Spring 2019

Generalizations of Markov Chain Discretizations

John E. Kampmeyer III
Elizabethtown College, kampmeyerj1@etown.edu

Generalizations of Markov Chain Discretizations

MA400: Senior Thesis

John E. Kampmeyer, III

Advisor: Dr. Bogdan Doytchinov
Department of Mathematical Sciences
Elizabethtown College
2019

Generalizations of Markov Chain Discretizations

MA400: Senior Thesis

John E. Kampmeyer, III

Abstract

Markov chains have famously been a crucial tool in understanding stochastic processes and queuing systems, among many other applications. Both discrete-time chains and continuous-time chains have been important centers in both research and application. These two cases are described by transition matrices. Continuous-time chains are difficult to model because this matrix is rather hard to compute in general.

One attack to this problem is approximating a continuous-time chain with one that evolves in discrete time. The transition matrix is still difficult to compute exactly but can also be approximated to any order. The first-order approximation of this quantity is well-known. In 2008, Rachel Irby studied the second-order approximation and compared its performance to that of the first-order counterpart. In this paper, we will generalize the results outlined in Rachel's paper by using an approximation of any order. Specifically, we will study direct comparisons using norms. In addition, we will compare the continuous-time chain to its approximate model through the study of stationary and limiting distributions.

Contents

1 Introduction

Many will already be familiar with the mathematical object known as a Markov chain. As a refresher, we give the following definition:

Definition 1.1. *A stochastic process X_t, which takes values in a state space S, is a Markov chain if for all times $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1}$ and states $x_0, x_1, \ldots, x_n, x_{n+1} \in S$,*

$$\mathbb{P}\{X_{t_{n+1}} = x_{n+1} \mid X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\} = \mathbb{P}\{X_{t_{n+1}} = x_n \mid X_{t_n} = x_n\}. \tag{1}$$

Condition (1) is known as the Markov property and essentially tells us that the future evolution of the system depends only on the present state. The state space S in the definition need not be finite, and in fact, the results herein described are applicable to both finite- and infinite-dimensional Markov chains.

The probabilities given in Definition 1.1 depend on a the time parameter t_i, and it is important noting that this parameter can be discrete or continuous. While this paper studies the behavior of continuous-time Markov chains, discrete-time chains serve as a familiar baseline. We now consider both types.

1.1 Discrete-Time Markov Chains

In discrete-time Markov chains, the time parameter t_i as given in Definition 1.1 can be 0 or any natural number n. At time $t = 0$, the row vector $p(0) = (p_1, p_2, p_3, \ldots)$ describes the initial probability distribution of the system; that is, the system has probability p_1 of being in state 1, probability p_2 of being in state 2, probability p_3 of being in state 3, etc.

As the system evolves, this distribution changes from each discrete moment in time to the next according to what are known as transition probabilities. For states $i, j \in S$, and $n \in \mathbb{N}$,

$$\mathbb{P}\{X_{n+1} = j \mid X_n = i\} =: p_{ij}.$$

The double-index notation hints that we can arrange these transition probabilities into the following one-step transition matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

By the way p_{ij} is defined above, the individual elements correspond to the probabilities that the system transitions from the state having the same-numbered row to the state with the same-numbered column. This is a standard convention in the study of Markov chains.

Consider the ith row of the matrix P as given above, $(p_{i1}, p_{i2}, p_{i3}, \ldots)$. This describes the probability distribution of the system the moment of time after it enters state i. Because every state is represented as a column is this vector, we must have that $p_{i1} + p_{i2} + p_{i3} + \ldots = 1$. Furthermore, every element of the matrix P must be non-negative as a probability. This gives us the following definition:

Definition 1.2. *A square matrix $P = (p_{ij})$ is a transition matrix if and only if*

1. $\forall\, i, j : p_{ij} \geq 0$
2. $\forall\, i : \sum_j p_{ij} = 1.$

The matrix P describes how the system evolves from time $t = m$ to time $t = m + 1$. Repeated matrix multiplication, therefore, tells us the evolution of the system after multiple time steps. Given an initial probability distribution $p(0)$, we can find the distribution at time $t = n$, which we denote as $p(n)$, by the following:

$$p(n) = p(0)P^n.$$

1.2 Continuous-Time Markov Chains

In continuous-time Markov chains, the time parameter t_i is 0 or any positive real number. Just like in the discrete case, $p(0)$ describes the distribution of the system at time $t = 0$. One important difference, however, is the fact that the positive reals have no smallest element, so there is no notion of the "next" moment in time. When studying how the system evolves over time, we are therefore forced to work with probabilities like

$$p_{ij}(t) := \mathbb{P}\{X_{s+t} = j \mid X_s = i\},$$

which describes the probability of the system being in state j at time t after some earlier time s.

If we let $P(t)$ represent the transition matrix, which is dependent of on our continuous time, for a continuous-time Markov chain, then it is not hard to show that the distribution of the system at time t, denoted by $p(t)$, is

$$p(t) = p(0)P(t).$$

We also have that for all times s and t,

$$p(0)P(s+t) = p(0)P(s)P(t). \tag{2}$$

So far, we have formulations which are analogous to the discrete-time case. The pressing question, however, is how we can calculate the transition matrix. Remember, this is not a one-step matrix, since there is no "smallest" time (at least, not mathematically speaking).

The key to this problem is a matrix of the following form:

$$Q = \begin{pmatrix} -\lambda_1 & q_{12} & q_{13} & \cdots \\ q_{21} & -\lambda_2 & q_{23} & \cdots \\ q_{31} & q_{32} & -\lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\lambda_i = \sum_{j \neq i} q_{ij}$ and for all $i \neq j, q_{ij} \geq 0$. This matrix is known as the *infinitesimal generator matrix*, and its elements are commonly called the *intensities* of the Markov chain. The main diagonal elements λ_i represent the instantaneous rate at which the system can transition out of a state i, while the off-diagonal elements $q_{ij}, i \neq j$ represent the instantaneous rate at which the system can transition from state i to state j.

Knowing the matrix Q allows us to represent the system's transition matrix using the matrix exponential; that is,

$$P(t) = e^{tQ} = I + tQ + \frac{1}{2!}t^2Q^2 + \frac{1}{3!}t^3Q^3 + \ldots \tag{3}$$

Calculating this quantity exactly is generally a difficult task and is computationally expensive, using (3) in its raw form. [3] describes various alternatives that increase the efficiency of the numerical computation; however, this will not be the focus of our research. Instead, we will use discrete-time Markov chains to approximate the continuous-time chains and study how these approximations behave in relation to the original chain.

1.3 Discrete-Time Approximations

Suppose we have a continuous-time Markov chain with transition matrix $P(t) = e^{tQ}$ which evolves over the closed, bounded interval $[0, t]$. Again, because this time interval is continuous, there is no next smallest time after time $t = 0$. Therefore, we discretize this interval into N sub-intervals of equal length $\delta := t/N$. We thus have a discrete-time Markov chain with *one-step* transition matrix

$$e^{\delta Q} = I + \delta Q + \frac{1}{2!}\delta^2Q^2 + \frac{1}{3!}\delta^3Q^3 + \ldots =: \tilde{P}. \tag{4}$$

Notice that we once again have a matrix exponential, which we approximate by taking terms up to order m and denoting this by $\tilde{P}_m$, i.e.

$$\tilde{P}_m := I + \delta Q + \frac{1}{2!}\delta^2Q^2 + \ldots + \frac{1}{m!}\delta^mQ^m. \tag{5}$$

Because $\left(e^{\delta Q}\right)^N = e^{tQ}$, $\tilde{P}_m^N$ serves as our one-step, discrete approximation to the continuous-time transition matrix, e^{tQ}. Before we can arrive at any results, we must first find a δ small enough for it to be a transition matrix. It is sufficient to determine the conditions under which $\tilde{P}_m$ is a transition matrix because a transition matrix raised to any integer power is itself a transition matrix.

The first-order approximation of $\tilde{P}$, $\tilde{P}_1 = I + \delta Q$, has long been the center of study and considered a fairly accurate estimation for computational purposes. In her senior thesis (see [1]), Rachel Irby went one step further in studying the behavior of the second-order approximation, $\tilde{P}_2 = I + \delta Q + \frac{1}{2!}\delta^2 Q^2$. The conditions under which both approximations are transition matrices are therein listed. The aim of our research is studying the behavior of $\tilde{P}_m$ for any order m. Must we then produce a bound on δ for every m?

As we will show later in this paper, we only need to require that $\tilde{P}_1$ is a transition matrix for it to follow that $\tilde{P}_m$ is a transition matrix! With that being said, we now reproduce the derivation of the bound on δ which guarantees $\tilde{P}_1$ being a transition matrix.

Recall from Definition 1.2 that a square matrix is a transition matrix if and only if all elements are non-negative and every row sum is 1. In the infinitesimal generator matrix Q, all the off-diagonal elements are non-negative, and so it is clear that all the off-diagonal elements in $\tilde{P}_1 = I + \delta Q$ are non-negative. We therefore must ensure that the main diagonal elements of $\tilde{P}_1$ are non-negative, i.e.

$$1 - \delta\lambda_i \geq 0,$$

where, again, the λ_i are the intensities of the system and i runs over the state space S. In order for the main diagonal elements to be non-negative, we must have

$$0 < \delta \leq \frac{1}{\sup_i \lambda_i}. \tag{6}$$

The condition that every row sum in a transition matrix P must be 1 can be rewritten as $P\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is a column vector of 1's. Likewise, $Q\mathbf{1} = \mathbf{0}$ by the properties of Q, where $\mathbf{0}$ denotes a column vector of 0's. Thus,

$$\tilde{P}_1\mathbf{1} = (I + \delta Q)\mathbf{1} = I\mathbf{1} + \delta Q\mathbf{1} = \mathbf{1} + \delta\mathbf{0} = \mathbf{1}.$$

Therefore, (6) is our desired bound on δ that assures $\tilde{P}_1$ is a transition matrix.

2 Norms

Our desire to measure how close our discrete-time approximations are to the continuous-time matrix exponential necessitates the introduction of a metric on the space of transition matrices and on the space of probability distributions in which we are working. The natural way to accomplish this is to define norms (see [2]) on these spaces, from which we get the metrics induced by these norms, given for $\mathbf{x}, \mathbf{y} \in V$ as the norm of the difference between $\mathbf{x}$ and $\mathbf{y}$.

Let us first begin with a general vector space V of any finite or infinite dimension over a field. As we wish to make our way to the space of probability distributions, we can restrict ourselves to be working over the field of the reals. A norm on V is a function $\|\cdot\| : V \to \mathbb{R}$ such that:

1. $\forall\, \mathbf{v} \in V : \|\mathbf{v}\| \geq 0$ and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$.

2. $\forall\, c \in \mathbb{R}, \forall\, \mathbf{v} \in V : \|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$

3. $\forall\, \mathbf{v}_1, \mathbf{v}_2 \in V : \|\mathbf{v}_1 + \mathbf{v}_2\| \leq \|\mathbf{v}_1\| + \|\mathbf{v}_2\|$.

One family of normed vector spaces that are of particular popularity are the l_p spaces. For $p \geq 1$, these are defined to be the set of all $\mathbf{x} = (x_i)$, where i runs over the dimension of the space, such that $\sum_i |x_i|^p < \infty$. On this set, we define the function $\|\cdot\|_p : l_p \to \mathbb{R}$ by

$$\|\mathbf{x}\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}.$$

Minkowski's inequality gives us that this is indeed a norm on the space l_p.

Among all the l_p norms, the most popular are those corresponding to the values $p = 1, 2$, and ∞. For $p = 1$, we have the "Taxicab" norm,

$$\|\mathbf{x}\|_1 = \sum_i |x_i|,$$

for $p = 2$, we have the Euclidean norm,

$$\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2},$$

and as $p \longrightarrow \infty$, we approach the l_∞ norm,

$$\|\mathbf{x}\|_\infty = \sup_i |x_i|.$$

Let us choose to work with the l_1 norm for the remainder of our research. We do this for three reasons, the first of which we outline now below.

Suppose we have a sequence of vectors $\mathbf{x}^{(n)} = \left(x_i^{(n)}\right)$ converging to $\mathbf{x} = (x_i)$. Then

$$\left\|\mathbf{x}^{(n)} - \mathbf{x}\right\|_1 = \sum_i |x_i^{(n)} - x_i| \longrightarrow 0$$

as $n \longrightarrow \infty$. This implies that $\sup_i |x_i^{(n)} - x_i| \longrightarrow 0$, which implies that for all i, $|x_i^{(n)} - x_i| \longrightarrow 0$. All of these implications are strictly strictly one-sided, and it is worth noting that

$$\sup_i |x_i^{(n)} - x_i| \longrightarrow 0$$

describes the l_∞ norm on the vector space. This means that in general, the l_1 is the strongest of all the l_p norms. Of course, in the finite-dimensional case, this does not mean much, as all norms are equivalent there.

The second reason we choose the l_1 norm is made evident when we look at the space of probability distributions. If we represent a distribution by the vector $\mathbf{p} = (p_i)$ such that all the $p_i \geq 0$ and $\sum_i p_i = 1$, then we get that the space of probability distributions is a subspace of the general vector space over the reals. It is therefore natural that we should have the l_1 norm as the norm induced on this subspace, for

$$\|\mathbf{p}\|_1 = \sum_i |p_i| = \sum_i p_i = 1 \tag{7}$$

The third reason we choose the l_1 norm is made evident as we turn our attention to the space of linear operators $A : V \to V$. This space is again a linear space. We can define the norm of a linear operator A to be

$$\|A\| = \sup \{\|A\mathbf{x}\| \mid \mathbf{x} \in V \text{ and } \|\mathbf{x}\| = 1\}.$$

It can be shown that this is indeed a norm on the space of linear operators along with the following properties:

$$\forall\, \mathbf{x} \in V : \|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\| \tag{8}$$

$$\forall\, A, B : V \to V : \|AB\| \leq \|A\| \cdot \|B\| \tag{9}$$

Furthermore, it can be shown that the analogue of the l_1 norm of a linear operator $A = (a_{ij})$ is

$$\|A\| = \sup_i \sum_j |a_{ij}|. \tag{10}$$

Again, the space of transition matrices of a Markov chain is a subspace of the general space of linear operators with this norm induced on it. This tells us that for a transition matrix $P = (p_{ij})$,

$$\|P\| = \sup_i \sum_j |p_{ij}| = \sup_i \{1\} = 1.$$

The fact that the norm of a transition matrix is 1 will be very useful in future calculations.

Another type of linear operator is the infinitesimal generator matrix $Q = (q_{ij})$. Recall that this matrix has the defining property that the elements along the main diagonal is the opposite of the sum of the elements within the same row, i.e. for all i,

$$q_{ii} = -\lambda_i = -\sum_{j \neq i} q_{ij}.$$

The norm of Q, by (10), is therefore

$$\begin{aligned} \|Q\| &= \sup_i \sum_j |q_{ij}| \\ &= \sup_i 2\lambda_i, \end{aligned}$$

or in other words,

$$\|Q\| = 2 \sup_i \lambda_i. \tag{11}$$

For finite-dimensional Q, it is clear that $\|Q\|$ is bounded.

For infinite-dimensional Q, we can ignore the cases in which $\|Q\|$ is unbounded, for we believe there will be sufficiently many situations in which an infinite Q will have interesting cases. A more rigorous argument can be constructed by considering a system with n servers. We will assume that the arrival time of customers is a Poisson process with parameter λ and the service time for the servers is exponentially distributed all with parameter μ. Our state space S will be the number of customers in the whole system.

As new customers arrive in the system, we transition from state 0 to state 1, from state 1 to state 2, etc. all in accordance with the parameter λ. Note that if all servers in our system are busy and a new customer arrives, they are placed in a queue and we transition from state n to state $n+1$. In fact, it is possible for the line to grow infinitely long, and this is precisely the scenario we wish to consider for our infinite-dimensional argument.

As each server completes its service to a customer, the system transitions back one state. From state 1 to state 0, it transitions with parameter μ. What about from state 2 to state 1? In general, suppose we have two servers busy with a customer in the system, and the times of service are $T_1 \sim \exp \mu_1$ and $T_2 \sim \exp \mu_2$. Then $\min\{T_1, T_2\} \sim \exp(\mu_1 + \mu_2)$, that is, the minimum of those two times signify the transition from state 2 to state 1 with parameter $\mu_1 + \mu_2$. In our specific Markov chain in which the service parameters are all the same, the system transitions from state 2 to state 1 with parameter 2μ. Similarly, the system transitions from state 3 to state 2 with parameter 3μ, and from state n to state $n-1$ with parameter $n\mu$.

What about the transition from state $n+k$ to state $n+k-1$ for some $k \geq 1$? In this case, all n servers are busy, and so the system still transitions with parameter $n\mu$.

If we construct the infinitesimal generator matrix Q from this system, we see the following for the first $n+1$ rows, corresponding to the transitions between states 0 through n:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots & 0 & 0 & 0 \\ \mu & -\mu-\lambda & \lambda & 0 & \dots & 0 & 0 & 0 \\ 0 & 2\mu & -2\mu-\lambda & \lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & n\mu & -n\mu-\lambda & \lambda \end{pmatrix}.$$

By our above argument that state $n+k$ transitions to state $n+k-1$, $k \geq 1$, with parameter $n\mu$, we know that this last row in the above matrix Q is repeated (and shifted over with the addition of a 0) from row $n+2$ ad infinitum. We therefore have that the main diagonal elements are

$$-\lambda, -\mu-\lambda, -2\mu-\lambda, \dots, -n\mu-\lambda, -n\mu-\lambda, \dots$$

By (11), we therefore conclude that

$$\|Q\| = 2(n\mu + \lambda) < \infty.$$

As a final, but important, remark, observe that (6), which guarantees that $\tilde{P}_1$ is a transition matrix, can be rewritten as the following:

$$0 < \delta \leq \frac{2}{\|Q\|}. \tag{12}$$

3 Short-term Behavior

Recall that we discretize the continuous interval $[0, t]$, where $t > 0$ is finite, into N equal sub-intervals so that the matrix exponential $e^{\delta Q}, \delta = t/N$, is the transition matrix for a discrete-time Markov chain approximation to the continuous case. Therefore, the N-th power of the m-th order approximation to $e^{\delta Q}$, $\tilde{P}_m$, approximates the transition matrix $e^{tQ} = (e^{\delta Q})^N$ of the continuous-time Markov chain. We wish to compare these two matrices using the norms we established in section 2.

More generally, we must have a way to compare the N-th power of two transition matrices. For transition matrices A and B that don't necessarily commute, we have that $\|A\| = \|B\| = 1$ and

$$\begin{aligned} A^n - B^n &= A^n - BA^{n-1} + BA^{n-1} - B^2A^{n-2} + B^2A^{n-2} - \ldots + B^{n-1}A - B^n \\ &= (A - B)A^{n-1} + B(A - B)A^{n-2} + \ldots + B^{n-1}(A - B). \end{aligned}$$

By the triangle inequality, this implies that

$$\begin{aligned} \|A^n - B^n\| &\leq \|A - B\| \left\|A^{n-1}\right\| + \|B\| \|A - B\| \left\|A^{n-2}\right\| + \ldots + \left\|B^{n-1}\right\| \|A - B\| \\ &= n \|A - B\|, \end{aligned}$$

since the fact that A and B are transition matrices implies that any power of A and B is a transition matrix and thus has norm 1. In all, we have that

$$\|A^n - B^n\| \leq n \|A - B\| \tag{13}$$

Before we jump into comparing e^{tQ} and $\tilde{P}_m^N$, we will take a step back and recall some previous results that both motivate, and highlight the significance of, our estimation.

3.1 Comparing e^{tQ} and $\tilde{P}_1^N$

First, we reproduce the proof of a well-known result, as stated in the following:

Theorem 3.1. *Let Q be the infinitesimal generator matrix of a continuous-time Markov chain and let a $t > 0$ be given. Choose an $N \in \mathbb{N}$ such that $0 < \delta \leq 2/\|Q\|$, where $\delta := t/N$. Let $\tilde{P}_1 = I + \delta Q$ be the first-order approximation to $e^{\delta Q}$. Then there exists a real constant C, whose value depends on t and $\|Q\|$ but not δ, such that*

$$\left\|e^{tQ} - \tilde{P}_1^N\right\| \leq C\delta.$$

In other words, under the conditions that $\tilde{P}_1$ is a transition matrix, $\tilde{P}_1^N$, the transition matrix of the discrete-time approximation Markov chain, is a first-order approximation to e^{tQ}, the transition matrix of the continuous-time Markov chain. Let us now look at the proof.

Proof. We first look at the difference between $e^{\delta Q}$ and $\tilde{P}_1$:

$$e^{\delta Q} - \tilde{P}_1 = \frac{1}{2!}\delta^2 Q^2 + \frac{1}{3!}\delta^3 Q^3 + \frac{1}{4!}\delta^4 Q^4 + \ldots,$$

which implies that

$$\begin{aligned}
\left\|e^{\delta Q}-\tilde{P}_1\right\| &= \left\|\frac{1}{2!}\delta^2Q^2+\frac{1}{3!}\delta^3Q^3+\frac{1}{4!}\delta^4Q^4+\dots\right\| \\
&\le \frac{1}{2!}\delta^2\left\|Q^2\right\|+\frac{1}{3!}\delta^3\left\|Q^3\right\|+\frac{1}{4!}\delta^4\left\|Q^4\right\|+\dots \text{ (by the triangle inequality)} \\
&\le \frac{1}{2!}\delta^2\left\|Q\right\|^2+\frac{1}{3!}\delta^3\left\|Q\right\|^3+\frac{1}{4!}\delta^4\left\|Q\right\|^4+\dots \text{ (by (9))} \\
&= \frac{1}{2!}\delta^2\left\|Q\right\|^2\left(1+\frac{2!}{3!}\delta\left\|Q\right\|+\frac{2!}{4!}\delta^2\left\|Q\right\|^2+\dots\right) \\
&\le \frac{1}{2!}\delta^2\left\|Q\right\|^2\left(1+\frac{1}{1!}\delta\left\|Q\right\|+\frac{1}{2!}\delta^2\left\|Q\right\|^2+\dots\right) \\
&= \frac{1}{2!}\delta^2\left\|Q\right\|^2e^{\delta\|Q\|} \\
&\le \frac{e^2}{2!}\left\|Q\right\|^2\delta^2,
\end{aligned}$$

where the inequality asserted in the last line comes from the fact that

$$0<\delta\le\frac{2}{\|Q\|}$$

implies that $\delta\|Q\|\le 2$. In all, we have that

$$\left\|e^{\delta Q}-\tilde{P}_1\right\|\le\frac{e^2}{2!}\left\|Q\right\|^2\delta^2.$$

By (13), and the definition $\delta := t/N$, we obtain

$$\begin{aligned}
\left\|e^{\delta QN}-\tilde{P}_1^N\right\| &\le N\frac{e^2}{2!}\left\|Q\right\|^2\delta^2 \\
&= \frac{t}{\delta}\frac{e^2}{2!}\left\|Q\right\|^2\delta^2 \\
&= t\frac{e^2}{2!}\left\|Q\right\|^2\delta.
\end{aligned}$$

Therefore, we have that

$$\left\|e^{tQ}-\tilde{P}_1^N\right\|\le C\delta,$$

where $C=t(e^2/2!)\left\|Q\right\|^2$. □

3.2 Comparing e^{tQ} and $\tilde{P}_2^N$

We now reproduce a similar result that arises when comparing e^{tQ} with $\tilde{P}_2^N$. More specifically, we see that the latter matrix, under the conditions that it is a transition matrix, is a second-order approximation to the former. This is given more formally in the following:

Theorem 3.2. *Let Q be the infinitesimal generator matrix of a continuous-time Markov chain and let a $t>0$ be given. Choose an $N\in\mathbb{N}$ such that $0<\delta\le 2/\|Q\|$, where $\delta := t/N$. Let $\tilde{P}_2=I+\delta Q+\frac{1}{2!}\delta^2Q^2$ be the second-order approximation to $e^{\delta Q}$. Then there exists a real constant C, whose value depends on t and $\|Q\|$ but not δ, such that*

$$\left\|e^{tQ}-\tilde{P}_2^N\right\|\le C\delta^2.$$

We now give Irby's proof of this theorem.

Proof. We first look at the difference between $e^{\delta Q}$ and $\tilde{P}_2$:

$$e^{\delta Q} - \tilde{P}_2 = \frac{1}{3!}\delta^3 Q^3 + \frac{1}{4!}\delta^4 Q^4 + \frac{1}{5!}\delta^5 Q^5 + \dots,$$

which implies that

$$\begin{aligned}
\left\| e^{\delta Q} - \tilde{P}_2 \right\| &= \left\| \frac{1}{3!}\delta^3 Q^3 + \frac{1}{4!}\delta^4 Q^4 + \frac{1}{5!}\delta^5 Q^5 + \dots \right\| \\
&\leq \frac{1}{3!}\delta^3 \left\|Q^3\right\| + \frac{1}{4!}\delta^4 \left\|Q^4\right\| + \frac{1}{5!}\delta^5 \left\|Q^5\right\| + \dots \text{ (by the triangle inequality)} \\
&\leq \frac{1}{3!}\delta^3 \|Q\|^3 + \frac{1}{4!}\delta^4 \|Q\|^4 + \frac{1}{5!}\delta^5 \|Q\|^5 + \dots \text{ (by (9))} \\
&= \frac{1}{3!}\delta^3 \|Q\|^3 \left(1 + \frac{3!}{4!}\delta \|Q\| + \frac{3!}{5!}\delta^2 \|Q\|^2 + \dots\right) \\
&\leq \frac{1}{3!}\delta^3 \|Q\|^3 \left(1 + \frac{1}{1!}\delta \|Q\| + \frac{1}{2!}\delta^2 \|Q\|^2 + \dots\right) \\
&= \frac{1}{3!}\delta^3 \|Q\|^3 e^{\delta\|Q\|} \\
&\leq \frac{e^2}{3!} \|Q\|^3 \delta^3,
\end{aligned}$$

where the inequality asserted in the last line comes from the fact that

$$0 < \delta \leq \frac{2}{\|Q\|}$$

implies that $\delta \|Q\| \leq 2$. In all, we have that

$$\left\| e^{\delta Q} - \tilde{P}_2 \right\| \leq \frac{e^2}{3!} \|Q\|^3 \delta^3.$$

By (13), and the definition $\delta := t/N$, we obtain

$$\begin{aligned}
\left\| e^{\delta Q N} - \tilde{P}_2^N \right\| &\leq N\frac{e^2}{3!} \|Q\|^3 \delta^3 \\
&= \frac{t}{\delta}\frac{e^2}{3!} \|Q\|^3 \delta^3 \\
&= t\frac{e^2}{3!} \|Q\|^3 \delta^2.
\end{aligned}$$

Therefore, we have that

$$\left\| e^{tQ} - \tilde{P}_2^N \right\| \leq C\delta^2,$$

where $C = t(e^2/3!) \|Q\|^3$. □

3.3 Comparing e^{tQ} and $\tilde{P}_m^N$

We now report the results of our research, which generalize the previous two by establishing a similar inequality between e^{tQ} and $\tilde{P}_m^N$ for any natural number m. More specifically, we will show that the latter matrix is indeed an m-th order approximation of the former, as would be expected. Furthermore, it will be shown that our result gives a tighter bound than the above estimations.

Recall from the previous two subsections that establishing the respective inequalities first required our checking that $\tilde{P}_1$ and $\tilde{P}_2$ were transition matrices. The condition that $0 < \delta \leq 2/\|Q\|$ assures this for both matrices. But does this hold for the general $\tilde{P}_m$?

The answer to this question is in the affirmative, and we show why it is so by establishing that the m-th order approximation can by represented as a convex sum of powers of the *first*-order approximation. In other words, such a representation tells us that $\tilde{P}_1$ being a transition matrix implies that $\tilde{P}_m$ is a transition matrix!

We formalize our claim in the following:

Theorem 3.3. *Let*

$$\tilde{P} = c_0 I + c_1 \delta Q + c_2 \delta^2 Q^2 + \ldots + c_m \delta^m Q^m$$

where $c_0, c_1, c_2, \ldots, c_m$ *satisfy the following conditions:*

$$c_m \geq 0, \quad c_{m-1} \geq m c_m, \quad c_{m-2} \geq (m-1) c_{m-1}, \quad \ldots, \quad c_1 \geq 2c_2, \quad c_0 \geq c_1.$$

Then there exist $a_0, a_1, a_2, \ldots, a_m \geq 0$ *such that*

$$\tilde{P} = a_0 I + a_1 (I + \delta Q) + a_2 (I + \delta Q)^2 + \ldots + a_m (I + \delta Q)^m.$$

Furthermore, $a_0 + a_1 + a_2 + \ldots + a_m = c_0$.

Proof. Observe first that the conditions on the c_i imply that for every i, $c_i \geq 0$.

We will first show that the real-valued polynomial

$$f(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_m x^m$$

can be represented as

$$g(x) = a_0 + a_1(1+x) + a_2(1+x)^2 + \ldots + a_m(1+x)^m,$$

where $a_0, a_1, a_2, \ldots, a_m \geq 0$ and the c_i satisfy the aforementioned conditions in the theorem. The condition that the a_i add up to c_0 is an automatic consequence of this, since

$$g(0) = a_0 + a_1 + a_2 + \ldots + a_m = f(0) = c_0.$$

If we let $1 + x = y$, then we have that $f(x) = g(1+x)$ if and only if $g(y) = f(y-1)$ and

$$\begin{aligned}
f(y-1) &= c_0 + c_1(y-1) + c_2(y-1)^2 + \ldots + c_m(y-1)^m \\
&= c_0 \\
&\quad + c_1(-1) + c_1(y) \\
&\quad + c_2 \begin{pmatrix} 2 \\ 0 \end{pmatrix} (1) - c_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} y + c_2 \begin{pmatrix} 2 \\ 2 \end{pmatrix} y^2 \\
&\quad + c_3 \begin{pmatrix} 3 \\ 0 \end{pmatrix} (-1) + c_3 \begin{pmatrix} 3 \\ 1 \end{pmatrix} y - c_3 \begin{pmatrix} 3 \\ 2 \end{pmatrix} y^2 + c_3 \begin{pmatrix} 3 \\ 3 \end{pmatrix} y^3 \\
&\quad + \ldots \\
&\quad + c_m \begin{pmatrix} m \\ 0 \end{pmatrix} (-1)^m + c_m \begin{pmatrix} m \\ 1 \end{pmatrix} (-1)^{m-1} y + \ldots + c_m \begin{pmatrix} m \\ m \end{pmatrix} y^m,
\end{aligned}$$

where $\begin{pmatrix} \alpha \\ k \end{pmatrix}$ is the generalized binomial coefficient, given by

$$\begin{pmatrix} \alpha \\ k \end{pmatrix} = \frac{\alpha(\alpha-1)(\alpha-2)\ldots(\alpha-k+1)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot k}.$$

Because $g(y) = a_0 + a_1 y + a_2 y^2 + \ldots + a_m y^m$, we have the linear system

$$\mathbf{A} = \mathbf{B}\mathbf{C},$$

where $\mathbf{A}$ is a column vector of the a_i, $\mathbf{C}$ is a column vector of the c_i, and $\mathbf{B}$ is an $(m+1)\times(m+1)$ matrix with an arbitrary element b_{ij} given by

$$b_{ij} = \left(\begin{array}{c} j \\ i \end{array} \right) (-1)^{i+j},\ 0 \le i,j \le m.$$

This gives us the following formula for the a_i in terms of the c_i:

$$a_i = \sum_{j=0}^{m} (-1)^{i+j} \left(\begin{array}{c} j \\ i \end{array} \right) c_j,\ i = 0,1,2,\ldots,m.$$

We can rewrite this sum for $i \le j \le m$ (since the binomial coefficient is zero for $j < i$) and use an equivalent formula for the binomial coefficient for these values, namely

$$\left(\begin{array}{c} j \\ i \end{array} \right) = \frac{j!}{i!(j-i)!}.$$

This gives us

$$\begin{aligned} a_i = \sum_{j=0}^{m} (-1)^{i+j} \left(\begin{array}{c} j \\ i \end{array} \right) c_j &= \sum_{j=i}^{m} (-1)^{i+j} \frac{j!}{i!(j-i)!} c_j \\ &= \frac{1}{i!} \sum_{j=i}^{m} (-1)^{j-i} \frac{j!}{(j-i)!} c_j \\ &= \frac{1}{i!} \sum_{k=0}^{m-i} (-1)^{k} \frac{(i+k)!}{k!} c_{i+k},\ k = j-i. \end{aligned}$$

Consider an a_i coefficient of the form a_{m-p}, $0 \le p \le m$. Then

$$\begin{aligned} a_{m-p} &= \frac{1}{(m-p)!} \left[\sum_{k=0}^{p} (-1)^k \frac{(m-p+k)!}{k!} c_{m-p+k} \right] \\ &= \frac{1}{(m-p)!} \left[\frac{(m-p)!}{0!} c_{m-p} - \frac{(m-p+1)!}{1!} c_{m-p+1} \right. \\ &\quad + \frac{(m-p+2)!}{2!} c_{m-p+2} - \frac{(m-p+3)!}{3!} c_{m-p+3} \\ &\quad + \ldots \\ &\quad + \frac{(m-p+2k)!}{(2k)!} c_{m-p+2k} - \frac{(m-p+2k+1)!}{(2k+1)!} c_{m-p+2k+1} \\ &\quad + \ldots \\ &\quad \left. + (-1)^p \frac{m!}{p!} c_m \right]. \end{aligned}$$

We can rewrite this sum as the sum of differences of positive numbers by grouping together every other term. The form of this resulting sum depends on the parity of p.

Case 1: Suppose p is odd. Then using our conditions on the c_i, that is, $c_{m-q} \geq (m-q+1)c_{m-q+1}$ for $1 \leq q \leq m$,

$$\begin{aligned}
a_{m-p} &= \frac{1}{(m-p)!}\sum_{j=0}^{\frac{p-1}{2}}\left[\frac{(m-p+2j)!}{(2j)!}c_{m-p+2j} - \frac{(m-p+2j+1)!}{(2j+1)!}c_{m-p+2j+1}\right] \\
&\geq \frac{1}{(m-p)!}\sum_{j=0}^{\frac{p-1}{2}}\left[\frac{(m-p+2j+1)!}{(2j)!}c_{m-p+2j+1} - \frac{(m-p+2j+1)!}{(2j+1)!}c_{m-p+2j+1}\right] \\
&= \frac{1}{(m-p)!}\sum_{j=0}^{\frac{p-1}{2}}(m-p+2j+1)!c_{m-p+2j+1}\left[\frac{1}{(2j)!} - \frac{1}{(2j+1)!}\right] \\
&\geq 0.
\end{aligned}$$

Case 2: Suppose p is even. For $p = 0$, we have that $a_m \geq 0$. For p even, $p \geq 2$, we have

$$\begin{aligned}
a_{m-p} &= \frac{1}{(m-p)!}\left[\frac{m!}{p!}c_m + \sum_{j=0}^{\frac{p}{2}-1}\left[\frac{(m-p+2j)!}{(2j)!}c_{m-p+2j} - \frac{(m-p+2j+1)!}{(2j+1)!}c_{m-p+2j+1}\right]\right] \\
&= \frac{m!}{p!(m-p)!}c_m + \frac{1}{(m-p)!}\sum_{j=0}^{\frac{p}{2}-1}\left[\frac{(m-p+2j)!}{(2j)!}c_{m-p+2j} - \frac{(m-p+2j+1)!}{(2j+1)!}c_{m-p+2j+1}\right] \\
&\geq \frac{m!}{p!(m-p)!}c_m + \frac{1}{(m-p)!}\sum_{j=0}^{\frac{p}{2}-1}(m-p+2j+1)!c_{m-p+2j+1}\left[\frac{1}{(2j)!} - \frac{1}{(2j+1)!}\right] \\
&\geq 0.
\end{aligned}$$

Thus, $a_0, a_1, a_2, \ldots, a_m \geq 0$.

Our above argument establishes that we indeed can represent

$$f(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_m x^m$$

as

$$g(x) = a_0 + a_1(1+x) + a_2(1+x)^2 + \ldots + a_m(1+x)^m,$$

where $a_0, a_1, a_2, \ldots, a_m \geq 0$ and $a_0 + a_1 + a_2 + \ldots + a_m = c_0$. Observe that these properties hold if we evaluate f at δQ, that is

$$f(\delta Q) = c_0 I + c_1(\delta Q) + c_2(\delta Q)^2 + \ldots + c_m(\delta Q)^m = \tilde{P}$$

has the representation

$$g(\delta Q) = a_0 I + a_1(I+\delta Q) + a_2(I+\delta Q)^2 + \ldots + a_m(I+\delta Q)^m.$$

□

Having established this result, we now can show that the following is true:

Corollary 3.3.1. *If $\tilde{P}_1 = I + \delta Q$ is a transition matrix, then*

$$\tilde{P}_m = I + \delta Q + \frac{1}{2!}\delta^2 Q^2 + \ldots + \frac{1}{m!}\delta^n Q^m$$

is a transition matrix.

Proof. Consider the coefficients $c_i = 1/i!$, $0 \le i \le m$. Observe that

$$\begin{aligned} c_1 &= \frac{1}{1!} = \frac{1}{0!} = c_0, \\ 2c_2 &= \frac{2}{2!} = \frac{1}{1!} = c_1, \\ 3c_3 &= \frac{3}{3!} = \frac{1}{2!} = c_2, \\ &\dots, \\ mc_m &= \frac{m}{m!} = \frac{1}{(m-1)!} = c_{m-1}, \\ \text{and } c_m &= \frac{1}{m!} \ge 0. \end{aligned}$$

This tells us that by Theorem 3.3 that there exist $a_0, a_1, a_2, \dots, a_m \ge 0$ such that $a_0 + a_1 + a_2 + \dots + a_m = c_0 = 1$ and

$$\tilde{P}_m = a_0 I + a_1(I + \delta Q) + a_2(I + \delta Q)^2 + \dots + a_m(I + \delta Q)^m.$$

Whenever $\tilde{P}_1 = I + \delta Q$ is a transition matrix, we have that $\tilde{P}_m$ is a convex sum of powers of transition matrices and is thus itself a transition matrix. □

Because $\tilde{P}_m$ is a transition matrix whenever $\tilde{P}_1$ is, $\tilde{P}_m^N$ is also a transition matrix and we are finally ready to compare it to e^{tQ}:

Theorem 3.4. *Let Q be the infinitesimal generator matrix of a continuous-time Markov chain and let a $t > 0$ be given. Choose an $N \in \mathbb{N}$ such that $0 < \delta \le 2/\|Q\|$, where $\delta := t/N$. Let*

$$\tilde{P}_m = I + \delta Q + \frac{1}{2!}\delta^2 Q^2 + \dots + \frac{1}{m!}\delta^m Q^m$$

be the m-th order approximation to $e^{\delta Q}$, where $m \in \mathbb{N}$. Then there exists a real constant C, whose value depends on t and $\|Q\|$ but not δ, such that

$$\left\| e^{tQ} - \tilde{P}_m^N \right\| \le C\delta^m.$$

Proof. We first look at the difference between $e^{\delta Q}$ and $\tilde{P}_m$:

$$e^{\delta Q} - \tilde{P}_m = \frac{\delta^{m+1}}{(m+1)!}Q^{m+1} + \frac{\delta^{m+2}}{(m+2)!}Q^{m+2} + \frac{\delta^{m+3}}{(m+3)!}Q^{m+3} + \dots,$$

which implies that

$$\begin{aligned} \left\| e^{\delta Q} - \tilde{P}_m \right\| &= \left\| \frac{\delta^{m+1}}{(m+1)!}Q^{m+1} + \frac{\delta^{m+2}}{(m+2)!}Q^{m+2} + \frac{\delta^{m+3}}{(m+3)!}Q^{m+3} + \dots \right\| \\ &= \left\| \frac{\delta^{m+1}}{(m+1)!}Q^{m+1} \left(I + \frac{(m+1)!}{(m+2)!}\delta Q + \frac{(m+1)!}{(m+3)!}\delta^2 Q^2 + \dots \right) \right\| \\ &\le \left\| \frac{\delta^{m+1}}{(m+1)!}Q^{m+1} \right\| \left\| I + \frac{(m+1)!}{(m+2)!}\delta Q + \frac{(m+1)!}{(m+3)!}\delta^2 Q^2 + \dots \right\|, \end{aligned}$$

where the inequality in the last line comes from (9). An interesting fact that we will now show is that the matrix in the second factor of the above inequality,

$$P = I + \frac{(m+1)!}{(m+2)!}\delta Q + \frac{(m+1)!}{(m+3)!}\delta^2 Q^2 + \dots,$$

is in fact a transition matrix!

Consider the sequence of matrices P_k defined by

$$P_k = I + \frac{(m+1)!}{(m+2)!}\delta Q + \frac{(m+1)!}{(m+3)!}\delta^2 Q^2 + \ldots + \frac{(m+1)!}{(m+k+1)!}\delta^k Q^k.$$

The coefficients of this series can be rewritten as

$$c_0 = 1, c_1 = \frac{1}{m+2}, c_2 = \frac{1}{(m+2)(m+3)}, \ldots, c_k = \frac{1}{(m+2)(m+3)\ldots(m+k+1)}.$$

Observe that

$$\begin{aligned}
c_1 &< c_0 \\
2c_2 = \frac{2}{(m+2)(m+3)} &< \frac{m+3}{(m+2)(m+3)} = \frac{1}{m+2} = c_1 \\
3c_3 = \frac{3}{(m+2)(m+3)(m+4)} &< \frac{m+4}{(m+2)(m+3)(m+4)} = c_2 \\
&\ldots \\
kc_k = \frac{k}{(m+2)(m+3)\ldots(m+k+1)} &< \frac{m+k+1}{(m+2)(m+3)\ldots(m+k+1)} = c_{k-1} \\
c_k = \frac{1}{(m+2)(m+3)\ldots(m+k+1)} &> 0,
\end{aligned}$$

so the P_k satisfy the conditions for representation as a convex combination of powers of $\tilde{P}_1 = I + \delta Q$ by Theorem 3.3 and are thus transition matrices whenever $\tilde{P}_1$ is a transition matrix. As we let k go to infinity, we see that the sequence $P_1, P_2, \ldots, P_k, \ldots$ converges to P by the Weierstrass Convergence Theorem. Using our knowledge that P_k is a transition matrix for all k, we now show that P is also a transition matrix.

The fact that the P_k converge to P can be rewritten as

$$\|P_k - P\| \longrightarrow 0.$$

By (10), we have that

$$\|P_k - P\| = \sup_i \sum_j |p_{ij}^{(k)} - p_{ij}|,$$

where $p_{ij}^{(k)}$ is the i,j-th element of P_k and p_{ij} is the i,j-th element of P. The fact that this quantity goes to 0 means that for all i,

$$\sum_j |p_{ij}^{(k)} - p_{ij}| \longrightarrow 0,$$

which means that for all i, j,

$$|p_{ij}^{(k)} - p_{ij}| \longrightarrow 0,$$

i.e. $p_{ij}^{(k)} \longrightarrow p_{ij}$. Because $p_{ij}^{(k)} \geq 0$ for every i, j, and k, it follows that $p_{ij} \geq 0$ for every i and j.

To show that each row sum of P is 1, we introduce the notion of the positive part and negative part of a real number x. Respectively, these are defined to be

$$x^+ = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

$$x^- = \begin{cases} -x, & \text{if } x < 0 \\ 0, & \text{if } x \geq 0 \end{cases}.$$

It follows from these two definitions that $|x| = x^+ + x^-$ and $x = x^+ - x^-$. In particular,

$$|p_{ij}^{(k)} - p_{ij}| = (p_{ij}^{(k)} - p_{ij})^+ + (p_{ij}^{(k)} - p_{ij})^- \tag{14}$$

$$\text{and} \quad p_{ij}^{(k)} - p_{ij} = (p_{ij}^{(k)} - p_{ij})^+ - (p_{ij}^{(k)} - p_{ij})^-. \tag{15}$$

This implies that for all i,

$$\sum_j |p_{ij}^{(k)} - p_{ij}| = \sum_j (p_{ij}^{(k)} - p_{ij})^+ + \sum_j (p_{ij}^{(k)} - p_{ij})^-. \tag{16}$$

Because we have observed that the sum on the left-hand side goes to zero, it follows that the two sums on the right-hand side must also go to zero. By (15), we have that for all i,

$$\sum_j (p_{ij}^{(k)} - p_{ij})^+ - \sum_j (p_{ij}^{(k)} - p_{ij})^- = \sum_j (p_{ij}^{(k)} - p_{ij}) = \sum_j p_{ij}^{(k)} - \sum_j p_{ij}.$$

The left-hand side goes to zero and, as we have observed earlier, the first sum on the right-hand side is 1. This implies that the second sum on the right-hand side must also be 1. Therefore,

$$P = I + \frac{(m+1)!}{(m+2)!}\delta Q + \frac{(m+1)!}{(m+3)!}\delta^2 Q^2 + \dots$$

is a transition matrix and thus has norm equal to 1. Thus,

$$\begin{aligned}\left\| e^{\delta Q} - \tilde{P}_m \right\| &\leq \left\| \frac{\delta^{m+1}}{(m+1)!} Q^{m+1} \right\| \left\| I + \frac{(m+1)!}{(m+2)!}\delta Q + \frac{(m+1)!}{(m+3)!}\delta^2 Q^2 + \dots \right\| \\ &= \frac{1}{(m+1)!}\delta^{m+1} \left\| Q \right\|^{m+1}.\end{aligned}$$

We now apply (13) and use the definition $\delta := t/N$ to obtain

$$\begin{aligned}\left\| e^{\delta Q N} - \tilde{P}_m^N \right\| &\leq N \frac{1}{(m+1)!} \left\| Q \right\|^{m+1} \delta^{m+1} \\ &= \frac{t}{\delta} \frac{1}{(m+1)!} \left\| Q \right\|^{m+1} \delta^{m+1} \\ &= t \frac{1}{(m+1)!} \left\| Q \right\|^{m+1} \delta^{m}.\end{aligned}$$

Therefore, we have that

$$\left\| e^{tQ} - \tilde{P}_m^N \right\| \leq C\delta^m,$$

where $C = t \left\| Q \right\|^{m+1} / (m+1)!$. □

As a final note, we emphasize that Theorem 3.4 works for all natural m; in particular, it works for $m = 1, 2$. Recall that whenever $\tilde{P}_1$ is a transition matrix, Theorem 3.1 tells us that

$$\left\| e^{tQ} - \tilde{P}_1^N \right\| \leq \frac{te^2 \left\| Q \right\|^2}{2!}\delta$$

and Theorem 3.2 tells us that

$$\left\| e^{tQ} - \tilde{P}_2^N \right\| \leq \frac{te^2 \left\| Q \right\|^3}{3!}\delta^2.$$

Theorem 3.4 give us a (slightly) tighter bound by producing a smaller constant of proportionality without the e^2 factor, i.e.

$$\left\| e^{tQ} - \tilde{P}_1^N \right\| \leq \frac{t \left\| Q \right\|^2}{2!}\delta \quad \text{and} \quad \left\| e^{tQ} - \tilde{P}_2^N \right\| \leq \frac{t \left\| Q \right\|^3}{3!}\delta^2.$$

4 Long-term Behavior

In the previous sections, we considered the closed, bounded interval $[0, t]$ and a continuous-time Markov chain defined on this interval with infinitesimal generator matrix Q and transition matrix $P(t) = e^{tQ}$. We approximated $P(t)$ by discretizing $[0, t]$ into N steps of size $\delta := t/N$, taking $\tilde{P}_m$ to be the m-th order approximation of the one-step transition matrix on the discretized interval, $e^{\delta Q}$, and raising the result to the N-th power to compare it with $P(t) = e^{tQ} = (e^{\delta Q})^N$.

We saw that the measure of closeness between e^{tQ} and $\tilde{P}_m^N$ is described by the following inequality:

$$\left\| e^{tQ} - \tilde{P}_m^N \right\| \leq \frac{t \left\| Q \right\|^{m+1}}{(m+1)!} \delta^m.$$

Observe that this inequality is not very useful as $t \longrightarrow \infty$. For this long-term behavior, we must take a different approach to our research.

4.1 Stationary Distributions

Consider a Markov chain with transition matrix $P(t)$. As a quick aside, remember that in the discrete-time case, $P(t) = P^t$, where $t \in \mathbb{N}$ and P is the one-step transition matrix; in the continuous-time case, $P(t) = e^{tQ}$, where Q is the infinitesimal generator matrix. Given an initial distribution $p(0)$, it is often the case that $p(t) = p(0)P(t)$ "stabilizes" as $t \longrightarrow \infty$. In other words, the following limit exists:

$$\pi = \lim_{t \to \infty} p(0)P(t). \tag{17}$$

The distribution π is known as a steady-state or stationary distribution and is defined as follows:

Definition 4.1. *Let π be a probability distribution for a Markov chain with transition matrix $P(t)$. Then π is called a stationary distribution if*

$$\pi = \pi P(t) \text{ for all } t \geq 0. \tag{18}$$

Condition (18) in this definition comes from the limit in (17) in tandem with (2):

$$\pi = \lim_{s \to \infty} p(0)P(s + t) = \lim_{s \to \infty} p(0)P(s)P(t) = \pi P(t).$$

This condition makes it relatively easier to find stationary distributions for a given Markov chain. In fact, every chain has a stationary distribution, albeit not necessarily a unique one. We can make the process of finding these distributions even easier by deriving simpler, equivalent criteria.

In the discrete-time case, we have the one-step transition matrix P and $P(t) = P^t$ for $t \in \mathbb{N}$. It is a simple exercise to verify by induction that

$$\pi = \pi P^t \text{ if and only if } \pi = \pi P. \tag{19}$$

In the continuous-time case, we have the infinitesimal generator matrix Q and $P(t) = e^{tQ}$. If we assume that $t > 0$, we can rearrange (18) to arrive at

$$\pi \frac{P(t) - I}{t} = 0.$$

If we allow $t \longrightarrow 0$, then we obtain a factor of Q on the left-hand side, i.e.

$$0 = \pi \lim_{t \to 0} \frac{P(t) - I}{t} = \pi Q.$$

On the other hand, if we assume that $\pi Q = 0$, then $\pi Q^k = 0$ for all $k \in \mathbb{N}$ and

$$\begin{aligned}
\pi P(t) &= \pi e^{tQ} \\
&= \pi \left(I + tQ + \frac{1}{2!}t^2Q^2 + \frac{1}{3!}t^3Q^3 + \dots \right) \\
&= \pi + t(\pi Q) + \frac{1}{2!}t^2(\pi Q^2) + \frac{1}{3!}t^3(\pi Q^3) + \dots \\
&= \pi + 0 + 0 + 0 + \dots \\
&= \pi.
\end{aligned}$$

In summary, for the continuous-time case, we have the criterion

$$\pi = \pi P(t) \text{ if and only if } \pi Q = 0.$$

Having established some useful criteria for determining stationary distributions, let us turn our attention to studying the long-term behavior of the m-th order approximations of $e^{\delta Q}$, $\tilde{P}_m$. Irby gives an elegant proof for the following theorem:

Theorem 4.1. *Let Q be the infinitesimal generator matrix for a continuous-time Markov chain and let δ be a real number such that $0 < \delta \leq 2/\|Q\|$. Let $\tilde{P}_1 = I + \delta Q$ and $\tilde{P}_2 = I + \delta Q + \frac{1}{2!}\delta^2 Q^2$ be the first- and second-order approximations, respectively, of $e^{\delta Q}$. Then $e^{\delta Q}$, $\tilde{P}_1$, and $\tilde{P}_2$ all have the same stationary distributions.*

From this theorem, it follows that $\left(e^{\delta Q}\right)^N = e^{tQ}$, $\tilde{P}_1^N$, and $\tilde{P}_2^N$ all have the same stationary distributions by (19). It is an amazing, and quite surprising, result that the long-term behavior of these three matrices is identical. We will not reproduce Irby's proof here, for it turns out that the same is true of all other approximations to $e^{\delta Q}$, and we can instead focus our attention on the general case. We formalize our claim in the following theorem:

Theorem 4.2. *Let Q be the infinitesimal generator matrix of a continuous-time Markov chain and δ be a positive real number such that $0 < \delta < R/\|Q\|$, where R is the positive solution to $\frac{1}{x}(e^x - 1 - x) = 1$. For $m \in \mathbb{N}$, let*

$$\tilde{P}_m = I + \delta Q + \frac{1}{2!}\delta^2 Q^2 + \dots + \frac{1}{m!}\delta^m Q^m$$

be the m-th order approximation of $e^{\delta Q}$. Then $\tilde{P}_m$ and $e^{\delta Q}$ have the same stationary distributions.

Proof. As an opening remark, note that the positive solution to $\frac{1}{x}(e^x - 1 - x) = 1$ is approximately 1.25643. Our oddly more stringent condition on δ thus assures us that $\tilde{P}_1$ is a transition matrix, since

$$0 < \delta < R/\|Q\| \leq 2/\|Q\|,$$

and by Corollary 3.3.1, we are assured that $\tilde{P}_m$ is also a transition matrix.

Let π be a stationary distribution for the m-th order approximation $\tilde{P}_m$. Then

$$\begin{aligned}
\pi \tilde{P}_m = \pi &\iff \pi + \delta \pi Q + \frac{1}{2!}\delta^2 \pi Q^2 + \dots + \frac{1}{m!}\delta^m \pi Q^m = \pi \\
&\iff \delta \pi Q + \frac{1}{2!}\delta^2 \pi Q^2 + \dots + \frac{1}{m!}\delta^m \pi Q^m = 0 \\
&\iff \delta \pi Q \left(I + \frac{1}{2!}\delta Q + \dots + \frac{1}{m!}\delta^{m-1} Q^{m-1} \right) = 0 \\
&\iff \pi Q \left(I + \frac{1}{2!}\delta Q + \dots + \frac{1}{m!}\delta^{m-1} Q^{m-1} \right) = 0. \qquad (20)
\end{aligned}$$

We have that

$$\begin{aligned}
&\left\| \frac{1}{2!}\delta Q + \frac{1}{3!}\delta^2 Q^2 + \ldots + \frac{1}{m!}\delta^{m-1} Q^{m-1} \right\| \\
&\leq \frac{1}{2!}\delta \|Q\| + \frac{1}{3!}\delta^2 \left\|Q^2\right\| + \ldots + \frac{1}{m!}\delta^{m-1} \left\|Q^{m-1}\right\|, \qquad \text{by the triangle inequality} \\
&\leq \frac{1}{2!}\delta \|Q\| + \frac{1}{3!}\delta^2 \|Q\|^2 + \ldots + \frac{1}{m!}\delta^{m-1} \|Q\|^{m-1}, \qquad \text{by (9)} \\
&< \frac{1}{2!}\delta \|Q\| + \frac{1}{3!}\delta^2 \|Q\|^2 + \ldots + \frac{1}{m!}\delta^{m-1} \|Q\|^{m-1} + \frac{1}{(m+1)!}\delta^m \|Q\|^m + \ldots,
\end{aligned}$$

since $\delta \|Q\| > 0$. Denoting $\delta \|Q\|$ by x_0 makes it clearer to see that

$$\begin{aligned}
\left\| \frac{1}{2!}\delta Q + \frac{1}{3!}\delta^2 Q^2 + \ldots + \frac{1}{m!}\delta^{m-1} Q^{m-1} \right\| &< \frac{x_0}{2!} + \frac{x_0^2}{3!} + \ldots + \frac{x_0^{m-1}}{m!} + \frac{x_0^m}{(m+1)!} + \ldots \\
&= \frac{1}{x_0}(e^{x_0} - 1 - x_0) \\
&< 1
\end{aligned}$$

since $x_0 = \delta \|Q\| < R$. Thus, we have that $I + \frac{1}{2!}\delta Q + \ldots + \frac{1}{m!}\delta^{m-1}Q^{m-1}$ is invertible, and we can multiply both sides of (20) by its inverse. Therefore,

$$\pi \tilde{P}_m = \pi \iff \pi Q = 0,$$

i.e. π is a stationary distribution of $\tilde{P}_m$ if and only if it is a stationary distribution of $e^{\delta Q}$. □

Again, it follows from (19) that e^{tQ} has the same stationary distribution as $\tilde{P}_m^N$ for $m \in \mathbb{N}$.

4.2 Limiting Distributions

Another type of behavior that arises as $t \longrightarrow \infty$ is that of limiting distributions, defined as follows:

Definition 4.2. *A probability distribution π^* is called a limiting distribution for a Markov chain with transition matrix $P(t)$ if there exists an initial distribution $p(0)$ such that*

$$\lim_{t \to \infty} p(0)P(t) = \pi^*.$$

Compare this definition with the limit given in (17), which says that given an initial distribution $p(0)$,

$$\lim_{t \to \infty} p(0)P(t) = \pi$$

gives rise to the stationary distribution π. The nuance here is that, with stationary distributions, we start with an initial distribution and let the system stabilize to π as $t \longrightarrow \infty$; with limiting distributions, we start with a distribution π^* and assert that it is a limiting distribution if we can find an initial distribution $p(0)$ to "get there."

It is worth noting that if a distribution π^* is a limiting distribution, then it is a stationary distribution, i.e. Definition 4.2 implies Definition 4.1. Why is this so? Consider a limiting distribution π^* for a Markov chain. Then there exists an initial distribution $p(0)$ for which

$$p(0)P(s) \longrightarrow \pi^*$$

as $s \longrightarrow \infty$. Then for all $t \geq 0$,

$$p(0)P(s+t) \longrightarrow \pi^*.$$

On the other hand,

$$p(0)P(s+t) = p(0)P(s)P(t) \longrightarrow \pi^* P(t),$$

so π^* is a stationary distribution.

The converse, on the other hand, is not always true. In the discrete case, consider a Markov chain with one-step transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then the distribution $\pi = (1/2, 1/2)$ is a stationary distribution, since $\pi P = \pi$. Due to the periodic nature of P, however, the only way we can assert that π is a limiting distribution is if we start at the initial distribution $p(0) = \pi$.

We also note that if π^* is a limiting distribution for a transition matrix $P(t)$, then it is a limiting distribution for $P(t)^k$ for some $k \in \mathbb{N}$.

We now turn our attention to studying how close the limiting distributions of a continuous-time Markov chain are to those of its discrete-time approximations. We first remind the reader of Chebyshev's Inequality, given in the following theorem:

Theorem 4.3. *(Chebyshev's Inequality) Let X be a random variable with finite mean μ and finite variance σ^2. Then for all $k > 0$,*

$$\mathbb{P}\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

From this follows the following lemma, given by Irby:

Lemma 4.4. *Let X be a random variable whose distribution depends on a positive number θ and assume that $\mathbb{E}(X) = \theta$ and $\mathbb{V}ar(X) \leq C\theta$, where the constant C does not depend on θ. Then for every real number a,*

$$\lim_{\theta \to \infty} \mathbb{P}\{X \leq a\} = 0.$$

Using this lemma and the uniformization formula (see [4]),

$$e^{tQ} = \sum_{k=0}^{\infty} \frac{(t/\delta)^k}{k!} e^{-t/\delta} (I + \delta Q)^k, \tag{21}$$

Irby proves the following theorem:

Theorem 4.5. *Let $\tilde{P}_1 = I + \delta Q$ be the first-order approximation to $e^{\delta Q}$ and let $p(0)$ be an initial distribution for which there exists the limit*

$$\lim_{n \to \infty} p(0)\tilde{P}_1^n =: \pi^*.$$

Then, for the same distribution $p(0)$, we have

$$\lim_{t \to \infty} p(0)e^{tQ} = \pi^*.$$

In other words, if π^* is a limiting distribution for the transition matrix $\tilde{P}_1$, then it is a limiting distribution for e^{tQ}. Note, however, that we are comparing $\tilde{P}_1$ and e^{tQ}, the first of which does not approximate the second; rather, $\tilde{P}_1^N$ approximates e^{tQ}. A better argument is to show that $\tilde{P}_1$ and $e^{\delta Q}$ have the same limiting distribution π^*, from which it follows that $\tilde{P}_1^N$ and e^{tQ} have the same limiting distribution. The details of the proof for Theorem 4.5 essentially do not change.

Using a formula similar to (21) and Lemma 4.4, Irby also proves that if π^* is a limiting distribution for $\tilde{P}_1$, then it is a limiting distribution for $\tilde{P}_2$, i.e. the following theorem is true:

Theorem 4.6. *Let $\tilde{P}_1 = I + \delta Q$ and $\tilde{P}_2 = I + \delta Q + \frac{1}{2}\delta^2 Q^2$ be the first- and second-order approximations to $e^{\delta Q}$, respectively. Let $p(0)$ be an initial distribution for which there exists the limit*

$$\lim_{n\to\infty} p(0)\tilde{P}_1^n =: \pi^*.$$

Then, for the same distribution $p(0)$, we have

$$\lim_{n\to\infty} p(0)\tilde{P}_2^n = \pi^*.$$

It is upon this result that we wish to seek to prove a generalization that if π^* is a limiting distribution for $\tilde{P}_1$, then it is a limiting distribution for $\tilde{P}_m$. We formalize this below:

Theorem 4.7. *Let $\tilde{P}_1 = I + \delta Q$ denote the transition matrix of the first order approximation and let $\tilde{P}_m = I + \delta Q + \frac{1}{2!}\delta^2 Q^2 + \ldots + \frac{1}{m!}\delta^m Q^m$ denote the transition matrix of the m-th order approximation. Let π^* be a distribution for which there exists an initial distribution $p(0)$ such that*

$$\lim_{n\to\infty} p(0)\tilde{P}_1^n =: \pi^*.$$

Then, for the same distribution $p(0)$, we have

$$\lim_{n\to\infty} p(0)\tilde{P}_m^n = \pi^*.$$

Proof. By Corrollary 3.3.1, we have that there exist $a_0, a_1, a_2, \ldots, a_m \geq 0$ such that

$$\tilde{P}_m = a_0 + a_1(I + \delta Q) + a_2(I + \delta Q)^2 + \ldots + a_m(I + \delta Q)^m.$$

Consider the i.i.d. random variables $X_1, X_2, \ldots X_n$ each with probability mass function $p(i) = a_i, 0 \leq i \leq m$. Additionally, let $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}\text{ar}(X_i) = \sigma^2$. If we denote the probability generating function $a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m$ of the X_i by $g_{X_i}(t)$, then it follows that

$$\tilde{P}_m = g_{X_i}(I + \delta Q).$$

Define $Z_n := X_1 + X_2 + \ldots + X_n$. Then

$$\begin{aligned}
\tilde{P}_m^n &= \left(g_{X_i}(I + \delta Q)\right)^n \\
&= g_{Z_n}(I + \delta Q) \\
&= \sum_{k=0}^{mn} \mathbb{P}\{Z_n = k\}(I + \delta Q)^k \\
&= \sum_{k=0}^{\infty} \mathbb{P}\{Z_n = k\}(I + \delta Q)^k,
\end{aligned}$$

where all but finitely many probabilities are zero.

We are given that

$$p(0)\tilde{P}_1^n \longrightarrow \pi^* \text{ as } n \to \infty,$$

so for an arbitrary $\varepsilon > 0$, we can find an n_0 such that for all $n \geq n_0$,

$$\left\| p(0)\tilde{P}_1^n - \pi^* \right\| < \frac{\varepsilon}{2}.$$

Note that since π^* is a limiting distribution for $\tilde{P}_1$, it is a stationary distribution. By Theorem 4.2, it is therefore a stationary distribution for $\tilde{P}_m$, i.e.

$$\pi^* = \pi^* \tilde{P}_m.$$

Then

$$\begin{aligned}\left\|p(0)\tilde{P}_m^n - \pi^*\right\| &= \left\|p(0)\tilde{P}_m^n - \pi^*\tilde{P}_m^n\right\| \\ &= \left\|(p(0)-\pi^*)\tilde{P}_m^n\right\| \\ &= \left\|(p(0)-\pi^*)\sum_{k=0}^{\infty}\mathbb{P}\{Z_n = k\}(I+\delta Q)^k\right\| \\ &= \left\|\sum_{k=0}^{\infty}\mathbb{P}\{Z_n = k\}(p(0)-\pi^*)(I+\delta Q)^k\right\| \\ &\le \sum_{k=0}^{\infty}\mathbb{P}\{Z_n = k\}\left\|(p(0)-\pi^*)\tilde{P}_1^k\right\|.\end{aligned}$$

Note that

$$\left\|(p(0)-\pi^*)\tilde{P}_1^k\right\| = \left\|p(0)\tilde{P}_1^k - \pi^*\right\| \le \|p(0)\|\left\|\tilde{P}_1^k\right\| + \|\pi^*\| = 2.$$

Using this and our bound on $||p(0)\tilde{P}_1^n - \pi^*||$ for $n \ge n_0$, we see that

$$\begin{aligned}\left\|p(0)\tilde{P}_m^n - \pi^*\right\| &\le \sum_{k=0}^{\infty}\mathbb{P}\{Z_n = k\}\left\|(p(0)-\pi^*)\tilde{P}_1^k\right\| \\ &\le \sum_{k=0}^{n_0}\mathbb{P}\{Z_n = k\}\cdot 2 + \sum_{k=n_0+1}^{\infty}\mathbb{P}\{Z_n = k\}\cdot\frac{\varepsilon}{2} \\ &= 2\mathbb{P}\{Z_n \le n_0\} + \frac{\varepsilon}{2}\mathbb{P}\{Z_n > n_0\} \\ &\le 2\mathbb{P}\{Z_n \le n_0\} + \frac{\varepsilon}{2},\end{aligned}$$

by virtue of the fact that $\mathbb{P}\{Z_n > n_0\} \le 1$ as a probability. Observe that

$$\begin{aligned}\mathbb{P}\{Z_n \le n_0\} &= \mathbb{P}\{Z_n - \mathbb{E}(Z_n) \le n_0 - \mathbb{E}(Z_n)\} \\ &= \mathbb{P}\{Z_n - n\mu \le n_0 - n\mu\} \\ &= \mathbb{P}\{n\mu - Z_n \ge n\mu - n_0\} \\ &\le \mathbb{P}\{|n\mu - Z_n| \ge n\mu - n_0\} \\ &\le \frac{\mathbb{V}\mathrm{ar}(n\mu - Z_n)}{(n\mu - n_0)^2}, \quad \text{by Chebyshev's inequality} \\ &= \frac{\mathbb{V}\mathrm{ar}(Z_n)}{(n\mu - n_0)^2} \\ &= \frac{n\sigma^2}{(n\mu - n_0)^2} \\ &= \frac{\sigma^2}{n(\mu - \frac{n_0}{n})^2},\end{aligned}$$

which goes to zero as $n \to \infty$. In particular, this quantity can be made less than $\varepsilon/4$, and so for n large enough,

$$\left\|p(0)\tilde{P}_m^n - \pi^*\right\| \le 2\mathbb{P}\{Z_n \le n_0\} + \frac{\varepsilon}{2} < 2\cdot\frac{\varepsilon}{4} + \frac{\varepsilon}{2} = \varepsilon.$$

Therefore, $p(0)\tilde{P}_m^n \longrightarrow \pi^*$ as $n \to \infty$. □

5 Conclusion and Future Considerations

Our research proved to be fruitful with many results showing the promise of application to further study. We were able to successfully generalize the results from past research which provided a new insight into the study of continuous-time Markov chains.

Our result that a convex sum of powers of the first-order approximation $\tilde{P}_1$ (as a transition matrix) can be used to represent other transition matrices was crucial to our success in understanding how close $\tilde{P}_m^N$ was to e^{tQ}. This representation even allowed us to derive a bound on the normed difference between these two matrices that is more stringent than that of previous research.

Using a more strict inequality on δ, we were able to establish the equivalence of the stationary distributions between e^{tQ} and $\tilde{P}_m^N$ for all $m \in \mathbb{N}$. Using this, our convex sum representation, and a known result from probability, we were able establish one more generalization of previous research, namely that a limiting distribution for $\tilde{P}_1$ is also a limiting distribution for $\tilde{P}_m$.

Presently, we have the one-sided implications that a limiting distribution for $\tilde{P}_1$ is a limiting distribution for e^{tQ} and also for $\tilde{P}_m$. Future considerations may include trying to establish equivalence between these three matrices, i.e. considering if it's possible for a limiting distribution of e^{tQ} to also be a limiting distribution of $\tilde{P}_1$ or if a limiting distribution of $\tilde{P}_m$ can be a limiting distribution of $\tilde{P}_1$. We initially tried to tackle either of these questions from the approach of our powerful convex sum representation. One major hindrance we encountered was the fact that a representation of $\tilde{P}_1$ in terms of either e^{tQ} or $\tilde{P}_m$ would have to be non-convex. Other methods may therefore need to be implemented in order to solve these problems.

6 Acknowledgements

I would like to take this time to thank Dr. Doytchinov not only for his guidance and collaboration on this project, but also for being a mentor to me during my time at Elizabethtown College. Learning from him has been critical to my growth as a mathematician. I will truly miss hearing his daily stories and jokes when I am gone from this institution.

References

[1] Irby, R. M. *Time Discretization of Markov Chains.* Bachelor's thesis, Elizabethtown College, 2008.

[2] Kreyszig, E. *Introductory functional analysis with applications*, vol. 1. wiley New York, 1978.

[3] Moler, C., and Van Loan, C. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review 45*, 1 (2003), 3–49.

[4] Stewart, W. J. *Introduction to the numerical solution of Markov chains.* Princeton University Press, 1994.