Applied Mathematical Modelling 61 (2018) 561–576

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Approximate adaptive uniformization of continuous-time Markov chains

Alexander Andreychenko, Werner Sandmann*, Verena Wolf

Department of Computer Science, Saarland University, Campus E1 3, Saarbrücken 66123, Germany

ARTICLE INFO

Article history:
Received 23 October 2017
Revised 19 April 2018
Accepted 15 May 2018
Available online 22 May 2018

Keywords:
Continuous-time Markov chain
Transient probability distribution
Uniformization
Randomization
Discrete-time conversion
Dynamic state space truncation

ABSTRACT

We consider the approximation of transient (time dependent) probability distributions of discrete-state continuous-time Markov chains on large, possibly infinite state spaces. A framework for approximate adaptive uniformization is provided, which generalizes the well-known uniformization technique and many of its variants. Based on a birth process and a discrete-time Markov chain a computationally tractable approximating process/model is constructed. We investigate the theoretical properties of this process and prove that it yields computable lower and upper bounds for the desired transient probabilities. Finally, we discuss different specific ways of performing approximate adaptive uniformization and analyze the corresponding approximation errors. The application is illustrated by an example of a stochastic epidemic model.

1. Introduction

Continuous-time Markov chains (CTMCs) with large (possibly countably infinite) discrete state spaces are widely used to model real-world systems and processes in many areas such as engineering and environmental processes, computer and communication networks, transportation, logistics, production and manufacturing systems, healthcare systems, epidemic diseases, multiphysics systems, or biochemically reacting systems, amongst many others. Often the interest is in transient (time dependent) system behavior or performance properties, respectively, and the analysis requires the determination of transient probability distributions. For instance, recent real-world applications in which transient probability distributions of CTMCs are highly relevant include such diverse areas as epidemics and the spread of diseases in human and animal populations as well as in computer networks [1–4], upstream inventory information sharing in supply chain networks [5], evacuation route selection of large-scale crowds under emergencies [6], allelopathic interactions between competing phytoplankton species in marine environments [7], or the impact of calcium fluxes on cells in the brain [8].

We consider a regular (conservative and non-explosive) CTMC $X = \{X(t), t \geq 0\}$, that is a discrete-state Markov jump process with only finitely many jumps in any finite time interval and right-continuous sample paths. Without loss of generality we assume that the state space is $S \subseteq \mathbb{N} = \{0, 1, 2, \ldots\}$. The infinitesimal generator of the CTMC is expressed by the matrix $Q = (q_{ij})_{i,j \in S}$, where $q_{ij} < \infty$ for $i \neq j$ are the state transition rates and $q_{ii} := -\sum_{i \neq j} q_{ij}$ such that all row sums of Q are zero, that is the CTMC is conservative. We define $q_i := -q_{ii}$. At any time $t \geq 0$ the transient (time-dependent) distribution is the collection of all state probabilities at that time, represented by the row vector $p(t)$. For any initial distribution $p(0)$ it is given

* Corresponding author.
E-mail addresses: alexander.andreychenko@uni-saarland.de (A. Andreychenko), werner.sandmann@uni-saarland.de (W. Sandmann), verena.wolf@uni-saarland.de (V. Wolf).

https://doi.org/10.1016/j.apm.2018.05.009
0307-904X/

by

$$p(t) = p(0)e^{Qt} = p(0)\sum_{k=0}^{\infty}\frac{(Qt)^k}{k!} \tag{1}$$

as the unique solution of the corresponding system of Kolmogorov (backward and forward) differential equations, see [9, pp. 80–81], [10, pp. 251–252].

Since analytical solutions are only possible in special cases, numerical techniques are commonly applied [11,12], but conceptually exact approaches are often computationally intractable due to prohibitively large (possibly infinite) state spaces.

Uniformization, also known as randomization or discrete-time conversion, provides a numerically stable computational scheme to approximate the transient probability distributions of CTMCs with uniformly bounded transition rates. It was originally proposed by Jensen [13], has been used to study computational and theoretical aspects of CTMCs [14–22] and has become one of the most prevalent approaches to the numerical transient analysis of CTMCs.

The underlying principle is to consider a *subordinated discrete-time Markov chain* (DTMC) $Y = \{Y(n), n \in \mathbb{N}\}$ with initial distribution $p(0)$ and transition probability matrix $P = Q/\lambda + I$ where $\sup_i q_i \le \lambda < \infty$. Hence, $Q = \lambda(P - I)$ such that the transient distribution of the CTMC can be written as

$$p(t) = p(0)e^{-\lambda t}e^{\lambda Pt} = p(0)e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k P^k}{k!} = \sum_{k=0}^{\infty} p(0)P^k\frac{(\lambda t)^k e^{-\lambda t}}{k!}. \tag{2}$$

Note that $p(0)P^k =: \pi^{(k)}$ is the probability distribution of the DTMC after k steps and $(\lambda t)^k e^{-\lambda t}/k!$ is the probability that a Poissonian random variable with parameter λt takes the value k, which equals the probability of exactly k events during a time period of length $t \ge 0$ in a Poisson process with rate λ. Hence, the principle of uniformization can be viewed as the construction of a Poisson process $\{N(t),\ t \ge 0\}$ and a DTMC $\{Y(n), n \in \mathbb{N}\}$ such that $Y(N(t))$ has the same distribution as $X(t)$ for every $t \ge 0$.

In numerical computations $\pi^{(k)} = \pi^{(k-1)}P$ can be computed iteratively and the infinite sum in (2) must be truncated, for which a priori truncation error bounds can be set and a corresponding truncation point can be determined based on the cumulated sum of Poisson probabilities, cf. [23], [12, pp. 410–413]. This technique, henceforth referred to as standard uniformization (SU), often works well, but it has several serious drawbacks in case of large and stiff models. In particular, it requires uniformly bounded transition rates and a global *uniformization rate* λ that is a finite upper bound on the maximum outrate of states, which is typically not available in the case of an infinite state space.

Adaptive uniformization (AU) [24] is an approach to cope with large state spaces without assuming uniformly bounded transition rates by adapting the uniformization rate. More precisely, it is shown in [24] that depending on the set of *active states* with nonzero probabilities after a particular number of jumps one can construct a birth process $\{B(t),\ t \ge 0\}$ and a subordinated DTMC $\{Y(n), n \in \mathbb{N}\}$ such that $Y(B(t))$ has the same distribution as $X(t)$ for every $t \ge 0$. The n-th transition rate of the birth process corresponds to the 'fastest' state visited by the DTMC at step n with positive probability. This differs from the (constant) rate λ of the Poisson process in that the birth process jumps at rates smaller than or equal to λ. As a result, the number of vector-matrix multiplications required by AU may be lower than for SU, in particular if the average rate of the birth process is much less than the jump rate λ of the Poisson process in SU. However, in many cases AU is less efficient than SU.

The computation of the probabilities $Pr(B(t) = n)$ required with AU [24] is more difficult than the computation of Poisson probabilities for SU. Furthermore, in general, the rates of the birth process defined by AU converge to the supremum of the transition rates of all non-transient states of the CTMC X. Thus, the potential computational gain that can be achieved by AU as compared to SU is anyway marginal for sufficiently long time horizons. Besides, the initial set of active states can be already large or the adapted set of active states can grow quickly, that is, the number of states with nonzero initial probability is large, or a large number of states have a nonzero probability after only a few steps in the subordinated DTMC. In such cases, even for small and moderate time horizons AU is less efficient than SU.

Another way of applying the uniformization idea to Markov chains with large state spaces is by state space truncation, which effectively reduces the number of nonzero entries in the transition probability matrix of the respective subordinated DTMC Y, thereby reducing the effort required for the vector-matrix multiplications in numerical computations. Such truncation approaches often allow to compute accurate approximations to the transient distributions of infinite-state CTMCs with relatively moderate execution times and memory costs. For instance, approximate uniformization [25] is a truncated version of SU using a fixed state space truncation such that the assumption of uniformly bounded transition rates can be relaxed. Similarly, inexact uniformization [26] proceeds SU while ignoring states with small probabilities below a pre-specified threshold, which in fact corresponds to truncating the state space.

Fast adaptive uniformization [27] combines state space truncation and the use of adaptive uniformization rates. In this approach, the truncation of the state space not only decreases the computational effort of the vector-matrix multiplication, but also allows the use of smaller transition rates for the birth process than with AU. This leads to a further reduction in the number of vector-matrix multiplications required to approximate transient distributions of CTMCs. The method has been applied in [27] to stochastic chemical kinetics by heuristically choosing sets of active states as sets of 'temporarily significant' states with a probability below a certain threshold similarly as in [26] but with adapted uniformization rates

rather than with a constant uniformization rate. A detailed study of the probabilistic structure and interpretation as well as an error analysis is lacking and highly desirable.

In this paper, we provide a general framework for approximate adaptive uniformization that covers and generalizes all the uniformization approaches outlined above. We generalize the notion of active sets to allow for an arbitrary non-explosive birth process randomizing the time between transitions in the corresponding subordinated DTMC. We use this framework to investigate the theoretical/structural and computational properties of the resulting stochastic process $\bar{X} = \{\bar{X}(t), t \geq 0\}$ where $\bar{X}(t) := Y(B(t))$ for all $t \geq 0$. We prove that the transient distributions of $\bar{X}$ yield strict upper and lower bounds for the transient distributions of X, that is, for all states i and times t,

$$Pr(\bar{X}(t) = i) \leq Pr(X(t) = i) \leq Pr(\bar{X}(t) = i) + Pr(\bar{X}(t) = \perp), \tag{3}$$

where $\perp$ is a special error state, which we adjoin to the state space of X. Additionally, we provide an approximation error analysis and discuss different strategies for the dynamic truncation of the state space.

The general framework for approximate adaptive uniformization is given in the next section, including all necessary definitions of the birth process, the subordinated DTMC, and the combined stochastic process that is used to approximate the desired transient distribution of the CTMC under consideration. Then we have all prerequisites required to state (3) formally as our main theorem and to deduce respective corollaries covering standard uniformization and adaptive uniformization. The proof of this main theorem, in the course of which we prove a couple of intermediate results that reveal important theoretical insights into the properties of the constructed stochastic process, is given in Section 3. Subsequently, in Section 4 we investigate different strategies for choosing the active sets used in approximate adaptive uniformization and prove error bounds for the corresponding approximations. In Section 5, the application is illustrated by an example of a stochastic epidemic model. Finally, Section 6 concludes the paper.

2. General framework for approximate adaptive uniformization

Our general framework builds upon and generalizes the probabilistic interpretations of SU and AU, respectively, in that we consider the construction of a CTMC that is a combination of a birth process and a subordinated nonhomogeneous DTMC. While with AU this construction is based on specifically defined active sets with nonzero probabilities we consider in general an arbitrary infinite sequence $A_0, A_1, \ldots$ of subsets of $\mathbb{N}$ such that $\sup_{i \in A_n} q_i < \infty$ for all $n \in \mathbb{N}$. We say that a state i is *active* at the n-th *epoch* if $i \in A_n$ and we call A_n the n-th active set. Moreover, let $\lambda_0 \geq 0, \lambda_1 \geq 0, \ldots$ be such that $\lambda_n \geq \sup_{i \in A_n} q_i$. We refer to λ_n as the nth uniformization rate. If used for computing an approximation to the transient distribution of a given CTMC, then the sets $A_0, A_1, \ldots$ are typically chosen to be finite truncations of the state space (cf. Section 4).

Now, we specify a birth process $B = \{B(t), t \geq 0\}$, a nonhomogeneous DTMC $Y = \{Y(n), n \in \mathbb{N}\}$ and a combined CTMC $\bar{X} = \{\bar{X}(t), t \geq 0\}$. Essentially, Y keeps track of the current state of X and B relates epochs with time. We will show that the combined CTMC $\bar{X}$ can be used to approximate the desired transient probability distribution of X. This approximation is exact for certain choices of the active sets, in particular for the choices corresponding to SU and AU (see Corollaries 1 and 2 below).

The birth process. The birth process B is a pure birth process that is independent of X with birth rates $\lambda_0 \geq 0, \lambda_1 \geq 0, \ldots$ and state space $\mathbb{N}$. For an infinitesimal time interval $[t, t + \Delta)$, with $\Delta > 0$ and any state $n \in \mathbb{N}$, we find transition probabilities

$$Pr(B(t + \Delta) = m \mid B(t) = n) = \begin{cases} 1 - \lambda_n \Delta + o(\Delta), & \text{if } m = n, \\ \lambda_n \Delta + o(\Delta), & \text{if } m = n + 1, \\ o(\Delta), & \text{if } m > n + 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

where $o(\Delta)$ is a function such that $o(0) = 0$ and $\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} = 0$. Moreover, we have for the initial probability distribution that

$$Pr(B(0) = 0) = 1, \tag{5}$$

and all other initial probabilities are zero. Note that even if X is non-explosive, B may be explosive depending on the chosen sequence of uniformization rates. In the sequel, we assume that the uniformization rates are chosen such that B is non-explosive. More details are given in Section 4.

The subordinated DTMC. The DTMC Y is independent of X and B and has state space $\mathbb{N} \cup \{\perp\}$ where $\perp$ is a new distinguished state. For any state $i \in \mathbb{N} \cup \{\perp\}$ and any time epoch $n \in \mathbb{N}$, define $p_i(n) := Pr(Y(n) = i)$. All active states $i \in A_0$ have initial probability $p_i(0) > 0$ and all remaining states, including the state $\perp$, have initially probability zero. The idea of approximate adaptive uniformization is to observe X at the time instances where B jumps. The DTMC Y keeps track of the current state of X conditioned on the jumps of B. Whenever X jumps to a state that is not active in the subsequent epoch, the DTMC Y

will jump to the special state $\perp$. For $i, j \in \mathbb{N} \cup \{\perp\}$, the one-step transition probabilities of Y are given by

$$Pr(Y(n+1) = j \mid Y(n) = i) = \begin{cases} q_{ij}/\lambda_n, & \text{if } i \in A_n,\ j \in A_{n+1},\ i \neq j, \\ \sum_{k \notin A_{n+1}} q_{ik}/\lambda_n, & \text{if } i \in A_n,\ j = \perp, \\ 1 - \sum_{k \neq i} q_{ik}/\lambda_n, & \text{if } i \in A_n,\ j \in A_{n+1},\ i = j, \\ 1, & \text{if } i \notin A_n,\ i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Note that $i \in A_n$, $j = \perp$ implies $i \neq j$ and that all infinite sums in the equation above converge since $\sum_{k \neq i} q_{ik} = q_i \leq \lambda_n$ for $i \in A_n$. The above transition probabilities depend on the current epoch n, which means that Y is in general nonhomogeneous in time. Clearly, if all A_n and λ_n are identical, then Y is homogeneous in time.

The combined stochastic process. We define the stochastic process $\bar{X} = \{\bar{X}(t), t \geq 0\}$ by $\bar{X}(t, \omega) = Y(B(t, \omega), \omega)$ for any outcome ω. This construction is also called the subordination of Y to B. The process $\bar{X}$ has indeed the same state space as Y. Since Y and B are independent, we have

$$Pr(\bar{X}(t) = i) = Pr(Y(B(t)) = i) = \sum_{n=0}^{\infty} Pr(Y(n) = i) Pr(B(t) = n). \tag{7}$$

Lemma 1. *The stochastic process $\bar{X}$ is a continuous-time Markov chain. That is, for states $i, j_1, \ldots, j_m$ and time points $t > t_1 > \ldots > t_m$,*

$$Pr(\bar{X}(t) = i \mid \bar{X}(t_1) = j_1, \ldots, \bar{X}(t_m) = j_m) = Pr(\bar{X}(t) = i \mid \bar{X}(t_1) = j_1).$$

This follows directly from the independence of Y and B and the fact that both Y and B are Markov chains.

As a main result of the paper, we prove that if $\bar{X}$ is constructed based on an *arbitrary* sequence $A_0, A_1, \ldots$, then the transient state probabilities of $\bar{X}$ are lower bounds of the corresponding transient probabilities of X and that adding the transient probability of $\perp$ yields upper bounds. More precisely:

Theorem 1. *For any sequence of active sets $A_0, A_1, \ldots$, uniformization rates $\lambda_0, \lambda_1, \ldots$, and all $t \geq 0$, $i \in \mathbb{N}$, if B does not explode, then*

$$Pr(\bar{X}(t) = i) \leq Pr(X(t) = i) \leq Pr(\bar{X}(t) = i) + Pr(\bar{X}(t) = \perp).$$

In particular, this mitigates the drawbacks of adaptive uniformization because fast states (i.e., states having high exit rates) with small probability can be excluded from being active. In this case the speed of the birth process remains slow or can even become slower as the dynamics of the underlying systems slows down. Similar as for inexact uniformization [26], the number of non-zero entries in the transition probability matrix of Y remains manageable for large or infinite Markov chains.

Before we proceed to the proof of Theorem 1 we formulate corollaries addressing special cases corresponding to uniformization approaches outlined in the introduction.

Corollary 1 ***(Standard Uniformization).*** *Assume that $\sup_i q_i < \infty$. Choose $A_n = \mathbb{N}$ and $\lambda_n = \lambda < \infty$ for all $n \in \mathbb{N}$ where $\lambda \geq \sup_i q_i$. Then for all $t \geq 0$, $i \in \mathbb{N}$,*

$$Pr(\bar{X}(t) = i) = Pr(X(t) = i).$$

Note that Corollary 1 obviously addresses standard uniformization since the birth process with $\lambda_n = \lambda$ for all $n \in \mathbb{N}$ is the Poisson process with rate λ. It is clear that with $A_n = \mathbb{N}$ for all $n \in \mathbb{N}$ all states are always active such that the sums in (6) with $j = \perp$ are empty, $Pr(Y(n) = \perp) = 0$ for all $n \in \mathbb{N}$ and (7) yields $Pr(\bar{X}(t) = \perp) = 0$ for all $t \geq 0$.

Corollary 2 ***(Adaptive Uniformization**, Theorem 1 **in** [24]).* *For all $n \in \mathbb{N}$ choose $A_n = \{i \in S \mid p_i(n) > 0\}$ and $\lambda_n = \sup_{i \in A_n} q_i$. If B does not explode, then for all $t \geq 0$, $i \in \mathbb{N}$,*

$$Pr(\bar{X}(t) = i) = Pr(X(t) = i).$$

Here, since $A_{n+1} = \{i \in S \mid p_i(n+1) > 0\}$, obviously $(i \in A_n) \wedge (k \notin A_{n+1}) \Rightarrow q_{ik} = 0$, because otherwise $p_k(n+1)$ would be positive and thus k be active at epoch $n+1$ and contained in A_{n+1}. Hence, for all $n \in \mathbb{N}$,

$$Pr(Y(n+1) = \perp \mid Y(n) = i) = \sum_{k \notin A_{n+1}} \frac{q_{ik}}{\lambda_n} = 0, \quad i \in A_n,$$

such that according to (6) we have $Pr(Y(n) = \perp) = 0$ for all $n \in \mathbb{N}$ and (7) yields $Pr(\bar{X}(t) = \perp) = 0$ for all $t \geq 0$.

Note that the bounds in Theorem 1 hold in particular if $\sup_i q_i < \infty$ and $\lambda_n = \lambda < \infty$ for all $n \in \mathbb{N}$ where $\lambda \geq \sup_i q_i$. A similar result was shown by van Dijk [25]. In practice, it is used to truncate large or infinite state spaces. In fact, applying a fixed state space truncation as in approximate uniformization [25] can be expressed as a special case of our general framework by formally choosing a constant set of active states. Similarly, dynamically choosing small active sets that contain the

most 'significant' states and ignoring states with small probabilities below a pre-specified threshold as in inexact unifomization [26] corresponds to a specific choice of the active sets in our general framework. The advantage of choosing small sets A_n is that the computational complexity of the approximation algorithm is reduced. Of course, this comes at the expense of reduced accuracy.

3. Proof of Theorem 1

In [24], the proof of our Corollary 2, which is Theorem 1 in [24], is established by first proving that the residence time distribution of a state i is identical in X and $\bar{X}$. Then, the equality of the jump probabilities is shown. Similar proof arguments cannot be used to prove our more general Theorem 1 since the dependence of the birth process $B(t)$ on $\bar{X}(t)$ cannot be avoided. Instead, we will relate the transient distributions of X and $\bar{X}$ indirectly by comparing them to an intermediate process $\widetilde{X}$ which keeps track of the discarded probability mass in the state $\perp$.

We first investigate the infinitesimal transition probabilities of $\bar{X}$. We will then introduce an auxiliary 'intermediate' Markov chain $\widetilde{X}$ and compare the Kolmogorov forward equations of X, $\bar{X}$, and $\widetilde{X}$ to establish our Theorem 1.

3.1. Transition probabilities of $\bar{X}$

We now consider the transition probabilities of the Markov chain $\bar{X}$ for an infinitesimal time interval $[t, t+\Delta)$. These probabilities will depend on the probability $e(n, i, t)$ that the birth process B occupies an epoch n under the condition that the Markov chain $\bar{X}$ occupies a state $i \in \mathbb{N}$ such that $Pr(\bar{X}(t) = i) > 0$, i.e.

$$e(n, i, t) := Pr(B(t) = n \mid \bar{X}(t) = i).$$

Note that for any time point $t \geq 0$ and any state $i \in \mathbb{N}$, we have $\sum_{n=0}^{\infty} e(n, i, t) = 1$. Moreover, for any epoch n we have that $Pr(Y(n) = i) = 0$ implies $e(n, i, t) = 0$.

Theorem 2. *Let $[t, t+\Delta)$ be an infinitesimal time interval, $i, j \in \mathbb{N}$ with $i \neq j$, $t \geq 0$, and $Pr(\bar{X}(t) = i) > 0$. Then,*

(a) *the transition probability $Pr(\bar{X}(t+\Delta) = j \mid \bar{X}(t) = i)$ equals*

$$\left(q_{ij} \sum_{n: j \in A_{n+1}} e(n, i, t) \right) \Delta + o(\Delta),$$

(b) *the transition probability $Pr(\bar{X}(t+\Delta) = i \mid \bar{X}(t) = i)$ equals*

$$1 - \left(q_i \sum_{n: i \in A_{n+1}} e(n, i, t) + \sum_{n: i \notin A_{n+1}} \lambda_n e(n, i, t) \right) \Delta + o(\Delta),$$

and

(c) *the transition probability $Pr(\bar{X}(t+\Delta) = \perp \mid \bar{X}(t) = i)$ equals*

$$\left(\sum_{n: i \notin A_{n+1}} (\lambda_n - q_i) e(n, i, t) + \sum_{k \neq i} q_{ik} \sum_{n: k \notin A_{n+1}} e(n, i, t) \right) \Delta + o(\Delta).$$

For state $\perp$ we find $Pr(\bar{X}(t+\Delta) = \perp \mid \bar{X}(t) = \perp) = 1$ and for any state $i \in \mathbb{N}$ we have $Pr(\bar{X}(t+\Delta) = i \mid \bar{X}(t) = \perp) = 0$.

Proof. We first prove part a). By definition, we have

$$\begin{aligned} Pr(\bar{X}(t+\Delta) = j \mid \bar{X}(t) = i) &= Pr(Y(B(t+\Delta)) = j \mid Y(B(t)) = i) \\ &= \sum_{n=0}^{\infty} Pr(Y(n) = j, B(t+\Delta) = n \mid Y(B(t)) = i) \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} Pr(Y(n) = j, B(t+\Delta) = n \mid Y(m) = i, B(t) = m) \\ &\quad \times Pr(Y(m) = i, B(t) = m \mid \bar{X}(t) = i). \end{aligned}$$

Note that the probability that B makes more than two jumps within $[t, t+\Delta)$ is $o(\Delta)$. On the other hand, the probability that Y moves from i to j in zero jumps is obviously zero. We then only need to consider the case that the birth process B makes exactly one jump in the time interval $[t, t+\Delta)$. Finally, for the last probability we have that $\bar{X}(t) = i$ and $B(t) = m$

imply $Y(m) = i$. Therefore, the above double sum simplifies and we obtain

$$\begin{aligned} Pr(\bar{X}(t+\Delta) = j \mid \bar{X}(t) = i) &= \sum_{n=0}^{\infty} Pr(Y(n+1) = j, B(t+\Delta) = n+1 \mid Y(n) = i, B(t) = n) e(n,i,t) + o(\Delta) \\ &= \sum_{n=0}^{\infty} Pr(Y(n+1) = j \mid Y(n) = i) Pr(B(t+\Delta) = n+1 \mid B(t) = n) e(n,i,t) + o(\Delta) \\ &= \sum_{n:i\in A_n, j\in A_{n+1}} \frac{q_{ij}}{\lambda_n} (\lambda_n \Delta + o(\Delta)) e(n,i,t) + o(\Delta), \end{aligned}$$

where the second equality holds because of the independence of Y and B and the third equality by substituting (4) and (6) into the above. For an epoch n such that $i \notin A_n$ we have $Pr(Y(n) = i) = 0$ and then also $e(n,i,t) = 0$. Therefore,

$$Pr(\bar{X}(t+\Delta) = j \mid \bar{X}(t) = i) = \left(q_{ij} \sum_{n: j\in A_{n+1}} e(n,i,t) \right) \Delta + o(\Delta).$$

For part b) we find, by a similar reasoning as above,

$$Pr(\bar{X}(t+\Delta) = i \mid \bar{X}(t) = i) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} Pr(Y(n) = i, B(t+\Delta) = n \mid Y(m) = i, B(t) = m) e(m,i,t).$$

Again, the birth process takes two jumps in $[t, t+\Delta)$ with probability $o(\Delta)$ and we only need to consider the case that B makes zero jumps or one jump in $[t, t+\Delta)$. Thus,

$$\begin{aligned} Pr(\bar{X}(t+\Delta) = i \mid \bar{X}(t) = i) &= \sum_{n=0}^{\infty} Pr(Y(n) = i, B(t+\Delta) = n \mid Y(n) = i, B(t) = n) e(n,i,t) \\ &\quad + \sum_{n=0}^{\infty} Pr(Y(n+1) = i, B(t+\Delta) = n+1 \mid Y(n) = i, B(t) = n) e(n,i,t) + o(\Delta) \\ &= \sum_{n=0}^{\infty} Pr(Y(n) = i \mid Y(n) = i) Pr(B(t+\Delta) = n \mid B(t) = n) e(n,i,t) \\ &\quad + \sum_{n=0}^{\infty} Pr(Y(n+1) = i \mid Y(n) = i) Pr(B(t+\Delta) = n+1 \mid B(t) = n) e(n,i,t) + o(\Delta) \\ &= \sum_{n=0}^{\infty} (1 - \lambda_n \Delta + o(\Delta)) e(n,i,t) \; + \sum_{n:i\in A_{n+1}} \frac{\lambda_n - q_i}{\lambda_n} (\lambda_n \Delta + o(\Delta)) e(n,i,t) + o(\Delta), \end{aligned}$$

where the second equality holds because of the independence of Y and B and the third equality by substituting (6) and (4). Finally, recall that $\sum_{n=0}^{\infty} e(n,i,t) = 1$ such that the above further simplifies to

$$\begin{aligned} Pr(\bar{X}(t+\Delta) = i \mid \bar{X}(t) = i) &= 1 - \sum_{n=0}^{\infty} \lambda_n \Delta + \sum_{n:i\in A_{n+1}} (\lambda_n - q_i) \Delta e(n,i,t) + o(\Delta) \\ &= 1 - \left(\sum_{n:i\notin A_{n+1}} \lambda_n e(n,i,t) + q_i \sum_{n:i\in A_{n+1}} e(n,i,t) \right) \Delta + o(\Delta). \end{aligned}$$

For part (c), we find, by similar reasoning as above,

$$\begin{aligned} Pr(\bar{X}(t+\Delta) = \perp \mid \bar{X}(t) = i) &= \sum_{n=0}^{\infty} Pr(Y(n+1) = \perp \mid Y(n) = i) Pr(B(t+\Delta) = n+1 \mid B(t) = n) e(n,i,t) + o(\Delta) \\ &= \left(\sum_{n:i\in A_{n+1}} \frac{\sum_{k\notin A_{n+1}} q_{ik}}{\lambda_n} \right) (\lambda_n \Delta + o(\Delta)) e(n,i,t) \\ &\quad + \left(\sum_{n:i\notin A_{n+1}} \frac{\lambda_n - \sum_{k\in A_{n+1}} q_{ik}}{\lambda_n} \right) (\lambda_n \Delta + o(\Delta)) e(n,i,t) + o(\Delta) \\ &= \left(\sum_{n:i\notin A_{n+1}} (\lambda_n - q_i) e(n,i,t) + \sum_{k\neq i} q_{ik} \sum_{n:k\notin A_{n+1}} e(n,i,t) \right) \Delta + o(\Delta). \end{aligned}$$

The transition probabilities for state $\perp$ follow directly from the fact that $\perp$ is absorbing. Thus, the proof of Theorem 2 is completed. □

An immediate consequence of Theorem 2 is that $\bar{X}$ has the infinitesimal generator function $\bar{Q}(t) \equiv \left(\bar{q}_{ij}(t) : i, j \in \mathbb{N} \cup \{\perp\}\right)$ for $t \geq 0$ with

$$\bar{q}_{ij}(t) = \begin{cases} q_{ij} \sum_{n: j \in A_{n+1}} e(n,i,t), & \text{if } i \neq j,\ i, j \in \mathbb{N}, \\ -\left(\sum_{n: i \in A_{n+1}} q_i e(n,i,t) + \sum_{n: i \notin A_{n+1}} \lambda_n e(n,i,t)\right), & \text{if } i = j,\ i \in \mathbb{N}, \\ \sum_{n: i \notin A_{n+1}} (\lambda_n - q_i) e(n,i,t) + \sum_{k \neq i} q_{ik} \sum_{n: k \notin A_{n+1}} e(n,i,t), & \text{if } i \in \mathbb{N},\ j = \perp, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

As usual we denote the exit rate $-\bar{q}_{ii}(t)$ as $\bar{q}_i(t)$ for any state $i \in \mathbb{N} \cup \{\perp\}$ and any time point $t \geq 0$. Note that $\bar{Q}(t)$ is conservative for all finite t, i.e., $\sum_{j \in \mathbb{N} \cup \{\perp\}} \bar{q}_{ij}(t) = 0$.

Corollary 3. *For any two distinct states $i, j \in \mathbb{N}$ and any time point $t \geq 0$ we have $\bar{q}_{ij} \leq q_{ij}$.*

Corollary 3 follows from Theorem 2 and the fact that $\sum_{n: j \in A_{n+1}} e(n,i,t) \leq 1$.

In the following, we study the relationship between the transient distributions of X and $\bar{X}$. For this purpose we will construct an auxiliary Markov chain $\widetilde{X}$ and then establish relationships between the transient distributions of $\widetilde{X}$ and $\bar{X}$ as well as between the transient distributions of $\widetilde{X}$ and X.

3.2. The shadow process $\widetilde{X}$

Let $\widetilde{X} = \{\widetilde{X}(t), t \geq 0\}$ be a nonhomogeneous CTMC taking values in the state space $\mathbb{N} \cup \{\perp_i : i \in \mathbb{N}\}$. This CTMC $\widetilde{X}$, called shadow process, behaves similar to $\bar{X}$, except that the special state $\perp$ is split into states $\perp_i$ for $i \in \mathbb{N}$. The idea is that the states $\perp_i, i \in \mathbb{N}$, mimic the behavior of the original Markov chain X.

Let the shadow process $\widetilde{X}$ have the following initial distribution. For $i \in \mathbb{N}$,

$$Pr(\widetilde{X}(0) = i) = \begin{cases} Pr(\bar{X}(0) = i), & \text{if } i \in A_0, \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

and

$$Pr(\widetilde{X}(0) = \perp_i) = \begin{cases} 0, & \text{if } i \in A_0, \\ Pr(X(0) = i), & \text{otherwise.} \end{cases} \tag{10}$$

Furthermore, let the shadow process have the infinitesimal generator function $\tilde{Q}(t) \equiv \left(\tilde{q}_{ij}(t) : i, j \in \mathbb{N} \cup \{\perp_i : i \in \mathbb{N}\}\right)$ for $t \geq 0$ such that for $i, j \in \mathbb{N}$,

$$\tilde{q}_{ij}(t) = \bar{q}_{ij}(t), \tag{11}$$

$$\tilde{q}_{i\perp_j}(t) = \begin{cases} q_{ij} \sum_{n: j \notin A_{n+1}} e(n,i,t), & \text{if } i \neq j, \\ \sum_{n: i \notin A_{n+1}} (\lambda_n - q_i) e(n,i,t), & \text{if } i = j, \end{cases} \tag{12}$$

and

$$\tilde{q}_{\perp_i \perp_j}(t) = q_{ij}. \tag{13}$$

All other entries of $\tilde{Q}(t)$ are zero. Note that we have $\sum_{j \in \mathbb{N}} \tilde{q}_{i\perp_j}(t) = \bar{q}_{i\perp}(t)$ for any state $i \in \mathbb{N}$. Moreover, since Q and $\bar{Q}(t)$ are conservative, so is $\tilde{Q}(t)$. Similarly, the non-explosiveness of X and $\bar{X}$ ensures the non-explosiveness of $\widetilde{X}$. Thus, the generator function $\tilde{Q}$ uniquely determines the transient distributions of $\widetilde{X}$. This follows from the fact that if $\widetilde{X}$ makes infinitely many jumps in a finite amount of time with probability greater than zero, then either this occurs within the subset of states $\mathbb{N}$, and then $\bar{X}$ must explode, or explosion occurs within the subset of states $\{\perp_i : i \in \mathbb{N}\}$, but then X must explode.

Theorem 3. *For all $t \geq 0$ and any $i \in \mathbb{N}$,*

$$\begin{aligned} Pr(\bar{X}(t) = i) &= Pr(\widetilde{X}(t) = i), \\ Pr(X(t) = i) &= Pr(\widetilde{X}(t) = i) + Pr(\widetilde{X}(t) = \perp_i). \end{aligned}$$

Moreover, for all $t \geq 0$,

$$Pr(\bar{X}(t) = \perp) = \sum_{k \in \mathbb{N}} Pr(\widetilde{X}(t) = \perp_k).$$

Proof. First we consider the probabilities to be in state $k \in \mathbb{N}$ both for $\widetilde{X}$ and $\bar{X}$. We write $\bar{p}(t)$ and $\tilde{p}(t)$ for the transient probability distribution of $\bar{X}$ and $\widetilde{X}$, respectively. Moreover, let i, j, k, m, n range over $\mathbb{N}$. Then the forward Kolmogorov equations of $\bar{X}$ are as follows.

$$\begin{aligned}\frac{d}{dt}\bar{p}_k(t) &= \sum_{i\neq k}\bar{q}_{ik}(t)\bar{p}_i(t) - \sum_{j\neq k}\bar{q}_{kj}(t)\bar{p}_k(t) - \bar{q}_{k\perp}(t)\bar{p}_k(t)\\ &= \sum_{i\neq k}\bar{q}_{ik}(t)\bar{p}_i(t) - \sum_{j\neq k}\bar{q}_{kj}(t)\bar{p}_k(t)\\ &\quad - \left(\sum_{n:k\notin A_{n+1}}(\lambda_n - q_k)e(n,k,t) + \sum_{m\neq k} q_{km}\sum_{n:m\notin A_{n+1}} e(n,k,t)\right)\bar{p}_k(t).\end{aligned}$$

For the shadow process we find, using (11) and (12)

$$\begin{aligned}\frac{d}{dt}\tilde{p}_k(t) &= \sum_{i\neq k}\tilde{q}_{ik}(t)\tilde{p}_i(t) - \sum_{j\neq k}\tilde{q}_{kj}(t)\tilde{p}_k(t) - \tilde{q}_{k\perp_k}(t)\tilde{p}_k(t) - \sum_{m\neq k}\tilde{q}_{k\perp_m}(t)\tilde{p}_k(t)\\ &= \sum_{i\neq k}\bar{q}_{ik}(t)\tilde{p}_i(t) - \sum_{j\neq k}\bar{q}_{kj}(t)\tilde{p}_k(t)\\ &\quad - \left(\sum_{n:k\notin A_{n+1}}(\lambda_n - q_k)e(n,k,t) + \sum_{m\neq k} q_{km}\sum_{n:m\notin A_{n+1}} e(n,k,t)\right)\tilde{p}_k(t).\end{aligned}$$

Thus, the forward Kolmogorov equations of $\bar{X}$ and $\widetilde{X}$ are identical. Since $\bar{X}$ and $\widetilde{X}$ are non-explosive these forward equations have unique solutions. Moreover, (9) implies $\bar{p}_k(0) = \tilde{p}_k(0)$ for all $k \in \mathbb{N}$. Thus, it follows $\bar{p}_k(t) = \tilde{p}_k(t)$ for all $t \geq 0$.

Now we consider the relationship between the $\perp$-states of $\widetilde{X}(t)$ and the $\perp$-state of $\bar{X}(t)$. For the shadow process, we get

$$\begin{aligned}\frac{d}{dt}\sum_k \tilde{p}_{\perp_k}(t) &= \sum_k \frac{d}{dt}\tilde{p}_{\perp_k}(t)\\ &= \sum_k\left[\sum_{i\neq k}\tilde{q}_{\perp_i\perp_k}(t)\tilde{p}_{\perp_i}(t) - \sum_{j\neq k}\tilde{q}_{\perp_k\perp_j}(t)\tilde{p}_{\perp_k}(t) + \sum_m \tilde{q}_{m\perp_k}(t)\tilde{p}_m(t)\right]\\ &= \sum_k\sum_{i\neq k}\tilde{q}_{\perp_i\perp_k}(t)\tilde{p}_{\perp_i}(t) - \sum_j\sum_{k\neq j}\tilde{q}_{\perp_k\perp_j}(t)\tilde{p}_{\perp_k}(t) + \sum_k\sum_m \tilde{q}_{m\perp_k}(t)\tilde{p}_m(t).\end{aligned}$$

The first two sums are identical. Splitting the third term according to the cases $m \neq k$ and $m = k$, respectively, and applying (12) yields that the above equals

$$\sum_k\sum_{m\neq k} q_{mk}\sum_{n:k\notin A_{n+1}} e(n,m,t)\tilde{p}_m(t) + \sum_k\sum_{n:k\notin A_{n+1}}(\lambda_n - q_k)e(n,k,t)\tilde{p}_k(t).$$

For $\bar{X}(t)$, we get

$$\begin{aligned}\frac{d}{dt}\bar{p}_\perp(t) &= \sum_{k\in\mathbb{N}}\bar{q}_{k\perp}(t)\bar{p}_k(t)\\ &= \sum_k\sum_{m\neq k} q_{km}\sum_{n:m\notin A_{n+1}} e(n,k,t)\bar{p}_k(t) + \sum_k\sum_{n:k\notin A_{n+1}}(\lambda_n - q_k)e(n,k,t)\bar{p}_k(t)\\ &= \sum_k\sum_{m\neq k} q_{mk}\sum_{n:k\notin A_{n+1}} e(n,m,t)\bar{p}_m(t) + \sum_k\sum_{n:k\notin A_{n+1}}(\lambda_n - q_k)e(n,k,t)\bar{p}_k(t).\end{aligned}$$

Given that $\bar{p}_k(t) = \tilde{p}_k(t)$ for all $k \in \mathbb{N}$ and $t \geq 0$, we have $\frac{d}{dt}\bar{p}_\perp(t) = \frac{d}{dt}\sum_{k\in\mathbb{N}}\tilde{p}_{\perp_k}(t)$. Furthermore, (10) yields $\bar{p}_\perp(0) = \sum_{k\in\mathbb{N}}\tilde{p}_{\perp_k}(0)$. The non-explosiveness of $\bar{X}$ and $\widetilde{X}$ then gives us $\bar{p}_\perp(t) = \sum_k \tilde{p}_{\perp_k}(t)$.

Finally, by a similar approach we show $\tilde{p}_k(t) + \tilde{p}_{\perp_k}(t) = p_k(t)$. For the shadow process, we get

$$\begin{aligned}\frac{d}{dt}\left(Pr(\widetilde{X}(t) = k) + Pr(\widetilde{X}(t) = \perp_k)\right) &= \frac{d}{dt}\tilde{p}_k(t) + \frac{d}{dt}\tilde{p}_{\perp_k}(t)\\ &= \sum_{i\neq k}\tilde{q}_{ik}(t)\tilde{p}_i(t) - \sum_{j\neq k}\tilde{q}_{kj}(t)\tilde{p}_k(t) - \sum_{m\neq k}\tilde{q}_{k\perp_m}(t)\tilde{p}_k(t) - \tilde{q}_{k\perp_k}(t)\tilde{p}_k(t)\\ &\quad + \sum_{i\neq k}\tilde{q}_{\perp_i\perp_k}(t)\tilde{p}_{\perp_i}(t) - \sum_{j\neq k}\tilde{q}_{\perp_k\perp_j}(t)\tilde{p}_{\perp_k}(t) + \sum_{m\neq k}\tilde{q}_{m\perp_k}(t)\tilde{p}_m(t) + \tilde{q}_{k\perp_k}(t)\tilde{p}_k(t)\end{aligned}$$

$$= \sum_{i\neq k} \tilde{q}_{ik}(t)\tilde{p}_i(t) + \sum_{i\neq k} \tilde{q}_{i\perp_k}(t)\tilde{p}_i(t) + \sum_{i\neq k} \tilde{q}_{\perp_i\perp_k}(t)\tilde{p}_{\perp_i}(t)$$
$$- \sum_{j\neq k} \tilde{q}_{kj}(t)\tilde{p}_k(t) - \sum_{j\neq k} \tilde{q}_{k\perp_j}(t)\tilde{p}_k(t) - \sum_{j\neq k} \tilde{q}_{\perp_k\perp_j}(t)\tilde{p}_{\perp_k}(t).$$

Applying (11)–(13) yields that the above equals

$$\sum_{i\neq k} q_{ik}\tilde{p}_i(t)\left[\sum_{n:k\in A_{n+1}} e(n,i,t) + \sum_{n:k\notin A_{n+1}} e(n,i,t)\right] + \sum_{i\neq k} q_{ik}\tilde{p}_{\perp_i}(t)$$
$$- \sum_{j\neq k} q_{kj}\tilde{p}_k(t)\left[\sum_{n:j\in A_{n+1}} e(n,k,t) + \sum_{n:j\notin A_{n+1}} e(n,k,t)\right] - \sum_{j\neq k} q_{kj}\tilde{p}_{\perp_k}(t).$$

Since $\sum_{n=0}^{\infty} e(n,i,t) = 1$ for any state $i \in \mathbb{N}$, we can simplify this and obtain

$$\frac{d}{dt}\left(\tilde{p}_k(t) + \tilde{p}_{\perp_k}(t)\right) = \sum_{i\neq k} q_{ik}(\tilde{p}_i(t) + \tilde{p}_{\perp_i}(t)) - \sum_{j\neq k} q_{kj}\left(\tilde{p}_k(t) + \tilde{p}_{\perp_k}(t)\right).$$

For the Markov chain X we have

$$\frac{d}{dt}p_k(t) = \sum_{i\neq k} q_{ik}p_i(t) - \sum_{j\neq k} q_{kj}p_k(t).$$

From (10) it follows $p_k(0) = \tilde{p}_k(0) + \tilde{p}_{\perp_k}(0)$. Now the non-explosiveness of X and $\widetilde{X}$ yields $p_k(t) = \tilde{p}_k(t) + \tilde{p}_{\perp_k}(t)$ for any $t \geq 0$. □

From Theorem 3 it follows that for any state $i \in \mathbb{N}$ and any time point $t \geq 0$,

$$Pr(\bar{X}(t) = i) = Pr(\widetilde{X}(t) = i) \leq Pr(\widetilde{X}(t) = i) + Pr(\widetilde{X}(t) = \perp_i) = Pr(X(t) = i)$$

and

$$Pr(X(t) = i) = Pr(\widetilde{X}(t) = i) + Pr(\widetilde{X}(t) = \perp_i)$$
$$\leq Pr(\widetilde{X}(t) = i) + \sum_{j\in\mathbb{N}} Pr(\widetilde{X}(t) = \perp_j) = Pr(\bar{X}(t) = i) + Pr(\bar{X}(t) = \perp),$$

which then implies Theorem 1. Thus, the proof of Theorem 1 is completed.

4. Choosing the active sets

In this section, we discuss different strategies for choosing the sets A_n and the uniformization rates λ_n when computing the transient probability distribution of $\bar{X}$ in order to approximate the transient probability distribution of a CTMC X. For a given time point $t \geq 0$, we use (7) to compute the transient distribution of $\bar{X}$. In the sequel, we will focus on

(1). ensuring that the resulting birth process B is non-explosive, such that we can apply Theorem 1, and
(2). keeping the total approximation error $Pr(\bar{X}(t) = \perp)$ small.

First of all, we remark that once the sets A_n are chosen, the rates $\lambda_n = \sup_{i\in A_n} q_i$ yield the slowest possible birth process [24]. In Section 4.4, we will see that this is advantageous from a computational point of view since the slower B the fewer non-zero terms are required in (7). Furthermore, we note that the active sets can be chosen dynamically. That is, given sets $A_0, \ldots, A_{n-1}$ and uniformization rates $\lambda_0, \ldots, \lambda_{n-1}$ it is possible to compute $Pr(Y(m) = i)$ and $Pr(B(t) = m)$ for any epoch $m < n$, state $i \in \mathbb{N} \cup \{\perp\}$, and time point $t \geq 0$, regardless of the choices of later active sets and later uniformization rates. Given that $n > 0$ and the probabilities $Pr(Y(n-1) = i)$ are known, let $f_j^{(n)}$ be a prediction of the probability $Pr(Y(n) = j)$ when $j \in \mathbb{N}$ is chosen to be active in epoch n, that is,

$$f_j^{(n)} = \sum_{i\neq j} Pr(Y(n-1) = i)\frac{q_{ij}}{\lambda_{n-1}} + Pr(Y(n-1) = j)\frac{\lambda_{n-1} - q_j}{\lambda_{n-1}}. \tag{14}$$

For $n = 0$ we set $f_j^{(n)} = p_j(0)$. Then indeed for all $n \in \mathbb{N}$,

$$Pr(Y(n) = j) = \begin{cases} f_j^{(n)}, & \text{if } j \in A_n, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Furthermore, we can also use these predictions to find the amount of probability moving to $\perp$ in every epoch, that is,

$$Pr(Y(n+1) = \perp) - Pr(Y(n) = \perp) = \sum_{j\notin A_{n+1}} f_j^{(n+1)}. \tag{16}$$

Finally, we have

$$Pr(\bar{X}(0) = \perp) = Pr(Y(0) = \perp) = \sum_{j \notin A_0} f_j^{(0)}. \tag{17}$$

Now we can choose active sets and uniformization rates in the following way, where for $n \in \mathbb{N}$ we denote by $f^{(n)}$ the probability vector with components $f_j^{(n)}$.

(1). Compute the prediction vector $f^{(0)}$ using the initial distribution of X.
(2). Choose the active set A_0 based on $f^{(0)}$.
(3). Choose the uniformization rate $\lambda_0 = \sup_{i \in A_0} q_i$.
(4). Compute the probabilities $Pr(Y(0) = i)$ for all $i \in \mathbb{N}$ according to Eq. (15).
(5). Compute the prediction vector $f^{(1)}$ according to Eq. (14).
(6). Iterate the above scheme.

4.1. State-based probability threshold

The simplest way to choose the active sets A_n is to include only those states that have a probability above a certain threshold [26,27]. This corresponds to the use of inexact vector-matrix multiplications, that is, during the computation of the probabilities $Pr(Y(n) = i)$ we remove those entries in the corresponding vector that are below a certain threshold. Given a threshold $\delta > 0$ we choose

$$A_n = \{i \in \mathbb{N} : f_i^{(n)} > \delta\}, \qquad n \in \mathbb{N}.$$

We now prove that this strategy of choosing the active sets ensures that the birth process has bounded rates if X is ergodic and we choose $\lambda_n = \sup_{i \in A_n} q_i$ for all n. To accomplish this we first show that the embedded Markov chain of the subordinated chain Y can be used to 'underapproximate' the distribution of the embedded Markov chain of X in the sense that lower bounds for all transient probabilities of the latter can be obtained. Recall that the embedded Markov chain (or jump chain) of a (discrete- or continuous-time) Markov chain Z is the DTMC Z_e with $Z_e(n) = Z(J_n)$ for all $n \in \mathbb{N}$ where J_n is the nth jump time of Z [9].

Lemma 2. *Given any sequence of active sets and uniformization rates, the transient distribution of the embedded Markov chain of Y is an 'underapproximation' of the transient distribution of the embedded Markov chain of X, which means that for all $n \in \mathbb{N}$ and $i \in \mathbb{N}$,*

$$Pr(Y_e(n) = i) \leq Pr(X_e(n) = i).$$

Proof. Initially, we have

$$Pr(Y(0) = i) \leq Pr(X(0) = i)$$

for any state $i \in \mathbb{N}$. The probability to jump from a state $i \in \mathbb{N}$ to a state $j \in \mathbb{N}$ with $i \neq j$ in the embedded Markov chain of X equals q_{ij}/q_i. For the subordinated Markov chain Y it holds

$$Pr(Y(n+1) = j \mid Y(n) = i, Y(n+1) \neq Y(n), Y(n+1) \neq \perp) = \frac{q_{ij}/\lambda_n}{q_i/\lambda_n} = q_{ij}/q_i$$

and, thus,

$$Pr(Y(n+1) = j \mid Y(n) = i, Y(n+1) \neq Y(n)) \leq q_{ij}/q_i. \tag{18}$$

This means that, regardless of the epoch in which Y jumps to a new state, it always performs this jump with a probability that is at most q_{ij}/q_i. It follows that the jump probabilities of the embedded Markov chain of Y are lower bounds for the jump probabilities of the embedded Markov chain of X. Finally, induction over the number of jumps yields Lemma 2. □

The embedded Markov chains of X and Y are important, because a state in a Markov chain is recurrent if and only if it is recurrent in the embedded Markov chain [9, pp. 185–188]. Furthermore, the ergodicity of a Markov chain requires all its states to be recurrent. This leads us to the following result.

Lemma 3. *Let X be ergodic. For some threshold $\delta > 0$ choose active sets*

$$A_n = \{i \in \mathbb{N} : f_i^{(n)} > \delta\}$$

and uniformization rates $\lambda_n = \sup_{i \in A_n} q_i$, $n \in \mathbb{N}$. Then the rates of the birth process B are bounded.

Proof. We prove Lemma 3 by contradiction and assume that the rates of B are unbounded. Since for any epoch n the rate λ_n is chosen as the supremum exit rate over the states in A_n there exists a sequence of states $\{i_n \in \mathbb{N} : n \in \mathbb{N}\}$ such that

$$q_{i_n} = \lambda_n, \ \ i_n \in A_n, \ \text{ and } \sup_{n \in \mathbb{N}} q_{i_n} = \infty.$$

Given our choice of active sets we further have

$$Pr(Y(n) = i_n) \geq \delta.$$

In other words, the Markov chain Y "goes to infinity" with a probability of at least δ. Since the exit rates of the states $\{i_n\}$ are unbounded, it must follow that all states in A_0 are transient in Y. If a state in A_0 were non-transient (i.e., recurrent) in Y then the probability to return to such a state after a finite amount of epochs would be one [28], but this is a contradiction with the fact that, with probability at least δ, Y goes to infinity as n increases. It follows that the embedded Markov chain of Y goes to infinity as well and then, by Lemma 2, so does the embedded Markov chain of X. We then have that all the states in A_0 are transient in the embedded Markov chain of X and then the same holds for X itself. This completes the proof of Lemma 3, because it is a contradiction with the assumption that X is ergodic. □

Note that the above result ensures that Theorem 1 can be applied if X is ergodic and if we choose the active sets by using a state-wise probability threshold δ. The boundedness of the rates of B ensures that B is non-explosive.

4.2. Global probability threshold

Although using a state-based probability threshold can be easily and efficiently implemented, it suffers from an obvious drawback. If the probability distribution of Y is spread out in such a way that many states have probability less that δ, then the total approximation error may become very large. In the worst case, it may happen that initially all states have probability less than the threshold and then we have $Pr(\bar{X}(t) = \perp) = 1$ for any time point $t \geq 0$. A solution to this problem is to use a global probability threshold $\epsilon > 0$. Therefore, we consider the problem of choosing the A_n given predicted probabilities $f^{(n)}$ such that during one epoch of Y the probability of the state $\perp$ grows by at most ϵ.

Let E_n be the set of states with non-zero predicted probability for epoch n, that is,

$$E_n = \{i \in \mathbb{N} : f_i^{(n)} > 0\}.$$

Since the infinitesimal generator matrix of X has finitely many non-zero entries in each row and A_0 is finite, E_n is also finite for all $n \in \mathbb{N}$. Now, the set of candidate sets $\mathcal{B}_n$ consists of all sets that ensure that at most ϵ probability-mass is "lost". Thus,

$$\mathcal{B}_n = \left\{ B : B \subset E_n, \sum_{j \notin B} f_j^{(n)} \leq \epsilon \right\}.$$

We then choose a subset A_n from $\mathcal{B}_n$ such that the maximal rate in A_n is minimized, that is,

$$\max_{i \in A_n} q_i = \inf_{B \in \mathcal{B}_n} \max_{i \in B} q_i.$$

Note that a set A_n that fulfills the above condition can be found by ordering the states in E_n according to their exit rates. However, this ordering may be costly if E_n is large.

A result similar to Lemma 3 can be easily shown for the global probability threshold. We omit to state it explicitly here.

4.3. Towards a priori error bounds

The global error threshold approach still does not lead to a priori error bounds. For any epoch n,

$$Pr(Y(n) = \perp) \leq (n+1)\epsilon,$$

but this probability may still reach one after $\lceil 1/\epsilon \rceil - 1$ epochs. Furthermore, for a sufficiently large time point $t \geq 0$,

$$Pr(B(t) \geq \lceil 1/\epsilon \rceil) \approx 1,$$

such that the total approximation error will be arbitrarily close to one. It is clear that to find an a priori bound on the total approximation error $Pr(\bar{X}(t) = \perp)$ we must consider both the rates of the birth process B and the time point of interest t.

Theorem 4. *Assume that the birth process B does not explode. Given an error bound $\epsilon > 0$, a time point $t > 0$, active sets A_n, and birth rates λ_n such that for all $n \in \mathbb{N}$,*

$$\lambda_n \sum_{j \notin A_{n+1}} f_j^{(n+1)} \leq \frac{\epsilon}{t}, \tag{19}$$

we have that

$$Pr(\bar{X}(t) = \perp) \leq \sum_{j \notin A_0} f_j^{(0)} + \epsilon. \tag{20}$$

Proof. For the derivative of the error probability $Pr(\bar{X}(t) = \perp)$ we get

$$\begin{aligned}\frac{d}{dt}Pr(\bar{X}(t) = \perp) &= \frac{d}{dt}\sum_{n=0}^{\infty} Pr(Y(n) = \perp)Pr(B(t) = n) \\ &= \sum_{n=0}^{\infty} Pr(Y(n) = \perp)\frac{d}{dt}Pr(B(t) = n). \end{aligned} \tag{21}$$

For the derivatives of the probabilities $Pr(B(t) = n)$ of the birth process B we get

$$\frac{d}{dt}Pr(B(t) = n) = \begin{cases} -\lambda_0 Pr(B(t) = 0), & \text{if } n = 0, \\ \lambda_{n-1}Pr(B(t) = n-1) - \lambda_n Pr(B(t) = n), & \text{otherwise.} \end{cases} \tag{22}$$

Applying (22) to (21) yields

$$\begin{aligned}\frac{d}{dt}Pr(\bar{X}(t) = \perp) &= \sum_{n=1}^{\infty} Pr(Y(n) = \perp)\lambda_{n-1}Pr(B(t) = n-1) - \sum_{n=0}^{\infty} Pr(Y(n) = \perp)\lambda_n Pr(B(t) = n) \\ &= \sum_{n=0}^{\infty} (Pr(Y(n+1) = \perp) - Pr(Y(n) = \perp))\lambda_n Pr(B(t) = n) \\ &= \sum_{n=0}^{\infty} \lambda_n \sum_{j \notin A_{n+1}} f_j^{(n+1)} Pr(B(t) = n) \end{aligned}$$

where the last equality follows from Eq. (16). Using $\sum_{n\in\mathbb{N}} Pr(B(t) = n) = 1$ we obtain

$$\frac{d}{dt}Pr(\bar{X}(t) = \perp) \leq \max_{n\in\mathbb{N}} \lambda_n \sum_{j \notin A_{n+1}} f_j^{(n+1)},$$

and by Eq. (17) it follows

$$Pr(\bar{X}(t) = \perp) \leq \sum_{j \notin A_0} f_j^{(0)} + t \max_{n\in\mathbb{N}} \lambda_n \sum_{j \notin A_{n+1}} f_j^{(n+1)}.$$

Applying the assumption (19) then yields (20). □

Note that it is no problem to choose active sets and uniformization rates to ensure that (19) holds (for instance, by choosing the sets A_n as for adaptive uniformization). However, it must also be ensured that the birth process B does not explode. Therefore, it will be interesting to find necessary and sufficient conditions that ensure non-explosion when the active sets and uniformization rates satisfy (19). However, a detailed discussion of this issue is beyond the scope of the present paper. For more information we refer to Dayar et al. [29].

4.4. Truncating the birth process

Given active sets A_n and uniformization rates λ_n we can find state-wise approximations of the transient probability distribution of X at a time $t \geq 0$ by computing the infinite sum (7). In practice, however, the sum must be truncated since usually analytical solutions are not available. In the case of standard uniformization, where the birth process is a Poisson process, it is well-known that such a truncation is possible for any desired error bound [23]. We will show that the same holds for general, non-explosive birth processes.

Lemma 4. *Let the birth process B be non-explosive. Then for any threshold $\epsilon > 0$ and any time point $t \geq 0$ there exists a right truncation point $R \in \mathbb{N}$ such that*

$$\sum_{n=R+1}^{\infty} Pr(B(t) = n) < \epsilon.$$

Proof. We prove the statement by contradiction. Assume that for some threshold $\epsilon > 0$ and a time point $t \geq 0$ we have

$$\sum_{n=R+1}^{\infty} Pr(B(t) = n) \geq \epsilon$$

for all $R \in \mathbb{N}$. Note that the sum on the left hand side gives the probability of at least $R+1$ jumps in t time-units. Hence,

$$0 < \epsilon \leq \sum_{n=R+1}^{\infty} Pr(B(t) = n) = Pr(B(t) \geq R+1) = 1 - Pr(B(t) < R+1).$$

Thus,

$$Pr(B(t) < R+1) \leq 1-\epsilon < 1$$

for all $R \in \mathbb{N}$. Then it follows

$$\lim_{R\to\infty} Pr(B(t) < R+1) \leq 1-\epsilon < 1.$$

Hence, the probability of B performing finitely many jumps in t time-units is at most $1-\epsilon < 1$ and therefore the probability of B performing infinitely many jumps in t time-units is at least $\epsilon > 0$, which means that B explodes. □

It is important to note that the smaller the birth rates of B, the smaller R will be for any time-point t. Allowing smaller birth rates and thus decreasing R is the main advantage of adaptive uniformization over standard uniformization.

We can now combine Theorem 4 and Lemma 4 to arrive at a computable approximation of the transient distribution of X with a priori error bounds.

Theorem 5. *Given error bounds $\epsilon_1 > 0$, $\epsilon_2 > 0$, active sets A_n, and uniformization rates λ_n, such that, choosing $\epsilon = \epsilon_1$, (19) holds for all $n \in \mathbb{N}$ and the resulting birth process B is non-explosive, there exists a truncation point R such that the active sets A'_n with*

$$A'_n = \begin{cases} A_n, & \text{if } n \leq R, \\ \emptyset, & \text{otherwise,} \end{cases}$$

and uniformization rates λ'_n with

$$\lambda'_n = \begin{cases} \lambda_n, & \text{if } n \leq R, \\ 0, & \text{otherwise,} \end{cases}$$

yield an approximation $\bar{X}'$ with global approximation error

$$Pr(\bar{X}'(t) = \perp) \leq \sum_{j \notin A_0} f_j^{(0)} + \epsilon_1 + \epsilon_2. \tag{23}$$

Proof. Let R be the smallest natural number such that

$$\sum_{n=R+1}^{\infty} Pr(B(t) = n) < \epsilon_2.$$

Since the birth process B is non-explosive, by Lemma 4 such an R exists. Let $\bar{X} = Y(B)$ and $\bar{X}' = Y'(B')$, respectively, be the Markov chains derived from active sets A_n and A'_n, and uniformization rates λ_n and λ'_n, respectively. As the active sets and uniformization rates of $\bar{X}$ and $\bar{X}'$ concur for the first $R+1$ epochs, we have for all $n \in \mathbb{N}$,

$$n \leq R \Rightarrow Pr(Y(n) = i) = Pr(Y'(n) = i),$$

for all $i \in \mathbb{N}$ and

$$n \leq R \Rightarrow Pr(B(t) = n) = Pr(B'(t) = n).$$

Thus, we obtain the total approximation error of $\bar{X}'$ as

$$\begin{aligned} Pr(\bar{X}'(t) = \perp) &= \sum_{n=0}^{\infty} Pr(Y'(n) = \perp) Pr(B'(t) = n) \\ &= \sum_{n=0}^{R} Pr(Y(n) = \perp) Pr(B(t) = n) + \sum_{n=R+1}^{\infty} Pr(Y'(n) = \perp) Pr(B'(t) = n) \\ &\leq \sum_{n=0}^{\infty} Pr(Y(n) = \perp) Pr(B(t) = n) + \sum_{n=R+1}^{\infty} Pr(B'(t) = n) \\ &= Pr(\bar{X}(t) = \perp) + \left(1 - \sum_{n=0}^{R} Pr(B(t) = n)\right). \end{aligned}$$

Applying Theorem 4 and Lemma 4 to the above yields (23), which completes the proof of Theorem 5. □

5. Application example

We consider a stochastic SIR (Susceptible → Infected → Removed) model [30, Ch. 2] with closed homogeneous uniformly mixing population of constant size $N \in \mathbb{N}$ where the individuals in the population are classified according to the standard terminology in epidemic modeling as susceptibles, infectives, and removals (recovered, permanently immune). Susceptibles can become infected when in contact with infectives. Infectives become permanently immune (removals) when being

Table 1
Expectations and standard deviations of the population sizes for all epidemiological classes, and absorption probabilities for the SIR example at different times.

	$t = 10$	$t = 20$	$t = 50$	$t = 100$	$t = 200$
$E[S(t)]$	992.18	970.52	805	733	731.84
$\sigma[S(t)]$	10.4	46.83	254.3	327.25	329.28
$E[I(t)]$	3.67	11.38	26	1.19	0.000661
$\sigma[I(t)]$	6.22	19.99	36.19	4.42	0.07
$E[R(t)]$	4.13	18.09	167	265	268.15
$\sigma[R(t)]$	4.72	27.57	224.58	325.96	329.27
P(absorption)	0.5255	0.5844	0.6035	0.8001	0.9979

Table 2
Comparison of SU, AU, and variants of AAU for the SIR example.

Comparison of methods for time horizon $t = 100$					
Method	$P_\perp$	Runtime (s)	Av. $\|S\|$	Max $\|S\|$	Jumps
SU	0	5361	496,761	496,761	23,856
AU	0	4743	472,321	496,750	20,675
AAU-GPT, $\epsilon = 10^{-15}$	7.9×10^{-12}	3971	161,996	197,384	8716
AAU-AEB, $\epsilon = 10^{-15}$	9.2×10^{-16}	4430	158,291	191,611	9559
AAU-SPT, $\delta = 10^{-50}$	2.1×10^{-44}	2209	250,038	283,918	13,781
AAU-SPT, $\delta = 10^{-30}$	2.0×10^{-24}	1506	185,194	218,161	11,518
AAU-SPT, $\delta = 10^{-15}$	1.6×10^{-09}	754	112,763	143,923	8632
AAU-SPT, $\delta = 10^{-10}$	1.3×10^{-04}	319	74,378	103,329	6815
Comparison of methods for time horizon $t = 200$					
Method	$P_\perp$	Runtime (s)	Av. $\|S\|$	Max $\|S\|$	Jumps
SU	0	49,531	496,761	496,761	47,209
AU	0	49,083	484,797	496,750	42,255
AAU-GPT, $\epsilon = 10^{-15}$	1.4×10^{-11}	11,239	149,183	197,384	14,894
AAU-AEB, $\epsilon = 10^{-15}$	9.6×10^{-16}	14,888	146,367	192,317	16,930
AAU-SPT, $\delta = 10^{-50}$	4.6×10^{-44}	17,574	245,418	283,918	26,473
AAU-SPT, $\delta = 10^{-30}$	4.1×10^{-24}	10,762	175,804	218,161	21,394
AAU-SPT, $\delta = 10^{-15}$	2.9×10^{-09}	3443	98,572	143,923	14,461
AAU-SPT, $\delta = 10^{-10}$	1.8×10^{-04}	443	61,035	103,329	9565

recovered. At any time $t \geq 0$ the discrete random variables $S(t)$, $I(t)$, and $R(t) = N - S(t) - I(t)$ denote the numbers of susceptibles, infectives, and removals, respectively. The epidemic is described by a CTMC $\{(S(t), I(t)), t \geq 0\}$ with finite state space $\{(s, i) : 0 \leq s \leq s_0, 0 \leq i \leq i_0 + s_0 + s\}$ where s_0 and i_0 denote the initial numbers of susceptibles and infectives, respectively. The possible state transitions are from a state (s, i) to a state $(s-1, i+1)$, $s, i > 0$, corresponding to an infection, with transition rate $\beta si/N$ where $\beta > 0$ is a constant contact rate and from a state (s, i) to a state $(s, i-1)$, $i > 0$, corresponding to a recovery from infection, with transition rate γi where $\gamma > 0$ is a constant recovery rate per infected individual. This implies in particular that states $(s, 0)$, $0 \leq s \leq s_0$, are absorbing; all other states are transient.

It is clear that this CTMC with finite state space can be analyzed with SU, AU, and approximate adaptive uniformization, henceforth referred to as AAU. We consider the three variants of AAU introduced in the previous section, that is, AAU with global probability threshold (AAU-GPT), AAU with a priori error bound (AAU-AEB) and AAU with state-based probability threshold (AAU-SPT). The state space is mapped to the set $\mathbb{N}$ according to a numbering of the states that may be different for the different methods. Of course, in this paper our goal is not to present in-depth studies of the model for different parameter values, but the purpose of the example is to illustrate the application and the performance of the different uniformization methods (and corresponding computer implementations) with a particular focus on the number of states that have to be handled by the different methods. For extensive treatments of stochastic epidemic models we refer the reader to, e.g., [30,31]. We emphasize that stochastic epidemic models are not restricted to biological or medical applications such as human or animal diseases, but they also apply to, e.g., the spread of computer viruses and malicious objects in computer networks [2–4]. Note that in order to compare with SU we have to consider an example with finite (but large) state space, because SU cannot deal with infinite state spaces.

As a specific instance of the SIR model we consider a constant population of size $N = 1000$ with initially one infected individual. Hence, the initial state of the CTMC is (999,1). For the contact rate and the recovery rate per infected individual we choose $\beta = 1/3$ and $\gamma = 1/5$. For a comparison of SU, AU, and the three variants of AAU we have computed the transient probability distribution of the SIR model at different times with all methods, where for AAU we have chosen different values for the respective probability thresholds and error bounds, respectively, as described in Sections 4.1–4.3.

Since our focus is not on an in-depth study of the SIR model, we omit to present extensive tables or figures of the whole probability distribution at different times, but with regard to the analysis of the specific model we restrict ourselves to

some representative properties. Table 1 contains for different times the values of the expectations and standard deviations, respectively, of the numbers of susceptibles, infectives, and removals as well as the absorption probability. We emphasize that up to at least six digits there is no difference in these values obtained from the transient probability distributions computed with any of the methods under study. Comparisons of the performance of these methods with regard to the error $P_{\perp}$, the runtime, the average and the maximum number of significant states to be processed, and the number of jumps in the corresponding DTMC when computing transient probability distributions for time horizons $t = 100$ and $t = 200$ are given in Table 2, where AAU-GPT, AAU-AEB, and AAU-SPT refer to AAU with global probability threshold, a priori error bound, and state-based probability threshold, respectively.

As can be seen the runtime can be significantly decreased by AAU while maintaining a reasonable accuracy, e.g. for AAU with $\delta_s = 10^{-10}$ the truncation error is of order 10^{-4}, which is sufficiently small for many if not most applications, and in the case of $t = 200$ the runtime as compared to SU is decreased by a factor of more than one hundred. Hence, AAU yields a great efficiency improvement even for models with a finite state space. Furthermore, as shown by the detailed theoretical foundations in the previous sections, AAU can also deal with infinite state spaces, which is not possible with SU.

6. Conclusion

We have provided a general framework for approximate adaptive uniformization of continuous-time Markov chains, which generalizes and includes as special cases standard uniformization as well as the adaptive uniformization approach based on active sets introduced by van Moorsel and Sanders [24]. A large body of theoretical foundations has been developed, including error bounds for the respective approximations of transient probability distributions. Furthermore, strategies for the choice of our extension of the concept of active sets are given along with proofs that show how certain error bounds can be guaranteed. It has been shown that the general non-explosive birth process used in approximate adaptive uniformization supports truncation of the infinite sum to be computed in a way that allows for approximations up to a predefined accuracy, which means that a priori error bounds can be set and are then met. Approximate adaptive uniformization is particularly useful for numerically computing approximations to the transient probability distributions of multivariate continuous-time Markov chains with huge, possibly infinite state spaces. Thus, it substantially enhances the numerical transient analysis of complex systems and networks that can be modeled as a continuous-time Markov chain. The application and the performance of the different uniformization approaches have been demonstrated by an example of a stochastic epidemic model.

The construction of a computationally tractable approximating model for multivariate continuous-time Markov chains, along with computable error bounds and application guidelines, as developed in this paper, constitutes a flexible and powerful means of gaining new insights into many real-world problems. While much theory and many efficient computational methods for analyzing the long run behavior of complex Markov models that are ubiquitious in real-world applications are available, in particular state space truncation approaches for CTMCs with infinite state spaces [32–34], similarly flexible and powerful approaches to the transient analysis of complex CTMC models have been lacking and are highly desirable. Approximate adaptive uniformization of continuous-time Markov chains renders possible to efficiently analyze transient probability distributions at arbitrary time points as well as the evolution of these probability distributions over time, which has many important managerial implications. For instance, in epidemics it is obviously valuable to have information about the evolution of the epidemic over time rather than only of long run averages. Such time dependent information can be useful in deciding if and when specific medical treatments such as vaccination or actions like quarantine should be conducted. In the computer virus case, network administrators and system managers can activate antivirus programs or update firewall configurations. In call centers and customer contact centers, which are often modeled by complex CTMCs, managers can optimize staffing policies based on, e.g., time dependent probabilities of large customer backlogs or high traffic intensity. In emergency situations the evacuation route selection of large-scale crowds can be based on the transient analysis of corresponding CTMC models. Clearly, any company that operates complex systems, networks or processes can benefit a lot from a proper analysis of transient performance indicators and their changes over time. Hence, approximate adaptive uniformization is useful with regard to many design issues, required actions and decisions in a broad spectrum of real-world systems.

References

[1] C. Ji, D. Jiang, Threshold behaviour of a stochastic SIR model, Appl. Math. Model. 38 (2014) 5067–5079.
[2] J. Amador, The SEIQS stochastic epidemic model with external source of infection, Appl. Math. Model. 40 (2016) 8352–8365.
[3] Z. Wang, X. Fan, Q. Han, Global stability of deterministic and stochastic multigroup SEIQR models in computer network, Appl. Math. Model. 37 (2013) 8673–8686.
[4] L.X. Yang, X. Yang, Y. Wu, The impact of patch forwarding on the prevalence of computer virus: a theoretical assessment approach, Appl. Math. Model. 43 (2017) 110–125.
[5] S. Srivathsan, M. Kamath, Understanding the value of upstream inventory information sharing in supply chain networks, Appl. Math. Model. 54 (2018) 393–412.
[6] J. Wang, J. Sun, S. Lo, Randomness in the evacuation route selection of large-scale crowds under emergencies, Appl. Math. Model. 39 (2015) 5693–5706.
[7] P.S. Mandal, L.J.S. Allen, M. Banerjee, Stochastic modeling of phytoplankton allelopathy, Appl. Math. Model. 38 (2014) 1583–1596.
[8] X. Ding, X. Zhang, L. Ji, Contribution of calcium fluxes to astrocyte spontaneous calcium oscillations in deterministic and stochastic models, Appl. Math. Model. 55 (2018) 371–382.
[9] W.J. Anderson, Continuous-time Markov Chains: An Applications-Oriented Approach, Springer, 1991.
[10] R. Serfozo, Basics of Applied Stochastic Processes, Springer-Verlag, 2009.

[11] E. de Souza e Silva, R. Gail, Transient solutions for Markov chains, in: W.K. Grassmann (Ed.), Computational Probability, Kluwer Academic Publishers, 2000, pp. 43–79.
[12] W.J. Stewart, Introduction to the Numerical Solution of Markov Chains, Princeton University Press, 1995.
[13] A. Jensen, Markoff chains as an aid in the study of Markoff processes, Skand. Aktuarietidskr. 36 (1953) 87–91.
[14] M. Arns, P. Buchholz, A. Panchenko, On the numerical analysis of inhomogeneous continuous time Markov chains, INFORMS J. Comput. 22 (3) (2010) 416–432.
[15] W.K. Grassmann, Transient solutions in Markovian queueing systems, Comput. Oper. Res. 4 (1977) 47–56.
[16] W.K. Grassmann, Finding transient solutions in Markovian event systems through randomization, in: W.J. Stewart (Ed.), Numerical Solution of Markov Chains, Marcel Dekker, New York, 1991, pp. 357–372.
[17] D. Gross, D.R. Miller, The randomization technique as a modeling tool and solution procedure for transient Markov processes, Oper. Res. 32 (2) (1984) 926–944.
[18] A. Hordijk, D.L. Iglehart, R. Schassberger, Discrete time methods for simulating continuous time Markov chains, Adv. Appl. Probab. 8 (1976) 772–788.
[19] J. Keilson, Markov Chain Models – Rarity and Exponentiality, Springer-Verlag, 1979.
[20] A. Reibman, K.S. Trivedi, Numerical transient analysis of Markov models, Comput. Oper. Res. 15 (1) (1988) 19–36.
[21] W. Sandmann, Discrete-time stochastic modeling and simulation of biochemical networks, Comput. Biol. Chem. 32 (4) (2008) 292–297.
[22] B.S. Yoon, J.G. Shanthikumar, Bounds and approximations for the transient behavior of continuous-time Markov chains, Probab. Eng. Inf. Sci. 3 (1989) 175–198.
[23] B.L. Fox, P.W. Glynn, Computing Poisson probabilities, Commun. ACM 31 (4) (1988) 440–445.
[24] A.P.A. van Moorsel, W.H. Sanders, Adaptive uniformization, Stoch. Model. 10 (3) (1994) 619–648.
[25] N.M. van Dijk, Approximate uniformization for continuous-time Markov chains with an application to performability analysis, Stoch. Process. Appl. 40 (2) (1992) 339–357.
[26] R.B. Sidje, K. Burrage, S. MacNamara, Inexact uniformization method for computing transient distributions of Markov chains, SIAM J. Sci. Comput. 29 (6) (2007) 2562–2580.
[27] M. Mateescu, V. Wolf, F. Didier, T.A. Henzinger, Fast adaptive uniformisation of the chemical master equation, IET Syst. Biol. 4 (6) (2010) 441–452.
[28] E. Çinlar, Introduction to Stochastic Processes, Prentice-Hall, 1975.
[29] T. Dayar, H. Hermanns, D. Spieler, V. Wolf, Bounding the equilibrium distribution of Markov population models, Numer. Linear Algebr. Appl. 18 (2011) 931–946.
[30] H. Andersson, T. Britton, Stochastic Epidemic Models and Their Statistical Analysis, Springer, 2000.
[31] M.J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, 2008.
[32] H. Baumann, W. Sandmann, Bounded truncation error for long run averages in infinite Markov chains, J. Appl. Probab. 52 (3) (2015) 609–621.
[33] G. Latouche, P.G. Taylor, Truncation and augmentation of level-independent QBD processes, Stoch. Process. Appl. 99 (2002) 53–80.
[34] R.L. Tweedie, Truncation approximation of invariant measures for Markov chains, J. Appl. Probab. 35 (3) (1998) 517–536.