Uniform Acceleration Expansions for Markov Chains with Time-Varying Rates[1]

William A. Massey
Bell Laboratories, Room 2C-320
Murray Hill, NJ 07974-0636
will@research.bell-labs.com

Ward Whitt
AT&T Laboratories, Room A117,
180 Park Avenue, Building 103,
Florham Park, NJ 07932-0971
wow@research.att.com

December 10, 1997

Abstract

We study uniform acceleration (UA) expansions of finite-state continuous-time Markov chains with time-varying transition rates. The UA expansions can be used to justify, evaluate, and refine the pointwise stationary approximation, which is the steady-state distribution associated with the time-dependent generator at the time of interest. We obtain UA approximations from these UA asymptotic expansions. We derive a time-varying analog to the uniformization representation of transition probabilities for chains with constant transition rates, and apply it to establish asymptotic results related to the UA asymptotic expansion. These asymptotic results can serve as appropriate time-varying analogs to the notions of stationary distributions and limiting distributions. We illustrate the UA approximations by doing a numerical example for the time-varying Erlang loss model.

[1]Accepted for publication in the Annals of Applied Probability.
AMS 1991 *subject classifications.* 60J27, 60K30.
Keywords and phrases: Time-inhomogeneous Markov chains, time-dependent behavior, nonstationary queueing models, pointwise stationary approximation, eigenvalue bounds, nonnegative matrices, asymptotic expansions, Poisson's equation, Erlang loss model.

1 Introduction

In many applied settings, such as with queueing systems, physical reality indicates that it is appropriate to use nonstationary models. For example, arrival rates in service systems typically vary substantially by time of day; e.g., see p. 259 of Hall [7]. However, if the rate of change is sufficiently slow, then it is natural to approximate the time-dependent distribution at any time t by the steady-state distribution of the model with transition characteristics at that time t. In particular, for a nonstationary *continuous-time Markov chain (CTMC)* with time-dependent generator $\{\mathbf{A}(t) : t \geq 0\}$, we would approximate the time-dependent probability vector $\mathbf{p}(t)$ at time t by the steady-state probability vector $\boldsymbol{\pi}(t)$ associated with $\mathbf{A}(t)$, obtained by solving $\boldsymbol{\pi}(t)\mathbf{A}(t) = 0$ and $\boldsymbol{\pi}(t)\mathbf{1}^{\mathsf{T}} = 1$, with the usual regularity conditions guaranteeing a unique solution. (We regard vectors as row vectors, so that $\mathbf{1}^{\mathsf{T}}$ is a column vector of 1's with T the matrix transpose.)

Some variant of the approximation procedure just described is routinely used in the performance analysis of telecommunications systems and in many other applied settings. It has been studied and called the *pointwise stationary approximation* (PSA) by Green and Kolesar [6], Whitt [26] and Eick, Massey and Whitt [4]. For example, Whitt [26] proved that PSA is asymptotically correct for $M_t/M_t/s$ queues and more general time-dependent birth-and-death processes as the birth and death rates increase, which is equivalent (by a change of time scale) to having the rates change more slowly.

In this paper we propose a way to quantitatively evaluate PSA and develop refinements to it (without having to solve for the actual time-dependent distribution). In particular, we focus on a class of asymptotic approximations called *uniform acceleration* (UA) *asymptotic expansions* for CTMCs. The UA framework provides a natural way to justify, evaluate and refine the PSA because the PSA is the first term of the UA expansion. When the next few terms are relatively small, we can be confident that the PSA is a good approximation, but when they are not, then the PSA can be regarded as unreliable. The first few terms of the UA expansion provide a convenient check on PSA because they are essentially no more difficult to compute than PSA itself.

In particular, suppose that $\mathbf{A}(t)$ is the time-dependent generator for a nonstationary CTMC. Then the *UA approximation of order* n for the distribution at time t is

$$\mathbf{p}(t) \approx \sum_{k=0}^{n} \boldsymbol{\pi}^{(k)}(t) \ , \tag{1.1}$$

where, for each k, the vector $\boldsymbol{\pi}^{(k)}$ is a solution of *Poisson's equation*

$$\boldsymbol{\pi}^{(k)}(t)\mathbf{A}(t) = \mathbf{y}^{(k)}(t) \ , \tag{1.2}$$

with

$$\mathbf{y}^{(0)}(t) = 0, \quad \boldsymbol{\pi}^{(0)}(t)\mathbf{1}^{\mathsf{T}} = 1, \tag{1.3}$$

$$\mathbf{y}^{(k)}(t) = \frac{d}{dt}\boldsymbol{\pi}^{(k-1)}(t) \quad \text{and} \quad \boldsymbol{\pi}^{(k)}(t)\mathbf{1}^{\mathsf{T}} = 0 \ \text{ for } \ k \geq 1 \ . \tag{1.4}$$

From (1.2) and (1.3), we see that $\boldsymbol{\pi}^{(0)}(t)$ is indeed the stationary distribution associated with $\mathbf{A}(t)$, so that $\boldsymbol{\pi}^{(0)}(t)$ is the PSA. However, to calculate the higher-order terms $\boldsymbol{\pi}^{(k)}(t)$

for $k \geq 1$, we need the derivatives of $\boldsymbol{\pi}^{(k)}(t)$ for $k \geq 0$, but these derivatives can also be calculated by solving Poisson equations. For example, by differentiating (1.2) for $k = 0$, we see that

$$\frac{d\boldsymbol{\pi}^{(0)}}{dt}(t)\mathbf{A}(t) = -\boldsymbol{\pi}^{(0)}(t)\frac{d\mathbf{A}}{dt}(t) \ . \tag{1.5}$$

Similarly,

$$\frac{d\boldsymbol{\pi}^{(1)}}{dt}(t)\mathbf{A}(t) = \frac{d^2}{dt^2}\boldsymbol{\pi}^{(0)}(t) - \boldsymbol{\pi}^{(1)}(t)\frac{d}{dt}\mathbf{A}(t) \tag{1.6}$$

and

$$\frac{d^2\boldsymbol{\pi}^{(0)}}{dt^2}(t)\mathbf{A}(t) = -2\frac{d\boldsymbol{\pi}^{(0)}}{dt}(t)\frac{d\mathbf{A}}{dt}(t) - \boldsymbol{\pi}^{(0)}(t)\frac{d^2\mathbf{A}}{dt^2}(t) \ . \tag{1.7}$$

More generally (by induction),

$$\frac{d^j\boldsymbol{\pi}^{(n)}}{dt^j}(t)\mathbf{A}(t) = \frac{d^{j+1}\boldsymbol{\pi}^{(n-1)}}{dt^{j+1}}(t) - \sum_{k=0}^{j-1}\binom{j}{k}\frac{d^k\boldsymbol{\pi}^{(n)}}{dt^k}(t)\frac{d^{j-k}\mathbf{A}}{dt^{j-k}}(t) \ , \tag{1.8}$$

where $\boldsymbol{\pi}^{(-i)}(t) \equiv 0$. In order for the $\boldsymbol{\pi}^{(n)}$ vectors to have the required derivatives and for the UA approximation to be well defined, we assume that the generator possesses all the required derivatives at t.

We remark that the derivatives of $\boldsymbol{\pi}^{(0)}(t)$ are also of interest in the *sensitivity analysis* of the stationary distribution to changes in the generator $\mathbf{A}(t)$, now regarded as the generator for a stationary CTMC subject to possible change in the transition intensities. It is known that such sensitivity analysis can be performed by solving Poisson's equation; e.g., see Schweitzer [23].

Thus, to calculate the first $n+1$ terms $\boldsymbol{\pi}^{(0)}(t), \boldsymbol{\pi}^{(1)}(t), \ldots, \boldsymbol{\pi}^{(n)}(t)$ in the UA approximation (1.1), we need to solve Poisson's equation $(n+1)(n+2)/2$ times with different (known) righthand sides. In (1.1) we are not interested in large n, because the full series is not a convergent series. It is instead an asymptotic expansion; see (3.12) below. Indeed, higher-order terms are likely to be sensitive to fine structure, so we only want to consider small n in (1.1). Since we are primarily interested in small n, e.g. $1 \leq n \leq 4$, it is essentially no more difficult to calculate the UA approximation than it is to calculate the PSA, i.e., to calculate the steady-state distribution for a single generator. We can use the same algorithm each time we need to solve Poisson's equation.

As reviewed in Whitt [27], there are efficient algorithms for solving Poisson's equation. For finite-state CTMCs, the solution of Poisson's equation can be expressed explicitly in terms of the fundamental matrix, but that is usually not the best way to proceed computationally. For birth-and-death processes and skip-free CTMC's, the solution can easily be calculated recursively; see Remark 1 on p. 287 of Whitt [27]. For more on numerical methods for solving Markov chains, see Stewart [25]. Thus, it tends to be much easier to compute the UA approximation for some time point t than it is to numerically solve the time-dependent differential equation. The UA approximation allows us to focus on a single time point t without having to calculate the probability vectors at previous time points.

However, for time-dependent birth-and-death processes and other relatively simple time-dependent one-dimensional CTMCs, it is actually not extraordinarily difficult to numerically solve for the complete time-dependent distribution (even though this is not often done in practice). Indeed, we do so to evaluate the performance of UA approximations. From that more sophisticated computational perspective, PSA and the UA approximations become more important to analyze larger time-dependent systems, such as time-dependent queueing networks and loss networks. We are unaware of any attempts to calculate the actual time-dependent distributions of such nonstationary networks with more than two or three nodes.

The UA approximation might also be useful for hybrid numerical schemes. We might opt to solve the Kolmogorov equations, but only in an interval $[t-\Delta, t]$ before a time t of interest. We could then use the UA approximation for the initial distribution at time $t-\Delta$.

We also gain insight without performing any calculations. We can see that PSA is asymptotically correct as the arrival rates change more slowly. To see this, suppose that the derivative $\mathbf{A}'(t)$ depends on a parameter γ by $\mathbf{A}'_\gamma(t) = \gamma \mathbf{A}'_1(t)$. From (1.5), we see that then $\boldsymbol{\pi}^{(0)\prime}_\gamma(t) = \gamma \boldsymbol{\pi}^{(0)\prime}_1(t)$. Then, from (1.2) and (1.4) for $k=1$, we see that $\boldsymbol{\pi}^{(1)}_\gamma(t) = \gamma \boldsymbol{\pi}^{(1)}_1(t)$, so that $\boldsymbol{\pi}^{(1)}_\gamma(t) \to 0$ as $\gamma \to 0$. Similar reasoning applies to higher order terms.

In practice, we rarely know high derivatives of the time-dependent generator $\mathbf{A}(t)$ precisely. Thus, we may wish to approximate the time-dependent generator $\mathbf{A}(t)$ at a given time point t by a linear or quadratic function. Then we can use estimation procedures as in Massey, Parker and Whitt [15] to estimate the parameters of the approximating generator from data. When $\mathbf{A}(t)$ is linear, (1.8) simplifies to

$$\frac{d^j \boldsymbol{\pi}^{(n)}}{dt^j}(t)\mathbf{A}(t) = \frac{d^{j+1}\boldsymbol{\pi}^{(n-1)}}{dt^{j+1}}(t) - j\frac{d^{j-1}\boldsymbol{\pi}^{(n)}}{dt^{j-1}}(t)\frac{d\mathbf{A}}{dt}(t) \ . \tag{1.9}$$

We have yet to explain where the UA approximations in (1.1)–(1.9) come from. They are derived from UA asymptotic expansions, which we will explain in the rest of this paper. UA expansions were first developed and applied to the time-dependent M/M/1 queue, which we denote by $M_t/M_t/1$ queue, by Massey [13], Keller [11] and Massey [14]. The UA technique was developed to create a mathematical framework that justified and refined the analysis of time-dependent queues by Newell [21]. That application dramatically shows the insights provided by the UA expansion. Under appropriate regularity conditions, the two-term UA approximations for the mean and the probability that the server is busy at time t are

$$\mathsf{E}[Q(t)] \approx \frac{\rho(t)}{1-\rho(t)} - \frac{\rho'(t)(1+\rho(t))}{\mu(t)(1-\rho(t))^4} \tag{1.10}$$

and

$$\mathsf{P}(Q(t)>0) \approx \rho(t) - \frac{\rho'(t)}{\mu(t)(1-\rho(t))^2} \ , \tag{1.11}$$

where $\rho(t) \equiv \lambda(t)/\mu(t)$ with $\lambda(t)$ being the arrival rate and $\mu(t)$ the service rate.

To have (1.10) and (1.11), we need to assume that $\rho(t) < 1$ and more: for stability at t, we need to assume that $\rho^*(t) < 1$, where

$$\rho^*(t) = \sup_{0 \le s < t} \frac{\int_s^t \lambda(u)du}{\int_s^t \mu(u)du} \ ; \tag{1.12}$$

see Massey [14].

Consistent with intuition, formulas (1.10) and (1.11) show that PSA tends to overestimate (underestimate) congestion when the time-dependent traffic intensity $\rho(t)$ is increasing (decreasing). Reasoning heuristically, in this context we might judge PSA (the first terms) and the UA refinements to be good approximations if each succeeding term is no more than 10% of the preceding term. From (1.10) and (1.11), we can see when this occurs for different functions $\lambda(t)$ and $\mu(t)$.

The fourth power of $1-\rho(t)$ in the denominator of the second term in (1.10) is reminiscent of the fourth power in the asymptotic variance of the sample mean for the stationary M/M/1 queue; see Example 1 on p. 281 of Whitt [27]. The similar result can be understood by the role of Poisson's equation in both contexts.

It is also possible to analyze the $M_t/M_t/1$ queue in more detail when it is unstable, as shown by Mandelbaum and Massey [12]. They combine the UA analysis with strong approximations to establish functional strong laws of large numbers and functional central limit theorems.

The UA expansions have also been applied to the $M_t/G/\infty$ queue and networks of such infinite-server queues by Eick, Massey and Whitt [4] and Massey and Whitt [16], [17], p. 323. The linearity of infinite-server models produces very appealing simple formulas for the UA expansion. As shown in Massey and Whitt [20], these formulas provide useful insight into the behavior of associated finite-server systems, with or without additional waiting space.

Our goal here is to develop the UA theory for finite-state CTMCs in general. In order to motivate the UA expansion, we start in Section 2 by reviewing the constant-rate CTMC theory. In Section 3 we introduce UA expansions and state three fundamental CTMC theorems that generalize basic results for the constant-rate case. There is some overlap between our theoretical results in Section 3 and a recent paper by Khasminskii, Yin and Zhang [10], which we discovered after completing our work. The proofs are quite different and Khasminskii et al. do not discuss computation; e.g., they do not discuss (1.1) - (1.8). We also take a more probabilistic approach to UA analysis and explore further its significance to time-varying Markov chains. In Section 4 we work out an explicit example of uniform acceleration for the special case of a two-state CTMC. In Section 5 we obtain results for time-varying finite-state birth-and-death processes. As an illustrative numerical example, in Section 6 we apply the UA theory to treat the time-dependent Erlang loss model. Finally, in Section 7 we give proofs of the theorems in Sections 3 and 5.

2 Background on the Constant-Rate Case

One of our goals is to relate the UA expansions for nonstationary CTMCs to classical theorems for stationary CTMCs. Hence in this section we briefly review the theory for stationary CTMCs. Let $\{\, Q(t) \mid t \geq 0 \,\}$ be a time-homogeneous CTMC with state space $\{0, 1, \ldots, \ell\}$. For all $t \geq 0$, we represent the distribution of $Q(t)$ as a *probability vector* $\mathbf{p}(t)$, where

$$\mathbf{p}(t) \equiv \sum_{i=0}^{\ell} \mathsf{P}(Q(t) = i)\mathbf{e}_i, \tag{2.1}$$

with $\mathbf{e}_i$ being the i-th unit basis vector. The Kolmogorov *forward equations* for the distribution of $Q(t)$ can be written as

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\mathbf{A}, \tag{2.2}$$

where the off-diagonal terms of the *operator* (square matrix) $\mathbf{A}$ are nonnegative and interpreted as transition rates: a_{ij}, the (i,j) entry of $\mathbf{A}$, is the instantaneous transition rate from i to j when $i \neq j$. The i-th diagonal term satisfies

$$a_{ii} = -a_i \equiv - \sum_{j:j\neq i} a_{ij}. \tag{2.3}$$

Thus a_i is the reciprocal of the mean holding time in state i. The operator $\mathbf{A}$ is called the *transition rate matrix* or the *infinitesimal generator* of the CTMC $\{\, Q(t) \mid t \geq 0 \,\}$. We assume that it is irreducible; i.e., with positive probability you can get from any state to any other state.

Using the matrix exponential, we can write the solution for $\mathbf{p}(t)$ as

$$\mathbf{p}(t) = \mathbf{p}(0) \exp(t\mathbf{A}). \tag{2.4}$$

In this setting there are three fundamental theorems for time-homogeneous or constant-rate CTMC transition probabilities. The first theorem is the *uniformization property*; e.g., see Chapter 2 of Keilson [9].

Theorem 2.1 (Uniformization) *If* $\lambda > \sup_{0\leq i\leq \ell} a_i$, *then* $\mathbf{P}_\lambda \equiv \mathbf{I} + \mathbf{A}/\lambda$ *is an aperiodic, stochastic matrix and*

$$\exp(t\mathbf{A}) = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \mathbf{P}_\lambda^n. \tag{2.5}$$

Formula 2.5 says that the random sample paths of a CTMC can be represented as a discrete-time Markov chain (DTMC) with single-step transition matrix $\mathbf{P}_\lambda$, with the discrete steps occurring at the jump times of a time-homogeneous Poisson process having rate λ.

Since the generator $\mathbf{A}$ is irreducible, there exists a unique probability vector $\boldsymbol{\pi}$ such that

$$\boldsymbol{\pi}\mathbf{A} = \mathbf{0}. \tag{2.6}$$

Also note that (2.2) and (2.6) are special cases of *Poisson's equation*. In general, if we are given a vector $\mathbf{y}$ whose components sum to zero, then there exists a vector $\mathbf{x}$ such that

$$\mathbf{y} = \mathbf{x}\mathbf{A}. \tag{2.7}$$

Moreover, $\mathbf{x}$ is unique up to the vector addition of some scalar multiple of $\boldsymbol{\pi}$. This is immediate (see [27]) when we write the solution as

$$\mathbf{x} = \mathbf{y} \int_0^\infty \Big(\mathbf{1}^\mathsf{T}\boldsymbol{\pi} - \exp(t\mathbf{A}) \Big) dt + (\mathbf{x} \cdot \mathbf{1}^\mathsf{T})\boldsymbol{\pi}. \tag{2.8}$$

The second theorem is the *stationarity property*.

Theorem 2.2 (Stationarity) *If* $\mathbf{p}(0) = \boldsymbol{\pi}$, *then* $\mathbf{p}(t) = \boldsymbol{\pi}$ *for all* $t > 0$.

Theorem 2.2 states that if the CTMC is initialized to be in equilibrium, then it stays in equilibrium for all time. (For the basic CTMC chain theory here, see Chung [2].)

Finally, the third theorem is the *ergodic property*.

Theorem 2.3 (Ergodicity) *For* any *initial probability vector* $\mathbf{p}(0)$,

$$\lim_{t\to\infty} \mathbf{p}(t) = \boldsymbol{\pi}. \tag{2.9}$$

In the next section we develop time-varying analogs of these three theorems.

3 Uniform Acceleration Expansions

Now let $\{ Q(t) \mid t \geq 0 \}$ be a CTMC with time-varying rates, again with state space $\{0, 1, \ldots, \ell\}$. If $\mathbf{p}(t)$ is the probability vector for the distribution of $Q(t)$, then instead of (2.2) we have

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\mathbf{A}(t), \tag{3.1}$$

where we now have a family of generators $\{ \mathbf{A}(t) \mid t \geq 0 \}$, which we assume is measurable (with respect to the Borel field on $[0, \infty)$) and bounded as a function of time. Just like (2.2), (2.6), and (2.7), (3.1) is a form of Poisson's equation, but we do not know the derivative $\frac{d}{dt}\mathbf{p}(t)$. Thus the exact solution of (3.1) corresponds the solving a system of ordinary differential equations.

In terms of a formal solution, we can express $\mathbf{p}(t)$ as

$$\mathbf{p}(t) = \mathbf{p}(0)\mathbf{E}_{\mathbf{A}}(t), \tag{3.2}$$

where $\mathbf{E}_{\mathbf{A}}(t)$ is the *time-ordered exponential* of $\{ \mathbf{A}(s) \mid 0 \leq s \leq t \}$; see Dollard and Friedman [3]. The time-ordered exponential $\mathbf{E}_{\mathbf{A}}(t)$ is the unique matrix solution to

$$\frac{d}{dt}\mathbf{E}_{\mathbf{A}}(t) = \mathbf{E}_{\mathbf{A}}(t)\mathbf{A}(t) \tag{3.3}$$

and $\mathbf{E}_{\mathbf{A}}(0) = \mathbf{I}$, the identity matrix. We must define $\mathbf{E}_{\mathbf{A}}(t)$ in this manner since in general

$$\mathbf{E}_{\mathbf{A}}(t) \neq \exp\left(\int_0^t \mathbf{A}(s)ds\right). \tag{3.4}$$

We now want to give a generalization of the uniformization formula (2.5). Just as in the constant-rate case, the time-inhomogeneous CTMC can be represented as a time-inhomogeneous DTMC where the discrete steps occur at the jump times of a homogeneous Poisson process with rate λ. The DTMC will have a transition matrix $\mathbf{P}_{\lambda}(s) \equiv \mathbf{I} + \mathbf{A}(s)/\lambda$ if the jump is at time s. Since the generators are assumed to be bounded, we can find one rate λ uniformizing the transition rates at all times.

We now want to find an expression for the time-dependent transition probability matrix, i.e., the time-ordered exponential $\mathbf{E}_{\mathbf{A}}(t)$. This can be done by averaging over the number of points in $[0, t]$ and their location. The number of points has a Poisson distribution. Conditional on there being exactly n such points in $[0, t]$, they are distributed as the order statistics of n i.i.d. random variables uniformly distributed over $[0, t]$. Given that there are n jumps at the points $s_1, \ldots, s_n$, the transition probability is $\mathbf{P}_{\lambda}(s_1)\mathbf{P}_{\lambda}(s_2)\ldots\mathbf{P}_{\lambda}(s_n)$. Hence, we obtain the following formula, which will play a key role in our convergence proof for UA expansions. Later, we give a direct analytical proof of this formula. The analytical proofs show that Theorem 3.1 extends beyond the probabilistic setting in which $\mathbf{A}(t)$ is a generator.

Theorem 3.1 (Time-Varying Uniformization) *Let $\{ \mathbf{A}(s) \mid 0 \leq s \leq t^* \}$ be a bounded family of generators. If λ is a positive constant with*

$$\lambda > \sup_{0 \leq s \leq t^*} \sup_{0 \leq i \leq \ell} a_i(s) \tag{3.5}$$

then for all $t \in [0, t^*)$,

$$\mathbf{E_A}(t) = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \int_{0 \le s_1 \le \cdots \le s_n \le t} \cdots \int \prod_{i=1}^{n} \mathbf{P}_\lambda(s_i) \; \frac{n!}{t^n} ds_1 \cdots ds_n, \tag{3.6}$$

where $\mathbf{P}_\lambda(s) \equiv \mathbf{I} + \mathbf{A}(s)/\lambda$ *is a stochastic matrix.*

For a stationary model (with constant transition rates), we typically focus on the long-run limiting behavior. We do so, not because we are interested in the distant future, but because we believe the limiting behavior will adequately describe the present or near future. We anticipate that the system of interest will currently be in equilibrium or steady state. The fact that the limiting distribution is also the equilibrium distribution gives us a reasonable model of current and near future behavior.

With time-varying rates, we could also focus on long-run limiting behavior. For a "stable" system with time-varying rates, long-run limits of subsequences and averages describe the range of excursions over time and the long-run average behavior; e.g., see Heyman and Whitt [8]. For an unstable system, long-run limits describe the way the stochastic processes grow as time evolves; e.g., see Massey and Whitt [18].

However, with time-varying rates, limits as $t \to \infty$ tend to be less relevant for describing the present or near future. If we let $t \to \infty$ within this situation, then we would be approximating the current behavior of the system with arrival and service rates that have not yet happened. Hence, it is natural to consider a different kind of asymptotics.

Now suppose that the rates of the generator $\mathbf{A}(t)$ vary so slowly in time that the process $Q(t)$ can achieve equilibrium before there is any significant change in the rates. We can formalize this by changing $\mathbf{A}(t)$ to $\mathbf{A}(\epsilon t)$, where $\epsilon > 0$. (In replacing $\mathbf{A}(t)$ by $\mathbf{A}(\epsilon t)$ we are focusing on the behavior in the neighborhood of time 0. If we wanted to focus on the behavior in the neighborhood of time t_0, then we would replace $\mathbf{A}(t)$ by $\mathbf{A}(t_0 + \epsilon t)$.) We will let $\mathbf{p}(t;\epsilon)$ be the corresponding probability vector that solves the differential equation

$$\frac{d}{dt}\mathbf{p}(t;\epsilon) = \mathbf{p}(t;\epsilon)\mathbf{A}(\epsilon t). \tag{3.7}$$

When $\epsilon = 1$, we have our original process, but as $\epsilon \downarrow 0$, we have a process with *slowly varying* rates. In fact if we let $\epsilon = 0$, then we could simply let $t \to \infty$ and use the steady-state analysis associated with the generator $\mathbf{A}(0)$, i.e., we would be using PSA at time $t = 0$.

In the formalism of asymptotic expansions, this would give us the leading term of the *inner expansion* for the transition probabilities. What we will call uniform acceleration will correspond to the *outer expansion*. (For a discussion of inner and outer limits in the context of boundary layer theory, we refer the reader to Chapter 9 of Bender and Orszag [1].) We still take the limit as $t \to \infty$, but we simultaneously let $\epsilon \downarrow 0$ such that $\tau = \epsilon t$ for some fixed $\tau > 0$. In effect we are holding the time scale for $\mathbf{A}$ fixed as we obtain a steady-state limit for Q. If we now switch to the time scale of τ, then the probability vector $\mathbf{p}(\tau;\epsilon)$ will solve the forward equations

$$\epsilon\frac{d}{d\tau}\mathbf{p}(\tau;\epsilon) = \mathbf{p}(\tau;\epsilon)\mathbf{A}(\tau). \tag{3.8}$$

Note that (3.8) corresponds to uniformly accelerating the rates, i.e., replacing $\mathbf{A}(t)$ by $\mathbf{A}(t)/\epsilon$ in (3.1). What we have done is to switch from the time scale of the Markov chain to the time scale of the generators. Observe that if $\mathbf{A}$ is constant over time, then we can write the solution of (3.8) as

$$\mathbf{p}(\tau;\epsilon) = \mathbf{p}(0)\exp\left(\frac{\tau}{\epsilon}\mathbf{A}\right) \tag{3.9}$$

and we see that in the constant-rate case, the uniform acceleration limit of taking $\epsilon \downarrow 0$ gives exactly the same results as the steady-state limit of $\tau \to \infty$.

We remark that Khasminskii et al. [10] study both the inner and outer expansions. However, unlike for the outer expansion, the inner expansion seems to offer no computational advantage in this setting over solving for the time-dependent probabilities in the original system.

We can use the uniformization expansion to give a probabilistic interpretation of uniform acceleration. First, in general the solution to (3.8) is

$$\mathbf{p}(\tau;\epsilon) = \mathbf{p}(0)\mathbf{E}_{\mathbf{A}}(\tau;\epsilon), \tag{3.10}$$

where $\mathbf{E}_{\mathbf{A}}(\tau;\epsilon)$ is the time-ordered exponential associated with $\mathbf{A}(\tau)/\epsilon$. Now if we use a Poisson process with rate λ/ϵ to perform the uniformization, we obtain by Theorem 3.1

Corollary 3.2 *Under the hypotheses of Theorem 3.1, we have for all $\epsilon > 0$*

$$\mathbf{E}_{\mathbf{A}}(\tau,\epsilon) = \sum_{n=0}^{\infty} \frac{e^{-\lambda\tau/\epsilon}(\lambda\tau/\epsilon)^n}{n!} \int_{0\le s_1\le\cdots\le s_n\le\tau}\!\!\cdots\int \prod_{i=1}^{n}\mathbf{P}_\lambda(s_i)\,\frac{n!}{\tau^n}ds_1\cdots ds_n. \tag{3.11}$$

Hence the parameter ϵ only appears in the uniformizing Poisson rate λ/ϵ and not in the DTMC stochastic matrices $\mathbf{P}(s)$.

We say that $\mathbf{p}(\epsilon)$ is an *asymptotic probability vector* if

$$\mathbf{p}(\epsilon) \cong \sum_{n=0}^{\infty}\epsilon^n\mathbf{p}_n \quad \text{as } \epsilon\downarrow 0, \tag{3.12}$$

which is shorthand for

$$\lim_{\epsilon\downarrow 0}\frac{\mathbf{p}(\epsilon) - [\mathbf{p}_0 + \epsilon\mathbf{p}_1 + \cdots + \epsilon^n\mathbf{p}_n]}{\epsilon^{n+1}} = \mathbf{p}_{n+1} \tag{3.13}$$

for all non-negative integers n. In (3.12), $\mathbf{p}_0$ is a probability vector, but $\mathbf{p}_n$ satisfies

$$\mathbf{p}_n\mathbf{1}^\top = 0 \quad \text{for all } n\ge 1, \tag{3.14}$$

where $\top$ denotes matrix transpose. As noted in the introduction, the series in (3.12) need not be convergent for any $\epsilon > 0$.

We state the remaining two theorems here and prove them in the final section. First, we will say that $\{\, \mathbf{A}(s) \mid 0 \le s \le \tau \,\}$ is a smooth family of operators if every entry of $\mathbf{A}(\cdot)$ is an infinitely differentiable function of time.

Theorem 3.3 (Time-Varying Stationarity) *Suppose that $\{\, \mathbf{A}(s) \mid 0 \le s \le \tau^* \,\}$ is a smooth family of irreducible Markov generators.*

1. *For all $\tau \in [0, \tau^*)$, we can construct a unique sequence of vectors $\boldsymbol{\pi}^{(n)}(\tau)$ for $n = 0, 1, \ldots$ which are solutions to the following set of Poisson's equations:*

$$\boldsymbol{\pi}^{(0)}(\tau)\mathbf{A}(\tau) = \mathbf{0}, \tag{3.15}$$

where $\boldsymbol{\pi}^{(0)}(\tau) \cdot \mathbf{1}^{\mathsf{T}} = 1$ and

$$\boldsymbol{\pi}^{(n)}(\tau)\mathbf{A}(\tau) = \frac{d}{d\tau}\boldsymbol{\pi}^{(n-1)}(\tau), \quad \text{for all } n \geq 1, \tag{3.16}$$

where $\boldsymbol{\pi}^{(n)}(\tau) \cdot \mathbf{1}^{\mathsf{T}} = 0$.

2. *Let $\mathbf{p}(\epsilon)$ be an asymptotic probability vector of the form*

$$\mathbf{p}(\epsilon) \cong \sum_{n=0}^{\infty} \epsilon^n \boldsymbol{\pi}^{(n)}(0) \quad \text{as } \epsilon \downarrow 0, \tag{3.17}$$

where $\boldsymbol{\pi}^{(n)}(0)$ satisfies (3.15) and (3.16) for $\tau = 0$, and let $\mathbf{p}(\tau;\epsilon)$ be the unique probability vector that solves the forward equations (3.8) with $\mathbf{p}(0;\epsilon) = \mathbf{p}(\epsilon)$.

Then, for all $\tau \in [0, \tau^*)$, $\mathbf{p}(\tau;\epsilon)$ is also an asymptotic probability vector of the form,

$$\mathbf{p}(\tau;\epsilon) \cong \sum_{n=0}^{\infty} \epsilon^n \boldsymbol{\pi}^{(n)}(\tau). \quad \text{as } \epsilon \downarrow 0, \tag{3.18}$$

where $\boldsymbol{\pi}^{(n)}$ satisfies (3.15) and (3.16).

As we will see later, the proof of this theorem is straightforward. It is analogous to stationary behavior in the sense that the vector functions $\boldsymbol{\pi}^{(n)}$, which comprise the coefficients of the asymptotic probability vector, when used initially (at time 0) will give the future asymptotic expansion terms for all time. Below, we show that no other collection of vector functions has this property and the initial distribution used has no relevance to the asymptotic behavior of the transition probabilities.

Theorem 3.4 (Time-Varying Ergodicity) *Let $\{ \mathbf{A}(s) \mid 0 \leq s \leq \tau^* \}$ be a smooth family of generators, let $\mathbf{p}(\tau;\epsilon)$ be the probability vector that solves (3.8) for $\tau \in [0, \tau^*)$, and let $\mathbf{p}(0;\epsilon)$ be any asymptotic probability vector. Then, for all $\tau \in [0, \tau^*)$, (3.18) holds where $\boldsymbol{\pi}^{(n)}$ satisfies (3.15) and (3.16).*

In Theorem 3.4 we have assumed that $\mathbf{p}(0;\epsilon)$ is an asymptotic probability vector. This assumption includes an ordinary probability vector as a special case. Then $\mathbf{p}_n = 0$ for $n \geq 1$ in (3.12). The fact that the initial distribution has no impact on the UA expansion may be disturbing. The idea is that the relevant history before t is captured by $\mathbf{A}(t)$ and the derivatives of $\mathbf{A}$ at t. Previous time-varying ergodicity results are contained in Zeifman [28] and Khasminskii et al. [10].

The UA asymptotic expansion based of Theorems 3.3 and 3.4 is (3.18). The associated UA approximation is the first n terms of (3.18) for some n. In our UA approximations for a fixed time-dependent generator $\mathbf{A}(t)$, we simply set $\epsilon = 1$ in the UA asymptotic expansion as in (1.1). If we want to see how the quality of the UA approximations improve as $\epsilon \downarrow 0$, then we can start with a fixed generator $\mathbf{A}(t)$ and consider the family of CTMC's indexed by ϵ with infinitesimal generators $\mathbf{A}(t)/\epsilon$. Note that the approximations with the pair $(\mathbf{A}(t)/\epsilon, \epsilon)$ are independent of ϵ, so that there is no loss of generality in letting $\epsilon = 1$ above. With $\epsilon = 1$ set, the UA approximation will perform better when $\mathbf{A}(t)$ changes more slowly.

4 Time-Varying Two-State Markov Chain

To illustrate the behavior of UA asymptotics, consider a two-state Markov chain. Its forward equations are

$$\frac{d}{dt}p_0(t) = a_1(t)p_1(t) - a_0(t)p_0(t) \tag{4.1}$$
$$\frac{d}{dt}p_1(t) = a_0(t)p_0(t) - a_1(t)p_1(t). \tag{4.2}$$

Since $p_0(t)+p_1(t)=1$, we can write $p_1(t)$ as the solution of an ordinary differential equation of degree 1 and obtain

$$p_1(t) = \int_0^t a_0(s)\, \exp\left(-\int_s^t [a_0+a_1](r)dr\right) ds + p_1(0)\exp\left(-\int_0^t [a_0+a_1](s)ds\right). \tag{4.3}$$

By induction, we can write $p_1(t)$ as

$$\begin{aligned} p_1(t) = &\sum_{n=0}^{m}\left[\frac{D^n a_0(t)}{a_0(t)+a_1(t)} - \frac{D^n a_0(0)}{a_0(0)+a_1(0)}\exp\left(-\int_0^t [a_0+a_1](s)ds\right)\right] \\ &+\int_0^t D^{m+1}a_0(s)\, \exp\left(-\int_s^t [a_0+a_1](r)dr\right) ds + p_1(0)\exp\left(-\int_0^t [a_0+a_1](s)ds\right), \end{aligned} \tag{4.4}$$

for all integers $m \geq 0$, where D is the differential operator

$$Df(x) \equiv -\frac{d}{dx}\frac{f(x)}{a_0(x)+a_1(x)} \tag{4.5}$$

and D^0 is the identity operator.

Now, if we apply the UA scaling, then we obtain

$$\begin{aligned} p_1(t;\epsilon) = &\sum_{n=0}^{m}\epsilon^n\left[\frac{D^n a_0(t)}{a_0(t)+a_1(t)} - \frac{D^n a_0(0)}{a_0(0)+a_1(0)}\exp\left(-\frac{1}{\epsilon}\int_0^t [a_0+a_1](s)ds\right)\right] \\ &+\epsilon^m\int_0^t D^{m+1}a_0(s)\, \exp\left(-\frac{1}{\epsilon}\int_s^t [a_0+a_1](r)dr\right) ds \\ &+p_1(0;\epsilon)\exp\left(-\frac{1}{\epsilon}\int_0^t [a_0+a_1](r)dr\right). \end{aligned} \tag{4.6}$$

Hence in the limit as $\epsilon \downarrow 0$, we obtain the following UA expansion,

$$p_1(t;\epsilon) \cong \sum_{n=0}^{\infty}\epsilon^n\frac{D^n a_0(t)}{a_0(t)+a_1(t)}. \tag{4.7}$$

Notice that $p_1(0;\epsilon)$ does not appear in the UA expansion of $p_1(t;\epsilon)$. Given the explicit form of the solution for $p_1(t,\epsilon)$ in (4.6) we see that $p_1(t;\epsilon)$ is *not* an analytic function of ϵ as ϵ approaches 0, because an expression like $\exp(-1/\epsilon)$ will have an essential singularity at $\epsilon = 0$.

5 Time-Varying Finite-State Birth-Death Models

In this section we give some results for the special case of finite-state birth-death processes (with time-varying rates). The following complements characterizations of the solutions of Poisson's equation in [27].

Proposition 5.1 *Let $\mathbf{A}$ be the generator for a birth-death process on the state space $\{0,1,\ldots,\ell\}$ with birth rate vector $\boldsymbol{\lambda} = [\lambda(0),\ldots,\lambda(\ell-1),0]$ and death rate $\boldsymbol{\mu} = [0,\mu(1),\ldots,\mu(\ell)]$. If $\mathbf{y} \equiv [y(0),y(1),\ldots,y(\ell)]$ is an arbitrary $(\ell+1)$-dimensional vector whose components sum to zero, and $\mathbf{x} \equiv [x(0),x(1),\ldots,x(\ell)]$ is the unique solution to Poisson's equation $\mathbf{xA} = \mathbf{y}$ where the components of $\mathbf{x}$ sum to zero also, then*

$$x(\ell) = \beta(\ell) \sum_{i=0}^{\ell-1} \frac{y^*(i)}{\lambda(i)\beta(i)}, \tag{5.1}$$

where $y^*(i) \equiv y(0) + \cdots + y(i)$ and $\beta(i)$ is the blocking probability (probability of i) for the embedded birth-death process on the smaller state space $\{0,1,\ldots,i\}$ which can be expressed as

$$\beta(i) = \prod_{k=0}^{i-1} \frac{\lambda(k)}{\mu(k+1)} \Big/ \sum_{j=0}^{i} \prod_{k=0}^{j-1} \frac{\lambda(k)}{\mu(k+1)} \,. \tag{5.2}$$

Corollary 5.2 *Using the previous hypothesis, consider birth-death processes with rates*

$$\lambda(t;i) = \lambda(t) r_i \quad \text{and} \quad \mu(t;i) = \mu(t) s_i \tag{5.3}$$

for all $i = 0,1,\ldots,\ell$, where $\lambda(t)$, $\mu(t)$, r_i, and s_i are all non-negative quantities. If we set $\rho(t) \equiv \lambda(t)/\mu(t)$, then as $\epsilon \downarrow 0$, we have

$$\mathsf{P}\left(Q(t;\epsilon) = i\right) = \pi_i^{(0)}(t;\ell) + O(\epsilon) \tag{5.4}$$

for all $i = 0,1,\ldots,\ell$, where

$$\pi_i^{(0)}(t;\ell) = \rho(t)^i \prod_{k=0}^{i-1} \frac{r_k}{s_{k+1}} \Big/ \sum_{j=0}^{\ell} \rho(t)^j \prod_{k=0}^{j-1} \frac{r_k}{s_{k+1}}. \tag{5.5}$$

Moreover,

$$\mathsf{P}\left(Q(t;\epsilon) = \ell\right) = \beta_\ell(\rho(t)) \left[1 - \epsilon \frac{\rho'(t)}{\lambda(t)\rho(t)} \sum_{i=0}^{\ell-1} \frac{q_\ell(\rho(t)) - q_i(\rho(t))}{r_i \beta_i(\rho(t))} \prod_{j=i+1}^{\ell} \overline{\beta}_j(\rho(t)) + O(\epsilon^2)\right], \tag{5.6}$$

where

$$\beta_i(\rho(t)) \equiv \pi_i^{(0)}(t;i), \ \overline{\beta}_i(\rho(t)) \equiv 1 - \beta_i(\rho(t)), \tag{5.7}$$

and

$$q_i(\rho(t)) \equiv \sum_{j=0}^{i} j \pi_j^{(0)}(t;i). \tag{5.8}$$

Note that as in (1.10) and (1.11), the second term of (5.6) is of the form $-\epsilon\rho'(t)c(t)$ for $c(t) > 0$.

The time-dependent Erlang model is the special case of a time-dependent birth-death process with birth rates $\lambda(t;k) \equiv \lambda(t)$ for $0 \leq k \leq \ell - 1$, and death rates $\mu(t;k) \equiv k\mu(t)$ for some non-negative functions $\lambda(t)$ and $\mu(t)$. Without loss of generality (by performing a deterministic time transformation), it suffices to let $\mu(t) = \mu$ for all $t \geq 0$. Hence we construct the time-dependent Erlang model by defining the off-diagonal terms for its family of generators $\{\, \mathbf{A}(t) \mid t \geq 0 \,\}$ to be, for all $i \neq j$,

$$a_{ij}(t) = \begin{cases} \lambda(t) & j = i+1 \text{ and } i < \ell, \\ i\mu & j = i-1 \text{ and } i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

The proof of the next result follows immediately from Corollary 5.2.

Corollary 5.3 *For the time-varying Erlang model, we have as $\epsilon \downarrow 0$,*

$$\mathsf{P}(Q(t;\epsilon) = \ell) = \beta_\ell(\rho(t)) \left[1 - \epsilon \frac{\rho'(t)}{\lambda(t)} \sum_{i=0}^{\ell-1} \left(1 - \frac{\beta_\ell(\rho(t))}{\beta_i(\rho(t))} \right) \prod_{j=i+1}^{\ell} \overline{\beta}_j(\rho(t)) + O(\epsilon^2) \right], \tag{5.10}$$

where for all $x \geq 0$, $\beta_i(x)$ is the Erlang blocking formula for i channels with offered load x, i.e.,

$$\beta_i(x) \equiv \frac{x^i}{i!} \Big/ \sum_{j=0}^{i} \frac{x^j}{j!}\,. \tag{5.11}$$

We can use the asymptotic expansion for $\mathsf{P}(Q(t;\epsilon) = \ell)$ to compute the corresponding expansion for $\mathsf{E}[Q(t;\epsilon)]$ by using the identity.

$$\mathsf{E}[Q(t;\epsilon)] = \rho(t)\Big(1 - \mathsf{P}(Q(t;\epsilon) = \ell)\Big) - \frac{\epsilon}{\mu(t)} \frac{d}{dt} \mathsf{E}[Q(t;\epsilon)]. \tag{5.12}$$

6 A Numerical Example

We now consider an example of the time-varying Erlang loss model. We show how the probability $\mathsf{P}(Q(t) = \ell)$ can be approximated by our UA asymptotics. To test how well the approximations work, we will compare them to $\mathsf{P}(Q(t) = \ell)$ itself, obtained by numerically integrating the forward equations (3.1) that the vector $\mathbf{p}(t)$ satisfies. This results in numerically solving $\ell + 1$ coupled ordinary differential equations over the time interval $[0, t]$. For the UA approximation, we exploit (1.1)-(1.8). In a subsequent paper, we will discuss in greater detail the computational aspects of UA expansions and how various numerical tricks can be used to efficiently compute UA expansions for time-varying birth-death processes. Such calculations are significantly faster than numerically integrating forward equations.

For our numerical example, the specific model parameters are $\lambda(t) = 15 + 5\sin(.5t)$, $\mu = 1$ and $\ell = 20$. We start the process at $t = 0$ and run it for 22 time units. We treat the first 12 time units as a warmup period and only plot the results for the final 10 time units. There is no special reason for the warmup period of 12; it seemed adequate to eliminate the

GRAPH KEY	FUNCTION PLOTTED
Exact	$\mathsf{P}(Q(12+\cdot)=\ell)$
Approx 0	$\pi_\ell^{(0)}(12+\cdot)$
Approx 1	$\pi_\ell^{(0)}(12+\cdot)+\pi_\ell^{(1)}(12+\cdot)$
Approx 2	$\pi_\ell^{(0)}(12+\cdot)+\pi_\ell^{(1)}(12+\cdot)+\pi_\ell^{(2)}(12+\cdot)$

Table 1: Key to graph labels

initial effect, but that is not necessary. Figure 1 compares the exact blocking probability to graphs of the first three terms of the UA expansion. (Approximation 2 is the dashed dark line, while the exact value is the solid dark line). The legend keys for the graph are given in Table 1.

Recall that the first term $\pi_\ell^{(0)}$ in the UA expansion is the PSA approximation. Hence Figure 1 also contains a comparison between the exact blocking probability and its PSA estimate. Now this estimate is merely the Erlang blocking formula, which is a monotone function applied to $\lambda(\cdot)$. Therefore the PSA approximation and the arrival rate function will have the same times for extreme values (times for maxima and minima.) We immediately see that PSA can fail as a good estimate of the exact blocking probability because it does not account for the natural lag between the times of peak blocking and peak arrivals. (See [20] for more discussion.) In the two-term UA approximation, we see that the lag disappears. Finally the three-term UA approximation corrects for the height of the approximate peak. This phenomenon does not always occur though. We intend to evaluate the quality of UA expansions as approximations in greater depth in a subsequent paper.

7 Proofs of Theorems

We now prove the theorems and corollaries in Sections 3 and 5. For this purpose, we introduce norms on vectors and matrices. If $\mathbf{x}$ is an $(\ell+1)$-dimensional row vector, then our choice of vector norms will be the ℓ_1-norm, i.e.,

$$|\mathbf{x}| \equiv \sum_{i=0}^{\ell} |x_i|. \tag{7.1}$$

If $\mathbf{p}$ is a probability vector, then $|\mathbf{p}|=1$. Our choice of operator norms will be the one induced by the ℓ_1-norm on row vectors, or

$$|\mathbf{A}| \equiv \sup_{|\mathbf{x}|\le 1} |\mathbf{x}\mathbf{A}| = \max_{0\le i\le \ell} \sum_{j=0}^{\ell} |a_{ij}|. \tag{7.2}$$

When $\mathbf{P}$ is a stochastic matrix, $|\mathbf{P}|=1$. So when $\{\, \mathbf{A}(s) \mid 0\le s\le t \,\}$ is a family of Markov generators, we have $|\mathbf{E}_{\mathbf{A}}(t)|=1$.

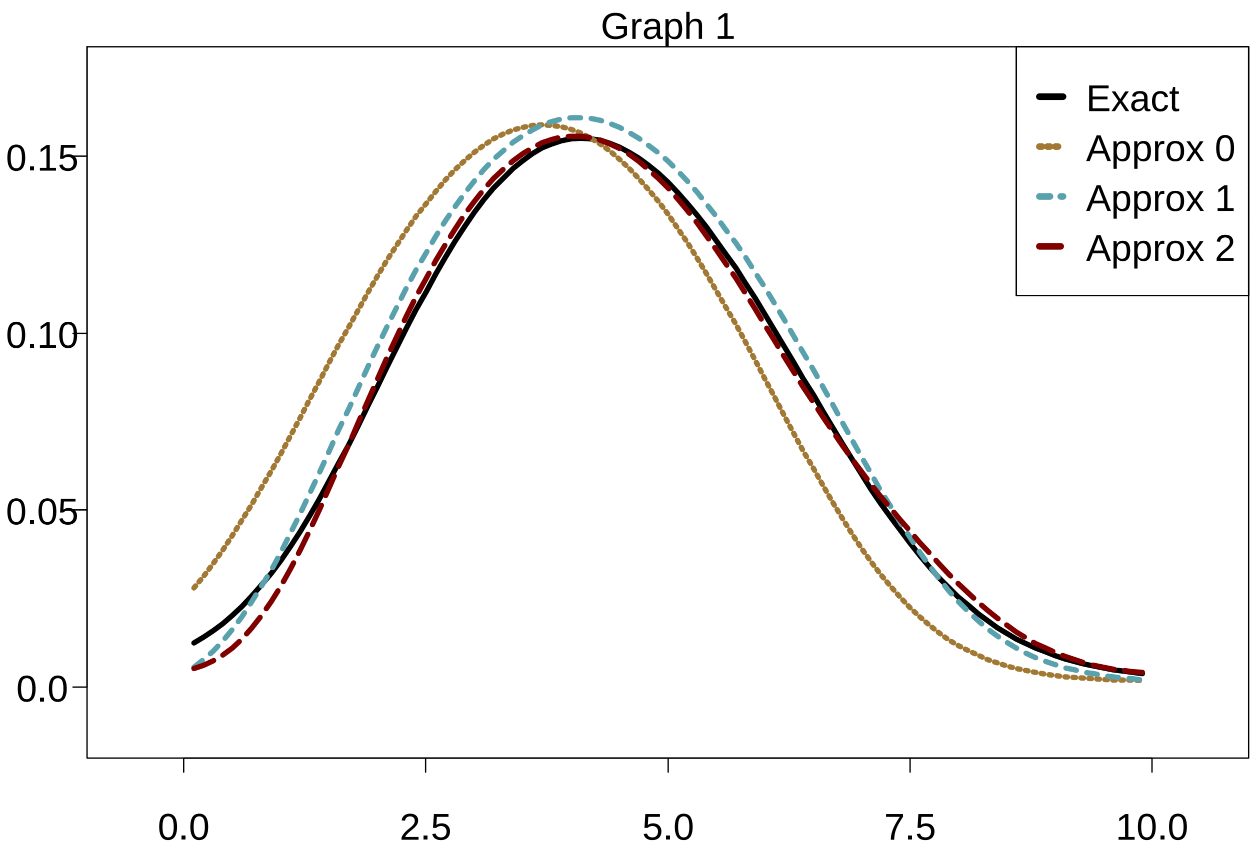

Figure 1: Comparing the exact distribution to one, two and three terms in the UA expansion.

We now give the proof of our time-varying analogue to the uniformization expansion. Our proof exploits the probabilistic argument, but the argument can be extended beyond probabilities; see Chapter 14 of Gantmacher [5].

Proof of Theorem 3.1: First we show that it is sufficient to have for all nonnegative integers m,

$$\begin{aligned}\mathbf{E}_{\mathbf{A}}(t) &= \sum_{n=0}^{m} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \int_{0\le s_1\le\cdots\le s_n\le t}\!\!\cdots\int \prod_{i=1}^{n} \mathbf{P}_\lambda(s_i)\ \frac{n!}{t^n} ds_1\cdots ds_n, \\ &+\lambda^{m+1} \int_{0\le s_1\le\cdots\le s_{m+1}\le t}\!\!\cdots\int e^{-\lambda(t-s_1)}\mathbf{E}_{\mathbf{A}}(s_1)\prod_{i=1}^{m}\mathbf{P}_\lambda(s_{i+1})\ ds_1\cdots ds_{m+1}. \end{aligned} \tag{7.3}$$

The theorem follows from (7.3) by taking the limit as $m \to \infty$, since

$$\left|\lambda^{m+1} \int_{0\le s_1\le\cdots\le s_{m+1}\le t}\!\!\cdots\int e^{-\lambda(t-s_1)}\mathbf{E}_{\mathbf{A}}(s_1)\prod_{i=1}^{m}\mathbf{P}_\lambda(s_{i+1})\ ds_1\cdots ds_{m+1}\right| \le \frac{(\lambda t)^{m+1}}{(m+1)!}. \tag{7.4}$$

The identity (7.3) can be shown by induction, making the observation that

$$\frac{d}{dt}\mathbf{E}_{\mathbf{A}}(t) = -\lambda\mathbf{E}_{\mathbf{A}}(t) + \lambda\mathbf{E}_{\mathbf{A}}(t)\mathbf{P}_\lambda(t) \tag{7.5}$$

yields the formula

$$\mathbf{E}_{\mathbf{A}}(t) = e^{-\lambda t}\mathbf{I} + \int_0^t \lambda e^{-\lambda(t-s)}\mathbf{E}_{\mathbf{A}}(s)\mathbf{P}_\lambda(s)ds \tag{7.6}$$

which completes the proof. ∎

Now we state and prove the following lemma and corollary to give a simple proof of the time-varying analogue to stationarity.

Lemma 7.1 *For any bounded family of generators* $\{\, \mathbf{A}(s) \mid 0 \le s \le t \,\}$, $\mathbf{E}_{\mathbf{A}}(t)$ *is an invertible matrix and*

$$\frac{d}{dt}\mathbf{E}_{\mathbf{A}}(t)^{-1} = -\mathbf{A}(t)\mathbf{E}_{\mathbf{A}}(t)^{-1}. \tag{7.7}$$

Proof: Let α be any positive constant strictly greater than $\sup_{0\le s\le t}|\mathbf{A}(s)|$. Integrating (3.3), we have for all s where $0 \le s \le t$,

$$\mathbf{E}_{\mathbf{A}}(s) = \mathbf{I} + \int_0^s \mathbf{E}_{\mathbf{A}}(r)\mathbf{A}(r)dr. \tag{7.8}$$

Now by Theorem 3.1 it is clear that $\mathbf{E}_{\mathbf{A}}(r)$ is always a stochastic matrix, and so

$$|\int_0^s \mathbf{E}_{\mathbf{A}}(r)\mathbf{A}(r)dr| \le \int_0^s |\mathbf{A}(r)|dr < \alpha s. \tag{7.9}$$

Since the spectral radius of an operator is always bounded above by its operator norm, by (7.8) and (7.9), we see that $\mathbf{E}_{\mathbf{A}}(s)$ will always be invertible whenever $0 \le s \le 1/\alpha$.

By induction on n it follows that $\mathbf{E}_{\mathbf{A}}(s)$ is invertible for all s belonging to the interval $[n/\alpha, (n+1)/\alpha]$ if $(n+1)/\alpha < t$. If we let $s = u + r$, then we can define by induction hypothesis

$$\hat{\mathbf{E}}(u) \equiv \mathbf{E}_{\mathbf{A}}(r)^{-1}\mathbf{E}_{\mathbf{A}}(u+r) \tag{7.10}$$

where $0 \leq u \leq 1/\alpha$ and $(n-1)/\alpha \leq r \leq n/\alpha$. Differentiating with respect to u, we obtain

$$\frac{d}{du}\hat{\mathbf{E}}(u) = \hat{\mathbf{E}}(u)\mathbf{A}(u+r) \quad \text{and} \quad \hat{\mathbf{E}}(0) = \mathbf{I}. \tag{7.11}$$

Now it follows that $\hat{\mathbf{E}}(u)$ is a time ordered exponential that is also invertible for all $u \leq 1/\alpha$. Thus $\mathbf{E}_{\mathbf{A}}(s)$ is invertible and (7.7) easily follows. ∎

Corollary 7.2 *Under the same hypothesis as Lemma 7.1, if* $\mathbf{x}$ *and* $\mathbf{y}$ *are vector processes such that for all* $t \geq 0$ *we have*

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{x}(t)\mathbf{A}(t) + \mathbf{y}(t), \tag{7.12}$$

then

$$\mathbf{x}(t) = \mathbf{x}(0)\mathbf{E}_{\mathbf{A}}(t) + \int_0^t \mathbf{y}(s)\mathbf{E}_{\mathbf{A}}(s)^{-1}\mathbf{E}_{\mathbf{A}}(t)ds. \tag{7.13}$$

Proof: By the uniqueness of the solution to the ordinary differential equation (7.12) we need only differentiate (7.13) and show that (7.12) holds. ∎

Proof of Theorem 3.3: First, the various Poisson equations have unique solutions, e.g., see Section 4 of [27]. From the form of the solution, and the assumed differentiability of $\mathbf{A}$ at t, the vectors $\boldsymbol{\pi}^{(n)}$ have derivatives of all orders with 0 sums (except $\boldsymbol{\pi}^{(0)}(t)\mathbf{1}^{\mathsf{T}} = 1$). Now define the following sequence of vector processes

$$\mathbf{r}^{(n)}(t;\epsilon) \equiv \mathbf{p}(t;\epsilon) - \sum_{j=0}^{n} \epsilon^j \boldsymbol{\pi}^{(j)}(t). \tag{7.14}$$

Proving the theorem reduces to proving that $\mathbf{r}^{(n)}(t;\epsilon) = O(\epsilon^{n+1})$. Using (3.8), (3.15), and (3.16), we have

$$\begin{aligned}
\epsilon\frac{d}{dt}\mathbf{r}^{(n)}(t;\epsilon) &= \epsilon\frac{d}{dt}\mathbf{p}(t;\epsilon) - \epsilon\frac{d}{dt}\sum_{j=0}^{n}\epsilon^j\boldsymbol{\pi}^{(j)}(t) \\
&= \mathbf{p}(t;\epsilon)\mathbf{A}(t) - \sum_{j=0}^{n}\epsilon^{j+1}\boldsymbol{\pi}^{(j+1)}(t)\mathbf{A}(t) \\
&= \mathbf{r}^{(n+1)}(t;\epsilon)\mathbf{A}(t) \\
&= \mathbf{r}^{(n)}(t;\epsilon)\mathbf{A}(t) - \epsilon^{n+1}\boldsymbol{\pi}^{(n+1)}(t)\mathbf{A}(t).
\end{aligned} \tag{7.15}$$

If we set $t = 0$, then

$$\mathbf{r}^{(n)}(0;\epsilon) = \mathbf{p}(0;\epsilon) - \sum_{j=0}^{n}\epsilon^j\boldsymbol{\pi}^{(j)}(0) = O(\epsilon^{n+1}). \tag{7.16}$$

For all $0 \leq s \leq t$, use Lemma 7.1 to construct the operator $\mathbf{E}_{\mathbf{A}}(s,t;\epsilon) \equiv \mathbf{E}_{\mathbf{A}}(s;\epsilon)^{-1}\mathbf{E}_{\mathbf{A}}(t;\epsilon)$. Also by Lemma 7.1, we see that it solves the differential equation

$$\epsilon\frac{d}{ds}\mathbf{E}_{\mathbf{A}}(s,t;\epsilon) = -\mathbf{A}(s)\mathbf{E}_{\mathbf{A}}(s,t;\epsilon). \tag{7.17}$$

If we apply Corollary 7.2 to the differential equation for $\mathbf{r}^{(n)}(t;\epsilon)$ given by its derivative equalling (7.15), we obtain

$$\begin{aligned}
\mathbf{r}^{(n)}(t;\epsilon) &= -\epsilon^n \int_0^t \boldsymbol{\pi}^{(n+1)}(s)\mathbf{A}(s)\mathbf{E}_{\mathbf{A}}(s,t;\epsilon)ds + \mathbf{r}^{(n)}(0;\epsilon)\mathbf{E}_{\mathbf{A}}(t;\epsilon) \\
&= \epsilon^{n+1} \int_0^t \boldsymbol{\pi}^{(n+1)}(s)\frac{d}{ds}\mathbf{E}_{\mathbf{A}}(s,t;\epsilon)ds + O(\epsilon^{n+1}) \\
&= \epsilon^{n+1} \left[\boldsymbol{\pi}^{(n+1)}(t) - \boldsymbol{\pi}^{(n+1)}(0)\mathbf{E}_{\mathbf{A}}(t;\epsilon) - \int_0^t \left(\frac{d}{ds}\boldsymbol{\pi}^{(n+1)}(s)\right)\mathbf{E}_{\mathbf{A}}(s,t;\epsilon)ds\right] \\
&\quad + O(\epsilon^{n+1}) \\
&= O(\epsilon^{n+1}),
\end{aligned}$$

The second and third steps follow from using (7.17) and integration by parts, which completes the proof. ∎

To prove our theorem for the time-varying analogue of ergodicity, we introduce the *coefficient of ergodicity*, which is essentially Birkhoff's contraction coefficient; see Seneta [22] page 83, L4.3 on page 139, and page 145. For any square matrix $\mathbf{P}$, let

$$\tau_1(\mathbf{P}) = \frac{1}{2} \max_{0\le i<j\le \ell} \sum_{k=0}^{\ell} |p_{ik} - p_{jk}|\,. \tag{7.18}$$

Lemma 7.3 *For all square matrices $\mathbf{P}_1$ and $\mathbf{P}_2$,*

$$\tau_1(\mathbf{P}_1 + \mathbf{P}_2) \le \tau_1(\mathbf{P}_1) + \tau_1(\mathbf{P}_2) \tag{7.19}$$

and, for all scalars λ,

$$\tau_1(\lambda\mathbf{P}) = |\lambda|\tau_1(\mathbf{P}). \tag{7.20}$$

Moreover, if $\mathbf{P}$ is stochastic, then

$$\tau_1(\mathbf{P}) = \sup_{\mathbf{x}:\mathbf{x}\cdot\mathbf{1}^{\mathsf{T}}=0,\,\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{x}\mathbf{P}|}{|\mathbf{x}|} = 1 - \min_{0\le i<j\le\ell} \sum_{k=0}^{\ell} \min\left(p_{ik}, p_{jk}\right). \tag{7.21}$$

Hence, for all stochastic matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, we have

$$\tau_1(\mathbf{P}_1\mathbf{P}_2) \le \tau_1(\mathbf{P}_1)\tau_1(\mathbf{P}_2). \tag{7.22}$$

Relations (7.19) and (7.20) in Lemma 7.3 imply that τ_1 acts as a vector space norm on square matrices. By inequality (7.22), τ_1 acts as if it were an operator norm when we restrict it to products of stochastic matrices. It follows from (7.21) that the coefficient $\tau_1(\mathbf{P})$ is an upper bound on the moduli of all the eigenvalues of a stochastic matrix $\mathbf{P}$ *except* the largest one, which is 1. Moreover, for any two probability vectors $\mathbf{p}$ and $\mathbf{q}$, we have

$$|(\mathbf{p}-\mathbf{q})\mathbf{P}| \le \tau_1(\mathbf{P})|\mathbf{p}-\mathbf{q}|. \tag{7.23}$$

Proof of Theorem 3.4: The proof for this theorem reduces to showing that for all probability vectors $\mathbf{p}$ and $\mathbf{q}$, we have

$$(\mathbf{p}-\mathbf{q})\mathbf{E}_{\mathbf{A}}(t;\epsilon) \cong 0 \tag{7.24}$$

as $\epsilon \downarrow 0$. This limit is established by the following result, which gives us our time-varying UA analogue for ergodicity.

Theorem 7.4 *If* $\{ \mathbf{p}(s;\epsilon) \mid 0 \le s \le t \}$ *and* $\{ \mathbf{q}(s;\epsilon) \mid 0 \le s \le t \}$ *are both probability vectors that solve the same set of accelerated forward equations for the family of continuous, irreducible, Markov generators* $\{ \mathbf{A}(s) \mid 0 \le s \le t \}$ *with* $\mathbf{p}(0;\epsilon) = \mathbf{p}$ *and* $\mathbf{q}(0;\epsilon) = \mathbf{q}$, *then there exists two positive constants* α_0 *and* α_1 *for all positive* t *and* ϵ *such that*

$$|\mathbf{p}(t;\epsilon) - \mathbf{q}(t;\epsilon)| = |(\mathbf{p}-\mathbf{q})\mathbf{E_A}(t;\epsilon)| \le \alpha_0 e^{-\alpha_1 t/\epsilon}|\mathbf{p}-\mathbf{q}|. \tag{7.25}$$

Moreover, if $\lambda > \sup_{0\le s\le t}\sup_{0\le i\le \ell} a_i(s)$, *then there exists a constant* $0 < \sigma < 1$, *such that we can set* $\alpha_0 = 1 + 1/\sigma^2$ *and for all* $0 < \lambda_* < \lambda$

$$\alpha_1 = \min\Big(\lambda - \lambda_* + \lambda_* \log(\lambda_*/\lambda)\,,\ \lambda_* |\log\sigma|/\ell_* \Big). \tag{7.26}$$

Proof: To show that (7.25) holds, suppose that for any constant $\lambda > \sup_{0\le s\le t}\sup_{0\le i\le \ell} a_i(s)$, there exists some positive integer ℓ_*, depending only on ℓ, such that

$$\sigma \equiv \sup_{0\le s_1 \le \cdots \le s_{\ell_*} \le t} \tau_1\left(\prod_{i=1}^{\ell_*} \mathbf{P}_\lambda(s_i)\right) < 1 \tag{7.27}$$

and for all $0 \le \lambda_* < \lambda$, we have

$$\mathsf{P}(N(\lambda t) < \lambda_* t) \le e^{-\left(\lambda - \lambda_* + \lambda_* \log(\lambda_*/\lambda)\right)t}, \tag{7.28}$$

where $\{ N(t) \mid t \ge 0 \}$ is a standard Poisson process (mean 1). We will prove both of these below. We can then complete the proof since it follows from (7.27) that

$$\begin{aligned}
\tau_1(\mathbf{E_A}(t;\epsilon)) &\le \sum_{n=0}^{\infty} \mathsf{P}(N(\lambda t/\epsilon) = n) \int\cdots\int_{0\le s_1\le\cdots\le s_n\le t} \tau_1\left(\prod_{i=1}^{n}\mathbf{P}(s_i)\right) \frac{n!}{t^n} ds_1\cdots ds_n \\
&\le \mathsf{P}(N(\lambda t/\epsilon) < m) + \mathsf{P}(N(\lambda t/\epsilon) \ge m)\sigma^{\lfloor m/\ell_* \rfloor}
\end{aligned}$$

for all m. Finally we set $m = \lfloor \lambda_* t/\epsilon \rfloor$ where $\lambda_* < \lambda$, and obtain

$$\tau_1(\mathbf{E_A}(t;\epsilon)) \le \mathsf{P}(N(\lambda t/\epsilon) < \lambda_* t/\epsilon) + \sigma^{\lambda_* t/(\ell_*\epsilon) - 2}. \tag{7.29}$$

By (7.28), we have

$$\tau_1(\mathbf{E_A}(t;\epsilon)) \le e^{-\left(\lambda - \lambda_* + \lambda_* \log(\lambda_*/\lambda)\right)t/\epsilon} + \sigma^{\lambda_* t/(\ell_*\epsilon) - 2}, \tag{7.30}$$

and so (7.25) follows once we set

$$\alpha_0 = 1 + \frac{1}{\sigma^2} \ \text{ and } \ \alpha_1 = \min\Big(\lambda - \lambda_* + \lambda_* \log(\lambda_*/\lambda)\,,\ \lambda_* |\log\sigma|/\ell_* \Big), \tag{7.31}$$

where $0 \le \lambda_* \le \lambda$. Once we establish (7.27) and (7.28), this completes the proof. ∎

Since λ_* is arbitrary (within constraints), then a judicious choice of λ_* will maximize α_1. This maximum value cannot be attained when $\lambda_* = 0$ or $\lambda_* = \lambda$, so the optimum value for λ_* is in $(0, \lambda)$. Now we establish lemmas to prove (7.27) and (7.28).

Lemma 7.5 *If* $\{ \mathbf{A}(s) \mid 0 \le s \le t \}$ *is a continuous family of irreducible generators for the state space* $\{0, 1, \ldots, \ell\}$ *and* λ *is a constant where*

$$\lambda > \sup_{0 \le s \le t} \sup_{0 \le i \le \ell} a_i(s), \tag{7.32}$$

then there exists some positive integer ℓ_**, depending only on* ℓ*, such that*

$$\sigma \equiv \sup_{0 \le s_1 \le \cdots \le s_{\ell_*} \le t} \tau_1 \left(\prod_{i=1}^{\ell_*} \mathbf{P}_\lambda(s_i) \right) < 1, \tag{7.33}$$

where $\mathbf{P}_\lambda(s) = \mathbf{I} + \mathbf{A}(s)/\lambda$.

Proof: If $\mathbf{P}$ is an irreducible stochastic matrix, then there is an integer $n \le \ell$ such that $p_{ij}^{(n)} > 0$. For the special case of $\mathbf{P} = \mathbf{I} + \mathbf{A}/\lambda$, where $\mathbf{A}$ is a CTMC generator and $\lambda > \max_{0 \le i \le \ell} a_i$, all the diagonal entries of $\mathbf{P}$ are positive. So if $\mathbf{P}$ is irreducible and has this special form then it follows that $\mathbf{P}^\ell$ is a strictly positive matrix since $p_{ij}^{(\ell)} \ge p_{ij}^{(n)} p_{jj}^{(\ell-n)} > 0$.

Now for any stochastic matrix $\mathbf{P}$ on the set of states $\{0, 1, \ldots, \ell\}$, define $\mathcal{I}(\mathbf{P})$ to be its *incidence matrix* where every entry p_{ij} is replaced by 1 if $p_{ij} > 0$ and 0 otherwise. For any two stochastic matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, we have

$$\mathcal{I}(\mathbf{P}_1\mathbf{P}_2) = \mathcal{I}(\mathcal{I}(\mathbf{P}_1)\mathcal{I}(\mathbf{P}_2)). \tag{7.34}$$

When $\mathbf{P}$ has the above form, then it is aperiodic and its incidence matrix contains the identity matrix $\mathbf{I}$ as a submatrix or

$$\mathcal{I}(\mathbf{P}) \ge \mathbf{I}, \tag{7.35}$$

where the ordering relation denotes that every entry of $\mathcal{I}(\mathbf{P})$ is greater than or equal to every entry of $\mathbf{I}$. This means that the incidence matrix of an ordered product of stochastic matrices $\mathbf{P}_1, \ldots, \mathbf{P}_n$ of this type will have the following property:

$$\mathcal{I}\left(\prod_{\alpha=1}^{n} \mathbf{P}_\alpha \right) \ge \mathcal{I}\left(\prod_{\alpha \in \Gamma} \mathbf{P}_\alpha \right), \tag{7.36}$$

where the product $\prod_{\alpha=1}^{n} \mathbf{P}_\alpha$ is defined to equal $\mathbf{P}_1 \cdots \mathbf{P}_n$, Γ is any ordered subset of the indices $\{1, 2, \ldots, n\}$, and $\prod_{\alpha \in \Gamma} \mathbf{P}_\alpha$ is the corresponding ordered product of matrices.

Finally, let $|\mathcal{I}_\ell|$ equal the number of distinct incidence matrices that correspond to irreducible stochastic matrices for the state space $\{0, 1, \ldots, \ell\}$. Whenever we form a product of $\ell_* = \ell \cdot |\mathcal{I}_\ell|$ such irreducible, aperiodic stochastic matrices, then by the pigeonhole principle, there must be a subsequence of at least ℓ of these matrices with the same incidence matrix. By the inequality (7.36), all the entries of such a matrix product will be strictly positive. This shows that for all times $0 \le s_1 \le \cdots \le s_{\ell_*} \le t$, we have

$$\tau_1 \left(\prod_{i=1}^{\ell_*} \mathbf{P}_\lambda(s_i) \right) < 1. \tag{7.37}$$

Since this is a continuous function of the s_i, which range over a compact subset of $[0, t]$, the relation (7.33) holds. (The supremum is attained, but it must be less than 1.) ∎

Now we show that (7.28) holds. This follows from applying the Chernoff bound. (See Chapter 1 of Shwartz and Weiss [24].)

Lemma 7.6 *Let* $\{ N(t) \mid t \geq 0 \}$ *be a standard (rate 1) Poisson process. For all* λ *and* λ_* *with* $0 \leq \lambda_* < \lambda$, *we have*

$$\mathsf{P}(N(\lambda t) < \lambda_* t) \leq e^{-\left(\lambda - \lambda_* + \lambda_* \log(\lambda_*/\lambda)\right)t}, \tag{7.38}$$

and so the probability $\mathsf{P}(N(\lambda t) < \lambda_* t)$ *decays exponentially fast as* $t \to \infty$.

Proof: If $\theta > 0$, we can use Chebychev's inequality to obtain

$$\mathsf{P}\left(N(\lambda t) < \lambda_* t\right) = \mathsf{P}\left(e^{-\theta N(\lambda t)} > e^{-\theta \lambda_* t}\right) \leq \frac{\mathsf{E}\left[e^{-\theta N(\lambda t)}\right]}{e^{-\theta \lambda_* t}} \leq e^{-(\lambda - \lambda e^{-\theta} - \theta \lambda_*)t}. \tag{7.39}$$

We can minimize this upper bound by setting $\theta = \log \lambda/\lambda_*$, which gives us (7.38).

By the Cauchy mean value theorem, we have

$$\frac{1}{\lambda} < \frac{\log \lambda - \log \lambda_*}{\lambda - \lambda_*} < \frac{1}{\lambda_*}. \tag{7.40}$$

Therefore,

$$\begin{aligned} \lambda - \lambda_* + \lambda_* \log(\lambda_*/\lambda) &= \lambda - \lambda_* - \lambda_*(\log \lambda - \log \lambda_*) && (7.41) \\ &= (\lambda - \lambda_*)\left[1 - \lambda_* \frac{\log \lambda - \log \lambda_*}{\lambda - \lambda_*}\right] > 0, && (7.42) \end{aligned}$$

which completes the proof. ∎

Proof of Proposition 5.1: Poisson's equation is

$$\mathbf{x}\mathbf{A} = \mathbf{y} \quad \text{and} \quad \mathbf{x} \cdot \mathbf{1}^{\mathsf{T}} = 0. \tag{7.43}$$

Define $\mathbf{R}$ to be the matrix corresponding to the *right shift operator* on row vectors and $\mathbf{L}$ to be the *left shift operator* on row vectors. For any given vector $\mathbf{x}$, let $\mathbf{\Delta}(\mathbf{x})$ be the diagonal matrix such that the n-th diagonal term of $\mathbf{\Delta}(\mathbf{x})$ equals the n-th component of $\mathbf{x}$. If we set $\lambda(\ell) = \mu(0) = 0$, then

$$\mathbf{A} = \mathbf{\Delta}(\boldsymbol{\lambda})\mathbf{R} + \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{L} - \mathbf{\Delta}(\boldsymbol{\lambda} + \boldsymbol{\mu}). \tag{7.44}$$

Now $\mathbf{R}$ and $\mathbf{L}$ have various algebraic properties such as $\mathbf{RLR} = \mathbf{R}$ and $\mathbf{LRL} = \mathbf{L}$. Moreover since $\mu(0) = 0$, we have $\mathbf{\Delta}(\boldsymbol{\mu}) = \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{LR}$. If we redefine $\lambda(\ell)$ to be strictly positive but use $\mathbf{\Delta}(\boldsymbol{\lambda})\mathbf{RL}$ in place of $\mathbf{\Delta}(\boldsymbol{\lambda})$ we get

$$\begin{aligned} \mathbf{A} &= \mathbf{\Delta}(\boldsymbol{\lambda})\mathbf{RLR} + \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{LRL} - \mathbf{\Delta}(\boldsymbol{\lambda})\mathbf{RL} - \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{LRLR} \\ &= \mathbf{\Delta}(\boldsymbol{\lambda})\mathbf{RL}(\mathbf{R} - \mathbf{I}) + \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{LRL}(\mathbf{I} - \mathbf{R}) \\ &= \Big(\mathbf{\Delta}(\boldsymbol{\lambda}) - \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{L}\Big)\mathbf{RL}\Big(\mathbf{R} - \mathbf{I}\Big) \end{aligned}$$

If we set

$$\mathbf{x}^* \equiv \mathbf{x}\Big(\mathbf{\Delta}(\boldsymbol{\lambda}) - \mathbf{\Delta}(\boldsymbol{\mu})\mathbf{L}\Big) \quad \text{and} \quad \mathbf{y}^* = \mathbf{y}(\mathbf{I} - \mathbf{R})^{-1}, \tag{7.45}$$

then Poisson's equation reduces to

$$\mathbf{x}^*\mathbf{RL} = -\mathbf{y}^* \tag{7.46}$$

and

$$\mathbf{x}^*\Big(\boldsymbol{\Delta}(\boldsymbol{\lambda}) - \boldsymbol{\Delta}(\boldsymbol{\mu})\mathbf{L}\Big)^{-1}\mathbf{1}^\top = 0. \tag{7.47}$$

The solution for all but $x^*(\ell)$ is immediate, namely

$$x^*(i) = -y^*(i) \quad \text{for } i = 0, \ldots, \ell - 1. \tag{7.48}$$

Since $x^*(\ell) = x(\ell)\lambda(\ell)$ and $y^*(\ell) = \sum_{i=0}^{\ell} y(i) = 0$, we can write $\mathbf{x}^*$ as

$$\mathbf{x}^* = -\mathbf{y}^* + x^*(\ell)\mathbf{e}_\ell = -\mathbf{y}^* + x(\ell)\lambda(\ell)\mathbf{e}_\ell. \tag{7.49}$$

Since $\mathbf{e}_\ell \mathbf{L}^i = \mathbf{0}$ for $i > \ell$, we have

$$\begin{aligned} x(\ell)\lambda(\ell)\mathbf{e}_\ell\Big(\boldsymbol{\Delta}(\boldsymbol{\lambda}) - \boldsymbol{\Delta}(\boldsymbol{\mu})\mathbf{L}\Big)^{-1} &= x(\ell)\lambda(\ell)\mathbf{e}_\ell\Big(\mathbf{I} - \boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{-1}\boldsymbol{\Delta}(\boldsymbol{\lambda})^{-1} && (7.50)\\ &= x(\ell)\lambda(\ell)\sum_{i=0}^{\ell}\mathbf{e}_\ell\Big(\boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{i}\boldsymbol{\Delta}(1/\boldsymbol{\lambda}). && (7.51) \end{aligned}$$

Using (7.50) and substituting (7.49) into (7.47), we have

$$\begin{aligned} x(\ell)\lambda(\ell)\mathbf{e}_\ell\Big(\boldsymbol{\Delta}(\boldsymbol{\lambda}) - \boldsymbol{\Delta}(\boldsymbol{\mu})\mathbf{L}\Big)^{-1}\cdot\mathbf{1}^\top &= x(\ell)\lambda(\ell)\sum_{i=0}^{\ell}\mathbf{e}_\ell\Big(\boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{i}\boldsymbol{\Delta}(1/\boldsymbol{\lambda})\mathbf{1}^\top \\ &= \sum_{i=0}^{\ell}\mathbf{y}^*\Big(\boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{i}\boldsymbol{\Delta}(1/\boldsymbol{\lambda})\mathbf{1}^\top \\ &= \sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-1} y^*(j)\mathbf{e}_j\Big(\boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{i}\boldsymbol{\Delta}(1/\boldsymbol{\lambda})\mathbf{1}^\top \\ &= \sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-1} y^*(j)\mathbf{e}_j\Big(\boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{i}\boldsymbol{\Delta}(1/\boldsymbol{\lambda})\mathbf{1}^\top \\ &= \sum_{j=0}^{\ell-1}\sum_{i=0}^{j} y^*(j)\mathbf{e}_j\Big(\boldsymbol{\Delta}(\boldsymbol{\mu}/\boldsymbol{\lambda})\mathbf{L}\Big)^{i}\boldsymbol{\Delta}(1/\boldsymbol{\lambda})\mathbf{1}^\top. && (7.52) \end{aligned}$$

For all $i \le j$, define

$$\lambda(i,j) \equiv \begin{cases} \prod_{k=i}^{j}\lambda(k) & \text{when } i < j, \\ \lambda(i) & \text{when } i = j. \end{cases} \tag{7.53}$$

If $\mu(i,j)$ is defined similarly, then evaluating the operators on both sides of (7.52) gives

$$x(\ell)\lambda(\ell)\sum_{i=0}^{\ell}\frac{\mu(\ell-i+1,\ell)}{\lambda(\ell-i,\ell)} = \sum_{j=0}^{\ell-1}\sum_{i=0}^{j} y^*(j)\frac{\mu(j-i+1,j)}{\lambda(j-i,j)}, \tag{7.54}$$

which simplifies to

$$x(\ell)\sum_{i=0}^{\ell}\frac{\mu(i+1,\ell)}{\lambda(i,\ell-1)} = \sum_{j=0}^{\ell-1}\frac{y^*(j)}{\lambda(j)}\sum_{i=0}^{j}\frac{\mu(i+1,j)}{\lambda(i,j-1)}. \tag{7.55}$$

For all j, we have

$$\sum_{i=0}^{j}\frac{\mu(i+1,j)}{\lambda(i,j-1)} = \frac{\mu(1,j)}{\lambda(0,j-1)}\sum_{i=0}^{j}\frac{\lambda(0,i-1)}{\mu(1,i)} = \frac{1}{\beta(j)}, \tag{7.56}$$

so that (7.55) simplifies to

$$\frac{x(\ell)}{\beta(\ell)} = \sum_{j=0}^{\ell-1} \frac{y^*(j)}{\lambda(j)\beta(j)}, \tag{7.57}$$

which completes the proof. ∎

Proof of Corollary 5.2: Observe that the solution for $\pi_i^{(0)}(t;\ell)$ follows from the steady state distribution for birth death processes. Using (5.7) and (5.8) gives us

$$\overline{\beta}_i(\rho(t)) = 1 - \beta_i(\rho(t)) = \sum_{j=0}^{i-1} \rho(t)^j \prod_{k=0}^{j-1} \frac{r_k}{s_{k+1}} \Big/ \sum_{j=0}^{i} \rho(t)^j \prod_{k=0}^{j-1} \frac{r_k}{s_{k+1}}. \tag{7.58}$$

and

$$q_i(\rho(t)) = \sum_{j=0}^{i} j\,\rho(t)^j \prod_{k=0}^{j-1} \frac{r_k}{s_{k+1}} \Big/ \sum_{j=0}^{i} \rho(t)^j \prod_{k=0}^{j-1} \frac{r_k}{s_{k+1}}. \tag{7.59}$$

for all $i = 0, 1, \ldots, \ell$.

From (5.5) and (7.58) we obtain the identity

$$\sum_{j=0}^{i} \pi_j^{(0)}(t;\ell) = \prod_{j=i+1}^{\ell} \overline{\beta}_j(\rho(t)). \tag{7.60}$$

Moreover, from (7.58), (7.59), and using the notion of the logarithmic derivative, we can show that the $\overline{\beta}_i$'s satisfy the identity

$$\frac{x}{\overline{\beta}_i(x)} \frac{d}{dx} \overline{\beta}_i(x) = q_{i-1}(x) - q_i(x). \tag{7.61}$$

Applying the logarithmic derivative again to (7.60) and using the identity (7.61), we have

$$\begin{aligned}
\frac{d}{dt} \sum_{j=0}^{i} \pi_j^{(0)}(t;\ell) &= \left(\sum_{j=0}^{i} \pi_j^{(0)}(t;\ell)\right) \cdot \sum_{j=i+1}^{\ell} \frac{1}{\overline{\beta}_j(\rho(t))} \frac{d}{dt} \overline{\beta}_j(\rho(t)) \\
&= \left(\sum_{j=0}^{i} \pi_j^{(0)}(t;\ell)\right) \cdot \frac{\rho'(t)}{\rho(t)} \sum_{j=i+1}^{\ell} q_{j-1}(\rho(t)) - q_j(\rho(t)) \\
&= \left(\sum_{j=0}^{i} \pi_j^{(0)}(t;\ell)\right) \cdot \frac{\rho'(t)}{\rho(t)} \Big(q_i(\rho(t)) - q_\ell(\rho(t))\Big) \\
&= -\left(\prod_{j=i+1}^{\ell} \overline{\beta}_j(\rho(t))\right) \cdot \frac{\rho'(t)}{\rho(t)} \Big(q_\ell(\rho(t)) - q_i(\rho(t))\Big).
\end{aligned}$$

Using (5.1), we now have

$$\begin{aligned}
\pi_\ell^{(1)}(t) &= \beta_\ell(\rho(t)) \sum_{i=0}^{\ell-1} \frac{\frac{d}{dt} \sum_{j=0}^{i} \pi_j^{(0)}(t;\ell)}{\lambda(t;i)\beta_i(\rho(t))} \\
&= -\beta_\ell(\rho(t)) \sum_{i=0}^{\ell-1} \frac{\rho'(t)\Big(\beta_\ell(\rho(t)) - \beta_i(\rho(t))\Big)}{\lambda(t)\rho(t) r_i \beta_i(\rho(t))} \prod_{j=i+1}^{\ell} \overline{\beta}_j(\rho(t)) \\
&= -\beta_\ell(\rho(t)) \frac{\rho'(t)}{\lambda(t)\rho(t)} \sum_{i=0}^{\ell-1} \frac{\beta_\ell(\rho(t)) - \beta_i(\rho(t))}{r_i \beta_i(\rho(t))} \prod_{j=i+1}^{\ell} \overline{\beta}_j(\rho(t)),
\end{aligned}$$

which completes the proof. ∎

References

[1] Bender, C. M. and Orszag, S. A. 1978. *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill Book Company, New York, NY.

[2] Chung, K. L. 1967. *Markov Chains*, Springer-Verlag, New York.

[3] Dollard, J. D. and Friedman, C. N. 1979. *Product Integration with Applications to Differential Equations.* Encyclopedia Math. Appl. 10, Addison-Wesley, Reading, MA.

[4] Eick, S. G., Massey, W. A. and Whitt, W. 1993. The physics of the $M_t/G/\infty$ queue. *Oper. Res.* 41, 731–742.

[5] Gantmacher, F. R. *Applications of the Theory of Matrices*, Interscience Publishers, New York, 1959.

[6] Green, L. and Kolesar, P. 1991. The pointwise stationary approximation for queues with nonstationary arrivals. *Management Sci.* 37, 84–97.

[7] Hall, R. W. 1991. *Queueing Methods for Services and Manufacturing*, Prentice Hall, Englewood Cliffs, NJ 07632.

[8] Heyman, D. P. and Whitt, W. 1984. The asymptotic behavior of queues with time-varying arrival rates. *J. Appl. Prob.* 21, 143–156.

[9] Keilson, J. 1979. *Markov Chain Models - Rarity and Exponentiality*, Springer-Verlag, New York.

[10] Khasminskii, R. Z., Yin, G. and Zhang, Q. 1996. Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains. *SIAM J. Appl. Math.* 56, 277–293.

[11] Keller, J. B. 1982. Time-dependent queues. *SIAM Rev.* 24, 401–412.

[12] Mandelbaum, A. and Massey, W. A. 1995. Strong approximations for time-dependent queues. *Math. Oper. Res.* 20, 33-64.

[13] Massey, W. A. 1981. Nonstationary Queues, Ph.D. dissertation, Stanford University.

[14] Massey, W. A. 1985. Asymptotic analysis for the time-dependent M/M/1 queue. *Math. Oper. Res.* 10, 305–327.

[15] Massey, W. A., Parker, G. A. and Whitt, W. 1996. Estimating the parameters of a nonhomogeneous Poisson process with linear rate. *Telecommunication Systems* 5, 361-388.

[16] Massey, W. A. and Whitt, W. 1993a. A probabilistic generalization of Taylor's theorem. *Statistics and Probability Letters* 16, 51–54.

[17] Massey, W. A. and Whitt, W. 1993b. Networks of infinite server queues with nonstationary Poisson input. *Queueing Systems* 13, 183–250.

[18] Massey, W. A. and Whitt, W. 1994a. Unstable asymptotics for nonstationary queues. *Math. Oper. Res.* 19, 267–291.

[19] Massey, W. A. and Whitt, W. 1994b. An analysis of the modified-offered-load approximation for the nonstationary Erlang loss model. *Ann. Appl. Prob.* 4, 1145–1160.

[20] Massey, W. A. and Whitt, W. 1997. Peak congestion in multi-server service systems with slowly varying arrival rates. *Queueing Systems*, 25, 157–172.

[21] Newell, G. F. 1968. Queues with time-dependent arrival rates, I, II and III. *J. Appl. Prob.* 5, 436–451, 579-606.

[22] Seneta, E. 1981. *Non-negative Matrices and Markov Chains*, second ed., Springer-Verlag, New York.

[23] Schweitzer, P. J. 1968. Perturbation Theory and Finite Markov Chains. *J. Appl. Prob.* 5, 401-413.

[24] Shwartz, A. and Weiss, A. 1995. *Large Deviations for Performance Analysis*, Chapman and Hall, London.

[25] Stewart, W. J. *Introduction to the Numerical Solutions of Markov Chains*, Princeton University Press, Princeton, NJ.

[26] Whitt, W. 1991. The pointwise stationary approximation for $M_t/M_t/s$ queues is asymptotically correct as the rates increases. *Management Sci.* 37, 307–314.

[27] Whitt, W. 1992. Asymptotic formulas for Markov processes with applications to simulation. *Oper. Res.* 40, 279–291.

[28] Zeifman, A. I. 1989. Properties of a system with losses in the case of variable rates. *Automation and Remote Control*, 1, 82–87.