Queueing Systems manuscript No.
(will be inserted by the editor)

Analysis of Adaptive Queueing Policies via Adiabatic Approach

Leena Zacharias · **Thinh Nguyen** ·
Yevgeniy Kovchegov · **Kyle Bradford**

Received: date / Accepted: date

Abstract We introduce an adiabatic framework to study adaptive queueing policies. The adiabatic framework provides analytical tools for stability analysis of slowly changing systems that can be modeled as time inhomogeneous Markov chains. We apply this appproach to study an adaptive queueing policy, where service rate is adaptively changed based on the estimated arrival rates that tend to vary with time. This makes the packet distribution in the queue over time behave like a time inhomogeneous reversible Markov chain. Our results provide an upper bound on the time for an initial distribution of packets in the queue to converge to a stationary distribution corresponding to some pre-specified queueing policy. This approach is useful to analyse and design adaptive queueing policies and can be readily extended to any system that can be modeled as a time inhomogeneous reversible Markov chain. We provide simulations that confirm our theoretical results.

Keywords Queueing policy · Adiabatic time · Continuous time Markov chain

1 Introduction

A Markov chain is a finite or countable state space random process where the probability of any particular future behavior of the process, when its present state is known exactly, is not altered by additional knowledge concerning its past behavior [5]. If these probabilities, called transition probabilities, are independent of time, it is called a stationary or time homogeneous Markov chain or else, a nonstationary or time inhomogeneous Markov chain. A Markov chain can be discrete or continuous depending

This work was supported in part by the National Science Foundation under Grant 0845476 and Grant 0834775.

Leena Zacharias, Thinh Nguyen
School of Electrial Engineering and Computer Science, Oregon State University, Corvallis, Oregon-USA
E-mail: thinhq@eecs.oregonstate.edu

Yevgeniy Kovchegov, Kyle Bradford
Dept. of Mathematics, Oregon State University, Corvallis, Oregon-USA

on whether the time index set is discrete or continuous. The theory of Markov chains is extensively used in queueing theory [6] and consequently in communication networks [7], among other fields.

It is a well-known result that an irreducible, aperiodic Markov chain converges to a unique stationary distribution [3]. The time taken to converge to this distribution is of interest in applications like Monte Carlo Markov chain [3]. After a large number of steps the Markov chain must be "ε-close" according to some distance notion to its stationary distribution. This is the idea of Markov chain mixing and the time is quantified by mixing time [9]. In the case of time inhomogeneous Markov chains, the evolution must be slow enough such that the distribution is "ε-close" to the stationary distribution of the final transition matrix. This is the so-called adiabatic approach and this time is quantified by adiabatic time [8].

The adiabatic approach follows along the lines of the adiabatic evolution in quantum mechanics [10]. The quantum adiabatic theorem studies the evolution of a system from an initial Hamiltonian to a final Hamiltonian (operator corresponding to total energy of the system) and says that if the evolution is slow enough, the system will be "ε-close" in ℓ^2 norm to the ground state of the final Hamiltonian. We consider ℓ^1 norms in our study of Markov chains.

In this paper, we study the convergence, in adiabatic sense, of time inhomogeneous Markov chains, where the time inhomogeneity can be due to an underlying nonstationary process or uncertainties in measurement of an underlying stationary process. We apply the above convergence results to a Markovian queueing model where the time inhomogeneity arises due to the latter of the two conditions mentioned above.

1.1 Application to Queueing Models

A queue is typically defined by the arrival rate, service or departure rate, number of servers and buffer size. In a finite buffer size queue, we are interested in maintaining a distribution which is more biased towards smaller queue lengths, since otherwise we will have high blocking probabilities. Such a stable queue is achieved by keeping the departure rate strictly above the arrival rate. In a network situation for example, there might be other constraints also on the departure rate. If the departure rate is too large, it might cause congestion somewhere else in the network. Or in the case of multiple queues of different kinds of traffic, each of which has to satisfy some quality of service conditions, sending packets from just one queue will affect the other queues. In wireless networks, we might require to maintain such a departure rate and nothing more, due to power restrictions. Also, in a network of multiple such nodes we might have problems of collision in the wireless medium. Hence it becomes necessary to monitor the departure rate and keep it at such a level to keep our queue stable and at the same time achieve the desired network objectives.

We consider a queue in which packets arrive at a fixed but unknown rate and we use an estimate of this arrival rate to decide a queueing policy designed to ensure stable queues. In particular, we let the departure rate be always higher than the estimated arrival rate by a fixed multiplicative constant, anticipating that the estimate will be correct in the long-run. Since the estimated arrival rate changes and is more accurate with time, the departure rate also changes. The packets in the queue evolve according to a time inhomogeneous Markov chain dictated by this adaptive departure policy. We study the time required for the queue to reach this stable distribution using the above outlined adiabatic approach and under suitable estimation and departure policies for the unknown arrival rate.

1.2 Related Work

The adiabatic theorem in quantum mechanics was first stated by Born and Fock [1]. A version of the adiabatic theorem [10] considers two Hamiltonians and the evolution of the system from the initial to final Hamiltonian. The theorem states that for sufficiently large time, the final state of the system and the ground state of the final Hamiltonian will be ε-close in ℓ^2 norm. The lower bound on time was found to be inversely proportional to the cube of the least spectral gap of the Hamiltonian over all time.

The adiabatic theorem for Markov chains was studied by Kovchegov [8] where the adiabatic evolution was studied for discrete time and continuous time Markov chains. The linear evolution of a time inhomogeneous Markov chain from an initial to a final probability transition matrix was studied and the adiabatic time was found to be proportional to the square of the mixing time or inversely proportional to the square of the spectral gap of the final transition matrix. This result was generalized to a more general adiabatic dynamics in [2].

1.3 Overview

Section 2 gives the mathematical preliminaries including definitions and general results whcih we use in the analysis. Section 3 defines the evolution of continuous time Markov chains with bounded generators where the changes happen at fixed time intervals and gives an upper bound on the distance between its distribution with the stationary distribution at large enough time. In section 4, we apply the upper bound to an M/M/1/K queue with unknown but constant arrival rate. We also define an adaptive queueing policy where we estimate the arrival rate and decide the departure rate to ensure an eventually stable queue. The main result gives the upper bound on time needed for the queue to converge to the eventual stable distribution. Simulations are also provided to compare the distance predicted by our upper bounds to the actual distance.

2 Preliminaries

Now we look at the main definitions and some results which will be useful in the rest of the paper. A vector x is a column vector whose ith entry is denoted by $x(i)$ and transpose denoted by x^T. A matrix P is a square matrix whose (i,j)th entry is denoted by P_{ij}.

– We use total variation distance to measure the distance between two probability distributions.

Definition 1 (Total variation distance) For any two probability distributions ν and π on a finite state space Ω, we define

$$\|\nu-\pi\|_{TV} = \frac{1}{2}\sum_{i\in\Omega}|\nu(i)-\pi(i)|.$$

□

Note that the above function measures the distance on a scale of 0 to 1.

– Define the following scalar products and the corresponding norms.

Definition 2 Let π be a strictly positive probability distribution on a finite state space Ω and let $\ell^2(\pi)$ be the real vector space $\mathbb{R}^r$ with the following scalar product and norm,

$$\langle x,y\rangle_\pi = \sum_{i\in\Omega} x(i)y(i)\pi(i),$$

$$\|x\|_\pi = \left(\sum_{i\in\Omega} x(i)^2\pi(i)\right)^{\frac{1}{2}},$$

and let $\ell^2(\frac{1}{\pi})$ be the real vector space $\mathbb{R}^r$ with the following scalar product and norm,

$$\langle x,y\rangle_{\frac{1}{\pi}} = \sum_{i\in\Omega}\frac{x(i)y(i)}{\pi(i)},$$

$$\|x\|_{\frac{1}{\pi}} = \left(\sum_{i\in\Omega}\frac{x(i)^2}{\pi(i)}\right)^{\frac{1}{2}}.$$

□

– The following result is used to relate total variation distance to the above defined norm and is proved in Chapter 6, Theorem 3.2 of [3, p. 209].

Proposition 1 *For any two probability distributions* ν *and* π *(strictly positive),*

$$\|\nu-\pi\|_{TV} \le \frac{1}{2}\|\nu-\pi\|_{\frac{1}{\pi}}.$$

□

– **Definition 3 (Reversibility)** Let P be a transition matrix and π a strictly positive probability distribution on Ω. The pair (P,π) is reversible if the detailed balance equations

$$\pi(i)P_{ij} = \pi(j)P_{ji}$$

hold for all $i, j \in \Omega$. □

If P is irreducible, then π is the unique stationary distribution of P.

– The following result is from [3, p. 202], Chapter 6, Theorem 2.1.

Proposition 2 (Necessary and sufficient condition of reversibility) *The pair (P, π) is reversible if and only if*

$$P^* = D^{\frac{1}{2}} P D^{-\frac{1}{2}},$$

is a symmetric matrix where

$$D = diag\{\pi(1), \ldots, \pi(r)\}.$$

Moreover, a reversible matrix P has real eigenvalues with right eigenvectors orthonormal in $\ell^2(\pi)$ and left eigenvectors orthonormal in $\ell^2(\frac{1}{\pi})$. □

This proposition is used in the proof of the following result. See [3].

– The following result is from Chapter 6, Theorem 3.3 of [3, p. 209].

Proposition 3 *Let P be a reversible irreducible transition matrix on the finite state space Ω, with the stationary distribution π and second largest eigenvalue modulus $|\lambda_2(P)|$. Then for any probability distribution ν on Ω and for all $n \geq 1$*

$$\left\| \nu^T P^n - \pi^T \right\|_{\frac{1}{\pi}} \leq |\lambda_2(P)|^n \left\| \nu - \pi \right\|_{\frac{1}{\pi}}.$$

□

– For a continuous time Markov chain $\{X_t\}_{t \geq 0}$ on a finite state space Ω with a bounded generator matrix $Q = [Q_{ij}]_{i,j \in \Omega}$, and $q = \max_{i \in \Omega} \sum_{j: j \neq i} Q_{ij}$, the upper bound on the departure rates of all states, uniformization [11] gives transition probabilities to be

$$P(t) = \sum_{n=0}^{\infty} e^{-qt} \frac{(qt)^n}{n!} P^n = e^{Qt}, \tag{1}$$

where the matrix $P = I + \frac{1}{q} Q$. The matrix $P(t)$ denotes the transition probabilities in time t.

– The following result is used to bound the distance of the continuous chain in terms of a discrete one and is based on [3, p. 364]. The proof follows along similar lines and can be found in [12].

Proposition 4 *For a continuous time Markov chain on a finite state space Ω with generator matrix $Q = q(P - I)$ with reversible, irreducible P and stationary distribution π, for any probability distribution ν on Ω,*

$$\left\| \nu^T e^{Qt} - \pi^T \right\|_{\frac{1}{\pi}} \leq \left\| \nu - \pi \right\|_{\frac{1}{\pi}} e^{-q(1 - |\lambda_2(P)|)t},$$

where $|\lambda_2(P)|$ is the second largest eigenvalue modulus of P. □

– The matrix given below arises in the queueing example which we consider in Section 4 and its eigenvalues are stated in the below result which is proved in [4].

Proposition 5 *The eigenvalues of an $r \times r$ matrix,*

$$P = \begin{pmatrix} 1-\beta & \beta & 0 & 0 & \dots \\ 1-\beta & 0 & \beta & 0 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \dots & 0 & 1-\beta & 0 & \beta \\ \dots & 0 & 0 & 1-\beta & \beta \end{pmatrix},$$

where $0 < \beta < 1$ are given by $1, 2\sqrt{\beta(1-\beta)}\cos\left(\frac{\pi j}{r}\right)$, $j = 1, \dots, r-1$ and therefore the second largest eigenvalue modulus is $|\lambda_2(P)| = 2\sqrt{\beta(1-\beta)}\cos\left(\frac{\pi}{r}\right)$. □

– Mixing time of a Markov chain measures the time needed for the Markov chain to converge to its stationary distribution.

Definition 4 For a continuous time Markov chain $P(t)$, with stationary distribution π, given an $\varepsilon > 0$, the mixing time $t_{mix}(\varepsilon)$ is defined as

$$t_{mix}(\varepsilon) = \inf\left\{t : \|\nu^T P(t) - \pi^T\|_{TV} \leq \varepsilon, \text{for all probability distributions } \nu\right\}.$$

□

– To define adiabatic time, consider a linear evolution [8] of generator matrices described by

$$Q\left(\frac{t}{T}\right) = \left(1 - \frac{t}{T}\right) Q_{initial} + \frac{t}{T} Q_{final},$$

for $T > 0$, $0 \leq t \leq T$ and bounded generators $Q_{initial}$ and Q_{final}. If π_f is the unique stationary distribution for Q_{final}, adiabatic time measures the time needed for the chain to converge to π_f.

Definition 5 Given the above transitions generating a continuous time Markov chain $P(0,T)$ and $\varepsilon > 0$, adiabatic time is defined as

$$T_\varepsilon = \inf\left\{T : \|\nu^T P(0,T) - \pi_f^T\|_{TV} \leq \varepsilon, \text{for all probability distributions } \nu\right\}.$$

We will look at a different evolution of generator matrices in the next section. □

3 Adiabatic Framework

In this section we define an evolution of continuous time Markov generator matrices and look at the convergence in Definition 5 in terms of this new evolution. Consider the following evolution: time is divided into slots of size Δt and the generator matrix changes at these intervals. The bounded generator matrix Q_i determines the transition probabilities in the time interval $(i\Delta t, (i+1)\Delta t]$. The method of uniformization gives the corresponding transition probability matrix $P(i\Delta t, (i+1)\Delta t)$ as in (1). The upper bound on departure rates over all states is q_i for each Q_i. Therefore,

$$P(i\Delta t, (i+1)\Delta t) = e^{Q_i \Delta t} = e^{q_i (P_i - I)\Delta t},$$

where $P(t_1,t_2)$ denotes the matrix of transition probabilities from time t_1 to t_2. Let the matrix P_i be irreducible and reversible with stationary distribution π_i and second largest eigenvalue modulus $|\lambda_2(P_i)|$. Note that this evolution can be the result of any kind of time inhomogeneity in the system resulting in a changing Q_i which is updated at fixed intervals of time. As noted before, the time inhomogeneity can be due to the nature of the underlying process or due to uncertainties in measurements of parameters and the following theorem captures either kind.

Let ν_n be the distribution of the chain at time $n\Delta t$. We are interested in the distance between ν_n and the stationary distribution π_n corresponding to matrix P_n at time $n\Delta t$. The following theorem gives an upper bound on the distance at time $n\Delta t$ in terms of the distance at $n_0\Delta t$ for $n_0 < n$.

Theorem 1 *For the time inhomogeneous Markov chain generated by the matrices* $\{Q_i\}_{i=0}^n = \{q_i(P_i - I)\}_{i=0}^n$ *from time* 0 *to* $n\Delta t$*,*

$$\|\nu_n - \pi_n\|_{TV} \le \frac{1}{2}\|\nu_{n_0} - \pi_{n_0}\|_{\frac{1}{\pi_{n_0}}} \prod_{i=n_0}^{n-1} e^{-q_i(1-|\lambda_2(P_i)|)\Delta t} \sqrt{\max_{k\in\Omega} \frac{\pi_i(k)}{\pi_{i+1}(k)}}$$
$$+ \frac{1}{2}\sum_{i=n_0}^{n-1}\left[\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}} \prod_{j=i+1}^{n-1} e^{-q_j\left(1-|\lambda_2(P_j)|\right)\Delta t}\sqrt{\max_{k\in\Omega}\frac{\pi_j(k)}{\pi_{j+1}(k)}}\right],$$

where ν_n *is the distribution at time* $n\Delta t$, ν_{n_0} *is the distribution at time* $n_0\Delta t$ *for* $n_0 < n$ *and* $\{P_i\}_i$ *are irreducible and reversible with stationary distribution* π_i *and second largest eigenvalue modulus* $|\lambda_2(P_i)|$.

Proof We start with the $\frac{1}{\pi}$ norm so that we can use the result in Proposition 4.

$$\|\nu_n - \pi_n\|_{\frac{1}{\pi_n}} = \|\nu_{n-1}^T e^{Q_{n-1}\Delta t} - \pi_n^T\|_{\frac{1}{\pi_n}}$$
$$\le \|\nu_{n-1}^T e^{Q_{n-1}\Delta t} - \pi_{n-1}^T\|_{\frac{1}{\pi_n}} + \|\pi_{n-1} - \pi_n\|_{\frac{1}{\pi_n}} \tag{2}$$
$$\le \|\nu_{n-1}^T e^{Q_{n-1}\Delta t} - \pi_{n-1}^T\|_{\frac{1}{\pi_{n-1}}} \sqrt{\max_{k\in\Omega}\frac{\pi_{n-1}(k)}{\pi_n(k)}} + \|\pi_{n-1} - \pi_n\|_{\frac{1}{\pi_n}} \tag{3}$$
$$\le \|\nu_{n-1} - \pi_{n-1}\|_{\frac{1}{\pi_{n-1}}} e^{-q_{n-1}(1-|\lambda_2(P_{n-1})|)\Delta t}\sqrt{\max_{k\in\Omega}\frac{\pi_{n-1}(k)}{\pi_n(k)}}$$
$$+\|\pi_{n-1} - \pi_n\|_{\frac{1}{\pi_n}}, \tag{4}$$

where (2) follows from triangle inequality, (3) follows from the definition of the norm and (4) follows from Proposition 4. We can expand the above iteratively using triangle inequality for every time step which gives us

$$\|\nu_n - \pi_n\|_{\frac{1}{\pi_n}} \le \|\nu_{n_0} - \pi_{n_0}\|_{\frac{1}{\pi_{n_0}}} \prod_{i=n_0}^{n-1} e^{-q_i(1-|\lambda_2(P_i)|)\Delta t}\sqrt{\max_{k\in\Omega}\frac{\pi_i(k)}{\pi_{i+1}(k)}}$$
$$+\sum_{i=n_0}^{n-1}\left[\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}}\prod_{j=i+1}^{n-1} e^{-q_j(1-|\lambda_2(P_j)|)\Delta t}\sqrt{\max_{k\in\Omega}\frac{\pi_j(k)}{\pi_{j+1}(k)}}\right].$$

The result follows from Proposition 1. □

4 Application to Queueing Model

In this section we apply the above defined adiabatic evolution model to a queueing process. In particular, we consider time inhomogeneity due to uncertainty in a parameter. Consider an M/M/1/K queue with unknown packet arrival rate λ per unit time. We estimate λ at time $i\Delta t$ denoted by $\hat{\lambda}_i$ and decide packet departure rate, $\mu_i = f(\hat{\lambda}_i)$ based on this estimate.

Definition 6 Queueing policy is defined as the sequence $\{\hat{\lambda}_i, \mu_i = f(\hat{\lambda}_i)\}_{i\geq 1}$ where $f : \mathbb{R}_+ \to \mathbb{R}_+$ and μ_i is applied for time from $(i\Delta t, (i+1)\Delta t]$. □

This decides the following generator matrix from $(i\Delta t, (i+1)\Delta t]$:

$$Q_i = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu_i & -(\mu_i+\lambda) & \lambda & 0 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \dots & 0 & \mu_i & -(\mu_i+\lambda) & \lambda \\ \dots & 0 & 0 & \mu_i & -\mu_i \end{pmatrix}. \tag{5}$$

The corresponding transition probability matrix $P(i\Delta t, (i+1)\Delta t)$ is obtained as in (1). The upper bound on departure rates over all states is $\lambda + \mu_i$. Therefore,

$$P(i\Delta t, (i+1)\Delta t) = e^{Q_i \Delta t} = e^{(\lambda+\mu_i)(P_i - I)\Delta t},$$

where the matrix, P_i is

$$P_i = \begin{pmatrix} 1-\beta_i & \beta_i & 0 & 0 & \dots \\ 1-\beta_i & 0 & \beta_i & 0 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \dots & 0 & 1-\beta_i & 0 & \beta_i \\ \dots & 0 & 0 & 1-\beta_i & \beta_i \end{pmatrix}, \tag{6}$$

where $\beta_i = \frac{\lambda}{\mu_i+\lambda}$. The P_i's are reversible and irreducible with stationary distribution given by

$$\pi_i = \frac{1}{\sum_{r=0}^{K} \rho_i^r} \left[1, \rho_i, \rho_i^2, \dots, \rho_i^{K-1}, \rho_i^K\right]^T,$$

where $\rho_i = \frac{\beta_i}{1-\beta_i} = \frac{\lambda}{\mu_i}$. The second largest eigenvalue modulus is given by $|\lambda_2(P_i)| = 2\frac{\sqrt{\rho_i}}{(1+\rho_i)} \cos(\frac{\pi}{K+1})$. Also, the matrix Q_0 could be decided from a random departure rate and corresponds to the transition probability matrix $P(0, \Delta t)$.

4.1 Performance of an Adaptive Queueing Policy

Now we look at a specific queueing policy determined by the time average of number of packets arrived.

Definition 7

$$\hat{\lambda}_i = \frac{1}{i\Delta t}\sum_{k=1}^{i} X_k,$$
$$\mu_i = f(\hat{\lambda}_i) = (1+\delta)\hat{\lambda}_i,$$

where $X_k \sim \text{Poisson}(\lambda\Delta t)$ is the number of packets in the kth slot of duration Δt and $\delta > 0$ is a constant. □

This particular queueing policy ensures that the departure rate is always higher than the estimated arrival rate and since the estimated arrival rate itself must approach the actual one, should ensure a stable queue.

With this queueing policy, we have the adiabatic evolution generated by the matrices Q_i in (5). The corresponding P_i's are given by (6). The ratio $\rho_i = \frac{\lambda}{\mu_i} = \frac{\lambda}{(1+\delta)\hat{\lambda}_i}$. With full knowledge of the arrival rate, the above ratio becomes $\rho = \frac{1}{1+\delta}$ and let the corresponding matrix P has stationary distribution π. We redefine adibatic time in this setting.

Definition 8 Given the above transitions generating a continuous time Markov chain $P(0,n\Delta t)$ and $\varepsilon > 0$, adiabatic time is defined as

$$T_{\varepsilon,\gamma} = \Delta t \cdot \inf\{n : \Pr\{\|\nu^T P(0,n\Delta t) - \pi^T\|_{TV} < \varepsilon\} > 1-\gamma\,,\ \text{for all probability distributions } \nu\}\,.$$

where π is the stationary distribution corresponding to the queueing policy $\{\lambda,(1+\delta)\lambda\}$.

Theorem 2 (Main result) *Given $0 < \varepsilon < 1$, $0 < \gamma < 1$ and λ, the unknown arrival rate, the queueing policy in Definition 7 with $\delta > 0$ for an M/M/1/K queue has*

$$T_{\varepsilon,\gamma} \le \frac{2\log\frac{2}{\gamma_1}}{\lambda(\varepsilon_0^2-\varepsilon_0^3)} + \frac{\log\left(2\left[(1+\varepsilon_0)(1+\delta)\right]^{\frac{K}{2}+1}\right) - \log(\varepsilon\delta)}{\frac{1}{2\Delta t}\log\left(\frac{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}\right) + \lambda\left(\sqrt{(1+\delta)(1-\varepsilon_0)}-1\right)^2},$$

where ε_0 satisfies

$$e^{-\lambda\Delta t(\sqrt{(1+\delta)(1-\varepsilon_0)}-1)^2}\sqrt{\frac{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}} < 1,$$

$$\frac{\frac{\sqrt{(1-\varepsilon_0)(1+\delta)}\varepsilon_1}{|(1-\varepsilon_0)^2(1+\delta)-(1-\varepsilon_0+\varepsilon_1)|}}{1-e^{-\lambda\Delta t(\sqrt{(1+\delta)(1-\varepsilon_0)}-1)^2}\sqrt{\frac{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}}} \le \frac{\varepsilon}{2},$$

$$\varepsilon_0 \le \frac{\delta\varepsilon}{(1+\delta)(1+\varepsilon)},$$

and $\varepsilon_1 = \frac{\lambda\Delta t(\varepsilon_0^2-\varepsilon_0^3)}{2\log\frac{2}{\gamma_1}}\left(\frac{1}{\sqrt{\lambda\Delta t\gamma_2}}+\varepsilon_0\right)$, $0<\gamma_1<\gamma$, $\gamma_2 = \gamma-\gamma_1$. □

The above theorem gives an upper bound on the time we must wait before the distribution of the queue length converges to the desired stationary distribution π. Note that π is decided by δ and can be designed to give a stable stationary distribution. Hence, for given small ε and γ, the theorem gives the sufficient amount of time to converge to a stable distribution within ε with high probability of $1-\gamma$. The choice of ε_0 to be the largest which satisfies all three conditions will give the lowest lower bound in the theorem. At large enough time the estimated arrival rate must approach the actual arrival rate and the difference can be bounded by ε_0 with a high probability. Furthermore two consecutive estimates, can differ only by a maximum of ε_1.

4.2 Proof of Theorem 2

To prove Theorem 2, we need to bound the terms in Theorem 1. Note that Theorem 1 involves terms from time $n_0\Delta t$ onwards. The motive behind this was, at large enough n_0 we can put an upper bound on all the terms in Theorem 1. Since $\hat{\lambda}_i\Delta t$ is an empirical average of i iid Poisson($\lambda\Delta t$) distributed random variables, it follows from the law of large numbers that it must approach the actual value of $\lambda\Delta t$ at large enough i. Furthermore, two consecutive estimates must be very near to each other. We need this fact to bound terms which have both i and $i+1$ in them like $\sqrt{\max_{k\in\Omega}\frac{\pi_i(k)}{\pi_{i+1}(k)}}$ and $\|\pi_i-\pi_{i+1}\|_{\frac{1}{\pi_{i+1}}}$. The above facts are made precise in the following lemma.

Lemma 1 *For* $0<\varepsilon_0<1$ *and* $0<\gamma_1<1$, *there exists* $n_0=\frac{2\log\frac{2}{\gamma_1}}{\lambda\Delta t(\varepsilon_0^2-\varepsilon_0^3)}$ *such that*

– $\forall i\geq n_0$, *with probability at least* $1-\gamma_1$

$$|\hat{\lambda}_i-\lambda|\leq\lambda\varepsilon_0, \tag{7}$$

$$e^{-(\lambda+\mu_i)(1-|\lambda_2(P_i)|)\Delta t}<e^{-\lambda\Delta t(\sqrt{(1+\delta)(1-\varepsilon_0)}-1)^2}, \tag{8}$$

$$\|\pi_i-\pi\|_{TV}<\frac{1}{2}\frac{\varepsilon_0(1+\delta)}{\delta-\varepsilon_0(1+\delta)}. \tag{9}$$

– *For* $\varepsilon_1=\frac{1}{n_0}(\frac{1}{\sqrt{\lambda\Delta t\gamma_2}}+\varepsilon_0)$ *and* $0<\gamma_2<1$ *and* $\forall i\geq n_0$, *with probability at least* $1-\gamma=1-\gamma_1-\gamma_2$

$$|\hat{\lambda}_{i+1}-\hat{\lambda}_i|<\lambda\varepsilon_1, \tag{10}$$

$$\sqrt{\max_{k\in\{0,1,\ldots,K\}}\frac{\pi_i(k)}{\pi_{i+1}(k)}}<\sqrt{\frac{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}}, \tag{11}$$

$$\|\pi_i-\pi_{i+1}\|_{\frac{1}{\pi_{i+1}}}<\frac{\sqrt{(1-\varepsilon_0)(1+\delta)}\varepsilon_1}{|(1-\varepsilon_0)^2(1+\delta)-(1-\varepsilon_0+\varepsilon_1)|}. \tag{12}$$

– *With probability at least* $1-\gamma_1$,

$$\|\nu_{n_0}-\pi_{n_0}\|_{\frac{1}{\pi_{n_0}}}<\frac{\left[(1+\varepsilon_0)(1+\delta)\right]^{\frac{K}{2}+1}}{\delta}. \tag{13}$$

□

Refer Appendix A for proof.

Rewriting Theorem 1 for the queueing model we have described,

$$\|\nu_n-\pi_n\|_{TV} \leq \frac{1}{2}\|\nu_{n_0}-\pi_{n_0}\|_{\frac{1}{\pi_{n_0}}}\prod_{i=n_0}^{n-1} e^{-(\lambda+\mu_i)(1-|\lambda_2(P_i)|)\Delta t}\sqrt{\max_{k\in\{0,1,\ldots,K\}}\frac{\pi_i(k)}{\pi_{i+1}(k)}}$$
$$+\frac{1}{2}\sum_{i=n_0}^{n-1}\left[\|\pi_i-\pi_{i+1}\|_{\frac{1}{\pi_{i+1}}}\prod_{j=i+1}^{n-1} e^{-(\lambda+\mu_j)(1-|\lambda_2(P_j)|)\Delta t}\sqrt{\max_{k\in\{0,1,\ldots,K\}}\frac{\pi_j(k)}{\pi_{j+1}(k)}}\right].$$

According to Lemma 1, the above inequality can be bounded, with probability at least $1-\gamma_1-\gamma_2=1-\gamma$.

$$\|\nu_n-\pi_n\|_{TV} < \frac{1}{2}\frac{\left[(1+\varepsilon_0)(1+\delta)\right]^{\frac{K}{2}+1}}{\delta}\left[A\cdot B\right]^{n-n_0}+\frac{1}{2}\frac{C}{1-A\cdot B}\leq\frac{\varepsilon}{2}. \quad (14)$$

We use the notation below for brevity.

$$A = e^{-\lambda\Delta t(\sqrt{(1+\delta)(1-\varepsilon_0)}-1)^2},$$
$$B = \sqrt{\frac{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}},$$
$$C = \frac{\sqrt{(1-\varepsilon_0)(1+\delta)}\varepsilon_1}{(1-\varepsilon_0)^2(1+\delta)-(1-\varepsilon_0+\varepsilon_1)}.$$

(14) holds if ε_0 is chosen such that $A\cdot B<1$. Now the distance between ν_n and π which is the stationary distribution corresponding to the matrix P with full knowledge of λ is given by

$$\begin{aligned}\|\nu_n-\pi\|_{TV} &= \|\nu_n-\pi_n+\pi_n-\pi\|_{TV}\\ &\leq \|\nu_n-\pi_n\|_{TV}+\|\pi_n-\pi\|_{TV}\leq\varepsilon,\end{aligned} \quad (15)$$

using triangle inequality.

Our aim is to find the n such that $\|\nu_n-\pi\|_{TV}<\varepsilon$. (14) holds if both terms are $\leq\frac{\varepsilon}{4}$. For the first term we get,

$$\frac{\left[(1+\varepsilon_0)(1+\delta)\right]^{\frac{K}{2}+1}}{\delta}\left[A.B\right]^{n-n_0}\leq\frac{\varepsilon}{2}$$

$$n-n_0 \geq \frac{\log\left(2\left[(1+\varepsilon_0)(1+\delta)\right]^{\frac{K}{2}+1}\right)-\log(\varepsilon\delta)}{\frac{1}{2}\log\left(\frac{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}\right)+\lambda\Delta t(\sqrt{(1+\delta)(1-\varepsilon_0)}-1)^2},$$

which gives a condition on n. Adding with n_0 and multiplying by Δt gives the adiabatic time as defined in Definition (8). For the second term of (14),

$$\frac{\frac{\sqrt{(1-\varepsilon_0)(1+\delta)}\varepsilon_1}{(1-\varepsilon_0)^2(1+\delta)-(1-\varepsilon_0+\varepsilon_1)}}{1-e^{-\lambda\Delta t(\sqrt{(1+\delta)(1-\varepsilon_0)}-1)^2}\sqrt{\frac{\left[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\right]^{K+1}-1}{\left[(1-\varepsilon_0)(1+\delta)\right]^{K+1}-1}}} \leq \frac{\varepsilon}{2},$$

which gives a second condition on ε_0.

From (15) and (9),

$$\|\pi_n - \pi\|_{TV} < \frac{1}{2}\frac{\varepsilon_0(1+\delta)}{\delta-\varepsilon_0(1+\delta)} \leq \frac{\varepsilon}{2}, \tag{16}$$

which gives

$$\varepsilon_0 \leq \frac{\delta\varepsilon}{(1+\delta)(1+\varepsilon)},$$

and a third condition on ε_0. This completes the proof of Theorem 2.

4.3 Simulations

Now we look at the distance of the distribution of the queue from a stable distribution with increasing time. This distance must be small for large enough time from the discussions above. Here we look at distance as a function of time for a single sample path to verify whether the distance at the adiabatic time predicted is indeed lower than ε.

Simulation Setup 1 $\lambda = 10, \delta = 0.1, K = 100, \Delta t = 0.5, \varepsilon = 0.1, \gamma = 0.05, \gamma_1 = 0.04$.

Theorem 2 predicts $n_0 = 174391$ and $n = 175438$ for $\varepsilon_0 = 0.003$ (the highest which satisfies all the conditions in Theorem 2) or adiabatic time for $\varepsilon = 0.1, \gamma = 0.05$, $T_{0.1,0.05} = 87719$. Figure 1 shows the results of a simulation which measures the distance at each time slot vs. the distance upper bound given by triangle inequality in Theorem 1. The triangle inequality upper bound very closely follows the actual distance and both are much smaller than the distance upper bound predicted by (14) and (16), which is 0.056. Note that this is only a sample path. This indicates that even though Theorem 2 gives $T_{0.1,0.05} = 87719$ as a sufficient condition, the distance of $\varepsilon = 0.1$ is achieved much before. For example, Figure 2 shows the simulation results for $T = 5000$ or $n = 10000$. (The upper bound starts from $n_0 = 500$.) This was also found to be true for all 100 sample paths tested. The Figure 3 shows the histogram of distances in these 100 sample paths. Note that all of them have distance < 0.1.

Notice the initial increase in the upper bound due to Theorem 1 in both Figures 1 and 2. This is because of the second term of the bound which adds terms with time. After a sufficient number of time slots, however, the decrease in the first term overshadows the increase in the second and the overall bound starts coming down.

Now we look at the effects of increasing δ which decides how fast the departures are compared to the arrivals.

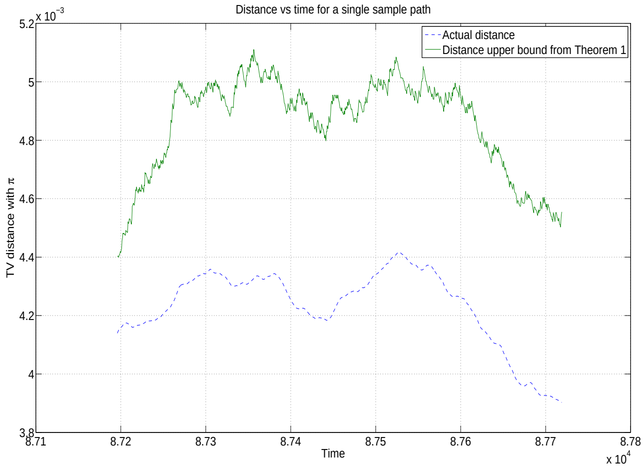

Fig. 1 Comparison of distance upper bound in Theorem 1 with actual distance

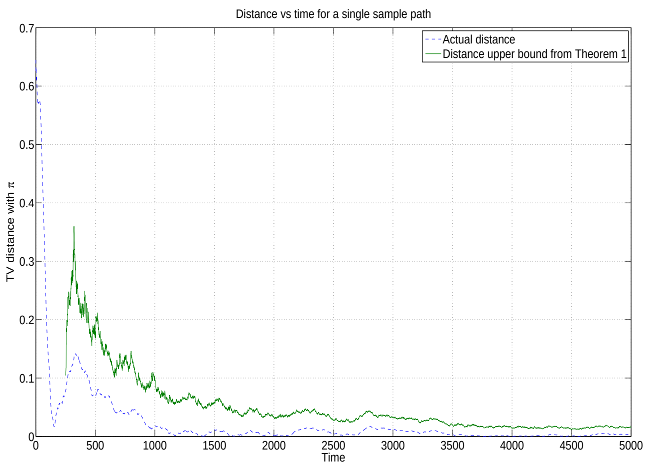

Fig. 2 Comparison of distance upper bound with actual distance at lesser time than predicted by Theorem 2

Simulation Setup 2 $\lambda = 10, K = 100, \Delta t = 0.5, \varepsilon = 0.1, \gamma = 0.05, \gamma_1 = 0.04$ with different $\delta = 0.1 : 0.1 : 0.9$.

Note that increase in δ corresponds to emptying out the queue faster or in other words, the stable distribution should be approached faster. Theorem 2 predictions are as shown in Figure 4 which shows that the adiabatic time decreases with increases in δ. Or, at the same time, the distance decreases with increasing δ. The distances at $T = 2500$ averaged over 100 sample paths are shown in Figure 5 which confirms this.

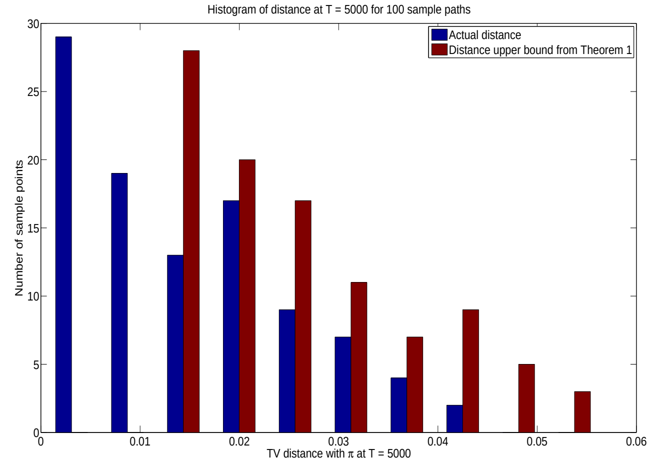

Fig. 3 Histogram of distances of 100 sample paths at $T = 5000$

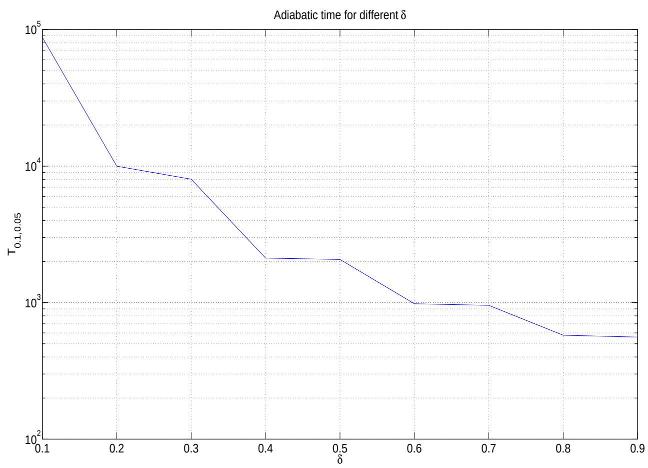

Fig. 4 Adiabatic time for different δ

Another parameter of interest is the sampling time Δt which decides how frequently the estimates are done.

Simulation Setup 3 $\lambda = 10, K = 100, \delta = 0.1, \varepsilon = 0.1, \gamma = 0.05, \gamma_1 = 0.04$ with different $\Delta t = 1, 5, 10, 20, 25, 50$.

The increase in Δt corresponds to applying the information gained from estimates less frequently and must result in an increased distance at the same time. The distances at time $T = 100$ averaged over 100 sample paths are shown in Figure 6 which shows that this is true. This shows the benefits of updates which are more frequent

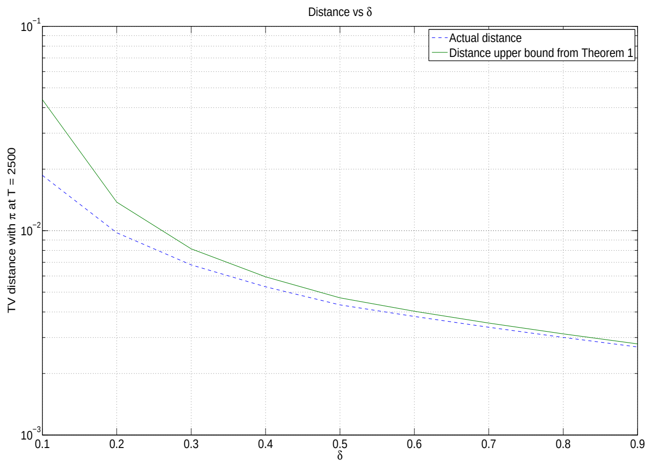

Fig. 5 Comparison of distance for different δ

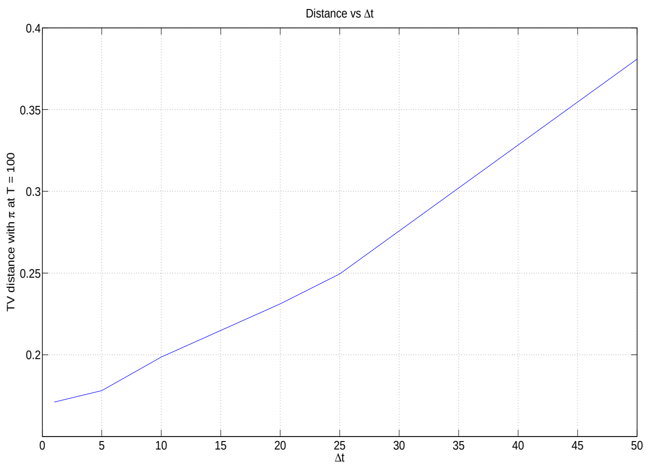

Fig. 6 Comparison of distance for different Δt.

and hence the system following a smoother path towards the stable distribution rather than attempting to reach there at one go.

5 Conclusion

We considered a time inhomogeneous Markov chain with its evolution governed by generator matrices which change at fixed time intervals. We followed the adiabatic approach to study this chain and bounded the distance to a stationary distribution. This is a general theorem which is applicable to evolutions of the above kind.

We used this model of time inhomogeneous Markov chains to characterize a birth-death chain in which the time inhomogeneity is due to uncertainty in parameters. We considered a Markovian queue with an adaptive queueing policy and derived sufficient conditions on time after which the distribution of the queue can be considered to be stable with high probability.

A Proof of Lemma 1

A.1 Proof of (7)

For a random variable X which is the empirical average of i iid random variables $\{X_k\}_{k=1}^{i}$, Chernoff bound for tail probabilities gives,

$$\begin{aligned}\Pr\{X \geq a\} &\leq \exp\Big(-i \sup_{t>0}\big(at - \log(\mathbb{E}[e^{tX_i}])\big)\Big),\\ \Pr\{X \leq a\} &\leq \exp\Big(-i \sup_{t>0}\big(-at - \log(\mathbb{E}[e^{-tX_i}])\big)\Big).\end{aligned}$$

Since X_i is Poisson$(\lambda\Delta t)$ in our case $\log(\mathbb{E}[e^{tX_i}]) = \lambda\Delta t(e^t - 1)$. Substituting this in the above, we get

$$\Pr\{\hat{\lambda}_i\Delta t \geq \lambda\Delta t(1+\varepsilon_0)\} \leq \exp\Big(-i\lambda\Delta t \sup_{t>0}\big((1+\varepsilon_0)t - (e^t - 1)\big)\Big).$$

The maximum above is attained at $t = \log(1+\varepsilon_0) > 0$. Similarly for

$$\Pr\{\hat{\lambda}_i\Delta t \leq \lambda\Delta t(1-\varepsilon_0)\} \leq \exp\Big(-i\lambda\Delta t \sup_{t>0}\big(-(1-\varepsilon_0)t - (e^{-t} - 1)\big)\Big).$$

The maximum here is attained at $t = \log(\frac{1}{1-\varepsilon_0}) > 0$. Using the above bounds, we can bound the probability of the complement of the event in (7).

$$\begin{aligned}\Pr\{|\hat{\lambda}_i - \lambda| \geq \lambda\varepsilon_0\} &= \Pr\{\hat{\lambda}_i\Delta t \geq \lambda\Delta t(1+\varepsilon_0)\} + \Pr\{\hat{\lambda}_i\Delta t \leq \lambda\Delta t(1-\varepsilon_0)\}\\ &\leq \exp\Big(-i\lambda\Delta t\big((1+\varepsilon_0)\log(1+\varepsilon_0) - \varepsilon_0\big)\Big)\\ &\quad + \exp\Big(-i\lambda\Delta t\big((1-\varepsilon_0)\log(1-\varepsilon_0) + \varepsilon_0\big)\Big)\\ &< 2\exp\Big(-i\lambda\Delta t\big((1+\varepsilon_0)\log(1+\varepsilon_0) - \varepsilon_0\big)\Big) &(17)\\ &< 2\exp\Big(-i\lambda\Delta t\big((1+\varepsilon_0)(\varepsilon_0 - \frac{\varepsilon_0^2}{2}) - \varepsilon_0\big)\Big) &(18)\\ &= 2\exp\Big(-i\lambda\Delta t\big(\frac{\varepsilon_0^2 - \varepsilon_0^3}{2}\big)\Big) \leq \gamma_1. &(19)\end{aligned}$$

(17) holds since $(1-\varepsilon_0)\log(1-\varepsilon_0) + \varepsilon_0 > (1+\varepsilon_0)\log(1+\varepsilon_0) - \varepsilon_0$ for $0 < \varepsilon_0 < 1$ and (18) is true since $\log(1+\varepsilon_0) > \varepsilon_0 - \frac{\varepsilon_0^2}{2}$ according to Taylor's theorem for $0 < \varepsilon_0 < 1$. From (19),

$$\begin{aligned}\exp\Big(-i\lambda\Delta t\big(\frac{\varepsilon_0^2 - \varepsilon_0^3}{2}\big)\Big) &\leq \frac{\gamma_1}{2}\\ i &\geq \frac{2\log\frac{2}{\gamma_1}}{\lambda\Delta t\big(\varepsilon_0^2 - \varepsilon_0^3\big)}.\end{aligned}$$

This implies (7) must hold for

$$n_0 = \frac{2\log\frac{2}{\gamma_1}}{\lambda\Delta t(\varepsilon_0^2 - \varepsilon_0^3)}.$$

A.2 Proof of (10)

We have the following equation from the estimation of λ,

$$\begin{aligned}\hat{\lambda}_{i+1}\Delta t &= \hat{\lambda}_i\Delta t\frac{i}{i+1}+\frac{1}{i+1}X_{i+1}.\\ \hat{\lambda}_{i+1}\Delta t-\hat{\lambda}_i\Delta t &= \frac{1}{i+1}(X_{i+1}-\hat{\lambda}_i\Delta t)\\ &< \frac{1}{n_0}(X_{i+1}-\lambda\Delta t(1-\varepsilon_0)), \qquad (20)\end{aligned}$$

with a probability of at least $1-\gamma_1, \forall i+1>n_0$ since we have $\hat{\lambda}_i \geq \lambda(1-\varepsilon_0)$ by (7). Since X_{i+1} is a Poisson($\lambda\Delta t$) random variable, we can find a constant c such that $\Pr\{|X_{i+1}-\lambda\Delta t| \geq c\lambda\Delta t\} \leq \gamma_2$. From Chebyshev's inequality, for Poisson($\lambda\Delta t$) random variable

$$\begin{aligned}\Pr\{|X_{i+1}-\lambda\Delta t| \geq c\lambda\Delta t\} &\leq \frac{1}{c^2\lambda\Delta t}\\ &= \gamma_2.\end{aligned}$$

A choice of $c=\frac{1}{\sqrt{\lambda\Delta t\gamma_2}}$ will satisfy the above equation. Therefore, (20) can be written as

$$\begin{aligned}\hat{\lambda}_{i+1}-\hat{\lambda}_i &< \frac{1}{n_0}(\lambda(1+c)-\lambda(1-\varepsilon_0))\\ &= \frac{\lambda}{n_0}(c+\varepsilon_0),\end{aligned}$$

with a probability of at least $1-\gamma=(1-\gamma_1)(1-\gamma_2)$. Similarly we have

$$\hat{\lambda}_{i-1}-\hat{\lambda}_i < \frac{\lambda}{n_0}(c+\varepsilon_0),$$

with a probability of at least $1-\gamma$. Therefore (10) holds for $\varepsilon_1=\frac{1}{n_0}(c+\varepsilon_0)$.

A.3 Proof of (8)

$$\begin{aligned}|\lambda_2(P_i)| &= 2\cos\left(\frac{\pi}{K+1}\right)\frac{\rho_i^{1/2}}{1+\rho_i}\\ &= 2\cos\left(\frac{\pi}{K+1}\right)\sqrt{1+\delta}\,\frac{\sqrt{\hat{\lambda}_i/\lambda}}{1+\frac{(1+\delta)\hat{\lambda}_i}{\lambda}}.\end{aligned}$$

$$\begin{aligned}e^{-(\lambda+\mu_i)(1-|\lambda_2(P_i)|)\Delta t} &= e^{\lambda\Delta t\left(2\cos\left(\frac{\pi}{K+1}\right)\sqrt{(1+\delta)\frac{\hat{\lambda}_i}{\lambda}}-1-(1+\delta)\frac{\hat{\lambda}_i}{\lambda}\right)}\\ &< e^{\lambda\Delta t\left(2\sqrt{(1+\delta)\frac{\hat{\lambda}_i}{\lambda}}-1-(1+\delta)\frac{\hat{\lambda}_i}{\lambda}\right)} \qquad (21)\\ &= e^{-\lambda\Delta t\left(\sqrt{(1+\delta)\frac{\hat{\lambda}_i}{\lambda}}-1\right)^2}. \qquad (22)\end{aligned}$$

(21) is an inequality since $\cos(\frac{\pi}{K+1})<1$ for $K>1$. From (7) it follows that $\forall i \geq n_0$, with a probability at least $1-\gamma_1$, $1-\varepsilon_0 \leq \frac{\hat{\lambda}_i}{\lambda} \leq 1+\varepsilon_0$. The RHS of (22) is maximized at $\frac{\hat{\lambda}_i}{\lambda}=1-\varepsilon_0$. Therefore (22) can be bounded with a probability of at least $1-\gamma_1$ by

$$e^{-(\lambda+\mu_i)(1-|\lambda_2(P_i)|)\Delta t} < e^{-\lambda\Delta t\left(\sqrt{(1+\delta)(1-\varepsilon_0)}-1\right)^2}.$$

A.4 Proof of (9)

$$\begin{aligned} \|\pi_i - \pi\|_{TV} &= \frac{1}{2}\sum_{k=0}^{K} |\pi_i(k) - \pi(k)| \\ &= \frac{1}{2}\sum_{k=0}^{K} \left| \frac{\rho_i^k}{\sum_{r=0}^K \rho_i^r} - \frac{\rho^k}{\sum_{r=0}^K \rho^r} \right| \\ &= \frac{1}{2}\sum_{k=0}^{K} \left| \frac{(\frac{\lambda}{(1+\delta)\hat{\lambda}_i})^k}{\sum_{r=0}^K (\frac{\lambda}{(1+\delta)\hat{\lambda}_i})^r} - \frac{\rho^k}{\sum_{r=0}^K \rho^r} \right|. \end{aligned} \tag{23}$$

The RHS of (23) is maximized at $\frac{\hat{\lambda}_i}{\lambda} = 1 - \varepsilon_0$ and therefore the above can be written as

$$\begin{aligned} \|\pi_i - \pi\|_{TV} &\leq \frac{1}{2}\sum_{k=0}^{K} \Big| \frac{(\frac{\rho}{1-\varepsilon_0})^k}{\sum_{r=0}^K (\frac{\rho}{1-\varepsilon_0})^r} - \frac{\rho^k}{\sum_{r=0}^K \rho^r} \Big| \\ &= \frac{1}{2}\sum_{k=0}^{K} (1+\delta)^{K-k} \Big| \frac{(1-\varepsilon_0)^{K-k}((1-\varepsilon_0)(1+\delta)-1)}{[(1-\varepsilon_0)(1+\delta)]^{K+1}-1} - \frac{\delta}{(1+\delta)^{K+1}-1} \Big| \\ &< \frac{1}{2}\sum_{k=0}^{K} (1+\delta)^{K-k} \Big| \frac{(1-\varepsilon_0)^{K-k}\delta}{[(1-\varepsilon_0)(1+\delta)]^{K+1}-1} - \frac{\delta}{(1+\delta)^{K+1}-1} \Big| \\ &= \frac{\delta}{2}\sum_{k=0}^{K} (1+\delta)^{K-k} \Big[\frac{(1-\varepsilon_0)^{K-k}}{[(1-\varepsilon_0)(1+\delta)]^{K+1}-1} - \frac{1}{(1+\delta)^{K+1}-1} \Big] \\ &= \frac{1}{2}\frac{\varepsilon_0(1+\delta)}{\delta - \varepsilon_0(1+\delta)}, \end{aligned} \tag{24}$$

$\frac{(1-\varepsilon_0)^{K-k}}{[(1-\varepsilon_0)(1+\delta)]^{K+1}-1} = \frac{1}{(1-\varepsilon_0)^{k+1}(1+\delta)^{K+1}-(1-\varepsilon_0)^{k-K}} > \frac{1}{(1-\varepsilon_0)^{k+1}(1+\delta)^{K+1}-1} > \frac{1}{(1+\delta)^{K+1}-1}$ and hence the quantity inside the modulus in (24) is always positive.

A.5 Proof of (11)

$$\begin{aligned} \max_{k\in\{0,1,\ldots,K\}} \frac{\pi_i(k)}{\pi_{i+1}(k)} &= \max_{k\in\{0,1,\ldots,K\}} \frac{\rho_i^k}{\rho_{i+1}^k} \frac{\sum_{r=0}^K \rho_{i+1}^r}{\sum_{r=0}^K \rho_i^r} \\ &= \max_{k\in\{0,1,\ldots,K\}} \Big(\frac{\hat{\lambda}_{i+1}}{\hat{\lambda}_i}\Big)^k \frac{\sum_{r=0}^K (\frac{\lambda}{(1+\delta)\hat{\lambda}_{i+1}})^r}{\sum_{r=0}^K (\frac{\lambda}{(1+\delta)\hat{\lambda}_i})^r}. \end{aligned} \tag{25}$$

From (7) it follows that $\forall i \geq i_0$, with a probability at least $1-\gamma_1$, $1-\varepsilon_0 \leq \frac{\hat{\lambda}_i}{\lambda}, \frac{\hat{\lambda}_{i+1}}{\lambda} \leq 1+\varepsilon_0$. In addition from (10), with a probability at least $1-\gamma$, $\frac{\hat{\lambda}_i}{\lambda} - \varepsilon_1 < \frac{\hat{\lambda}_{i+1}}{\lambda} < \frac{\hat{\lambda}_i}{\lambda} + \varepsilon_1$. The RHS of (25) is maximized at

$\frac{\hat{\lambda}_i}{\lambda} = 1-\varepsilon_0$, $\frac{\hat{\lambda}_{i+1}}{\lambda} = 1-\varepsilon_0+\varepsilon_1$. Therefore (25) can be bounded with a probability of at least $1-\gamma$ by

$$\begin{aligned}
\max_{k\in\{0,1,\ldots,K\}} \frac{\pi_i(k)}{\pi_{i+1}(k)} &\leq \max_{k\in\{0,1,\ldots,K\}} \Big(\frac{1-\varepsilon_0+\varepsilon_1}{1-\varepsilon_0}\Big)^k \frac{\sum_{r=0}^K (\frac{\rho}{1-\varepsilon_0+\varepsilon_1})^r}{\sum_{r=0}^K (\frac{\rho}{1-\varepsilon_0})^r} \\
&= \Big(\frac{1-\varepsilon_0+\varepsilon_1}{1-\varepsilon_0}\Big)^K \frac{\sum_{r=0}^K (\frac{\rho}{1-\varepsilon_0+\varepsilon_1})^r}{\sum_{r=0}^K (\frac{\rho}{1-\varepsilon_0})^r} \\
&= \frac{(1-\varepsilon_0+\varepsilon_1)^{K+1}-\rho^{K+1}}{(1-\varepsilon_0)^{K+1}-\rho^{K+1}} \frac{1-\varepsilon_0-\rho}{1-\varepsilon_0+\varepsilon_1-\rho} \\
&< \frac{(1-\varepsilon_0+\varepsilon_1)^{K+1}-\rho^{K+1}}{(1-\varepsilon_0)^{K+1}-\rho^{K+1}} \\
&= \frac{\big[(1-\varepsilon_0+\varepsilon_1)(1+\delta)\big]^{K+1}-1}{\big[(1-\varepsilon_0)(1+\delta)\big]^{K+1}-1},
\end{aligned} \tag{26}$$

where the inequality in (26) holds since $1-\varepsilon_0 < 1-\varepsilon_0+\varepsilon_1$ for $\varepsilon_1 > 0$.

A.6 Proof of (12)

$$\begin{aligned}
\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}} &= \Big(\sum_{k=0}^K \frac{\big(\pi_i(k)-\pi_{i+1}(k)\big)^2}{\pi_{i+1}(k)}\Big)^{1/2} \\
&= \Big(\sum_{k=0}^K \frac{\Big(\frac{\rho_i^k}{\sum_{r=0}^K \rho_i^r} - \frac{\rho_{i+1}^k}{\sum_{r=0}^K \rho_{i+1}^r}\Big)^2}{\frac{\rho_{i+1}^k}{\sum_{r=0}^K \rho_{i+1}^r}}\Big)^{1/2} \\
&= \left(\sum_{k=0}^K \frac{\left(\frac{\left(\frac{\lambda}{(1+\delta)\hat{\lambda}_i}\right)^k}{\sum_{r=0}^K \left(\frac{\lambda}{(1+\delta)\hat{\lambda}_i}\right)^r} - \frac{\left(\frac{\lambda}{(1+\delta)\hat{\lambda}_{i+1}}\right)^k}{\sum_{r=0}^K \left(\frac{\lambda}{(1+\delta)\hat{\lambda}_{i+1}}\right)^r}\right)^2}{\frac{\left(\frac{\lambda}{(1+\delta)\hat{\lambda}_{i+1}}\right)^k}{\sum_{r=0}^K \left(\frac{\lambda}{(1+\delta)\hat{\lambda}_{i+1}}\right)^r}}\right)^{1/2}.
\end{aligned} \tag{27}$$

As in the above, the RHS of (27) is maximized at $\frac{\hat{\lambda}_i}{\lambda} = 1-\varepsilon_0$, $\frac{\hat{\lambda}_{i+1}}{\lambda} = 1-\varepsilon_0+\varepsilon_1$. Therefore (27) can be bounded with a probability of at least $1-\gamma$ by

$$\begin{aligned}
\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}} &\leq \left(\sum_{k=0}^K \frac{\left(\frac{\left(\frac{\rho}{1-\varepsilon_0}\right)^k}{\sum_{r=0}^K \left(\frac{\rho}{1-\varepsilon_0}\right)^r} - \frac{\left(\frac{\rho}{1-\varepsilon_0+\varepsilon_1}\right)^k}{\sum_{r=0}^K \left(\frac{\rho}{1-\varepsilon_0+\varepsilon_1}\right)^r}\right)^2}{\frac{\left(\frac{\rho}{1-\varepsilon_0+\varepsilon_1}\right)^k}{\sum_{r=0}^K \left(\frac{\rho}{1-\varepsilon_0+\varepsilon_1}\right)^r}}\right)^{1/2} \\
&= \Big(\sum_{k=0}^K \frac{\Big(\frac{(\frac{\rho}{1-\varepsilon_0})^k}{a} - \frac{(\frac{\rho}{1-\varepsilon_0+\varepsilon_1})^k}{b}\Big)^2}{\frac{(\frac{\rho}{1-\varepsilon_0+\varepsilon_1})^k}{b}}\Big)^{1/2} \\
&= \Big(\frac{bc}{a^2}-1\Big)^{1/2},
\end{aligned} \tag{28}$$

where we define

$$
\begin{aligned}
a &= \sum_{k=0}^{K} \Big(\frac{\rho}{1-\varepsilon_0}\Big)^k \\
&= \frac{(1-\varepsilon_0)^{K+1} - \rho^{K+1}}{(1-\varepsilon_0)^K (1-\varepsilon_0-\rho)}, \\
b &= \sum_{k=0}^{K} \Big(\frac{\rho}{1-\varepsilon_0+\varepsilon_1}\Big)^k \\
&= \frac{(1-\varepsilon_0+\varepsilon_1)^{K+1} - \rho^{K+1}}{(1-\varepsilon_0+\varepsilon_1)^K (1-\varepsilon_0+\varepsilon_1-\rho)}, \\
c &= \sum_{k=0}^{K} \Big(\frac{\rho(1-\varepsilon_0+\varepsilon_1)}{(1-\varepsilon_0)^2}\Big)^k \\
&= \frac{(1-\varepsilon_0)^{2K+2} - \rho^{K+1}(1-\varepsilon_0+\varepsilon_1)^{K+1}}{(1-\varepsilon_0)^{2K}\big((1-\varepsilon_0)^2 - \rho(1-\varepsilon_0+\varepsilon_1)\big)}.
\end{aligned}
$$

Therefore, we can write using $x = 1-\varepsilon_0$, $y = 1-\varepsilon_0+\varepsilon_1$ for ease of notation

$$
\begin{aligned}
\frac{bc}{a^2} &= \frac{\big[y^{K+1} - \rho^{K+1}\big]\big[x^{2K+2} - \rho^{K+1}y^{K+1}\big]}{\big(x^{K+1} - \rho^{K+1}\big)^2 y^K} \frac{(x-\rho)^2}{(y-\rho)\big(x^2-\rho y\big)} \\
&= \frac{x^{2K+2}y^{K+1} + \rho^{2K+2}y^{K+1} - \rho^{K+1}\big[x^{2K+2} + y^{2K+2}\big]}{x^{2K+2}y^{K+1} + \rho^{2K+2}y^{K+1} - 2\rho^{K+1}x^{K+1}y^{K+1}} \frac{(x-\rho)^2 y}{(y-\rho)\big(x^2-\rho y\big)} \\
&< \frac{(x-\rho)^2 y}{(y-\rho)\big(x^2-\rho y\big)} \qquad (29) \\
&= \frac{x^2y + \rho^2 y - 2\rho xy}{x^2y + \rho^2 y - \rho(x^2+y^2)},
\end{aligned}
$$

where (29) is true since $x^{2K+2} + y^{2K+2} > 2x^{K+1}y^{K+1}$. Therefore (28) can be written as

$$
\begin{aligned}
\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}} &< \Big(\frac{x^2y + \rho^2 y - 2\rho xy}{x^2y + \rho^2 y - \rho(x^2+y^2)} - 1\Big)^{1/2} \\
&= \frac{\sqrt{\rho}\,\varepsilon_1}{\sqrt{(y-\rho)\big(x^2-\rho y\big)}} \\
&< \frac{\sqrt{\rho}\,\varepsilon_1}{\sqrt{(x-\rho)\big(x^2-\rho y\big)}} \qquad (30) \\
&= \frac{\sqrt{\rho x}\,\varepsilon_1}{\sqrt{(x^2-\rho x)\big(x^2-\rho y\big)}} \\
&< \frac{\sqrt{\rho x}\,\varepsilon_1}{|x^2-\rho y|} \qquad (31) \\
&= \frac{\sqrt{(1-\varepsilon_0)(1+\delta)}\,\varepsilon_1}{|(1-\varepsilon_0)^2(1+\delta) - (1-\varepsilon_0+\varepsilon_1)|},
\end{aligned}
$$

where (30) and (31) are true since $y > x$.

A.7 Proof of (13)

$$\begin{aligned}
\|\nu_{n_0}-\pi_{n_0}\|_{\frac{1}{\pi_{n_0}}} &= \Big[\sum_{k=0}^{K}\frac{\nu_{n_0}^2(k)}{\pi_{n_0}(k)}-1\Big]^{1/2} \\
&= \Big[\sum_{r=0}^{K}\rho_{n_0}^r\sum_{k=0}^{K}\frac{\nu_{n_0}^2(k)}{\rho_{n_0}^k}-1\Big]^{1/2} \\
&< \Big[\sum_{r=0}^{K}\rho_{n_0}^r\sum_{k=0}^{K}\frac{1}{\rho_{n_0}^k}\Big]^{1/2} \qquad (32)\\
&= \Big[\sum_{r=0}^{K}\Big(\frac{\lambda}{(1+\delta)\hat{\lambda}_{n_0}}\Big)^r\sum_{k=0}^{K}\Big(\frac{(1+\delta)\hat{\lambda}_{n_0}}{\lambda}\Big)^k\Big]^{1/2}, \qquad (33)
\end{aligned}$$

where (32) is true since $\nu_{n_0}(k)$ is a probability. The RHS of (33) is maximized at $\frac{\hat{\lambda}_{n_0}}{\lambda}=1+\varepsilon_0$. Therefore (33) can be bounded with a probability of at least $1-\gamma_1$ by

$$\begin{aligned}
\|\nu_{n_0}-\pi_{n_0}\|_{\frac{1}{\pi_{n_0}}} &\le \Big[\sum_{r=0}^{K}\Big(\frac{\rho}{1+\varepsilon_0}\Big)^r\sum_{k=0}^{K}\Big(\frac{1+\varepsilon_0}{\rho}\Big)^k\Big]^{1/2} \\
&= \Big[\frac{\big((1+\varepsilon_0)^{K+1}-\rho^{K+1}\big)^2}{(1+\varepsilon_0)^K\rho^K(1+\varepsilon_0-\rho)^2}\Big]^{1/2} \\
&= \frac{(1+\varepsilon_0)^{K+1}-\rho^{K+1}}{(1+\varepsilon_0)^{\frac{K}{2}}\rho^{\frac{K}{2}}(1+\varepsilon_0-\rho)} \\
&< \frac{(1+\varepsilon_0)^{K+1}}{(1+\varepsilon_0)^{\frac{K}{2}}\rho^{\frac{K}{2}}(1+\varepsilon_0-\rho)} \\
&= \frac{(1+\varepsilon_0)^{\frac{K}{2}+1}(1+\delta)^{\frac{K}{2}+1}}{(1+\varepsilon_0)(1+\delta)-1} \\
&< \frac{\big[(1+\varepsilon_0)(1+\delta)\big]^{\frac{K}{2}+1}}{\delta}.
\end{aligned}$$

This completes the proof of Lemma 7.

References

1. Born, M., Fock, V.: Beweis des adiabatensatzes. Zeitschrift Für Physik A Hadrons and Nuclei **51**(3-4), 165–180 (1928)
2. Bradford, K., Kovchegov, Y.: Adiabatic times for Markov chains and applications. Journal of Statistical Physics **143**, 955–969 (2011)
3. Brémaud, P.: Markov chains: Gibbs fields, Monte Carlo simulation, and queues. Springer-Verlag, New York (1999)
4. Feller, W.: An introduction to probability theory and its applications, volume 1, second edn. John Wiley & Sons, New York (1959)
5. Karlin, S., Taylor, H.M.: A first course in stochastic processes, second edn. Academic Press, New York (1975)
6. Kleinrock, L.: Queueing systems, volume 1: Theory. John Wiley & Sons, New York (1975)
7. Kleinrock, L.: Queueing systems, volume 2: Computer applications. John Wiley & Sons, New York (1976)
8. Kovchegov, Y.: A note on adiabatic theorem for Markov chains. Statistics and Probability Letters **80**, 186–190 (2010)
9. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov chains and mixing times, first edn. American Mathematical Society (2008)

10. Messiah, A.: Quantum mechanics, volume 2, first edn. John Wiley & Sons, New York (1962)
11. Ross, S.M.: Introduction to probability models, tenth edn. Academic Press (2010)
12. Zacharias, L.: An adiabatic approach to analysis of time inhomogeneous Markov chains: a queueing policy application. Master's thesis, Oregon State University, Corvallis, USA (2011)