A NEW APPROACH TO COMPUTING THE TRANSIENT-STATE PROBABILITIES IN TIME-INHOMOGENEOUS MARKOV CHAINS

Yoshiaki Inoue
Osaka University

(Received September 15, 2020; Revised October 15, 2021)

Abstract This paper considers the computation of the transient-state probabilities in time-inhomogenous continuous-time Markov chains. We first introduce a new class of time-inhomogenous Markov chains, which is closely related to the phase-type representation of non-negative probability distributions. We show that the introduced class of Markov chains covers a wide-class of time-inhomogenous Markov chains. We then develop a computational method of the transient-state probabilities in Markov chains of this class, which is an extension of the uniformization method in time-homogeneous Markov chains. The developed computational method has a remarkable feature that the time-discretization of the generator is not necessary, as opposed to previously known methods in the literature.

Keywords: Queue, time-inhomogeneous Markov chain, uniformization, algorithmic analysis, phase-type approximation

1. Introduction

A better understanding of congestion phenomena is of key importance for enhancing the quality of service facilities, such as hospitals, airports, restaurants, and event venues. In particular, the recent outbreak of infectious diseases has made it increasingly important for facility managers to control the flow of people so that they do not stay in one place. A common feature shared by many service facilities is the existence of a finite opening time, i.e., they accept customers at only certain periods of the day. This tends to introduce time-dependence in the arrival of customers, generating peak and off-peak periods.

The continuous-time time-inhomogeneous Markov chain is a basic mathematical tool used to formulate such situations. Suppose that the system state (typically the number of present customers) at time $t \in [0, T]$ is denoted by $X(t)$, where $T < \infty$ denotes a constant representing the end of acceptance period. The system state $X(t)$ takes its value on finite state space $\mathcal{M} := \{0, 1, \ldots, M-1\}$ and its dynamics is determined by a time-inhomogeneous infinitesimal generator $Q(t)$ ($t \in [0, T]$). Let $q_{i,j}(t) := [Q(t)]_{i,j}$ ($i, j \in \mathcal{M}$, $t \in [0, T]$) denote the $(i, j)$th element of $Q(t)$, and let $q_i(t) := -q_{i,i}(t) = \sum_{j \in \mathcal{M} \setminus \{i\}} q_{i,j}(t)$ ($i \in \mathcal{M}$, $t \in [0, T]$) denote the transition rate from state $i$ at time $t$. Throughout this paper, we assume that $Q(t)$ is Lebesgue measurable and it has bounded transition rates, i.e.,

$$q_{\sup} := \sup_{t \in [0,T]} \max_{i \in \mathcal{M}} q_i(t) < \infty. \tag{1.1}$$

The infinitesimal generator $Q(t)$ is thus Lebesgue integrable over $[0, T]$.

For time-inhomogeneous Markov chains, computation of the transient-state probabilities is of primary interest. Let $\pi_i(t) := \Pr(X(t) = i)$ ($i \in \mathcal{M}$) denote the probability that the chain is in state $i$ at time $t$. We define $\pi$ as the transient-state probability vector of
(X(t))_{t \in [0,T]}:
\pi(t) = (\pi_0(t), \pi_1(t), \ldots, \pi_{M-1}(t)).

Given the initial condition \( \pi(0) \) at time \( t = 0 \), \( \pi(t) \) \( t \in [0,T] \) is given by the solution of Kolmogorov’s differential equation
\[
\frac{d\pi(t)}{dt} = \pi(t)Q(t), \quad t \in [0,T].
\]

(1.2)

If the transition rates are time-homogeneous, i.e., if \( Q(t) = Q(0) \) holds for all \( t \in [0,T] \), the uniformization technique [12, Page 154] provides an efficient way to compute the transient probabilities \( \pi(t) \): in such a case, the solution of (1.2) is given explicitly by
\[
\pi(t) = \pi(0) \exp[Q(0)t],
\]

(1.3)

which is further rewritten as
\[
\pi(t) = \pi(0) \sum_{n=0}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!} \cdot (I + \theta^{-1}Q(0))^n,
\]

(1.4)

where \( I \) denotes a unit matrix and \( \theta \) denotes a non-negative number satisfying \( \theta \geq \max_{i \in M} q_i(0) \). The right-hand side of (1.4) is a probability-weighted average of stochastic matrices, so that it can be computed easily with high-precision.

Because the uniformization method relies on the matrix exponential representation (1.3) of \( \pi(t) \), its extension to a time-inhomogeneous case is not straightforward. Van Dijk [3] proposes an extension of the uniformization method to time-inhomogeneous Markov chains, which is developed based on the fact that the solution of (1.2) is represented as
\[
\pi(t) = \pi(0) \sum_{n=0}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!} \cdot \mathcal{Z}_t \cdot \mathcal{P}(u_1) \mathcal{P}(u_2) \cdots \mathcal{P}(u_n),
\]

(1.5)

where \( \theta \) denotes a non-negative number satisfying
\[
\theta \geq q_{\text{sup}},
\]

(1.6)

and \( \mathcal{P}(t) \) denotes a stochastic matrix defined as
\[
\mathcal{P}(t) = I + \theta^{-1}Q(t), \quad t \in [0,T].
\]

(1.7)

Computing \( \pi(t) \) directly using the expression (1.5) requires a time-discretization step, i.e., an approximation that \( \mathcal{P}(t) \) takes a constant value in \([n\Delta, (n + 1)\Delta) \) \( n = 0, 1, \ldots \) for some small \( \Delta > 0 \). An efficient implementation of such a computational procedure is further studied by Van Moorsel and Wolter [9]. Arns et al. [1] present another approach to extending the uniformization method to time-inhomogeneous Markov chains, which is based on an observation that occurrences of state-transitions are regarded to be governed by a time-inhomogeneous Poisson process. They also provide a detailed discussion about the effect of the time-discretization on the accuracy of computed transient-state probabilities.

Due to the necessity for the time-discretization step, the uniformization methods proposed in the previous works are essentially similar to approximating \((X(t))_{t \in [0,T]}\) by a time-inhomogeneous discrete-time Markov chain \((\hat{X}_n)_{n \in [0,1,\ldots,T/\Delta]}\), for which the transient-state...
probabilities can be computed readily. This reflects an essential difficulty of directly handling the time-inhomogeneity for continuous-time Markov chains in numerical computation. To the best of our knowledge, no methods for computing the transient-state probabilities that does not require the time-discretization is reported in the literature.

The main purpose of this paper is to introduce a new approach to the computation of the transient-state probabilities in time-inhomogeneous continuous-time Markov chains. The approach taken in this paper is closely related to the use of phase-type distributions for modeling non-negative probability distributions. Specifically, we introduce a special class of time-inhomogeneous continuous-time Markov chains, whose infinitesimal generators are represented in a form related to the phase-type approximations. For this class of Markov chains, we show that the transient-state probabilities are computable without the time-discretization of $P(t)$ (or equivalently $Q(t)$). The key observation leading to this result is that for the introduced class of Markov chains, the multiple integrals on the right-hand side of (1.5) are rewritten as probability-weighted averages of stochastic matrices, so that a high-precision computational procedure is readily constructed without the necessity of the time-discretization. Related to the versatility of phase-type distributions, the introduced class is also shown to cover a wide range of time-inhomogeneous Markov chains, which shows their usefulness as a modeling tool.

We note that there also exist other approaches to time-inhomogeneous continuous-time Markov chains, which are mainly aimed at those defined on the infinite time-horizon [4–8, 14, 15]. Ergodic properties and the rate of convergence in time-inhomogeneous birth-and-death processes are studied by Zeifman [15] and Granovsky and Zeifman [4]. The point-wise stationary approximation (PSA) for time-inhomogeneous birth-and-death processes is studied by Green and Kolesar [5] and Whitt [14], which approximates the long-run average of the transient probability distribution as a weighted average of stationary distributions calculated from instantaneous transition rates at each time instant. As an extension of the PSA, Massey and Whitt [7] further consider a technique called the uniform acceleration asymptotic expansion, which consists of solving Poisson’s equations iteratively and approximating the transient probability distribution by the sum of their solutions. Also, using functional law of large numbers and functional central limit theorem, process-approximation methods for time-inhomogeneous Markov chains arising from queueing-network models are studied by Mandelbaum et al. [6]. Massey and Pender [8] also consider a process-approximation for time-inhomogeneous queueing systems with abandonment based on estimating the mean, variance, and third cumulant moment. The distinguishing feature of the present paper from the related literature is that we introduce a versatile class of time-inhomogeneous Markov chains defined on a finite time-interval $[0, T]$, for which the transient probabilities can be computed directly without approximations or scaling limits.

The rest of this paper is organized as follows. In section 2, we introduce the special class of time-inhomogeneous continuous-time Markov chains and show its versatility. We then develop a computational method for Markov chains of this class in Section 3 and we present numerical examples in Section 4. Finally, we conclude this paper in Section 5.
2. A Special Class of Time-Inhomogeneous Markov Chains

2.1. Definition

In this paper, we consider time-inhomogeneous continuous-time Markov chains whose infinitesimal generator \( Q(t) \) is represented in the following form:

\[
Q(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot Q^{[m]}, \quad t \in [0,T],
\]

(2.1)

where \( \gamma > 0 \) denotes a positive real number and \( Q^{[m]} \) \( (m = 0, 1, \ldots) \) denotes the infinitesimal generator of a time-homogeneous Markov chain (defined on the state space \( \mathcal{M} \)) which satisfies

\[
\hat{q}_{\text{sup}} := \sup_{\mathcal{M}} \left\{ \max_{i \in \mathcal{M}} [-Q^{[m]}]_{i,i}; \ m = 0, 1, \ldots \right\} < \infty.
\]

(2.2)

By definition, we have \( Q^{[m]} e = 0 \) \( (m = 0, 1, \ldots) \), where \( e \) denotes a column vector of ones. In addition, we can verify that \( q_{\text{sup}} \leq \hat{q}_{\text{sup}} \), so that (1.1) is satisfied.

Infinitesimal generators with the infinite-series representation (2.1) form a versatile class in the set of time-inhomogeneous continuous-time Markov chains, as phase-type distributions do in the set of non-negative probability distributions. To see this, let \( D_{[0,T]} \) denote the set of non-negative bounded functions with domain \([0,T]\) such that each \( \phi \in D_{[0,T]} \) is represented as

\[
\phi(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot \hat{c}^{[m]}, \quad t \in [0,T],
\]

(2.3)

for some \( \gamma > 0 \) and \( \hat{c}^{[m]} \geq 0 \) \( (m = 0, 1, \ldots) \) with \( \sup \{ \hat{c}^{[m]} ; m = 0, 1, \ldots \} < \infty \). Note that the infinite series in (2.3) is always convergent under this bounding condition. By definition, an infinitesimal generator takes the form (2.1) if, and only if all of its non-diagonal elements belong to \( D_{[0,T]} \). To see this, note first that if a function \( \phi_0(t) \in D_{[0,T]} \) is given in terms of \( \gamma_0 \geq 0 \) and \( \hat{c}_0^{[m]} \geq 0 \) \( (m = 0, 1, \ldots) \) by

\[
\phi_0(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma_0 t](\gamma_0 t)^m}{m!} \cdot \hat{c}_0^{[m]}, \quad t \in [0,T],
\]

then we have for any \( \gamma > \gamma_0 \),

\[
\phi_0(t) = \sum_{i=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^i}{i!} \cdot \exp[(\gamma - \gamma_0)t] \cdot \hat{c}_0^{[i]}
\]

\[=
\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{\exp[-\gamma t](\gamma_0 t)^i}{i!} \cdot \frac{(\gamma - \gamma_0)^j t^j}{j!} \cdot \hat{c}_0^{[i]}
\]

\[=
\sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} \cdot \frac{(\gamma_0 t)^i (\gamma - \gamma_0)^{m-i}}{\gamma^m} \cdot \hat{c}_0^{[i]}
\]

\[=
\sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot \hat{c}_0^{[m]}, \quad t \in [0,T],
\]

(2.4)
Therefore, if non-diagonal elements $q_{i,j}(t)$ of $Q(t)$ all belong to $D_{[0,T]}$, then they are written in terms of the same parameter $\gamma$ as

$$q_{i,j}(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t] \gamma^m}{m!} \cdot \hat{c}_i^m, \quad i, j \in \mathcal{M}, i \neq j,$$

which in turn implies that $Q(t)$ satisfies (2.1) with $[Q^m]_{i,j} = q_{i,j}^m$ $(i \neq j)$ and $[Q^m]_{i,i} = -\sum_{j \in \mathcal{M} \setminus \{i\}} q_{i,j}$. The converse is immediate because if $Q(t)$ is given by (2.1), we have

$$[Q]_{i,j}(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t] \gamma^m}{m!} \cdot [Q^m]_{i,j} \in D_{[0,T]}.$$

### 2.2. Basic Properties of $D_{[0,T]}$

We see that $D_{[0,T]}$ contains functions of the following types at least:

(a) Constant functions $\phi(x) = c$ $(x \in [0, T], c \geq 0)$.

(b) The cumulative distribution functions (CDFs), the complementary CDFs, and the probability density functions (PDFs) of phase-type distributions.

(a) is readily verified by substituting $c_{[m]} = c$ in (2.3):

$$\phi(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t] \gamma^m}{m!} \cdot c = c, \quad t \in [0, T]. \quad (2.5)$$

Also, (b) follows from the fact that we have for any phase-type random variable $Y$ with representation $(\eta, S)$,

$$Pr(Y > t) = \eta \exp[S\eta t] e = \sum_{m=0}^{\infty} \frac{\exp[-\zeta t] \zeta^m}{m!} \cdot \eta (I + \zeta^{-1} S)^m e,$$

$$Pr(Y \leq t) = 1 - Pr(Y > t) = \sum_{m=0}^{\infty} \frac{\exp[-\zeta t] \zeta^m}{m!} \cdot (1 - \eta (I + \zeta^{-1} S)^m e),$$

$$\frac{d Pr(Y \leq t)}{dt} = \sum_{m=0}^{\infty} \frac{\exp[-\zeta t] \zeta^m}{m!} \cdot \eta (I + \zeta^{-1} S)^m (-S) e, \quad (2.6)$$

where $\zeta > 0$ denotes a constant greater than or equal to the maximum absolute value of diagonal elements of $S$.

In addition, $D_{[0,T]}$ has the following closure properties:

**Lemma 2.1.** (i) $D_{[0,T]}$ is closed under addition:

$$\phi_1, \phi_2 \in D_{[0,T]} \Rightarrow \phi_1 + \phi_2 \in D_{[0,T]}.$$

(ii) $D_{[0,T]}$ is closed under (pointwise) multiplication:

$$\phi_1, \phi_2 \in D_{[0,T]} \Rightarrow \phi_1 \phi_2 \in D_{[0,T]}.$$
(iii) $D_{[0,T]}$ is closed under convolution:

$$\phi_1, \phi_2 \in D_{[0,T]} \implies \phi_1 \ast \phi_2 \in D_{[0,T]}.$$  

(iv) $D_{[0,T]}$ is closed under integration over $[0,t]$:

$$\phi_1 \in D_{[0,T]} \implies \left( \int_0^t \phi_1(u) \, du; \ t \in [0,T] \right) \in D_{[0,T]}.$$  

(v) $D_{[0,T]}$ is closed under a complement:

$$\phi_1 \in D_{[0,T]} \implies c - \phi_1 \in D_{[0,T]},$$

where $c \geq \sup \{c^m; \ m = 0, 1, \ldots \}$.  

**Proof.** Without loss of generality, we assume that $\phi_1(t)$ and $\phi_2(t)$ ($t \in [0,T]$) are given in terms of a common rate parameter $\gamma > 0$ by (cf. (2.4))

$$\phi_i(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot c_i^m, \quad i \in \{1, 2\},$$

(2.7)

where $c_i^m \geq 0$ ($i \in \{1, 2\}$, $m = 0, 1, \ldots$). Lemma 2.1 (i) then immediately follows from (2.7). Also, Lemma 2.1 (v) is readily verified from (2.5) and (2.7). Furthermore, we have Lemma 2.1 (ii) and (iii) from

$$\phi_1(t)\phi_2(t) = \sum_{m=0}^{\infty} \frac{\exp[-2\gamma t](2\gamma t)^m}{m!} \sum_{m_1=0}^{m} \binom{m}{m_1} \left(\frac{1}{2}\right)^m \frac{c_1^{m_1} c_2^{m-m_1}}{\gamma},$$

$$\phi_1 \ast \phi_2(t) = \sum_{m=1}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \sum_{m_1=0}^{m-1} \frac{c_1^{m_1} c_2^{m-m_1-1}}{\gamma}.$$  

Finally, we obtain Lemma 2.1 (iv) by letting $\phi_2(t) = 1$ ($t \in [0,T]$) in (iii).  

**Remark 1.** For $\phi_1(t) \in D_{[0,T]}$, its derivative is given by

$$\phi_1'(t) := \frac{d}{dt} \phi_1(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot \gamma (c_1^{m+1} - c_1^m),$$

so that $\phi'(t) \in D_{[0,T]}$ if and only if $(c_1^m)_{m=0,1,\ldots}$ is non-decreasing.

From the closure properties shown above and the denseness of phase-type distributions in the set of all non-negative probability distributions [10, Page 242], we can see that $D_{[0,T]}$ covers a wide-range of non-negative functions defined on a finite interval $[0,T]$. In particular, any non-negative continuous function defined on $[0,T]$ can be approximated arbitrarily closely by a function in $D_{[0,T]}$, which is formally stated as follows:

**Theorem 2.1.** For any non-negative continuous function $f$ defined on $[0,T]$, there exists a sequence of functions $\phi_0 \in D_{[0,T]}, \phi_1 \in D_{[0,T]}, \ldots$ such that

$$f(t) = \lim_{n \to \infty} \phi_n(t), \quad t \in (0,T).$$
Proof. Since \( f(t) \) is continuous on the finite closed interval \([0,T]\), it is integrable over \([0,t] \subseteq [0,T]\). We then define \( F(t) \) \((t \in [0,T])\) as

\[
F(t) = \frac{\int_0^t f(x)dx}{\int_0^T f(x)dx}.
\]

(2.8)

By definition, \( F(t) \) is a continuous non-decreasing function bounded by 0 and 1. It thus represents a CDF of a non-negative random variable. Therefore, from the denseness of phase-type distributions, there exists a sequence of phase-type CDFs \( G_0, G_1, \ldots \) such that

\[
F(t) = \lim_{n \to \infty} G_n(t), \quad t \in [0,T].
\]

(2.9)

Let \( g_n(t) \) denote the PDF corresponding to \( G_n(t) \):

\[
G_n(t) = \int_0^t g_n(x)dx, \quad t \in [0,T].
\]

(2.10)

Note that \( g_n \in \mathcal{D}_{[0,T]} \) holds because the PDFs of phase-type distributions belong to \( \mathcal{D}_{[0,T]} \), as we have shown in (2.6). We further define \( \phi_n(t) \) as

\[
\phi_n(t) = g_n(t) \int_0^T f(x)dx,
\]

(2.11)

which obviously satisfies \( \phi_n \in \mathcal{D}_{[0,T]} \).

It then follows from (2.8), (2.9), (2.10), and (2.11) that

\[
\int_0^t f(x)dx = \left( \int_0^T f(x)dx \right) \cdot \lim_{n \to \infty} \int_0^t g_n(x)dx
\]

\[
= \lim_{n \to \infty} \int_0^t \phi_n(x)dx
\]

\[
= \int_0^t \lim_{n \to \infty} \phi_n(x)dx,
\]

where the last equality follows from the bounded convergence theorem. We thus obtain from the fundamental theorem of calculus,

\[
f(t) = \lim_{n \to \infty} \phi_n(t), \quad t \in [0,T].
\]

As mentioned above, an infinitesimal generator \( Q(t) \) takes the form (2.1) if and only if its non-diagonal elements are in \( \mathcal{D}_{[0,T]} \). Therefore, Theorem 2.1 suggests that infinitesimal generators of the form (2.1) define a versatile class of time-inhomogenous continuous-time Markov chains on the finite time-interval \([0,T]\).

3. Computation of Transient-State Probabilities

In this section, we develop a computational method for the transient-state probability vector \( \pi(t) \) \((t \in [0,T])\), under the assumption that \( Q(t) \) is given by (2.1). We choose \( \theta = \hat{q}_{\text{sup}} \) (cf.
(2.2) because it clearly satisfies (1.6). Let $P^{[m]} := I + \theta^{-1}Q^{[m]}$ ($m = 0, 1, \ldots$). It then follows from (1.7) that

$$P(t) = \sum_{m=0}^{\infty} \frac{\exp[-\gamma t](\gamma t)^m}{m!} \cdot P^{[m]}, \quad t \in [0, T].$$

For $t \in [0, T]$, $n = 0, 1, \ldots$, we define $B_n(t)$ as (cf. (1.5))

$$B_0(t) = I,$$

$$B_n(t) = \int_0^t \int_0^{u_1} \cdots \int_0^{u_{n-1}} P(u_1)P(u_2) \cdots P(u_n)du_n \cdots du_1, \quad n = 1, 2, \ldots.$$  

We can verify from (3.4) that $B_1(t), B_2(t), \ldots$ ($t \in [0, T]$) are determined by (3.2) and the following recursion:

$$B_n(t) = \int_0^t B_{n-1}(u)P(u)du, \quad n = 1, 2, \ldots.$$  

For $n = 1, 2, \ldots$ and $m = 0, 1, \ldots$, we define $\alpha_{n,m}(k)$ and $\beta_{n,m}(k)$ as

$$\alpha_{n,m}(k) = \binom{n + k}{n}/\binom{n + m + 1}{n + 1}, \quad k = 0, 1, \ldots, m,$$

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

By definition, $\beta_{n,m}(k)$ denotes the probability function of a binomial distribution with number of trials $m$ and success probability $n/(n+1)$. Also, we can verify that $\sum_{k=0}^{m} \alpha_{n,m}(k) = 1$, so that $\alpha_{n,m}(k)$ represents the probability function of a discrete distribution with parameters $n$ and $m$.

**Lemma 3.1.** $B_n(t)$ ($t \in [0, T], n = 1, 2, \ldots$) is given by

$$B_n(t) = \frac{t^n}{n!} \sum_{m=0}^{\infty} \frac{\exp[-n\gamma t](n\gamma t)^m}{m!} \cdot B_n^{[m]},$$

where $B_n^{[m]}$ ($n = 1, 2, \ldots, m = 0, 1, \ldots$) denotes a matrix determined recursively by

$$B_1^{[m]} = \frac{1}{m+1} \sum_{k=0}^{m} P^{[k]}, \quad m = 0, 1, \ldots,$$

$$B_n^{[m]} = \sum_{k=0}^{m} \alpha_{n-1,m}(k)B_{n-1}^{[k]}, \quad n = 2, 3, \ldots, m = 0, 1, \ldots,$$

$$B_n^{[k]} = \sum_{\ell=0}^{k} \beta_{n,k}(\ell)B_n^{[\ell]}P^{[k-\ell]}, \quad n = 1, 2, \ldots, k = 0, 1, \ldots, m.$$  

**Remark 2.** It is readily verified that $B_n^{[m]}$ ($n = 1, 2, \ldots, m = 0, 1, \ldots$) is a stochastic matrix, i.e., its elements are all non-negative and $B_n^{[m]} e = e$ holds.
The proof of Lemma 3.1 is provided in A.

The following theorem then immediately follows from (1.5), (3.3), and Lemma 3.1.

**Theorem 3.1.** For a time-inhomogeneous continuous-time Markov chain with infinitesimal generator \( Q(t) \) of the form (2.1), the transient-state probability vector \( \pi(t) \) \( (t \in [0, T]) \) is given by

\[
\pi(t) = \pi(0) \sum_{n=0}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{\infty} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!} B_n^{[m]},
\]

(3.12)

where \( B_0^{[m]} := I \) \( (m = 0, 1, \ldots) \) and \( B_n^{[m]} \) \( (n = 1, 2, \ldots, m = 0, 1, \ldots) \) is defined as in Lemma 3.1.

It is worth noting that most steps in the proof of Lemma 3.1 can be also carried out for various types of series-expansions of \( P(t) \) different from (3.1), e.g., Taylor series and Fourier series expansions. However, the key advantage of the expression (3.1) with Poisson weighted sum is that it leads to the formula (3.12) suitable for numerical computation. More specifically, (3.12) shows that the transient-state probabilities are given in terms of the probability-weighted average of stochastic matrices \( B_n^{[m]} \) (cf. Remark 2), where the computation of \( B_n^{[m]} \) consists of only additions and multiplications of non-negative numbers (cf. (3.9), (3.10), (3.11)). Therefore, the right-hand side of (3.12) can be computed readily without suffering from the loss of significance in floating-point numbers due to calculations involving negative numbers. If we were employed other types of expressions like Taylor or Fourier series expansions, it would result in a similar formula to (3.12) but with negative coefficients existing, which may lead to numerical instability.

With the formula (3.12), computation of \( \pi(t) \) \( (t \in [0, T]) \) is in fact quite straightforward. Once a truncation point \( (n^*, m^*) \) for the double infinite sums in (3.12) is determined, we only need to compute \( B_n^{[m]} \) \( (n = 0, 1, \ldots, n^*, m = 0, 1, \ldots, m^*) \) using the recursion in Lemma 3.1. Furthermore, we can determine the truncation point in the following way. Let \( \pi_{n^*,m^*}(t) \) denote the truncation approximation to (3.12) using a truncation point \( (n^*, m^*) \):

\[
\pi_{n^*,m^*}(t) = \pi(0) \sum_{n=0}^{n^*} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{m^*} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!} B_n^{[m]}, \quad t \in [0, T].
\]

(3.13)

Let \( \|\cdot\|_\infty \) denote the supremum norm of vectors, i.e., \( \|(x_0, x_1, \ldots, x_{M-1})\|_\infty = \max\{x_i; i = 0, 1, \ldots, M-1\} \). We have from (3.13),

\[
\|\pi_{n^*,m^*}(t) - \pi(t)\|_\infty \leq 1 - \sum_{n=0}^{n^*} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{m^*} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!}
\]

\[
= \sum_{n=0}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{\infty} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!} - \sum_{n=0}^{n^*} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{m^*} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!}
\]

\[
= \sum_{n=0}^{n^*} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{m^*} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!} + \sum_{n=n^*+1}^{\infty} \frac{\exp[-\theta t](\theta t)^n}{n!}
\]

\[
- \sum_{n=0}^{n^*} \frac{\exp[-\theta t](\theta t)^n}{n!} \sum_{m=0}^{m^*} \frac{\exp[-n \gamma t](n \gamma t)^m}{m!}.
\]
\[
= \sum_{n=n^*+1}^{\infty} \frac{\exp[\theta t](\theta t)^n}{n!} + \sum_{n=0}^{n^*} \frac{\exp[\theta t](\theta t)^n}{n!} \sum_{m=m^*+1}^{\infty} \frac{\exp[-n\gamma t](n\gamma t)^m}{m!}.
\]

Note that the Poisson distribution is stochastically increasing with its mean \[11, \text{Example 8.A.2}\], i.e.,
\[
\sum_{k=K}^{\infty} \frac{\exp[-\zeta_1 k^k]}{k!} \leq \sum_{k=K}^{\infty} \frac{\exp[-\zeta_2 k^k]}{k!}, \quad \zeta_1 \leq \zeta_2, \quad K = 0, 1, \ldots.
\]

It then follows from (3.14) that
\[
\sup_{t \in [0,T]} \|\pi^{n^*,m^*}(t) - \pi(t)\|_\infty \leq \sum_{n=n^*+1}^{\infty} \frac{\exp[-\theta T](\theta T)^n}{n!} + \sum_{m=m^*+1}^{\infty} \frac{\exp[-n^*\gamma T](n^*\gamma T)^m}{m!}.
\]

Therefore, in order to ensure an error bound
\[
\sup_{t \in [0,T]} \|\pi^{n^*,m^*}(t) - \pi(t)\|_\infty \leq \epsilon,
\]
for a given \(\epsilon > 0\), it is sufficient to choose the truncation point \((n^*, m^*)\) satisfying
\[
\sum_{n=n^*+1}^{\infty} \frac{\exp[-\theta T](\theta T)^n}{n!} \leq \frac{\epsilon}{2}, \quad \sum_{m=m^*+1}^{\infty} \frac{\exp[-n^*\gamma T](n^*\gamma T)^m}{m!} \leq \frac{\epsilon}{2},
\]

or equivalently,
\[
\sum_{n=0}^{n^*} \frac{\exp[-\theta T](\theta T)^n}{n!} \geq 1 - \frac{\epsilon}{2}, \quad \sum_{m=0}^{m^*} \frac{\exp[-n^*\gamma T](n^*\gamma T)^m}{m!} \geq 1 - \frac{\epsilon}{2},
\]

which can be verified readily.

Remark 3. In computing \(B_n^{[m]}\) \((n = 0, 1, \ldots, n^*, \ m = 0, 1, \ldots, m^*)\), we need to compute \(\alpha_{n,m}(k)\) defined in (3.6). For each \(n = 1, 2, \ldots\) and \(m = 0, 1, \ldots\), the probability function \(\alpha_{n,m}(k)\) is efficiently computed with
\[
\alpha_{n,m}(m) = \frac{n+1}{n+m+1}, \quad \alpha_{n,m}(k) = \frac{k+1}{n+k+1} \cdot \alpha_{n,m}(k+1), \quad k = 0, 1, \ldots, m-1.
\]

Note here that \(\alpha_{n,m}(m - \ell)\) is strictly decreasing in \(\ell\).

4. Numerical Examples

In this section, we present some numerical examples. We consider an \(M(t)/M/1/K\) queue with time-dependent arrival rate \(\lambda(t)\) \((t \in [0,T])\). We set the mean service rate \(\mu = 1\) and the system capacity \(K = 100\) throughout. Let \(L(t)\) \((t \in [0,T])\) denote the number of customers in the \(M(t)/M/1/K\) system at time \(t\). \((L(t))_{t \in [0,T]}\) forms a time-inhomogeneous Markov chain with infinitesimal generator
\[
Q(t) = Q_0 + \lambda(t)Q_1,
\]
where
\[ Q_0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \mu & -\mu & 0 & \cdots & 0 & 0 \\ 0 & \mu & -\mu & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\mu & 0 \\ 0 & 0 & 0 & \cdots & -\mu & -\mu \end{pmatrix}, \quad Q_1 = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \]

We consider three different arrival rate functions \( \lambda(t) \) shown in Figure 1, which have decreasing, unimodal, and bimodal shaped curve, respectively. These arrival rate functions are intended to imitate the arrival patterns in a day often observed in real service facilities:
- The decreasing arrival rate represents a situation where many customers arrive shortly after the opening, which is observed typically in facilities with large waiting times such as hospitals.
- The unimodal arrival rate represents a situation where the peak-time comes in the middle of the opening hours, which is often observed in facilities such as student/company canteens.
- The bimodal arrival rate represents a situation where two peak-times occur in a day. This pattern is observed in facilities such as municipal offices, where the first peak-time is reached in the morning, which is followed by a decrease in customer arrivals during the lunch break.

More specifically, the arrival rate functions in Figure 1 are given by the form (2.3) with common parameter \( \gamma = 0.2 \) and different sequences \( (c^{[m]})_{m=0,1,\ldots} \) given as follows:

**Decreasing:**
\[ c^{[m]} = c_{\text{dec}} \cdot 0.75^m + 0.7, \quad m = 0, 1, \ldots \]

**Unimodal:**
\[ c^{[m]} = c_{\text{uni}} \prod_{i=0}^{m-1} 1.3 \cdot 0.97^i, \quad m = 0, 1, \ldots \]

**Bimodal:**
\[ c^{[m]} = c_{\text{bi}} (f^{[m]}_1 + f^{[m]}_2), \quad m = 0, 1, \ldots, \]
\[ f^{[m]}_1 = \prod_{i=0}^{m-1} 1.5 \cdot 0.8^i, \quad m = 0, 1, \ldots, \]
\[ f^{[m]}_2 = \begin{cases} 0, & m = 0, 1, \ldots, 9, \\ \prod_{i=10}^{m-1} 1.5 \cdot 0.8^{i-10}, & m = 10, 11, \ldots. \end{cases} \]

The coefficients \( c_{\text{dec}}, c_{\text{uni}}, \) and \( c_{\text{bi}} \) are determined numerically so that the mean arrival rate (denoted by \( \bar{\lambda} \)) takes a common fixed value \( \bar{\lambda} = 1 \), where
\[
\bar{\lambda} := \frac{1}{T} \int_0^T \lambda(t) \, dt = \frac{1}{T} \sum_{m=0}^\infty \frac{1}{\gamma} \sum_{k=m+1}^\infty \frac{\exp[-\gamma T](\gamma T)^k}{k!} \cdot c^{[m]}.
\]
Approach to inhomogeneous Markov chains

\[
\lambda(t) = \sum_{k=0}^{\infty} \frac{\exp[-\gamma T](\gamma T)^k}{k!} \cdot \frac{1}{k+1} \sum_{m=0}^{k} c^{(m)}.
\]

It is readily verified that (4.1) is written in the form (2.1) under this condition, so that we can apply the computational method developed in the previous section. We use an error-bound \( \epsilon = 10^{-11} \) in the computation.

In Figure 2, we show the computed results of \( \mathbb{E}[L(t)] \) for the three different \( \lambda(t) \). For reference, we also plot approximate values of \( \mathbb{E}[L(t)] \) calculated using a discrete-time Markov chain with the transition probability matrix \( P_n = I + Q(t_n)h \), where \( h \) denotes the time-step size and \( t_n := nh \ (n = 1, 2, \ldots, \lfloor T/h \rfloor) \). We observe that as the number of discrete time points increases, the discrete approximation approaches to the value of \( \mathbb{E}[L(t)] \) computed with the method developed in this paper. We note that the loss rate \( \lambda_{\text{loss}}(t) := \lambda(t) \Pr(L(t) = K) \ (t \in [0,T]) \), which represents the expected number of lost customers per unit time, is sufficiently small \( (< 10^{-10}) \) in all cases, so that the transient behavior of the mean queue-length shown in Figure 2 can be regarded as that of the corresponding M(t)/M/1/\infty queue. Figure 3, Figure 4, and Figure 5 also show the comparison of the computed probability distribution of \( L(t) \) with the corresponding discrete approximations. From these figures, we observe that the computed distribution of \( L(t) \) is close to its discrete approximation with sufficiently large time steps. These results demonstrate the validity and feasibility of the computational method we have developed in this paper.

5. Concluding Remarks

In this paper, we introduced the class of time-inhomogeneous continuous-time Markov chains with the special representation (2.1) for infinitesimal generators. We showed that this class covers a wide range of time-inhomogeneous Markov chains, as the class of phase-type distributions does for non-negative probability distributions. We then developed a computational method for the transient-state probabilities in this class of Markov chains, which does not require time-discretization steps in the computation. Finally, we presented some numerical examples, which demonstrate the numerical feasibility of the developed computational method.

In this paper, we showed that infinitesimal generators of the form (2.1) define a versatile class of time-inhomogenous continuous-time Markov chains. In practice, one is required...
Figure 2: The mean number of customers in the system as a function of time $t$
Figure 3: The probability distribution of the number of customers in the system (Decreasing $\lambda(t)$)
Figure 4: The probability distribution of the number of customers in the system (Unimodal $\lambda(t)$)
Figure 5: The probability distribution of the number of customers in the system (Bimodal $\lambda(t)$)
to take a step of fitting \( Q(t) \) to some given data. It is usually the case that \( Q(t) \) can be decomposed into several matrices, each of which has time-dependency characterized by a one-dimensional function (cf. (4.1) and [1]):

\[
Q(t) = \sum_{\ell=1}^{L} f_{\ell}(t) A_{\ell},
\]

where \( L \in \{1, 2, \ldots \} \). The matrices \( A_{\ell} \) are determined by the system model being considered, while \( f_{\ell}(t) \) can be thought of as model parameters to be fitted from data. For example, if the system is modeled as a queue with time-dependent arrival and service rates (i.e., the \( M(t)/M(t)/c/K \) queue), then \( Q(t) \) takes the form (5.1) with \( f_{0}(t) \) and \( f_{1}(t) \) representing the arrival and service rates as functions of time \( t \). From this viewpoint, the fitting of \( Q(t) \) is reduced to that of individual components \( f_{\ell}(t) \), given the model structure \( A_{\ell} \).

A straightforward way to find a function \( f_{\ell} \in D_{[0,T]} \) approximating a given function \( g_{\ell}(t) \) would be to consider a cost functional like

\[
\int_{0}^{T} (f_{\ell}(t) - g_{\ell}(t))^{2}dt,
\]

and minimize it numerically with respect to \( \gamma \) and \( c^{[m]} \); because \( c^{[m]} \) \((m = 0, 1, \ldots)\) is an infinite sequence, we have to set \( c^{[m]} = 0 \) \((m \geq M)\) for some \( M \) or consider a parametric form of \( c^{[m]} \). Given the function \( g_{\ell}(t) \), we can evaluate the gradient of this cost functional with respect to \( \gamma \) and \( c^{[m]} \) numerically, so that standard gradient descent methods will work.

However, it is unlikely in practice that the exact form of the objective form \( g_{\ell}(t) \) of the function is available a priori. Instead, we need to perform the fitting of \( f_{\ell}(t) \) in a more ad hoc way. For example, the arrival rate function \( \lambda(t) \) is typically fitted from the set of arrival-time sequences \( \{(t_{d,1}, t_{d,2}, \ldots, t_{d,N_{d}}); d = 1, 2, \ldots, D\} \) recorded over \( D \) days. A possible approach to fitting \( \lambda(t) \) in this case is that we assume the time-dependent Poisson arrivals are generated from the mechanism that (i) the total number of arriving customers \( N \) is distributed according to a Poisson distribution and (ii) the arrival times of customers are independent and identically distributed according to a probability distribution \( F \) with support \([0, T]\). We can then approximate \( F \) by a phase-type distribution, via standard methods like the expectation-maximization (EM) algorithm [2, 13]. The resulting function is written as

\[
\lambda(t) = \Lambda \cdot \frac{dF(t)}{dt},
\]

where \( \Lambda \) equals \( \int_{0}^{T} \lambda(t)dt \) and it represents the expected number of arrivals in a day. As discussed in Section 2.2, this \( \lambda(t) \) belongs to \( D_{[0,T]} \), so that it can be used to construct an infinitesimal generator \( Q(t) \) of the form (2.1).

Acknowledgments
This research was supported in part by JSPS KAKENHI Grant Number JP18K18007.

A. Proof of Lemma 3.1
We prove Lemma 3.1 by induction. For \( n = 1 \), we have from (3.1) and (3.5),

\[
B_{1}(t) = \sum_{m=0}^{\infty} \int_{0}^{t} \frac{\exp[-\gamma u](\gamma u)^{m}}{m!}du \cdot P^{[m]}
\]
(3.8) thus follows for \( n = 1 \). We then assume that (3.8) holds for some \( n = j \) \((j = 1, 2, \ldots)\).

Under this assumption, it follows from (3.1) and (3.5) that

\[
B_{j+1}(t) = \frac{1}{j!} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} B_j^{[k_1]} P^{[k_2]} \cdot \int_0^t u^j \cdot \frac{\exp[-(j+1)\gamma u] j^{k_1} (\gamma u)^{k_1+k_2}}{k_1! k_2!} \, du
\]

\[
= \frac{1}{j!} \sum_{k=0}^{\infty} B_j^{[k]} \cdot \beta_{j,k}(k_1) \cdot \frac{(j+k)!}{k!((j+1)\gamma)^j} \cdot \int_0^t \frac{\exp[-(j+1)\gamma u] ((j+1)\gamma u)^{j+k}}{(j+k)!} \, du
\]

\[
= \sum_{k=0}^{\infty} \hat{B}_j^{[k]} \cdot \frac{(j+k)}{(j+1)\gamma} \cdot \frac{1}{j+1} \sum_{m=j+k+1}^{\infty} \frac{\exp[-(j+1)\gamma t] ((j+1)\gamma t)^m}{m!}
\]

\[
= t^{j+1} \sum_{k=0}^{\infty} \hat{B}_j^{[k]} \cdot \frac{\exp[-(j+1)\gamma t] ((j+1)\gamma t)^m}{m!(j+1)!} \alpha_{j,m}(k)
\]

so that we have (3.8) for \( n = j + 1 \), which proves that (3.8) holds for any \( n = 1, 2, \ldots \) \( \square \)

References

[1] M. Arns, P. Buchholz, and A. Panchenko: On the numerical analysis of inhomogeneous continuous-time Markov chains. *INFORMS Journal on Computing*, 22 (2010), 416–432.

[2] S. Asmussen, O. Nerman, and M. Olsson: Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23 (1996), 419–441.

[3] N.M. van Dijk: Uniformization for nonhomogeneous Markov chains. *Operations Research Letters*, 12 (1992), 283–291.

[4] B.L. Granovsky and A. Zeifman: Nonstationary queues: Estimation of the rate of convergence. *Queueing Systems*, 46 (2004), 363–388.

[5] L. Green and P. Kolesar: The pointwise stationary approximation for queues with nonstationary arrivals. *Management Science*, 37 (1991), 84–97.

[6] A. Mandelbaum, W.A. Massey, and M.I. Reiman, Strong approximations for Markovian service networks. *Queueing Systems*, 30 (1998), 149–201.

[7] W.A. Massey and W. Whitt: Uniform acceleration expansions for Markov chains with time-varying rates. *The Annals of Applied Probability*, 8 (1998), 1130–1155.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.
[8] W.A. Massey and J. Pender: Gaussian skewness approximation for dynamic rate multi-server queues with abandonment. *Queueing Systems, 75* (2013), 243–277.

[9] A.P.A. van Moorsel and K. Wolter: Numerical solution of non-homogeneous Markov processes through uniformization. *Proc. of the 12th European Simulation Multiconference*, (1998), 710–717.

[10] M.F. Neuts: *Structured Stochastic Matrices of M/G/1 Type and Their Applications* (Marcel Dekker, New York, 1989).

[11] M. Shaked and J.G. Shanthikumar: *Stochastic Orders* (Springer, New York, 2007).

[12] H.C. Tijms: *Stochastic Models, An Algorithmic Approach* (John Wiley & Sons, Chichester, 1994).

[13] A. Thummler, P. Buchholz, and M. Telek: A novel approach for phase-type fitting with the EM algorithm. *IEEE Transactions on Dependable and Secure Computing, 3* (2006), pp. 245–258.

[14] W. Whitt: The pointwise stationary approximation for M_t/M_t/s queues is asymptotically correct as the rates increase. *Management Science, 37*, 307–314.

[15] A.I. Zeifman: Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. *Stochastic Processes and their Applications, 95* (1995), 157–173.

Yoshiaki Inoue
Department of Information and Communications Technology
Graduate School of Engineering, Osaka University
Suita 565-0871, Japan
E-mail: yoshiaki@comm.eng.osaka-u.ac.jp