MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Randomized Approximation Scheme and Perfect Sampler for Closed Jackson Networks with Multiple Servers

Shuji KIJIMA and Tomomi MATSUI

(Communicated by Kokichi SUGIHARA)

METR 2006–34 June 2006

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Randomized Approximation Scheme and Perfect Sampler for Closed Jackson Networks with Multiple Servers

Shuji KIJIMA[1] and Tomomi MATSUI[2]

[1] Department of Mathematical Informatics,
Graduate School of Information Science and Technology,
University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan
kijima@misojiro.t.u-tokyo.ac.jp

[2] Department of Information and System Engineering,
Faculty of Science and Engineering, Chuo University,
Bunkyo-ku, Tokyo 112-8551, Japan
http://www.simplex.t.u-tokyo.ac.jp/~tomomi/

June 5, 2006

Abstract

In this paper, we propose a fully polynomial-time randomized approximation scheme (FPRAS) for the closed Jackson network. Our algorithm is based on Markov chain Monte Carlo (MCMC) method. Thus, our scheme returns an approximate solution, of which the size of error satisfies a given error rate. To our knowledge, the algorithm is the first polynomial time MCMC algorithm for closed Jackson networks with multiple servers. We propose two of new ergodic Markov chains, both of which have a unique stationary distribution that is the product form solution of closed Jackson networks. One of them is for approximate sampler, and we show it mixes rapidly. The other is for perfect sampler based on monotone coupling from the past (CFTP) algorithm proposed by Propp and Wilson, and we show it has a monotone *update function*.

1 Introduction

A Jackson network is one of the basic and significant models in queueing network theory. In the model, customers receive service at nodes with multiple exponential servers on first-come-first-served (FCFS) basis, and move stochastically to a next node when service is completed. In [14], Jackson showed that the network has a product-form solution as the steady-state distribution of customers in the network [15, 12]. By computing the normalizing constant of the product-form solution, we can obtain important performance measures like as throughput, rates of utilization of stations, and so on.

There is well-known Buzen's algorithm [5], which computes the normalizing constant of the product-form solution. However, the running time of Buzen's algorithm is pseudo-polynomial time depending on the number of customers in a closed network. Chen and O'Cnneide [7] proposed a randomized algorithm based on Markov chain Monte Carlo (MCMC), but in some very special cases it becomes a polynomial-time algorithm. In [24], Ozawa proposed a perfect sampler for closed Jackson networks with single servers, however his chain mixes in pseudo-polynomial time.

In this paper, we propose a fully polynomial-time randomized approximation scheme (FPRAS) for calculating the normalizing constant of the product form solution of a closed Jackson network. We deal with the model that a given network is strongly connected, a class of customers is unique, no customer leaves or enters the network, and each node has multiple servers. Our algorithm is

[1]He is supported by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

based on MCMC method, and the approximation rate of our scheme is theoretically guaranteed in a stochastic form. Precisely, for any given parameter ε and δ, satisfying $0 < \varepsilon < 1$, $0 < \delta < 1$, an approximate solution Z obtained by our algorithm satisfies

$$\Pr[|Z - A| \leq \varepsilon A] \geq 1 - \delta$$

where A is the exact solution.

We propose two ergodic Markov chains, both of which have a unique stationary distribution that is the product-form solution of a given closed Jackson network. Here we note that our chains are NOT a simulation of a given queueing network, but just have a unique stationary distribution which is the same as the product-form solution of a network. We show that the *mixing time* of our chain for approximate sampler is bounded by $n(n-1)\ln(K\varepsilon^{-1})/2$ for an arbitrary positive $\varepsilon < 1$, where n is the number of nodes and K is the number of customers. A key idea which derives polynomiality is not to simulate behavior of customers in a network, while both algorithms of [7] and [24] simulate behavior of customers. We estimate the mixing time by using a technique of *path coupling* introduces in [4]. On the other hand, we show the *monotonicity* of the other chain, and design a perfect sampler based on monotone coupling from the past (monotone CFTP) algorithm proposed by [25].

2 Jackson Network

We denote the set of real numbers (non-negative, positive real numbers) by $\mathbb{R}$ ($\mathbb{R}_+$, $\mathbb{R}_{++}$), and the set of integers (non-negative, positive integers) by $\mathbb{Z}$ ($\mathbb{Z}_+$, $\mathbb{Z}_{++}$), respectively. A closed Jackson network is a queueing network model satisfying the followings;
(i) The network has $n \in \mathbb{Z}_{++}$ nodes. Node $i \in \{1, \ldots, n\}$ contains $s_i \in \mathbb{Z}_{++}$ servers, thus at most s_i customers can receive services on node i at a time.
(ii) In each node, customers are served on first-come-first-served (FCFS) basis. The servicing time of every server on node $i \in \{1, \ldots, n\}$ is exponentially distributed with mean $1/\mu_i \in \mathbb{R}_{++}$.
(iii) Once served in node $i \in \{1, \ldots, n\}$, a customer goes to node $j \in \{1, \ldots, n\}$ with probability $W_{ij} \in \mathrm{R}_+$. We assume that the matrix $W = (W_{ij})$ of transition probability of customers is irreducible and aperiodic, so ergodic.
(iv) No customers leave or enter the network. Thus, we assume that there are always $K \in \mathbb{Z}_{++}$ customers in the network.

In queueing network theory, it is well-known that a closed Jackson network has a *product form solution*, described below, as a steady state distribution of customers in a network. First, we introduce the set of non-negative integer points

$$\Delta(K) \stackrel{\text{def.}}{=} \left\{ \boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{Z}_+^n \mid \textstyle\sum_{i=1}^n x_i = K \right\}$$

in an $n-1$ dimensional simplex. Clearly, a state of K customers on nodes in a network with n nodes is represented by $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \Delta(K)$. Since matrix W of the transition probability of customers is ergodic, 1 is an eigenvalue and corresponding eigenvector is unique, excluding constant factor. Let $\theta \in \mathbb{R}_{++}^n$ be an eigenvector for W with corresponding to the eigenvalue 1, i.e., $\theta W = \theta$. The steady-state distribution $J^K : \Delta(K) \to \mathbb{R}_{++}$ for the closed Jackson network is product form defined by

$$J^K(\boldsymbol{x}) = \frac{1}{G(K)} \prod_{i=1}^n \alpha_i(x_i), \tag{2.1}$$

where $\alpha_i : \mathbb{Z}_+ \to \mathbb{R}_{++}$ is a function defined by

$$\alpha_i(z) \stackrel{\text{def.}}{=} \frac{1}{\prod_{j=1}^{z} \min\{j, s_i\}} \left(\frac{\theta_i}{\mu_i}\right)^z \equiv \begin{cases} \dfrac{1}{z!}\left(\dfrac{\theta_i}{\mu_i}\right)^z & (z \le s_i), \\ \dfrac{1}{s_i^{z-s_i} s_i!}\left(\dfrac{\theta_i}{\mu_i}\right)^z & (z > s_i), \end{cases} \tag{2.2}$$

for $z \in \mathbb{Z}_+$, (we denote $\prod_{j=1}^{0} \min\{j, s_i\} \stackrel{\text{def.}}{=} 1$ and $0! \stackrel{\text{def.}}{=} 1$), and $G(K) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{x} \in \Delta} \prod_{i=1}^{n} \alpha_i(x_i)$ is the normalizing constant. For convenience we denote $\alpha(\boldsymbol{x}) \stackrel{\text{def.}}{=} \prod_{i=1}^{n} \alpha_i(x_i)$ for any $\boldsymbol{x} \in \mathbb{Z}_+^n$.

3 Randomized Approximation Scheme

In the following we consider a closed Jackson network with n nodes and K customers which has the product form solution (2.1) for any $\boldsymbol{x} \in \Delta(K)$.

In this section, we give an FPRAS for calculating the normalizing constant $G(K)$ of product form solution for a closed Jackson network. Our approximation scheme is a standard Jerrum-Sinclair type recursive algorithm [17, 16], while there are some technical points.

3.1 Outline of approximation scheme

In this subsection, we outline our approximation scheme.

For a non-negative integer $N \in \mathbb{Z}_+$, and a non-negative integer vector $\boldsymbol{c} \in \mathbb{Z}_+^n$, we define a constant $G(N; \boldsymbol{c})$ by

$$G(N; \boldsymbol{c}) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{x} \in \Delta(N)} \prod_{i=1}^{n} \alpha_i(x_i + c_i).$$

Clearly, $G(K; \boldsymbol{0}) = G(K)$, what we want to compute. Given a node j in the network, we define a set $\Xi_j(N) \subset \Delta(N)$ by

$$\Xi_j(N) \stackrel{\text{def.}}{=} \left\{\boldsymbol{x} \in \Delta(N) \mid x_j \ge \left\lceil \tfrac{N}{n} \right\rceil \right\}.$$

Note that the convex hull of $\Xi_j(N)$ is also a simplex. We denote the jth unit vector by $\boldsymbol{e}_j$, whose jth element is one and others are zeros. Given a node j in the network, there is a bijection between $\Xi_j(N)$ and $\Delta\left(N - \left\lceil \frac{N}{n} \right\rceil\right)$, since obviously the properties

$$\left[\forall \boldsymbol{y} \in \Delta\left(N - \left\lceil \tfrac{N}{n} \right\rceil\right),\ \boldsymbol{y} + \left\lceil \tfrac{N}{n} \right\rceil \boldsymbol{e}_j \in \Xi_j(N)\right] \text{ and } \left[\forall \boldsymbol{x} \in \Xi_j(N),\ \boldsymbol{x} - \left\lceil \tfrac{N}{n} \right\rceil \boldsymbol{e}_j \in \Delta\left(N - \left\lceil \tfrac{N}{n} \right\rceil\right)\right]$$

hold. The subset $\Xi_j(N)$ of $\Delta(N)$ naturally leads the constant

$$F_j(N; \boldsymbol{c}) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{x} \in \Xi_j(N)} \prod_{i=1}^{n} \alpha_i(x_i + c_i).$$

Then we can see that

$$\begin{aligned} F_j(N; \boldsymbol{c}) &= \sum_{\boldsymbol{x} \in \Xi_j(N)} \prod_{i=1}^{n} \alpha_i(x_i + c_i) \\ &= \sum_{\boldsymbol{y} \in \Delta\left(N - \left\lceil \frac{N}{n} \right\rceil\right)} \alpha_j\left(y_j + c_j + \left\lceil \tfrac{N}{n} \right\rceil\right) \prod_{i \ne j} \alpha_i(y_i + c_i) \\ &= G\left(N - \left\lceil \tfrac{N}{n} \right\rceil; \boldsymbol{c} + \left\lceil \tfrac{N}{n} \right\rceil \boldsymbol{e}_j\right). \end{aligned}$$

Thus,

$$G(N;\boldsymbol{c}) = \frac{G(N;\boldsymbol{c})}{F_j(N;\boldsymbol{c})} \cdot G\left(N - \left\lceil \tfrac{N}{n} \right\rceil; \boldsymbol{c} + \left\lceil \tfrac{N}{n} \right\rceil \boldsymbol{e}_j\right) \tag{3.1}$$

holds and we can compute $G(N;\boldsymbol{c})$ if we know the ratio

$$\frac{F_j(N;\boldsymbol{c})}{G(N;\boldsymbol{c})} = \sum_{\boldsymbol{x}\in\Xi_j(N)} \frac{\prod_{i=1}^n \alpha_i(x_i+c_i)}{G(N;\boldsymbol{c})}$$

and the constant $G\left(N - \left\lceil \frac{N}{n} \right\rceil; \boldsymbol{c} + \left\lceil \frac{N}{n} \right\rceil \boldsymbol{e}_j\right)$. We define a function $J^N_{\boldsymbol{c}} : \Delta(N) \to \mathbb{R}_{++}$ by

$$J^N_{\boldsymbol{c}}(\boldsymbol{x}) \stackrel{\text{def.}}{=} \frac{1}{G(N;\boldsymbol{c})} \prod_{i=1}^n \alpha_i(x_i+c_i), \tag{3.2}$$

which is a probability function on $\Delta(N)$ satisfying

$$\frac{F_j(N;\boldsymbol{c})}{G(N;\boldsymbol{c})} = \sum_{\boldsymbol{x}\in\Xi_j(N)} J^N_{\boldsymbol{c}}(\boldsymbol{x}).$$

Then we can estimate the ratio $F_j(N;\boldsymbol{c})/G(N;\boldsymbol{c})$ by the Monte Carlo method with a sampler for $J^N_{\boldsymbol{c}}$. By using the equation (3.1) recursively, we can compute the constant $G(K)$ as follows. Given a sequence of indices (nodes) $(i^0, i^1, \ldots, i^{R-1}) \in \{1,\ldots,n\}^R$, we obtain that

$$\begin{aligned}
G(K) &= G(K;\boldsymbol{0}) \\
&= G(K^1;\boldsymbol{c}^1) \cdot \frac{G(K^0;\boldsymbol{c}^0)}{F_{i^0}(K^0;\boldsymbol{c}^0)} \\
&= G(K^2;\boldsymbol{c}^2) \cdot \frac{G(K^1;\boldsymbol{c}^1)}{F_{i^1}(K^1;\boldsymbol{c}^1)} \cdot \frac{G(K^0;\boldsymbol{c}^0)}{F_{i^0}(K^0;\boldsymbol{c}^0)} \\
&= \cdots \\
&= G(K^R;\boldsymbol{c}^R) \cdot \prod_{r=0}^{R-1} \frac{G(K^r;\boldsymbol{c}^r)}{F_{i^r}(K^r;\boldsymbol{c}^r)}
\end{aligned}$$

where $K^0 \stackrel{\text{def.}}{=} K$, $\boldsymbol{c}^0 \stackrel{\text{def.}}{=} \boldsymbol{0}$, $K^r \stackrel{\text{def.}}{=} K^{r-1} - \left\lceil \frac{K^{r-1}}{n} \right\rceil$, and $\boldsymbol{c}^r \stackrel{\text{def.}}{=} \boldsymbol{c}^{r-1} + \left\lceil \frac{K^{r-1}}{n} \right\rceil \boldsymbol{e}_{i^{r-1}}$ for $r = 1, 2, \ldots, R$. If we set the sequence $(i^0, \ldots, i^{R-1})$ satisfying that $K^{R-1} > K^R = 0$, then

$$G(K) = \alpha(\boldsymbol{c}^R) \cdot \prod_{r=0}^{R-1} \frac{G(K^r;\boldsymbol{c}^r)}{F_{i^r}(K^r;\boldsymbol{c}^r)}. \tag{3.3}$$

The above equality gives an idea of an approximation scheme for computing the constant $G(K)$ in a recursive fashion. In Section 3.3, we describe the detail of our scheme and discuss the approximation ratio. In the rest of this section, we estimate R, the number of recursions.

Lemma 3.1 *The number of recursions R satisfying $K^{R-1} > K^R = 0$ is bounded by $n \ln K + 1$.*

Proof: If $n = 1$, then $R = 1$ and hence we obtain the claim. If $n \geq 2$ and $K = 1, 2$ then $R = 1, 2$, respectively, hence we also obtain the claim. In the following, we consider the case $n \geq 2$ and $K \geq 3$. We define R' by

$$R' \stackrel{\text{def.}}{=} \min \left\{ r \;\middle|\; K \left(\tfrac{n-1}{n}\right)^r < 1 \right\},$$

then clearly $R \leq R'$, since $K' - \lceil K'/n \rceil \leq K'(n-1)/n$ for any $K' \in \mathbb{Z}_{++}$. Thus it is enough to show that

$$K \left(\tfrac{n-1}{n}\right)^{n \ln K + 1} \leq 1.$$

With considering $\ln K > 0$ and $n \geq 2$,

$$\left(\tfrac{n-1}{n}\right)^{n \ln K + 1} < \left(\tfrac{n-1}{n}\right)^{n \ln K} = \left(\left(1 - \tfrac{1}{n}\right)^n\right)^{\ln K} \leq \left(\tfrac{1}{\mathrm{e}}\right)^{\ln K} = 1/K.$$

Thus we obtain the claim. $\square$

For a polynomial-time approximation scheme, we need to consider the followings;

- How to sample efficiently from $J^N_{\boldsymbol{c}}(\boldsymbol{x})$?
- How to choose the index i^r?
- How many samples do we need for Monte Carlo approximation?

In Section 3.2, we will propose a new sampler for $J^N_{\boldsymbol{c}}(\boldsymbol{x})$ based on Markov chain and discuss the mixing time of our chain.

3.2 Rapidly mixing Markov chain

Given a vector $\boldsymbol{c} \in \mathbb{Z}^n_+$ and a positive integer $N \in \mathbb{Z}_{++}$, we propose a new Markov chain $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ with state space $\Delta(N)$. A transition of $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ from a current state $X \in \Delta(N)$ to a next state X' is defined as follows. First, we chose a distinct pair of indices $\{j_1, j_2\} \subset \{1, 2, \ldots, n\}$ uniformly at random. Next, let $k = X_{j_1} + X_{j_2}$, and chose $l \in \{0, 1, \ldots, k\}$ with probability

$$\frac{\alpha_{j_1}(l + c_{j_1})\alpha_{j_2}(k - l + c_{j_2})}{\sum_{s=0}^{k} \alpha_{j_1}(s + c_{j_1})\alpha_{j_2}(k - s + c_{j_2})}$$
$$\left(\equiv \frac{\alpha_{j_1}(l + c_{j_1})\alpha_{j_2}(k - l + c_{j_2}) \prod_{j \notin \{j_1, j_2\}} \alpha_j(X_j + c_j)}{\sum_{s=0}^{k} \alpha_{j_1}(s + c_{j_1})\alpha_{j_2}(k - s + c_{j_2}) \prod_{j \notin \{j_1, j_2\}} \alpha_j(X_j + c_j)} \right)$$

then set

$$X'_i = \begin{cases} l & (\text{for } i = j_1), \\ k - l & (\text{for } i = j_2), \\ X_i & (\text{otherwise}). \end{cases}$$

Since $\alpha_i(x)$ is a positive function, the Markov chain $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ is irreducible and aperiodic, so ergodic, hence has a unique stationary distribution. Also, $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ satisfies *detailed balance equation*

$$J^N_{\boldsymbol{c}}(\boldsymbol{x}) P(\boldsymbol{x} \to \boldsymbol{y}) = J^N_{\boldsymbol{c}}(\boldsymbol{y}) P(\boldsymbol{y} \to \boldsymbol{x})$$

0	$\alpha_i^+(0)\alpha_j^+(k)/A$	$\alpha_i^+(1)\alpha_j^+(k-1)/A$	$\cdots$	$\alpha_i^+(k)\alpha_j^+(0)/A$	1

0	$\alpha_i^+(0)\alpha_j^+(k+1)/A'$	$\alpha_i^+(1)\alpha_j^+(k)/A'$	$\alpha_i^+(2)\alpha_j^+(k-1)/A'$	$\cdots$	$\alpha_i^+(k+1)\alpha_j^+(0)/A'$	1

Figure 1: A figure of alternating inequalities for a pair of indices (i, j) and a non-negative integer k. We denote $\alpha_i^+(z) \stackrel{\text{def.}}{=} \alpha_i(z + c_i)$, $\alpha_j^+(z) \stackrel{\text{def.}}{=} \alpha_j(z + c_j)$, $A \stackrel{\text{def.}}{=} \sum_{s=0}^{k} \alpha_i^+(s)\alpha_j^+(k-s)$, and $A' \stackrel{\text{def.}}{=} \sum_{s=0}^{k+1} \alpha_i^+(s)\alpha_j^+(k+1-s)$.

for any states $\boldsymbol{x}, \boldsymbol{y} \in \Delta(K)$, where $P(\boldsymbol{x} \to \boldsymbol{y})$ denotes the transition probability from $\boldsymbol{x}$ to $\boldsymbol{y}$. Thus the stationary distribution is equivalent to $J_{\boldsymbol{c}}^N(\boldsymbol{x})$ defined by (3.2). We can obtain a sample w.r.t. the probability function $J_{\boldsymbol{c}}^N(\boldsymbol{x})$ by simulating $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ sufficiently many steps. Next we discuss the *mixing time* (defined below) of $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$.

Given a pair of probability distributions ν_1 and ν_2 on a finite state space Ω, the *total variation distance* between ν_1 and ν_2 is defined by

$$\mathrm{d_{TV}}(\nu_1, \nu_2) \stackrel{\text{def.}}{=} \max_{A \subseteq \Omega} \sum_{x \in A} |\nu_1(x) - \nu_2(x)| \equiv \frac{1}{2} \sum_{x \in \Omega} |\nu_1(x) - \nu_2(x)|. \tag{3.4}$$

Given an arbitrary positive real ε, the *mixing time* of an ergodic Markov chain $\mathcal{M}$ is defined by

$$\tau(\varepsilon) \stackrel{\text{def.}}{=} \max_{x \in \Omega} \{\min\{t \mid \forall s \geq t,\ \mathrm{d_{TV}}(\pi, P_x^s) \leq \varepsilon\}\} \tag{3.5}$$

where π is the stationary distribution of $\mathcal{M}$ and P_x^s is the probability distribution of the chain $\mathcal{M}$ at time period $s \geq 0$ with initial state x (at time period 0).

In the rest of this subsection, we show the following theorem.

Theorem 3.2 *For $0 < \forall \varepsilon < 1$, the mixing time $\tau(\varepsilon)$ of Markov chain $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ satisfies*

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \ln(N\varepsilon^{-1}).$$

Let $\boldsymbol{c} = (c_1, \ldots, c_n)$ be an arbitrary non-negative integer vector, and we consider the cumulative distribution function $g_{ij}^{[k,\boldsymbol{c}]} : \{0, 1, \ldots, k\} \to \mathbb{R}_+$ defined by

$$g_{ij}^{[k,\boldsymbol{c}]}(l) \stackrel{\text{def.}}{=} \frac{\sum_{s=0}^{l} \alpha_i(s + c_i)\alpha_j(k - s + c_j)}{\sum_{s=0}^{k} \alpha_i(s + c_i)\alpha_j(k - s + c_j)}. \tag{3.6}$$

We also define $g_{ij}^{[k,\boldsymbol{c}]}(-1) \stackrel{\text{def.}}{=} 0$, for convenience. We abbreviate $g_{ij}^{[k,\boldsymbol{c}]}$ to g_{ij}^k, if there is no confusion. We can describe a transition of the Markov chain $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ by using the function g_{ij}^k as follows. First, choose a distinct pair $\{j_1, j_2\}$ of indices with probability $2/(n(n-1))$. Next, put $k = X_{j_1} + X_{j_2}$, generate a uniformly random real number $\Lambda \in [0, 1)$, choose l satisfying $g_{j_1 j_2}^k(l-1) \leq \Lambda < g_{j_1 j_2}^k(l)$, and set $X'_{j_1} = l$, $X'_{j_2} = k - l$ and $X'_i = X_i$ for any other indices $i \notin \{j_1, j_2\}$.

The following lemma plays an important role for our main theorems.

Lemma 3.3 *For any $i \in \{1, \ldots, n\}$, the function $\alpha_i : \mathrm{Z}_+ \to \mathrm{R}_{++}$ $(i \in \{1, 2, \ldots, n\})$ is log-concave, i.e.,*

$$\ln \alpha_i(z) - \ln \alpha_i(z-1) \geq \ln \alpha_i(z+1) - \ln \alpha_i(z) \qquad (\forall z \in \mathbb{Z}_{++}). \tag{3.7}$$

It implies that, for any pair of distinct indices (i,j) $(i,j \in \{1,2,\ldots,n\})$, for any $k \in \mathbb{Z}_+$, and for any $\mathbf{c} \geq \mathbf{0}$, the inequalities

$$g_{ij}^{[k+1,\boldsymbol{c}]}(l) \leq g_{ij}^{[k,\boldsymbol{c}]}(l) \leq g_{ij}^{[k+1,\boldsymbol{c}]}(l+1) \qquad (\forall l \in \mathbb{Z}_+), \tag{3.8}$$

called alternating inequalities*, holds.*

Figure 1 illustrates the definition of alternating inequalities. In the following, we denote $\alpha_i^+(z) \stackrel{\text{def.}}{=} \alpha_i(z+c_i)$ and $\alpha_j^+(z) \stackrel{\text{def.}}{=} \alpha_j(z+c_j)$.

Proof: First, we show the log-concavity of $\alpha_i(z)$. From the equations (2.2) of the function $\alpha_i(z)$,

$$\ln \alpha_i(z) = \begin{cases} z \ln\left(\frac{\theta_i}{\mu_i}\right) - \sum_{j=1}^{z} \ln j & (z \leq s_i), \\ z \ln\left(\frac{\theta_i}{\mu_i}\right) - (z-s_i)\ln s_i - \sum_{j=1}^{s_i} \ln j & (z > s_i), \end{cases}$$

hold. Thus

$$\ln \alpha_i(z) - \ln \alpha_i(z-1) = \begin{cases} \ln\left(\frac{\theta_i}{\mu_i}\right) - \ln z & (z \leq s_i), \\ \ln\left(\frac{\theta_i}{\mu_i}\right) - \ln s_i & (z > s_i). \end{cases}$$

From the above, the function α_i satisfies (3.7), and thus α_i is log-concave. Since the function $\alpha_i^+(z) \stackrel{\text{def.}}{=} \alpha_i(z+c_i)$, obtained by shifting the domain of α_i by c_i, is also log-concave and satisfies

$$\ln \alpha_i^+(z) - \ln \alpha_i^+(z-1) \geq \ln \alpha_i^+(z+1) - \ln \alpha_i^+(z) \qquad (\forall z \in \mathbb{Z}_{++}). \tag{3.9}$$

Next, we show the latter statement. When $k=0$, it is obvious. When we fix $k \in \mathbb{Z}_{++}$, the alternating inequalities (3.8) hold for any $l \in \{0,1,\ldots,k\}$, if and only if

$$\begin{aligned} &\left(\sum_{s=0}^{l} \alpha_i^+(s)\alpha_j^+(k+1-s)\right)\left(\sum_{s'=l+1}^{k} \alpha_i^+(s')\alpha_j^+(k-s')\right) \\ &\quad \leq \left(\sum_{s=0}^{l} \alpha_i^+(s)\alpha_j^+(k-s)\right)\left(\sum_{s'=l+1}^{k+1} \alpha_i^+(s')\alpha_j^+(k+1-s')\right), \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} &\left(\sum_{s=0}^{l} \alpha_i^+(k+1-s)\alpha_j^+(s)\right)\left(\sum_{s'=l+1}^{k} \alpha_i^+(k-s')\alpha_j^+(s')\right) \\ &\quad \leq \left(\sum_{s=0}^{l} \alpha_i^+(k-s)\alpha_j^+(s)\right)\left(\sum_{s'=l+1}^{k+1} \alpha_i^+(k+1-s')\alpha_j^+(s')\right), \end{aligned} \tag{3.11}$$

hold for any $l \in \{0,1,\ldots,k-1\}$. With considering the expansion of (3.10), it is enough to show that $\forall s, \forall s' \in \{0,1,\ldots,k\}$, $0 \leq s < s' \leq k$ implies that

$$\alpha_i^+(s)\alpha_j^+(k+1-s)\alpha_i^+(s')\alpha_j^+(k-s') \leq \alpha_i^+(s)\alpha_j^+(k-s)\alpha_i^+(s')\alpha_j^+(k+1-s'). \tag{3.12}$$

Since, α_j is log-concave for any index $j \in \{1,2,\ldots,n\}$, the inequalities $(k-s') < (k-s'+1) \leq (k-s) < (k-s+1)$ implies that

$$\ln \alpha_j^+(k-s') + \ln \alpha_j^+(k-s+1) \leq \ln \alpha_j^+(k-s'+1) + \ln \alpha_j^+(k-s)$$

holds. From the above, the inequality (3.12) hold $\forall s, \forall s' \in \{0,1,\ldots,k\}$ satisfying $0 \leq s < s' \leq k$. We obtain inequality (3.11) in the same way as (3.10) by interchanging i and j. $\square$

We show Theorem 3.2 by using the following path coupling technique proposed in [4].

Theorem 3.4 (Path coupling [4]) *Let $\mathcal{M}$ be a finite ergodic Markov chain with a state space Ω. Let $H = (\Omega, \mathcal{E})$ be a connected undirected graph with vertex set Ω and edge set $\mathcal{E} \subset \Omega^2$. Let the length of all edges be 1, and let the distance between x and y, denoted by $d(x,y)$ and/or $d(y,x)$, be the length of a shortest path between x and y. Suppose that there exists a joint process $(X,Y) \mapsto (X',Y')$ with respect to $\mathcal{M}$ satisfying that whose marginals are a faithful copy of $\mathcal{M}$. If there exists a positive real β, exactly less than one, satisfying*

$$\mathrm{E}[d(X',Y')] \leq \beta d(X,Y)$$

for any edge $\{X,Y\} \in \mathcal{E}$ of H, then the mixing time $\tau(\varepsilon)$ of the Markov chain $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq (1-\beta)^{-1} \ln(\varepsilon^{-1} D),$$

where $D \stackrel{\text{def.}}{=} \max\{d(x,y) \mid \forall x, \forall y \in \Omega\}$ is the diameter of the graph H. ■

Proof of Theorem 3.2 Let $\mathcal{H} = (\Delta(N), \mathcal{E})$ be an undirected simple graph with vertex set $\Delta(N)$ and edge set $\mathcal{E}$ defined as follows. A pair of vertices $\{\boldsymbol{x}, \boldsymbol{y}\}$ is an edge of $\mathcal{H}$ if and only if $(1/2)\sum_{i=1}^{n} |x_i - y_i| = 1$. Clearly the graph $\mathcal{H}$ is connected. We define the length of an edge $e \in \mathcal{E}$ as 1, and the distance $d_{\mathrm{A}}(\boldsymbol{x}, \boldsymbol{y})$ for each pair $(\boldsymbol{x}, \boldsymbol{y}) \in (\Delta(N))^2$ by the length of a shortest path from $\boldsymbol{x}$ to $\boldsymbol{y}$ on $\mathcal{H}$. Clearly, the diameter of $\mathcal{H}$ defined by $\max_{\boldsymbol{x}, \boldsymbol{y} \in \Delta(N)}\{d_{\mathrm{A}}(\boldsymbol{x}, \boldsymbol{y})\}$, is bounded by N.

We define a joint process $(X,Y) \mapsto (X',Y')$ for any pair $\{X,Y\} \in \mathcal{E}$. Pick a distinct pair of indices $\{i_1, i_2\}$ uniformly at random. Then put $k_X = X_{i_1} + X_{i_2}$ and $k_Y = Y_{i_1} + Y_{i_2}$, generate a uniformly random number $\Lambda \in [0,1)$, chose $l_X \in \{0,1,\ldots,k_X\}$ and $l_Y \in \{0,1,\ldots,k_Y\}$ which satisfy $g^{k_X}_{i_1 i_2}(l_X - 1) \leq \Lambda < g^{k_X}_{i_1 i_2}(l_X)$ and $g^{k_Y}_{i_1 i_2}(l_Y - 1) \leq \Lambda < g^{k_Y}_{i_1 i_2}(l_Y)$, and set $X'_{i_1} = l_X$, $X'_{i_2} = k_X - l_X$, $Y'_{i_1} = l_Y$ and $Y'_{i_2} = k_Y - l_Y$.

Now we show that

$$\beta \stackrel{\text{def.}}{=} 1 - \frac{2}{n(n-1)}$$

satisfies

$$\mathrm{E}[d_{\mathrm{A}}(Y',Y')] \leq \beta d_{\mathrm{A}}(X,Y)$$

for any pair $\{X,Y\} \in \mathcal{E}$. Here we suppose that $X, Y \in \mathcal{E}$ satisfies $|X_j - X_j| = 1$ for $j \in \{j_1, j_2\}$, and $|X_j - X_j| = 0$ for $j \notin \{j_1, j_2\}$.

<u>**Case 1**</u>: In case that neither of indices j_1 nor j_2 are chosen, i.e., $\{i_1, i_2\} \cap \{j_1, j_2\} = \emptyset$. Put $k = X_{i_1} + X_{i_2}$, then it is easy to see that $\Pr(X'_{i_1} = l) = \Pr(Y'_{i_1} = l)$ for any $l \in \{0, \ldots, k\}$ since $Y_{i_1} + Y_{i_2} = k$. By setting $X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}$, we have $d_{\mathrm{A}}(X',Y') = d_{\mathrm{A}}(X,Y)$.

<u>**Case 2**</u>: In case that both of indices j_1 and j_2 are chosen, i.e., $\{i_1, i_2\} = \{j_1, j_2\}$. In the same way as Case 1, we can set $X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}$. Hence $d_{\mathrm{A}}(X',Y') = 0$.

<u>**Case 3**</u>: In case that exactly one of indices j_1 and j_2 is chosen, i.e., $|\{i_1, i_2\} \cap \{j_1, j_2\}| = 1$. Without loss of generality, we can assume that $i_1 = j_1$ and that $X_{i_1} + 1 = Y_{i_1}$. Let $k = X_{i_1} + X_{i_2}$. Then $Y_{i_1} + Y_{i_2} = k + 1$ obviously. Let us consider a transition of the joint process with a random number $\Lambda \in [0,1)$. Let $l \in \{0,1,\ldots,k\}$ be a unique index satisfying $g^{k}_{i_1 i_2}(l-1) \leq \Lambda < g^{k}_{i_1 i_2}(l)$. Then alternating inequalities imply that $g^{k+1}_{i_1 i_2}(l-1) \leq \Lambda < g^{k+1}_{i_1 i_2}(l+1)$. Therefore, if $X'_{i_1} = l$ then Y'_{i_1}

should be in $\{l, l+1\}$ by the definition of the joint process. Thus we always obtain that $[X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} + 1 = Y'_{i_2}]$ or $[X'_{i_1} + 1 = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}]$. Hence $d_{\mathrm{A}}(X', Y') = d_{\mathrm{A}}(X, Y)$.

With considering that Case 2 occurs with probability $2/(n(n-1))$, we obtain that

$$\mathrm{E}[d_{\mathrm{A}}(X', Y')] \le \left(1 - \frac{2}{n(n-1)}\right) d_{\mathrm{A}}(X, Y).$$

Since the diameter of $\mathcal{H}$ is bounded by N, Theorem 3.4 (Path Coupling Theorem) implies that the mixing time $\tau(\varepsilon)$ satisfies

$$\tau(\varepsilon) \le \frac{n(n-1)}{2} \ln(N\varepsilon^{-1}).$$

□

3.3 Monte Carlo integration

In this section, we give an FPRAS for calculating the normalizing constant $G(K)$ of product form solution for a closed Jackson network. Since we already have an approximate sampler via the Markov chain $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$, we can estimate $G(K^r; \boldsymbol{c}^r)/F_i(K^r; \boldsymbol{c}^r)$ for $r \in \{0, 1, \ldots, R-1\}$ by the Monte Carlo method. The whole algorithm is as follows,

Algorithm 1 (Randomize Approximation Scheme with Approximate Sampler)

Step 0. Set $r := 0$, $N := K$, and $\boldsymbol{c} := \boldsymbol{0}$.
Step 1. While $N \ge 1$, repeat the followings from i to iv.
  i. Generate Q samples, each of which is obtained by simulating $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ for $T_{\mathrm{A}}(N)$ steps.
  ii. For each $i \in \{1, \ldots, n\}$, U_i denotes the number of samples satisfying $x_i \ge N/n$.
  iii. Let $I \in \{1, \ldots, n\}$ be an index with $U_I = \max\{U_1, \ldots, U_n\}$. Set $Z^r := U_I/Q$.
  iv. Set $\boldsymbol{c} := \boldsymbol{c} + \lceil N/n \rceil \cdot \boldsymbol{e}_I$, $N := N - \lceil N/n \rceil$, and $r := r + 1$.
Step 2. Output $Z := \alpha(\boldsymbol{c}) \cdot \prod_{r=0}^{R-1} 1/Z^r$.

Our algorithm generates Q samples by simulating $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ for $T_{\mathrm{A}}(N)$ steps for each sample. By setting $Q = 150nR^2\varepsilon^{-2}\ln(2nR/\delta)$ and $T_{\mathrm{A}}(k) = \left\lceil \frac{n(n-1)}{2} \ln \frac{10nRk}{\varepsilon} \right\rceil$, we obtain the following theorem. Note that, the definition of Step 1 iii of Algorithm 1 implies $U_I \ge N/n$ and $Z^r \ge 1/n$ for each iteration.

Theorem 3.5 *If we set $Q = 150nR^2\varepsilon^{-2}\ln(2nR/\delta)$ and $T_{\mathrm{A}}(k) = \left\lceil \frac{n(n-1)}{2} \ln \frac{10nRk}{\varepsilon} \right\rceil$, then our randomized approximation scheme (Algorithm 1) returns Z satisfying*

$$\Pr\left[|Z - G(K)| \le \varepsilon G(K)\right] \ge 1 - \delta. \tag{3.13}$$

Proof: First we consider r-th iteration (where $r \in \{0 \ldots, R-1\}$) of Step 1 in Algorithm 1 and show the following Claims 1 to 4. Suppose the index I, integer N and the vector $\boldsymbol{c}$ are chosen in the r-th iteration of Step 1 of Algorithm 1, and we define ω^r by

$$\omega^r \stackrel{\text{def.}}{=} \frac{F_I(N; \boldsymbol{c})}{G(N; \boldsymbol{c})}.$$

We denote p_i for $i \in \{1, \ldots, n\}$ as a probability that a sample satisfies $x_i \ge N/n$ where $\boldsymbol{x}$ is obtained by using Markov chian $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ after $T_{\mathrm{A}}(N)$ steps. Particularly we define $\widehat{\omega}^r \stackrel{\text{def.}}{=} p_I$. Note that $\widehat{\omega}^r$ approximates ω^r.

Claim 1. For each $r \in \{0, \dots, R-1\}$, $\Pr\left[\widehat{\omega}^r \le \frac{1}{2n}\right] < \frac{\delta}{2R}$ holds.

Let $S \subset \{1, \dots, n\}$ be the set of indices defined by $S \stackrel{\text{def.}}{=} \{i \mid p_i \le 1/(2n)\}$. Note that $|S| \le n-1$. For an arbitrary $i \in S$, the probability of the event $U_i/Q \ge 1/n$ satisfies

$$\begin{aligned}\Pr\left[\frac{U_i}{Q} \ge \frac{1}{n}\right] &= \Pr\left[\frac{U_i}{Q} \ge \left(1 + \left(\frac{1}{np_i} - 1\right)\right) p_i\right] \\ &\le \mathrm{e}^{-\frac{1}{3}\left(\frac{1}{np_i}-1\right)^2 Qp_i} \le \mathrm{e}^{-\frac{1}{3}1^2 Q\frac{1}{2n}} < \mathrm{e}^{-\ln\frac{2nR}{\delta}} \le \frac{\delta}{2nR}\end{aligned}$$

where the first inequality is obtained by using the Chernoff bound (see [23], p. 64 for example). Since $U_I/Q \ge 1/n$, the probability of the event $I \in S$ satisfies

$$\Pr[I \in S] = \sum_{i \in S} \Pr[I = i] \le \sum_{i \in S} \Pr\left[\frac{U_i}{Q} \ge \frac{1}{n}\right] \le \sum_{i \in S} \frac{\delta}{2nR} \le (n-1) \cdot \frac{\delta}{2nR} < \frac{\delta}{2R}$$

with considering $|S| \le n-1$, and we obtain Claim 1.

Claim 2. For each $r \in \{0, \dots, R-1\}$, if $\widehat{\omega}^r > \frac{1}{2n}$ then $|\omega^r - \widehat{\omega}^r| \le \frac{\varepsilon}{5R}\widehat{\omega}^r$ holds.

With considering the definition of the mixing time (3.5) and the total variation distance (3.4), Theorem 3.2 and the definition of $T_{\mathrm{A}}(N)$ imply

$$|\omega^r - \widehat{\omega}^r| \le \mathrm{d_{TV}}\left(P_{\boldsymbol{y}}^{T_{\mathrm{A}}(N)}, \pi\right) \le \frac{\varepsilon}{10nR}$$

where $P_{\boldsymbol{y}}^{T_{\mathrm{A}}(N)}$ denote the distribution of a sample obtained by Markov chain $\mathcal{M}_{\mathrm{A}}(N; \boldsymbol{c})$ after $T_{\mathrm{A}}(N)$ steps with an initial state $\boldsymbol{y} \in \Delta(N)$. Thus if $\widehat{\omega}^r > \frac{1}{2n}$, then

$$|\omega^r - \widehat{\omega}^r| \le \frac{\varepsilon}{10nR} = \frac{\varepsilon}{5R} \cdot \frac{1}{2n} \le \frac{\varepsilon}{5R}\widehat{\omega}^r$$

hold and we obtain Claim 2.

Claim 3. For each $r \in \{0, \dots, R-1\}$, if $\widehat{\omega}^r > \frac{1}{2n}$ then the conditional probability $\Pr\left[|Z^r - \widehat{\omega}^r| \ge \frac{\varepsilon}{5R}\widehat{\omega}^r \;\middle|\; \widehat{\omega}^r > \frac{1}{2n}\right] \le \frac{\delta}{2R}$ holds, where Z^r is defined in Step 1 iii of Algorithm 1.

By using Chernoff bound, $n \ge 2$ implies

$$\begin{aligned}\Pr\left[|Z^r - \widehat{\omega}^r| \ge \frac{\varepsilon}{5R}\widehat{\omega}^r \;\middle|\; \widehat{\omega}^r > \frac{1}{2n}\right] &\le 2\mathrm{e}^{-\left(\frac{\varepsilon}{5R}\right)^2 \frac{1}{3} 150nR^2\varepsilon^{-2}\ln\frac{2nR}{\delta}\widehat{\omega}^r} \\ &= 2\mathrm{e}^{-2n\ln\frac{2nR}{\delta}\widehat{\omega}^r} \le 2\mathrm{e}^{-2n\ln\frac{2nR}{\delta}\frac{1}{2n}} = 2\mathrm{e}^{-\ln\frac{2nR}{\delta}} = \frac{2\delta}{2nR} \le \frac{\delta}{2R}\end{aligned}$$

and we obtain the Claim 3.

From Claims 1, 2 and 3, we obtain the following claim.

Claim 4. For each $r \in \{0, \dots, R-1\}$, $\Pr\left[\left(1 + \frac{\varepsilon}{2R}\right)^{-1} \le \frac{\omega^r}{Z^r} \le \left(1 + \frac{\varepsilon}{2R}\right)\right] \ge 1 - \frac{\delta}{R}$ holds.

When $\widehat{\omega}^r > \frac{1}{2n}$, Claim 2 implies

$$\left(1 - \frac{\varepsilon}{5R}\right)\widehat{\omega}^r \le \omega^r \le \left(1 + \frac{\varepsilon}{5R}\right)\widehat{\omega}^r \qquad (\forall r \in \{1, \dots, R-1\}) \tag{3.14}$$

and, Claim 3 also implies

$$\Pr\left[\left(1+\tfrac{\varepsilon}{5R}\right)^{-1} Z^r \leq \widehat{\omega}^r \leq \left(1-\tfrac{\varepsilon}{5R}\right)^{-1} Z^r \;\middle|\; \widehat{\omega}^r > \tfrac{1}{2n}\right] \geq 1-\frac{\delta}{2R} \qquad (\forall r \in \{1,\ldots,R-1\})\,. \tag{3.15}$$

Thus if $\widehat{\omega}^r > \frac{1}{2n}$, inequalities (3.14) and (3.15) imply that for each $r \in \{1,\ldots,R-1\}$,

$$\left(1-\frac{\varepsilon}{5R}\right)\left(1+\frac{\varepsilon}{5R}\right)^{-1} \leq \frac{\omega^r}{Z^r} \leq \left(1+\frac{\varepsilon}{5R}\right)\left(1-\frac{\varepsilon}{5R}\right)^{-1} \tag{3.16}$$

hold with the probability higher than $1-\frac{\delta}{2R}$. The right-hand side of (3.16) satisfies

$$\text{(r.h.s.)} = \left(1+\frac{\varepsilon}{5R}\right)\left(1-\frac{\varepsilon}{5R}\right)^{-1} = \left(1+\frac{2\varepsilon}{5R-\varepsilon}\right) \leq \left(1+\frac{2\varepsilon}{4R}\right) = \left(1+\frac{\varepsilon}{2R}\right)$$

and the left-hand side of (3.16) satisfies

$$\text{(l.h.s.)} = \left(1-\frac{\varepsilon}{5R}\right)\left(1+\frac{\varepsilon}{5R}\right)^{-1} = \text{(r.h.s.)}^{-1} \geq \left(1+\frac{\varepsilon}{2R}\right)^{-1}.$$

Thus, from the above discussion and Claim 1,

$$\begin{aligned}
&\Pr\left[\left(1+\frac{\varepsilon}{2R}\right)^{-1} \leq \frac{\omega^r}{Z^r} \leq \left(1+\frac{\varepsilon}{2R}\right)\right] \\
\geq\ & \Pr\left[\left(1+\tfrac{\varepsilon}{5R}\right)^{-1} Z^r \leq \omega^r \leq \left(1+\tfrac{\varepsilon}{5R}\right) Z^r \;\middle|\; \widehat{\omega}^r > \tfrac{1}{2n}\right] \cdot \Pr\left[\widehat{\omega}^r > \tfrac{1}{2n}\right] \\
\geq\ & \left(1-\frac{\delta}{2R}\right)\left(1-\frac{\delta}{2R}\right) \ \geq\ 1-\frac{\delta}{R}
\end{aligned}$$

hold and we obtain Claim 4.

From Claim 4, we obtain the following.

Claim 5. The random variables $Z^0,\ldots,Z^{R-1}$ in Algorithm 1 satisfies

$$\Pr\left[(1-\varepsilon) \leq \frac{\omega^0\cdots\omega^{R-1}}{Z^0\cdots Z^{R-1}} \leq (1+\varepsilon)\right] \geq 1-\delta.$$

By multiplying inequalities in Claim 4, we obtain

$$\Pr\left[\left(1+\frac{\varepsilon}{2R}\right)^{-R} \leq \frac{\omega^0\cdots\omega^{R-1}}{Z^0\cdots Z^{R-1}} \leq \left(1+\frac{\varepsilon}{2R}\right)^{R}\right] \geq \left(1-\frac{\delta}{R}\right)^R \geq 1-\delta.$$

Since $\left(1+\frac{\varepsilon}{2R}\right)^R \leq 1+\varepsilon$ and $\left(1+\frac{\varepsilon}{2R}\right)^{-R} \geq (1+\varepsilon)^{-1} \geq 1-\varepsilon$, Claim 5 is now clear.

Lastly, we conclude the proof of Theorem 3.5. Let $\boldsymbol{c}$ be the vector in Step 2 of Algorithm 1, then the equality (3.3) implies $\alpha(\boldsymbol{c})/(\omega^0\cdots\omega^{R-1}) = G(K)$, and the output of Algorithm 1 is $Z = \alpha(\boldsymbol{c})/(Z^0\cdots Z^{R-1})$. Thus Claim 5 implies

$$\Pr\left[(1-\varepsilon) \leq \frac{Z}{G(K)} \leq (1+\varepsilon)\right] \geq 1-\delta.$$

Hence we obtain the theorem. $\square$

4 Perfect Sampler

In this section, we propose another sampler for $J^N_{\boldsymbol{c}}$, which is a perfect sampler based on monotone CFTP. We propose another Markov chain $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$, which is a modified version of $\mathcal{M}_{\mathrm{A}}(N;\boldsymbol{c})$, and show that the chain is *monotone*. In the proof of monotonicity, Lemma 3.3 in the previous section plays a key roll again. Note that the problem of sampling from the product-form solution $J^K (= J^K_{\mathbf{0}})$ of closed Jackson networks (2.1) is included as a special case.

4.1 Monotone coupling from the past

Here we review CFTP briefly [25]. Suppose that we have an ergodic Markov chain $\mathcal{M}$ with a finite state space Ω and a transition matrix P. The transition rule of the Markov chain $X \mapsto X'$ can be described by a deterministic function $\phi : \Omega \times [0,1) \to \Omega$, called *update function*, as follows. Given a random number Λ uniformly distributed over $[0,1)$, update function ϕ satisfies that $\Pr(\phi(x,\Lambda) = y) = P(x,y)$ for any $x,y \in \Omega$. We can realize the Markov chain by setting $X' = \phi(X,\Lambda)$. Clearly, update functions corresponding to the given transition matrix P are not unique. The result of transitions of the chain from the time t_1 to t_2 $(t_1 < t_2)$ with a sequence of random numbers $\boldsymbol{\lambda} = (\lambda[t_1], \lambda[t_1+1], \ldots, \lambda[t_2-1]) \in [0,1)^{t_2-t_1}$ is denoted by $\Phi^{t_2}_{t_1}(x,\boldsymbol{\lambda}) : \Omega \times [0,1)^{t_2-t_1} \to \Omega$ where $\Phi^{t_2}_{t_1}(x,\boldsymbol{\lambda}) \stackrel{\text{def.}}{=} \phi(\phi(\cdots(\phi(x,\lambda[t_1]),\ldots,\lambda[t_2-2]),\lambda[t_2-1])$. We say that a sequence $\boldsymbol{\lambda} \in [0,1)^{|T|}$ satisfies the *coalescence condition*, when $\exists y \in \Omega$, $\forall x \in \Omega$, $y = \Phi^0_T(x,\boldsymbol{\lambda})$.

Suppose that there exists a partial order "$\succeq$" on the set of states Ω, and that a unique pair of states $(x_{\max}, x_{\min})$ exists in the partially ordered set $(\Omega, \succeq)$, satisfying $x_{\max} \succeq x \succeq x_{\min}$, $\forall x \in \Omega$. A transition rule expressed by a deterministic update function ϕ is called *monotone* (with respect to "$\succeq$") if $\forall \lambda \in [0,1)$, $\forall x, \forall y \in \Omega$, $x \succeq y \Rightarrow \phi(x,\lambda) \succeq \phi(y,\lambda)$. We also say that a chain is *monotone* if the chain has a *monotone* update function.

With these preparations, a standard monotone CFTP algorithm is expressed as follows.

Algorithm 2 (Monotone CFTP Algorithm [25])

Step 1. Set the starting time period $T := -1$ to go back, and set $\boldsymbol{\lambda}$ be the empty sequence.

Step 2. Generate random real numbers $\lambda[T], \lambda[T+1], \ldots, \lambda[\lceil T/2 \rceil - 1] \in [0,1)$, and insert them to the head of $\boldsymbol{\lambda}$ in order, i.e., put $\boldsymbol{\lambda} := (\lambda[T], \lambda[T+1], \ldots, \lambda[-1])$.

Step 3. Start two chains from $x_{\max}$ and $x_{\min}$, respectively, at time period T, and run each chain to time period 0 according to the update function ϕ with the sequence of numbers in $\boldsymbol{\lambda}$. (Here we note that every chain uses the common sequence $\boldsymbol{\lambda}$.)

Step 4. [Coalescence check] The state obtained at time period 0 is denoted by $\Phi^0_T(x,\boldsymbol{\lambda})$.
 (a) If $\exists y \in \Omega$, $y = \Phi^0_T(x_{\max},\boldsymbol{\lambda}) = \Phi^0_T(x_{\min},\boldsymbol{\lambda})$, then return y.
 (b) Else, update the starting time period $T := 2T$, and go to Step 2.

Theorem 4.1 (Monotone CFTP Theorem [25]) *Suppose that a Markov chain defined by an update function ϕ is monotone with respect to a partially ordered set of states $(\Omega, \succeq)$, and $\exists x_{\max}, \exists x_{\min} \in \Omega$, $\forall x \in \Omega$, $x_{\max} \succeq x \succeq x_{\min}$. Then the monotone CFTP algorithm (Algorithm 2) terminates with probability 1, and obtained value is a realization of a random variable exactly distributed according to the stationary distribution.* ■

Theorem 4.1 says that Algorithm 2 is a (probabilistically) finite time algorithm for infinite time simulation.

4.2 Monotone Markov chain

In this section we propose another Markov chain $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$ with a state space $\Delta(N)$ for the probability distribution $J_{\boldsymbol{c}}^N$. We abbreviate $\Delta(N)$ as Δ in the following. The transition rule of $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$ is defined by the following *update function* $\phi : \Delta \times [1,n) \to \Delta$. For a current state $X \in \Delta$, the next state $X' = \phi(X,\lambda) \in \Delta$ with respect to a uniformly random number $\lambda \in [1,n)$ is defined by

$$X_i' \;=\; \begin{cases} l & (\text{for } i = \lfloor\lambda\rfloor), \\ k-l & (\text{for } i = \lfloor\lambda\rfloor + 1), \\ X_i & (\text{otherwise}), \end{cases}$$

where $k = X_{\lfloor\lambda\rfloor} + X_{\lfloor\lambda\rfloor+1}$ and $l \in \{0,1,\ldots,k\}$ satisfies

$$g^{[k,\boldsymbol{c}]}_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l-1) \le \lambda - \lfloor\lambda\rfloor < g^{[k,\boldsymbol{c}]}_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l),$$

where $g^{[k,\boldsymbol{c}]}_{ij}$ is defined by (3.6). In the following, we abbreviate $g^{[k,\boldsymbol{c}]}_{ij}$ as g^k_{ij} in the same manner as the previous section. Our chain $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$ is a modified version of $\mathcal{M}_{\mathrm{A}}(N;\boldsymbol{c})$, obtained by restricting to choose only a consecutive pair of indices. Clearly, $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$ is ergodic. The chain has a unique stationary distribution $J_{\boldsymbol{c}}^N$ defined by (3.2).

In the following, we show the monotonicity of $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$. Here we introduce a partial order "$\succeq$" on Δ. For any state $\boldsymbol{x} \in \Delta$, we define *cumulative sum vector* $h_{\boldsymbol{x}} = (h_{\boldsymbol{x}}(0), h_{\boldsymbol{x}}(1), \ldots, h_{\boldsymbol{x}}(n)) \in \mathbb{Z}_+^{n+1}$ by

$$h_{\boldsymbol{x}}(i) \stackrel{\text{def.}}{=} \begin{cases} 0 & (\text{for } i = 0), \\ \sum_{j=1}^{i} x_j & (\text{for } i \in \{1,2,\ldots,n\}). \end{cases}$$

For any pair of states $\boldsymbol{x},\boldsymbol{y} \in \Delta$, we say $\boldsymbol{x} \succeq \boldsymbol{y}$ if and only if $h_{\boldsymbol{x}} \ge h_{\boldsymbol{y}}$. Next, we define two special states $\boldsymbol{x}_{\max}, \boldsymbol{x}_{\min} \in \Delta$ by $\boldsymbol{x}_{\max} \stackrel{\text{def.}}{=} (N,0,\ldots,0)$ and $\boldsymbol{x}_{\min} \stackrel{\text{def.}}{=} (0,\ldots,0,N)$. Then we can easily see that $\forall \boldsymbol{x} \in \Delta$, $\boldsymbol{x}_{\max} \succeq \boldsymbol{x} \succeq \boldsymbol{x}_{\min}$.

Theorem 4.2 *Markov chain $\mathcal{M}_{\mathrm{P}}(N;\boldsymbol{c})$ is monotone with respect to the partially ordered set $(\Delta,\succeq)$, i.e., $\forall \lambda \in [1,n)$, $\forall X, \forall Y \in \Delta$, $X \succeq Y \Rightarrow \phi(X,\lambda) \succeq \phi(Y,\lambda)$.*

Proof: We say that a state $X \in \Delta$ *covers* $Y \in \Delta$ (at j), denoted by $X \cdot\!\!\succ Y$ (or $X \cdot\!\!\succ_j Y$), when

$$X_i - Y_i = \begin{cases} +1 & (\text{for } i = j), \\ -1 & (\text{for } i = j+1), \\ 0 & (\text{otherwise}). \end{cases}$$

We show that if a pair of states $X, Y \in \Delta$ satisfies $X \cdot\!\!\succ_j Y$, then $\forall \lambda \in [1,n)$, $\phi(X,\lambda) \succeq \phi(Y,\lambda)$. We denote $\phi(X,\lambda)$ by X' and $\phi(Y,\lambda)$ by Y' for simplicity. For any index $i \ne \lfloor\lambda\rfloor$, it is easy to see that $h_X(i) = h_{X'}(i)$ and $h_Y(i) = h_{Y'}(i)$, and so $h_{X'}(i) - h_{Y'}(i) = h_X(i) - h_Y(i) \ge 0$ since $X \succeq Y$. In the following, we show that $h_{X'}(\lfloor\lambda\rfloor) \ge h_{Y'}(\lfloor\lambda\rfloor)$.

<u>Case 1</u>: If $\lfloor\lambda\rfloor \ne j-1$ and $\lfloor\lambda\rfloor \ne j+1$. Let $k = X_{\lfloor\lambda\rfloor} + X_{\lfloor\lambda\rfloor+1}$, then it is easy to see that $Y_{\lfloor\lambda\rfloor} + Y_{\lfloor\lambda\rfloor+1} = k$. Accordingly $X'_{\lfloor\lambda\rfloor} = Y'_{\lfloor\lambda\rfloor} = l$ where l satisfies

$$g^k_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l-1) \le \lambda - \lfloor\lambda\rfloor < g^k_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l),$$

and hence $h_{X'}(\lfloor\lambda\rfloor) = h_{Y'}(\lfloor\lambda\rfloor)$.

Case 2: Consider the case that $\lfloor\lambda\rfloor = j-1$. Let $k+1 = X_{j-1} + X_j$, then $Y_{j-1} + Y_j = k$, since $X \cdot\!\succ_j Y$. From the definition of cumulative sum vector,

$$\begin{array}{rl} h_{X'}(\lfloor\lambda\rfloor) - h_{Y'}(\lfloor\lambda\rfloor) & \\ = & h_{X'}(j-1) - h_{Y'}(j-1) \\ = & h_{X'}(j-2) + X'_{j-1} - h_{Y'}(j-2) - Y'_{j-1} \\ = & h_X(j-2) + X'_{j-1} - h_Y(j-2) - Y'_{j-1} \\ = & X'_{j-1} - Y'_{j-1}. \end{array}$$

Thus, it is enough to show that $X'_{j-1} \geq Y'_{j-1}$. Now suppose that $l \in \{0, 1, \ldots, k\}$ satisfies $g^k_{(j-1)j}(l-1) \leq \lambda - \lfloor\lambda\rfloor < g^k_{(j-1)j}(l)$ for λ. Then $g^{k+1}_{(j-1)j}(l-1) \leq \lambda - \lfloor\lambda\rfloor < g^{k+1}_{(j-1)j}(l+1)$, since the alternating inequalities in Lemma 3.3 imply that $g^{k+1}_{(j-1)j}(l-1) \leq g^k_{(j-1)j}(l-1) < g^{k+1}_{(j-1)j}(l) \leq g^{k+1}_{(j-1)j}(l+1)$. Thus we have that if $Y'_{j-1} = l$ then X'_{j-1} is equal to l or $l+1$. In other words,

$$\binom{X'_{j-1}}{Y'_{j-1}} \in \left\{ \binom{0}{0}, \binom{1}{0}, \binom{1}{1}, \ldots, \binom{k}{k}, \binom{k+1}{k} \right\}$$

and hence $X'_{j-1} \geq Y'_{j-1}$ holds in all cases. Accordingly, we have that $h_{X'}(\lfloor\lambda\rfloor) \geq h_{Y'}(\lfloor\lambda\rfloor)$.

Case 3: Consider the case that $\lfloor\lambda\rfloor = j+1$. We can show $h_{X'}(\lfloor\lambda\rfloor) \geq h_{Y'}(\lfloor\lambda\rfloor)$ in a similar way to Case 2.

For any pair of states X, Y satisfying $X \succeq Y$, it is easy to see that there exists a sequence of states $Z_1, Z_2, \ldots, Z_r$ satisfying $X = Z_1 \cdot\!\succ Z_2 \cdot\!\succ \cdots \cdot\!\succ Z_r = Y$. Then applying the above claim repeatedly, we obtain that $\phi(X, \lambda) = \phi(Z_1, \lambda) \succeq \phi(Z_2, \lambda) \succeq \cdots \succeq \phi(Z_r, \lambda) = \phi(Y, \lambda)$. ☐

Since $\mathcal{M}_{\mathrm{P}}(N; \boldsymbol{c})$ is a monotone chain, we can design a perfect sampler based on monotone CFTP. We could also employ Wilson's read once algorithm [27] and Fill's interruptible algorithm [9, 10], each of which also gives a perfect sampler.

4.3 Expected running time

Here, we assume the following condition, which leads that our perfect sampling algorithm terminates in an expected polynomial time.

Condition 1 *For all* $i \in \{1, 2, \ldots, n-1\}$,

$$\sum_{l=0}^{k} \left(g^k_{i(i+1)}(l) - g^{k+1}_{i(i+1)}(l) \right) \geq \frac{1}{2}$$

hold for any $k \in \{0, 1, \ldots, K\}$.

As a particular case, when we deal with a closed Jackson network with single servers model (i.e., $s_i = 1$ for each $i \in \{1, \ldots, n\}$), we can assume Condition 1 by arranging indices of servers to satisfy $\theta_i/\mu_i \geq \theta_{i+1}/\mu_{i+1}$ $(\forall i \in \{1, \ldots, n-1\})$.

Theorem 4.3 *Under Condition 1, the expectation of coalescence time* $T_* \in \mathbb{Z}_{++}$ *of* $\mathcal{M}_{\mathrm{P}}(N; \boldsymbol{c})$ *defined by* $T_* \stackrel{\text{def.}}{=} \min\{t > 0 \mid \exists y \in \Delta,\ \forall x \in \Delta,\ y = \Phi^0_{-t}(x, \boldsymbol{\Lambda})\}$ *is bounded* $\mathrm{O}(n^3 \ln N)$. *The whole number of transitions in our perfect sampler is also bounded by* $\mathrm{O}(n^3 \ln N)$.

Note that the coalescence time T_* is a random variable.

Proof: Let $\mathcal{H} = (\Delta, \mathcal{E})$ be the graph defined in the proof of Theorem 3.2; i.e., $\mathcal{H}$ is an undirected simple and connected graph with vertex set Δ and edge set $\mathcal{E}$ defined as follows. A pair of vertices $\{X, Y\}$ is an edge if and only if $(1/2)\sum_{i=1}^{n} |X_i - Y_i| = 1$. For each edge $e = \{X, Y\} \in \mathcal{E}$, there exists a unique pair of indices $j_1, j_2 \in \{1, \ldots, n\}$, called the *supporting pair* of e, satisfying

$$|X_i - Y_i| = \begin{cases} 1 & (i = j_1, j_2), \\ 0 & (\text{otherwise}). \end{cases}$$

We define the length $l_{\mathrm{P}}(e)$ of an edge $e = \{X, Y\} \in \mathcal{E}$ by $l_{\mathrm{P}}(e) \stackrel{\text{def.}}{=} (1/(n-1))\sum_{i=1}^{j^*-1}(n-i)$ where $j^* = \max\{j_1, j_2\} \geq 2$ and $\{j_1, j_2\}$ is the supporting pair of e. Note that $1 \leq \min_{e \in \mathcal{E}} l_{\mathrm{P}}(e) \leq \max_{e \in \mathcal{E}} l_{\mathrm{P}}(e) \leq n/2$. For each pair $X, Y \in \Delta$, we define the distance $d_{\mathrm{P}}(X, Y)$ be the length of a shortest path between X and Y on $\mathcal{H}$. Clearly, the diameter of $\mathcal{H}$, i.e., $\max_{(X,Y) \in \Delta^2} d_{\mathrm{P}}(X, Y)$, is bounded by $Nn/2$, since $d_{\mathrm{P}}(X, Y) \leq (n/2)\sum_{i=1}^{n}(1/2)|X_i - Y_i| \leq (n/2)N$ for any $(X, Y) \in \Delta^2$. The definition of edge length implies that for any edge $\{X, Y\} \in \mathcal{E}$, $d_{\mathrm{P}}(X, Y) = l_{\mathrm{P}}(\{X, Y\})$.

Now we show that $\mathrm{E}[d_{\mathrm{P}}(\phi(X, \Lambda), \phi(Y, \Lambda))] \leq \left(1 - \frac{1}{n(n-1)^2}\right) \cdot d_{\mathrm{P}}(X, Y)$ for any pair $\{X, Y\} \in \mathcal{E}$. In the following, we denote the supporting pair of $\{X, Y\}$ by $\{j_1, j_2\}$, and $X' = \phi(X, \Lambda)$, $Y' = \phi(Y, \Lambda)$ for the convenience. Without loss of generality, we can assume that $j_1 < j_2$, and $X_{j_2} + 1 = Y_{j_2}$.

<u>Case 1:</u> When $\lfloor \Lambda \rfloor = j_2 - 1$, we will show that

$$\mathrm{E}[d_{\mathrm{P}}(X', Y') \mid \lfloor \Lambda \rfloor = j_2 - 1] \leq d_{\mathrm{P}}(X, Y) - (1/2)(n - j_2 + 1)/(n-1).$$

In case $j_1 = j_2 - 1$, $X' = Y'$ with conditional probability 1. Hence $d_{\mathrm{P}}(X', Y') = 0$. In the following, we consider the case $j_1 < j_2 - 1$. Put $k = X_{j_2-1} + X_{j_2}$ then $Y_{j_2-1} + Y_{j_2} = k + 1$ since $X_{j_2} + 1 = Y_{j_2}$. In the same argument of Case 2 in the proof of Theorem 4.2, Lemma 3.3 implies that when $l \in \{0, 1, \ldots, k\}$ satisfy $g^k_{(j-1)j}(l-1) \leq \Lambda - \lfloor \Lambda \rfloor < g^k_{(j-1)j}(l)$, then $X' = l$ and $Y' = l$ or $l + 1$, since $g^{k+1}_{(j-1)j}(l-1) \leq g^k_{(j-1)j}(l-1) \leq \Lambda - \lfloor \Lambda \rfloor < g^k_{(j-1)j}(l) \leq g^{k+1}_{(j-1)j}(l+1)$ hold. In other words,

$$\begin{pmatrix} X'_{j_2-1} \\ Y'_{j_2-1} \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ldots, \begin{pmatrix} k \\ k \end{pmatrix}, \begin{pmatrix} k \\ k+1 \end{pmatrix} \right\}$$

hold. If $X'_{j_2-1} = Y'_{j_2-1}$, the supporting pair of $\{X', Y'\}$ is $\{j_1, j_2\}$ and so $d_{\mathrm{P}}(X', Y') = d_{\mathrm{P}}(X, Y)$. If $X'_{j_2-1} \neq Y'_{j_2-1}$, the supporting pair of $\{X', Y'\}$ is $\{j_1, j_2 - 1\}$ and so $d_{\mathrm{P}}(X', Y') = d_{\mathrm{P}}(X, Y) - (n - j_2 + 1)/(n-1)$.

Condition 1 implies that

$$\begin{array}{rcll} \Pr\left[X'_{j_2-1} \neq Y'_{j_2-1} \mid \lfloor \Lambda \rfloor = j_2 - 1\right] & = & \sum_{l=0}^{k}\left(g^k_{i(i+1)}(l) - g^{k+1}_{i(i+1)}(l)\right) \geq 1/2, & \text{and,} \\ \Pr\left[X'_{j_2-1} = Y'_{j_2-1} \mid \lfloor \Lambda \rfloor = j_2 - 1\right] & & \leq 1/2. & \end{array}$$

Thus we obtain that

$$\begin{aligned} \mathrm{E}[d_{\mathrm{P}}(X', Y') | \lfloor \Lambda \rfloor = j_2 - 1] &\leq (1/2)d_{\mathrm{P}}(X, Y) + (1/2)(d_{\mathrm{P}}(X, Y) - (n - j_2 + 1)/(n-1)) \\ &= d_{\mathrm{P}}(X, Y) - (1/2)(n - j_2 + 1)/(n-1). \end{aligned}$$

<u>Case 2:</u> When $\lfloor \Lambda \rfloor = j_2$, we can show that $\mathrm{E}[d_{\mathrm{P}}(X', Y') | \lfloor \Lambda \rfloor = j_2] \leq d_{\mathrm{P}}(X, Y) + (1/2)(n - j_2)/(n-1)$ in a similar way to Case 1.

Case 3: When $\lfloor\Lambda\rfloor \neq j_2 - 1$ and $\lfloor\Lambda\rfloor \neq j_2$, it is easy to see that the supporting pair $\{j_1', j_2'\}$ of $\{X', Y'\}$ satisfies $j_2 = \max\{j_1', j_2'\}$. Thus $d_{\mathrm{P}}(X,Y) = d_{\mathrm{P}}(X',Y')$.

The probability of appearance of Case 1 is equal to $1/(n-1)$, and that of Case 2 is less than or equal to $1/(n-1)$. From the above,

$$\begin{aligned} \mathrm{E}[d_{\mathrm{P}}(X',Y')] &\leq d_{\mathrm{P}}(X,Y) - \frac{1}{n-1}\cdot\frac{1}{2}\cdot\frac{n-j_2+1}{n-1} + \frac{1}{n-1}\cdot\frac{1}{2}\cdot\frac{n-j_2}{n-1} = d_{\mathrm{P}}(X,Y) - \frac{1}{2(n-1)^2} \\ &\leq \left(1 - \frac{1}{2(n-1)^2}\cdot\frac{1}{\max_{\{X,Y\}\in\mathcal{E}}\{d_{\mathrm{P}}(X,Y)\}}\right) d_{\mathrm{P}}(X,Y) = \left(1 - \frac{1}{n(n-1)^2}\right) d_{\mathrm{P}}(X,Y) \end{aligned}$$

hold for any pair $(X,Y) \in \mathcal{E}$.

Next we estimate the expectation of the coalescence time T_* of $\mathcal{M}_{\mathrm{P}}$. Since the metric d_{P} is defined as the shortest path in the graph $\mathcal{H}$, we also obtain

$$d_{\mathrm{P}}(X',Y') \leq \left(1 - \tfrac{1}{n(n-1)^2}\right) d_{\mathrm{P}}(X,Y) \qquad (\forall X,Y \in \Delta)$$

hold with the analogous argument of path coupling [4]. We define $D \stackrel{\text{def.}}{=} d(\boldsymbol{x}_{\max}, \boldsymbol{x}_{\min})$ and $\tau_0 \stackrel{\text{def.}}{=} n(n-1)^2(1+\ln D)$, then we have

$$\begin{aligned} \Pr[T_* > \tau_0] &= \Pr\left[\Phi_{-\tau_0}^0(\boldsymbol{x}_{\max}, \boldsymbol{\Lambda}) \neq \Phi_{-\tau_0}^0(\boldsymbol{x}_{\min}, \boldsymbol{\Lambda})\right] = \Pr\left[\Phi_0^{\tau_0}(\boldsymbol{x}_{\max}, \boldsymbol{\Lambda}) \neq \Phi_0^{\tau_0}(\boldsymbol{x}_{\min}, \boldsymbol{\Lambda})\right] \\ &\leq \textstyle\sum_{(X,Y)\in\Delta^2} d_{\mathrm{P}}(X,Y)\Pr\left[X = \Phi_0^{\tau_0}(\boldsymbol{x}_{\max}, \boldsymbol{\Lambda}), Y = \Phi_0^{\tau_0}(\boldsymbol{x}_{\min}, \boldsymbol{\Lambda})\right] \\ &= \mathrm{E}\left[d_{\mathrm{P}}\left(\Phi_0^{\tau_0}(\boldsymbol{x}_{\max}, \boldsymbol{\Lambda}), \Phi_0^{\tau_0}(\boldsymbol{x}_{\min}, \boldsymbol{\Lambda})\right)\right] \leq \left(1 - \frac{1}{n(n-1)^2}\right)^{\tau_0} d_{\mathrm{P}}(\boldsymbol{x}_{\max}, \boldsymbol{x}_{\min}) \\ &= \left(1 - \frac{1}{n(n-1)^2}\right)^{n(n-1)^2(1+\ln D)} D \leq \mathrm{e}^{-1}\mathrm{e}^{-\ln D} D = \frac{1}{\mathrm{e}}. \end{aligned}$$

The *submultiplicativity* of coalescence time ([25]) implies that $\Pr[T_* > k\tau_0] \leq (\Pr[T_* > \tau_0])^k \leq (1/\mathrm{e})^k$ for any $k \in \mathbb{Z}_+$. Thus

$$\begin{aligned} \mathrm{E}[T_*] &= \textstyle\sum_{t=0}^{\infty} t \cdot \Pr[T_* = t] \leq \tau_0 + \tau_0 \cdot \Pr[T_* > \tau_0] + \tau_0 \cdot \Pr[T_* > 2\tau_0] + \cdots \\ &\leq \tau_0 + \tau_0/\mathrm{e} + \tau_0/\mathrm{e}^2 + \cdots = \tau_0/(1 - 1/\mathrm{e}) \leq 2\tau_0. \end{aligned}$$

Clearly $D \leq Nn$, and thus we obtain the result about the expected coalescence time that

$$\mathrm{E}[T_*] \;\leq\; 2n(n-1)^2(1 + \ln(Nn)) \;=\; \mathrm{O}(n^3 \ln Nn).$$

Finally we estimate the whole number of transitions required in our sampling algorithm. Put $m = \lceil \log_2 T_* \rceil$. Algorithm 2 terminates when we set the starting time period $T = -2^m$ at $(m+1)$st iteration. Then the total number of simulated transitions is bounded by $2\cdot(2^0 + 2^1 + 2^2 + \cdots + 2^m) < 2\cdot 2\cdot 2^m \leq 8T_*$, since we need to execute two chains from both $\boldsymbol{x}_{\max}$ and $\boldsymbol{x}_{\min}$. Thus the expectation of total number of transitions of $\mathcal{M}_{\mathrm{P}}(N; \boldsymbol{c})$ is bounded by $\mathrm{O}(\mathrm{E}[8T_*]) = \mathrm{O}(n^3 \ln Nn)$. □

Corollary 4.4 *Under Condition 1, the mixing time of $\mathcal{M}_{\mathrm{P}}(N; \boldsymbol{c})$ is bounded by $n(n-1)^2 \ln(Nn)$.* ■

References

[1] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, in Séminarie de Probabilitiés XVII 1981/1982, vol. 986 of Springer-Verlag Lecture Notes in Mathematics, D. Dold and B. Eckmann, (eds.), Springer-Verlag, New York, 1983, 243–297.

[2] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacious, Open, closed and mixed networks of queues with different classes of customers, Journal of ACM, **22** (1975), 248–260. Springer-Verlag, New York, 2001.

[3] R. Bubley, Randomized Algorithms: Approximation, Generation, and Counting, Springer-Verlag, New York, 2001.

[4] R. Bubley and M. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS 1997), 223–231.

[5] J. P. Buzen, Computational algorithms for closed queueing networks with exponential servers, Communications of the ACM, **16** (1973), 527–531.

[6] X. Chao, M. Miyazawa, and M. Pinedo, Queueing Networks, Customers, Signals and Product Form Solutions, John Wiley & Sons, Inc, 1999.

[7] W. Chen and C. A. O'Cinneide, Towards a polynomial-time randomized algorithm for closed product-form networks, ACM Transactions on Modeling and Computer Simulation, **8** (1998), 227–253.

[8] M. Dyer and C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables, Theoretical Computer Sciences, **246** (2000), 265–278.

[9] J. Fill, An interruptible algorithm for perfect sampling via Markov chains, The Annals of Applied Probability, **8** (1998), 131–162.

[10] J. Fill, M. Machida, D. Murdoch, and J. Rosenthal, Extension of Fill's perfect rejection sampling algorithm to general chains, Random Structures and Algorithms, **17** (2000), 290–316.

[11] E. Gelenbe and G. Pujolle, Introduction to Queueing Networks, Second Edition, New York: John Wiley & Sons, Inc, 1998.

[12] W. J. Gordon and G. F. Newell, Closed queueing systems with exponential servers. Operations Research, **15**-2 (1967), 254–265.

[13] O. Häggström, Finite Markov Chains and Algorithmic Application, London Mathematical Society, Student Texts **52**, Cambridge University Press, 2002.

[14] J. R. Jackson, Networks of waiting lines, The Journal of Operations Research Society of America, **5** (1957), 518–521.

[15] J. R. Jackson, Jobshop-like queueing systems, Management Science, **10**-1 (1963), 131–142.

[16] M. Jerrum, Counting, Sampling and Integrating: Algorithms and Complexity, Lectures in Mathematics, ETH Zürich, Birkhauser Verlag, Basel, 2003.

[17] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo method: an approach to approximate counting and integration, in Approximation Algorithm for NP-hard Problems, D. Hochbaum, (ed.), PWS, 1996, 482–520.

[18] S. Kijima and T. Matsui, Approximate counting scheme for $m \times n$ contingency tables , IEICE Transactions on Information and Systems, **E87-D** (2004), 308–314.

[19] S. Kijima and T. Matsui, Polynomial time perfect sampling algorithm for two-rowed contingency tables, Random Structures Algorithms, *to appear*.

[20] S. Kijima and T. Matsui, Approximate/perfect samplers for closed Jackson networks, Proceedings of the 2005 Winter Simulation Conference (2005), 862–868.

[21] T. Matsui, M. Motoki, and N. Kamatani, Polynomial time approximate sampler for discretized Dirichlet distribution, 14th International Symposium on Algorithms and Computation (ISAAC 2003), Lecture Notes in Computer Science, **2906** (2003), 676–685.

[22] T. Matsui, M. Motoki, N. Kamatani, and S. Kijima, Polynomial time approximate/perfect samplers for discretized Dirichlet distribution, METR 2006-09, Mathematical Engineering Technical Reports, University of Tokyo, 2006. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[23] M. Mitzenmacher and E. Upfal, Probability and Computing, Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, 2005.

[24] T. Ozawa, Perfect simulation of a closed Jackson network, The Operations Research Society of Japan Queueing Symposium, Hikone, Japan, 2004.

[25] J. Propp and D. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Structures Algorithms, **9** (1996), 223–252.

[26] K. W. Ross and J. Wang, Asymptotically optimal importance sampling for product-form queuing networks. ACM Transactions on Modeling and Computer Simulation, **3** (1993), 244–268.

[27] D. Wilson, How to couple from the past using a read-once source of randomness, Random Structures Algorithms, **16** (2000), 85–113.