-2} & P_{2,-1} & P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ P_{3,-3} & P_{3,-2} & P_{3,-1} & P_{3,0} & P_{3,1} & \cdots \\ P_{4,-4} & P_{4,-3} & P_{4,-2} & P_{4,-1} & P_{4,0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

ただし、 i 番目 ($i \in \mathcal{M}_k, k = 0, 1, \dots$) の要素が $-[Q_{k,0}]_{i,i}$ で与えられる $1 \times M_k$ ベクトル μ_k を用いて $P_{k,\ell}$ ($k = 0, 1, \dots, \ell = -k, -k+1, \dots$) は次式で与えられる。

$$P_{k,\ell} = \begin{cases} I + \text{diag}^{-1}(\mu_k) Q_{k,0}, & \ell = 0 \\ \text{diag}^{-1}(\mu_k) Q_{k,\ell}, & \ell = -k, -k+1, \dots, -1, 1, 2, \dots \end{cases}$$

定義より、隠れマルコフ連鎖 $\{(L_\tau, J_\tau); \tau = 0, 1, \dots\}$ も既約で正再帰的である。以下では、隠れマルコフ連鎖 $\{(L_\tau, J_\tau); \tau = 0, 1, \dots\}$ は定常であると仮定し、その定常分布を $\{p_k; k = 0, 1, \dots\}$ と記す。ただし、 p_k ($k = 0, 1, \dots$) は $1 \times M_k$ ベクトルであり、 j 番目 ($j \in \mathcal{M}_k$) の要素は $\Pr(L_0 = k, J_0 = j)$ で与えられる。定義より、 q_k ($k = 0, 1, \dots$) は p_k を用いて次式のように表すことができる。

$$q_k = p_k \text{diag}^{-1}(\mu_k) \Big/ \left(\sum_{\ell=0}^{\infty} p_\ell \text{diag}^{-1}(\mu_\ell) e \right) \quad (9)$$

次に、非負整数 n ($n = 0, 1, \dots$) に対して、 ${}_k f_{k,n}(i)$ ($k = n+1, n+2, \dots, i \in \mathcal{M}_k$) を $1 \times M_n$ ベクトルとし、その j 番目 ($j \in \mathcal{M}_n$) の要素は、 $(L_0, J_0) = (k, i)$ から開始したレベル n ($n = 0, 1, \dots, k-1$) への初到達時間がレベル k を訪れることなく状態 (n, j) で終了する確率を表すとする。さらに、 ${}_k F_{n,n}$ ($k = n+1, n+2, \dots$) を $M_n \times M_n$ 行列とし、その (i, j) 番目 ($i, j \in \mathcal{M}_n$) の要素は状態 (n, i) から開始したレベル n への再帰時間が、レベル k に遷移することなく状態 (n, j) で終了する確率を表すとする。これらを用いて $u_{k,n}(i)$ は次式のように書き換えることができる。

$$u_{k,n}(i) = \mu_k(i) {}_k f_{k,n}(i) (I - {}_k F_{n,n})^{-1} \text{diag}^{-1}(\mu_n) \quad (10)$$

なお、

$${}_k f_{k,n}(i) (I - {}_k F_{n,n})^{-1} = {}_k f_{k,n}(i) \sum_{l=0}^{\infty} {}_k F_{n,n}^l$$

の j 番目 ($j \in \mathcal{M}_k$) の要素は状態 (k, i) から開始したレベル k への再帰時間の間に状態 (n, j) を訪問した平均回数を表すことに注意する。

定義より ${}_k f_{k,n}(i) e > 0$ ($i \in \mathcal{M}_k$) は $u_{k,n}(i) e > 0$ と等価であることに注意する。よって $\bar{u}_{k,n}(i)$ は式 (10) を用いて以下のように書き換えることができる。

$$\bar{u}_{k,n}(i) = \frac{{}_k f_{k,n}(i) (I - {}_k F_{n,n})^{-1} \text{diag}^{-1}(\mu_n)}{{}_k f_{k,n}(i) (I - {}_k F_{n,n})^{-1} \text{diag}^{-1}(\mu_n) e} \quad \text{if } {}_k f_{k,n}(i) e > 0 \quad (11)$$

$F_{n,n}$ を $M_n \times M_n$ 行列とし、その (i, j) 番目 ($i, j \in \mathcal{M}_n$) の要素は状態 (n, i) から開始したレベル n への再帰時間が状態 (n, j) で終了する確率を表すとする。隠れマルコフ連鎖の既約性ならびに正再帰性から $F_{n,n}$ は既約な確率行列であることに注意する。

補題 5. 任意の n ($n = 0, 1, \dots$) について次式が成立する.

$$\lim_{k \rightarrow \infty} {}_k\mathbf{F}_{n,n} = \mathbf{F}_{n,n}$$

補題 5 の証明は付録 A で与える.

次に ${}_k\bar{\mathbf{F}}_n(j)$ ($k = n+1, n+2, \dots, j \in \mathcal{M}_n$) を次式で定義する.

$${}_k\bar{\mathbf{F}}_n(j) = {}_k\mathbf{F}_{n,n} + (\mathbf{F}_{n,n} - {}_k\mathbf{F}_{n,n})\mathbf{e}\mathbf{e}_j^T = {}_k\mathbf{F}_{n,n} + (\mathbf{I} - {}_k\mathbf{F}_{n,n})\mathbf{e}\mathbf{e}_j^T \quad (12)$$

ただし, $\mathbf{e}_j$ ($j \in \mathcal{M}_n$) は j 番目の要素が 1 である単位列ベクトルとし, $\cdot^T$ は転置を表す. 式 (12) で定義された ${}_k\bar{\mathbf{F}}_n(j)$ は $M_n \times M_n$ 確率行列であることに注意する. そこで ${}_k\bar{\mathbf{p}}_n(j)$ ($k = n+1, n+2, \dots, j \in \mathcal{M}_n$) を遷移確率行列 ${}_k\bar{\mathbf{F}}_n(j)$ の不変確率ベクトルであると定義する.

$${}_k\bar{\mathbf{p}}_n(j) = {}_k\bar{\mathbf{p}}_n(j){}_k\bar{\mathbf{F}}_n(j), \quad {}_k\bar{\mathbf{p}}_n(j)\mathbf{e} = 1 \quad (13)$$

補題 6. 任意の n ($n = 0, 1, \dots$) に対して次式が成立する.

$$\lim_{k \rightarrow \infty} {}_k\bar{\mathbf{p}}_n(j) = \bar{\mathbf{p}}_n \quad j \in \mathcal{M}_n, \quad (14)$$

ただし, $\bar{\mathbf{p}}_n = \mathbf{p}_n/(\mathbf{p}_n\mathbf{e})$ である.

補題 6 の証明は付録 B に与える.

式 (11) に現れる $(\mathbf{I} - {}_k\mathbf{F}_{n,n})^{-1}$ について考察する. 式 (12) を式 (13) に代入し, 整理すると次式を得る.

$${}_k\bar{\mathbf{p}}_n(j)(\mathbf{I} - {}_k\mathbf{F}_{n,n}) = {}_k\bar{\mathbf{p}}_n(j)(\mathbf{I} - {}_k\mathbf{F}_{n,n})\mathbf{e}\mathbf{e}_j^T, \quad j \in \mathcal{M}_n \quad (15)$$

ここで ${}_k\bar{\mathbf{\Pi}}_n$ ($k = n+1, n+2, \dots$) を次式で与えられる $M_n \times M_n$ 行列とする.

$${}_k\bar{\mathbf{\Pi}}_n = \begin{pmatrix} {}_k\bar{\mathbf{p}}_n(1) \\ {}_k\bar{\mathbf{p}}_n(2) \\ \vdots \\ {}_k\bar{\mathbf{p}}_n(M_n) \end{pmatrix}, \quad k = n+1, n+2, \dots$$

このとき, 式 (15) は ${}_k\bar{\mathbf{\Pi}}_n$ を用いて次のように書き換えることができる.

$${}_k\bar{\mathbf{\Pi}}_n(\mathbf{I} - {}_k\mathbf{F}_{n,n}) = \text{diag}({}_k\mathbf{a}_n), \quad k = n+1, n+2, \dots$$

ただし, ${}_k\mathbf{a}_n$ ($k = n+1, n+2, \dots$) は $M_n \times 1$ ベクトルであり, その j 番目 ($j \in \mathcal{M}_n$) の要素 $[{}_k\mathbf{a}_n]_j$ は

$$[{}_k\mathbf{a}_n]_j = {}_k\bar{\mathbf{p}}_n(j)(\mathbf{I} - {}_k\mathbf{F}_{n,n})\mathbf{e}$$

で与えられる. よって次式を得る.

$$(\mathbf{I} - {}_k\mathbf{F}_{n,n})^{-1} = \text{diag}^{-1}({}_k\mathbf{a}_n) \cdot {}_k\bar{\mathbf{\Pi}}_n, \quad k = n+1, n+2, \dots \quad (16)$$

なお, $\text{diag}^{-1}({}_k\mathbf{a}_n)$ の j 番目 ($j \in \mathcal{M}_n$) の対角要素 $[{}_k\bar{\mathbf{p}}_n(j)(\mathbf{I} - {}_k\mathbf{F}_{n,n})\mathbf{e}]^{-1} = \mathbf{e}_j(\mathbf{I} - {}_k\mathbf{F}_{n,n})^{-1}\mathbf{e}$ は, 状態 (n, j) からレベル k への初到達時間内にレベル n を訪問する平均回数を表しており, 隠れマルコフ連鎖の正再帰性から, 十分大きな k に対しては正かつ有限となる.

次に, 式 (16) を用いて式 (11) を以下のように書き換える.

$$\bar{\mathbf{u}}_{k,n}(i) = \frac{{}_k\mathbf{f}_{k,n}(i)\text{diag}^{-1}({}_k\mathbf{a}_n) \cdot {}_k\bar{\mathbf{\Pi}}_n \cdot \text{diag}^{-1}(\boldsymbol{\mu}_n)}{{}_k\mathbf{f}_{k,n}(i)\text{diag}^{-1}({}_k\mathbf{a}_n) \cdot {}_k\bar{\mathbf{\Pi}}_n \cdot \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}} \quad \text{if } \mathbf{u}_{k,n}(i)\mathbf{e} > 0 \quad (17)$$

ここで補題 6 より, 任意の非負整数 n と十分小さな任意の正数 ϵ に対して, ある自然数 $K := K(n, \epsilon)$ が存在し,

$$\mathbf{O} < \left(1 - \frac{\epsilon}{2 + \epsilon}\right) \mathbf{e}\bar{\mathbf{p}}_n < {}_k\bar{\mathbf{\Pi}}_n < \left(1 + \frac{\epsilon}{2 + \epsilon}\right) \mathbf{e}\bar{\mathbf{p}}_n, \quad \forall k \geq K \quad (18)$$

が成立する。また、式 (17) に現れる ${}_k\mathbf{f}_{k,n}(i)$, $\text{diag}^{-1}({}_k\mathbf{a}_n)$ および $\text{diag}^{-1}(\boldsymbol{\mu}_n)$ は非負である。よって、式 (17), 式 (18), ならびに $\mathbf{u}_{k,n}(1)\mathbf{e} > 0$ から、以下の不等式を得る。

$$\frac{{}_k\mathbf{f}_{k,n}(1)\text{diag}^{-1}({}_k\mathbf{a}_n) \cdot \left(1 - \frac{\epsilon}{2+\epsilon}\right) \mathbf{e}\bar{\mathbf{p}}_n \cdot \text{diag}^{-1}(\boldsymbol{\mu}_n)}{{}_k\mathbf{f}_{k,n}(1)\text{diag}^{-1}({}_k\mathbf{a}_n) \cdot \left(1 + \frac{\epsilon}{2+\epsilon}\right) \mathbf{e}\bar{\mathbf{p}}_n \cdot \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}} < \bar{\mathbf{u}}_{k,n}(1) < \frac{{}_k\mathbf{f}_{k,n}(1)\text{diag}^{-1}({}_k\mathbf{a}_n) \cdot \left(1 + \frac{\epsilon}{2+\epsilon}\right) \mathbf{e}\bar{\mathbf{p}}_n \cdot \text{diag}^{-1}(\boldsymbol{\mu}_n)}{{}_k\mathbf{f}_{k,n}(1)\text{diag}^{-1}({}_k\mathbf{a}_n) \cdot \left(1 - \frac{\epsilon}{2+\epsilon}\right) \mathbf{e}\bar{\mathbf{p}}_n \cdot \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}}, \quad \forall k \geq K$$

さらに上式は次のように整理される。

$$\frac{1}{1+\epsilon} \cdot \frac{\bar{\mathbf{p}}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)}{\bar{\mathbf{p}}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}} < \bar{\mathbf{u}}_{k,n}(1) < (1+\epsilon) \cdot \frac{\bar{\mathbf{p}}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)}{\bar{\mathbf{p}}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}}, \quad \forall k \geq K \quad (19)$$

十分小さな ϵ ($\epsilon > 0$) に対しては、

$$\frac{1}{1+\epsilon} > 1 - \epsilon$$

が成立する。また、式 (9) より

$$\frac{\bar{\mathbf{p}}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)}{\bar{\mathbf{p}}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}} = \frac{\mathbf{p}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)}{\mathbf{p}_n \text{diag}^{-1}(\boldsymbol{\mu}_n)\mathbf{e}} = \bar{\mathbf{q}}_n > 0$$

である。よって、式 (19) より、任意の非負整数 n と十分小さな任意の $\epsilon > 0$ に対して、ある自然数 K が存在し、

$$(1-\epsilon)\bar{\mathbf{q}}_n < \bar{\mathbf{u}}_{k,n}(1) < (1+\epsilon)\bar{\mathbf{q}}_n, \quad \forall k \geq K$$

が成立する。以上により、定理 2 が示された。

3. レベル依存する M/G/1 型マルコフ連鎖に対する数値計算アルゴリズム

本節では、系 3 を基礎とした、レベル依存する M/G/1 型マルコフ連鎖における条件付き定常分布の数値計算アルゴリズムを議論する。式 (6) で示したように、レベル依存する M/G/1 型マルコフ連鎖における $\mathbf{U}_{k,n}$ は、 $\mathbf{U}_{k,n} = \mathbf{Z}_{k-1}\mathbf{Z}_{k-2}\cdots\mathbf{Z}_n$ で与えられるため、 $\mathbf{Z}_k$ ($k = 0, 1, \dots$) を順次計算することで $\mathbf{U}_{k,n}$ が計算可能である。さらに、ある非負整数 N が与えられたとき、レベルが N であるという条件下での相の条件付き定常分布 $\bar{\mathbf{q}}_N$ が系 3 を利用して求まれば、式 (5) より、条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ を得ることができる。この考えに沿った数値計算アルゴリズムの概要をアルゴリズム 1 に示す。Step 1 では $\mathbf{Z}_k$ ($k = 0, 1, \dots, N-1$) を式 (7) と式 (8) に基づき計算している。Step 2 では $\mathbf{U}_{k,N}$ ($k = N+1, N+2, \dots$) を順次計算し、系 3 で示した極限に収束したか否かの判定を行い、 $\bar{\mathbf{q}}_N$ を求めている。最後に Step 3 では、条件付き定常分布 $\{\mathbf{q}_k(N); k = 0, 1, \dots, N\}$ を

$$\mathbf{q}_k(N) = \mathbf{q}_{k+1}(N)\mathbf{U}_{k+1,k} = \mathbf{q}_{k+1}(N)\mathbf{Z}_k, \quad k = 0, 1, \dots, N-1$$

を利用して計算している。

この数値計算アルゴリズムは、Step 2 において $\bar{\mathbf{q}}_N$ の候補 $\bar{\mathbf{u}}$ の選択ならびに停止条件が明示されていないことを除き、仮定 4 の下で、文献 [6] で提案されている数値計算アルゴリズムと同一である。仮定 4 の下では、全ての i ($i \in \mathcal{M}$) に対して $\mathbf{u}_{k,N}(i)\mathbf{e} > 0$ 、かつ $\bar{\mathbf{u}}_{k,N}(i)\mathbf{e}$ ($i \in \mathcal{M}$) で張られる $(M-1)$ -単体 $\mathcal{P}_{k,N}$ に対して $\bar{\mathbf{q}}_N \in \cap_{k=N+1}^{\infty} \text{ri } \mathcal{P}_{k,N}$ が成立する。よって、 $\mathcal{P}_{k,N}$ の内点、例えば重心

$$\bar{\mathbf{g}}_{k,N}^{(\text{grav})} = \frac{1}{M} \sum_{i \in \mathcal{M}} \bar{\mathbf{u}}_{k,N}(i)$$

を $\bar{\mathbf{q}}_N$ の近似解として採用すると

$$\lim_{k \rightarrow \infty} \|\bar{\mathbf{u}}_{k,N}(i) - \bar{\mathbf{g}}_{k,N}^{(\text{grav})}\| = 0 \quad (i \in \mathcal{M}),$$

$$\max_{i \in \mathcal{M}} \|\bar{\mathbf{u}}_{k,N}(i) - \bar{\mathbf{g}}_{k,N}^{(\text{grav})}\| \leq \epsilon \Rightarrow \|\bar{\mathbf{g}}_{k,N}^{(\text{grav})} - \bar{\mathbf{q}}_N\| \leq \epsilon$$

アルゴリズム 1 (条件付き定常分布 $\{q_k(N); k = 0, 1, \dots, N\}$ の計算)

Input: $Q_{0,l}$ ($l = 0, 1, \dots$), $Q_{k,l}$ ($k = 1, 2, \dots, l = -1, 0, 1, \dots$), N , ならびに ϵ

Output: K と $q_k(N)$ ($k = 0, 1, 2, \dots, N$)

Step 1: Z_n ($n = 0, 1, \dots, N-1$) の計算

$T := Q_{0,0}$

for $k = 1$ to N **do**

$Z_{k-1} := Q_{k,-1}(-T)^{-1}$; 次式で $T (= T_k)$ を計算

$$T := Q_{k,0} + \sum_{n=0}^{k-1} Z_{k-1} Z_{k-2} \cdots Z_n Q_{n,k-n} \quad (\text{A})$$

end for

$U := I$; $k := N$

Step 2: 正規化された境界ベクトル $\bar{q}_N$ の決定

$k := k+1$; $Z_{k-1} := Q_{k,-1}(-T)^{-1}$; $U := Z_{k-1}U$

U から $\bar{q}_N$ の候補 $\bar{u}$ を決定

if 基準 ϵ の下で $\bar{u}$ が $\bar{q}_N$ に収束していると判断されなければ **then**

式 (A) から T を計算し, Step 2 に戻る

end if

$K := k$

Step 3: $q_k(N)$ ($k = 0, 1, \dots, N$) の計算

$q_N(N) := \bar{u}$; $sum := q_N(N)e$

for $k = N-1$ to 0 by -1 **do**

$q_k(N) := q_{k+1}(N)Z_k$; $sum := sum + q_k(N)e$

end for

for $k = 0$ to N **do**

$q_k(N) := q_k(N)/sum$

end for

が成立する. すなわち, 仮定 4 の下では, $\bar{q}_N$ の近似解 $\bar{g}_{k,N}^{(\text{grav})}$ の精度を保証する十分条件が存在するため, それに従って収束判定を行っている.

以下では, $U_{k,n}$ ($k = 1, 2, \dots, n = 0, 1, \dots, k-1$) が計算可能なレベル依存する G/G/1 型連続時間マルコフ連鎖において, ある非負整数 N が与えられたとき, 同様の収束判定が可能であることを示す. その準備として, まず, $\bar{q}_N$ の近似解の候補を議論する. レベル依存する G/G/1 型連続時間マルコフ連鎖では, 全ての $\bar{u}_{k,N}(i)$ ($k = N+1, N+2, \dots, i \in \mathcal{M}_k$) が非零である保証がない. さらに, 各レベル内での相のラベル付けは任意であるため, 一般には, 特定の相 i について $u_{k,N}(i)$ の $k \rightarrow \infty$ における極限を議論することができない. そこで, $\bar{q}_N$ に対する近似解の候補 $\bar{u}_{k,N}$ (レベル依存する M/G/1 型マルコフ連鎖に対するアルゴリズム 1 では $\bar{u}$) を次のように定める.

$$\bar{u}_{k,N} = \frac{e^T U_{k,N}}{e^T U_{k,N} e} > 0 \quad (20)$$

$\mathcal{M}_{k,N}^+$ ($k = N+1, N+2, \dots$) を $u_{k,N}(i)e > 0$ ($i \in \mathcal{M}_k$) となるラベル集合とすると, 式 (20) は $\bar{u}_{k,N}$ が $\bar{u}_{k,N}(i)$ ($i \in \mathcal{M}_{k,N}^+$) の凸結合で与えられることを示している. まず, 定義より, $U_{k,N}$ の行の中で, $u_{k,N}(i)e = 0$ となる行 $i \in \mathcal{M}_k \setminus \mathcal{M}_{k,N}^+$ は $\bar{u}_{k,N}$ に貢献しない. そこで, 記法を簡略化するため

$$\bar{u}_{k,N}(i) = 0, \quad i \in \mathcal{M}_k \setminus \mathcal{M}_{k,N}^+$$

と定義する. さらに $\bar{U}_{k,N}$ ($k = N+1, N+2, \dots$) を $M_k \times M_N$ 行列とし, その i 番目 ($i \in \mathcal{M}_k$) の行は $\bar{u}_{k,N}(i)$ で与えられるとする. このとき, $U_{k,N}$ ($k = N+1, N+2, \dots$) は

$$U_{k,N} = \text{diag}(U_{k,N}e)\bar{U}_{k,N} \quad (21)$$

と書き換えることができるので, 式 (20) は

$$\bar{u}_{k,N} = \frac{1}{e^T U_{k,N} e} \cdot e^T \text{diag}(U_{k,N}e)\bar{U}_{k,N}$$

と等価である。そこで

$$\alpha_{k,N} = \begin{pmatrix} \alpha_{k,N}(1) & \alpha_{k,N}(2) & \cdots & \alpha_{k,N}(M_k) \end{pmatrix} = \frac{1}{\mathbf{e}^T \mathbf{U}_{k,N} \mathbf{e}} \cdot \mathbf{e}^T \text{diag}(\mathbf{U}_{k,N} \mathbf{e})$$

と定義すると、 $\alpha_{k,N} \geq \mathbf{0}$ かつ $\alpha_{k,N} \mathbf{e} = 1$ である。さらに、 $\mathcal{M}_{k,N}^+ \neq \emptyset$ であり、定義より、 $i \in \mathcal{M}_k \setminus \mathcal{M}_{k,N}^+$ なる i に対しては $\alpha_{k,N}(i) = 0$ となる。よって

$$\bar{\bar{\mathbf{u}}}_{k,N} = \sum_{i \in \mathcal{M}_k} \alpha_{k,N}(i) \bar{\mathbf{u}}_{k,N}(i) = \sum_{i \in \mathcal{M}_{k,N}^+} \alpha_{k,N}(i) \bar{\mathbf{u}}_{k,N}(i) \quad (22)$$

かつ

$$\alpha_{k,N}(i) > 0 \quad (i \in \mathcal{M}_{k,N}^+), \quad \sum_{i \in \mathcal{M}_{k,N}^+} \alpha_{k,N}(i) = 1$$

となるような $\alpha_{k,N}(i)$ ($i \in \mathcal{M}_{k,N}^+$) が存在することが分かる。

補題 7. レベル依存する G/G/1 型連続時間マルコフ連鎖では、任意の非負整数 N と任意の正数 ϵ に対して、ある自然数 $K := K(N, \epsilon)$ が存在し、全ての k ($k = K+1, K+2, \dots$) に対して次式が成立する。

$$\|\bar{\mathbf{u}}_{k,N}(i) - \bar{\bar{\mathbf{u}}}_{k,N}\| < \epsilon, \quad i \in \mathcal{M}_{k,N}^+$$

証明. 系 3 ならびに式 (22) より、任意の非負整数 N と任意の正数 ϵ に対して、ある自然数 $K := K(N, \epsilon)$ が存在し、全ての $k = K, K+1, \dots$ に対して

$$\|\bar{\mathbf{q}}_N - \bar{\bar{\mathbf{u}}}_{k,N}\| = \left\| \sum_{i \in \mathcal{M}_{k,N}^+} \alpha_{k,N}(i) (\bar{\mathbf{q}}_N - \bar{\mathbf{u}}_{k,N}(i)) \right\| \leq \sum_{i \in \mathcal{M}_{k,N}^+} \alpha_{k,N}(i) \|\bar{\mathbf{q}}_N - \bar{\mathbf{u}}_{k,N}(i)\| \leq \sum_{i \in \mathcal{M}_{k,N}^+} \alpha_{k,N}(i) \cdot \epsilon = \epsilon$$

が成立する。よって、系 3 より

$$\|\bar{\mathbf{u}}_{k,N}(i) - \bar{\bar{\mathbf{u}}}_{k,N}\| \leq \|\bar{\mathbf{u}}_{k,N}(i) - \bar{\mathbf{q}}_N\| + \|\bar{\mathbf{q}}_N - \bar{\bar{\mathbf{u}}}_{k,N}\| \leq 2\epsilon, \quad \forall k \geq K$$

が成立する。最後に $\epsilon := 2\epsilon$ と置き直すことで、補題 7 を得る。 $\square$

補題 8. レベル依存する G/G/1 型連続時間マルコフ連鎖では、任意の N, k の組 ($N = 0, 1, \dots, k = N+1, N+2, \dots$) に対して、 $\bar{\mathbf{q}}_N$ は $\bar{\mathbf{u}}_{k,N}(i)$ ($i \in \mathcal{M}_{k,N}^+$) の凸結合で与えられる。

$$\bar{\mathbf{q}}_N = \sum_{i \in \mathcal{M}_{k,N}^+} \beta_{k,N}(i) \bar{\mathbf{u}}_{k,N}(i)$$

ただし

$$\beta_{k,N}(i) > 0 \quad (i \in \mathcal{M}_{k,N}^+), \quad \sum_{i \in \mathcal{M}_{k,N}^+} \beta_{k,N}(i) = 1 \quad (23)$$

証明. 式 (3) と式 (21) から

$$\bar{\mathbf{q}}_N = \frac{1}{\mathbf{q}_k \mathbf{U}_{k,N} \mathbf{e}} \cdot \mathbf{q}_k \mathbf{U}_{k,N} = \frac{1}{\mathbf{q}_k \mathbf{U}_{k,N} \mathbf{e}} \cdot \mathbf{q}_k \text{diag}(\mathbf{U}_{k,N} \mathbf{e}) \bar{\mathbf{U}}_{k,N}$$

を得る。ここで

$$\begin{pmatrix} \beta_{k,N}(1) & \beta_{k,N}(2) & \cdots & \beta_{k,N}(M_k) \end{pmatrix} = \frac{1}{\mathbf{q}_k \mathbf{U}_{k,N} \mathbf{e}} \cdot \mathbf{q}_k \text{diag}(\mathbf{U}_{k,N} \mathbf{e})$$

とし、 $\mathbf{q}_k > \mathbf{0}$ に注意すると、式 (22) と同様の議論より補題 8 が得られる。 $\square$

定理 9. レベル依存する G/G/1 型連続時間マルコフ連鎖では、任意の N, k の組 ($N = 0, 1, \dots, k = N+1, N+2, \dots$) に対して次式が成立する。

$$\|\bar{\bar{\mathbf{u}}}_{k,N} - \bar{\mathbf{q}}_N\| \leq \max_{i \in \mathcal{M}_{k,N}^+} \|\bar{\mathbf{u}}_{k,N}(i) - \bar{\bar{\mathbf{u}}}_{k,N}\|$$

アルゴリズム 2 (アルゴリズム 1 における Step 2)

Step 2: 正規化された境界ベクトル $\bar{q}_N$ の決定

$k := k + 1$; $Z_{k-1} := Q_{k,-1}(-T)^{-1}$; $U := Z_{k-1}U$

$\bar{u}$ を式 (20) により計算する

if $\max_{\substack{i \in \mathcal{M}_k \\ u(i)e > 0}} \|\bar{u}(i) - \bar{u}\| > \epsilon$ **then**

式 (A) により T を計算し, Step 2 に戻る

end if

$K := k$

証明. 補題 8 から

$$\|\bar{u}_{k,N} - \bar{q}_N\| = \left\| \sum_{i \in \mathcal{M}_{k,N}^+} \beta_k(i) (\bar{u}_{k,N} - \bar{u}_{k,N}(i)) \right\| \leq \sum_{i \in \mathcal{M}_{k,N}^+} \beta_k(i) \|\bar{u}_{k,N} - \bar{u}_{k,N}(i)\| \leq \max_{i \in \mathcal{M}_{k,N}^+} \|\bar{u}_{k,N} - \bar{u}_{k,N}(i)\|$$

が成立する. □

補題 7 ならびに定理 9 より, $\bar{q}_N$ の近似解 $\bar{q}_N^{\text{approx}}$ を計算する際には, 適当な正数 ϵ に対して

$$\max_{i \in \mathcal{M}_{k,N}^+} \|\bar{u}_{k,N}(i) - \bar{u}_{k,N}\| \leq \epsilon$$

となるような k ($k = N+1, N+2, \dots$) が現れるまで, 順次 $U_{k,N}$ を計算していき, 上式が成立した時点で $\bar{q}_N$ の近似解を $\bar{q}_N^{\text{approx}} = \bar{u}_{k,N}$ と取れば, 真値との誤差 $\|\bar{q}_N^{\text{approx}} - \bar{q}_N\|$ が ϵ 以下となることが保証される. 以上をまとめて, レベル依存する M/G/1 型マルコフ連鎖に対するアルゴリズム 1 中の Step 2 をアルゴリズム 2 で置き換えた数値計算アルゴリズムを提案する. ただし $\bar{u}(i)$ は U の i 番目の行ベクトルを表している.

本節を終えるにあたり, 以下の系を与える.

系 10. レベル依存する G/G/1 型連続時間マルコフ連鎖において, ある N, k の組 ($N = 0, 1, \dots, k = N+1, N+2, \dots$) に対して $U_{k,N}$ のランクが 1 であるならば, 次式が成立する.

$$\bar{q}_N = \bar{u}_{k,N}$$

証明. $U_{k,N}$ のランクが 1 であるならば, 式 (22) から, $\bar{u}_{k,N}(i) = \bar{u}_{k,N}$ ($i \in \mathcal{M}_{k,N}^+$) となる. よって系 10 は定理 9 より直ちに従う. □

注意 11. $U_{k,N}$ のランクが 1 である場合, $U_{k,N} = U_{k,N}e \cdot \bar{u}_{k,N}$ と書くことができる. よって, 系 10 は式 (3) を用いて下記のように直接, 導出可能である.

$$\bar{q}_N = \frac{1}{q_N e} \cdot q_N = \frac{1}{q_N e} \cdot q_k U_{k,N} = \frac{1}{q_N e} \cdot q_k U_{k,N} e \cdot \bar{u}_{k,N} = \frac{1}{q_N e} \cdot q_N e \cdot \bar{u}_{k,N} = \bar{u}_{k,N}$$

また, レベル依存する M/G/1 型連続時間マルコフ連鎖において $U_{N+1,N}$ のランクが 1 である場合, 本稿で提案した数値計算アルゴリズムの Step 2 は 1 回だけ実行され, $K = N+1$ となる. なお, $|\mathcal{M}_{k,N}^+| = 1$ は $U_{k,N}$ のランクが 1 であるための十分条件である.

4. おわりに

本稿では, 文献 [6] において, 仮定 4 の下で提案されたレベル依存する M/G/1 型連続時間マルコフ連鎖の条件付き定常分布 $\{q_k(N); k = 0, 1, \dots, N\}$ に対する数値計算アルゴリズムが, 仮定 4 が成立しない任意のレベル依存する M/G/1 型連続時間マルコフ連鎖にも適用できることを示した. この数値計算アルゴリズムは, 境界ベクトル $\bar{q}_N$ の近似解がもつ精度の保証が可能であるという特徴をもつ. 本稿では割愛したが, 数値計算アルゴリズムで計算される条件付き定常分布の近似 $q^{\text{approx}}(N)$ と真値 $q(N)$ の誤差を $\|q^{\text{approx}}(N) - q(N)\|$ で評価した際, これが予め与えられた正数 ϵ 以下となるように, Step 2 の収束判定を行うことも可能である.

また, 本稿で提案した数値計算アルゴリズムでは, 系 10 で示したように, $U_{k,N}$ のランクが 1 となるような k ($k > N$) が存在する場合, 境界ベクトル $\bar{q}_N$ が厳密解となるため, 条件付き定常分布 $\{q_k; k = 0, 1, \dots, N\}$ の厳密解が計算可能である. ここで, 式 (6) ならびに式 (7) から明らかなように, レベル依存する M/G/1 型

連続時間マルコフ連鎖の場合、 $\mathbf{Q}_{N+1,-1}$ のランクが 1 であることは、 $\mathbf{U}_{N+1,N} = \mathbf{Z}_N$ のランクが 1 であるための必要十分条件となっている。すなわち、レベル依存する M/G/1 型連続時間マルコフ連鎖の場合、条件付き定常分布の厳密解が計算できるクラスが、マルコフ連鎖の遷移構造によって事前に判別可能である。例えば、サービス速度が系内客数に依存する集団到着 M/PH/1 待ち行列モデルがこの場合に該当する。

A. 補題 5 の証明

$\Omega_n(i)$ ($i \in \mathcal{M}_n$) を状態 (n, i) から開始したレベル n への再帰時間の標本路集合とする。また、 ${}_k\Omega_n(i)$ および ${}_{k+}\Omega_n(i)$ ($i \in \mathcal{M}_n$, $k = n+1, n+2, \dots$) をそれぞれ状態 (n, i) から開始したレベル n への再帰時間の標本路のうち、レベル k およびレベル k 以上に遷移しない標本路の集合とする。

ここで ${}_{k+}\mathbf{F}_{n,n}$ ($k = n+1, n+2, \dots$) を $M_n \times M_n$ 行列とし、その (i, j) 要素 ($i, j \in \mathcal{M}_n$) を状態 (n, i) から開始したレベル n への再帰時間がレベル k 以上に遷移することなく状態 (n, j) で終了する確率を表すものとする。定義から、 ${}_{k+}\Omega_n(i) \subset {}_k\Omega_n(i) \subset \Omega_n(i)$ が成立する。よって次式を得る。

$${}_{k+}\mathbf{F}_{n,n} \leq {}_k\mathbf{F}_{n,n} \leq \mathbf{F}_{n,n}, \quad k = n+1, n+2, \dots \quad (24)$$

さらに、 ${}_{k+}\Omega_n(i) \subset {}_{(k+1)+}\Omega_n(i)$ ($k = n+1, n+2, \dots$) であり、極限 $\lim_{k \rightarrow \infty} {}_{k+}\Omega_n(i)$ はいずれ終了する再帰時間の標本路全てからなる集合を表すため、 $\lim_{k \rightarrow \infty} {}_{k+}\Omega_n(i) = \Omega_n(i)$ が成立する。以上より、 $\lim_{k \rightarrow \infty} {}_{k+}\mathbf{F}_{n,n} = \mathbf{F}_{n,n}$ が成立し、式 (24) から補題が示される。

B. 補題 6 の証明

$\bar{\mathbf{p}}_n$ は $\mathbf{F}_{n,n}$ に対する不変確率ベクトルであることに注意する。

$$\bar{\mathbf{p}}_n = \bar{\mathbf{p}}_n \mathbf{F}_{n,n}, \quad \bar{\mathbf{p}}_n \mathbf{e} = 1$$

よって、式 (13) を用いると

$$(\bar{\mathbf{p}}_n(j) - \bar{\mathbf{p}}_n) (\mathbf{I} - \mathbf{F}_{n,n} + \mathbf{e}\bar{\mathbf{p}}_n) = {}_k\bar{\mathbf{p}}_n(j) (\mathbf{I} - \mathbf{F}_{n,n}) + \bar{\mathbf{p}}_n - \bar{\mathbf{p}}_n = {}_k\bar{\mathbf{p}}_n(j) ({}_k\bar{\mathbf{F}}_n(j) - \mathbf{F}_{n,n})$$

を得る。さらに $\mathbf{F}_{n,n}$ は既約であるので、 $\mathbf{I} - \mathbf{F}_{n,n} + \mathbf{e}\bar{\mathbf{p}}_n$ は正則である。よって

$${}_k\bar{\mathbf{p}}_n(j) = \bar{\mathbf{p}}_n + {}_k\bar{\mathbf{p}}_n(j) ({}_k\bar{\mathbf{F}}_n(j) - \mathbf{F}_{n,n}) (\mathbf{I} - \mathbf{F}_{n,n} + \mathbf{e}\bar{\mathbf{p}}_n)^{-1}, \quad k = n+1, n+2, \dots, j \in \mathcal{M}_n \quad (25)$$

を得る。一方、補題 5 から、 $\lim_{k \rightarrow \infty} (\mathbf{F}_{n,n} - {}_k\mathbf{F}_{n,n})\mathbf{e} = \mathbf{0}$ が成立する。よって、式 (12) から

$$\lim_{k \rightarrow \infty} {}_k\bar{\mathbf{F}}_n(j) = \lim_{k \rightarrow \infty} {}_k\mathbf{F}_{n,n} = \mathbf{F}_{n,n}, \quad j \in \mathcal{M}_n \quad (26)$$

を得る。 M_n は有限なので、式 (25) の両辺の $k \rightarrow \infty$ を取り、 ${}_k\bar{\mathbf{p}}_n(j) \geq \mathbf{0}$, ${}_k\bar{\mathbf{p}}_n(j)\mathbf{e} = 1$ ならびに式 (26) を用いると式 (14) が得られる。

謝辞 本研究の一部は JSPS 科研費 16K00034 の助成を受けている。

参考文献

- [1] L. Bright and P. G. Taylor, Calculating the equilibrium distribution in level-dependent quasi-birth-and-death process. *Stochastic Models*, 11 (1995) 497–526.
- [2] V. Klimenok and A. Dudin, Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory. *Queueing Systems*, 54 (2006) 245–259.
- [3] Q. L. Li, Z. T. Lian, and L. M. Liu, An RG-factorization approach for a BMAP/M/1 generalized processor-sharing queue. *Queueing Systems*, 21 (2005) 507–530.
- [4] H. Masuyama, The limit formula for the normalized fundamental matrix of the northwest-corner truncation of Markov chains: Matrix-infinite-product form solutions of block-Hessenberg Markov chains. Available online: <https://arxiv.org/abs/1603.07787> (2016).
- [5] M. F. Neuts, *Structured Stochastic Matrix of M/G/1 Type and Thier Applications*, Marcel Dekker, New York (1989).
- [6] T. Takine, Analysis and computation of the stationary distribution in a special class of level-dependent M/G/1-type and its application to BMAP/M/ ∞ and BMAP/M/c+M queues. *Queueing Systems*, 84 (2016) 49–77.

Geometric convergence of finite-level M/G/1-type Markov chains

Yosuke Katsumata

Hiroyuki Masuyama
Graduate School of Informatics
Kyoto University

Yutaka Takahashi

Abstract This paper considers the convergence of the stationary distribution vector of a finite-level M/G/1-type Markov chain as its maximum level approaches to infinity. The main contribution of this paper is to prove that the stationary distribution vector of the finite-level M/G/1-type Markov chain converges geometrically to that of the corresponding infinite-level M/G/1-type Markov chain under the condition that the level increment has a geometric decay. We first express the difference between the stationary distribution vectors of the finite- and infinite-level M/G/1-type Markov chains, using the deviation matrix of the infinite chain. Next, using a matrix-analytic expression of the deviation matrix, we present some asymptotic results on the difference between the stationary distribution vectors.

1. Introduction

This paper considers finite-level M/G/1-type Markov chains. Such Markov chains appear as the queue length processes in semi-Markovian queues with finite capacity (see, e.g., [4, 5, 7, 9, 14]). In fact, the finite-level M/G/1-type Markov chain is equivalent to the last-column-block-augmented truncation (LC-block-augmented truncation) of the corresponding infinite-level M/G/1-type Markov chain (see, e.g., [21, 25]).

We can compute, in principle, the stationary distribution vector of the finite-level M/G/1-type Markov chain by using general-purpose methods, such as Gauss-Seidel method and LU decomposition method. However, except for a few special cases [2, 17], the stationary distribution vector of the finite-level M/G/1-type Markov chain does not have an informative and numerically feasible expression. Thus, several researchers have studied the approximation and asymptotic analysis of finite-level M/G/1-type Markov chains and related ones that originate from finite semi-Markovian queues.

Miyazawa et al. [28] studied the asymptotics of the loss probability of a finite queue associated with a finite-level quasi-birth-and-death process, which is a special case of both finite-level M/G/1-type and GI/M/1-type Markov chains. J. Kim and B. Kim [19] extended Miyazawa et al. [28]'s result to the finite-level GI/M/1-type Markov chain. Baiocchi [5] derived a geometric asymptotic formula for the loss probability of a MAP/G/1/ K queue, which is a typical example of finite queues associated with finite-level M/G/1-type Markov chains. Baiocchi and Blefari-Melazzi [6] proposed the approximation method for the loss probability of the MAP/G/1/ K queue, using a partial fraction decomposition of the matrix generating function of the total arrival number in a single service time. Akar and Arikan [1] discussed the rational approximation of the virtual waiting time distribution of a MAP/D/1/ K queue.

Gouweleeuw [11] proposed an approximate formula for the loss probability of a BMAP/G/1/ K queue, which expresses the loss probability in terms of the queue length distribution of the corresponding BMAP/G/1/ ∞ queue, i.e., BMAP/G/1 queue. Gouweleeuw and Tijm [12] also applied the approximate method proposed in [11] to various discrete-time finite-buffer queues. Ishizaki and Takine [15] proved that such a connection between finite and infinite queues is exactly established for a special finite-level M/G/1-type Markov chain whose level-decreasing jumps are governed by a block matrix of rank one. Using this result, Ishizaki and Takine [16] also derived an approximate formula for the loss probability of a BMAP/D/1 queue in discrete time.

This paper considers the asymptotics of the difference between the stationary distribution vectors of a finite-level M/G/1-type Markov chain and the corresponding infinite-level M/G/1-type Markov chain as the maximum level of the finite chain approaches to infinity. This study intends to

evaluate the error of the approximate formulae for the loss probability of finite queues connected to finite-level M/G/1-type Markov chains. As mentioned in the beginning, a finite-level M/G/1-type Markov chain is the LC-block-augmented truncation of the corresponding infinite-level M/G/1-type Markov chain. Some total-variation-distance error bounds are established for the stationary distribution vector obtained by the LC-block-augmented truncation of block-structured Markov chain including the M/G/1-type Markov chain (see, e.g., [23, 24, 25, 26]). However, these bounds do not tell us an exact convergence speed of the difference between the two stationary distribution vectors. Katsumata et al. [18] derived a subgeometric asymptotic formula for the difference between the stationary distribution vectors of a finite-level M/G/1-type Markov chain and the corresponding infinite-level M/G/1-type Markov chain, under the condition that the infinite chain has level increments whose equilibrium distribution is long-tailed [10]. Katsumata et al. [18] also assumed that the level increments follow a subexponential distribution, under which the authors proved that the difference between the two stationary distributions asymptotically decays at the same speed as the stationary tail distribution vector of the infinite chain.

As far as we know, there have been no studies on the geometric convergence rates of the difference between the stationary distribution vectors of the finite-level M/G/1-type Markov chain and the corresponding infinite-level M/G/1-type Markov chain. In this paper, we assume that the infinite chain has level increments with an asymptotically geometric decay. We then investigate the convergence rates of the stationary distributions of the finite and infinite chains.

The rest of this paper is organized as follows. Section 2 describes finite-level and infinite M/G/1-type Markov chains and provides preliminary results on these Markov chains. Section 3 discusses the deviation matrix of the infinite-level M/G/1-type Markov chain, which plays an important role to derive the main result of this paper. Finally, Section 4 presents the main result of this paper.

2. M/G/1-type Markov chains

In this section, we describe the M/G/1-type Markov chain together with related definitions and assumptions.

Let $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ and $\mathbb{N} = \{1, 2, 3, \dots\}$. Let $x \wedge y = \min(x, y)$ for $x, y \in (-\infty, \infty)$. We then define $\{(X_\nu, J_\nu); \nu \in \mathbb{Z}_+\}$ as a discrete-time Markov chain such that, for all $\nu \in \mathbb{Z}_+$,

$$\begin{aligned} X_\nu &\in \mathbb{Z}_+, \\ J_\nu &\in \mathbb{M}_0 := \{1, 2, \dots, M_0\} \subset \mathbb{N} && \text{if } X_\nu = 0, \\ J_\nu &\in \mathbb{M}_1 := \{1, 2, \dots, M_1\} \subset \mathbb{N} && \text{if } X_\nu \in \mathbb{N}. \end{aligned}$$

Thus, the state space of the Markov chain $\{(X_\nu, J_\nu)\}$ is given by

$$\mathbb{F} := \bigcup_{k=0}^{\infty} \mathbb{L}(k),$$

where $\mathbb{L}(k) = \{k\} \times \mathbb{M}_{k \wedge 1}$ for $k \in \mathbb{Z}_+$. The subset $\mathbb{L}(k)$ of state space $\mathbb{F}$ is called *level k*.

Let $\mathbf{P}$ denote the transition probability matrix of the Markov chain $\{(X_\nu, J_\nu)\}$. We then assume that $\mathbf{P}$ is in the form:

$$\mathbf{P} = \begin{matrix} & \mathbb{L}(0) & \mathbb{L}(1) & \mathbb{L}(2) & \mathbb{L}(3) & \cdots \\ \begin{matrix} \mathbb{L}(0) \\ \mathbb{L}(1) \\ \mathbb{L}(2) \\ \mathbb{L}(3) \\ \vdots \end{matrix} & \begin{pmatrix} \mathbf{B}(0) & \mathbf{B}(1) & \mathbf{B}(2) & \mathbf{B}(3) & \cdots \\ \mathbf{B}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \mathbf{A}(2) & \cdots \\ \mathbf{O} & \mathbf{A}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \cdots \\ \mathbf{O} & \mathbf{O} & \mathbf{A}(-1) & \mathbf{A}(0) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{matrix}, \quad (1)$$

where $\mathbf{O}$ denotes the zero matrix. Note here that $\mathbf{A}(k)$ and $\mathbf{B}(k)$, $k \in \mathbb{Z}_+ \cup \{-1\}$, satisfy

$$\mathbf{B}(0)\mathbf{e} + \sum_{k=1}^{\infty} \mathbf{B}(k)\mathbf{e} = \mathbf{e}, \quad \mathbf{B}(-1)\mathbf{e} + \sum_{k=0}^{\infty} \mathbf{A}(k)\mathbf{e} = \mathbf{e}, \quad \sum_{k=-1}^{\infty} \mathbf{A}(k)\mathbf{e} = \mathbf{e},$$

where $\mathbf{e}$ denotes the column vector of ones whose order is determined according to the context. The Markov chain $\{(X_\nu, J_\nu)\}$ described above is called the *infinite-level M/G/1-type Markov chain* or *M/G/1-type Markov chain*, for short (see, e.g., [29]).

Throughout the paper, we assume the following.

Assumption 2.1 (i) The transition probability matrix $\mathbf{P}$ is irreducible and aperiodic; (ii) the stochastic $\mathbf{A}$ is irreducible; (iii) $\sum_{k=1}^{\infty} k\mathbf{B}(k)\mathbf{e} < \infty$; and (iv) $\sigma := \boldsymbol{\pi} \sum_{k=-1}^{\infty} k\mathbf{A}(k)\mathbf{e} < 0$, where $\boldsymbol{\pi}$ denotes the unique stationary distribution vector of $\mathbf{A}$.

Under Assumption 2.1, the M/G/1-type Markov chain $\{(X_\nu, J_\nu)\}$ is ergodic (i.e., irreducible, aperiodic and positive recurrent) [3, Chapter XI, Proposition 3.1] and thus the Markov chain has the unique stationary distribution vector, denoted by $\mathbf{x} = (x(k, i))_{(k, i) \in \mathbb{F}}$. For later use, we define $\mathbf{x}(k)$, $k \in \mathbb{Z}_+$, as the subvector of $\mathbf{x}$ corresponding to level k , i.e., $\mathbf{x}(k) = (x(\ell, i))_{(\ell, i) \in \mathbb{L}(k)}$.

Next we introduce a finite-level version of the (infinite-level) M/G/1-type Markov chain. Let $\{(X_\nu^{(N)}, J_\nu^{(N)}); \nu \in \mathbb{Z}_+\}$ denote a discrete-time Markov chain with state space $\mathbb{F}^{\leq N} := \bigcup_{k=0}^N \mathbb{L}(k)$, whose transition probability matrix $\mathbf{P}^{(N)}$ is given by

$$\mathbf{P}^{(N)} = \begin{matrix} & \mathbb{L}(0) & \mathbb{L}(1) & \mathbb{L}(2) & \cdots & \mathbb{L}(N-2) & \mathbb{L}(N-1) & \mathbb{L}(N) \\ \begin{matrix} \mathbb{L}(0) \\ \mathbb{L}(1) \\ \mathbb{L}(2) \\ \vdots \\ \mathbb{L}(N-2) \\ \mathbb{L}(N-1) \\ \mathbb{L}(N) \end{matrix} & \left(\begin{matrix} \mathbf{B}(0) & \mathbf{B}(1) & \mathbf{B}(2) & \cdots & \mathbf{B}(N-2) & \mathbf{B}(N-1) & \overline{\mathbf{B}}(N-1) \\ \mathbf{B}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \cdots & \mathbf{A}(N-3) & \mathbf{A}(N-2) & \overline{\mathbf{A}}(N-2) \\ \mathbf{O} & \mathbf{A}(-1) & \mathbf{A}(0) & \cdots & \mathbf{A}(N-4) & \mathbf{A}(N-3) & \overline{\mathbf{A}}(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \cdots & \mathbf{A}(0) & \mathbf{A}(1) & \overline{\mathbf{A}}(1) \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \cdots & \mathbf{A}(-1) & \mathbf{A}(0) & \overline{\mathbf{A}}(0) \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \cdots & \mathbf{O} & \mathbf{A}(-1) & \overline{\mathbf{A}}(-1) \end{matrix} \right) \end{matrix},$$

where

$$\overline{\mathbf{A}}(k) = \sum_{\ell=k+1}^{\infty} \mathbf{A}(\ell), \quad \overline{\mathbf{B}}(k) = \sum_{\ell=k+1}^{\infty} \mathbf{B}(\ell), \quad k \in \mathbb{Z}_+ \cup \{-1\}.$$

We call the Markov chain $\{(X_\nu^{(N)}, J_\nu^{(N)})\}$ the *finite-level M/G/1-type Markov chain*. Note that $\mathbf{P}^{(N)}$ is the *last-column-block-augmented truncation* (LC-block-augmented truncation) of $\mathbf{P}$ (see, e.g., [25, 26]). It thus follows from Assumption 2.1 and [26, Proposition 1.1] that $\mathbf{P}^{(N)}$ has at least one and at most M_1 closed communicating classes. Therefore, $\mathbf{P}^{(N)}$ has a stationary distribution vector, denoted by $\mathbf{x}^{(N)} = (x^{(N)}(k, i))_{(k, i) \in \mathbb{F}^{\leq N}}$. We extend $\mathbf{P}^{(N)}$ and $\mathbf{x}^{(N)}$ to the orders of $\mathbf{P}$ and $\mathbf{x}$, respectively, by appending zero elements. We also define $\mathbf{x}^{(N)}(k) = (x^{(N)}(\ell, i))_{(\ell, i) \in \mathbb{L}(k)}$ for $k \in \mathbb{Z}_+$, which is the subvector of $\mathbf{x}^{(N)}$ corresponding to level k .

In addition to Assumption 2.1, we assume the following.

Assumption 2.2 It holds that $\sum_{k=1}^{\infty} k^2 \mathbf{A}(k) < \infty$ and $\sum_{k=1}^{\infty} k^2 \mathbf{B}(k) < \infty$.

Under Assumptions 2.1 and 2.2, we can show (see Theorem 3.3 below) that there exists a matrix $\mathbf{D}$ such that

$$\mathbf{D} = \sum_{n=0}^{\infty} (\mathbf{P}^n - \mathbf{e}\mathbf{x}).$$

The matrix $\mathbf{D}$ is called the *deviation matrix* of the Markov chain $\{(X_\nu, J_\nu)\}$ and the transition probability matrix $\mathbf{P}$. It is known (see, e.g., [13]) that the difference between $\mathbf{x}^{(N)}$ and $\mathbf{x}$ is expressed as follows:

$$\mathbf{x}^{(N)} - \mathbf{x} = \mathbf{x}^{(N)}(\mathbf{P}^{(N)} - \mathbf{P})\mathbf{D}. \quad (2)$$

Using this equation, we discuss the convergence of $\{\mathbf{x}^{(N)}; N \in \mathbb{N}\}$ to $\mathbf{x}$ in Section 4. To this end, we provide a condition for the existence of $\mathbf{D}$ and then derive a matrix-analytic expression of $\mathbf{D}$ in the next section.

3. Deviation matrix

We define the first passage time to $\mathbb{A} \subset \mathbb{F}$ as $\tau_{\mathbb{A}} = \inf\{\nu \in \mathbb{N}; (X_\nu, J_\nu) \in \mathbb{A}\}$. For convenience, we introduce the notation:

$$\begin{aligned} \mathbb{E}_{(k,i)}[\cdot] &= \mathbb{E}[\cdot \mid (X_0, J_0) = (k, i)], \quad (k, i) \in \mathbb{F}, \\ \mathbb{E}_{\mathbf{x}}[\cdot] &= \sum_{(k,i) \in \mathbb{F}} x(k, i) \mathbb{E}_{(k,i)}[\cdot]. \end{aligned}$$

The following proposition shows a condition for the existence of $\mathbf{D}$.

Proposition 3.1 ([8, Lemma 2.6]) *Suppose that a discrete-time Markov chain $\{(X_\nu, J_\nu)\}$ is ergodic. The deviation matrix $\mathbf{D}$ exists if and only if $\mathbb{E}_{\mathbf{x}}[\tau_{\{(k,i)\}}] < \infty$ for some $(k, i) \in \mathbb{F}$. In addition, if $\mathbb{E}_{\mathbf{x}}[\tau_{\{(k,i)\}}] < \infty$ for some $(k, i) \in \mathbb{F}$, then $\mathbb{E}_{\mathbf{x}}[\tau_{\{(\ell,j)\}}] < \infty$ for all $(\ell, j) \in \mathbb{F}$.*

We need the following lemma to apply Proposition 3.1 to the M/G/1-type Markov chain.

Lemma 3.2 *Suppose that the M/G/1-type Markov chain $\{(X_\nu, J_\nu)\}$ satisfies Assumption 2.1. It then holds that Assumption 2.2 is equivalent to $\sum_{k=1}^{\infty} k\mathbf{x}(k) < \infty$. Furthermore, $\sum_{k=1}^{\infty} k\mathbf{x}(k) < \infty$ if and only if $\mathbb{E}_{\mathbf{x}}[\tau_{(\ell,i)}] < \infty$ for all $(\ell, i) \in \mathbb{F}$.*

Using Proposition 3.1 and Lemma 3.2, we obtain a necessary and sufficient condition for the existence of $\mathbf{D}$ of the M/G/1-type Markov chain $\{(X_\nu, J_\nu)\}$.

Theorem 3.3 *Suppose that the M/G/1-type Markov chain $\{(X_\nu, J_\nu)\}$ satisfies Assumption 2.1. The deviation matrix $\mathbf{D}$ exists if and only if Assumption 2.2 holds.*

Next we derive a matrix-analytic expression of $\mathbf{D}$. For any matrix (or vector) $\mathbf{F}$, we define $|\mathbf{F}|$ as the matrix (or the vector) obtained by taking the absolute values of the entries of $\mathbf{F}$. We then have Proposition 3.4 below, which shows that $\mathbf{D}$ is the unique solution of the following matrix equations.

Proposition 3.4 ([8, Lemma 2.7]) *Suppose that the M/G/1-type Markov chain $\{(X_\nu, J_\nu)\}$ is ergodic and $\mathbf{D}$ exists. Under these conditions, $\mathbf{D}$ equals the unique solution $\mathbf{\Delta}$ of the following matrix equations:*

$$(\mathbf{I} - \mathbf{P})\mathbf{\Delta} = \mathbf{I} - \mathbf{e}\mathbf{x}, \quad \mathbf{x}\mathbf{\Delta} = \mathbf{0}, \quad \mathbf{\Delta}\mathbf{e} = \mathbf{0}, \quad (3)$$

where $\mathbf{x}|\mathbf{\Delta}| < \infty$.

In what follows, we obtain a matrix-analytic expression of $\mathbf{D}$ by using the matrix equations in (3). For this purpose, we introduce some definitions and known results. Let $\underline{\mathbb{L}}(n) = \bigcup_{k=0}^n \mathbb{L}(k)$ and

$\bar{\mathbb{L}}(n) = \bigcup_{k=n+1}^{\infty} \mathbb{L}(k)$. Let $\Phi(0) := (\Phi_{i,j}(0))_{i,j \in \mathbb{M}_1}$, $\mathbf{G} := (G_{i,j})_{i,j \in \mathbb{M}_1}$ and $\mathbf{R}(k) := (R_{i,j}(k))_{i,j \in \mathbb{M}_1}$, $k \in \mathbb{N}$, whose (i, j) th entries are given by

$$\begin{aligned}\Phi_{i,j}(0) &= \mathbb{P} \left(\tau_{\bar{\mathbb{L}}(n)} < \infty, (X_{\tau_{\bar{\mathbb{L}}(n)}}, J_{\tau_{\bar{\mathbb{L}}(n)}}) = (n, j) \mid (X_0, J_0) = (n, i) \in \bar{\mathbb{L}}(0) \right), \\ G_{i,j} &= \mathbb{P} \left(\tau_{\bar{\mathbb{L}}(n)} < \infty, (X_{\tau_{\bar{\mathbb{L}}(n)}}, J_{\tau_{\bar{\mathbb{L}}(n)}}) = (n, j) \mid (X_0, J_0) = (n+1, i) \in \bar{\mathbb{L}}(1) \right), \\ R_{i,j}(k) &= \mathbb{E} \left[\sum_{\nu=1}^{\tau_{\bar{\mathbb{L}}(n+k-1)}} \mathbb{1}((X_{\nu}, J_{\nu}) = (n+k, j)) \mid (X_0, J_0) = (n, i) \in \bar{\mathbb{L}}(0) \right],\end{aligned}$$

where $\mathbb{1}(\cdot)$ denotes the indicator function of the event in the parentheses.

The following proposition presents the basic result on the structure of $\mathbf{G}$.

Proposition 3.5 ([20, Proposition 2.1]) *Suppose that the conditions (i) and (ii) of Assumption 2.1 are satisfied. After some permutations of the states, the matrix $\mathbf{G}$ is of the form:*

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{\bullet} & \mathbf{O} \\ \mathbf{G}_{\mathbb{T}, \bullet} & \mathbf{G}_{\mathbb{T}} \end{pmatrix}, \quad (4)$$

where $\mathbf{G}_{\bullet}$ is irreducible and $\mathbf{G}_{\mathbb{T}}$ is strictly lower-triangular.

We assume, without loss of generality, that $\mathbf{G}$ is of the form (4). It then follows from Proposition 3.5 that $\mathbf{G}$ has the unique stationary distribution vector, denoted by $\mathbf{g}$. We now partition $\mathbf{P}$ in the following manner:

$$\mathbf{P} = \begin{matrix} & \mathbb{L}(0) & \bar{\mathbb{L}}(0) \\ \begin{matrix} \mathbb{L}(0) \\ \bar{\mathbb{L}}(0) \end{matrix} & \begin{pmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{H} \end{pmatrix} \end{matrix},$$

where

$$\begin{aligned}\mathbf{T} &= \mathbf{B}(0), & \mathbf{U} &= (\mathbf{B}(1) \quad \mathbf{B}(2) \quad \mathbf{B}(3) \quad \cdots), \\ \mathbf{V} &= \begin{pmatrix} \mathbf{B}(-1) \\ \mathbf{B}(-2) \\ \mathbf{B}(-3) \\ \vdots \end{pmatrix}, & \mathbf{H} &= \begin{pmatrix} \mathbf{A}(0) & \mathbf{A}(1) & \mathbf{A}(2) & \cdots \\ \mathbf{A}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \cdots \\ \mathbf{A}(-2) & \mathbf{A}(-1) & \mathbf{A}(0) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.\end{aligned}$$

The above partition plays a role in solving the matrix equations in (3). We also define $\widehat{\mathbf{H}} := (\widehat{H}(k, i; \ell, j))_{(k,i),(\ell,j) \in \bar{\mathbb{L}}(0)}$ as the fundamental matrix of $\mathbf{H}$, that is,

$$\widehat{\mathbf{H}} = \sum_{m=0}^{\infty} \mathbf{H}^m = (\mathbf{I} - \mathbf{H})^{-1}.$$

Furthermore, we denote the (k, ℓ) th block of $\widehat{\mathbf{H}}$ by $\widehat{\mathbf{H}}(k; \ell) := (\widehat{H}(k, i; \ell, j))_{(i,j) \in \mathbb{M}_1 \times \mathbb{M}_1}$. Note that the (i, j) th entry $\widehat{H}(k, i; \ell, j)$ of $\widehat{\mathbf{H}}(k; \ell)$ is the expected number of visits to state $(\ell, j) \in \bar{\mathbb{L}}(0)$ before entering $\mathbb{L}(0)$, given that the Markov process starts in state $(k, i) \in \bar{\mathbb{L}}(0)$. We then have the following result.

Proposition 3.6 ([32, Theorem 9]) *The block matrices $\widehat{\mathbf{H}}(k; \ell)$, $k, \ell \in \mathbb{N}$, are determined by the recursion:*

$$\widehat{\mathbf{H}}(k; \ell) = \begin{cases} \mathbf{G}\widehat{\mathbf{H}}(k-1; \ell), & k > \ell, \\ \widehat{\mathbf{H}}(1; 1) + \mathbf{G}\widehat{\mathbf{H}}(k-1; k), & k = \ell \geq 2, \\ \sum_{n=1}^{\ell-1} \widehat{\mathbf{H}}(k; n)\mathbf{R}(\ell-n), & k < \ell, \end{cases}$$

where $\widehat{\mathbf{H}}(1; 1) = (\mathbf{I} - \Phi(0))^{-1}$ and $\Phi(0) = \sum_{m=0}^{\infty} \mathbf{A}(m)\mathbf{G}^m$.

We need some more definitions to describe the matrix-analytic expression of $\mathbf{D}$. We define $\boldsymbol{\mu}(k) = (\mu_i(k))_{i \in \mathbb{M}_1}$, $k \in \mathbb{N}$, as the column vector such that $\mu_i(k)$ is the mean first passage time to a state in $\mathbb{L}(0)$, given that the M/G/1-type Markov chain $\{(X_\nu, J_\nu)\}$ starts with state $(k, i) \in \mathbb{L}(k)$. It then follows that

$$\boldsymbol{\mu}(k) = \{\mathbf{I} - \mathbf{G}^{k-1} + (k-1)\mathbf{e}\mathbf{g}\} \{\mathbf{I} - \mathbf{A} + (\mathbf{e} - \boldsymbol{\alpha})\mathbf{g}\}^{-1} \mathbf{e} + \mathbf{G}^{k-1} \boldsymbol{\mu}(1), \quad k \in \mathbb{N},$$

where $\boldsymbol{\alpha} = \sum_{m=0}^{\infty} (m+1)\mathbf{A}(m)\mathbf{e}$. In addition, we define $\mathbf{P}^{(0)}$ as

$$\mathbf{P}^{(0)} = \mathbf{B}(0) + \mathbf{U}\widehat{\mathbf{H}}\mathbf{V}.$$

Note here that $\mathbf{P}^{(0)}$ is the transition probability matrix of a censored Markov chain with censoring set $\mathbb{L}(0)$ ([31, Theorem 2]). Note also that the stationary distribution vector of $\mathbf{P}^{(0)}$ is given by $\mathbf{x}^{(0)} = (\mathbf{x}(0)\mathbf{e})^{-1}\mathbf{x}(0)$. Furthermore, let

$$(\mathbf{I} - \mathbf{P}^{(0)})^\# = (\mathbf{I} - \mathbf{P}^{(0)} + \mathbf{e}\mathbf{x}^{(0)})^{-1} - \mathbf{e}\mathbf{x}^{(0)},$$

which is the *group inverse* of $\mathbf{I} - \mathbf{P}^{(0)}$ (see [27]).

The following theorem presents a matrix-analytic expression of the deviation matrix $\mathbf{D}$.

Theorem 3.7 *If Assumptions 2.1 and 2.2 hold, then the deviation matrix $\mathbf{D}$ exists and*

$$\mathbf{D} = (\mathbf{I} - \mathbf{e}\mathbf{x})\mathbf{K},$$

where $\mathbf{K} := (K(k, i; \ell, j))_{(k, i), (\ell, j) \in \mathbb{F}}$ is a matrix such that $\mathbf{K}(k; \ell) := (K(k, i; \ell, j))_{(i, j) \in \mathbb{M}_{k \wedge 1} \times \mathbb{M}_{\ell \wedge 1}}$, $k, \ell \in \mathbb{Z}_+$, is determined by

$$\begin{aligned} \mathbf{K}(0; 0) &= (\mathbf{I} - \mathbf{P}^{(0)})^\# \left\{ \mathbf{I} - \left(\mathbf{e} + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mathbf{B}(m)\widehat{\mathbf{H}}(m; n)\mathbf{e} \right) \mathbf{x}(0) \right\}, \\ \mathbf{K}(0; \ell) &= (\mathbf{I} - \mathbf{P}^{(0)})^\# \left\{ \sum_{m=1}^{\infty} \mathbf{B}(m)\widehat{\mathbf{H}}(m; \ell) - \left(\mathbf{e} + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mathbf{B}(m)\widehat{\mathbf{H}}(m; n)\mathbf{e} \right) \mathbf{x}(\ell) \right\}, \quad \ell \in \mathbb{N}, \\ \mathbf{K}(k; 0) &= \mathbf{G}^{k-1} \{\mathbf{I} - \Phi(0)\}^{-1} \mathbf{B}(-1)\mathbf{K}(0; 0) - \boldsymbol{\mu}(k)\mathbf{x}(0), \quad k \in \mathbb{N}, \\ \mathbf{K}(k; \ell) &= \widehat{\mathbf{H}}(k; \ell) + \mathbf{G}^{k-1} \{\mathbf{I} - \Phi(0)\}^{-1} \mathbf{B}(-1)\mathbf{K}(0; \ell) - \boldsymbol{\mu}(k)\mathbf{x}(\ell), \quad k, \ell \in \mathbb{N}. \end{aligned}$$

4. Geometric convergence

In this section, we discuss the geometric convergence of the sequence $\{\mathbf{x}^{(N)}; N \in \mathbb{N}\}$ under Assumptions 2.1 and 2.2. It follows from (2) and Theorem 3.7 that, for $k \in \mathbb{Z}_+^{\leq N} := \{0, 1, \dots, N\}$,

$$\begin{aligned} \mathbf{x}^{(N)}(k) - \mathbf{x}(k) &= \mathbf{x}^{(N)}(0) \left[\sum_{n=N+1}^{\infty} \mathbf{B}(n)\mathbf{S}(N, n, k) + \frac{1}{-\sigma} \overline{\mathbf{B}}(N-1)\mathbf{e}\mathbf{x}(k) \right] \\ &\quad + \sum_{\ell=1}^N \mathbf{x}^{(N)}(\ell) \left[\sum_{n=N-\ell+1}^{\infty} \mathbf{A}(n)\mathbf{S}(N, n+\ell, k) + \frac{1}{-\sigma} \overline{\mathbf{A}}(N-\ell-1)\mathbf{e}\mathbf{x}(k) \right], \quad (5) \end{aligned}$$

where

$$\overline{\mathbf{A}}(n) = \sum_{m=n+1}^{\infty} \mathbf{A}(m), \quad \overline{\mathbf{B}}(n) = \sum_{m=n+1}^{\infty} \mathbf{B}(m), \quad n \in \mathbb{Z}_+ \cup \{-1\},$$

and

$$\begin{aligned} \mathbf{S}(N, n, k) &= (\mathbf{G}^{N-k} - \mathbf{G}^{n-k})\widehat{\mathbf{H}}(k; k) + (\mathbf{G}^{N-1} - \mathbf{G}^{n-1}) \{\mathbf{I} - \Phi(0)\}^{-1} \mathbf{B}(-1)\mathbf{K}(0; k) \\ &\quad + (\mathbf{G}^{N-1} - \mathbf{G}^{n-1}) \left[\{\mathbf{I} - \mathbf{A} + (\mathbf{e} - \boldsymbol{\alpha})\mathbf{g}\}^{-1} \mathbf{e} - \boldsymbol{\mu}(1) \right] \mathbf{x}(k). \end{aligned}$$

Using (5), we obtain the following lemma, which provides a sufficient condition for the convergence of $\{\mathbf{x}^{(N)}; N \in \mathbb{N}\}$ to $\mathbf{x}$.

Lemma 4.1 *If Assumptions 2.1 and 2.2 hold, then $\mathbf{x}^{(N)}(k) \geq \mathbf{x}(k)$ for all $k \in \mathbb{Z}_+^{\leq N}$ and*

$$\lim_{N \rightarrow \infty} \mathbf{x}^{(N)} = \mathbf{x}.$$

For $z \in \mathbb{C}$, let $\hat{\mathbf{A}}(z) := \sum_{k=-1}^{\infty} z^k \mathbf{A}(k)$ and $\hat{\mathbf{B}}(z) := \sum_{k=1}^{\infty} z^k \mathbf{B}(k)$ with convergence radii r_A and r_B , respectively. We then assume the following.

Assumption 4.2 It holds that $r_A > 1$ and $r_B > 1$.

Assumption 4.3 There exists a solution of $\det(\mathbf{I} - \hat{\mathbf{A}}(z)) = 0$ for $z \in (1, r_A)$.

Remark 4.4 If Assumption 4.2 holds, then $\mathbf{x}$ is light-tailed [22, Theorem 2] and Assumption 2.2 is satisfied. Thus, it follows from Theorems 3.3 and 3.7 that if Assumptions 2.1 and 4.2 hold then the deviation matrix $\mathbf{D}$ exists and (5) holds.

Remark 4.5 There exists $a \in (1, r_A)$ such that $\delta(y) < 1$ for all $y \in (1, a)$, where $\delta_A(z)$ denotes the Perron-Frobenius eigenvalue of $\hat{\mathbf{A}}(z)$. Furthermore, $\delta_A(1) = 1$, and $\delta_A(y)$ is continuous for $y \in (1, r_A)$. Therefore, if $\lim_{y \nearrow r_A} \hat{\mathbf{A}}(y) > 1$, then Assumption 4.3 is satisfied.

Let θ denote

$$\begin{aligned} \theta &= \sup\{y \in (1, r_A); \det(\mathbf{I} - \hat{\mathbf{A}}(y)) = 0\} \\ &= \sup\{y \in (1, r_A); \delta_A(y) = 1\}. \end{aligned}$$

Let $\mathbf{v}_A(z) > \mathbf{0}$ denote a right eigenvector of $\hat{\mathbf{A}}(z)$ corresponding to eigenvalue $\delta_A(z)$. It is known (see [30, Lemmas 2.3 and 2.4]) that if Assumptions 2.1, 4.2 and 4.3 are satisfied then there exist $K \in \mathbb{N}$, $\gamma \in (0, 1)$ and $b \in (0, \infty)$ such that

$$\mathbf{P}^{(N)} \mathbf{v} \leq \mathbf{P} \mathbf{v} \leq \gamma \mathbf{v} + b \mathbf{1}_{\underline{\mathbb{L}}(K)}, \quad N \in \mathbb{N}, \quad (6)$$

where $\mathbf{v} := (v(\ell, i))_{(\ell, i) \in \mathbb{F}}$ is given by

$$\mathbf{v}(\ell) := (v(\ell, i))_{i \in \mathbb{M}_{\ell \wedge 1}} = \begin{cases} \zeta \mathbf{e}, & \ell = 0, \\ \zeta \eta^\ell \mathbf{v}_A(\eta), & \ell \in \mathbb{N}, \end{cases}$$

with $\eta \in (1, \theta \wedge r_B)$ and $\zeta = (1 - \gamma)^{-1}$. Pre-multiplying both sides of (6) by $\mathbf{x}^{(N)}$, we have $\mathbf{x}^{(N)} \mathbf{v} \leq b/(1 - \gamma) < \infty$. Thus, we obtain the following lemma.

Lemma 4.6 *If Assumptions 2.1, 4.2 and 4.3 hold, then*

$$\lim_{N \rightarrow \infty} \sum_{\ell=1}^N \eta^\ell \mathbf{x}^{(N)}(\ell) < \infty, \quad \eta \in (1, \theta \wedge r_B).$$

For any functions $f(n)$ and $g(n)$, matrix-functions $\mathbf{M}(n) = (M_{i,j}(n))_{i,j}$ and $\mathbf{N}(n) = (N_{i,j}(n))_{i,j}$ of the same order, we introduce the asymptotic notation:

$$\begin{aligned} f(n) &\overset{n}{\sim} g(n) \iff \lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1, \\ f(n) &\overset{n}{\asymp} g(n) \iff \liminf_{n \rightarrow \infty} \frac{f(n)}{g(n)} > 0 \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{f(n)}{g(n)} < \infty, \\ \mathbf{M}(n) &\overset{n}{\sim} \mathbf{N}(n) \iff M_{i,j}(n) \overset{n}{\sim} N_{i,j}(n), \quad \text{for all } i, j, \\ \mathbf{M}(n) &\overset{n}{\asymp} \mathbf{N}(n) \iff M_{i,j}(n) \overset{n}{\asymp} N_{i,j}(n), \quad \text{for all } i, j. \end{aligned}$$

We now make an assumption on $\mathbf{A}(n)$ and $\mathbf{B}(n)$ in order to discuss the geometric convergence rate of $\mathbf{x}^{(N)} - \mathbf{x}$.

Assumption 4.7 For some $k_A, k_B \in \mathbb{Z}$,

$$\mathbf{A}(n) \stackrel{n}{\sim} n^{k_A} r_A^{-n} \mathbf{C}_A, \quad \mathbf{B}(n) \stackrel{n}{\sim} n^{k_B} r_B^{-n} \mathbf{C}_B,$$

where $\mathbf{C}_A$ and $\mathbf{C}_B$ are nonnegative matrices.

Lemma 4.8 If Assumptions 4.2 and 4.7 are satisfied, then

$$\overline{\mathbf{A}}(n) \stackrel{n}{\asymp} n^{k_A} r_A^{-n} \mathbf{C}_A \stackrel{n}{\sim} \mathbf{A}(n), \quad \overline{\mathbf{B}}(n) \stackrel{n}{\asymp} n^{k_B} r_B^{-n} \mathbf{C}_B \stackrel{n}{\sim} \mathbf{B}(n),$$

and thus $\overline{\overline{\mathbf{A}}}(n) \stackrel{n}{\asymp} \mathbf{A}(n)$ and $\overline{\overline{\mathbf{B}}}(n) \stackrel{n}{\asymp} \mathbf{B}(n)$.

Finally, using (5), Lemmas 4.1, 4.6 and 4.8, we obtain the main results of this paper.

Theorem 4.9 Suppose that Assumptions 2.1, 4.2, 4.3 and 4.7 are satisfied. The statements (i) and (ii) below hold:

(i) If $r_B \leq \theta$, then

$$\lim_{N \rightarrow \infty} \frac{\mathbf{x}^{(N)} - \mathbf{x}}{\alpha^{-N}} = \mathbf{0}, \quad \alpha \in (1, r_B).$$

(ii) If $\theta < r_B$, then

$$\lim_{N \rightarrow \infty} \frac{\mathbf{x}^{(N)} - \mathbf{x}}{\alpha^{-N}} = \mathbf{0}, \quad \alpha \in (1, \theta).$$

In addition, if $\mathbf{G}$ is aperiodic, then the statements (iii) and (iv) below hold:

(iii) If $r_B \leq \theta$, then

$$\lim_{N \rightarrow \infty} \frac{|\mathbf{x}^{(N)} - \mathbf{x}|}{\beta^{-N}} = \infty, \quad \beta > r_B.$$

(iv) If $\theta < r_B$, then

$$\lim_{N \rightarrow \infty} \frac{|\mathbf{x}^{(N)} - \mathbf{x}|}{\beta^{-N}} = \infty, \quad \beta > \max(\theta, \lambda_2),$$

where λ_2 denotes the second largest eigenvalue of $\mathbf{G}$.

References

- [1] N. Akar and E. Arikan, A numerically efficient method for the MAP/D/1/K queue via rational approximations, *Queueing Systems*, **22**(1), 97–120, 1996.
- [2] N. Akar, N. C. Oğuz and K. Sohraby, Matrix-geometric solutions of M/G/1-type Markov chains: a unifying generalized state-space approach, *IEEE Journal on Selected Areas in Communications*, **16**(5), 626–639, 1998.
- [3] S. Asmussen, *Applied Probability and Queues, Second edition*, New York: Springer, 2003.
- [4] A. Baiocchi and N. Blefari-Melazzi, Steady-state analysis of the MMPP/GI/1/K queue, *IEEE Transactions on Communications*, **41**(4), 531–534, 1993.
- [5] A. Baiocchi, Analysis of the loss probability of the MAP/G/1/K queues, Part I: Asymptotic theory, *Stochastic Models*, **10**(4), 867–893, 1994.
- [6] A. Baiocchi and N. Blefari-Melazzi, Analysis of the loss probability of the MAP/G/1/K queues, Part II: Approximations and numerical results, *Stochastic Models*, **10**(4), 895–925, 1994.
- [7] A. Cuyt, R. B. Lenin, G. Willems, C. Blondia and P. Rousseeuw, Computing packet loss probabilities in multiplexer models using rational approximation, *IEEE Transactions on Computers*, **52**(5), 633–644, 2003.
- [8] S. Dendievel, G. Latouche and Y. Liu, Poisson’s equation for discrete-time quasi-birth-and-death processes, *Performance Evaluation*, **70**(9), 564–577, 2013.

- [9] A. N. Dudin, A. A. Shaban and V. I. Klimenok, Analysis of a queue in the BMAP/G/1/N system, *International Journal of Simulation Systems, Science & Technology*, **6**(1–2), 13–23, 2005.
- [10] S. Foss, D. Korshunov and S. Zachary, *An Introduction to Heavy-Tailed and Subexponential Distributions*, New York: Springer, 2011.
- [11] F. N. Gouweleeuw, Calculating the loss probability in a BMAP/G/1/N + 1 queue, *Stochastic Models*, **12**(3), 473–492, 1996.
- [12] F. N. Gouweleeuw and H. C. Tijms, Computing loss probabilities in discrete-time queues, *Operations Research*, **46**(1), 149–154, 1998.
- [13] B. Heidergott, A. Hordijk and N. Leder, Series expansions for continuous-time Markov processes, *Operations Research*, **58**(3), 756–767, 2010.
- [14] C. Herrmann, The complete analysis of the discrete time finite DBMAP/G/1/N queue, *Performance Evaluation*, **43**(2), 95–121, 2001.
- [15] F. Ishizaki and T. Takine, Loss probability in a finite discrete-time queue in terms of the steady state distribution of an infinite queue, *Queueing Systems*, **31**(3–4), 317–326, 1999.
- [16] F. Ishizaki and T. Takine, Cell loss probability approximations and their application to call admission control, *Advances in Performance Analysis*, **2**(3), 225–258, 1999.
- [17] B. Kim and J. Kim, Explicit solution for the stationary distribution of a discrete-time finite buffer queue, *Journal of health organization and management*, **12**(3), 1121–1133, 2016.
- [18] Y. Katsumata, H. Masuyama and Y. Takahashi, Subgeometric convergence of finite-level M/G/1-type Markov chains, *Dai 32-kai machigyousetsu-sinpojiumu houbunshu*, 96–103, 2015 (in Japanese).
- [19] J. Kim and B. Kim, Asymptotic analysis for loss probability of queues with finite GI/M/1 type structure, *Queueing Systems*, **57**(1), 47–55, 2007.
- [20] T. Kimura, K. Daikoku, H. Masuyama and Y. Takahashi, Light-tailed asymptotics of stationary tail probability vectors of Markov chains of M/G/1 type, *Stochastic Models*, **26**(4), 505–548, 2010.
- [21] H. Li and Y. Q. Zhao, Stochastic block-monotonicity in the approximation of the stationary distribution of infinite Markov chains, *Stochastic Models*, **16**(2), 313–333, 2000.
- [22] Q. -L. Li and Y. Q. Zhao, Light-tailed asymptotics of stationary probability vectors of Markov chains of GI/G/1 type, *Advances in Applied Probability*, **37**(4), 1075–1093, 2005.
- [23] H. Masuyama, Error bounds for augmented truncations of discrete-time block-monotone Markov chains under geometric drift conditions, *Advances in Applied Probability*, **47**(1), 83–105, 2015.
- [24] H. Masuyama, Error bounds for augmented truncations of discrete-time block-monotone Markov chains under subgeometric drift conditions, *SIAM Journal on Matrix Analysis and Applications*, **37**(3), 877–910, 2016.
- [25] H. Masuyama, Continuous-time block-monotone Markov chains and their block-augmented truncations, *Linear Algebra and Its Applications*, **514**, 105–150, 2017.
- [26] H. Masuyama, Error bounds for last-column-block-augmented truncations of block-structured Markov chains, To appear in *Journal of the Operations Research Society of Japan*, **60**(3), 2017.
- [27] C. D. Meyer, Jr., The role of the group generalized inverse in the theory of finite Markov chains, *SIAM Review*, **17**(3), 443–464, 1975.
- [28] M. Miyazawa, Y. Sakuma and S. Yamaguchi, Asymptotic behaviors of the loss probability for a finite buffer queue with QBD structure, *Stochastic Models*, **23**(1), 79–95, 2007.
- [29] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and their Applications*, New York: Marcel Dekker, 1989.
- [30] S. Sakaiya, Error bounds for jump-to-the-boundary-level-extended block-banded truncations of GI/G/1-Type Markov chains with geometric ergodicity, *Department of Systems Science, Graduate School of Informatics, Kyoto University*, Master’s thesis, 2016.

- [31] Y. Q. Zhao, Censoring technique in studying block-structured Markov chains, *Advances in Algorithmic Methods for Stochastic Models*, eds G. Latouche and P. G. Taylor, Notable Publications, 417–433, 2000.
- [32] Y. Q. Zhao, W. Li and W. J. Braun, Censoring, factorizations, and spectral analysis for transition matrices with block-repeating entries, *Methodology and Computing in Applied Probability*, **5**(1), 35–58, 2003.

セルラネットワークの空間確率モデルに対する ポアソン点過程近似について

*高橋雄太 三好直人
東京工業大学

概要 セルラネットワークに対して、基地局の位置を空間点過程を用いてモデル化し性能評価を行うことを考える。性能評価指標のひとつとして、ユーザーと基地局の通信可能性を表す被覆確率がある。一般の空間点過程を用いた場合、解析結果が複雑になり数値計算できたとしても非常に計算時間がかかる場合がある。本研究では、一般の空間点過程を用いたモデルに対して、ポアソン点過程近似と呼ばれる近似手法を用いて被覆確率を数値計算する。また、この近似に必要な「ギャップ」と呼ばれる値自体が複雑な形になる場合があるのでそれ自体の近似を考える。これにより被覆確率を解析的に求めるのが困難な場合でも、原点から一番近い点までの距離の分布を用いて近似を行うことが出来る。また、 α -ジニブル点過程を用いて数値実験を行い近似手法の有効性を検証する。

1. 序文

無線通信ネットワークに確率幾何モデルを適用し、性能評価を行う研究が活発に行われている。これは、受けるべき信号の強さが無線の配置に大きく依存するからである。たとえば、無線通信ネットワークの性能評価指標として被覆確率があるが、これはノード間の距離による信号の減衰や受信ノードの位置でのノイズを考慮した信号対干渉比がある閾値を超える確率であり、通信可能性を表す指標である。被覆確率を大きくするためには、無線ノードが規則的にバランスよく配置されていることが理想的であるが、地理的な問題などにより実際には不規則に配置されているとみなせる。この無線ノードの配置の不規則さを空間点過程を用いて表現できるところに確率幾何モデルのメリットがある。

セルラネットワークにおいても、Andrews *et al.* [1] では無線基地局の配置が Poisson 点過程に従ったモデルが提案され、被覆確率などの評価が行われている。しかしこの文献にも記述されているように Poisson 点過程では点同士の相関を考慮できないため、バランスを考えて配置されている基地局を表すモデルとして適していないという問題がある。この問題に対して Miyoshi and Shirai [2] では点同士が負の相関を持つ Ginibre 点過程を用いたモデルに対して解析を行っている。

一般の空間点過程を用いたモデルに対して、被覆確率の近似を行っているのが Haenggi [3] である。この論文では、比較的容易に数値計算できるポアソン点過程を用いたモデルを利用した近似手法（ポアソン点過程近似）を提案している。この手法では、ポアソン点過程を用いたモデルと一般の点過程を用いたモデルの被覆確率が平行移動させることによってほぼ一致することを用いて近似を行っている。この、近似では平行移動する量（ギャップ）が重要な役割を果たす。Wei *et al.* [4], [5] では、複数種類の基地局を用いた無線通信ネットワークにおいても同様のギャップを用いて近似が行えることを示している。そのため、ギャップの算出はより重要性を増すが、これらの論文ではギャップがシミュレーションで求められているだけである。

本研究では、ポアソン点過程近似を用いるために必要なギャップを一般の点過程、伝播損失関数に対して算出した。ギャップ自体の数値計算に時間がかかる場合があるのでギャップ自体の近似も提案した。この近似では一般の点過程の原点から一番近い点までの距離の分布を用いてギャップを表現することが出来る。

本研究の結果により、一番近い点までの距離の分布さえ分かれば計算時間を大きく短縮することができるようになったといえる。本稿の構成は次のとおりである。2 節で扱うモデルを述べ、3 節でポアソン点過程近似の説明とギャップの導出、その近似を行う。また、計算例として α -ジニブル点過程を用いた。4 節では今回の近似を適用した結果と解析的な結果の比較を行い、近似手法の有効性を検証する。

2. 確率幾何モデル

この節ではセルラネットワークを空間確率モデルを用いてモデル化し、性能評価指標である被覆確率を導入する。

$\Phi = \{X_i\}_{i=1}^{\infty}$ を基地局の配置を表す点過程とし、各基地局の通信電力は $P_i = P$ (一定) とする。ここでは Φ は単純、局所有限、定常な点過程とする。ユーザーは一番近い基地局 B_o と通信すると仮定する。つまり、基地局の Voronoi セル内のユーザーはその基地局と通信する。以下では点過程の定常性から、原点にいるユーザーを考える。

F_i は基地局 i のフェージングを表し、Rayleigh フェージングを仮定すると平均 1 の独立な指数分布となる。また、距離に応じた伝播損失関数 ℓ は単調減少関数とする。この仮定では、原点のユーザーが他の基地局から受ける累積干渉は $I_o = P \sum_{j \neq B_o} F_j \ell(|X_j|)$ と表現できる。 I_o を用いると、信号対干渉比 SIR は以下のように表せる。

$$SIR = \frac{PF_o \ell(|X_{B_o}|)}{I_{B_o}}$$

SIR がある閾値 θ を超えたとき原点にいるユーザーは基地局と通信できると仮定する。この確率を被覆確率とよび、 $p(\beta, \theta) = P(SIR > \theta)$ とする。このモデルでは被覆確率は以下のように表現できる.[2]

$$\begin{aligned} p(\beta, \theta) &= P(SIR > \theta) \\ &= E\left[\prod_{j \neq B_o} \left(1 + \theta \frac{\ell(X_j)}{\ell(X_{B_o})}\right)^{-1}\right] \end{aligned}$$

3. ポアソン点過程の被覆確率を用いた近似

3.1. ポアソン点過程近似

被覆確率に対するポアソン点過程近似とは、Haenggi [3] で提案されたセルラネットワークの下りリンクにおける被覆確率の近似方法である。この方法では図 1 のように、一般の点過程を用いて算出した被覆確率がポアソン点過程を用いて算出した被覆確率をある量だけ平行移動させることによって近い関数を得られる事を生かして近似を行っている。ポアソン点過程を用いたモデルの被覆確率と一般の点過程を用いた被覆確率の平行移動する量を G (ギャップ) とし、 p を一般の点過程を用いた被覆確率、 $p^{(ppp)}$ をポアソン点過程を用いた被覆確率とすると以下を満たすような G を用いて近似を行う。

$$\frac{p(G\theta)}{p^{(ppp)}(\theta)} \rightarrow 1, \theta \downarrow 0$$

G を決める際に様々な θ を用いることが出来るが、 θ が 0 付近の被覆確率をより厳密に評価したいため上記のような G を考える。また、分子と分母をテイラー展開すると以下が得られる。

$$\begin{aligned} \frac{p(G\theta)}{p^{(ppp)}(\theta)} &= \frac{1 + p'(0)G\theta + o(\theta)}{1 + p'^{(ppp)}(0)\theta + o(\theta)} \\ &\approx \frac{1 + p'(0)G\theta}{1 + p'^{(ppp)}(0)\theta} \end{aligned}$$

よって、 G は以下のとおりに表せる。

$$G = \frac{p'^{(ppp)}(0)}{p'(0)}$$

したがって、以下の形で近似を行うことが出来る。

$$p(G\theta) \approx p^{(ppp)}(\theta), \quad G = \frac{p^{(ppp)}(0)}{p'(0)}$$

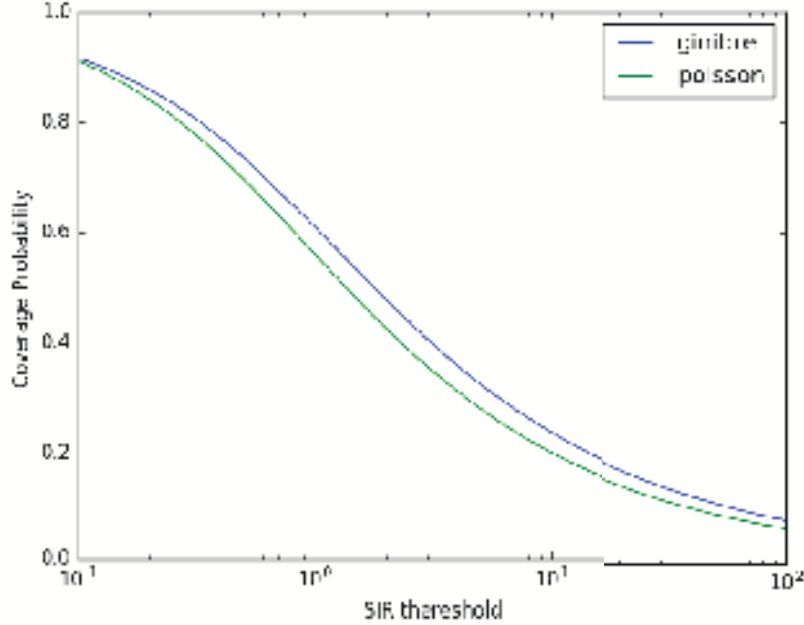


図 1 被覆確率

3.2. 一般の点過程に対するマクローリン係数

3.1 節でポアソン点過程近似について説明したがこの近似を用いるには G を算出する必要がある。 G を算出するには p と $p^{(ppp)}$ の一次のマクローリン係数が必要である。したがって、ここではマクローリン係数を求める。

2 節のモデルで、 $p(\beta, \theta) = E[\prod_{j \neq B_o} (1 + \theta \frac{\ell(|X_j|)}{\ell(|X_{B_o}|)})^{-1}]$ であったのでこの p について一次のマクローリン係数を考える。その際に用いる定理を 2 つ挙げる。[6]

定理 1 $\Phi = \{X_i\}_{i \in \mathbb{N}}$ を点過程とする。 $B_s = \{x \in \mathbb{R}^2; |x| \leq \ell^{-1}(\ell(|X_{B_o}|)/s)\}$ とすると、被覆確率は以下のように表現できる。

$$p(\beta, \theta) = E[\exp(-\theta \int_1^\infty \frac{\Phi(B_s) - 1}{s(\theta + s)} ds)]$$

定理 2 (一般の空間点過程におけるマクローリン係数) ℓ を狭義単調減少関数とすると、空間点過程 Φ に対する被覆確率が、

$$p(\beta, \theta) = E[\exp(-\theta \int_1^\infty \frac{\Phi(B_s) - 1}{s(\theta + s)} ds)]$$

のとき、その一次のマクローリン係数は以下で与えられる。

$$p'(0) = - \int_1^\infty \frac{E[\Phi(B_s) - 1]}{s^2} ds$$

ただし、 $B_s = \{x \in \mathbb{R}; |x| \leq \ell^{-1}(\ell(|X_{B_o}|)/s)\}$ とする。

ここで、一次のマクローリン係数が複雑になる場合に対しての近似を考える。

提案 1 (一般の空間点過程におけるマクローリン係数の近似) ℓ を狭義単調減少関数とすると、空間点過程 Φ に対する被覆確率が、

$$E\left[\prod_{j \neq B_o} \left(1 + \theta \frac{\ell(X_j)}{\ell(X_{B_o})}\right)^{-1}\right] = E\left[\exp\left(-\theta \int_1^\infty \frac{\Phi(B_s) - 1}{s(\theta + s)} ds\right)\right]$$

のとき、その一次のマクローリン係数の近似は以下で与えられる。

$$p'(0) \approx -\lambda\pi \int_1^\infty \int_0^\infty \frac{(\ell^{-1}(\frac{\ell(R)}{s}) - R^2)}{s^2} P(|X_{B_o}| \in dR) ds$$

ただし、 $B_s = \{x \in \mathbb{R}; |x| \leq \ell^{-1}(\ell(|X_{B_o}|)/s)\}$ とする。

$$\begin{aligned} p'(0) &= -\int_1^\infty \frac{E[\Phi(B_s) - 1]}{s^2} ds \\ &= -\int_1^\infty \int_0^\infty \frac{E[\Phi(|X_{B_o}|, \ell^{-1}(\frac{\ell(|X_{B_o}|)}{s}) | X_{B_o} = R)]}{s^2} P(|X_{B_o}| \in dR) ds \\ &\approx -\lambda\pi \int_1^\infty \int_0^\infty \frac{(\ell^{-1}(\frac{\ell(R)}{s}) - R^2)}{s^2} P(|X_{B_o}| \in dR) ds \end{aligned}$$

3 番目の式は一番近くの点が他の点と独立である場合を近似式とした。

特に、 $\ell(x) = x^{-2\beta}$ の時は s に依存しない近似式となる。

$$p'(0) \approx \frac{\lambda\pi}{(\beta - 1)} \int_0^\infty R^2 P(|X_{B_o}| \in dR)$$

3.3. ポアソン点過程近似の適用例 (α -ジニブル点過程)

この節では、 α -ジニブル点過程について簡単に紹介し、ポアソン点過程近似を用いるために一次のマクローリン係数の計算例を挙げる。

3.3.1. α -ジニブル点過程

これらの点過程は $\mathbb{C} \simeq \mathbb{R}$ 上の行列式点過程の一つである。

定義 1 (結合強度 [7]) Φ を $\mathbb{R}^d$ 上の単純な点過程とし、 μ を $(R^d, \mathcal{B}(R^d))$ 上の局所有限な測度とする。また、 $\rho^{(n)}$ を点過程 Φ の測度 μ に対する n 次の結合強度 (相関関数) とする。すなわち連続、対称、有界閉集合の台を持つ関数 f に対して以下の等式が成立する。

$$E\left[\sum_{\substack{X_1, \dots, X_n \in \Phi \\ X_i \neq X_j, i \neq j}} f(X_1, X_2, \dots, X_n)\right] = \int \int \dots \int_{R^{d \times n}} f(x_1, x_2, \dots, x_n) \rho^{(n)}(x_1, x_2, \dots, x_n) \prod_{i=1}^n \mu(dx_i)$$

定義 2 (行列式点過程 [7]) 結合強度 ρ^n が以下の等式を満たすとき、点過程 Φ は測度 μ に対して核 $K : R^d \times \mathbb{R}^d \rightarrow \mathbb{C}$ を持つ行列式点過程であるいう。

$$\rho^n(x_1, x_2, \dots, x_n) = \det(K(x_i, x_j))_{1 \leq i, j \leq n}$$

ただし、 $((K(z_i, z_j)))_{1 \leq i, j \leq n}$ は $\mathbb{R}^d \times \mathbb{R}^d$ で連続で正定値エルミート行列である。

定義 3 (α -ジニブル点過程 [7]) $\mathbb{C}$ 上の行列式点過程 Φ が測度 $\mu(dz) = \pi^{-1}e^{-|z|^2}m(dz)$ に対して, 以下のように定義される核 K を持つとき点過程 Φ を α -ジニブル点過程という.

$$K(z, w) = e^{z\bar{w}/\alpha}, \quad z, w \in \mathbb{C}, \quad \alpha \in (0, 1]$$

ただし, m は $\mathbb{C}$ 上のルベーグ測度である.

α -ジニブル点過程は, $\alpha = 1$ のときジニブル点過程と一致し $\alpha \rightarrow 0$ のとき定常ポアソン点過程に弱収束することが知られている.[8] また, α -ジニブル点過程の距離の 2 乗の分布に注目すると以下の命題が示せる.

定理 3 (Kostlan の定理 [7]) $X_i, i \in \mathbb{N}$ が α -ジニブル点過程の点を表すとする. このとき,

$$\{|X_1|^2, |X_2|^2, \dots\} \stackrel{d}{=} \{Y_1, Y_2, \dots\}$$

ただし, $Y_i, i \in \mathbb{N}$ は独立で $Y_i \sim \text{Gamma}(i, \pi\lambda/\alpha)$ である.

系 1 (ジニブル点過程からの α -ジニブル点過程の構成)) $\Phi_{\lambda/\alpha} = \{X_i\}_{i \in \mathbb{N}}$ を強度 λ/α のジニブル点過程とすると, α -ジニブル点過程 Φ は以下のように表せる.

$$\Phi(C) = \sum_{i \in \mathbb{N}} \xi_i 1_C(X_i), \quad C \in \mathcal{B}(\mathbb{C})$$

ただし, $\{\xi_i\}_{i \in \mathbb{N}}$ は互いに独立で $P(\xi_i = 1) = \alpha$ & $P(\xi_i = 0) = 1 - \alpha$ を満たすとする.

3.3.2. α -ジニブル点過程の一次のマクローリン係数

まず, 伝播損失関数が一般の狭義単調減少関数であると仮定して α -ジニブル点過程の一次のマクローリン係数を算出する. そのあと, 具体例としてよく用いられる $\ell(x) = x^{-2\beta}$ と, 原点付近で発散しないより現実に即した $\ell(x) = (1 + x^{2\beta})^{-1}$ を用いた場合を計算例としてあげる.

定理 4 ℓ が狭義単調減少関数で, g を $\ell(x) = g(x^2)$ を満たす関数であるとする. また, p を α -ジニブル点過程の被覆確率とすると, 一次の微分係数は次のように与えられる.

$$p'(0) = -\alpha^2 \sum_{i=1}^{\infty} \sum_{k \neq i} E_{Y_i}[(g(Y_i))^{-1} E_{Y_k}[g(Y_k) 1\{Y_i \leq Y_k\}]] \prod_{j \neq i, k} (1 - \alpha + \alpha \Gamma(j, Y_i))$$

ただし, $\{Y_i\}_{i \in \mathbb{N}}$ は独立で $Y_i \sim \text{Gamma}(i, 1)$ である.

(証明) $\Phi = \{X_i\}_{i \in \mathbb{N}}$ を α -ジニブル点過程 であるとし, $\{\xi_i\}_{i \in \mathbb{N}}$ は互いに独立で $P(\xi_i = 1) = \alpha$ & $P(\xi_i = 0) = 1 - \alpha$ を満たし, $\{Y_i\}_{i \in \mathbb{N}}$ とも互いに独立であるとする.

定理 2 を用いると, $B_s = \{x \in \mathbb{R}^2; |x| \leq g^{-1}(g(|X_{B_o}|^2)/s)\}$ より $\Phi(B_s) - 1 = \Phi(B_s \setminus B_1)$ が成り立つので,

$$\begin{aligned} p'(0) &= - \int_1^\infty \frac{E[\Phi(B_s) - 1]}{s^2} ds \\ &= - \int_1^\infty \frac{E[\Phi(B_s \setminus B_1)]}{s^2} ds \end{aligned}$$

ここで, g は単調減少であり逆関数を持つので,

$$\begin{aligned} E[\Phi(B_s \setminus B_1)] &= E\left[\sum_{k \neq B_o} 1\{|X_{B_o}| < |X_k| \leq g^{-1}\left(\frac{g(|X_{B_o}|^2)}{s}\right)\}\right] \\ &= \sum_{i=1}^{\infty} \sum_{k \neq i} E[1\{|X_i| < |X_k| \leq g^{-1}\left(\frac{g(|X_i|^2)}{s}\right)\} 1\{B_o = i\}] \end{aligned}$$

$|X_i|^2$ が与えられており, $|X_k|^2 \leq Y_k$ が分かっているとき,

$$\{B_o = i\} = \{|X_i|^2 < Y_j, \xi_j = 1, j \in \mathbb{N} \setminus \{i, k\}\} \cup \{\xi_j = 0\}$$

よって,

$$\begin{aligned} & E[1\{|X_i|^2 < |X_k|^2 \leq g^{-1}\left(\frac{g(|X_{B_o}|^2)}{s}\right)\} 1\{B_o = i\} |X_i|^2] \\ &= \alpha P(|X_i|^2 < Y_k \leq g^{-1}\left(\frac{g(|X_{B_o}|^2)}{s}\right)) \prod_{j \neq i, k} (1 - \alpha + \alpha \Gamma(j, |X_i|^2)) \end{aligned}$$

また, s に関する積分を行うと,

$$\begin{aligned} \int_1^\infty \frac{P(|X_i|^2 < |X_k|^2 \leq g^{-1}\left(\frac{g(|X_{B_o}|^2)}{s}\right))}{s^2} ds &= \int_1^\infty \frac{1}{s^2} \int_{|X_i|^2}^{g^{-1}\left(\frac{g(|X_{B_o}|^2)}{s}\right)} \frac{y^{k-1} e^{-y}}{(k-1)!} dy ds \\ &= \int_{|X_i|^2}^\infty \frac{y^{k-1} e^{-y}}{(k-1)!} \int_{\frac{g(|X_i|^2)}{g(y)}}^\infty \frac{1}{s^2} ds dy \\ &= (g(|X_i|^2))^{-1} E_{Y_k} [g(Y_k) 1\{|X_i|^2 \leq Y_k\}] \end{aligned}$$

したがって,

$$\begin{aligned} p'(0) &= -\alpha \sum_{i=1}^\infty \sum_{k \neq i} E_{|X_i|^2} [(g(|X_i|^2))^{-1} E_{Y_k} [g(Y_k) 1\{|X_i|^2 \leq Y_k\}]] \prod_{j \neq i, k} (1 - \alpha + \alpha \Gamma(j, |X_i|^2)) \\ &= -\alpha^2 \sum_{i=1}^\infty \sum_{k \neq i} E_{Y_i} [(g(Y_i))^{-1} E_{Y_k} [g(Y_k) 1\{Y_i \leq Y_k\}]] \prod_{j \neq i, k} (1 - \alpha + \alpha \Gamma(j, Y_i)) \end{aligned}$$

が成り立つ. □

次に先ほどあげた伝播損失関数を用いて計算する. α -ジニブル点過程の被覆確率を p としたとき, $\{Y_i\}_{i \in \mathbb{N}}$ が独立で $Y_i \sim \text{Gamma}(i, 1)$ であるとする. 一次の微分係数の計算例は以下のとおりである.

- $l(x) = x^{-2\beta}$ の場合

$$p'(0) = -\alpha^2 \sum_{i=1}^\infty \sum_{k \neq i} E_{Y_i} [Y_i^\beta E_{Y_k} [Y_k^{-\beta} 1\{Y_i \leq Y_k\}]] \prod_{j \neq i, k} (1 - \alpha + \alpha \Gamma(j, Y_i))$$

- $\ell(x) = (1 + x^{2\beta})^{-1}$ の場合

$$p'(0) = -\alpha^2 \sum_{i=1}^\infty \sum_{k \neq i} E_{Y_i} [(1 + Y_i)^\beta E_{Y_k} [(1 + Y_k)^{-\beta} 1\{Y_i \leq Y_k\}]] \prod_{j \neq i, k} (1 - \alpha + \alpha \Gamma(j, Y_i))$$

3.3.3. α -ジニブル点過程の一次のマクローリン係数の近似

- $\ell(x) = x^{-2\beta}$ の場合

α -ジニブル点過程の一次の微分係数の近似は以下のように表せる.

$$\begin{aligned} p'(0) &\approx \frac{\lambda\pi}{\beta-1} \int_0^\infty R^2 f_{X_{B_o}}(R) dR \\ &= \frac{\lambda\pi}{\beta-1} E[|X_{B_o}|^2] \\ &= \frac{\lambda\pi}{\beta-1} \int_0^\infty P(|X_{B_o}|^2 > R) dR \\ &= \frac{\lambda\pi}{\beta-1} \int_0^\infty \prod_{i \in \mathbb{N}} (1 - \alpha + \alpha \Gamma(i, R)) dR \end{aligned}$$

- $\ell(x) = (1 + x^{2\beta})^{-1}$ の場合

近似を用いるために、一番近い点の距離の分布に対する確率密度関数を考える。

$$\begin{aligned} P(|X_{B_o}| \leq R) &= 1 - P(|X_{B_o}| > R) \\ &= 1 - \prod_{i \in \mathbb{N}} (1 - \alpha + \alpha P(Y_i > R^2)) \end{aligned}$$

したがって、確率密度関数 $f_{X_{B_o}}$ は

$$\begin{aligned} f_{|X_{B_o}|}(R) &= \frac{d}{dR} P(|X_{B_o}| \leq R) \\ &= \alpha \sum_{i \in \mathbb{N}} f_{Y_i}(R^2) \prod_{j \neq i} (1 - \alpha + \alpha P(Y_j > R^2)) \end{aligned}$$

したがって、 α -ジニブル点過程の一次の微分係数の近似は以下のように表せる。

$$\begin{aligned} p'(0) &\approx -\lambda\pi \int_0^\infty \frac{(s(1 + R^{2\beta}) - 1)^{1/\beta} - R^2}{s^2} f_{X_{B_o}}(R) dR \\ &= -\lambda\pi \int_0^\infty \frac{(s(1 + R^{2\beta}) - 1)^{1/\beta} - R^2}{s^2} f_{X_{B_o}}(R) dR \\ &= -\lambda\pi \left(\int_0^\infty \int_{R^{2\beta}}^\infty \frac{u^{1/\beta}}{(1+u)^2} du (1 + R^{2\beta}) f_{|X_{B_o}|}(R) dR - \int_0^\infty R^2 f_{|X_{B_o}|}(R) dR \right) \end{aligned}$$

3.3.4. ポアソン点過程の一次のマクローリン係数

ポアソン点過程の一次の微分係数は以下のとおりである。

$$\begin{aligned} p^{(ppp)}(0) &= - \int_1^\infty \frac{E[\Phi(B_s) - 1]}{s^2} ds \\ &= - \int_1^\infty \int_0^\infty \frac{E[\Phi(B_s \setminus B_1) | |X_{B_o}| = R]}{s^2} dP(|X_{B_o}| \in dR) ds \\ &= - \int_1^\infty \int_0^\infty \frac{\lambda\pi R^2 (s^{1/\beta} - 1)}{s^2} dP(|X_{B_o}| \in dR) ds \\ &= - \int_1^\infty \int_0^\infty \frac{\lambda\pi R^2 (s^{1/\beta} - 1)}{s^2} 2\lambda\pi R e^{-\lambda\pi R^2} ds \\ &= - \int_1^\infty \frac{s^{1/\beta} - 1}{s^2} ds \\ &= \frac{1}{\beta - 1} \end{aligned}$$

4. 数値結果

ここでは、 $\ell(x) = x^{-2}$ に対するポアソン点過程と α -ジニブル点過程の被覆確率の解析結果を紹介する。

4.1. 被覆確率

- Poisson 点過程 [1]

$$\begin{aligned} p(\beta, \theta) &= \frac{1}{1 + \rho(\beta, \theta)} \\ \rho(\beta, \theta) &= \frac{\theta}{\beta} \int_1^\infty \frac{s^{1/\beta}}{s(s + \theta)} \end{aligned}$$

- α -ジニブル点過程 [2]

$$p(\beta, \theta) = \alpha \int_0^\infty e^{-v} M(\alpha, \beta, \theta, v) S(\alpha, \beta, \theta, v) dv$$

$$M(\alpha, \beta, \theta, v) = \prod_{j=1}^{\infty} \left(1 - \alpha + \alpha \frac{1}{(j-1)!} \int_v^\infty \frac{s^{j-1} e^{-s}}{1 + \theta(v/s)^\beta} ds \right)$$

$$S(\alpha, \beta, \theta, v) = \sum_1^{\infty} \frac{v^{i-1}}{(i-1)!} \left(1 - \alpha + \frac{\alpha}{(i-1)!} \int_v^\infty \frac{s^{i-1} e^{-s}}{1 + \theta(v/s)^\beta} ds \right)^{-1}$$

4.2. 精度の考察

$\beta = 2$ として α -Ginibre 点過程に対して解析結果とポアソン点過程近似の結果を比較する.

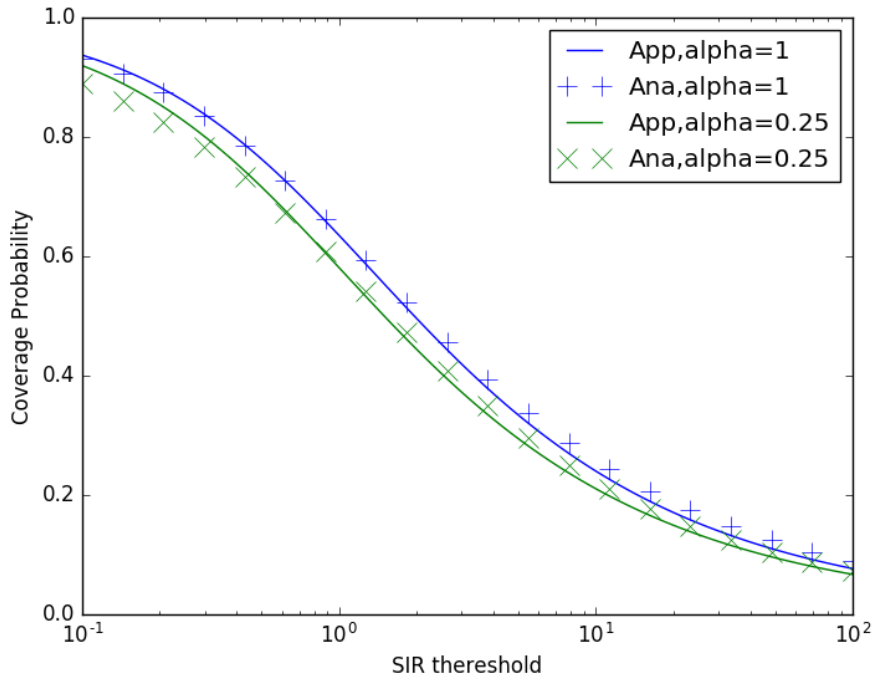


図 2 被覆確率

5. 結論と課題

被覆確率のポアソン点過程近似を行い、精度の考察を行った. α -ジニブル点過程における被覆確率の計算時間は近似を用いる事で 1/100 程度になることを確認した. 一般の定常空間点過程に対して原点から一番近い点の距離の分布を用いて近似できることを示した. このことにより, 一番近い点の距離の分布が分かれば被覆確率を近似することができる.

今後の課題としては, 他の空間点過程の原点から一番近い距離の分布を算出して実際に近似を行うとどの程度の精度で近似が行えるかシミュレーションと比較することにより確かめたい. また, 一番近くの点の距離の分布を算出することも煩雑になる場合があるので一番近い点の距離の分布自体の近似を考えることも意義があると考えられる. また, 近似の精度評価を数式上で行いたい.

参考文献

- [1] J. G. Andrews, F. Baccelli and R. K. Ganti. (2011). A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications*, 59(11), 3122-3134.
- [2] Miyoshi, N., Shirai, T. (2014). A cellular network model with Ginibre configured base stations. *Advances in Applied Probability*, 46(3), 832-845.
- [3] M. Haenggi. (2014). The mean interference-to-signal ratio and its key role in cellular and amorphous networks. *IEEE Wireless Communications Letters*, 3(6), 597-600.
- [4] H. Wei, N. Deng, W. Zhou and M. Haenggi. (2015, December). A simple approximative approach to the SIR analysis in general heterogeneous cellular networks. In *2015 IEEE Global Communications Conference (GLOBECOM)* (pp. 1-6). IEEE.
- [5] H. Wei, N. Deng, W. Zhou and M. Haenggi. (2016). Approximate SIR analysis in general heterogeneous cellular networks. *IEEE Transactions on Communications*, 64(3), 1259-1273.
- [6] H. Nagamatsu, N. Miyoshi and T. Shirai. (2014, May). Pad approximation for coverage probability in cellular networks. In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2014 12th International Symposium on (pp. 693-700). IEEE.
- [7] N. Miyoshi and T. Shirai. (2016). Spatial modeling and analysis of cellular networks using the Ginibre point process: A tutorial. *IEICE Transactions on Communications* E99-B(11), 2247-2255.
- [8] A. Goldman. (2010). The Palm measure and the Voronoi tessellation for the Ginibre process. *The Annals of Applied Probability*, 20(1), 90-128.

Maintenance Optimization for Multiple-unit Systems Based on Markov Decision Process

Lu Jin

Department of Informatics
University of Electro-Communications

Abstract Advances in technology have led to systems becoming complex and comprising several different units. Breakdowns of such systems are usually due to the deterioration of one or more units. This study first focused on the optimal maintenance decision policy for a deteriorating system consisting of two units for which the deteriorations are assumed to be different and mutually dependent. Three actions, keep operating, complete maintenance, and partial maintenance, are available for each unit. Since the units structurally form a system, maintenance of either unit may imply maintenance of the other unit as well. The decision-making problem is formulated as a Markov decision process for which keep operating or a joint maintenance is chosen at each decision epoch. We investigated the structure of the optimal maintenance policy for such a system which can minimize the expected total discounted long-run cost. A set of sufficient conditions for a monotone policy to be optimal was identified.

1. Introduction

Preventive maintenance that can prevent breakdowns of deteriorating systems before they occur has been widely investigated. In most of the proposed models, systems deteriorate randomly over time and, unless maintenance is taken, eventually fail. Maintenance managers decide when and how to perform maintenance actions, such as repair and overhaul, with the objective of reducing the overall operating and maintenance cost. Maintenance optimization finds the optimal balance between the costs and benefits of performing maintenance actions.

The optimal maintenance decision-making problem for deteriorating systems with single unit has been studied extensively, and one of the most important achievements in optimal maintenance problems has been to identify the structure of optimal maintenance policies. Derman [4] proposed a structure called a *control limit policy* and provided a sufficient condition for the optimality of a control limit policy for the keep and replace problem in the perfect observation case. Lam and Yeh [10] presented algorithms for deriving optimal maintenance policies to minimize the expected long-run cost for continuous-time Markov deteriorating systems. Elwany, Gebraeel, and Maillart [6] presented a model based on exponentially increasing deterioration with the problem formulated as an Markov decision process and showed that the optimal replacement policy is a control limit policy. Kurt and Kharoufeh [8] generalized Derman's model for deteriorating systems operating in a controllable environment. They obtained sufficient conditions for the optimality of a control limit policy. Byon, Ntamo, and Ding [2] investigated the optimal maintenance policy for wind turbines operating under stochastic weather conditions, focusing on the effect of weather conditions on maintenance lead time.

With the improvement of technology, systems become huge and complex. Such systems are composed of multiple units with possibly different types of deteriorations.

Many researchers have considered the optimal maintenance policies for multiple-unit system and some of the proposed models are for the case in which deterioration is independent between units. Thomas [13] investigated a deteriorating system with multiple units and compared the replacement of the whole system with the replacement of individual units. Sethi [12] considered a system consisting of two identical units for which the failure are occur independent. Each unit has two states (good and failure) and three actions (keep operating, replace one unit, replace both units). A monotone policy with respect to elapsed time was derived under certain conditions. Kawaguchi and Suzuki [7] investigated a multiple-unit system without deterioration. The true states in their model are not known directly and are monitored simultaneously using several monitors. They first defined monotone policies in the strict and weak senses and provided a sufficient condition under which a *weakly monotone policy* becomes optimal in view of a minimum total expected cost. Wijnmalen and Hontelez [15] and van der Duyn Schouten and Vanneste [5] discussed the optimal decision

problem for a multi-unit system under perfect observations on the basis of the results of simulation. Barbera, Schneider and Watson [1] considered a system of two independent units in series and studied the optimal solution to minimize the long-run expected cost of maintenance actions and failures.

During the past decades, there has been a growing interest in the modeling and optimization of maintenance of system consisting of multiple dependent units, because people have realized that interactions between units in a system cannot be neglected and should be taken into account in maintenance decisions. Interactions between units can be classified into three different types (Thomas [14]): economic dependence, structural dependence, and stochastic dependence. Economic dependence means that either costs can be saved when several units are jointly maintained instead of separately or the opposite. Dekker and Wildeman [3] reviewed the literature on multiple-unit maintenance models with economic dependence. Lai and Chen [9] investigated a maintenance policy for a two-unit system with failure rate interaction between units. The components in their model are stochastic dependence, and the failure of component 1 affects the failure rate of unit 2 while the failure of unit 2 causes unit 1 to instantaneously fail. They derived the long-run expected cost per unit time by introducing relative costs as a criterion of optimality and discussed the optimal period that minimizes that cost.

Most multiple-unit maintenance models consider independent case or only one of the three dependences, whereas in practice, the dependences in a multiple-unit system are more complicated. In this paper, we investigate the optimal maintenance decision problem for a deteriorating system with two dependent units. Besides keep operation, two types of maintenance actions, partial maintenance (repair) and complete maintenance (overhaul), are available for each unit. Due to the interaction between two units, a joint maintenance policy should be determined based on the information of both units. In the model, the economic dependence (interaction between maintenance costs), the structural dependence (interaction between the actions be implemented to the two units), and the stochastic dependence (interaction between deteriorations of two units) are taken into account. We formulate this problem as a Markov decision process (MDP) and analytically explore the properties of the optimal policy along any deteriorating path. We provide a set of conditions for the existence of the optimality of a *monotone policy* in a sense of component-wise order.

2. Model and Formulation

In this research, we investigate the optimal maintenance decision problem for a deteriorating system with two dependent units. Besides keep operation, two types of maintenance actions, repair and overhaul, are available for each unit. Due to the interaction between two units, a joint maintenance policy should be determined based on the information of both units. In the model, the economic dependence (interaction between maintenance costs), the structural dependence (interaction between the actions be implemented to the two units), and the stochastic dependence (interaction between deteriorations of two units) are taken into account. We formulate the decision-making problem as a Markov decision process and determine an optimal joint maintenance policy which minimizes the expected long-run cost of a two-unit system over an infinite horizon.

2.1. Model Description

Suppose that the system consists of two different units, U_1 and U_2 . The deterioration of unit U_k ($k = 1, 2$) can be measured as one of a finite number of states from space $\mathcal{S}_k = \{0, 1, \dots, n_k\}$. The states are ordered to reflect the level of deterioration. That is, 0 denotes a new state, and n_k refers the most worn state, of unit U_k . The deterioration level of the whole system is evaluated using a two-dimensional information vector $s = (s_1, s_2)$ in space $\mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2$, where s_k is the deterioration state of U_k . We call s the system's information vector. We introduce a partial order among vectors to compare information vectors of the system. For $s = (s_1, s_2)$ and $s' = (s'_1, s'_2)$, if $s_i \leq s'_i$ for $i = 1, 2$, then we say s' is greater than s in the sense of *component-wise ordering*, denoted by $s \leq_c s'$. In this research, s' represents a more deteriorated state of the system than s , if $s \leq_c s'$.

A decision is made whether to continue system's operation for an additional unit of time period, or to take one maintenance action from several options referring to the information $s = (s_1, s_2)$. If "keep operating" (without taking any action to either unit) is selected for information vector s , operation cost $K(s)$ is incurred and both units are deteriorating for the current time period. Note that any unit can not recover by

themselves. We assume that the deterioration of system follows $f(\mathbf{u}|\mathbf{s})$, which is a conditional probability that the system deteriorates to $\mathbf{u}(\geq_c \mathbf{s})$ from $\mathbf{s}$.

For a system which is formed by several units structurally, the maintenance of either unit implies maintenance of other units. Let a_k ($k = 1, 2$) represents an action implemented to unit U_k . a_k takes a value of 1 or 0. Here $a_k = 0$ represents a complete maintenance such as “overhaul” and $a_k = 1$ represents a partial maintenance such as “repair”. Overhaul recovers the deterioration of a system perfectly, while repair just recovers the deterioration partially. If $a_k = 0$, overhaul cost O_k is incurred for U_k , and the state of U_k is reset to 0. If $a_k = 1$, repair cost $R_k(s_k)$ is incurred for U_k , and the deterioration state of U_k will be improved to a prespecified level $r_k(< s_k)$. $R_k(s_k)$ depends on the current state s_k since the unit with a more deteriorated state usually needs a costly repair than that with a better state. Therefore, four types of maintenance actions, $\mathbf{a} = (a_1, a_2) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, are available for a two-unit system. The corresponding maintenance costs for action $\mathbf{a} = (a_1, a_2)$ is given as

$$C_{\mathbf{a}}(s_1, s_2) = \begin{cases} R_1(s_1) + R_2(s_2) & \text{for } \mathbf{a} = (1, 1) \\ R_1(s_1) + O_2 & \text{for } \mathbf{a} = (1, 0) \\ O_1 + R_2(s_2) & \text{for } \mathbf{a} = (0, 1) \\ O_{12} & \text{for } \mathbf{a} = (0, 0) \end{cases} \quad (1)$$

We assume that $O_{12} \leq O_1 + O_2$ since that the overhaul cost can be saved when two units are joint overhauled instead of separately.

Let $G(a_1, a_2) = 2 - (a_1 + a_2)$ be the size of action (a_1, a_2) which represents the scale of a maintenance action. As the size shown, for overhaul of the whole system, $G(0, 0) = 2$, and it is the largest maintenance action, while for repairs of both units, $G(1, 1) = 0$, is the smallest maintenance action for the system. For convenience, let $M(G)$, $G = 0, 1, 2$ denote the maintenance actions with three different size, respectively.

2.2. Formulation of Maintenance Decision-making Problem

The decision-making problem can be formulated as an MDP, which is a sequential decision making model. Union of the sets $\mathcal{S}_{\mathcal{H}}$ for all the paths forms the space of information vectors $\mathcal{S}$. The maintenance manager should decide whether continue operating the system or choose a maintenance action (a_1, a_2) to minimize system's total expected cost on the basis of $\mathbf{s} \in \mathcal{S}_{\mathcal{H}}$. Let

$$KV^{(N)}(\mathbf{s}) = K(\mathbf{s}) + \beta \sum_{\mathbf{u} \geq_c \mathbf{s}} f(\mathbf{u}|\mathbf{s}) V^{(N-1)}(\mathbf{u}) \quad (2)$$

be the total expected cost if keep operating is selected for information vector $\mathbf{s}$ and optimal decision policy is followed for the remaining $N - 1$ time periods. Here, β ($0 < \beta < 1$) is a discount factor.

Let

$$MV^{(N)}(\mathbf{s}) = \min_{\mathbf{a}} \{MV^{(N)}(\mathbf{s}|\mathbf{a})\} \quad (3)$$

be the minimal total expected maintenance cost of the process starting at state $\mathbf{s} = (s_1, s_2)$ over N time periods. Here,

$$MV^{(N)}(\mathbf{s}|\mathbf{a}) = C_{\mathbf{a}}(\mathbf{s}) + \beta V^{(N-1)}(\mathbf{a} \circ \mathbf{r}) \quad (4)$$

is the expected maintenance cost if action $\mathbf{a}$ is selected for $\mathbf{s}$. The maintenance action will take one time period and the deterioration of system is recovered to $\mathbf{a} \circ \mathbf{r}$. Here $\mathbf{a}$ is the action to be selected and $\mathbf{r} = (r_1, r_2)$, in which r_k is the prespecified level of U_k improved by a partially repair.

An optimal policy is a sequence of optimal actions for every possible $\mathbf{s}$. In the decision-making process, the optimal action that minimizes the total expected cost for each $\mathbf{s}$ can be determined by solving a set of equations:

$$V^{(N)}(\mathbf{s}) = \min \{KV^{(N)}(\mathbf{s}), MV^{(N)}(\mathbf{s})\}. \quad (5)$$

Let $V^{(0)}(\mathbf{s}) = 0$ for any $\mathbf{s}$. From the standard argument of contraction mapping theory [11], $KV^{(N)}(\mathbf{s})$ and $MV^{(N)}(\mathbf{s})$ converge to $KV(\mathbf{s})$ and $MV(\mathbf{s})$ respectively as N tends to infinity. Therefore, we have a recursive function for the optimal total expected cost function:

$$V(s) = \min \{KV(s), MV(s)\}. \quad (6)$$

3. Optimal Maintenance Decision-Making

In this section, we investigate the optimal maintenance decision-making problem for a two-unit deteriorating system. The properties of the optimal expected cost function are examined, and the optimal maintenance policy is found to be given by a *monotone policy*.

When we focus on a state space in which the deteriorating states ordered by a component-wise order, if the ordered state space can be divided into at most two regions, {keep region, maintenance region}, we call this kind of decision policy, a *monotone policy* in the sense of a component-wise order.

3.1. Assumptions

We made the following assumptions regarding the deterioration process. Note that we consider functions to be either increasing or decreasing in the weak sense throughout this paper.

(A-1) For any state $\mathbf{u}$ and $\mathbf{u}' \in \mathcal{S}$,

$$f(\mathbf{u}'|s)f(\mathbf{u}|s') \leq f(\mathbf{u} \wedge \mathbf{u}'|s)f(\mathbf{u} \vee \mathbf{u}'|s') \quad (7)$$

for any $s \leq_c s'$. Here $\mathbf{u} \wedge \mathbf{u}' = (\min\{u_1, u'_1\}, \min\{u_2, u'_2\})$ and $\mathbf{u} \vee \mathbf{u}' = (\max\{u_1, u'_1\}, \max\{u_2, u'_2\})$.

(A-2) $K(s)$ is increasing in s in the sense of a component-wise order.

(A-3) $C_a(s)$ is increasing in s in the sense of a component-wise order.

(A-4) $K(s) - C_a(s)$ is increasing in s in the sense of a component-wise order.

(A-5) For two maintenance actions $\mathbf{a}$ and $\mathbf{a}'$ for which $G(\mathbf{a}) < G(\mathbf{a}')$, $C_a(s) - C_{a'}(s)$ is increasing in s in the sense of a component-wise order.

Assumption (A-1) means that, as the system deteriorates, the system is more likely to move to a more deteriorated state. (A-1) is known as a weak multivariate monotone likelihood ratio order [16], which is the extension of monotone likelihood ratio of one-dimensional state transition probabilities. Assumption (A-2) implies that the operation cost per unit time period do not decrease with the deterioration of the system. Assumption (A-3) implies that, the repair cost of one unit depends on its current state. A more deteriorated unit needs more expensive repair cost to recover it to a certain level. Assumption (A-4) implies that, as the state deteriorates, the advantage of maintenance becomes larger than that of keeping it operating. Assumption (A-5) implies that, as the system deteriorates, the advantage of large scale maintenance becomes larger than that of small scale maintenance.

3.2. Properties of Optimal Maintenance Policies

Under Assumptions (A-1) to (A-5), we derive the following theorems.

Lemma 1:

For any deteriorating state s , $KV(s) - MV(s|\mathbf{a})$ for any $\mathbf{a}$ is increasing in s in the sense of a component-wise order.

Theorem 1:

Under Assumptions (A-1) to (A-5), the optimal policy is given by a *monotone policy*.

Theorem 2:

Under assumptions (A-1) through (A-5), the scale of optimal maintenance action becomes larger as the system deteriorates.

Lemma 1 ensures the establishment Theorem 1 under Assumptions (A-1) to (A-5). Theorems 1 and 1 define the properties of the optimal maintenance decision. Theorem 1 shows that under some reasonable assumptions, there is exist at most one threshold for keep operation and maintenance. Theorem 2 shows that the assumptions presented in section 3.1 are sufficient for an increasing trend in the scale of the optimal maintenance.

4. Numerical Example

Table 1 Operation cost $K(s_1, s_2)$

		s_1				
s_2		0	1	2	3	4
	0	30	40	50	80	150
	1	40	50	60	90	160
	2	50	60	70	100	170
	3	80	90	100	110	180
	4	150	160	170	180	250

Here we present a numerical example. Consider a system consisting of two units (U_1 and U_2) for which the deterioration can be classified into one of five states: $s_1, s_2 \in \{0, 1, 2, 3, 4\}$ where 4 means the greatest deterioration. We assume that operation cost $K(s_1, s_2)$ for every (s_1, s_2) pair is given as shown in Table 1. Repair can recover the state of both units to state 1; the repair costs are given by $R_1(s_1) = 16 + 10s_1$ and $R_2(s_2) = 16 + 10s_2$. If the system is in a state better than (1, 1), repair is not selected. The overhaul costs for U_1 and U_2 are $O_1 = 114$ and $O_2 = 164$. The cost of overhauling the whole system is $O_{12} = 227$ since overhauling the two units simultaneously reduces some of the cost. These cost functions satisfy assumptions (A-2)-(A-5). We assume that discount factor β is 0.96.

The system deteriorates in accordance with

$$f(u|s)_{s_1, s_2 \in \{0, 1, 2, 3, 4\}} = \begin{cases} 0.9 & u = s \\ 0.1 & u = (\min(s_1 + 1, 4), \min(4, s_2 + 1)) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

and assumption (A-1) holds for $f(u|s)$.

Table 2 Optimal action for every (s_1, s_2)

		s_1				
s_2		0	1	2	3	4
	0	K	K	K	$a = (0, 1)$	$a = (0, 1)$
	1	K	K	$a = (1, 1)$	$a = (0, 1)$	$a = (0, 1)$
	2	K	$a = (1, 1)$	$a = (1, 1)$	$a = (0, 0)$	$a = (0, 0)$
	3	$a = (1, 1)$	$a = (1, 1)$	$a = (0, 0)$	$a = (0, 0)$	$a = (0, 0)$
	4	$a = (1, 1)$	$a = (0, 0)$	$a = (0, 0)$	$a = (0, 0)$	$a = (0, 0)$

We calculated $KV(s_1, s_2)$ and $MV(s_1, s_2|a)$ for every (s_1, s_2) pair on the basis of these parameters. Taking the action that has the minimal value for every state as the optimal action, we obtain the optimal action for every (s_1, s_2) , as shown in Table 2. The state space is divided into two regions: “keep operating” (cells with “K”) and “perform maintenance” (cells with “a”). The different types of maintenance represents different sizes of the maintenance action to be taken. Since the size of the action represents its scale, we can see that a maintenance action with a larger scale is more likely to be implemented as a system deteriorates in the sense of component-wise ordering. For example, if we focus on the states on one sample path $\{(2, 0) \rightarrow (3, 1) \rightarrow (4, 2) \rightarrow (4, 3) \rightarrow (4, 4)\}$ which is one pattern of system deterioration starting from (2, 0), the optimal action changes from “Keep operating” to “Maintenance” once. Furthermore, the corresponding optimal actions are $\{“K” \rightarrow “M(1) = (0, 1)” \rightarrow “M(2) = (0, 0)” \rightarrow “M(2) = (0, 0)” \rightarrow “M(2) = (0, 0)”\}$. The action $M(1) = (0, 1)$ represents a maintenance action with size 1, i.e., the the overhaul of U_1 and repair of U_2 , and $M(2) = (0, 0)$ represents a maintenance action with size 2, i.e., the overhaul of both units. This numerical example shows that the optimal action for a two-unit system is given by a monotone policy, and the scale of optimal maintenance action becomes larger as the system deteriorates.

5. Conclusion

While optimal decision-making problems for multiple-unit systems of which the units deteriorate independently have been investigated in previous studies, we investigated an optimal maintenance decision-making policy for a system with two dependent units. In this research, the economic dependence, the structural dependence, and the stochastic dependence are taken into account. We formulated the decision-making problem as a Markov decision process and derived a set of sufficient conditions for the optimal joint maintenance policy to be given by a monotone policy. Furthermore, we found that it is preferable to implement a maintenance action with a larger scale as the system deteriorates.

References

- [1] F. Barbera, H. Schneider and E. Watson: A condition based maintenance model for a two-unit series system, *European Journal of Operational Research*, **116** (1999), 281–290.
- [2] E. Byon, L. Ntamo and Y. Ding: Optimal maintenance strategies for wind turbine systems under stochastic weather conditions, *IEEE Transactions on Reliability*, **59** (2010), 393–404.
- [3] R. Dekker. and R. E. Wildeman: A review of multi-component maintenance models with economic dependence, *Mathematical Methods of Operations Research*, **45** (1997), 411–435.
- [4] C. Derman: On Optimal Replacement Rules When Changes of State Are Markovian, *Mathematical Optimization Techniques (edited by R. Bellman)*, University of California Press (1963), 201–210.
- [5] F. A. van der Duyn Schouten and S. G. Vanneste: Analysis and computation of (n, N) -strategies for maintenance of a two-component system, *European Journal of Operational Research*, **48** (1990), 260–274.
- [6] A. H. Elwany, N. Z. Gebraeel and L. M. Maillart: Structured replacement policies for components with complex degradation processes and dedicated sensors, *Operations Research*, **59** (2011), 684–695.
- [7] Y. Kawaguchi, K. Suzuki: Optimum maintenance policy using simultaneous monitoring equipments on multiple units, *Journal of the Japanese Society for Quality Control*, **26** (1996), 93–103. (in Japanese)
- [8] M. Kurt and J. P. Kharoufeh: Monotone optimal replacement policies for a Markovian deteriorating system in a controllable environment, *Operations Research Letters*, **38** (2010), 273–279.
- [9] M-T. Lai and Y-C. Chen: Optimal periodic replacement policy for a two-unit system with failure rate interaction. *The International Journal of Advanced Manufacturing Technology*, **29** (2006), 367–371.
- [10] C. T. Lam and R. H. Yeh: Optimal maintenance-policies for deteriorating systems under various maintenance strategies, *IEEE Transactions on Reliability*, **43** (1994), 423–430.
- [11] M. L. Puterman: *Markov decision processes: discrete stochastic dynamic programming*, Wiley (1994).
- [12] D. P. Sethi: *Opportunistic replacement policies for maintained systems*, Operations Research Center, University of California, Berkeley, (1976).
- [13] L. C. Thomas: Replacement of systems and components in renewal decision problems, *Operations Research*, **33** (1985), 404–411.
- [14] L. C. Thomas: A survey of maintenance and replacement models for maintainability and reliability of multi-item systems, *Reliability Engineering*, **16** (1986), 297–309.
- [15] D. J. Wijnmalen and J. A. Hontelez: Coordinated condition-based repair strategies for components of a multi-component maintenance system with discounts, *European Journal of Operational Research*, **98** (1997), 52–63.
- [16] W. Whitt: Multivariate monotone likelihood ratio and uniform conditional stochastic order, *Journal of Applied Probability*, **19** (1982), 695–701.

Monotone properties of an optimal maintenance policy for a Markovian deteriorating system with imperfect repair and incomplete information

Nobuyuki Tamura

Department of Industrial and Systems Engineering
Hosei University

Abstract This paper considers a system whose deterioration is modeled as a discrete-state and discrete-time Markov chain. At each discrete time point, one of the following three actions can be taken: operation (wait), repair, or replacement. After operation, the system deteriorates according to the Markov chain and it is monitored to obtain some incomplete information on the degree of the deterioration. We propose a partially observable Markovian decision process (POMDP) model with imperfect repair. When imperfect repair is repeated, the system is less likely to move to better states after completion of repair. We show that, under certain assumptions, a control-limit policy is optimal. Furthermore, we provide several structural properties on the optimal maintenance policy focused on repetition of imperfect repair.

1. Introduction

Since any system necessarily deteriorates due to usage or age, it cannot remain in good operating conditions and will eventually fail without maintenance. Therefore, it is important to decide when and how to take maintenance actions. Many previous studies have been conducted on solving maintenance problems mathematically. The papers by Pierskalla and Voelker [12], Sherif and Smith [15], Valdez-Flores and Feldman [18], Cho and Parlar [1], Dekker [2], and Wang [19] are excellent reviews of the subject. In the case of studies on the idea of imperfect maintenance and its applications to optimal maintenance problems, the paper by Pham and Wang [11] provides many examples.

If we assume that the deterioration levels of a system correspond to a finite set of non-negative integers, then the deterioration process of the system can be described as a Markov process with an absorbing state which expresses failure of the system. This model is called a Markovian deteriorating system. Derman [3] first studied the Markovian deteriorating system and derived sufficient conditions that a control-limit policy holds. Douer and Yechiali [4]), and Tamura [16] studied a Markovian deteriorating system for which repair or replacement can be selected as a maintenance action and showed that a generalized control-limit policy holds under several reasonable assumptions. These models assume that the true state of the system can be identified with certainty at any given time.

Unfortunately, we do not always detect the true state of a system. If it is costly and/or time consuming to observe a system directly, we use partial(incomplete) information on the deterioration of the system. For this situation, partially observable Markov decision process (POMDP) models are appropriate. POMDP models consider a system that is monitored by some mechanism which gives partial information on the state of the system but does not necessarily tell its true state. Monahan [8] gave an excellent review of this subject. Ohnishi et al. [9] proposed the maintenance problem via a POMDP model and derived sufficient conditions that the optimal maintenance policy satisfies the so-called “monotonic properties” by using the theory of stochastic order relations. Also, Ivy et al. [5] and Tamura et al. [17] studied a POMDP model with imperfect repair. However, these models assumed that repetition of repair does not influence degree of its imperfection.

The present paper considers a system whose deterioration follows a discrete-state and discrete-time Markov chain with an absorbing state. At each discrete time point, the true state of the system is not identified but is monitored instead to obtain partial information on the deterioration of the system. For this system, we propose a POMDP model with imperfect repair. As the system repeatedly undergoes repair, it becomes less likely to move to a better state after each completion

of repair. For the model, we provide the total discounted cost for an unbounded horizon and examine under what assumptions a control-limit policy is optimal. Furthermore, we examine the monotonicity of the optimal maintenance policy.

This remainder of the paper is organized as follows. In the next section, we explain the structure of the model in detail. In section 3, we provide the mathematical formulation as a partially Markov decision process model. In section 4, we derive several structural properties of the optimal maintenance policy that minimizes the expected total discounted cost for an unbounded horizon. Finally, we present our conclusions in section 5. The proofs for the theorems and lemmas in section 4 are summarized in the appendices.

2. Model description

We consider a system whose deterioration follows a discrete-time and discrete-state Markov chain with an absorbing state. The system can be classified into one of $N + 1$ states. State 0 represents the process before any deterioration takes place, that is, the initial new state, whereas N is the failure state of the system. The intermediate states $1, \dots, N - 1$ are ordered to reflect the relative degree of deterioration in ascending order. At each discrete time point, we can select from one of three possible actions: operation(wait), repair, or replacement.

Let p_{ij} be the transition probability of the Markov chain and $\mathbf{P}$ be the transition probability matrix. $\mathcal{S} = \{0, 1, \dots, N\}$ is the state space of the Markov chain. We let $q_{ij}(m)$ denote the repair probability that the system is returned to state j from state i through repair when the system has already undergone m repairs. Also, we let x_i denote the probability that the true state of a system, which has already undergone m repairs, is $i \in \mathcal{S} = \{0, 1, \dots, N\}$ and set $\mathbf{x} = (x_0, x_1, \dots, x_N)$. Hereafter, we call $\mathbf{x}$ the state vector of the system. Also, $\mathbf{e}_i$ means the $N + 1$ dimensional vector that has i th element 1 and all other elements zero. We consider actions to be selected based on the state vector and the number of repairs instead of the true state.

When wait (W) is selected, the system is operated for one period. Thus, a system in state i moves to state j with probability p_{ij} at the next time point according to the Markov chain. After operation, we monitor the system and obtain the outcome which has some relation to the true state of the system. We assume that the relationship between the outcome of the monitor θ and the true state of the system i is given as follows.

$$\gamma_{i\theta} = \Pr\{\text{the outcome is } \theta \mid \text{the operated system stays in state } i\}.$$

We define $(\mathbf{\Gamma})_{i\theta} = \gamma_{i\theta}$ and assume $\theta \in \mathcal{M} = \{1, \dots, M\}$. As a result, operating cost u_i is incurred. As defined above, when repair (R_R) is selected, a system in state i , which has already undergone m repairs moves to state j with probability $q_{ij}(m)$. We define $(\mathbf{Q}(m))_{ij} = q_{ij}(m)$. In this case, repair cost $r_i(m)$ is incurred. After completion of repair, the true state of the system is identified and an optimal action is selected. When replacement (R_T) is selected, a system in state i moves to state 0 without fail and replacement cost c_i is incurred.

We define

$$\begin{aligned} \mathbf{u} &= (u_0, \dots, u_N)^\top, \mathbf{r}(m) = (r_0(m), \dots, r_N(m))^\top, \mathbf{c} = (c_0, \dots, c_N)^\top, \\ \mathcal{F}^{N+1} &= \{\mathbf{f} \in \mathcal{R}^{N+1} \mid f_{i-1} \leq f_i \ (1 \leq i \leq N)\}, \end{aligned}$$

where $\top$ means transportation.

We analyze our model based on the properties of the stochastic order relations (e.g. see Karlin [6], Kijima [7]). The properties are very important in the study of these types of problems.

As part of our definitions, we also assume that a nonnegative matrix $\mathbf{B}$ is defined on the state space $\mathcal{N}$, where $\mathcal{N} = \{0, 1, 2, \dots, n\}$.

Definition 1. The matrix $\mathbf{B} \in TP_2$ if and only if

$$b_{hj}b_{ik} - b_{hk}b_{ij} \geq 0$$

for $h \leq i$ and $j \leq k$.

Definition 2. The matrix $\mathbf{B} \in SI$ if and only if

$$\sum_{j=h}^n b_{ij} \leq \sum_{j=h}^n b_{kj}$$

for $i \leq k$.

$\mathbf{B} \in TP_2$ means $\mathbf{B}$ has the property of totally positive of order 2. Similarly, $\mathbf{B} \in SI$ means $\mathbf{B}$ has the property of stochastic increasing. Hereafter, if a matrix $\begin{pmatrix} \mathbf{a}^\top \\ \mathbf{b}^\top \end{pmatrix} \in TP_2$ for vectors $\mathbf{a}$ and $\mathbf{b}$, then we write $\mathbf{a} \stackrel{T}{\prec} \mathbf{b}$. Also, if

$$\sum_{i=k}^{s+1} a_i \leq \sum_{i=k}^{s+1} b_i$$

then we write $\mathbf{a} \stackrel{S}{\prec} \mathbf{b}$.

For the model, we impose the following assumptions on the costs and the probabilities.

Assumption 1. For any m , $\mathbf{u} \in \mathcal{F}^{N+1}$, $\mathbf{r}(m) \in \mathcal{F}^{N+1}$, and $\mathbf{c} \in \mathcal{F}^{N+1}$.

Assumption 2. For any i , $r_i(m)$ is increasing in m .

Assumption 3. For any m , $\mathbf{u} - \mathbf{r}(m) \in \mathcal{F}^{N+1}$, $\mathbf{u} - \mathbf{c} \in \mathcal{F}^{N+1}$, and $\mathbf{r}(m) - \mathbf{c} \in \mathcal{F}^{N+1}$.

Assumption 1 means that as the system deteriorates, it becomes more costly to operate, repair, or replace the system. This assumption would appropriately reflect real situations. Assumption 2 means that it becomes more costly to repair the system with repetition of repair. Thus, this is related to a technical difficulty with repetition of repair. We can interpret Assumption 3 as follows. As the system deteriorates, the merit of repair or replacement increases relative to that of operation. In addition, the merit of replacement increases relative to that of repair with deterioration.

Assumption 4. For any m , $\mathbf{P}$ and $\mathbf{Q}(m)$ are totally positive of order 2 (TP_2).

If $\mathbf{P} \in TP_2$ then $\mathbf{P} \in SI$. Hence, Assumption 4 indicates that as the system deteriorates, it becomes more likely to move to a worse state. Similarly, since $\mathbf{Q}(m) \in SI$ because of $\mathbf{Q}(m) \in TP_2$, the system becomes less likely to move to a better state after completion of repair with deterioration.

Let $\mathbf{q}_i(m)$ denote the i th row vector of the matrix $\mathbf{Q}(m)$. This comprises the probability that the system in state i , which has already undergone m repairs, moves to different states after completion of repair. Then we impose the following assumption to express the influence of repetition of repair on the system.

Assumption 5. For any m and $i \in \mathcal{S}$, $\mathbf{q}_i(m) \stackrel{T}{\prec} \mathbf{q}_i(m+1)$.

Assumption 5 indicates that as the system repeatedly undergoes repair, it becomes less likely to move to a better state after completion of repair.

3. Formulation

Our problem can be modeled as a partially observable Markov decision process. Let β ($0 < \beta < 1$) be a discount factor.

Although we consider a Markov chain with stationary probabilities, the repair probabilities depend upon the number of repairs. For the formulation, we define a set $\mathbf{X}$ by

$$\mathbf{X} = \left\{ \mathbf{x} \in \mathcal{R}^{N+1} \mid x_i \geq 0 \text{ for } i \in \mathcal{S}, \sum_{i=0}^N x_i = 1 \right\}.$$

Furthermore, we let $V(\mathbf{x}, m)$ denote the optimal total expected discounted cost for an unbounded horizon when the state vector is $\mathbf{x}$ and the number of repairs is m .

When the system is operated at state vector $\mathbf{x}$, the probability that the outcome of the monitor is θ is given by

$$P(\theta|\mathbf{x}) = \sum_i \sum_j x_i p_{ij} \gamma_{j\theta}.$$

We let $T(j|\mathbf{x}, \theta)$ denote the probability that the true state of the system is j given that the outcome is θ after the system is operated at state vector $\mathbf{x}$. $T(j|\mathbf{x}, \theta)$ is expressed as

$$T(j|\mathbf{x}, \theta) = \frac{\sum_i x_i p_{ij} \gamma_{j\theta}}{\sum_i \sum_j x_i p_{ij} \gamma_{j\theta}},$$

via Bayes' formula. Thus, when Action W is selected at state vector $\mathbf{x}$, the state vector at the next time point becomes $\mathbf{T}(\mathbf{x}, \theta)$ where $(\mathbf{T}(\mathbf{x}, \theta))_j = T(j|\mathbf{x}, \theta)$. We let $W(\mathbf{x}, m)$ denote the total expected discounted cost when Action W is taken at state vector $\mathbf{x}$, the number of repairs is m , and the optimal maintenance policy is employed afterwards. Hence, $W(\mathbf{x}, m)$ is given by

$$W(\mathbf{x}, m) = \mathbf{x}^\top \mathbf{u} + \beta \sum_{\theta} P(\theta|\mathbf{x}) V(\mathbf{T}(\mathbf{x}, \theta), m). \quad (1)$$

When the system is repaired at state vector $\mathbf{x}$ and has already undergone m repairs, it moves to state j with probability $(\mathbf{x}^\top \mathbf{Q}(m+1))_j$. Afterwards, the true state is identified through inspection and an optimal action is selected. Thus, we let $R_R(\mathbf{x}, m)$ denote the total expected discounted cost when Action R_R is selected at state vector $\mathbf{x}$, the number of repairs is m , and the optimal maintenance policy is employed afterwards. Hence, $R_R(\mathbf{x}, m)$ is given by

$$R_R(\mathbf{x}, m) = \mathbf{x}^\top \mathbf{r}(m+1) + \beta \sum_j (\mathbf{x}^\top \mathbf{Q}(m+1))_j V(\mathbf{e}_j, m+1). \quad (2)$$

When the system is replaced at state vector $\mathbf{x}$ and has already undergone repair m times, it is returned to state 0 without fail and an optimal action is selected. Thus, we let $R_T(\mathbf{x}, m)$ denote the total expected discounted cost when Action R_T is selected at state vector $\mathbf{x}$, the number of repairs is m , and the optimal maintenance policy is employed afterwards. Hence, $R_T(\mathbf{x}, m)$ is given by

$$R_T(\mathbf{x}, m) = \mathbf{x}^\top \mathbf{c} + \beta V(\mathbf{e}_0, 0). \quad (3)$$

From the theory of dynamic programming (Ross [14]), $V(\mathbf{x}, m)$ is expressed by the following recursive equation.

$$V(\mathbf{x}, m) = \min [W(\mathbf{x}, m), R_R(\mathbf{x}, m), R_T(\mathbf{x}, m)]. \quad (4)$$

4. Structure of the optimal maintenance policy

This section provides several structural properties of the optimal maintenance policy. As a preliminary, we give some results below.

Theorem 1. *For any m , $V(\mathbf{x}, m)$ is a concave function of $\mathbf{x}$, and if $\mathbf{x}_1 \stackrel{T}{\prec} \mathbf{x}_2$, then $V(\mathbf{x}_1, m) \leq V(\mathbf{x}_2, m)$. Also, for any $\mathbf{x} \in \mathbf{X}$, $V(\mathbf{x}, m)$ is increasing in m .*

We can prove Theorem 1 by using mathematical induction. From the argument of the proof of Theorem 1, we have the following corollaries.

Corollary 1. *For any m , $W(\mathbf{x}, m)$ is concave in $\mathbf{x}$.*

Corollary 2. *For any m , $W(\mathbf{x}, m)$ and $R_R(\mathbf{x}, m)$ are increasing in $\mathbf{x} \in \mathbf{X}$ with respect to TP_2 . Also, for any $\mathbf{x} \in \mathbf{X}$, $W(\mathbf{x}, m)$ and $R_R(\mathbf{x}, m)$ are increasing in m .*

Using these results, some lemmas are obtained as follows.

Lemma 1. *For any m , if $\mathbf{x}_1 \stackrel{T}{\prec} \mathbf{x}_2$, then*

$$\begin{aligned} W(\mathbf{x}_1, m) - R_T(\mathbf{x}_1, m) &\leq W(\mathbf{x}_2, m) - R_T(\mathbf{x}_2, m), \\ R_R(\mathbf{x}_1, m) - R_T(\mathbf{x}_1, m) &\leq R_R(\mathbf{x}_2, m) - R_T(\mathbf{x}_2, m). \end{aligned}$$

Lemma 2. *For any m , $W(\mathbf{x}, m) - R_R(\mathbf{x}, m)$ is concave in $\mathbf{x}$.*

Lemma 3. *For any $\mathbf{x} \in \mathbf{X}$, $W(\mathbf{x}, m) - R_T(\mathbf{x}, m)$ and $R_R(\mathbf{x}, m) - R_T(\mathbf{x}, m)$ are increasing in m .*

These results are useful for deriving structural properties of the optimal maintenance policy. Now we let $D(\mathbf{x}, m)$ denote an optimal action at state vector $\mathbf{x}$ and number of repairs m . By using Lemmas 1, 1, and 2, a control-limit policy is optimal as follows.

Theorem 2. Suppose any $\tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}$ ($\tilde{\mathbf{x}} \prec^T \hat{\mathbf{x}}$). Then, for any m , there exist real numbers $\tilde{\alpha}_m$, $\bar{\alpha}_m$, and $\hat{\alpha}_m$ such that

$$D(\mathbf{x}_\alpha, m) = \begin{cases} W & \text{for } 0 \leq \alpha < \tilde{\alpha}_m, \\ R_R & \text{for } \tilde{\alpha}_m \leq \alpha < \hat{\alpha}_m, \\ W & \text{for } \hat{\alpha}_m \leq \alpha < \bar{\alpha}_m, \\ R_T & \text{for } \bar{\alpha}_m \leq \alpha \leq 1, \end{cases}$$

where $\mathbf{x}_\alpha = (1 - \alpha)\tilde{\mathbf{x}} + \alpha\hat{\mathbf{x}}$ and $0 \leq \tilde{\alpha}_m \leq \bar{\alpha}_m \leq \hat{\alpha}_m \leq 1$.

Also, according to Lemma 3, we can derive a property of the control-limit $\bar{\alpha}_m$ under repetition of repair. In this paper, the control-limit means the state vector such that replacement is optimally selected for the system which has already undergone m repairs.

Theorem 3. $\bar{\alpha}_m$ is decreasing in m .

Theorem 3 means that, since it becomes more costly to repair the system and less likely to be repaired to a better state with repetition of repair, to select replacement is more preferable than repair for a system which has undergone more repairs at smaller state vectors with respect to TP_2 .

Furthermore, by using Lemma 3 and Theorems 2 and 3, we obtain that the optimal maintenance policy satisfies the following monotone property.

Theorem 4. There exist integers m' and m^* such that, if $0 \leq m < m'$, then

$$D(\mathbf{x}_\alpha, m) = \begin{cases} W & \text{for } 0 \leq \alpha < \tilde{\alpha}_m, \\ R_R & \text{for } \tilde{\alpha}_m \leq \alpha < \hat{\alpha}_m, \\ W & \text{for } \hat{\alpha}_m \leq \alpha < \bar{\alpha}_m, \\ R_T & \text{for } \bar{\alpha}_m \leq \alpha \leq 1, \end{cases}$$

and if $m' \leq m < m^*$, then

$$D(\mathbf{x}_\alpha, m) = \begin{cases} W & \text{for } 0 \leq \alpha < \tilde{\alpha}_m, \\ R_R & \text{for } \tilde{\alpha}_m \leq \alpha < \bar{\alpha}_m, \\ R_T & \text{for } \bar{\alpha}_m \leq \alpha \leq 1, \end{cases}$$

if $m = m^*$ then

$$D(\mathbf{x}_\alpha, m) = \begin{cases} W & \text{for } 0 \leq \alpha < \bar{\alpha}_m, \\ R_T & \text{for } \bar{\alpha}_m \leq \alpha \leq 1, \end{cases}$$

where $\mathbf{x}_\alpha = (1 - \alpha)\tilde{\mathbf{x}} + \alpha\hat{\mathbf{x}}$, $0 \leq \tilde{\alpha}_m \leq \bar{\alpha}_m \leq \hat{\alpha}_m \leq 1$, and $0 \leq m' \leq m^*$.

Hence, m^* is the optimal number of repairs for the system because Action R_R is not selected when the number of repairs reaches m^* . Also, we find that the optimal maintenance policy may be characterized by four regions at most. If $m' = m^*$, then the structure of the optimal maintenance policy agrees with that of Ohnishi et al. [9].

5. Conclusion

In this paper, we have proposed a POMDP model for a Markovian deteriorating system with imperfect repair. As the system repeatedly undergoes imperfect repair, it becomes less likely to move to a better state after completion of repair. We have derived the total expected discounted cost for an unbounded horizon. For the model, we have shown that the control-limit policy is optimal under certain reasonable assumptions. Furthermore, we have investigated the behavior of the optimal maintenance policy with repetition of repair. As a result, we have shown that the optimal maintenance policy is monotonic and may be characterized by four regions at most. Finally, we have found that there exists an optimal number of repairs. These results are useful for numerically determining an optimal maintenance policy. However, since imperfect repair influences

the stochastic behavior of the system, we have adopted stricter assumptions than SI for the repair probabilities.

Future work includes deriving conditions under which the number of regions of the optimal maintenance policy is at most three. This means that we have to obtain conditions such that $m' = m^*$. Also, to numerically solve our problem more efficiently, it is important to find a finite upper bound on the optimal number of repairs.

References

- [1] D.I. Cho and M. Parlar: A Survey of Maintenance Models for Multi-Unit Systems, *European Journal of Operational Research*, **51** (1991), 1–23.
- [2] R. Dekker: Applications of Maintenance Optimization Models: A Review and Analysis, *Reliability Engineering and System Safety*, **51** (1996), 229–240.
- [3] C. Derman: On Optimal Replacement Rules When Changes of States are Markovian, In: R. Bellman (Ed.). *Mathematical Optimization Techniques*, (1963) 201–210, University of California Press, Berkley, California.
- [4] N. Douer and U. Yechiali: Optimal Repair and Replacement in Markovian Systems. *Communications in Statistics -Stochastic Models-*, **10** (1994), 253–270.
- [5] J.S. Ivy and S.M. Pollock: Marginally monotonic maintenance policies for a multi-state deteriorating machine with probabilistic monitoring, and silent failures, *IEEE Transactions on Reliability*, **54** (2005), 489–497.
- [6] S. Karlin: *Total Positivity* (1968), Stanford University Press, Stanford, California,
- [7] M. Kijima: *Markov Processes for Stochastic Modeling* (1997), Chapman & Hall, Cambridge.
- [8] G.E. Monahan: A survey of partially observable Markovian decision processes, theory, models and algorithms, *Management Science*, **28** (1982), 1–16.
- [9] M. Ohnishi, H. Kawai and H. Mine, An optimal inspection and replacement policy under incomplete state information, *European Journal of Operational Research* **27** (1986), 117–128.
- [10] S. Osaki(Ed.): *Stochastic Models in Reliability and Maintenance* (2001), Springer, Berlin.
- [11] H. Pham and H. Wang: Imperfect Maintenance, *European Journal of Operational Research*, **94** (1996), 425–438.
- [12] W.P. Pierskalla and J.A. Voelker: A survey of maintenance models: the control and surveillance of deteriorating systems, *Naval Research Logistics Quarterly*, **23** (1976), 353–388.
- [13] S.M. Ross: Quality control under Markovian deterioration, *Management Science*, **17** (1971), 587–596.
- [14] S.M. Ross: *Applied Probability Models with Optimization Applications* (1970), Dover, New York.
- [15] Y.S. Sherif and M.K. Smith: Optimal maintenance models for systems subject to failure - a review, *Naval Research Logistics Quarterly*, **28** (1981), 47–74.
- [16] N. Tamura: On a Markovian deteriorating system with uncertain repair and replacement, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E90-A** (2007), 467–473.
- [17] N. Tamura, K. Hayashi, T. Yuge and S. Yanagi, Monotone properties of optimal maintenance policy for two-state partially observable Markov decision process model with multiple observations. *International Journal of Operations Research*, **7** (2011), 23–34.
- [18] C. Valdez-Flores and R.M. Feldman: A survey of preventive maintenance models for stochastically deteriorating single-unit systems, *Naval Research Logistics*, **36** (1989), 419–446.
- [19] H. Wang: A survey of maintenance policies of deteriorating systems, *European Journal of Operational Research*, **139** (2002), 469–489.
- [20] C. White: A Markov quality control process subject to partial observation, *Management Science*, **23** (1978), 843–852.

契約電力超過確率最小化に関する研究

*小柳 淳二, 古林 雄貴, 河野 大地 (鳥取大学)

概要：本研究では，電力料金に影響を与える契約電力を超過する確率を最小にする運用政策について考察する．電力料金の値上げにより，企業などでは光熱費の削減が重要な問題である．契約電力を抑えることは，1年間の基本料金の削減につながり，夏季の冷房，冷凍装置の制御により，契約電力を抑えることができれば，短期間の節電以上の光熱費削減効果が期待できる．電気機器のオンオフをシステムの状態変化につながるアクションととらえることで，マルコフ決定過程として，契約電力超過確率の最小化問題を定式化し，最適政策の性質を分析する．

1. はじめに

電力の料金の削減は，企業などにおいて経費削減の方法として重要である．電力会社では30分間ごとに利用された電力の平均値（平均電力）（図1）を出し，そのうち月間で最も大きな数値がその月の最大需要電力であり，契約電力は当月を含む過去1年間で最大需要電力により決定される（図2）[1]．あるいは，多くの電力を使う企業などでは，契約電力を電力会社と契約を結ぶことで決定することもある．電力料金の基本料金は，契約電力に比例して1年間にわたって増加する，あるいは事前に契約している場合には，契約電力を超えることにより，超過分のペナルティ料金を請求される場合などがある．したがって，契約電力を低く抑える，あるいは，契約電力を超えないように電気を使用することは電力料金を抑えるために必要である．

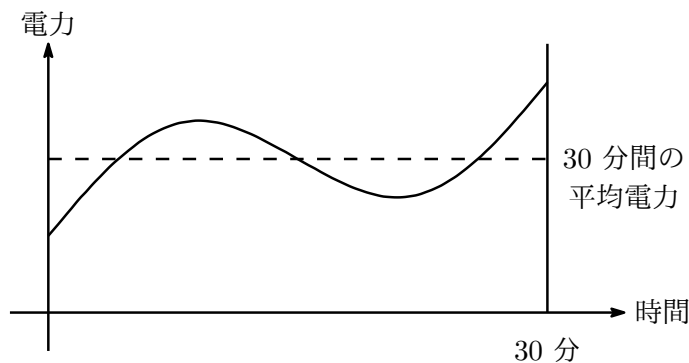


図 1: 30 分間の平均電力

2. システムの状態と目的

冷房装置により，冷却される n 個の部屋を持つシステムを考え，各部屋の温度レベルを表現する状態の番号 $(0, 1, \dots, l, 0$ が最低温度， l が最高温度) を考える．離散時間システムを考え，1単位時間ごとに，各部屋の温度レベルを知ることができ，各部屋ごとに冷房のオンオフを制御できるものとする．状態としては，部屋の温度レベルの状態 $\vec{i} = (i_1, \dots, i_n)$ に対し，アクションとして，部屋 k ($k = 0, \dots, n$) の冷房のオンオフ $s_k = 0, 1$ (0 がオフ， 1 がオン) を決定する問題となる．1期間オンオフを維持することで，次の期間の温度レベルが変化する．現在の状態 $\vec{i}$ とアクション $\vec{s} = (s_1, \dots, s_n)$ に対して次の状態が $\vec{j} = (j_1, \dots, j_n)$ になる確率を $p(\vec{j}|\vec{i}, \vec{s})$ で表現できると仮定する．

本研究では，簡単に各部屋の温度変化は独立に生じるものとし，部屋 k の温度レベルが i で

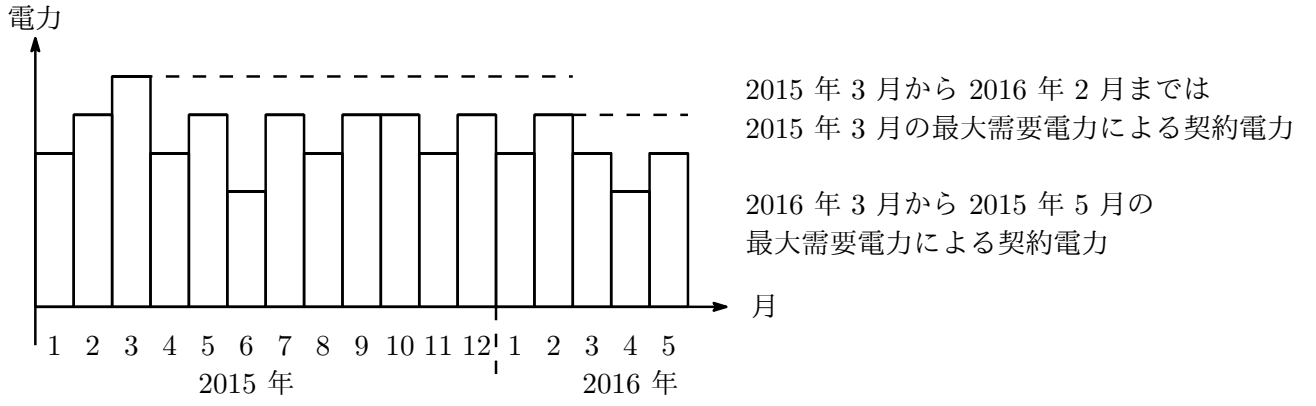


図 2: 最大需要電力と契約電力

1. 冷房オンのときには確率 $d_k(i)$ で温度レベルが $i-1$ に変化し, $1-d_k(i)$ で温度レベル i のままとする. ($d_k(0) = 0$)
 2. 冷房オフのときには確率 $u_k(i)$ で温度レベルが $i+1$ に変化し, 確率 $1-u_k(i)$ で温度レベルが i のままとする. ただし, 温度レベルが l のときには必ず冷房をオンにするものとする.
- n 個の部屋のうち, 契約電力内で冷房をオンにできる数を f 個までとし, 温度レベル l の部屋においては必ず冷房をオンにしなければならない条件のもとで, 指定期間内で $f+1$ 個以上の部屋で冷房をオンにする確率を最小にすることが目的となる.

3. 定式化

指定期間を残し $t+1$ 期間として, 現在の状態 $\vec{i}$ に対して指定期間内で契約電力以上を使用する確率を $V_t(\vec{i})$ とする. $V_0(\vec{i})$ として, $\vec{i}$ のうち, $f+1$ 個以上の成分が l の状態であれば, $V_0(\vec{i}) = 1$ そうでなければ, $V_0(\vec{i}) = 0$ とする, これは契約電力以上になる確率を表している.

$V_t(\vec{i})$ に対して, 各部屋 k の温度レベルの推移確率 $p(j_k|i_k, s_k)$ を考える. これは部屋 k の現在の状態が i_k , 部屋 k の冷房のオンオフが s_k のもとで, 次の状態が j_k になる確率を表す.

$p(j_k|i_k, s_k)$ は

1. $\vec{i}$ のうち $f+1$ 個以上が温度レベル l のときは $p(i_k|i_k, s_k) = 1$ とする. (吸収状態)
2. $\vec{i}$ のうち f 個以下が温度レベル l のときは

$$p(i_k+1|i_k, 0) = u_k(i) \quad (\text{冷房オフで次の状態 } i_k+1)$$

$$p(i_k|i_k, 0) = 1 - u_k(i) \quad (\text{冷房オフで次の状態 } i_k)$$

$$p(i_k-1|i_k, 1) = d_k(i) \quad (\text{冷房オンで次の状態 } i_k-1)$$

$$p(i_k|i_k, 1) = 1 - d_k(i) \quad (\text{冷房オンで次の状態 } i_k)$$

最適性方程式 [2] から,

$$V_t(\vec{i}) = \min_{s_k} \left\{ \sum_{\vec{j}} \prod_{j_k} p(j_k|i_k, s_k) V_{t-1}(\vec{j}) \right\}$$

によって, 契約電力を超過する確率を最小にする最適政策を求めることができる.

4. 数値例

部屋の数 $n = 3$, 温度レベル $l = 3$ の場合について

1. 部屋 0 に対して, スイッチオフのとき, $u_0(0) = 0.75, u_0(1) = 0.8, u_0(2) = 0.85$, (状態 3 でオフはありえない) オンのとき $d_0(0) = 0, d_0(1) = 0.6, d_0(2) = 0.5, d_0(3) = 0.5$.

2. 部屋 1 に対して, スイッチオフのとき, $u_1(0) = 0.75, u_1(1) = 0.8, u_1(2) = 0.9$, オンのとき $d_1(0) = 0, d_1(1) = 0.7, d_1(2) = 0.7, d_1(3) = 0.65$.
3. 部屋 2 に対して, スイッチオフのとき, $u_2(0) = 0.77, u_2(1) = 0.9, u_2(2) = 0.9$, オンのとき $d_2(0) = 0, d_2(1) = 0.7, d_2(2) = 0.7, d_2(3) = 0.65$.

とすると, 48 期間で, すべての部屋の温度状態が 2 から始まったとき, 2 つの部屋の冷房をオンにしていくことにより, 契約電力を超過する確率は, 1.8% 程度になることが計算でわかる. 他の数値例や政策との比較については, 当日発表する予定である.

参考文献

- [1] 中国電力「契約電力の決定方法」,
http://www.energia.co.jp/elec/b_menu/h_volt3/jituryo_kettei.html
- [2] Ross S.M., *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970.

Revisiting the Heavy-Traffic Asymptotics of the GI/G/1-type Markov Chain: a Complete Proof of the Asymptotic Formula for the Stationary Distribution

Tatsuaki Kimura^{*†}

Hiroyuki Masuyama[†]

Yutaka Takahashi[†]

^{*}NTT Network Technology Laboratories, NTT Corporation

[†]Department of Systems Science, Graduate School of Informatics, Kyoto University

Abstract This paper studies the heavy-traffic limit of the stationary distribution and its moments of the GI/G/1-type Markov chain. The authors presented the same topic at the previous year's conference, however, the proof and result for the heavy-traffic limit of the moments contain errors. We first correct Kimura et al (2012)'s proof of the heavy-traffic asymptotic formula for the stationary distribution. We then present a corrected asymptotic formula for the moments of the stationary distribution based on this correction.

1. Introduction

This paper studies the heavy-traffic limits of the stationary distribution and moments of the GI/G/1-type Markov chain. The GI/G/1-type Markov chain is a mathematical model for the analysis of various semi-Markovian queueing models. Several important classes of structured Markov chains, such as quasi-birth-and-death processes (QBDs), GI/M/1-type and M/G/1-type Markov chains are included in the GI/G/1-type Markov chains. Although the stationary distributions of the GI/G/1-type Markov chains can be obtained by matrix analytical methods, they cannot be expressed in analytical forms. Thus, it is difficult to understand how system parameters impact on the stationary distribution in general. In addition, numerical methods may require high computational costs when considering heavy-traffic conditions. Therefore, the heavy-traffic asymptotics of various queueing models has been studied in decades.

To begin with, we define the GI/G/1-type Markov chain and provide some necessary definitions and assumptions. Let $\{(X_n, S_n); n \in \mathbb{Z}_+ := 0, 1, \dots\}$ denote a discrete-time Markov chain with state space $\mathbb{F} := (\{0\} \times \mathbb{M}_0) \cup (\mathbb{N} \times \mathbb{M})$, where $\mathbb{N} := \{1, 2, 3, \dots\}$, $\mathbb{M}_0 = \{1, 2, \dots, M_0\}$ and $\mathbb{M} = \{1, 2, \dots, M\}$. We call X_n and S_n a *level variable* and a *phase variable*, respectively. Arranging the states in lexicographical order, the transition probability matrix $\mathbf{T}$ of the GI/G/1-type Markov chain $\{(X_n, S_n)\}$ is given by

$$\mathbf{T} = \begin{pmatrix} \mathbf{B}(0) & \mathbf{B}(1) & \mathbf{B}(2) & \mathbf{B}(3) & \cdots \\ \mathbf{B}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \mathbf{A}(2) & \cdots \\ \mathbf{B}(-2) & \mathbf{A}(-1) & \mathbf{A}(0) & \mathbf{A}(1) & \cdots \\ \mathbf{B}(-3) & \mathbf{A}(-2) & \mathbf{A}(-1) & \mathbf{A}(0) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (1.1)$$

where $\mathbf{A}(k) := (A_{i,j}(k))_{(i,j) \in \mathbb{M}^2}$ for $k \in \mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$; $\mathbf{B}(0) = (B_{i,j}(0))_{(i,j) \in \mathbb{M}_0^2}$; $\mathbf{B}(k) := (B_{i,j}(k))_{(i,j) \in \mathbb{M}_0 \times \mathbb{M}}$ for $k \in \mathbb{N}$; and $\mathbf{B}(-k) := (B_{i,j}(-k))_{(i,j) \in \mathbb{M} \times \mathbb{M}_0}$ for $k \in \mathbb{N}$.

We assume the following assumption throughout this paper.

Assumption 1.1

- (a) $\mathbf{T}$ is irreducible and stochastic;
- (b) $\mathbf{A} := \sum_{k \in \mathbb{Z}} \mathbf{A}(k)$ is irreducible and stochastic;
- (c) $\boldsymbol{\beta}_B := \sum_{k=1}^{\infty} k \mathbf{B}(k) \mathbf{e}$ is a finite vector, where $\mathbf{e}$ denotes a column vector of ones with an appropriate order according to the context; and
- (d) $\sigma := -\boldsymbol{\pi} \boldsymbol{\beta}_A > 0$, where $\boldsymbol{\beta}_A = \sum_{k \in \mathbb{Z}} k \mathbf{A}(k) \mathbf{e}$ and $\boldsymbol{\pi}$ is the unique probability vector such that $\boldsymbol{\pi} \mathbf{A} = \boldsymbol{\pi}$.

Under Assumption 1.1, $\mathbf{T}$ is irreducible and positive-recurrent and thus has the unique and positive stationary probability vector [1, Chapter XI, Proposition 3.1]. We define $\mathbf{x} = (x_i(k))_{(k,i) \in \mathbb{F}} > \mathbf{0}$ as the stationary probability vector of $\mathbf{T}$, which is partitioned level-wise, i.e., $\mathbf{x} = (\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), \dots)$, where $\mathbf{x}(\ell)$ denotes a subvector of $\mathbf{x}$ corresponding to a level $\ell \in \mathbb{Z}_+$.

We now briefly review the previous studies. Asmussen [2] proved that the diffusion-scaled level process converges weakly to a reflected Brownian motion as the mean drift in level $-\sigma$ goes to zero under the condition that

$$\sum_{k=1}^{\infty} k^2 \mathbf{B}(k) < \infty, \quad \sum_{k \in \mathbb{Z}} |k^3| \mathbf{A}(k) < \infty. \quad (1.2)$$

As a corollary, an asymptotic formula for the stationary distribution was also presented as follows:

$$\mathbb{P}(\sigma X > x, S = i) \rightarrow e^{-\gamma x} \pi_i, \quad x \geq 0, \quad i \in \mathbb{M}, \quad \text{as } \sigma \downarrow 0, \quad (1.3)$$

where $\gamma > 0$ is a certain parameter and (X, S) denotes a random vector distributed according to the stationary distribution of $\{(X_n, S_n)\}$, i.e., $\mathbb{P}(X = k, S = i) = x_i(k)$ for $(k, i) \in \mathbb{F}$. Falin [3] proved the heavy-traffic limit (1.3) for the M/G/1-type Markov chain assuming that

$$\begin{aligned} \sum_{k=1}^{\infty} k \mathbf{B}(k) < \infty, \quad \sum_{k \in \mathbb{Z}} k^2 \mathbf{A}(k) < \infty, \\ \mathbf{x}(0) \rightarrow \mathbf{0}, \quad \text{as } \sigma \downarrow 0, \quad \mathbf{A} = \mathbf{B}(0) + \sum_{k=1}^{\infty} \mathbf{B}(k). \end{aligned} \quad (1.4)$$

Kimura et al. [4] presented the heavy-traffic asymptotic formula (1.3) for the GI/G/1-type Markov chain without Falin's additional conditions (1.4), however, their proof includes some minor errors and implicitly assumes a necessary condition on the continuity of the transition probability matrix $\mathbf{T}$ with respect to the parameter introduced to approach it to the heavy-traffic limit. Kimura et al. [5] showed a heavy-traffic formula for the moments of the stationary distribution of the GI/G/1-type Markov chain, while the presented formula and its proof include errors related to those of [4].

In this paper, we first provide a *complete* proof for the heavy-traffic asymptotic formula (1.3). We then assume that

$$\sum_{k=1}^{\infty} k^m \mathbf{B}(k) < \infty, \quad \sum_{k=1}^{\infty} k^m \mathbf{A}(k) < \infty.$$

Under this assumption, we show a corrected version of the heavy-traffic asymptotic formula of the moments in [5] such that, for $i \in \mathbb{M}$ and $m \in \mathbb{N}$,

$$\mathbb{E}[(\sigma X)^m \mathbb{1}(S = i)] \rightarrow m! \gamma^m \pi_i, \quad \text{as } \sigma \downarrow 0,$$

where $\mathbb{1}(\cdot)$ denotes the indicator function.

2. Preliminaries

To describe the main results of this paper, we provide several important definitions related to the GI/G/1-type Markov chain. Let $\mathbf{T}^{[0]}$ denote the transition probability matrix of the censored Markov chain obtained by observing $\{(X_n, S_n)\}$ only when it is in the level 0. Let $\boldsymbol{\kappa} = (\kappa_i)_{i \in \mathbb{M}_0} > \mathbf{0}$ as the (unique) stationary probability vector of the irreducible stochastic matrix $\mathbf{T}^{[0]}$. For any fixed $\nu \in \mathbb{N}$, let $[\Phi(0)]_{i,j}$ denote the probability of hitting state (ν, j) for the first time before entering the levels $0, 1, \dots, \nu - 1$, given that it starts with state (ν, i) , i.e.,

$$[\Phi(0)]_{i,j} = \mathbb{P}(S_{T_{\downarrow \nu}} = j \mid X_0 = \nu, S_0 = i),$$

where $T_{\downarrow l} = \inf\{n \in \mathbb{N}; X_n = l < X_m \ (m = 1, 2, \dots, n - 1)\}$. Furthermore, for any fixed $\nu \in \mathbb{N}$, let $[\mathbf{G}(k)]_{i,j}$ denote the probability of hitting state (ν, j) when the Markov chain $\{(X_n, S_n)\}$ enters the levels $0, 1, \dots, \nu + k - 1$ for the first time, given that it starts with state $(\nu + k, i)$, i.e.,

$$[\mathbf{G}(k)]_{i,j} = \mathbb{P}(X_{T_{< \nu+k}} = \nu, S_{T_{< \nu+k}} = j \mid X_0 = \nu + k, S_0 = i), \quad k \in \mathbb{N},$$

where $T_{<l} = \inf\{n \in \mathbb{N}; X_n < l \leq X_m \text{ } (m = 1, 2, \dots, n-1)\}$. We also define $M_0 \times M$ and $M \times M$ matrices $\mathbf{R}_0(k)$ and $\mathbf{R}(k)$ ($k \in \mathbb{Z}_+$) such that, for $k \in \mathbb{N}$,

$$\begin{aligned} [\mathbf{R}_0(k)]_{i,j} &= \mathbb{E} \left[\sum_{n=1}^{T_{<k}} \mathbb{1}(X_n = k, S_n = j) \middle| X_0 = 0, S_0 = i \right], \\ [\mathbf{R}(k)]_{i,j} &= \mathbb{E} \left[\sum_{n=1}^{T_{<k+\nu}} \mathbb{1}(X_n = k + \nu, S_n = j) \middle| X_0 = \nu \in \mathbb{N}, S_0 = i \right]. \end{aligned}$$

We now define $\tilde{\mathbf{A}}(\xi) = \sum_{k \in \mathbb{Z}} e^{i\xi k} \mathbf{A}(k)$, $\tilde{\mathbf{R}}_0(\xi) = \sum_{k=1}^{\infty} e^{i\xi k} \mathbf{R}_0(k)$, $\tilde{\mathbf{R}}(\xi) = \sum_{k=1}^{\infty} e^{i\xi k} \mathbf{R}(k)$, and $\tilde{\mathbf{G}}(\xi) = \sum_{k=1}^{\infty} e^{-i\xi k} \mathbf{G}(k)$, respectively, where $i = \sqrt{-1}$. It then holds that

$$\mathbf{I} - \tilde{\mathbf{A}}(\xi) = (\mathbf{I} - \tilde{\mathbf{R}}(\xi))(\mathbf{I} - \Phi(0))(\mathbf{I} - \tilde{\mathbf{G}}(\xi)), \quad (2.1)$$

which is known as the *RG-factorization* [6, Theorem 14]. Let $\tilde{\mathbf{x}}(\xi) = \sum_{k=1}^{\infty} e^{i\xi k} \mathbf{x}(k)$. We then have

$$\tilde{\mathbf{x}}(\xi)(\mathbf{I} - \tilde{\mathbf{R}}(\xi)) = \mathbf{x}(0)\tilde{\mathbf{R}}_0(\xi). \quad (2.2)$$

Combining (2.1) and (2.2) leads to

$$\tilde{\mathbf{x}}(\xi) = \mathbf{x}(0)\tilde{\mathbf{R}}_0(\xi)(\mathbf{I} - \Phi(0))(\mathbf{I} - \tilde{\mathbf{G}}(\xi)) \frac{\text{adj}(\mathbf{I} - \tilde{\mathbf{A}}(\xi))}{\phi(\xi)}, \quad (2.3)$$

where $\text{adj}(\cdot)$ denotes the adjugate matrix of the square matrix in the brackets and

$$\phi(\xi) = \det(\mathbf{I} - \tilde{\mathbf{A}}(\xi)), \quad \xi \in \mathbb{R}. \quad (2.4)$$

Let $\delta(\xi)$ ($\xi \in \mathbb{R}$) denote a maximum-modulus eigenvalue of $\tilde{\mathbf{A}}(\xi)$, whose imaginary part is nonnegative and whose real part is not less than those of the other eigenvalues of maximum modulus. In addition, we define $\Gamma(\xi)$ such that

$$\phi(\xi) = (1 - \delta(\xi))\Gamma(\xi).$$

In what follows, we assume the following.

Assumption 2.1 $\sum_{k \in \mathbb{Z}} k^2 \mathbf{A}(k) < \infty$.

Under Assumption 2.1, $(d^2/d\xi^2)\tilde{\mathbf{A}}(\xi)|_{\xi=0}$ exists and thus $\delta(\xi)$ and $\Gamma(\xi)$ are all twice differentiable at $\xi = 0$. As a result, we obtain the following expression of $\delta''(0)$:

$$\delta''(0) = 2\sigma^2 - \pi \sum_{k \in \mathbb{Z}} k^2 \mathbf{A}(k) \mathbf{e} - 2\pi \sum_{k \in \mathbb{Z}} k \mathbf{A}(k) (\mathbf{I} - \mathbf{A} + \mathbf{e}\pi)^{-1} \beta_A. \quad (2.5)$$

3. Main Results

In this section, we show the heavy-traffic limit of the stationary distribution and moments.

3.1. α -Parametrization

To consider the heavy-traffic asymptotics, we parameterize $\{(X_n, S_n)\}$ with a parameter $\alpha \geq 0$, which is denoted by $\{({}_{(\alpha)}X_n, {}_{(\alpha)}S_n)\}$. All the vectors, matrices and functions associated with $\{({}_{(\alpha)}X_n, {}_{(\alpha)}S_n)\}$ are also denoted with a subscript “ (α) ”, e.g., ${}_{(\alpha)}\mathbf{x}(k)$, ${}_{(\alpha)}\mathbf{T}$, ${}_{(\alpha)}\tilde{\mathbf{A}}(\xi)$ etc. In addition, we make the following assumption hereafter.

Assumption 3.1

(a) Assumptions 1.1 (a)–(c) and 2.1 hold for all $\alpha \geq 0$;

- (b) Assumption 1.1 (d) holds for all $\alpha > 0$;
- (c) as $\alpha \downarrow 0$, $_{(\alpha)}\sigma$ converges to zero from above and $_{(0)}\sigma = 0$;
- (d) $_{(\alpha)}\mathbf{A} = \mathbf{A}$ for all $\alpha \geq 0$.
- (e) $\sup_{\alpha \geq 0} \sum_{k \in \mathbb{Z}}^{\infty} k^2 {}_{(\alpha)}\mathbf{A}(k) < \infty$; and
- (f) the sequences of the matrices $\{{}_{(\alpha)}\mathbf{A}(k); k \in \mathbb{Z}\}$ and $\{{}_{(\alpha)}\mathbf{B}(k); k \in \mathbb{Z}\}$ are uniformly right-continuous at $\alpha = 0$, i.e., for any $\varepsilon > 0$, there exists some $\alpha_0 > 0$ such that, for all $k \in \mathbb{Z}$ and $0 \leq \alpha < \alpha_0$,

$$|{}_{(\alpha)}\mathbf{A}(k) - {}_{(0)}\mathbf{A}(k)| < \varepsilon, \quad |{}_{(\alpha)}\mathbf{B}(k) - {}_{(0)}\mathbf{B}(k)| < \varepsilon.$$

Under Assumption 3.1, we can show the boundedness and right-continuity at $\alpha = 0$ of the vectors, matrices related to the Markov chain $\{{}_{(\alpha)}X_n, {}_{(\alpha)}S_n\}$ as follows.

Lemma 3.1 *Under Assumption 3.1, the following are true:*

- (i) $\sup_{\alpha \geq 0} (\mathbf{I} - {}_{(\alpha)}\tilde{\Phi}(0))^{-1} < \infty$.
- (ii) $\lim_{\alpha \downarrow 0} (\mathbf{I} - {}_{(\alpha)}\tilde{\Phi}(0))^{-1} = (\mathbf{I} - {}_{(0)}\tilde{\Phi}(0))^{-1}$.
- (iii) For any $k \in \mathbb{N}$, $_{(\alpha)}\mathbf{R}(k)$, $_{(\alpha)}\mathbf{R}_0(k)$, and $_{(\alpha)}\mathbf{G}(k)$ are right-continuous at $\alpha = 0$.
- (iv) For any $\xi \in \mathbb{R}$, $_{(\alpha)}\tilde{\mathbf{A}}(\xi)$, $_{(\alpha)}\tilde{\mathbf{R}}(\xi)$, $_{(\alpha)}\tilde{\mathbf{R}}_0(\xi)$, and $_{(\alpha)}\tilde{\mathbf{G}}(\xi)$ are right-continuous at $\alpha = 0$.
- (v) $\lim_{\alpha \downarrow 0} {}_{(\alpha)}\boldsymbol{\kappa} = {}_{(0)}\boldsymbol{\kappa}$.

3.2. Heavy-Traffic Limit of Stationary Distribution

In this subsection, we consider the heavy-traffic limit for the stationary distribution. To do this, we provide several lemmas necessary for the proof of the main theorem.

Lemma 3.2 *Under Assumption 3.1, $_{(0+)}\tilde{\mathbf{x}}(0) := \lim_{\alpha \downarrow 0} {}_{(\alpha)}\tilde{\mathbf{x}}(0) = \boldsymbol{\pi}$.*

Lemma 3.3 *Under Assumption 3.1,*

$$\lim_{\alpha \downarrow 0} \frac{({}_{(\alpha)}\sigma)^2}{{}_{(\alpha)}\phi({}_{(\alpha)}\sigma\xi)} = \frac{1}{i\xi + (i\xi)^2 \frac{{}_{(0)}\delta''(0)}{2}} \frac{1}{{}_{(0)}\Gamma(0)}. \quad (3.1)$$

Lemma 3.4 *Under Assumption 3.1,*

$$\lim_{\alpha \downarrow 0} \frac{1}{{}_{(\alpha)}\sigma} (\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{G}}({}_{(\alpha)}\sigma\xi)) \text{adj}(\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{A}}({}_{(\alpha)}\sigma\xi)) = i\xi {}_{(0)}\Gamma(0) {}_{(0)}\mathbf{d}_0 \boldsymbol{\pi}, \quad (3.2)$$

where

$${}_{(0)}\mathbf{d}_0 = {}_{(0)}\boldsymbol{\beta}_G + (\mathbf{I} - {}_{(0)}\mathbf{G})(\mathbf{I} - \mathbf{A} + \mathbf{e}\boldsymbol{\pi})^{-1} {}_{(0)}\boldsymbol{\beta}_A. \quad (3.3)$$

Lemma 3.5 *Under Assumption 3.1,*

$$\lim_{\alpha \downarrow 0} \frac{{}_{(\alpha)}\mathbf{x}(0)}{{}_{(\alpha)}\sigma} = \nu_* \cdot {}_{(0)}\boldsymbol{\kappa},$$

where ν_* is a finite positive number such that

$$\nu_* = [{}_{(0)}\boldsymbol{\kappa} {}_{(0)}\mathbf{R}_0(\mathbf{I} - {}_{(0)}\tilde{\Phi}(0)) {}_{(0)}\mathbf{d}_0]^{-1}. \quad (3.4)$$

Remark 3.1 Lemma 3.5 is a corrected version of Lemma 3.4 in [4], in which the expression of C (corresponding to ν_* in (3.4)) is incorrect. This error is mainly due to the error in Proposition 2.3 in [4]. Fortunately, the main result shown in Theorem 3.1 still holds and is not affected by this correction.

The following theorem shows the heavy-traffic asymptotic formula for the stationary distribution.

Theorem 3.1 Under Assumption 3.1,

$$\lim_{\alpha \downarrow 0} {}_{(\alpha)}\tilde{\mathbf{x}}({}_{(\alpha)}\sigma\xi) = \frac{1}{1 - i\xi\gamma} \boldsymbol{\pi}, \quad \xi \in \mathbb{R}, \quad (3.5)$$

where

$$\gamma = \frac{1}{2} \boldsymbol{\pi} \sum_{k \in \mathbb{Z}} k^2 {}_{(0)}\mathbf{A}(k) \mathbf{e} + \boldsymbol{\pi} \sum_{k \in \mathbb{Z}} k {}_{(0)}\mathbf{A}(k) (\mathbf{I} - \mathbf{A} + \mathbf{e}\boldsymbol{\pi})^{-1} {}_{(0)}\boldsymbol{\beta}_A. \quad (3.6)$$

Remark 3.2 Theorem 3.1 shows that

$$\lim_{\alpha \downarrow 0} \mathbb{P}({}_{(\alpha)}\sigma({}_{(\alpha)}X > x, {}_{(\alpha)}S = i) = e^{-\gamma x} \pi_i, \quad x \geq 0, \quad i \in \mathbb{M}.$$

Proof. Lemma 3.2 shows that (3.5) holds for $\xi = 0$. Thus, we consider the case of $\xi \neq 0$. From (2.3), we have

$$\begin{aligned} \lim_{\alpha \downarrow 0} {}_{(\alpha)}\tilde{\mathbf{x}}({}_{(\alpha)}\sigma\xi) &= \lim_{\alpha \downarrow 0} \frac{{}_{(\alpha)}\mathbf{x}(0)}{{}_{(\alpha)}\sigma} {}_{(\alpha)}\tilde{\mathbf{R}}_0({}_{(\alpha)}\sigma\xi) (\mathbf{I} - {}_{(\alpha)}\boldsymbol{\Phi}(0)) \\ &\quad \times \frac{\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{G}}({}_{(\alpha)}\sigma\xi)}{{}_{(\alpha)}\sigma} \text{adj}(\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{A}}({}_{(\alpha)}\sigma\xi)) \frac{({}_{(\alpha)}\sigma)^2}{{}_{(\alpha)}\phi({}_{(\alpha)}\sigma\xi)}. \end{aligned} \quad (3.7)$$

Applying Lemmas 3.1 and 3.3–3.5 to (3.7) yields

$$\lim_{\alpha \downarrow 0} {}_{(\alpha)}\tilde{\mathbf{x}}({}_{(\alpha)}\sigma\xi) = \frac{1}{1 - i\xi \frac{{}_{-(0)}\delta''(0)}{2}} \boldsymbol{\pi},$$

from which and (2.5), we obtain (3.5) and (3.6). $\square$

Remark 3.3 Kimura et al. [4] implicitly assumed that Assumptions 3.1 (e) and (f) hold and all vectors, matrices appeared in the proof of Theorem 3.1 are bounded and right-continuous at $\alpha = 0$. These facts are proved in Lemma 3.1.

4. Heavy-Traffic Limit of Moments

In this section, we provide the heavy traffic asymptotic formula for the moments of the stationary distribution. In what follows, for any function f (including vectors and matrix functions), we write $f^{(n)}$ ($n \in \mathbb{N}$) as an n times differential of f .

Before giving the main theorem, we provide two lemmas.

Lemma 4.1 If Assumption 3.1 (a)–(c) hold, then the following are true for any $m \in \mathbb{N}$:

- (i) If $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{A}(k) < \infty$, then $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^m {}_{(\alpha)}\mathbf{R}(k) < \infty$.
- (ii) If $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{A}(-k) < \infty$, then $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^m {}_{(\alpha)}\mathbf{G}(k) < \infty$.
- (iii) If $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{B}(k) < \infty$, then $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^m {}_{(\alpha)}\mathbf{R}_0(k) < \infty$.

Lemma 4.2 Suppose that Assumption 3.1 holds. If $\sup_{\alpha \geq 0} \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{A}(k) < \infty$ for some $m \in \mathbb{Z}_+$, then,

$$\lim_{\alpha \downarrow 0} {}_{(\alpha)}\sigma \frac{d^m}{d\xi^m} \left(\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{R}}({}_{(\alpha)}\sigma\xi) \right)^{-1} = \frac{m!}{(1 - i\xi\gamma)^{m+1}} (i\gamma)^m (\mathbf{I} - {}_{(0)}\boldsymbol{\Phi}(0)) {}_{(0)}\mathbf{d}_0 \boldsymbol{\pi}, \quad (4.1)$$

where γ and ${}_{(0)}\mathbf{d}_0$ are given in (3.6) and (3.3), respectively.

Remark 4.1 Lemma 4.2 is a corrected version of Lemma 4.2 in [5]. This error comes from the error in Lemma 3.4 in [4] (see also Remark 3.1).

Theorem 4.1 Suppose that Assumption 3.1 holds. If

$$\sup \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{A}(k) < \infty, \quad \sup \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{B}(k) < \infty,$$

for some $m \in \mathbb{Z}_+$, then,

$$\lim_{\alpha \downarrow 0} \frac{d^m}{d\xi^m} {}_{(\alpha)}\tilde{\mathbf{x}}({}_{(\alpha)}\sigma\xi) \Big|_{\xi=0} = m! (i\gamma)^m \cdot \boldsymbol{\pi}, \quad (4.2)$$

where γ is given in (3.6).

Remark 4.2 Theorem 4.1 in [5] states that the parameter (denoted as η) in the heavy-traffic limit formula of the moments is not equal to γ in (3.6), however, this is incorrect as shown in Theorem 4.1.

Proof. By differencing m times (2.2) with respect to ξ , we obtain

$$\begin{aligned} \frac{d^m}{d\xi^m} {}_{(\alpha)}\tilde{\mathbf{x}}^{(m)}({}_{(\alpha)}\sigma\xi) &= {}_{(\alpha)}\mathbf{x}(0) \sum_{\ell=0}^m \binom{m}{\ell} ({}_{(\alpha)}\sigma)^\ell {}_{(\alpha)}\tilde{\mathbf{R}}_0^{(\ell)}({}_{(\alpha)}\sigma\xi) \\ &\quad \times \frac{d^{m-\ell}}{d\xi^{m-\ell}} \left(\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{R}}({}_{(\alpha)}\sigma\xi) \right)^{-1}. \end{aligned} \quad (4.3)$$

According to Lemma 4.1, $\sup \sum_{k=1}^{\infty} k^{m+1} {}_{(\alpha)}\mathbf{B}(k) < \infty$ implies that $\sup \sum_{k=1}^{\infty} k^\ell {}_{(\alpha)}\mathbf{R}_0(k) < \infty$ for all $\ell \leq m$. Thus, applying the dominated convergence theorem, Lemmas 3.5 and 4.2 to (4.3), we obtain

$$\begin{aligned} &\lim_{\alpha \downarrow 0} \frac{d^m}{d\xi^m} {}_{(\alpha)}\tilde{\mathbf{x}}^{(m)}({}_{(\alpha)}\sigma\xi) \\ &= \lim_{\alpha \downarrow 0} \frac{{}_{(\alpha)}\mathbf{x}(0)}{{}_{(\alpha)}\sigma} \sum_{\ell=0}^m \binom{m}{\ell} ({}_{(\alpha)}\sigma)^\ell {}_{(\alpha)}\tilde{\mathbf{R}}_0^{(\ell)}({}_{(\alpha)}\sigma\xi) ({}_{(\alpha)}\sigma) \frac{d^{m-\ell}}{d\xi^{m-\ell}} \left(\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{R}}({}_{(\alpha)}\sigma\xi) \right)^{-1} \\ &= \lim_{\alpha \downarrow 0} \frac{{}_{(\alpha)}\mathbf{x}(0)}{{}_{(\alpha)}\sigma} {}_{(\alpha)}\tilde{\mathbf{R}}_0({}_{(\alpha)}\sigma\xi) \cdot \lim_{\alpha \downarrow 0} \left[({}_{(\alpha)}\sigma) \frac{d^m}{d\xi^m} \left(\mathbf{I} - {}_{(\alpha)}\tilde{\mathbf{R}}({}_{(\alpha)}\sigma\xi) \right)^{-1} \right] \\ &= \frac{m!(i\gamma)^m \nu_*}{(1 - i\xi\gamma)^{m+1}} {}_{(0)}\boldsymbol{\kappa}_{(0)} \mathbf{R}_0 (\mathbf{I} - {}_{(0)}\boldsymbol{\Phi}(0)) {}_{(0)}\mathbf{d}_0 \boldsymbol{\pi} = \frac{m!(i\gamma)^m}{(1 - i\xi\gamma)^{m+1}} \boldsymbol{\pi}, \end{aligned}$$

where we use (3.4) in the last equality. Consequently, letting $\xi = 0$ in the above equation yields (4.2).

References

- [1] S. Asmussen, *Applied Probability and Queues*, Second Edition, Springer, New York, 2003.
- [2] S. Asmussen, The heavy traffic limit of a class of Markovian queueing models, *Operations Research Letters* 6 (1987) 301–306.
- [3] G. Falin, Heavy traffic analysis of a random walk on a lattice semi-strip, *Stochastic Models* 11 (1995) 395–409.
- [4] T. Kimura, H. Masuyama and T. Takahashi, Heavy-traffic asymptotics for stationary GI/G/1-type Markov chains, *Operations Research Letters* 40 (2012) 185–189.
- [5] T. Kimura, H. Masuyama and T. Takahashi, Heavy-traffic limits of the moments of the stationary distribution of the GI/G/1-type Markov chain, In *Proc. of Queue Symposium*, (2016).
- [6] Y. Q. Zhao, W. Li, W. J. Braun, Censoring, factorizations, and spectral analysis for transition matrices with block-repeating entries, *Methodology and Computing in Applied Probability* 5 (2003) 35–58.

マルチエージェントモデルを用いたセルフレジ導入による レジサービスへの影響と解析

平田直也 中桐齊之
兵庫県立大学 環境人間学科

概要：現代の販売店では、通常よく目にする有人レジとは別に自分で商品のスキャンを行うセルフレジが存在する。人件費の削減やレジサービスの向上などを目的として導入が進められているが、実際にはセルフレジを利用する人はあまり見ない。そこで、マルチエージェントモデルを用いてセルフレジのシミュレーションを作った。その結果、雇用コスト削減においては効果があるが、レジサービスの効率化という観点からはあまり影響がないことがわかった。

1. はじめに

スーパーマーケットやコンビニエンスストアなどは、買い物をするために多くの人に利用されている。それらの店舗の中には、顧客が購入した商品を精算するためのレジが存在するが、その顧客が店舗に集中する時間帯等ではしばしば行列ができることがある。

そこで、近年、スーパーマーケットをはじめとして、レンタルDVDショップ、ファストフード店などではセルフレジを導入する店舗が増加してきている。ここで、セルフレジとは、レジにおける商品のスキャン、精算を客自身ですべて行うレジのことを指す。この導入による効果としては、従業員の雇用コストを削減できることやレジサービスの効率化、人の手による現金管理の手間の省略などがある[1]。

しかし、このセルフレジにおいては、顧客によっては、顧客自身で商品のスキャンを行うことに対して億劫に感じる客も存在し、既存の有人レジに好んで並ぶことがある。また、顧客自身がスキャンと清算を行うため、顧客がセルフレジの扱いが不慣れであるときは、他の顧客と比較して非常に多くの時間がかかってしまうこともある。したがって、セルフレジは、雇用コスト削減においては効果があるが、レジサービスの効率化という観点からは、その効果がわかりにくい。

こういった、店舗のレジにおける研究としては、従来の有人レジにおけるレジ周りの行動シミュレータの開発[2,3]や、その待ち行列に関する研究[4]などが存在する。しかし、近年、広まりつつあるセルフレジや、そのさらなる発展版として注目されているセミセルフレジに関する研究は少なく、その効果については、まだ詳しく解析されていない。

そこで、本研究では、このセルフレジやセミセルフレジの待ち行列モデルを可視化し、その効果を解析できるシミュレータを開発することとした。また、その結果から、現在のセルフレジの課題を発見し、その解決策を提案できるのではないかと考えた。

2. セルフレジとは

セルフレジとは、一般的に4台のレジが1組となったレジサーバーのことを指し、1度に4人までの客の精算処理を行うことができる。複数のレジに対して1列で並び、空いたレジに随時客が入っていく。レジには店員が配置されておらず、レジに入った顧客が顧客自身で商品をスキャンし、精算を行い、その後の袋詰め等を全て自分で行う。ここで、セルフレジには、複数台のレジを監視する店員が基本的に1名常駐しているが、その店員が行う業務は「お買い上げシールを貼る」「機器トラブルのサポート」「年齢確認」などに限定され、作業量は少ない。ゆえに従業員は1名で十分であるため、従業員の雇用コストが少なく済む。人件費削減においては非常に有効な手段と言えるシステムである。しかしながら、顧客が精算を行うため億劫に感じてセルフレジを避ける人や、セルフレジで非常に操作が遅い人が時折存在する。商品のスキャンに熟達した店員がいる有人レジと異なり、セルフレジにおいては、顧客自身のスキャン速度に依存してサービス時間が変化する。

レジ待機列の並び方には、2種類が存在する。1つ目は、レジ一つに対して1つの列を形成するM/M/1型モデルと呼ばれるもので、基本的なスーパーマーケット等の有人レジにおいて見られる列

がある。もう一つは、複数のレジに対して1つの列を形成し、空いたレジから一人ずつ客が入っていく $M/M/s$ 型モデルと呼ばれるものである。こちらは、主にコンビニエンスストア等で見られ、通常、セルフレジはこの並び方を採用している。 $M/M/1$ 型も $M/M/s$ 型も、客がレジへ到達する時間、レジ台数、サービスにかかる時間（商品のスキャン時間+精算にかかる時間）というパラメータを用いて、待ち行列としてモデルを構築することができる。店舗によっては、人が混み合う夕方など、時間帯によっては列形成が積極的になるため、客の混雑度や稼働するレジ台数によって待ち行列の状態が変化することもある。

3.モデル

本研究では顧客をエージェントとして動かすマルチエージェントモデルを構築することとした。

2次元の空間上にレジを設置し、その空間上で顧客(以下、エージェントと呼ぶ)を配置する。エージェントは、入店したとき、既に商品を選択し終えており、精算するためのレジを探して空間上を移動し、レジの選択、列に並ぶ、商品をスキャンする、退店するといった行動を行う。具体的に、エージェントは次の①～⑤のどれかの行動をとる。

- 1 待ち行列の選択 エージェントは、入店後、すぐに、全てのレジに並んでいるエージェントの数を調べ、最も人数が少ないレジを目的地に設定する。そのレジに誰も並んでいないときは、レジの位置を、人が並んでいるときは、そのレジにできている待ち行列の最後尾の地点を目的地に設定する。
- 2 移動 エージェントは、目的地に向かって移動を行う。このとき、エージェントが待機列に存在する場合は、前方の座標を常に参照し、エージェントがいれば停止、いなくなれば、前へ進むとする。
- 3 待ち行列の移動 エージェントは、現在並んでいる待ち行列に隣接するすべてのレジに並んでいるエージェントの数を数え、隣接するレジの待ち行列が短かった場合、その列へと移動する。
- 4 スキャンと清算 エージェントはレジに到着したとき、商品のスキャンと精算を行う。このとき、1ステップごとに自分の持っている商品数 C を1つ減らしていく。商品数が $C=0$ になった時点で、精算処理(Sステップ)を行い、レジを出る。
- 5 退店 レジを出たエージェントは、退店したとみなし、空間上からエージェントを削除する。

これら①～⑤の行動を入店した全ての客エージェントが行い、時間(T)を1増加させる。

図1に、客エージェントの行動のフローチャートを示す。

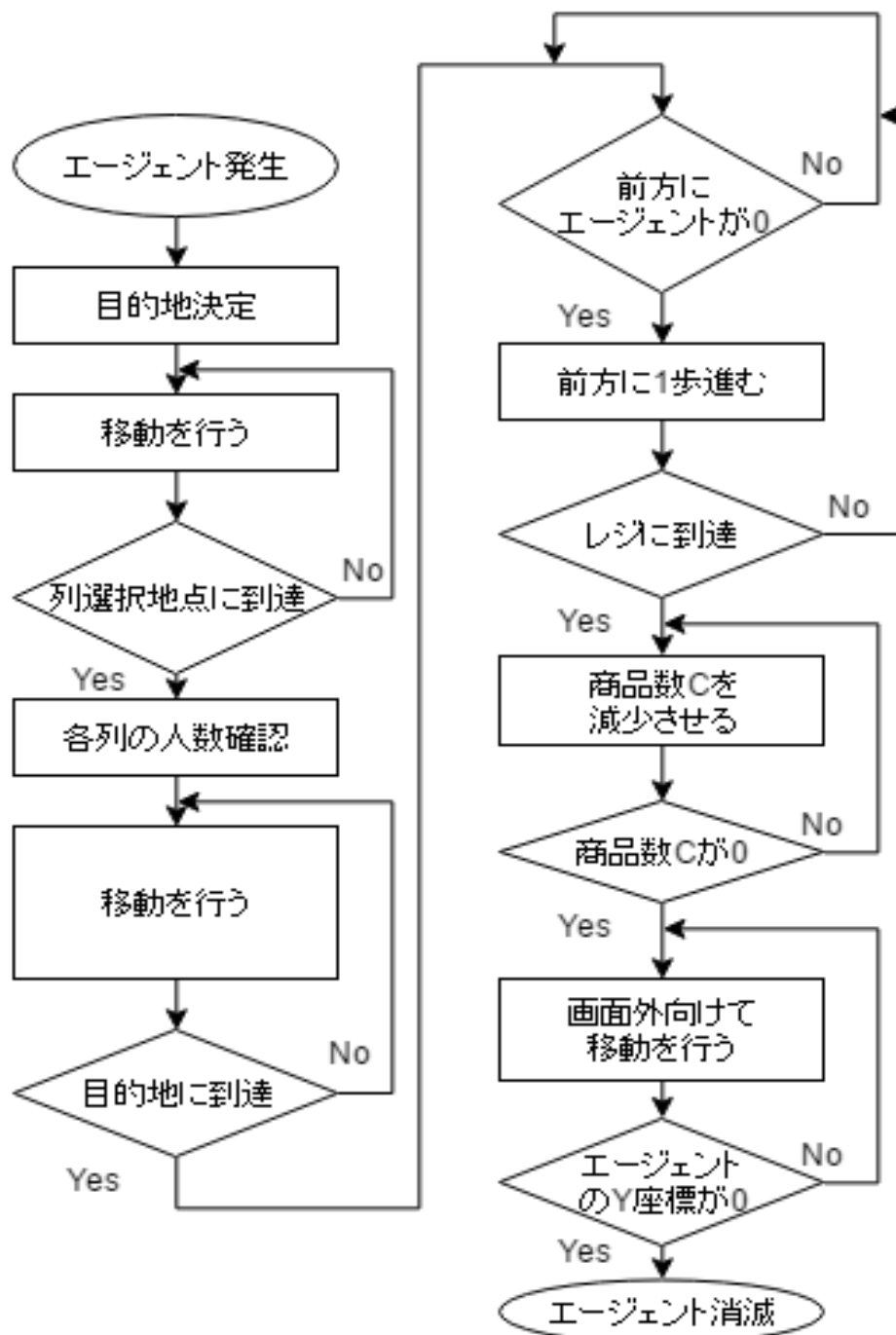


図1. エージェントの行動

ここで、エージェントには個性をもたせるため、以下の要素を持ったエージェントが出現することとした。A,B,Cの要素を持つ確率は、それぞれA=10%,B=2%,C=5%とする。

A. 隣が空いていても自分の位置から最も近い位置のレジに並ぶ

B. 商品数が極端に多い

C. スキャンの速度が非常に遅い

Bはいわゆる爆買いをしていく外国人、Cは年配の方などに対応している。

次に、本研究を行う上で、様々なシチュエーションに対応するために、客やレジなどが持つ数値をパラメータとして設定した。表1にパラメータを示す。

パラメータ	説明
レジ台数 N	レジの台数を表しており、値は2～9の範囲内で変化させることができる。
最小商品数 C_{min}	客エージェントが持つ商品数の下限。設定した商品数より少ない商品を持つ客は出現しない。値は1～5の範囲内で変化させることができる。
最多商品数 C_{max}	客エージェントが持つ商品数の上限。設定した商品数より多い商品を持つ客は出現しない。値は6～20の範囲内で変化させることができる。
商品数 C	上記の C_{min}, C_{max} の範囲内で、各エージェントはそれぞれの商品数を持つ。

表 1. パラメーター一覧

左右で来店の割合を変化させているのは、店内のレイアウト等によって客の人流が変化してくることがあるためである[6]。例えば出入り口がレジから見て右側にあるスーパーであれば、店内を一周した後にレジへと向かうため、基本的に左から人が多く流れてくることがとなる。その店の状況や条件下でシミュレーションを行うために、任意で割合を設定できるようにした。

空間には、エージェント以外にレジオブジェクトが存在する。これは、レジの数が設定した数によって、店員の数も変化するため、それを可視化するために設置した。オブジェクトではあるが、レジ業務を行うレジには、スキャン速度が存在し、各レジでスキャンをする店員がベテランか初心者かなどによって異なる。ゆえに、それぞれのレジには速度 V_n を設け、レジの存在するレジによって商品数 C が減少する速度が V_n によって異なるようにした。

実装は、構造計画研究所が制作したマルチエージェントシミュレータ「artisoc 4.0」[5]を用いて行い、コンピュータを用いて、レジサービスについての客の行動をシミュレーション実験することとした。

4. シミュレーション実験

セルフレジの有無によってレジサービスにどのような影響が及ぼされるかを調べるため、シミュレーション実験を条件①②③について行った。セルフレジを4台一組としたとき、レジの横並びのスペースとしてはレジ2台分の幅で4台がほぼ収まるため、有人レジ2台分につきセルフレジ4台が設置可能とした。このとき、比較用として、有人レジのみの方法（①）、有人レジ2台分のスペースをそのままセルフレジに置き換える方法（②）と、有人レジのみの時のレジサーバーの数が同じになるよう調製した方法（③）の3条件で実験を行った。具体的なレジ台数は以下のとおりである。

- ①セルフレジ無し ($N_p = 0$)、有人レジ9台 ($N_s = 9$) （合計9台）
- ②セルフレジ4台 ($N_p = 4$)、有人レジ7台 ($N_s = 7$) （合計11台）
- ③セルフレジ4台 ($N_p = 4$)、有人レジ5台 ($N_s = 5$) （合計9台）

シミュレーションは客エージェント $P=2000$ [人]が来店し、全ての客が退店を行うまで実験を行った。その際、列に並んだり商品をスキャンするために客エージェントが停止している時間を待ち時間 W として計測した。

レジ台数 N を除く各パラメータは以下の通りとしてシミュレーションそれぞれ10回ずつ行った。

最小商品数 $C_{min} = 1$	客総数 $P = 2000$
最大商品数 $C_{max} = 20$	有人レジスキャン速度 $v = 1$

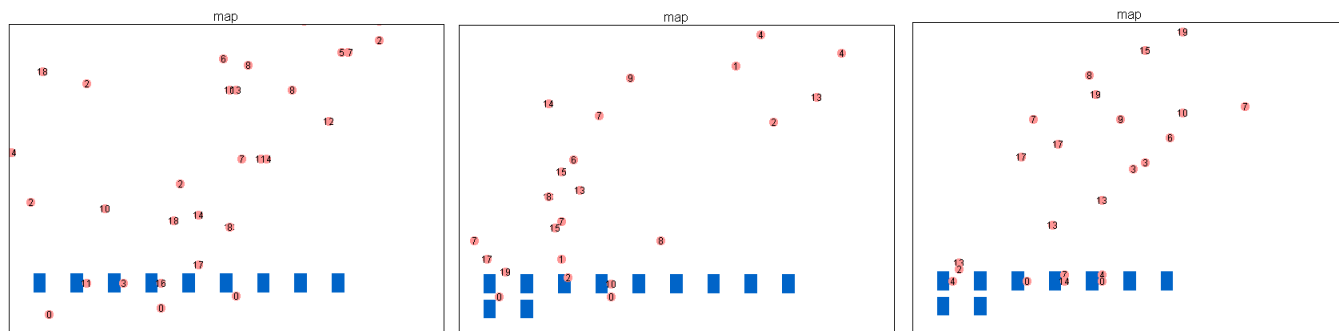


図2. $T=100$ のときのシミュレーション画面(左から条件①、②、③)

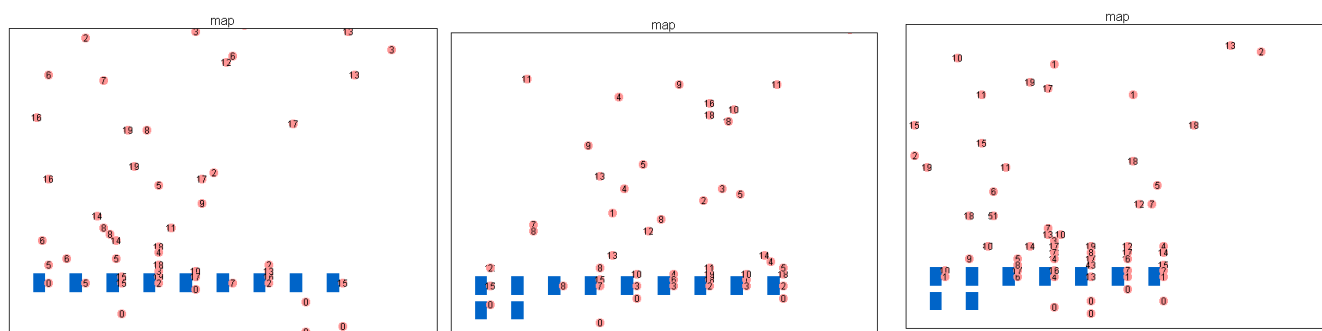


図3. $T=500$ のときのシミュレーション画面(左から条件①、②、③)

すべてのエージェントが画面から消えた時の時間の平均値 $T_{\max}$ を図4に、エージェント一人当たりの平均待ち時間 W を図5に示す。

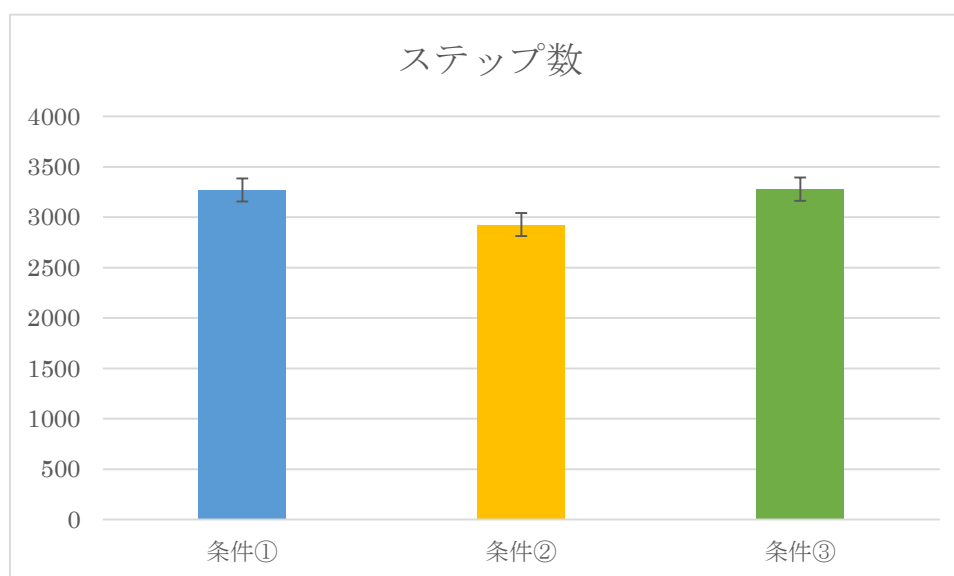


図4. 平均ステップ数

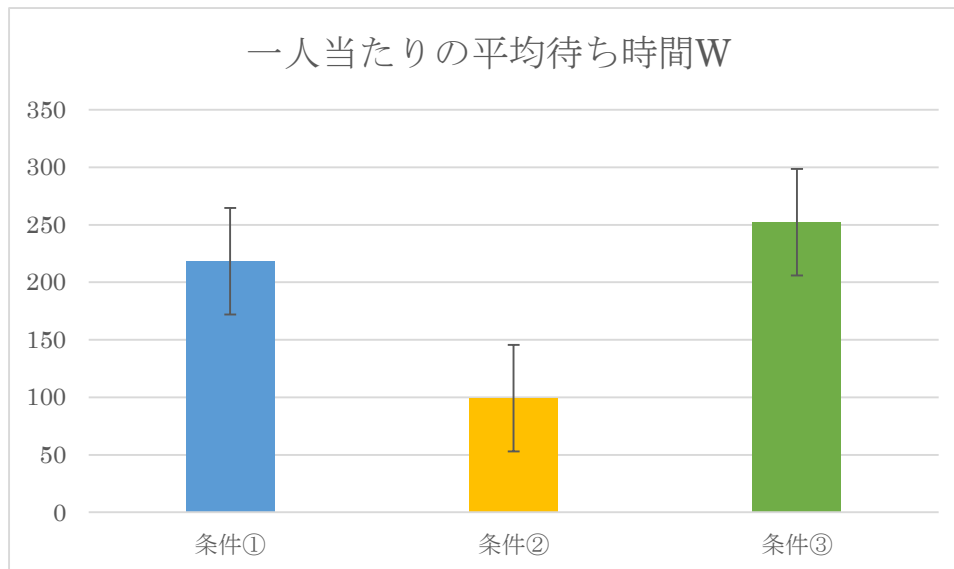


図5. エージェント一人当たりの平均待ち時間 W

図4より、すべてのエージェントが精算処理される時間 T_{max} にはあまり差がないことがわかる。

図5より、エージェント一人当たりの平均待ち時間 W は②のみ少ないことがわかる。

また、セルフレジ単体の効力を調べるために、「セルフレジ4台のみ」と「有人レジ4台のみ」とし、客エージェント数 P を $P=500$ [人]として同様にシミュレーションを行った。結果を図6,7に示す。

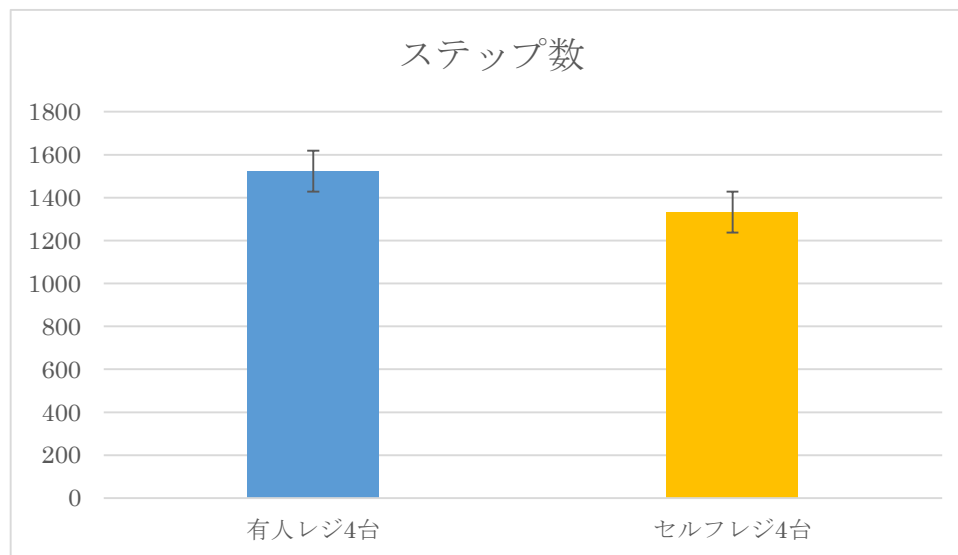


図6. すべてのエージェント($P=500$ [人])が処理されるまでの時間 T_{max}

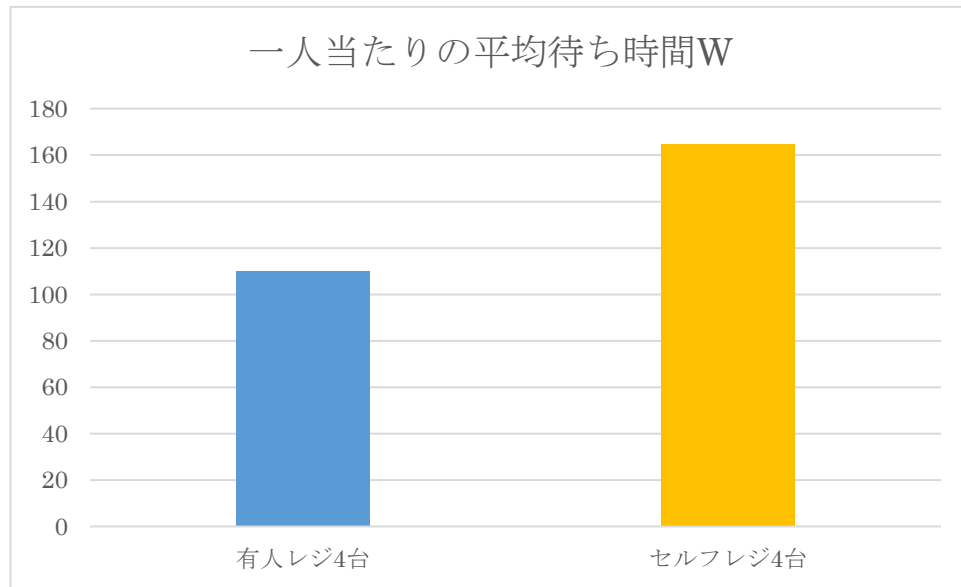


図7. エージェント一人当たりの平均待ち時間 W

図6より、セルフレジのほうが T_{max} が小さいことがわかる。また、図7よりセルフレジのほうが一人当たりの待ち時間 W が大きいことがわかる。

5. 考察

セルフレジは有人レジよりもレジサービスの処理能力が高いのか、また、効率的よくレジサービスを行うことができるにはどうすればいいのかを解析するため、本シミュレーションを行った。図4,5より、有人レジとセルフレジとでは処理能力に関してあまり大きな差は見られなかった。これはレジの違いではなく、レジサーバーの数によって処理能力が高まっていくためである。よって、レジサービスの効率という点においてはにおいて有人レジもセルフレジも特に差異はないといえる。また、図6より、セルフレジ単体で見れば客を捌いていく速度はセルフレジのほうが速いように思えるが、図7からは、客一人一人に待たせる時間が長くなっている。これはセルフレジが M/M/s 型の待ち行列であるため、1列で客を並ばせる分待ち時間が長くなってしまったと考える。

セルフレジは客の処理能力にあまり差は見られなかったが、言い換えると有人レジと同等の能力を有していることとなる。セルフレジは本来4台のセルフレジに対して1人の監視スタッフが常駐しているが、その一人分で4台分のレジサーバーを管理することができるため、同じ時間レジを稼働させていても、約3人分の人件費削減となる。また、図2,3のシミュレーション画面より、有人レジをセルフレジに変えることで、同じレジ台数でも店内にスペースが生まれ、その場所を商品陳列や新たなレジ設置に用いることが可能であり、レジサービスや売上の向上にもつながる。

6. 結論

本研究ではセルフレジのサービス効果について検証を行ったが、シミュレーション上では同等の能力を有していることがわかった。また、人件費削減や店内レイアウトなど主に店舗側への利点を発見することができた。今後は時間条件で客の来店確率を変化させたり、商品数が少ない客を優先的にセルフレジへ向かわせるとどうなるかなど、様々な条件下で実験を行うことで、より現実的なシミュレーションの開発をしていきたい。

参考文献

- [1] 住友信託銀行, 産業界の動き～小売業のコスト削減が期待されるセルフレジ, 調査日報 2010年3月号
- [2] 芹沢 良, マルチエージェントを使用したレジにおける混雑解消法の検証, 東京工科大学メディア学部2005年度卒業論文, 2006年3月
- [3] 加藤 菜美絵, 作業能力を考慮した人員配置問題を解くマルチエージェントモデルの構築, 群馬大学社会情報学部社会情報学科2006年度卒業論文, 2007年4月
- [4] 森 雅俊, スーパーの食料品売り場のレジにおける待ち行列の解析, 南山大学数理情報学部数理科学科2003年度卒業論文, 2004年
- [5] 「artisoc4.0」(構造計画研究所)
MAS コミュニティ - artisoc4.0, <http://mas.kke.co.jp/modules/tinyd0/index.php?id=13>
(2016年12月10日アクセス)
- [6] 山坂 つばさ, マルチエージェントシミュレーションを用いたファッションアパレル店舗モデルに関する研究 信州大学繊維学部2015年度卒業論文

An arrival-time distribution for the equilibrium mean waiting time of a discrete-time single-server queue with acceptance period and Poissonian population of customers

Yutaka Sakuma¹

Masahiro Kobayashi²

Hiroyuki Masuyama³

¹School of Electrical and Computer Engineering, National Defense Academy

²Graduate School of Science, Tokai University

³Graduate School of Informatics, Kyoto University

Abstract We study a discrete-time first-come first-served (FCFS) single-server queue with an acceptance period and no early arrival. The number of arriving customers is Poisson distributed, and their service times are generally distributed. Customers choose their arrival times with the goal of minimizing their expected waiting times. In this study, we show an arrival-time distribution of customers for the equilibrium mean waiting time.

1. Introduction

Many real-life queueing systems have acceptance periods for arriving customers, that is, the system accepts arriving customers only during the period between the opening and closing times. Typical examples of such queueing systems are restaurant in lunch time, service counter at bank (or government office), and rush-hour congestion in transportation networks. Customers of these systems face with the decision problem of when to arrive at the systems so as to achieve a certain goal, e.g., to minimize the waiting time for service. Customers' decisions on arrival, naturally, interact with each other, and thus the arrival times of customers are endogenously determined.

Glazer and Hassin [1]'s work is a pioneer study on the decision problem on when to arrive at the queueing system with an acceptance period. Glazer and Hassin studied a continuous-time first-come first-served (FCFS) single-server queue with an acceptance period, where a Poisson-distributed number of homogeneous customers arrive at the system and customers may arrive at the system before its opening time, i.e., early arrival of customers is allowed. The service times are assumed to be independently and identically distributed according to an exponential distribution. Glazer and Hassin assumed that customers choose their arrival times with the goal of minimizing their expected waiting times. More specifically, they studied the queueing model by a non-cooperative game with a random number of players, and obtained an equilibrium strategy of arriving customers as a mixed strategy.

Similar queueing systems have been studied. Hassin and Kleiner [2] studied a queueing system without early arrival, and obtained an equilibrium strategy of arriving customers. Hassin and Kleiner [2] also reported that no early arrivals reduce the mean waiting time in equilibrium especially when the system is heavily loaded (see Figure 3 therein). Haviv [3] and Ravner [4] studied queueing models with a tardiness cost and an arrival order cost, respectively, as well as waiting cost.

In this paper, we consider a discrete-time FCFS single-server queue with an acceptance period and no early arrivals, and discuss an arriving-time distribution achieving the equilibrium mean waiting time, which is referred to as the *equilibrium arriving-time distribution*. As in the preceding studies [1, 2, 3, 4], we assume that the number of arriving customers is Poisson distributed. However, our study has three differences from these preceding studies. First, our model is in discrete time whereas the existing models in the preceding studies are in continuous time. The assumption of discrete time facilitates the computation of an equilibrium mean waiting time and the corresponding equilibrium arriving-time distribution. Second, the service times in our model are generally distributed whereas those in the existing models are exponentially distributed. Thus, our model enables us to investigate the effect of the service time distribution on the arrival strategy of customers. Third, the waiting cost of our model is the mean *actual* waiting time whereas that

of the existing models is the mean *virtual* waiting time.

The rest of the paper is organized as follows. Section 2 describes the queueing model studied in this paper, and provides some fundamental results on the mean workload and the actual waiting time. Section 3 introduces the notion of equilibrium mean waiting time and equilibrium arriving-time distribution, and then presents a procedure for calculating an equilibrium mean waiting time and the corresponding equilibrium arriving-time distribution.

2. Preliminaries

2.1. Model description

We consider a discrete-time first-come, first-served (FCFS) queueing system with infinite waiting room and one server as follows. The time axis of the system is divided into fixed-length time intervals, where each time interval is referred to as slot t for nonnegative integer $t \in \mathbb{Z}_+ := \{0, 1, 2, \dots\}$. Without loss of generality, each time slot is assumed to have length one. The system is open for arriving customers during an acceptance period, which is defined by a set of slots $\mathcal{T} := \{0, 1, 2, \dots, T\}$ for a positive integer T . The arriving customers entering during the same time-slot are served in random order. We assume that the server is available to continue service until all customers who arrived during the acceptance period $\mathcal{T}$ are served.

In what follows, some probabilistic assumptions on our model are listed.

Assumption 2.1 (Population and arrival time distribution) The number of customers seeking service from the system is denoted by a Poisson random variable A with positive mean λ , and each arriving customers independently chooses its arrival time from the acceptance period $\mathcal{T}$ with a common probability distribution $\mathcal{P} := \{p_t; t \in \mathcal{T}\}^1$.

Assumption 2.2 (Arrival instants of customers) In each slot $t \in \mathcal{T}$, the arrival of customers can occur immediately after the slot starts. Let A_t , $t \in \mathcal{T}$, denote the number of arrivals in slot t , then we have from Assumption 2.1, for $t \in \mathcal{T}$,

$$\mathbb{P}(A_t = n) = e^{-\lambda p_t} \frac{(\lambda p_t)^n}{n!}, \quad n \in \mathbb{Z}_+. \quad (2.1)$$

Assumption 2.3 (Service requirement and processing instants) The service times of customers arriving at the system in slots 0 through T are independent and identically distributed (i.i.d.) with discrete distribution $\{b(k); k \in \mathbb{N}\}$ having finite positive mean $\bar{b}$. Thus, let B denote a generic random variable for the service time. It then follows that $\mathbb{P}(B = k) = b(k)$ for $k \in \mathbb{N} := \{1, 2, 3, \dots\}$ and

$$\bar{b} := \mathbb{E}[B] = \sum_{k=1}^{\infty} k b(k). \quad (2.2)$$

If the server holds a customer in the middle of slot t ($t \in \mathbb{Z}_+$), it processes the customer's service requirement by 1 at the end of the slot.

We note that the system is always stable because the acceptance period $\mathcal{T}$ is a finite set and the expected total workload into the system is finite, i.e., $\mathbb{E}[A]\bar{b} = \lambda\bar{b} < \infty$. For $t \in \mathcal{T}$, let $B_{t,i}$'s, $i \in \{1, 2, \dots, A_t\}$, denote the service times of the customers arriving at slot t . For $t \in \mathcal{T}$, let X_t denote the total service times of the customers arriving at slot t , and $x_t(k) = \mathbb{P}(X_t = k)$ for $k \in \mathbb{N}$. It then follows from Assumption 2.2 and 2.3 that

$$X_t = \sum_{i=1}^{A_t} B_{t,i}, \quad t \in \mathcal{T}. \quad (2.3)$$

¹⁾The common arrival distribution $\mathcal{P}$ is referred to as a symmetric arrival strategy profile in Haviv 2013.

and thus

$$\sum_{k=0}^{\infty} z^k x_t(k) = \sum_{n=0}^{\infty} e^{-\lambda p_t} \frac{(\lambda p_t)^n}{n!} (\tilde{b}(z))^n = \exp \left\{ -\lambda p_t (1 - \tilde{b}(z)) \right\}, \quad (2.4)$$

where

$$\tilde{b}(z) = \sum_{k=1}^{\infty} z^k b(k).$$

Furthermore, since $\mathbb{E}[B] = \bar{b}$ and $\mathbb{E}[A_t] = \lambda p_t$ for $t \in \mathcal{T}$, we have

$$\mathbb{E}[X_t] = \mathbb{E}[A_t] \mathbb{E}[B] = \lambda p_t \bar{b}, \quad t \in \mathcal{T}. \quad (2.5)$$

Finally, we assume the following.

Assumption 2.4 The system is empty at the beginning of slot 0, i.e., immediately before the arrivals of customers (if any) in slot 0.

2.2. Workload in system and mean waiting time

We first consider workload in system. Let V_{t-} , $t \in \mathcal{T}$, denote the total unfinished workload immediately before the beginning of slot t . Let $v_t(k) = \mathbb{P}(V_{t-} = k)$ for $t \in \mathcal{T}$ and $k \in \mathbb{Z}_+$. Assumption 2.4 shows that $V_{0-} = 0$, i.e., $v_0(0) = 1$, and by definition, we have

$$V_{t-} = (V_{(t-1)-} + X_{t-1} - 1)^+, \quad t \in \{1, 2, \dots, T\}, \quad (2.6)$$

where $(x)^+ = \max(x, 0)$ for $x \in (-\infty, \infty)$. Thus, we obtain the following result.

Lemma 2.1 We have $v_0(0) = 1$ and $v_0(k) = 0$ for $k \in \mathbb{N}$, and

$$v_t(k) = \begin{cases} v_{t-1}(0)x_{t-1}(0) + v_{t-1}(0)x_{t-1}(1) + v_{t-1}(1)x_{t-1}(0), & k = 0, \\ \sum_{\ell=0}^{k+1} v_{t-1}(\ell)x_{t-1}(k+1-\ell), & k \in \mathbb{N}. \end{cases} \quad (2.7)$$

for $t = 1, 2, \dots, T$. Particularly, we have for $t = 1$,

$$v_1(k) = \begin{cases} x_0(0) + x_0(1), & k = 0, \\ x_0(k+1), & k \in \mathbb{N}, \end{cases} \quad (2.8)$$

Proof. Equation (2.7) is obvious by Assumptions 2.3, 2.4, and the recursion formula (2.6) because the unfinished workload (if any) is processed by 1 at the end of each slot. $\square$

Lemma 2.2

$$\mathbb{E}[V_{1-}] = \lambda p_0 \bar{b} + 1 - e^{-\lambda p_0}, \quad (2.9)$$

$$\mathbb{E}[V_{t-}] = \mathbb{E}[V_{(t-1)-}] + \lambda p_{t-1} \bar{b} - (1 - e^{-\lambda p_{t-1}} v_{t-1}(0)), \quad t = 2, 3, \dots, T. \quad (2.10)$$

Proof. From (2.6), we have for $t \in \mathcal{T}$,

$$\begin{aligned} \mathbb{E}[V_{t-}] &= \mathbb{E}[(V_{(t-1)-} + X_{t-1} - 1)\mathbb{I}(V_{(t-1)-} + X_{t-1} \geq 1)] \\ &= \mathbb{E}[V_{(t-1)-}] + \mathbb{E}[X_{t-1}] - (1 - \mathbb{P}(V_{(t-1)-} + X_{t-1} = 0)). \end{aligned} \quad (2.11)$$

Furthermore, we note that

$$\mathbb{P}(V_{(t-1)-} + X_{t-1} = 0) = \mathbb{P}(V_{(t-1)-} = 0) \mathbb{P}(X_{t-1} = 0) \quad (2.12)$$

because $V_{(t-1)-}$ and X_{t-1} are independent and nonnegative random variables. Equations (2.11) and (2.12) complete the proof. $\square$

Next, we consider the (actual) waiting time of an arbitrary customer arriving in a slot. Let $\mathcal{T}^+ = \{t \in \mathcal{T}; p_t > 0\}$. We then define W_t , $t \in \mathcal{T}^+$, as the waiting time of an arbitrary customer arriving in slot t .

Lemma 2.3

$$\mathbb{E}[W_t] = \mathbb{E}[V_{t-}] + \frac{\bar{b}}{2} \left(\frac{\lambda p_t}{1 - e^{-\lambda p_t}} - 1 \right), \quad t \in \mathcal{T}^+. \quad (2.13)$$

Proof. For $t \in \mathcal{T}^+$, let $q_t(k)$, $k \in \mathbb{Z}_+$, denote the probability that a customer randomly chosen from the ones arriving in slot t enters the server after the $k - 1$ members of them receive service and leave the system. Recalling that the arriving customers during the same slot are randomly ordered, then it follows from (2.1) that

$$\begin{aligned} q_t(k) &= \sum_{\ell=k+1}^{\infty} \frac{1}{\ell} \frac{\mathbb{P}(A_t = \ell)}{\mathbb{P}(A_t \geq 1)} = \sum_{\ell=k+1}^{\infty} \frac{1}{\ell} \frac{\mathbb{P}(A_t = \ell)}{1 - \mathbb{P}(A_t = 0)} \\ &= \sum_{\ell=k+1}^{\infty} \frac{1}{\ell} \frac{e^{-\lambda p_t}}{1 - e^{-\lambda p_t}} \frac{(\lambda p_t)^\ell}{\ell!}, \quad k \in \mathbb{Z}_+. \end{aligned} \quad (2.14)$$

Thus, for $t \in \mathcal{T}^+$, we have

$$\begin{aligned} \mathbb{E}[W_t] &= \mathbb{E}[V_{t-}] + \sum_{k=0}^{\infty} k q_t(k) \times \bar{b} \\ &= \mathbb{E}[V_{t-}] + \sum_{\ell=1}^{\infty} \frac{1}{\ell} \frac{e^{-\lambda p_t}}{1 - e^{-\lambda p_t}} \frac{(\lambda p_t)^\ell}{\ell!} \sum_{k=0}^{\ell-1} k \bar{b} \\ &= \mathbb{E}[V_{t-}] + \frac{\bar{b}}{2} \sum_{\ell=1}^{\infty} (\ell - 1) \frac{e^{-\lambda p_t}}{1 - e^{-\lambda p_t}} \frac{(\lambda p_t)^\ell}{\ell!} \\ &= \mathbb{E}[V_{t-}] + \frac{\bar{b}}{2} \left(\frac{\lambda p_t}{1 - e^{-\lambda p_t}} - 1 \right), \end{aligned}$$

which shows that (2.13) holds. $\square$

3. Arrival-Time Distribution for Equilibrium Mean Waiting Time

We introduce the *equilibrium mean waiting time* and *equilibrium arrival-time distribution*, and present a procedure for calculating them.

Definition 3.1 A positive value w^* is said to be an *equilibrium mean waiting time* if

$$\mathbb{E}[W_t] = w^*, \quad t \in \mathcal{T}^+, \quad (3.1)$$

$$\mathbb{E}[V_{t-}] > w^*, \quad t \in \mathcal{T} \setminus \mathcal{T}^+. \quad (3.2)$$

The arrival-time distribution $\mathcal{P} = \{p_t; t \in \mathcal{T}\}$ is said to be *equilibrium* if it ensures the existence of an equilibrium mean waiting time w^* .

Let $\mathcal{P}^*$ denote the set of equilibrium arrival-time distributions. Let $\mathcal{P}^* := \{p_t^*; t \in \mathcal{T}\}$ denote an arbitrarily element of $\mathcal{P}^*$. In the rest of this paper, we fix $\mathcal{P} = \mathcal{P}^*$, i.e., $p_t = p_t^*$ for all $t \in \mathcal{T}$. It follows from $V_{0-} = 0$, Definition 3.1 and Lemma 2.3 that $0 \in \mathcal{T}^+$, i.e.,

$$p_0^* > 0, \quad (3.3)$$

otherwise, i.e., if $0 \in \mathcal{T} \setminus \mathcal{T}^+$, we have $0 = \mathbb{E}[V_{0-}] > w^*$ by (3.2), which contradicts to $w^* > 0$. We then have

$$w^* = \mathbb{E}[W_0] = \frac{\bar{b}}{2} \left(\frac{\lambda p_0^*}{1 - e^{-\lambda p_0^*}} - 1 \right). \quad (3.4)$$

Remark 3.1 It is easy to see that $x/(1 - e^{-x})$ is increasing in $x \geq 0$ and

$$\lim_{x \rightarrow 0} \frac{x}{1 - e^{-x}} = 1. \quad (3.5)$$

Therefore, (3.4) implies that w^* is positive and increasing with $p_0^* \in (0, 1)$.

It follows from Lemma 2.3, Definition 3.1 and (3.4) that p_t^* ($t = 1, 2, \dots, T$) satisfies following equation.

$$\begin{aligned} \frac{\bar{b}}{2} \left(\frac{\lambda p_t^*}{1 - e^{-\lambda p_t^*}} - 1 \right) &= (w^* - \mathbb{E}[V_{t-}])^+ \\ &= \frac{\bar{b}}{2} \left(\frac{\lambda p_0^*}{1 - e^{-\lambda p_0^*}} - 1 - \frac{2\mathbb{E}[V_{t-}]}{\bar{b}} \right)^+, \quad t = 1, 2, \dots, T, \end{aligned} \quad (3.6)$$

which results in the following theorem.

Theorem 3.1 For $t = 1, 2, \dots, T$, the probability p_t^* is a solution of the following equation:

$$\frac{\lambda p_t^*}{1 - e^{-\lambda p_t^*}} = \left(\frac{\lambda p_0^*}{1 - e^{-\lambda p_0^*}} - 1 - \frac{2\mathbb{E}[V_{t-}]}{\bar{b}} \right)^+ + 1, \quad t = 1, 2, \dots, T, \quad (3.7)$$

$$\mathbb{E}[V_{t-}] = \mathbb{E}[V_{(t-1)-}] + \lambda p_{t-1}^* \bar{b} - (1 - e^{-\lambda p_{t-1}^*} v_{t-1}(0)), \quad t = 1, 2, \dots, T, \quad (3.8)$$

$$\sum_{t \in \mathcal{T}} p_t^* = 1, \quad (3.9)$$

where the last equation is the normalizing condition for the probability distribution $\{p_t^*; t \in \mathcal{T}\}$.

Proof. Equations (3.7) and (3.8) are immediately obtained from (3.6) and (2.10). $\square$

We close this section by summarizing the computational procedure for the equilibrium arrival-time distribution $\mathcal{P}^* = \{p_t^*; t \in \mathcal{T}\}$.

- Step 0: Set $p_0^* = \varepsilon$ for a small $\varepsilon > 0$, and choose $\delta > 0$ as a precision parameter.
- Step 1: Compute w^* by (3.4).
- Step 2: For $t = 1, 2, \dots, T$, compute p_t^* by (3.7), where $\mathbb{E}[V_{t-}]$ and $v_{t-1}(0)$ are recursively computed by (3.8) and (2.7), respectively.
- Step 3: If $|1 - \sum_{t \in \mathcal{T}} p_t^*| < \delta$, return $\{p_t^*; t \in \mathcal{T}\}$ as the equilibrium solution, otherwise $p_0^* := p_0^* + \varepsilon$ and go to Step 1.

Acknowledgments

The research of the first author was supported by JSPS KAKENHI Grant No. 16K21704. The research of the third author was supported by JSPS KAKENHI Grant No. 15K00034.

References

- [1] A. Glazer and R. Hassin: ?/M/1: on the equilibrium distribution of customer arrivals. *European Journal of Operational Research*. **13** (1983), 146–150.
- [2] R. Hassin and Y. Kleiner: Equilibrium and optimal arrival patterns to a server with opening and closing times. *IIE Transactions*. **43** (2011), 164–175.
- [3] M. Haviv: When to arrive at a queue with tardiness costs? *Performance Evaluation*. **70** (2013), 387–399.
- [4] L. Ravner: Equilibrium arrival times to a queue with order penalties. *European Journal of Operational Research*. **239** (2014), 456–468.

A Finite Bottleneck Game with Homogeneous Commuters

Ryo Kawasaki

Department of Industrial Engineering and Economics, School of Engineering,
Tokyo Institute of Technology.

Hideo Konishi

Department of Economics, Boston College.

Junki Yukawa

Department of Social Engineering, Graduate School of Decision Science and Technology,
Tokyo Institute of Technology.

Abstract This paper investigates the finite bottleneck game, in which we assume a finite set of commuters and a finite set of departing time slots. We show that the set of Nash equilibria is equivalent to the set of strong Nash equilibria when we assume homogeneous commuters in their preferences. We also show that pure-strategy Nash equilibria do not exist in general in this setting. Moreover, when we allow commuters to differ in their preferences, we show that Nash equilibria may not exist, and the equivalence result no longer follows.

1. Introduction

A bottleneck model is used in analyzing a rush-hour traffic congestion, where commuters depart from their origins (e.g. their houses) to their destinations (e.g. their workplaces). The simplest model was independently analyzed by Vickrey (1969) and Hendrickson and Kocur (1981), where a continuum of commuters depart from a single origin to a single destination connected by a single road. Along the road, there is a bottleneck in which a queue forms if there are too many commuters in the bottleneck at a given time. In these papers, commuters decide on the departure time based on the trade-offs between congestion and their optimal arrival time.

Moreover, these studies assume homogeneous commuters in that all commuters have the same preferred time of arrival and a specific form of the trip cost function. Their contribution was that departure time decision made by commuters are endogenously determined by means of including trade-off between their travel time and their arrival time.

Other subsequent papers, such as Smith (1983), Daganzo (1985) and Arnott et al. (1990), also consider a continuum of commuters and a continuous time horizon. However, a finite set of commuters and a discrete time horizon seem closer to real-life situations in which the population of a city is finite. The situation corresponds to where there is a relatively small number of commuters, each of which can cause congestion to occur.

Our model, which we call the finite bottleneck game, is endowed with a finite set of commuters and a finite set of time periods, each of which is called a slot. Commuters have preference on two arguments: her departure time and the queue-length which she have to wait through the bottleneck, where in this model the capacity is the maximum number of commuters that can pass through it in each slot. We assume homogeneous commuters as in the previous studies above, but we do not give a specific form of trip costs function. In this sense, our model is an abstract generalization of models of the aforementioned papers.

Mathematically, our model is also an extension of the congestion game (c.f. Rosenthal (1973)). The congestion game considers a situation in which n players choose a combination of primary factors out of t alternatives. Each player's payoff is determined by the sum of the costs of each primary factor she chooses, while the cost of each primary factor depends on the number of players who choose it, and not on the players' names. Rosenthal (1973) proved that there always exists at least one pure-strategy Nash equilibrium by constructing a potential function, which is later formalized by Monderer and Shapley (1996).

Though Rosenthal (1973) and Monderer and Shapley (1996) assume that the cost functions, hence payoff functions, have the same form among the players who take same factors, Milchtaich (1996) allows payoff functions to be different between players in his model, where players choose only one factor from a common set of factors. In Milchtaich (1996), it was shown that a Nash equilibrium always exists in pure strategies. Moreover, Konishi et al. (1997a) shows that in the same model, the set of strong Nash equilibria, which is a stronger concept than Nash equilibria, is nonempty.

Specifically, Konishi et al. (1997a) describes the games in the above class using the following three properties: anonymity [A], partial rivalry [PR] and independence of irrelevant choices [IIC]¹. First, [A] requires that the payoff of each player depends on the number of players who choose each action and not

¹[IIC] condition is also called no spillovers [NS] in Konishi et al. (1997b)

on the players' names. [PR] states that the payoff of each player increases if another player who had chosen the same strategy chooses a different strategy. Finally, [IIC] states that the payoff of a player is not affected even if another player that chooses a different strategy from hers switches to another strategy that is also different strategy from hers.

In relation to congestion games, our model does not satisfy [IIC], whereas the other two conditions hold. Specifically, [IIC] would be violated in the case where a player who had departed later then switched to an earlier departure time and thereby possibly creating a longer queue for some of those players which she leaps over.

In this paper, we restrict our attention to the special case where commuters are homogeneous. Then, we can show that a set of Nash equilibria coincides with that of strong Nash equilibria (Proposition 3.1), while pure-strategy Nash equilibria do not exist in general. It can be stated that this proposition partly explains the difficulty in the existence of a Nash equilibrium. In addition, we illustrate that the equivalence of the set of Nash equilibria and that of strong Nash equilibria does not hold when the homogeneity assumption is dropped.

The rest of the paper is organized as follows: in Section 2, we define the model and notations. In Section 3, we state the main result, and Section 4 compares it and the heterogeneous case.

2. Model

We consider a bottleneck model with finite numbers of commuters and time slots. Let $t = 1, \dots, T$ be the available time slots for departure. Each time intervals can be every minute or every five minutes, for example. Let the set of time slots be $\mathcal{T} = \{1, \dots, T\}$. At each time slot, c cars can go through a bottleneck, where c is a positive integer. If in the end of period $t - 1$ the length of queue q_{t-1} , and if m_t cars arrives at the bottleneck in period t , then the length of the queue in the end of period t is given by $q_t = \max\{0, \tilde{q}_t\}$, where $\tilde{q}_0 \equiv 0$ and $\tilde{q}_t \equiv q_{t-1} + m_t - c$ for $t \geq 1$ is called the effective length of queue in period t . Effective and real queue length vectors are denoted $\tilde{q} = (\tilde{q}_1, \dots, \tilde{q}_T)$ and $q = (q_1, \dots, q_T)$, respectively. Let $i = 1, \dots, n$ be commuters, and let the set of commuters be denoted by $N = \{1, \dots, n\}$. Commuter i 's choice (strategy) of departing time is denoted $\tau_i \in \mathcal{T}$. Given a strategy profile $\tau = (\tau_1, \dots, \tau_n) \in \mathcal{T}^N$, and the resulting departure pattern is given by $m(\tau) = (m_1(\tau), \dots, m_T(\tau)) \in \mathbb{Z}_+^T$ and resulting effective and real queue length vectors are given by $\tilde{q}(\tau)$ and $q(\tau)$, respectively. Each commuter's payoff function is written as $u_i(t, q_t)$, where we assume that $u_i(t, q_t) > u_i(t, q_t + 1)$ for all $t \in \mathcal{T}$ and $q_t \in \mathbb{Z}_+$. That is, each commuter would prefer the situation of departing at time t with less congestion. We assume strict preferences that is generic case.

De nition 2.1 (Strict Preferences).

For all $i = 1, \dots, n$, all $t, t' \in \mathcal{T}$ with $t \neq t'$, and all $q_t, q_{t'} \in \mathbb{Z}_+$, $u_i(t, q_t) \neq u_i(t', q_{t'})$.

De nition 2.2 (Nash Equilibrium).

A strategy profile τ is a Nash equilibrium if for all $i \in N$ and all $t \in \mathcal{T}$, $u_i(\tau, q_{\tau_i}(\tau)) \geq u_i(t, q_t(\tau, \tau_{-i}))$.

Before we give a characterization of Nash equilibria, we introduce new terms that is used for its characterization.

De nition 2.3 (Basin and Terrace).

1. A single slot t is said to be a **basin** at $\tau \in \mathcal{T}^N$ if $\tilde{q}_t(\tau) < 0$ and $\tilde{q}_{t-1}(\tau) \leq 0$.
2. A single slot t is a **single terrace** at $\tau \in \mathcal{T}^N$ if $\tilde{q}_t(\tau) = 0$ and $\tilde{q}_{t-1}(\tau) \leq 0$.
3. Consecutive slots $[t_1, t_2]$ with $1 \leq t_1 < t_2$ is a **connected terrace** at $\tau \in \mathcal{T}^N$ if $\tilde{q}_{t'}(\tau) > 0$ for all $t' \in [t_1, t_2]$ and $\tilde{q}_{t_2}(\tau) \leq 0$.

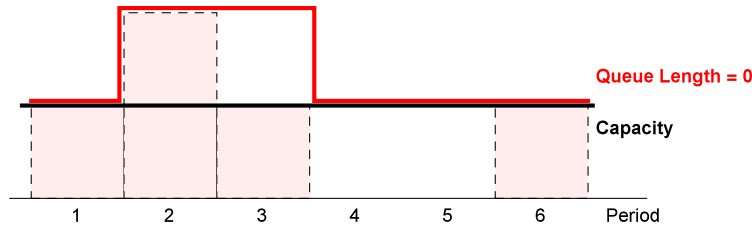


Figure 1: Basin and Terrace

Figure 1 is an example, where there are five commuters, six slots and $c = 1$, and the departure pattern is painted pink. In this example, $[2, 4]$ is a connected terrace while $t = 1$ and $t = 6$ each is a single terrace. In addition, $t = 6$ is a basin.

With these notions, the following is a characterization of Nash equilibria in this game.

Proposition 1.

A strategy profile τ is a Nash equilibrium if and only if for all $i \in N$,

1. $u_i(\tau_i, q_{\tau_i}(\tau)) \geq u_i(t', \max\{\tilde{q}_{t'}(\tau) + 1, 0\})$ for all $t' < \tau_i$,
2. for $t' > \tau_i$,
 - 2a. $u_i(\tau_i, q_{\tau_i}(\tau)) \geq u_i(t', \max\{\tilde{q}_{t'}(\tau), 0\})$ for all $t' \in [t_1, t_2]$ with $t' > \tau_i$, where $[t_1, t_2]$ is a connected terrace at τ and $\tau_i \in [t_1, t_2]$.
 - 2b. Otherwise, $u_i(\tau_i, q_{\tau_i}(\tau)) \geq u_i(t', \max\{\tilde{q}_{t'}(\tau) + 1, 0\})$.

Proof. First, suppose that profile τ is a Nash equilibrium. By definition, for all $i \in N$ and all $t' \in \mathcal{T}$,

$$u_i(\tau, q_{\tau_i}(\tau)) \geq u_i(t', q_{t'}(t', \tau_{-i})).$$

We show that for the both cases of $t' < \tau_i$ and $t' > \tau_i$, the queue-length at slot t' , $q_{t'}(t', \tau_{-i})$ satisfies the above.

Note that profile (t', τ_{-i}) satisfies the following:

$$m_{\tau_i}(t', \tau_{-i}) = m_{\tau_i}(\tau) - 1, \quad (1)$$

$$m_{t'}(t', \tau_{-i}) = m_{t'}(\tau) + 1, \quad (2)$$

$$m_t(t', \tau_{-i}) = m_t(\tau) \quad \forall t \neq \tau_i, t'. \quad (3)$$

1. When $t' < \tau_i$:

Since $q_{t'-1}(\tau) = q_{t'-1}(t', \tau_{-i})$ by (3), we obtain

$$\begin{aligned} \tilde{q}_{t'}(t', \tau_{-i}) &= q_{t'-1}(t', \tau_{-i}) + m_{t'}(t', \tau_{-i}) - c \\ &= q_{t'-1}(\tau) + (m_{t'}(\tau) + 1) - c \\ &= (q_{t'-1}(\tau) + m_{t'}(\tau) - c) + 1 \\ &= \tilde{q}_{t'}(\tau) + 1. \end{aligned}$$

Thus, $q_{t'}(t', \tau_{-i}) = \max\{\tilde{q}_{t'}(\tau) + 1, 0\}$.

2. When $t' > \tau_i$:

- 2a. When $t' \in [t_1, t_2]$, the connected terrace to which t' belongs:

Note that since $\tilde{q}_{t''}(\tau) > 0$, so $\tilde{q}_{t''}(\tau) \geq 1$ for all $t'' \in [\tau_i + 1, t' - 1]$, $\tilde{q}_{t''}(t', \tau_{-i}) = \tilde{q}_{t''}(\tau) - 1 \geq 0$. It follows that

$$\begin{aligned} \tilde{q}_{t'}(t', \tau_{-i}) &= q_{t'-1}(t', \tau_{-i}) + m_{t'}(t', \tau_{-i}) - c \\ &= (q_{t'-1}(\tau) - 1) + (m_{t'}(\tau) + 1) - c \\ &= q_{t'-1}(\tau) + m_{t'}(\tau) - c \\ &= \tilde{q}_{t'}(\tau). \end{aligned}$$

Thus, $q_{t'}(t', \tau_{-i}) = \max\{\tilde{q}_{t'}(\tau), 0\}$.

- 2b. Otherwise:

We first show $q_{t''}(\tau) = q_{t''}(t', \tau_{-i})$ for some $t'' \in [\tau_i, t')$. Suppose not. Then, $q_t(\tau) > q_t(t', \tau_{-i}) \geq 0$ for all $t \in [\tau_i, t')$, implying that t' belongs to the same connected terrace as τ_i . This is a contradiction, since we have already considered in the case 2a..

When $t'' = t' - 1$, we immediately obtain $q_{t'-1}(\tau) = q_{t'-1}(t', \tau_{-i})$.

When $t'' \neq t' - 1$, using (3), we obtain

$$\begin{aligned} \tilde{q}_{t''+1}(t', \tau_{-i}) &= q_{t''}(t', \tau_{-i}) + m_{t''}(t', \tau_{-i}) - c \\ &= q_{t''}(\tau) + m_{t''}(\tau) - c \\ &= \tilde{q}_{t''+1}(\tau). \end{aligned}$$

Thus, $q_{t''+1}(t', \tau_{-i}) = q_{t''+1}(\tau)$. Similarly,

$$\begin{aligned} q_{t''+2}(\tau) &= q_{t''+2}(t', \tau_{-i}), \\ &\vdots \\ q_{t'-1}(\tau) &= q_{t'-1}(t', \tau_{-i}). \end{aligned}$$

With this assertion, we obtain

$$\begin{aligned} \tilde{q}_{t'}(t', \tau_{-i}) &= q_{t'-1}(t', \tau_{-i}) + m_{t'}(t', \tau_{-i}) - c \\ &= q_{t'-1}(\tau) + (m_{t'}(\tau) + 1) - c \\ &= (q_{t'-1}(\tau) + m_{t'}(\tau) - c) + 1 \\ &= \tilde{q}_{t'}(\tau) + 1. \end{aligned}$$

Thus, $q_{t'}(t', \tau_{-i}) = \max\{\tilde{q}_{t'}(\tau) + 1, 0\}$.

Suppose next that profile τ satisfies the above but it is not a Nash equilibrium. Then, there exists some $i \in N$ who can improve by switching to another strategy t' . However, this implies that either case 1., case 2a., or case 2b. does not hold. This is a contradiction. ■

We also introduce a stronger concept, strong Nash equilibria. A strong Nash equilibrium is a strategy profile, which is immune to any coalitional deviation — that is, for any coalitional deviation, there is at least one player whose payoff does not improve by the deviation. Formally,

Definition 2.4 (Coalitional Deviation).

A **coalitional deviation** (S, τ'_S) from profile τ is a pair of a nonempty subset of players $S \subseteq N$ and their strategy profile in S , $\tau'_S = (\tau'_i)_{i \in S} \in \mathcal{T}^S$ such that $u_i(\tau'_S, \tau_{-S}) > u_i(\tau)$ for all $i \in S$.

Definition 2.5 (Strong Nash Equilibrium).

A profile τ is a **strong Nash equilibrium** if there is no coalitional deviation from τ .

Notice that a set of a strong Nash equilibrium is always a subset of a Nash equilibrium by definition.

3. The Analysis

In the following analysis, we assume homogeneous commuters. That is, we assume $u \equiv u_i$ for all $i \in N$. First, we show that the homogeneity in preferences is not enough to establish the general existence of Nash equilibria.

Example 1.

Consider the following three commuter problem: $N = \{1, 2, 3\}$, $\mathcal{T} = \{1, 2, 3\}$, $c = 1$, and players have the following preferences:

$$u(2, 0) > u(1, 0) > u(1, 1) > u(3, 0) > u(2, 1) > u(1, 2).$$

Then, there is no Nash equilibrium in this game.

Proof. There can be the following 9 strategy profiles.

1. $\tau^1 = (1, 1, 1)$ and $q(\tau^1) = (2, 1, 0)$. Then, player 1 moves to $t = 3$.
2. $\tau^2 = (1, 1, 2)$ and $q(\tau^2) = (1, 1, 0)$. Then, player 3 moves to $t = 3$.
3. $\tau^3 = (1, 1, 3)$ and $q(\tau^3) = (1, 0, 1)$. Then, player 2 moves to $t = 2$.
4. $\tau^4 = (1, 2, 2)$ and $q(\tau^4) = (0, 1, 0)$. Then, player 3 moves to $t = 3$.
5. $\tau^5 = (1, 2, 3)$ and $q(\tau^5) = (0, 0, 0)$. Then, player 3 moves to $t = 3$.
6. $\tau^6 = (2, 2, 2)$ and $q(\tau^6) = (0, 2, 1)$. Then, player 1 moves to $t = 1$.
7. $\tau^7 = (2, 2, 3)$ and $q(\tau^7) = (0, 1, 1)$. Then, player 1 moves to $t = 1$.
8. $\tau^8 = (2, 3, 3)$ and $q(\tau^8) = (0, 0, 1)$. Then, player 1 moves to $t = 1$.
9. $\tau^9 = (3, 3, 3)$ and $q(\tau^9) = (0, 0, 2)$. Then, player 1 moves to $t = 1$.

Hence, there is no Nash equilibrium. ■

Therefore, we need to restrict the domain of the preferences to establish the general existence of Nash equilibria.

However, we show a theorem which asserts that the set of Nash equilibria and the set of strong Nash equilibria coincide. Thus, it can be said that when a Nash equilibrium exists, it exhibits strong stability in the context of a coalitional deviation.

Theorem 3.1.

Assume that preferences are homogeneous. Then, the set of Nash equilibria coincides with the set of strong Nash equilibria.

Proof. The proof is omitted due to space constraint. ■

4. Heterogeneous Commuter Case

In this section, we analyze a more generalized case, commuters with heterogeneous preferences. We observe that there do not exist Nash equilibria in general, and also show that Proposition 3.1 is no longer true when assuming heterogeneous commuters.

Example 2.

Consider the following four commuter problem: $N = \{1, 2, 3, 4\}$, $\mathcal{T} = \{1, 2, 3, 4\}$, $c = 1$, and players have the following preferences:

$$\begin{aligned} u_1(1, 2) &> u_1(2, 1) > u_1(3, 1) > u_1(2, 2) > u_1(1, 3), \\ u_2(2, 1) &> u_2(1, 2) > u_2(1, 3) > u_2(2, 2), \end{aligned}$$

and for player $i = 3, 4$, $u_i(1, 3) > u_i(t, 0)$ for all $t = 2, 3, 4$.

Proof. In this example, players 3 and 4 always choose $t = 1$. Player 1 may choose $t = 1, 2, 3$, and player 2 may choose $t = 1, 2$. There can be the following 6 strategy profiles.

1. $\tau^1 = (1, 1, 1, 1)$ and $q(\tau^1) = (3, 2, 1, 0)$. In this case, player 1 moves to $t = 3$.
2. $\tau^2 = (2, 1, 1, 1)$ and $q(\tau^2) = (2, 2, 1, 0)$. In this case, player 1 moves to $t = 3$.
3. $\tau^3 = (3, 1, 1, 1)$ and $q(\tau^3) = (2, 1, 1, 0)$. In this case, player 2 moves to $t = 2$.
4. $\tau^4 = (1, 2, 1, 1)$ and $q(\tau^4) = (2, 2, 1, 0)$. In this case, player 2 moves to $t = 1$.
5. $\tau^5 = (2, 2, 1, 1)$ and $q(\tau^5) = (1, 2, 1, 0)$. In this case, player 2 moves to $t = 1$.
6. $\tau^6 = (3, 2, 1, 1)$ and $q(\tau^6) = (1, 1, 1, 0)$. In this case, player 1 moves to $t = 1$.

Hence, there is no Nash equilibrium in pure strategies. ■

The next example shows that when commuters' preferences differ from each other, the set of Nash equilibria may not coincide with the set of strong Nash equilibria.

Example 3.

Consider the following five commuter problem: $N = \{1, 2, 3, 4, 5\}$, $\mathcal{T} = \{1, 2, 3, 4, 5\}$, $c = 1$, and players have the following preferences:

$$\begin{aligned} u_1(1, 1) &> u_1(4, 1) > u_1(1, 2), \\ u_2(2, 0) &> u_2(1, 1) > u_2(1, 2) > u_2(2, 1), \\ u_3(3, 0) &> u_3(1, 2) > u_3(4, 1) > u_3(3, 1), \\ u_4(1, 4) &> u_2(t, 0) \quad t = 2, 3, 4, 5, \\ u_5(4, 4) &> u_5(t, 0) \quad t = 1, 2, 3, 5. \end{aligned}$$

Proof. There is only one Nash equilibrium $\tau^* = (1, 1, 4, 1, 4)$ in this game. However, either $S = \{1, 3\}$ or $S = \{1, 2, 3\}$ can deviate from τ^* with $\tau'_{\{1,3\}} = (\tau'_1, \tau'_3) = (4, 1)$, or $\tau'_{\{1,2,3\}} = (\tau'_1, \tau'_2, \tau'_3) = (4, 2, 3)$, respectively. ■

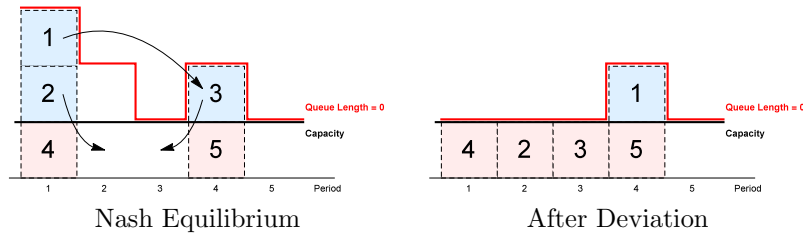


Figure 2: Deviation by $S = \{1, 2, 3\}$

5. Concluding Remarks

We have investigated the finite bottleneck game with homogeneous commuters. The main result of this paper is that the set of Nash equilibria coincides with that of strong Nash equilibria with this setting. In other words, the set of Nash equilibria “shrinks” to that of strong Nash equilibria, which may explain why the general existence of a Nash equilibrium is difficult to establish. We also have shown that this result cannot be applied to the heterogeneous commuter case.

In this regard, our model is quite contrasting to Milchtaich (1996) and Konishi et al. (1997a), where at least one Nash equilibrium always exists, and moreover the set of strong Nash equilibria is nonempty

under fair conditions that characterize the “congestion (of each strategy)”—anonymity, independence from irrelevant choice and partial rivalry—even though preferences of players are heterogeneous.

As our examples show, we do not give the sufficient condition for the general existence of a Nash equilibrium. Hence, the remaining task is to establish such condition.

References

- Arnott, R., A. De Palma, and R. Lindsey (1990). Economics of a bottleneck. *Journal of Urban Economics* 27(1), 111–130.
- Daganzo, C. F. (1985). The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation Science* 19(1), 29–37.
- Hendrickson, C. and G. Kocur (1981). Schedule delay and departure time decisions in a deterministic model. *Transportation Science* 15(1), 62–77.
- Konishi, H., M. Le Breton, and S. Weber (1997a). Equilibria in a model with partial rivalry. *Journal of Economic Theory* 72(1), 225–237.
- Konishi, H., M. Le Breton, and S. Weber (1997b). Equivalence of strong and coalition-proof nash equilibria in games without spillovers. *Economic Theory* 9(1), 97–113.
- Milchtaich, I. (1996). Congestion games with player-specific payoff functions. *Games and Economic Behavior* 13(1), 111–124.
- Monderer, D. and L. S. Shapley (1996). Potential games. *Games and Economic Behavior* 14(1), 124 – 143.
- Rosenthal, R. W. (1973). A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory* 2(1), 65–67.
- Smith, M. J. (1983). The existence and calculation of traffic equilibria. *Transportation Research Part B: Methodological* 17(4), 291–303.
- Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review* 59(2), 251–260.

上界と平均の上界が与えられた確率過程の特徴づけ

高田 寛之, 長崎大学工学研究科, htakada@cis.nagasaki-u.ac.jp

概要：本論文では上界及び下界、平均の上界が与えられた離散有限時間過程について考え、二項最大化関数の期待値の上界を二変量ベルヌーイ型確率変数の最大化関数の期待値で構成する。このベルヌーイ型確率変数もまた同じ下界、上界、平均の上界を持つ離散有限時間過程である。

1. はじめに

コンピュータ技術とネットワーク技術の飛躍的な発展と爆発的な普及によって、TCPIP ネットワークは社会基盤ネットワークとして変貌をとげつつある。しかしながら、現行インターネットには、通信の安定性の問題があり、人の命に関わる遠隔手術のようなサービスは、まだ展開することができない。

この安定性の問題を解決する一つの方法として、通品品質保証技術がある。この技術は、ユーザから通行計画書を受取り、経路上の混み具合を予想して、受け入れの可否を判断する。ところが、通行計画書が伝えるトラフィックの情報はパケット損失確率などを正確に計算するのに必要な情報をすべて含んでいるとは限らず、確率の計算を容易にする仮定はほとんど利用できない。

このような制約において、どのような方法が採用されてきたのかをみるために、ネットワーク算法の研究をおおまかに紹介する。初期のネットワーク算法 [1, 2] は、決定的ネットワーク算法と呼ばれる。創始者である Cruz は [3] において、次のようなトラフィックの特徴づけを行った。

$$\mathcal{A}_1(r, b) = \{A; A(t) - A(s) \leq r(t - s) + b \ (0 \leq s \leq t), A \text{ は非減少} \}.$$

$r > 0$ と $b > 0$ はこのトラフィックを特徴づけるパラメータである。これに加え、min-plus 代数における制御理論 [4] を導入することで、体系的にネットワークの混雑指標（待ち行列長や遅延）の最大値を求めることができるようになった。単純な例は、入力累積過程が A で、固定処理率 $c > 0$ の仕事保存待ち行列の待ち行列長 $Q(t)$ である。

$$Q(t) = \max_{s=0}^t A(t) - A(s) - c(t - s)$$

項 $X(s) = A(t) - A(s) - c(t - s)$ はネットフローと呼ばれ、 $A \in \mathcal{A}_1(r, b)$ であるとき、ネットフローもまた有界である。 $-c(t - s) \leq X(s) \leq r(t - s) + b - c(t - s)$ のように、有界で離散有限時間ネットフロー過程を第一種有界過程と呼び、その過程の集まり（クラス）を以下に定める。

$$\mathcal{T}_1(\mathbf{x}_0, \mathbf{x}_1) = \{(X_1, \dots, X_n); \forall i = 1, \dots, n, x_{0,i} \leq X_i \leq x_{1,i}\}, \quad \mathbf{x}_k = (x_{k,1}, \dots, x_{k,n}), \quad (k = 0, 1).$$

決定的ネットワーク算法は計算理論としては非常に有用であるが、過剰評価の問題があった。これはネットワーク規模に関して指数的に上界が爆発してしまう問題であり、無視できない問題である。具体的な例は [5] で見ることができる。原因は、最悪状況の積み重ねによって統計的多重効果を考慮しておらず、稀な事象も含めて評価していることにある。このようにして、統計的多重効果を考慮できる確率的ネットワーク算法の必要性が高まった [1, 6–11]。上記文献は代表的文献またはサーベイ論文のみであり、ここで引用していない論文もたくさんある。必要であれば、これらの文献の参考文献を更に辿ってほしい。

確率ネットワーク算法は、混雑指標の最大値を求める代わりに、指標の補分布の上界を求める。入力過程は確率過程として扱われ、それに伴い過程に関する確率的な条件を付け加える必要がある。どのような条件を付け加えるべきかには議論が必要であるが、本論文では、既存の実機に組み込まれた仕組みを維持するために、期待値についての条件を付け加えた $\mathcal{A}_1$ のサブクラス $\mathcal{A}_2$ を考慮する。

$$\mathcal{A}_2(\alpha, \beta) = \{A; A(t) - A(s) \leq \alpha(t - s), E[A(t) - A(s)] \leq \beta(t - s) \ (0 \leq s \leq t), A \text{ は非減少} \}.$$

ここで α, β は入力過程を特徴づけるパラメータ関数である。クラス $\mathcal{A}_2$ に属する例は、期待値が陽にわかる場合 [12–15] や期待値の上界を仮定する [16–18] がある。ここで述べた [18] 以外の文献は、クラス $\mathcal{A}_2$ の周辺積率母関

数の上界としてベルヌーイ分布の積率母関数を用いている。[18] はこれを多変量の場合に拡張して、多変量ベルヌーイ分布の積率母関数を用いている。クラス $\mathcal{A}_2$ に属するトラフィックに対してネットフロー過程のクラスは次のようにかける。

$$\begin{aligned}\mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y}) &= \{(X_1, \dots, X_n); \forall i = 1, \dots, n, x_{0,i} \leq X_i \leq x_{1,i}, E[X_i] \leq y_i\}, \\ \mathbf{x}_k &= (x_{k,1}, \dots, x_{k,n}), \quad (k = 0, 1), \quad \mathbf{y} = (y_1, \dots, y_n).\end{aligned}$$

パラメータ $\mathbf{x}_0$ は下界ベクトル、 $\mathbf{x}_1$ は上界ベクトル、 $\mathbf{y}$ は平均上界と呼び、 $\mathcal{T}_2$ に属する離散有限時間確率過程を第二種有界過程と呼ぶ。

なぜ平均値の上界を付与するだけで実機の仕組みが維持できるかという疑問を持つ人がいるかもしれない。通信品質保証技術の仕組みの中で、ユーザにトラフィックの平均レート $r > 0$ とバースト性 $b > 0$ を申告させる、すなわち送信しようとするトラフィックの調査をユーザに要請するので、そこから $\alpha(t) = rt + b, \beta(t) = rt$ とわかるからである。

最近の関連研究動向は、従来の確率ネットワーク算法における精度の問題に取り組むものが見られる。例えば、[19–21] は、最大順序統計量の精度の良い評価のために劣マルチンゲール性を仮定して、より精度の高い上界評価法を提案している。しかし、先に述べたように、追加的な条件として劣マルチンゲール性を選ぶということは、そのトラフィックが劣マルチンゲール性を持つことを何らかの方法で保証しなければならず、別の困難さを生むように思える。ブラウン運動を仮定したりマルコフ性を仮定する場合においても同様の問題が発生する。

Ciucu 等 [22] によれば、最大順序統計量についての評価は、定常エルゴード過程のクラスに対して、一般的に成立する公式の使用だけでは精度を向上することはできず、つまり統計的多重効果は得られないが、考慮するトラフィッククラスの特徴を注意深く選ぶことで、統計的多重効果を得ることができるかもしれないと述べている。例外的にフロー数漸近解析では [16–18] が示すように最大順序統計量に関する漸近確率の評価は正確かつ単純な評価法があるので、このような問題は起きない。しかし、漸近解析では、漸近対数補分布の上界しか得られないため、フロー数に関する多項式項の情報が得られない。仮にフロー数が有限で十分大きい場合に近似値として用いるにしても誤差が評価されているわけでもないの、判定にそのまま用いるのはかなり乱暴である。

以上のことを動機にして、本論文はパラメータ $\mathbf{x}_0, \mathbf{x}_1, \mathbf{y}$ を持つ第二種有界過程 $\mathbf{X} = (X_1, \dots, X_n)$ に対して、その最大順序統計量の補分布の上界を直接求める問題について考える。本論文では簡単のため $n = 1, 2$ の場合について限定する。

本論文の構成は以下の通りである。第 2 節では、予備的な知識としてチェルノフ不等式と一般化チェルノフ不等式、確率順序 [23, 24] における概念、多変数関数としての $\max$ の性質及び多変量ベルヌーイ型分布の定義を与える。第 3 節では、第二種の有界確率過程 ($n = 1, 2$) の最大順序統計量の期待値の上界を構成し、特に一般的な公式を用いる場合の上界との比較を行う。第 4 節で本研究のまとめと課題について述べる。

2. 準備

本節では、予備的な知識をまとめる。ベクトルの次数を n に固定し、多変数関数を考える時は n 変数関数を考える。添字の集合を $N = \{1, \dots, n\}$ と定める。 $\mathbf{e}_i$ を次数 n の第 i 単位ベクトル、 $\mathbf{e}$ は全ての項が 1 のベクトルとする。ベクトル間の不等式は項別不等式を表す。上付き添字 T はベクトルの転置を表す。

2.1. チェルノフ不等式

ここではチェルノフ不等式と一般化チェルノフ不等式について述べる。

チェルノフ不等式は非常に有名な不等式である。この不等式を使うことで確率変数 X の補分布の上界を X の積率母関数 $E[e^{\theta X}]$ を使って与えることができる。

Lemma 1 (Chernoff inequality).

$$\log \Pr[X > x] \leq \inf_{\theta > 0} -\theta x + \log E[e^{\theta X}], \quad (x \in \mathbb{R}).$$

一般化チェルノフ不等式はチェルノフ不等式の多変量一般化であり、凸最適化の結果である。

Lemma 2 (Generalized Chernoff inequality (Section 7.4.2 (pp.379-380) [25])). $\mathbb{R}^n$ を n 次元ユークリッド空間とする。 $C \subset \mathbb{R}^n$ と n 次元確率ベクトル $\mathbf{X} = (X_1, \dots, X_n)^T$ に対して、以下の不等式が成立する。

$$\log \Pr[\mathbf{X} \in C] \leq \inf_{\boldsymbol{\theta} \in \mathbb{R}^n} \sup_{\mathbf{z} \in C} (-\boldsymbol{\theta}^T \mathbf{z}) + \log E[e^{\boldsymbol{\theta}^T \mathbf{X}}].$$

ここで $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$ である。

[18] では一般化チェルノフ不等式を用いて、最小順序統計量についての補分布を次のよう評価した。

Lemma 3.

$$\begin{aligned} \log \Pr[\min(X_1, \dots, X_n) > x] &= \log \Pr[X_1 > x, \dots, X_n > x] \\ &\leq \inf_{\boldsymbol{\theta} \geq 0} - \sum_{i=1}^n \theta_i x + \log \mathbb{E}[e^{\sum_{i=1}^n \theta_i X_i}] \\ &= \inf_{\boldsymbol{\theta} > 0} -\theta x + \log \inf_{\mathbf{p} \geq 0; \mathbf{p}^t \mathbf{e} = 1} \mathbb{E}[e^{\sum_{i=1}^n \theta p_i X_i}]. \end{aligned}$$

ここで最後の等号は $\theta = \sum_{i=1}^n \theta_i, p_i = \theta_i / \theta$ とおいた。

この結果は不等式 $\min(x_1, \dots, x_n) \leq \sum_{i=1}^n p_i x_i$ とチェルノフ不等式の適用によっても求められる。ここで $(p_1, \dots, p_n)$ は $\sum_{i=1}^n p_i = 1$ を満たす任意の非負ベクトルである。

このようにして補分布は積率母関数や結合積率母関数の上界を求めることで上界を構成できる。

2.2. 確率順序

本節では確率順序に関する概念を述べる。

ある関数のクラス $\mathcal{F}$ について、確率ベクトル $\mathbf{X}$ が $\mathcal{F}$ -確率順序の意味で確率ベクトル $\mathbf{Y}$ に抑えられるとは次の不等式を満たすことである。

$$\forall F \in \mathcal{F}, \quad \mathbb{E}[F(\mathbf{X})] \leq \mathbb{E}[F(\mathbf{Y})].$$

代表的な $\mathcal{F}$ のクラスには、特別な名称が与えられている。例えば単調増加関数クラスに関しては通常確率順序 (st) 凸関数クラスに関しては凸確率順序 (cx)、単調増加凸関数クラスに関しては icx 確率順序優モジュラ関数クラス (後述) に関しては sm 確率順序といった名称がある。

多変数関数 F が単調増加であるとは、ベクトル間不等式を保存することである。

$$\mathbf{x} \leq \mathbf{y} \implies F(\mathbf{x}) \leq F(\mathbf{y}).$$

明らかに F が単調増加であることと全ての $\mathbf{x}, i, \delta$ についての差分 $\Delta_i^\delta F(\mathbf{x})$ が非負であることは同値である。ここで差分は以下のように定義される。

$$\Delta_i^\delta F(\mathbf{x}) = F(\mathbf{x} + \delta \mathbf{e}_i) - F(\mathbf{x}).$$

多変数関数 F が凸であるとは、任意のベクトル $\mathbf{x}, \mathbf{y}$ と内分比 $0 \leq \lambda \leq 1$ に対して、

$$F(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda F(\mathbf{x}) + (1 - \lambda) F(\mathbf{y})$$

が成り立つことである。

二階差分が常に 0 のとき F はモジュラ関数であるという。二階差分が常に非負値 (非正值) をとるとき F は優 (劣) モジュラ関数であるという。二階差分 $\Delta_i^{\delta_i} \Delta_j^{\delta_j} F(\mathbf{x})$ は差分の差分として定義される。

$$\Delta_i^{\delta_i} \Delta_j^{\delta_j} F(\mathbf{x}) = \Delta_j^{\delta_j} F(\mathbf{x} + \delta_i \mathbf{e}_i) - \Delta_j^{\delta_j} F(\mathbf{x}) = F(\mathbf{x} + \delta_i \mathbf{e}_i + \delta_j \mathbf{e}_j) - F(\mathbf{x} + \delta_i \mathbf{e}_i) - F(\mathbf{x} + \delta_j \mathbf{e}_j) + F(\mathbf{x}).$$

差分の順序は交換可能であることに注意する。同様にして m 階差分も逐次的に定義される。

本論文における独自の記号を導入する。 δ_i, δ_j が明確である場合には省略することがある。例えば、 $n = 2$ の場合に F の引数として項 i が $x_{0,i}, x_{1,i}$ の二値しかとらないベクトルだけを考えるとときには、考慮する引数は $(x_{0,1}, x_{0,2}), (x_{0,1}, x_{1,2}), (x_{1,1}, x_{0,2}), (x_{1,1}, x_{1,2})$ の 4 通りしかない。 $\delta_i = x_{1,i} - x_{0,i}$ とおけば、 $\Delta_1 F(\mathbf{x}_0) = F(x_{1,1}, x_{0,2}) - F(x_{0,1}, x_{0,2})$ のように、 δ_i を明示しない。この省略記法は、 m 階差分のように差分をとる項が増えた時に記述が複雑になることを防ぐ。

差分をとる項の添字の集合を使って

$$\Delta_{\{i\}} = \Delta_i^{\delta_i}, \quad \Delta_{\{i,j\}} = \Delta_i^{\delta_i} \Delta_j^{\delta_j}$$

と定める。ここで添字の集合を用いた時、 $j = i$ の場合は除外する。 $N = \{1, \dots, n\}$ において、添字の部分集合 $K \subset N$ についての F の差分 $\Delta_K F(\mathbf{x})$ は次のように定義される。

$$\Delta_K F(\mathbf{x}) = \sum_{J \subset K} (-1)^{|J|+1} F\left(\mathbf{x} + \sum_{i \in J} \delta_i \mathbf{e}_i\right).$$

ここで $\sum_{J \subset K}$ は K の部分集合 J を動かして和をとるという意味である。 $|J|$ は J の要素の数を表す。 $|\emptyset| = 0$ と定める。

$K \subset N$ に対してベクトル $\mathbf{x}(K)$ を次のようにさだめる。 $i \in N$ に対して $\mathbf{x}(K)$ の第 i 項を $\mathbf{x}(K)_i$ と表す。

$$\mathbf{x}(K)_i = \begin{cases} x_{0,i} & (i \in K^c) \\ x_{1,i} & (i \in K) \end{cases}.$$

この記号を使うと $F(\mathbf{x}(K))$ は添字集合 K を引数にとる集合関数 $G(K) = F(\mathbf{x}(K))$ として表せる。互いに背反な集合 $K, L \subset N$ に対して $\Delta_K G(L)$ のように集合関数に対しての差分も考えることができる。

2.3. 多変数関数 $\max$ の性質

ここでは多変数関数 $\max(\mathbf{x}) = \max_{i=1}^n x_i$ ($\mathbf{x} = (x_1, \dots, x_n)^T$) の性質について述べる。

Lemma 4. $\max$ は明らかに単調増加で凸である。 $m \geq 1$ に対して、 $G(L) = \max(\mathbf{x}(L))$ とおくと、 G は $2m-1$ 階差分が非負であり、 $2m$ 階差分は非正である。即ち次の性質が成り立つ。

$$(-1)^{|K|+1} \Delta_K G(L) \geq 0. \quad (1)$$

証明は省略する。

次の等式は minimum-maximum identity として知られる。

Lemma 5 (minimum-maximum identity [26]). $a_1, \dots, a_n$ は任意の実数とする。このとき

$$\begin{aligned} \min_{i=1}^n a_i &= \sum_{K \subset N; K \neq \emptyset} (-1)^{|K|+1} \max_{i \in K} a_i, \\ \max_{i=1}^n a_i &= \sum_{K \subset N; K \neq \emptyset} (-1)^{|K|+1} \min_{i \in K} a_i. \end{aligned}$$

特に $n = 2$ のときは $\min(a_1, a_2) + \max(a_1, a_2) = a_1 + a_2$ である。

次の不等式は $\max$ に関連してよく使われる。

Lemma 6 (Bool inequality). $\Pr(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \Pr(A_i)$.

例えば $\Pr(\max_{i=1}^n X_i > x) = \Pr(\bigcup_{i=1}^n \{X_i > x\}) \leq \sum_{i=1}^n \Pr(X_i > x)$. である。

次の不等式もチェルノフ不等式と共に用いられる。

Lemma 7. $\mathbf{X} \geq 0$ のとき $E[\max \mathbf{X}] \leq \sum_{i=1}^n E[X_i]$.

最後の二つの不等式は、一般的に成立するが、その上界の精度が非常に悪い。

2.4. ベルヌーイ型分布

確率ベクトル $\mathbf{Y} = (Y_1, \dots, Y_n)$ が多変量ベルヌーイ型分布であるとは第 i 項 Y_i が二値しかとらないことをいう。例えば、

$$(Y_1, Y_2) = \begin{cases} (-1, -2) & \text{w. p. } 0.3 \\ (1, -2) & \text{w. p. } 0.2 \\ (-1, 2) & \text{w. p. } 0.4 \\ (1, 2) & \text{w. p. } 0.1 \end{cases}$$

は Y_1 は $1, -1$ のどちらかの値しかとらないし、 Y_2 は $2, -2$ のどちらかの値しかとらないので二変量ベルヌーイ型分布である。この周辺分布は一変量ベルヌーイ分布になる。

$$Y_1 = \begin{Bmatrix} -1 & 0.7 \\ 1 & 0.3 \end{Bmatrix}, \quad Y_2 = \begin{Bmatrix} -2 & 0.5 \\ 2 & 0.5 \end{Bmatrix}.$$

多変量ベルヌーイ型分布の最大順序統計量の期待値は容易に計算できる。上の例の場合、

$$\begin{aligned} E[\max(Y_1, Y_2)] &= \max(-1, -2) * 0.3 + \max(1, -2) * 0.2 + \max(-1, 2) * 0.4 + \max(1, 2) * 0.1 \\ &= -1 * 0.3 + 1 * 0.2 + 2 * 0.4 + 2 * 0.1 = 0.9. \end{aligned}$$

一方、Lemma 7 は Y_1, Y_2 が非負確率変数ではないのでそもそも適用できない。

3. 第二種有界過程の $\max$ の期待値の上界 ($n \leq 2$)

本節では第二種有界過程の $\max$ の期待値の上界を与える。簡単のため $n = 1, 2$ の場合について考える。

3.1. $n = 1$ の場合

$n = 1$ の場合は $\max(X_1) = X_1$ であるから X_1 の期待値の上界を求める問題である。 $x_{0,1} \leq x_{1,1}$ なる実数 $x_{0,1}, x_{1,1}$ に対して、確率変数 X_1 が次の仮定を満たすとする $x_{0,1} \leq X_1 \leq x_{1,1}$ および $E[X_1] \leq y_1$. $P_1 = (X_1 - x_{0,1}) / (x_{1,1} - x_{0,1})$, $p_1 = (y_1 - x_{0,1}) / (x_{1,1} - x_{0,1})$ とおく。明らかに $0 \leq P_1 \leq 1$ かつ $E[P_1] \leq p_1$ が成り立つ。 $X_1 = (1 - P_1) * x_{0,1} + P_1 * x_{1,1} = x_{0,1} + P_1 * (x_{1,1} - x_{0,1})$ とかけるので

$$E[X_1] = x_{0,1} + E[P_1] * (x_{1,1} - x_{0,1}) \leq x_{0,1} + p_1 * (x_{1,1} - x_{0,1}) = (1 - p_1) * x_{0,1} + p_1 * x_{1,1} = E[Y_1],$$

である。ここに Y_1 は確率 p_1 で $x_{1,1}$ の値をとり、確率 $1 - p_1$ で $x_{0,1}$ の値をとるベルヌーイ型確率変数である。まとめると

Lemma 8. $X_1 \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})$ に対して

$$E[\max X_1] \leq E[\max Y_1]$$

を満たすベルヌーイ型確率変数 $Y_1 \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})$ が見つかる。

本質的には $\max$ を考えたことにはならないが、 X_1 をその値の下界と上界を端点として、内分比 $P_1 : 1 - P_1$ の凸結合として表すアイデアは次数を拡張したときにも使える。

3.2. $n = 2$ の場合

第二種有界確率過程 $\mathbf{X} = (X_1, X_2)$ を考える。 $i = 1, 2$ に対して、 $x_{0,i} \leq y_i \leq x_{1,i}$ において、

$$x_{0,i} \leq X_i \leq x_{1,i}, \quad (2)$$

$$E[X_i] \leq y_i, \quad (3)$$

を仮定する。 $P_i = (X_i - x_{0,i}) / (x_{1,i} - x_{0,i})$, $p_i = (y_i - x_{0,i}) / (x_{1,i} - x_{0,i})$ とおくと、 $0 \leq p_i \leq 1$ であり、

$$0 \leq P_i \leq 1, \quad (4)$$

$$E[P_i] \leq p_i, \quad (5)$$

が導かれる。 $x^+ = \max(0, x)$ とおく。一方、ベルヌーイ型確率ベクトル $\mathbf{Y} = (Y_1, Y_2)$ を次のように定める。

$$(Y_1, Y_2) = \begin{cases} (x_{0,1}, x_{0,2}) & \text{w. p. } 1 - \min(1, p_1 + p_2) = (1 - p_1 - p_2)^+ \\ (x_{1,1}, x_{0,2}) & \text{w. p. } \min(1, p_1 + p_2) - p_2 = \min(p_1, 1 - p_2) \\ (x_{0,1}, x_{1,2}) & \text{w. p. } \min(1, p_1 + p_2) - p_1 = \min(1 - p_1, p_2) \\ (x_{1,1}, x_{1,2}) & \text{w. p. } p_1 + p_2 - \min(1, p_1 + p_2) = (p_1 + p_2 - 1)^+ \end{cases}. \quad (6)$$

Y_i の周辺分布は次のように計算できる。

$$Y_i = \begin{cases} x_{0,i} & \text{w. p. } 1 - p_i \\ x_{1,i} & \text{w. p. } p_i \end{cases}. \quad (7)$$

従って、 $\mathbf{Y}$ は次の条件を満たすので $\mathbf{Y} \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})$ である。

$$x_{0,i} \leq Y_i \leq x_{1,i}, \quad (8)$$

$$E[Y_i] = y_i, \quad (9)$$

Theorem 1. $n = 2$ の場合について、 $\mathbf{X} \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})$ ならば、

$$E[\max \mathbf{X}] \leq E[\max \mathbf{Y}]$$

を満たす二変量ベルヌーイ型分布 $\mathbf{Y} \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})$ が見つかる。

Proof. $\mathbf{X} = (X_1, X_2)^T$ は次のように表現できる。

$$\begin{aligned} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} &= \begin{pmatrix} (1 - P_1)x_{0,1} + P_1x_{1,1} \\ (1 - P_2)x_{0,2} + P_2x_{1,2} \end{pmatrix} \\ &= (1 - \max(P_1, P_2)) \begin{pmatrix} x_{0,1} \\ x_{0,2} \end{pmatrix} + (\max(P_1, P_2) - P_2) \begin{pmatrix} x_{1,1} \\ x_{0,2} \end{pmatrix} \\ &\quad + (\max(P_1, P_2) - P_1) \begin{pmatrix} x_{0,1} \\ x_{1,2} \end{pmatrix} + (P_1 + P_2 - \max(P_1, P_2)) \begin{pmatrix} x_{1,1} \\ x_{1,2} \end{pmatrix}. \end{aligned}$$

この右辺の係数は非負で総和をとると 1 であることは容易に確認できる。max の凸性を用いることで、次の不等式を得る。

$$\begin{aligned} \max(X_1, X_2) &\leq (1 - \max(P_1, P_2)) \max(x_{0,1}, x_{0,2}) + (\max(P_1, P_2) - P_2) \max(x_{1,1}, x_{0,2}) \\ &\quad + (\max(P_1, P_2) - P_1) \max(x_{0,1}, x_{1,2}) + (P_1 + P_2 - \max(P_1, P_2)) \max(x_{1,1}, x_{1,2}). \end{aligned}$$

更に係数が 1, P_1 , P_2 , $\max(P_1, P_2)$ についてまとめなおすと右辺は次のようになる。

$$\max(x_{0,1}, x_{0,2}) + P_1 \Delta_{\{1\}} \max(x_{0,1}, x_{1,2}) + P_2 \Delta_{\{2\}} \max(x_{1,1}, x_{0,2}) - \max(P_1, P_2) \Delta_{\{1,2\}} \max(x_{0,1}, x_{0,2}).$$

$E[P_i] \leq p_i$ であることを思い出そう。また、Lemma 7 より $\max(P_1, P_2) \leq P_1 + P_2$ であるから両辺の期待値をとることで $E[\max(P_1, P_2)] \leq p_1 + p_2$ である。 $P_1 \leq 1$ かつ $P_2 \leq 1$ から $\max(P_1, P_2) \leq 1$ が成り立つので $E[\max(P_1, P_2)] \leq 1$ も成り立つ。これらを合わせて $E[\max(P_1, P_2)] \leq \min(1, p_1 + p_2)$ が成り立つ。また Lemma 4 より

$$\Delta_{\{1\}} \max(x_{0,1}, x_{1,2}) \geq 0, \quad \Delta_{\{2\}} \max(x_{1,1}, x_{0,2}) \geq 0, \quad \Delta_{\{1,2\}} \max(x_{0,1}, x_{0,2}) \leq 0$$

であるから、

$$\begin{aligned} E[\max(X_1, X_2)] &\leq (1 - \min(1, p_1 + p_2)) \max(x_{0,1}, x_{0,2}) + (\min(1, p_1 + p_2) - p_2) \max(x_{1,1}, x_{0,2}) \\ &\quad + (\min(1, p_1 + p_2) - p_1) \max(x_{0,1}, x_{1,2}) + (p_1 + p_2 - \min(1, p_1 + p_2)) \max(x_{1,1}, x_{1,2}) \end{aligned}$$

が得られる。この上界は $E[\max(Y_1, Y_2)]$ そのものである。□

本研究で与えた上界は、 $\mathbf{X}$ が非負でない場合でも評価が可能である。また、 $\mathbf{Y} \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})$ と合わせると、 $\max_{\mathbf{X} \in \mathcal{T}_2(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y})} E[\max(\mathbf{X})] = E[\max(\mathbf{Y})]$ であることがわかり、これ以上は $E[\max(\mathbf{X})]$ の上界を改善することはできない。注意として $\Pr(\max(\mathbf{X}) > x)$ の最善の上界を求めたわけではない。

この節の最後に非負有界確率ベクトル $\mathbf{X} = (X_1, X_2)$ について Lemma 7 による上界 $y_1 + y_2$ との誤差を求める。 $x_{0,1} \geq 0$ かつ $x_{0,2} \geq 0$ と仮定する。

$p_1 + p_2 \leq 1$ とする。このとき

$$\begin{aligned} y_1 + y_2 - E[\max(Y_1, Y_2)] &= \min(x_{0,1}, x_{0,2}) + p_1 ((x_{0,2} - x_{0,1})^+ - (x_{0,2} - x_{1,1})^+) + p_2 ((x_{0,1} - x_{0,2})^+ - (x_{0,1} - x_{1,2})^+). \end{aligned}$$

である。

$p_1 + p_2 > 1$ とする。このとき

$$\begin{aligned} y_1 + y_2 - E[\max(Y_1, Y_2)] &= (p_1 + p_2 - 1) \min(x_{1,1}, x_{1,2}) + (1 - p_1) \min(x_{0,1}, x_{1,2}) + (1 - p_2) \min(x_{1,1}, x_{0,2}). \end{aligned}$$

である。

4. まとめ

本論文では、 $n = 1, 2$ のときに第二種有界過程の最大順序統計量の期待値が、そのパラメータを使って構成した多変量ベルヌーイ型確率変数の期待値によって抑えられることを示した。証明の肝となるアイデアは有界確率ベクトルの取りうる直方体領域の端点に着目し、その内分比を使って max の凸性に関する不等式を構成する部分である。内分比について Lemma 7 と確率が 1 以下であることを併用することで、偶然の発見であるが、ベルヌーイ型

分布の $\max$ の期待値に帰着できた。このベルヌーイ型確率変数もまた第二種有界過程であるから、この上界は第二種有界過程のクラスでの最大値である。(これより小さい上界は探せない。) 注意として $\Pr(\max(\mathbf{X}) > x)$ の最善の上界を求めたわけではない。

$n \geq 3$ については、今後の課題であるが、現状で判明していることは、同じような議論を行うとベルヌーイ型分布の確率の部分が負になってしまう問題があって、第二種有界クラスに収まるようにうまく上界をとることができていない。単純に良い上界の構成方法を知らないだけなのか、どんなに頑張ってもそのような都合の良い上界を構成できないのかはわからない。 $\max$ は結合法則が成り立つので 2 つの場合を逐次的に適用することも別のアプローチとして考えられる。

参考文献

- [1] C.-S. Chang, *Performance guarantees in communication networks*. Springer Science & Business Media, 2000.
- [2] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queueing systems for the internet*. Springer Science & Business Media, 2001, vol. 2050.
- [3] R. L. Cruz, “A calculus for network delay. i. network elements in isolation,” *Information Theory, IEEE Transactions on*, vol. 37, no. 1, pp. 114–131, 1991.
- [4] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity: an algebra for discrete event systems*. John Wiley & Sons Ltd, 1992.
- [5] J. B. Schmitt, F. Zdarsky, M. Fidler *et al.*, “Delay bounds under arbitrary multiplexing: When network calculus leaves you in the lurch...” in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008, pp. 1669–1677.
- [6] M. Fidler, “An end-to-end probabilistic network calculus with moment generating functions,” in *Quality of Service, 2006. IWQoS 2006. 14th IEEE International Workshop on*. IEEE, 2006, pp. 261–270.
- [7] S. Mao and S. Panwar, “A survey of envelope processes and their applications in quality of service provisioning,” *Communications Surveys & Tutorials, IEEE*, vol. 8, no. 3, pp. 2–20, 2006.
- [8] Y. Jiang and Y. Liu, *Stochastic network calculus*. Springer, 2008, vol. 1.
- [9] M. Fidler, “Survey of deterministic and stochastic service curve models in the network calculus,” *Communications Surveys & Tutorials, IEEE*, vol. 12, no. 1, pp. 59–86, 2010.
- [10] Y. Jiang, “Stochastic network calculus for performance analysis of internet networks - an overview and outlook,” in *Computing, Networking and Communications (ICNC), 2012 International Conference on*. IEEE, 2012, pp. 638–644.
- [11] M. Fidler and A. Rizk, “A guide to the stochastic network calculus,” *Communications Surveys & Tutorials, IEEE*, vol. 17, no. 1, pp. 92–105, 2015.
- [12] F. Kelly, “Notes on effective bandwidths. stochastic networks: Theory and applications, royal statistical society lecture notes series, vol. 4,” 1996.
- [13] J. Liebeherr, S. Patek, and A. Burchard, “A calculus for end-to-end statistical service guarantees,” *University of Virginia, Charlottesville, VA*, 2001.
- [14] J. Liebeherr, S. D. Patek, and A. Burchard, “Statistical per-flow service bounds in a network with aggregate provisioning,” in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. IEEE, 2003, pp. 1680–1690.
- [15] C. Li, A. Burchard, and J. Liebeherr, “A network calculus with effective bandwidth,” *Networking, IEEE/ACM Transactions on*, vol. 15, no. 6, pp. 1442–1453, 2007.
- [16] K. Kobayashi, Y. Takahashi, and H. Takada, “A stochastic network calculus for many flows,” in *Teletraffic Congress, 2009. ITC 21 2009. 21st International*. IEEE, 2009, pp. 1–8.
- [17] —, “Asymptotic end-to-end stochastic evaluation for tandem networks with many flows,” in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*.

- ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 32.
- [18] H. Takada, “Joint moment generating function approach for stochastic network calculus,” in *the Queue Symposium 2015*, jan 2016.
 - [19] Y. Jiang, “A note on applying stochastic network calculus,” 2010.
 - [20] F. Ciucu, F. Poloczek, and J. Schmitt, “Sharp bounds in stochastic network calculus,” in *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1. ACM, 2013, pp. 367–368.
 - [21] —, “Sharp per-flow delay bounds for bursty arrivals: The case of fifo, sp, and edf scheduling,” in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 1896–1904.
 - [22] F. Ciucu and J. Schmitt, “Perspectives on network calculus: no free lunch, but still good value,” in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*. ACM, 2012, pp. 311–322.
 - [23] J. M. Felix Belzunce, Carolina Martinez-Riquelme, *An Introduction to Stochastic Orders*. Academic press, 2016.
 - [24] J. G. S. Moshe Shaked, *Stochastic Orders*. Springer-Verlag New York, 2007.
 - [25] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
 - [26] S. Ross, *A first course in probability*. Pearson, 2012.