Chapter 6

Continuous Time Markov Chains

In Chapter 3, we considered stochastic processes that were discrete in both time and space, and that satisfied the Markov property: the behavior of the future of the process only depends upon the current state and not any of the rest of the past. Here we generalize such models by allowing for time to be continuous. As before, we will always take our state space to be either finite or countably infinite.

A good mental image to have when first encountering continuous time Markov chains is simply a discrete time Markov chain in which transitions can happen at any time. We will see in the next section that this image is a very good one, and that the Markov property will imply that the jump times, as opposed to simply being integers as in the discrete time setting, will be exponentially distributed.

6.1 Construction and Basic Definitions

We wish to construct a continuous time process on some countable state space S that satisfies the Markov property. That is, letting $\mathcal{F}_{X(s)}$ denote all the information pertaining to the history of X up to time s, and letting $j \in S$ and $s \leq t$, we want

$$P\{X(t) = j \mid \mathcal{F}_{X(s)}\} = P\{X(t) = j \mid X(s)\}. \tag{6.1}$$

We also want the process to be time-homogeneous so that

$$P\{X(t) = j \mid X(s)\} = P\{X(t-s) = j \mid X(0)\}. \tag{6.2}$$

We will call any process satisfying (6.1) and (6.2) a time-homogeneous *continuous time Markov chain*, though a more useful equivalent definition in terms of transition rates will be given in Definition 6.1.3 below. Property (6.1) should be compared with the discrete time analog (3.3). As we did for the Poisson process, which we shall see is the simplest (and most important) continuous time Markov chain, we will attempt to understand such processes in more than one way.

Before proceeding, we make the technical assumption that the processes under consideration are right-continuous. This implies that if a transition occurs "at time t", then we take $X(t)$ to be the new state and note that $X(t) \neq X(t-)$.

Example 6.1.1. Consider a two state continuous time Markov chain. We denote the states by 1 and 2, and assume there can only be transitions between the two states (i.e. we do not allow $1 \to 1$). Graphically, we have

$$1 \rightleftarrows 2.$$

Note that if we were to model the dynamics via a discrete time Markov chain, the tansition matrix would simply be

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

and the dynamics are quite trivial: the process begins in state 1 or 2, depending upon the initial distribution, and then deterministically transitions between the two states. At this point, we do not know how to understand the dynamics in the continuous time setting. All we know is that the distribution of the process should only depend upon the current state, and not the history. This does not yet tell us when the firings will occur. □

Motivated by Example 6.1.1, our first question pertaining to continuous time Markov chains, and one whose answer will eventually lead to a general construction/simulation method, is: how long will this process remain in a given state, say $x \in S$? Explicitly, suppose $X(0) = x$ and let T_x denote the time we transition away from state x. To find the distribution of T_x, we let $s, t \geq 0$ and consider

$$\begin{aligned}
&P\{T_x > s + t \mid T_x > s\} \\
&\quad = P\{X(r) = x \text{ for } r \in [0, s+t] \mid X(r) = x \text{ for } r \in [0, s]\} \\
&\quad = P\{X(r) = x \text{ for } r \in [s, s+t] \mid X(r) = x \text{ for } r \in [0, s]\} \\
&\quad = P\{X(r) = x \text{ for } r \in [s, s+t] \mid X(s) = x\} && \text{(Markov property)} \\
&\quad = P\{X(r) = x \text{ for } r \in [0, t] \mid X(0) = x\} && \text{(time homogeneity)} \\
&\quad = P\{T_x > t\}.
\end{aligned}$$

Therefore, T_x satisfies the loss of memory property, and is therefore exponentially distributed (since the exponential random variable is the only continuous random variable with this property). We denote the parameter of the exponential holding time for state x as $\lambda(x)$. We make the useful observation that

$$\mathbb{E}T_x = \frac{1}{\lambda(x)}.$$

Thus, the higher the rate $\lambda(x)$, representing the rate *out* of state x, the smaller the expected time for the transition to occur, which is intuitively pleasing.

Example 6.1.2. We return to Example 6.1.1, though now we assume the rate from state 1 to state 2 is $\lambda(1) > 0$, and the rate from state 2 to state 1 is $\lambda(2) > 0$. We commonly incorporate these parameters into the model by placing them next to the transition arrow in the graph:

$$1 \underset{\lambda(2)}{\overset{\lambda(1)}{\rightleftarrows}} 2.$$

The dynamics of the model are now clear. Assuming $X(0) = 1$, the process will remain in state 1 for an exponentially distributed amount of time, with parameter $\lambda(1)$, at which point it will transition to state 2, where it will remain for an exponentially distributed amount of time, with parameter $\lambda(2)$. This process then continuous indefinitely. □

Example 6.1.2 is deceptively simple as it is clear that when the process transitions out of state 1, it must go to state 2, and vice versa. However, consider the process with states 1, 2, and 3 satisfying

$$1 \rightleftarrows 2 \rightleftarrows 3.$$

Even if you are told the holding time parameter for state 2, without further information you can not figure out wether you transition to state 1 or state 3 after you leave state 2. Therefore, we see we want to study the *transition probabilities* associated with the process, which we do now.

Still letting T_x denote the amount of time the process stays in state x after entering state x, and which we now know is exponentially distributed with a parameter of $\lambda(x)$, we define for $y \neq x$

$$p_{xy} \stackrel{\text{def}}{=} P\{X(T_x) = y \mid X(0) = x\},$$

to be the probability that the process transitions to state y after leaving state x. It can be shown that the time of the transition, T_x, and the value of the new state, y, are independent random variables. Loosely, this follows since if the amount of time the chain stays in state x affects the transition probabilities, then the Markov property (6.1) is not satisfied as we would require to know both the current state *and* the amount of time the chain has been there to know the probabilities associated with ending up in the different states.

We next define

$$\lambda(x, y) \stackrel{\text{def}}{=} \lambda(x)p_{xy}.$$

Since T_x is exponential with parameter $\lambda(x)$, we have that

$$P\{T_x < h\} = 1 - e^{-\lambda(x)h} = \lambda(x)h + o(h), \text{ as } h \to 0.$$

Combining the above, for $y \neq x$ and mild assumptions on the function λ,[1] we have

$$\begin{aligned} P\{X(h) = y \mid X(0) = x\} &= P\{T_x < h, X(T_x) = y \mid X(0) = x\} + o(h) \\ &= \lambda(x)hp_{xy} + o(h) \qquad (6.3) \\ &= \lambda(x, y)h + o(h), \end{aligned}$$

[1] For example, we do not want to let $\lambda(z) = \infty$ for any $z \in E$

as $h \to 0$, where the $o(h)$ in the first equality represents the probability of seeing two or more jumps (each with an exponential distribution) in the time window $[0, h]$. Therefore, $\lambda(x, y)$ yields the *local rate*, or intensity, of transitioning from state x to state y. It is worth explicitly pointing out that for $x \in S$

$$\sum_{y \neq x} \lambda(x, y) = \sum_{y \neq x} \lambda(x) p_{xy} = \lambda(x).$$

Note that we also have

$$\begin{aligned} P\{X(h) = x \mid X(0) = x\} &= 1 - \sum_{y \neq x} P\{X(h) = y \mid X(0) = x\} \\ &= 1 - \sum_{y \neq x} \lambda(x, y) h + o(h) \\ &= 1 - \lambda(x) h \sum_{y \neq x} p_{xy} + o(h) \\ &= 1 - \lambda(x) h + o(h). \end{aligned} \tag{6.4}$$

Similarly to our consideration of the Poisson process, it can be argued that any time homogeneous process satisfying the local conditions (6.3) and (6.4) also satisfies the Markov property (6.1). This is not surprising as the conditions (6.3)-(6.4) only make use of the current state of the system and ignore the entire past. This leads to a formal definition of a continuous time Markov chain that incorporates all the relevant parameters of the model and is probably the most common definition in the literature.

Definition 6.1.3. A time-homogeneous *continuous time Markov chain* with transition rates $\lambda(x, y)$ is a stochastic process $X(t)$ taking values in a finite or countably infinite state space S satisfying

$$\begin{aligned} P\{X(t+h) = x \mid X(t) = x\} &= 1 - \lambda(x) h + o(h) \\ P\{X(t+h) = y \mid X(t) = x\} &= \lambda(x, y) h + o(h), \end{aligned}$$

where $y \neq x$, and $\lambda(x) = \sum_{y \neq x} \lambda(x, y)$.

When only the local rates $\lambda(x, y)$ are given in the construction of the chain, then it is important to recognize that the transition probabilities of the chain can be recovered via the identity

$$p_{xy} = \frac{\lambda(x, y)}{\lambda(x)} = \frac{\lambda(x, y)}{\sum_{y \neq x} \lambda(x, y)}.$$

Example 6.1.4. Let N be a Poisson process with intensity $\lambda > 0$. As N satisfies

$$\begin{aligned} P\{N(t+h) = j+1 \mid N(t) = j\} &= \lambda h + o(h) \\ P\{N(t+h) = j \mid N(t) = j\} &= 1 - \lambda h + o(h), \end{aligned}$$

we see that it is a continuous time Markov chain. Note also that any Poisson process is the continuous time version of the deterministically monotone chain from Chapter 3. □

Example 6.1.5. Consider again the three state Markov chain

$$1 \underset{\lambda(2,1)}{\overset{\lambda(1,2)}{\rightleftarrows}} 2 \underset{\lambda(3,2)}{\overset{\lambda(2,3)}{\rightleftarrows}} 3,$$

where the local transition rates have been placed next to their respective arrows. Note that the holding time in state two is an exponential random variable with a parameter of

$$\lambda(2) \stackrel{\text{def}}{=} \lambda(2,1) + \lambda(2,3),$$

and the probability that the chain enters state 1 after leaving state 2 is

$$p_{21} \stackrel{\text{def}}{=} \frac{\lambda(2,1)}{\lambda(2,1) + \lambda(2,3)},$$

whereas the probability that the chain enters state 3 after leaving state 2 is

$$p_{23} \stackrel{\text{def}}{=} \frac{\lambda(2,3)}{\lambda(2,1) + \lambda(2,3)}.$$

This chain could then be simulated by sequentially computing holding times and transitions. □

An algorithmic construction of a general continuous time Markov chain should now be apparent, and will involve two building blocks. The first will be a stream of unit exponential random variables used to construct our holding times, and the second will be a discrete time Markov chain, denoted X_n, with transition probabilities p_{xy} that will be used to determine the sequence of states. Note that for this discrete time chain we necessarily have that $p_{xx} = 0$ for each x. We also explicitly note that the discrete time chain, X_n, is different than the continuous time Markov chain, $X(t)$, and the reader should be certain to clarify this distinction. The discrete time chain is often called the *embedded* chain associated with the process $X(t)$.

Algorithm 1. (Algorithmic construction of continuous time Markov chain)
Input:

- Let X_n, $n \geq 0$, be a discrete time Markov chain with transition matrix Q. Let the initial distribution of this chain be denoted by α so that $P\{X_0 = k\} = \alpha_k$.
- Let $E_n, n \geq 0$, be a sequence of independent unit exponential random variables.

Algorithmic construction:

1. Select $X(0) = X_0$ according to the initial distribution α.
2. Let $T_0 = 0$ and define $W(0) = E_0/\lambda(X(0))$, which is exponential with parameter $\lambda(X(0))$, to be the waiting time in state $X(0)$.
3. Let $T_1 = T_0 + W(0)$, and define $X(t) = X(0)$ for all $t \in [T_0, T_1)$.

4. Let X_1 be chosen according to the transition matrix Q, and define $W(1) = E_1/\lambda(X_1)$.

5. Let $T_2 = T_1 + W(1)$ and define $X(t) = X_1$ for all $t \in [T_1, T_2)$.

6. Continue process.

Note that two random variables will be needed at each iteration of Algorithm 1, one to compute the holding time, and one to compute the next state of the discrete time Markov chain. In the biology/chemistry context, the algorithm implicit in the above construction is typically called the *Gillespie algorithm*, after Dan Gillespie. However, it (and its natural variants) is also called, depending on the field, the *stochastic simulation algorithm*, *kinetic Monte Carlo*, *dynamic Monte Carlo*, the *residence-time algorithm*, the *n-fold way*, or the *Bortz-Kalos-Liebowitz algorithm*; needless to say, this algorithm has been discovered many times and plays a critical role in many branches of science.

As the future of the process constructed in Algorithm 1 only depends upon the current state of the system, and the current holding time is exponentially distributed, it satisfies the Markov property (6.1). Further, for $y \neq x$ we have

$$\begin{aligned} P\{X(h) = y \mid X(0) = x\} &= P\{X(T_1) = y, T_1 \le h \mid X(0) = h\} + o(h) \\ &= \lambda(x) h p_{xy} + o(h) \\ &= \lambda(x,y) h, \end{aligned}$$

showing we also get the correct local intensities. Therefore, the above construction via a stream of exponentials and an embedded discrete time Markov chain could be taken to be another alternative definition of a continuous time Markov chain.

One useful way to think about the construction in Algorithm 1 is in terms of alarm clocks:

1. When the chain enters state x, independent "alarm clocks" are placed at each state y, and the yth is programed to go off after an exponentially distributed amount of time with parameter $\lambda(x,y)$.

2. When the first alarm goes off, the chain moves to that state, all alarm clock are discarded, and we repeat the process.

Note that to prove that this algorithm is, in fact, equivalent to the algorithmic construction above, you need to recall that the minimum of exponential random variables with parameters $\lambda(x,y)$ is itself exponentially distributed with parameter

$$\lambda(x) = \sum_y \lambda(x,y),$$

and that it is the yth that went off with probability

$$\frac{\lambda(x,y)}{\sum_{j \neq x} \lambda(x,j)} = \frac{\lambda(x,y)}{\lambda(x)}.$$

See Propositions 2.3.18 and 2.3.19.

We close this section with three examples.

Example 6.1.6. We consider again a random walker on $S = \{0, 1, \dots\}$. We suppose the transition intensities are

$$\begin{aligned} \lambda(i, i+1) &= \lambda \\ \lambda(i, i-1) &= \mu, \qquad \text{if } i > 0, \end{aligned}$$

and $\lambda(0, -1) = 0$. Therefore, the probability of the embedded discrete time Markov chain transitioning up if the current state is $i \neq 0$, is $\lambda/(\lambda+\mu)$, whereas the probability of transitioning down is $\mu/(\lambda + \mu)$. When $i \neq 0$, the holding times will always be exponentially distributed with a parameter of $\lambda + \mu$.

Example 6.1.7. We generalize Example 6.1.6 by allowing the transition rates to depend upon the current state of the system. As in the discrete time setting this leads to a birth and death process. More explicitly, for $i \in \{0, 1, \dots, \}$ we let

$$\begin{aligned} \lambda(i, i+1) &= B(i) \\ \lambda(i, i-1) &= D(i), \end{aligned}$$

where $\mu_0 = 0$. Note that the transition rates are now state dependent, and may even be unbounded as $i \to \infty$. Common choices for the rates include

$$\begin{aligned} B(i) &= \lambda i \\ D(i) &= \mu i, \end{aligned}$$

for some scalar $\lambda, \mu > 0$. Another common model would be to assume a population satisfies a logistical growth model,

$$\begin{aligned} B(i) &= ri \\ D(i) &= \frac{r}{K} i^2. \end{aligned}$$

where K is the *carrying capacity.*

Analogously to Example 5.2.18, if we let $X(t)$ denote the state of the system at time t, we have that $X(t)$ solves the stochastic equation

$$X(t) = X(0) + Y_1 \left(\int_0^t B(X(s)) ds \right) - Y_2 \left(\int_0^t D(X(s)) ds \right), \qquad (6.5)$$

where Y_1 and Y_2 are independent unit-rate Poisson processes. As in Example 5.2.18, it is now an exercise to show that the solution to (6.5) satisfies the correct local intensity relations of Definition 6.1.3. For example, denoting

$$A(t) \stackrel{\text{def}}{=} Y_1 \left(\int_0^t B(X(s)) ds \right) D(t) \stackrel{\text{def}}{=} Y_2 \left(\int_0^t D(X(s)) ds \right),$$

we see that

$$\begin{aligned} P\{X(t+h) = x+1 \mid X(t) = x\} \\ &= P\{A(t+h) - A(t) = 1, D(t+h) - D(t) = 0 \mid X(t) = x\} + o(h) \\ &= B(x)h(1 - D(x)h) + o(h) \\ &= B(x)h + o(h). \end{aligned}$$

□

Example 6.1.8. We will model the dynamical behavior of a single gene, the mRNA molecules it produces, and finally the resulting proteins via a continuous time Markov chain. It is an entirely reasonable question to ask whether it makes sense to model the reaction times of such cellular processes via exponential random variables. The answer is almost undoubtably "no," however the model should be interpreted as an approximation to reality and has been quite successful in elucidating cellular dynamics. It is also a much more realistic model than a classical ODE approach, which is itself a crude approximation to the continuous time Markov chain model (we will discuss this fact later).

Consider a single gene that is producing mRNA (this process is called *transcription*) at a constant rate of λ_1, where the units of time are hours, say. Further, we suppose the mRNA molecules are producing proteins (this process is called *translation*) at a rate of $\lambda_2 \cdot (\#\text{mRNA})$, for some $\lambda_2 > 0$. Next, we assume that the mRNA molecules are being degraded at a rate of $d_m \cdot (\#\text{mRNA})$, and proteins are being degraded at a rate of $d_p \cdot (\#\text{proteins})$. Graphically, we may represent this system via

$$\begin{aligned} G &\overset{\lambda(1)}{\to} G + M \\ M &\overset{\lambda(2)}{\to} M + P \\ M &\overset{d_m}{\to} \emptyset \\ P &\overset{d_p}{\to} \emptyset. \end{aligned}$$

It is important to note that this is not the only way to write down these reactions. For example, many in the biological communities would write $M \to P$, as opposed to $M \to M + P$. However, we feel it is important to stress, through the notation $M \to M + P$, that the mRNA molecule is not lost during the course of the reaction.

As the number of genes in the model is assumed to be constant in time, the state space should be taken to be $\mathbb{Z}^2_{\geq 0}$. Therefore, we let $X(t) \in \mathbb{Z}^2_{\geq 0}$ be the state of the process at time t where the first component gives the number of mRNA molecules and the second gives the number of proteins.

Now we ask: what are the possible transitions in the model, and what are the rates? We see that the possible transitions are given by addition of the *reaction vectors*

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -1 \end{bmatrix},$$

with respective rates

$$\lambda_1, \quad \lambda_2 X_1(t), \quad d_m X_1(t), \quad d_p X_2(t).$$

Note that the rate of reaction 3, respectively 4, will be zero when $X_1(t) = 0$, respectively $X_2(t) = 0$. Therefore, non-negativity of the molecules is assured. □

6.2 Explosions

Now that we have a good idea of what a continuous time Markov chain is, we demonstrate a behavior that is not possible in the discrete time setting: explosions. Recall that in Algorithm 1, which constructs a continuous time Markov chain, the value T_n represents the time of the nth transition of the chain. Therefore, the chain so constructed is only defined up until the (random) time

$$T_\infty \stackrel{\text{def}}{=} \lim_{n\to\infty} T_n.$$

If $T_\infty < \infty$, then we say that an *explosion* has happened.

Definition 6.2.1. If

$$P_i\{T_\infty = \infty\} \stackrel{\text{def}}{=} P\{T_\infty = \infty \mid X(0) = i\} = 1, \quad \text{for all } i \in S,$$

than we will say the process is *non-explosive*. Otherwise we will say the process is *explosive*.

Note that a process could be explosive even if

$$P_i\{T_\infty = \infty\} = 1,$$

for some $i \in S$; see Example 6.2.4. It is not too difficult to construct an explosive process. To do so, we will first need the following result pertaining to exponential random variables.

Proposition 6.2.2. *Suppose that $\{E_n\}$, $n \geq 1$, are independent exponential random variables with respective parameters λ_n. Then,*

$$P\left\{\sum_n E_n < \infty\right\} = 1 \quad \iff \quad \sum_n \frac{1}{\lambda_n} < \infty.$$

Proof. We will prove one direction of the implication (the one we will use). For the other direction, see [13, Section 5.1]. We suppose that $\sum_n \frac{1}{\lambda_n} < \infty$. Because $\sum_n E_n \geq 0$ and

$$\mathbb{E}(\sum_n E_n) = \sum_n \mathbb{E}E_n = \sum_n \frac{1}{\lambda_n} < \infty,$$

we may conclude that $\sum_n E_n < \infty$ with probability one. □

Thus, we see that we can construct an explosive birth process by requiring that the holding times satisfy $\sum_n 1/\lambda(X_n) < \infty$.

Example 6.2.3. Consider a pure birth process in which the embedded discrete time Markov chain is the deterministically monotone chain of Example 3.1.5. Suppose that the holding time parameter in state i is $\lambda(i)$. Finally, let $X(t)$ denote the state of the continuous time process at time t. Note that the stochastic equation satisfied by X is

$$X(t) = X(0) + N\left(\int_0^t \lambda(X(s))ds\right).$$

Suppose that $\lambda(n) = \lambda n^2$ for some $\lambda > 0$ and that $X(0) = 1$. Then the nth holding time is determined by an exponential random variable with parameter λn^2, which we denote by E_n. Since

$$\sum_n \frac{1}{\lambda n^2} < \infty,$$

we may conclude by Proposition 6.2.2 that

$$P\left\{\sum_n E_n < \infty\right\} = 1,$$

and the process is explosive. The stochastic equation for this model is

$$X(t) = X(0) + N\left(\lambda \int_0^t X(s)^2 ds\right),$$

and should be compared with the deterministic ordinary differential equation

$$x'(t) = \lambda x^2(t) \qquad \Longleftrightarrow \qquad x(t) = x(0) + \lambda \int_0^t x(s)^2 ds,$$

which also explodes in finite time. □

Example 6.2.4. Consider a continuous time Markov chain with state space {-2,-1,0,1,2,...}. We suppose that the graph of the model is

$$-2 \underset{1}{\overset{1}{\rightleftarrows}} -1 \overset{2}{\leftarrow} 0 \overset{1}{\rightarrow} 1 \overset{1}{\rightarrow} 2 \overset{2^2}{\rightarrow} 3 \overset{3^2}{\rightarrow} \cdots,$$

where, in general, the intensity of $n \to n+1$, for $n \geq 1$ is $\lambda(n) = n^2$. From the previous example, we know this process is explosive. However, if $X(0) \in \{-2, -1\}$, then the probability of explosion is zero[2], whereas if $X(0) = 0$, the probability of explosion is 1/3. □

The following proposition characterizes the most common ways in which a process is non-explosive. A full proof can be found in [13].

[2]This is proven by the next proposition, but it should be clear

Proposition 6.2.5. *For any $i \in S$,*

$$P_i\{T_\infty < \infty\} = P_i\left\{\sum_n \frac{1}{\lambda(X_n)} < \infty\right\},$$

and therefore, the continuous time Markov chain is non-explosive iff

$$\sum_n \frac{1}{\lambda(X_n)} = \infty,$$

P_i- almost surely for every $i \in S$. In particular,

(1) If $\lambda(i) \leq c$ for all $i \in S$ for some $c > 0$, then the chain is non-explosive.

(2) If S is a finite set, then the chain is non-explosive.

(3) If $T \subset S$ are the transient states of $\{X_n\}$ and if

$$P_i\{X_n \in T, \forall n\} = 0,$$

for every $i \in S$, then the chain is non-explosive.

Proof. The equivalence of the probabilities is shown in [13, Section 5.2]. Will prove the results 1,2,3. For (1), simply note that

$$\sum_n \frac{1}{\lambda(X(n))} \geq \sum_n \frac{1}{c} = \infty.$$

To show (2), we note that if the state space is finite, we may simply take $c = \max\{\lambda_i\}$, and apply (1).

We will now show (3). If $P_i\{X_n \in T, \forall n\} = 0$, then entry into T^c is assured. There must, therefore, be a state $i \in T^c$, which is hit infinitely often (note that this value can be different for different realizations of the process). Let the infinite sequence of times when $X_n = i$ be denoted by $\{n_j\}$. Then,

$$\sum_n 1/\lambda(X_n) \geq \sum_j 1/\lambda(X_{n_j}) = \sum_j 1/\lambda(i) = \infty.$$

□

We will henceforth have a running assumption that unless otherwise explicitly stated, all processes consider are non-explosive. However, we will return to explosiveness later and prove another useful condition that implies a process is non-explosive. This condition will essentially be a linearity condition on the intensities. This condition is sufficient to prove the non-explosiveness of most processes in the queueing literature. Unfortunately, the wold of biology is not so easy and most processes of interest are highly non-linear and it is, in general, quite a difficult (and open) problem to characterize which systems are non-explosive.

6.3 Forward Equation, Backward Equation, and the Generator Matrix

We note that in each of the constructions of a continuous time Markov chain, we are given only the local behavior of the model. Similarly to when we studied the Poisson process, the question now becomes: how do these local behaviors determine the global behavior? In particular, how can we find terms of the form

$$P_{ij}(t) = P\{X(t) = j \mid X(0) = i\},$$

for $i, j \in S$, the state space, and $t \geq 0$?

We begin to answer this question by first deriving the Kolmogorov forward equations, which are a system of ordinary differential equations governing the behaviors of the probabilities $P_{ij}(t)$. We note that the forward equations are only valid if the process is non-explosive as we will derive them by conditioning on the state of the system "directly before" our time of interest. If that time is $T_\infty < \infty$, then this question does not really make sense, for what is the last jump before T_∞?

Proceeding, we have

$$\begin{aligned}
P_{ij}'(t) &= \lim_{h \to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} \\
&= \lim_{h \to 0} \frac{1}{h} \left(P\{X(t+h) = j \mid X(0) = i\} - P\{X(t) = j \mid X(0) = i\} \right) \\
&= \lim_{h \to 0} \frac{1}{h} \Bigg(\sum_{y \in S} P\{X(t+h) = j \mid X(t) = y, X(0) = i\} P\{X(t) = y \mid X(0) = i\} \\
&\qquad - P\{X(t) = j \mid X(0) = i\} \Bigg).
\end{aligned}$$

However,

$$\begin{aligned}
&\sum_{y \in S} P\{X(t+h) = j \mid X(t) = y, X(0) = i\} P\{X(t) = y \mid X(0) = i\} \\
&\quad = P\{X(t+h) = j \mid X(t) = j, X(0) = i\} P\{X(t) = j \mid X(0) = i\} \\
&\qquad + \sum_{y \neq j} P\{X(t+h) = j \mid X(t) = y, X(0) = i\} P\{X(t) = y \mid X(0) = i\} \qquad (6.6) \\
&\quad = (1 - \lambda(j)h) P_{ij}(t) + \sum_{y \neq j} \lambda(y, j) h P_{iy}(t) + o(h), \qquad (6.7)
\end{aligned}$$

and so

$$\begin{aligned}
P_{ij}'(t) &= \lim_{h \to 0} \frac{1}{h} \left((1 - \lambda(j)h - 1) P_{ij}(t) + \sum_{y \neq j} \lambda(y, j) P_{iy}(t) h + o(h) \right) \\
&= -\lambda(j) P_{ij}(t) + \sum_{y \neq j} \lambda(y, j) P_{iy}(t).
\end{aligned}$$

Thus,

$$P'_{ij}(t) = -\lambda(j)P_{ij}(t) + \sum_{y \neq j} P_{iy}(t)\lambda(y, j). \tag{6.8}$$

These are the *Kolmogorov forward equations* for the process. In the biology literature this system of equations is termed the *chemical master equation*.

We point out that there was a small mathematical "slight of hand" in the above calculation. To move from (6.6) to (6.7), we had to assume that

$$\sum_y P_{iy}(t)o_y(h) = o(h),$$

where we write $o_y(h)$ to show that the size of the error can depend upon the state y. This condition is satisfied for all systems we will consider.

Definition 6.3.1. Let $X(t)$ be a continuous time Markov chain on some state space S with transition intensities $\lambda(i, j) \geq 0$. Recalling that

$$\lambda(i) = \sum_{j \neq i} \lambda(i, j),$$

The matrix

$$A_{ij} = \begin{cases} -\lambda(i), & \text{if } i = j \\ \lambda(i, j), & \text{if } i \neq j \end{cases} = \begin{cases} -\sum_j \lambda(i, j), & \text{if } i = j \\ \lambda(i, j), & \text{if } i \neq j \end{cases}$$

is called the *generator*, or *infinitesimal generator*, or *generator matrix* of the Markov chain.

We see that the Kolmogorov forward equations (6.8) can be written as the matrix differential equation

$$P'(t) = P(t)A,$$

since

$$\begin{aligned}(P(t)A)_{ij} &= \sum_y P_{iy}(t)A_{yj} = P_{ij}A_{jj} + \sum_{y \neq j} P_{iy}A_{yj} \\ &= -\lambda(j)P_{ij}(t) + \sum_{y \neq j} P_{iy}\lambda(y, j).\end{aligned}$$

At least formally, this system can be solved

$$P(t) = P(0)e^{tA} = e^{tA},$$

where e^{tA} is the matrix exponential and we used that $P(0) = I$, the identity matrix. recall that the matrix exponential is defined by

$$e^{At} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{t^n A^n}{n!}.$$

This solution is always valid in the case that the state space is finite.

We make the following observations pertaining to the generator A:

1. The elements on the main diagonal are all strictly negative.
2. The elements off the main diagonal are non-negative.
3. Each row sums to zero.

We also point out that given a state space S, the infinitesimal generator A completely determines the Markov chain as it contains all the local information pertaining to the transitions: $\lambda(i,j)$. Thus, it is sufficient to characterize a chain by simply providing a state space, S, and generator, A.

Example 6.3.2. A molecule transitions between states 0 and 1. The transition rates are $\lambda(0,1) = 3$ and $\lambda(1,0) = 1$. The generator matrix is

$$A = \begin{bmatrix} -3 & 3 \\ 1 & -1 \end{bmatrix}.$$

□

Example 6.3.3. Consider a mathematician wandering between three coffee shops with graphical structure

$$A \underset{\lambda_1}{\overset{\mu_1}{\rightleftarrows}} B \underset{\lambda_2}{\overset{\mu_2}{\rightleftarrows}} C.$$

The infinitesimal generator of this process is

$$A = \begin{bmatrix} -\mu_1 & \mu_1 & 0 \\ \lambda_1 & -(\lambda_1 + \mu_2) & \mu_2 \\ 0 & \lambda_2 & -\lambda_2 \end{bmatrix},$$

and the transition matrix for the embedded Markov chain is

$$P = \begin{bmatrix} 0 & 1 & 0 \\ \lambda_1/(\lambda_1 + \mu_1) & 0 & \mu_2/(\lambda_1 + \mu_1) \\ 0 & 1 & 0 \end{bmatrix}.$$

□

Example 6.3.4. For a unit-rate Poisson process, we have

$$A = \begin{bmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & 0\dots \\ 0 & 0 & -1 & 1 \\ \vdots & \vdots & \ddots & \ddots \end{bmatrix}.$$

□

If we are given an initial condition, α, then $\alpha P(t)$ is the vector with jth element

$$(\alpha P(t))_j = \sum_i \alpha_i P_{ij} = \sum_i P\{X(t) = j \mid X(0) = i\} P\{X(0) = i\} \overset{\text{def}}{=} P_\alpha\{X(t) = j\},$$

giving the probability of being in state j at time t given and initial distribution of α. Thus, we see that if α is given, we have

$$\alpha P(t) = P_\alpha(t) = \alpha e^{tA}. \tag{6.9}$$

Backward equation

Before attempting to solve a system using Kolmogorov's forward equations, we introduce another set of equations, called *Kolmogorov's backward equations*, which are valid for all continuous time Markov chains. The derivation below follows that of [13].

We begin by finding an integral equation satisfied by $P_{ij}(t)$. We will then differentiate it to get the backward equations.

Proposition 6.3.5. *For all $i, j \in S$ and $t \geq 0$, we have*

$$P_{ij}(t) = \delta_{ij} e^{-\lambda(i)t} + \int_0^t \lambda(i) e^{-\lambda(i)s} \sum_{k \neq i} Q_{ik} P_{kj}(t-s) ds,$$

where, as usual,

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

is the Kronecker delta function, and Q is the transition matrix of the embedded discrete time Markov chain.

Proof. Conditioning on the first jump time of the chain, T_1, we have

$$\begin{aligned} &P\{X(t) = j \mid X(0) = i\} \\ &\qquad = P\{X(t) = j, T_1 > t \mid X(0) = i\} + P\{X(t) = j, T_1 \leq t \mid X(0) = i\}. \end{aligned}$$

We handle these terms separately. For the first term on the right hand side of the above equation, a first transition has not been made. Thus, $X(t) = j$ iff $j = i$ and does so with a probability of one. That is,

$$\begin{aligned} P\{X(t) = j, &T_1 > t \mid X(0) = i\} \\ &= P\{X(t) = j \mid T_1 > t, X(0) = i\} P\{T_1 > t \mid X(0) = i\} \\ &= \delta_{ij} P_i\{T_1 > t\} \\ &= \delta_{ij} e^{-\lambda(i)t}. \end{aligned}$$

For the second term, we will condition on the time of the first jump happening in $(s, s + \Delta)$, for small Δ (we will eventually take $\Delta \to 0$). As the holding time is exponential with parameter $\lambda(i)$, this event has probability

$$\int_s^{s+\Delta} \lambda(i) e^{-\lambda(i)r} dr = \lambda(i) e^{-\lambda(i)s} \Delta + O(\Delta^2).$$

We let $s_n = nt/N$ for some large N, denote $\Delta = t/N$, and see

$$
\begin{aligned}
P\{X(t) = j, T_1 \le t \mid X(0) = i\} &= \sum_{n=0}^{N-1} P\{X(t) = j, T_1 \in (s_n, s_{n+1}) \mid X(0) = i\} \\
&= \sum_{n=0}^{N-1} P\{X(t) = j \mid X(0) = i, T_1 \in (s_n, s_{n+1})\} P\{T_1 \in (s_n, s_{n+1}) \mid X(0) = i\} \\
&= \sum_{n=0}^{N-1} P\{X(t) = j \mid X(0) = i, T_1 \in (s_n, s_{n+1})\} \left[\lambda(i) e^{\lambda(i) s_n} \Delta + O(\Delta^2)\right] \\
&= \sum_{n=0}^{N-1} \lambda(i) e^{\lambda(i) s_n} \sum_{k \ne i} P\{X(t) = j, X_1 = k \mid X(0) = i, T_1 \in (s_n, s_{n+1})\} \Delta + O(\Delta) \\
&= \sum_{n=0}^{N-1} \lambda(i) e^{\lambda(i) s_n} \sum_{k \ne i} \Big[P\{X(t) = j \mid X_1 = k, X(0) = i, T_1 \in (s_n, s_{n+1})\} \\
&\qquad \times P\{X_1 = k \mid X(0) = i, T_1 \in (s_n, s_{n+1})\} \Big] \Delta + O(\Delta) \\
&\approx \sum_{n=0}^{N-1} \lambda(i) e^{\lambda(i) s_n} \sum_{k \ne i} Q_{ik} P_{kj}(t - s_n) \Delta + O(\Delta) \\
&\to \int_0^t \lambda(i) e^{\lambda(i) s} \sum_{k \ne i} Q_{ik} P_{kj}(t - s) ds,
\end{aligned}
$$

as $\Delta \to 0$. Combining the above shows the result. □

Proposition 6.3.6. *For all $i, j \in S$, we have that $P_{ij}(t)$ is continuously differentiable and*

$$P'(t) = AP(t), \tag{6.10}$$

which in component form is

$$P'_{ij}(t) = \sum_k A_{ik} P_{kj}(t).$$

The system of equations (6.10) is called the *Kolmogorov backwards equations.* Note that the difference with the forward equations is the order of the multiplication of $P(t)$ and A. However, the solution of the backwards equation is once again seen to be

$$P(t) = e^{tA},$$

agreeing with previous results.

Proof. Use the substitution $u = t - s$ in the integral equation to find that

$$\begin{aligned}
P_{ij}(t) &= \delta_{ij}e^{-\lambda(i)t} + \int_0^t \lambda(i)e^{-\lambda(i)s}\sum_{k\neq i} Q_{ik}P_{kj}(t-s)ds \\
&= \delta_{ij}e^{-\lambda(i)t} + \int_0^t \lambda(i)e^{-\lambda(i)(t-u)}\sum_{k\neq i} Q_{ik}P_{kj}(u)ds \\
&= e^{-\lambda(i)t}\left[\delta_{ij} + \int_0^t \lambda(i)e^{\lambda(i)u}\sum_{k\neq i} Q_{ik}P_{kj}(u)ds\right].
\end{aligned}$$

Differentiating yields

$$\begin{aligned}
P'_{ij}(t) &= -\lambda(i)e^{-\lambda(i)t}\left[\delta_{ij} + \int_0^t \lambda(i)e^{\lambda(i)u}\sum_{k\neq i} Q_{ik}P_{kj}(u)ds\right] \\
&\quad + e^{-\lambda(i)t}\cdot\lambda(i)e^{\lambda(i)t}\sum_{k\neq i} Q_{ik}P_{kj}(t) \\
&= -\lambda(i)P_{ij}(t) + \lambda(i)\sum_{k\neq i} Q_{ik}P_{kj}(t) \\
&= \sum_k(-\lambda(i)\delta_{ik}P_{kj}(t)) + \sum_k \lambda(i)Q_{ik}P_{kj}(t) \\
&= \sum_k(-\lambda(i)\delta_{ik} + \lambda(i)Q_{ik})P_{kj}(t) \\
&= \sum_k A_{ik}P_{kj}(t).
\end{aligned}$$

□

Both the forward and backward equations can be used to solve for the associated probabilities as the next example demonstrates.

Example 6.3.7. We consider a two state, $\{0, 1\}$, continuous time Markov chain with generator matrix

$$A = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}.$$

We will use both the forwards and backwards equations to solve for $P(t)$.

Approach 1: Backward equation. While we want to compute $P_{ij}(t)$ for each pair $i, j \in \{0, 1\}$, we know that

$$P_{00}(t) + P_{01}(t) = P_{10}(t) + P_{11}(t) = 1,$$

for all $t \geq 0$, and so it is sufficient to solve just for $P_{00}(t)$ and $P_{10}(t)$.

The backwards equation is $P'(t) = AP(t)$, yielding the equations

$$\begin{aligned}
P'_{00}(t) &= \lambda[P_{10}(t) - P_{00}(t)] \\
P'_{10}(t) &= \mu[P_{00}(t) - P_{10}(t)].
\end{aligned}$$

We see that

$$\mu P_{00}'(t) + \lambda P_{10}'(t) = 0 \implies \mu P_{00}(t) + \lambda P_{10}(t) = c.$$

We know that $P(0) = I$, so we see that

$$\mu P_{00}(0) + \lambda P_{10}(0) = c \iff \mu = c.$$

Thus,

$$\mu P_{00}(t) + \lambda P_{10}(t) = \mu \implies \lambda P_{10}(t) = \mu - \mu P_{00}(t).$$

Putting this back into our differential equations above we have that

$$P_{00}'(t) = \mu - \mu P_{00}(t) - \lambda P_{00}(t) = \mu - (\mu + \lambda) P_{00}(t).$$

Solving, with $P_{00}(t) = 1$ yields

$$P_{00}(t) = \frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} e^{-(\mu+\lambda)t}.$$

Of course, we also have that

$$\begin{aligned} P_{01}(t) &= 1 - P_{00}(t) \\ P_{10}(t) &= \frac{\mu}{\lambda} - \frac{\mu}{\lambda}\left(\frac{\mu}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)t}\right) = \frac{\mu}{\mu+\lambda} - \frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)t}. \end{aligned}$$

Approach 1: Forward equation. This is easier. We want to solve

$$P'(t) = P(t)A.$$

We now get

$$\begin{aligned} P_{00}'(t) &= -P_{00}(t)\lambda + P_{01}(t)\mu = -P_{00}(t)\lambda + (1 - P_{00}(t))\mu = \mu - (\lambda + \mu)P_{00}(t) \\ P_{10}'(t) &= -\lambda P_{10}(t) + \mu P_{11}(t) = -\lambda P_{10}(t) + \mu(1 - P_{10}(t)) = \mu - (\lambda + \mu)P_{00}(t), \end{aligned}$$

and the solutions above follow easily.

Note that, as in the discrete time setting, we have that

$$\lim_{t\to\infty} P(t) = \frac{1}{\lambda + \mu}\begin{bmatrix} \mu & \lambda \\ \mu & \lambda \end{bmatrix},$$

yielding a common row vector which can be interpreted as a limiting distribution. □

There is a more straightforward way to make the above computations: simply solve the matrix exponential.

Example 6.3.8 (Computing matrix exponentials)**.** Suppose that A is an $n \times n$ matrix with n distinct eigenvectors. Then, letting D be a diagonal matrix consisting of the eigenvalues of A, we can decompose A into

$$A = QDQ^{-1},$$

where Q consists of the eigenvectors of A (ordered similarly to the order of the eigenvalues in D). In this case, we get the very nice identity

$$e^{At} = \sum_{n=0}^{\infty} \frac{t^n (QDQ^{-1})^n}{n!} = Q \left(\sum_{n=0}^{\infty} \frac{t^n D^n}{n!} \right) Q^{-1} = Qe^{Dt}Q^{-1},$$

where e^{Dt}, because D is diagonal, is a diagonal matrix with diagonal elements $e^{\lambda_i t}$ where λ_i is the ith eigenvalue.

Example 6.3.9. We now solve the above problem using the matrix exponential. Supposing, for concreteness, that $\lambda = 3$ and $\mu = 1$, we have that the generator matrix is

$$A = \begin{bmatrix} -3 & 3 \\ 1 & -1 \end{bmatrix}$$

It is easy to check that the eigenvalues are $0, -4$ and the associated eigenvalues are $[1, 1]^t$ and $[-3, 1]^t$. Therefore,

$$Q = \begin{bmatrix} 1 & -3 \\ 1 & 1 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 1/4 & 3/4 \\ -1/4 & 1/4 \end{bmatrix},$$

and

$$e^{tA} = \begin{bmatrix} 1/4 + (3/4)e^{-4t} & 3/4 - (3/4)e^{-4t} \\ 1/4 - (1/4)e^{-4t} & 3/4 + (1/4)e^{-4t} \end{bmatrix}.$$

You should note that

$$\lim_{t \to \infty} e^{tA} = \begin{bmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{bmatrix},$$

which has a common row. Thus, for example, in the long run, the chain will be in state zero with a probability of 1/4. □

6.4 Stationary Distributions

In this section we will parallel our treatment of stationary distributions for discrete time Markov chains. We will aim for intuition, as opposed to attempting to prove everything, and point the interested reader to [13] and [11] for the full details of the proofs.

6.4.1 Classification of states

We start by again classifying the states of our process. Viewing a continuous time Markov chain as an embedded discrete time Markov chain with exponential holding times makes the classification of states, analogous to Section 3.4 in the discrete time setting, easy. We will again denote our state space as S.

Definition 6.4.1. The communication classes of the continuous time Markov chain $X(t)$ are the communication classes of the embedded Markov chain X_n. If there is only one communication class, we say the chain is *irreducible*; otherwise it is said to be *reducible*.

Noting that $X(t)$ will return to a state i infinitely often if and only if the embedded discrete time chain does (even in the case of an explosion!) motivates the following.

Definition 6.4.2. State $i \in S$ is called *recurrent* for $X(t)$ if i is recurrent for the embedded discrete time chain X_n. Otherwise, i is *transient*.

Definition 6.4.3. Let T_1 denote the first jump time of the continuous time chain. We define

$$\tau_i \stackrel{\text{def}}{=} \inf\{t \geq T_1 \ : \ X(t) = i\},$$

and set $m_i = \mathbb{E}_i \tau_i$. We say that state i is *positive recurrent* if $m_i < \infty$.

Note that, perhaps surprisingly, we *do not* define i to be positive recurrent if i is positive recurrent for the discrete time chain. In Example 6.4.10 we will demonstrate that i may be positive recurrent for X_n, while not for $X(t)$.

As in the discrete time setting, recurrence, transience, and positive recurrence are class properties.

Note that the concept of periodicity no longer plays a role, or even makes sense to define, as time is no longer discrete. In fact, if $P(t)$ is the matrix with entries $P_{ij}(t) = P\{X(t) = j \mid X(0) = i\}$ for an irreducible continuous time chain, then for every $t > 0$, $P_{ij}(t)$ has strictly positive entries because there is necessarily a path between i and j, and a non-zero probability of moving along that path in time $t > 0$.

6.4.2 Invariant measures

Recall that equation (6.9) states that if the initial distribution of the process is α, then $\alpha P(t)$ is the vector whose ith component gives the probability that $X(t) = i$. We therefore define an invariant measure in the following manner.

Definition 6.4.4. A measure $\eta = \{\eta_j, j \in S\}$ on S is called *invariant* if for all $t > 0$

$$\eta P(t) = \eta.$$

If this measure is a probability distribution (i.e. sums to one), then it is called a *stationary distribution*.

Note, therefore, that if the initial distribution is η, then $P_\eta\{X(t) = i\} = \eta_i$, for all $t \geq 0$, demonstrating why such a measure is called *invariant*.

The following theorem gives us a nice way to find stationary distributions of continuous time Markov chains.

Theorem 6.4.5. *Let $X(t)$ be an irreducible and recurrent continuous time Markov chain. Then the following statements are equivalent:*

1. *$\eta A = 0$;*
2. *$\eta P(t) = \eta$, for all $t \geq 0$.*

Proof. The proof of this fact is easy in the case of a finite state space, which is what we will assume here. Recall Kolmogorov's backward equation

$$P'(t) = AP(t).$$

Assume that $\eta A = 0$. Multiplying the backwards equation on the left by η shows

$$0 = \eta AP(t) = \eta P'(t) = \frac{d}{dt}\eta P(t).$$

Thus,

$$\eta P(t) = \eta P(0) = \eta,$$

for all $t \geq 0$.

Now assume that $\eta P(t) = \eta$ for al $t \geq 0$. Then, for all $h > 0$, we have

$$\eta P(h) = \eta \quad \implies \quad \eta(P(h) - I) = 0 \quad \implies \quad \frac{\eta}{h}(P(h) - I) = 0.$$

Taking $h \to 0$ now shows that

$$0 = \eta P'(0) = \eta A,$$

where we have used that $P'(0) = A$, which follows from either the forward or backward equations.

The interchange above of differentiation with summation can not in general be justified in the infinite dimensional setting, and different proof is needed and we refer the reader to [11, Section 3.5]. $\square$

Theorem 6.4.6. *Suppose that $X(t)$ is irreducible and recurrent. Then $X(t)$ has an invariant measure η, which is unique up to multiplicative factors. Moreover, for each $k \in S$, we have*

$$\eta_k = \pi_k/\lambda(k),$$

where π is the unique invariant measure of the embedded discrete time Markov chain X_n. Finally, η satisfies

$$0 < \eta_j < \infty, \quad \forall j \in S,$$

and if $\sum_i \eta_i < \infty$ then η can normalize by $1/\sum_k \eta_k$ to give a stationary distribution.

Proof. By Theorem 6.4.5, we must only show that there is a solution to $\eta A = 0$, satisfying all the desired results, if and only if there is an invariant measure to the discrete time chain. We first recall that π was an invariant measure for a discrete time Markov chain if and only if $\pi Q = \pi$, where Q is the transition matrix. By Theorem 3.5.16, such a π exists, and is unique up to multiplicative constants, if X_n is irreducible and recurrent.

Recall that if $j \neq k$, then $A_{jk} = \lambda(j)Q_{jk}$ and that $A_{jj} = -\lambda(j)$. We now simply note that

$$\eta' A = 0 \iff \sum_j \eta_j A_{jk} = 0, \quad \forall k \iff \sum_{j \neq k} \eta_j \lambda(j) Q_{jk} - \eta_k \lambda(k) = 0.$$

However, this holds if and only if

$$\sum_{j \neq k} \eta_j \lambda(j) Q_{jk} = \eta_k \lambda(k) \iff \pi Q = \pi, \quad \text{where } \pi_k \stackrel{\text{def}}{=} \lambda(k)\eta_k.$$

That is, the final equation (and hence all the others) holds if and only if π is invariant for the Markov matrix Q. Such a π exists, and satisfies all the desired properties, by Theorem 3.5.16. Further, we see the invariant measure of the continuous time Process satisfies $\eta_k = \pi_k / \lambda(k)$, as desired. □

Example 6.4.7. Consider the continuous time Markov chain with generator matrix

$$A = \begin{bmatrix} -5 & 3 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 2 & 1 & -4 & 1 \\ 0 & 2 & 2 & -4 \end{bmatrix}.$$

The unique left eigenvector of A with eigenvalue 0, i.e. the solution to $\eta A = 0$, normalized to sum to one is

$$\eta = \left[\frac{14}{83}, \frac{58}{83}, \frac{6}{83}, \frac{5}{83}\right].$$

Further, note that the transition matrix for the embedded discrete time Markov chain is

$$P = \begin{bmatrix} 0 & 3/5 & 1/5 & 1/5 \\ 1 & 0 & 0 & 0 \\ 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

Solving for the stationary distribution of the embedded chain, i.e. solving $\pi P = \pi$, yields

$$\pi = \left[\frac{35}{86}, \frac{29}{86}, \frac{6}{43}, \frac{5}{43}\right].$$

Finally, note that

$$
\begin{aligned}
[\eta_1\lambda(1), \eta_2\lambda(2), \eta(3)\lambda(3), \eta(4)\lambda(4)] &= \left[5\cdot\frac{14}{83}, \frac{58}{83}, 4\cdot\frac{6}{83}, 4\cdot\frac{5}{83}\right] \\
&= \left[\frac{70}{83}, \frac{58}{83}, \frac{24}{83}, \frac{20}{83}\right] \\
&= \frac{172}{83}\left[\frac{35}{86}, \frac{29}{86}, \frac{6}{43}, \frac{5}{43}\right] \\
&= \frac{172}{83}\pi,
\end{aligned}
$$

as predicted by the theory. □

We now consider the positive recurrent case. We recall that $m_i = \mathbb{E}_i\tau_i$, the expected first return time to state i. The following result should not be surprising at this point. See [11] for a proof.

Theorem 6.4.8. *Let A be the generator matrix for an irreducible continuous time Markov chain. Then the following are equivalent*

1. *Every state is positive recurrent.*
2. *Some state is positive recurrent.*
3. *A is non-explosive and has an invariant distribution η.*

Definition 6.4.9. We call the non-explosive continuous time Markov chain $\{X(t)\}$ *ergodic* if $\{X_n\}$ is recurrent and irreducible and a stationary distribution exists.

Note, therefore, that $X(t)$ is ergodic if and only if the chain is irreducible and positive recurrent.

The following example shows that positive recurrence of X_n does not guarantee existence of stationary distribution for $X(t)$. That is, $X(t)$ may not be positive recurrent.

Example 6.4.10. We consider a continuous time Markov chain whose embedded discrete time Markov chain has state space $S = \{0, 1, 2, \dots\}$ and transition matrix

$$
Q = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ q & 0 & p & 0 & \cdots \\ q & 0 & 0 & p & \\ \vdots & & \ddots & & \end{pmatrix},
$$

where $p + q = 1$. This is the "success run chain" and we showed in Problem 2.11 that the discrete time chain is positive recurrent. Let $\lambda(i)$ be the holding time parameter for state i of the associated continuous time Markov chain, and let E_m, $m \geq 0$, denote a sequence of independent unit exponential random variables, which are also independent of the embedded discrete time Markov chain. Finally, assuming that

$X_0 = 0$, let T_1 denote the first return time to state 0 of the *embedded chain*. For example, if $T_1 = 3$, then $X_0 = 0, X_1 = 1, X_2 = 2$, and $X_3 = 0$. More generally, we have $X_0 = 0, X_1 = 1, \dots, X_{T_1-1} = T_1 - 1$, and $X_T = 0$. For $m < T_1$, we let $W(m) = E_m/\lambda(m)$ be the holding time in state m. We have

$$\begin{aligned} m_0 = \mathbb{E}_0 \tau_0 = \mathbb{E}_0 \sum_{m=0}^{T_1-1} W(m) \\ = \mathbb{E} \sum_{m=0}^{\infty} W(m) 1_{\{m<T_1\}} \\ = \sum_{m=0}^{\infty} \mathbb{E}[W(m) 1_{\{m<T_1\}}]. \end{aligned}$$

However, we know that the holding times and the embedded chain are independent. Thus, as $1_{\{m<T_1\}}$ is simply a statement pertaining to the embedded chain,

$$\mathbb{E}[W(m)1_{\{m<T_1\}}] = [\mathbb{E}W(m)][\mathbb{E}1_{\{m<T_1\}}] = \frac{1}{\lambda(m)} P_0\{m < T_1\}.$$

Combining the above,

$$\begin{aligned} m_0 &= \sum_{m=0}^{\infty} \frac{1}{\lambda(m)} P_0\{m < T_1\} \\ &= \frac{1}{\lambda(0)} + \sum_{m=1}^{\infty} \frac{1}{\lambda(m)} P_0\{m < T_1\}. \end{aligned}$$

For $m \geq 1$,

$$P\{m < T_1\} = \sum_{n=m+1}^{\infty} P\{T_1 = n\} = \sum_{n=m+1}^{\infty} p^{n-2} q = q p^{m-1} \sum_{n=0}^{\infty} p^n = p^{m-1}.$$

Thus,

$$m_0 = \frac{1}{\lambda(0)} + \sum_{m=1}^{\infty} \frac{1}{\lambda(m)} p^{m-1}.$$

Of course, we have not chosen $\lambda(m)$ yet. Taking $\lambda(m) = p^m$, we see

$$m_0 = \frac{1}{\lambda(0)} + \sum_{m=1}^{\infty} \frac{1}{p^m} p^{m-1} = 1 + \sum_{m=1}^{\infty} \frac{1}{p} = \infty.$$

So, $\{X_n\}$ is positive recurrent, but $X(t)$ is not. □

The following example, taken from [11], shows two things. First, it demonstrates that a transient chain *can* have an invariant measure. Further, it even shows stranger behavior is possible: a transient chain can have an *invariant distribution*! Of course, the previous theorems seem to suggest that this is not possible. However, there is a catch: the chain could be explosive. In fact, if a transient chain is shown to have a stationary distribution, then the chain must be explosive for otherwise Theorem 6.4.8 is violated.

Example 6.4.11. Consider a discrete time random walker on $S = \{0, 1, 2, \dots\}$. Suppose that the probability of moving to the right is $p > 0$ and to the left is $q = 1 - p$. To convert this into a continuous time chain, we suppose that $\lambda(i)$ is the holding time parameter in state i. More specifically, we assume $X(t)$ is a continuous time Markov chain with generator matrix A satisfying

$$A = \begin{pmatrix} -\lambda(0)p & \lambda(0)p & 0 & 0 & 0 & \cdots \\ q\lambda(1) & -\lambda(1) & p\lambda(1) & 0 & 0 & \cdots \\ 0 & q\lambda(2) & -\lambda(2) & p\lambda(2) & 0 & \cdots \\ 0 & 0 & q\lambda(3) & -\lambda(3) & p\lambda(3) & \\ \vdots & \ddots & & \ddots & & \end{pmatrix}$$

We know that this chain is transient if $p > q$ since the discrete time chain is. We now search for an invariant measure satisfying

$$\eta A = 0,$$

which in component form is

$$\begin{aligned} -\lambda(0)p\eta_0 + q\lambda(1)\eta_1 &= 0 \\ \lambda(i-1)p\eta_{i-1} - \lambda(i)\eta_i + \lambda(i+1)q\eta_{i+1} &= 0 \qquad i > 0. \end{aligned}$$

We will confirm that η satisfying

$$\eta(i) = \frac{1}{\lambda(i)} \left(\frac{p}{q}\right)^i,$$

is a solution. The case $i = 0$ is easy to verify

$$\lambda(0)p\eta_0 = \lambda(0)p\frac{1}{\lambda(0)} = p = q\lambda(1)\frac{1}{\lambda(1)}\frac{p}{q} = q\lambda(1)\eta_1.$$

The $i > 0$ case follows similarly.

Therefore, there is always an invariant distribution, regardless of the values p and q. Taking $p > q$ and $\lambda(i) = 1$ for all i, we see that the resulting continuous time Markov chain is transient, and has an invariant distribution

$$\eta(i) = \left(\frac{p}{q}\right)^i,$$

which can not be normalized to provide an invariant *distribution*.

Now, consider the case when $p > q$, with $1 < p/q < 2$, and take $\lambda(i) = 2^i$. Define $\alpha \stackrel{\text{def}}{=} p/q < 2$. Then,

$$\sum_{i=0}^{\infty} \eta(i) = \sum_{i=0}^{\infty} \left(\frac{\alpha}{2}\right)^i = \frac{1}{1-\alpha/2} = \frac{2}{2-\alpha} < \infty,$$

Therefore, we can normalize to get a stationary distribution. Since we already know this chain is transient, we have shown that it must, in fact, explode. □

6.4.3 Limiting distributions and convergence

We have found conditions for the existence of a unique stationary distribution to a continuous time Markov chain: irreducibility and positive recurrence (i.e. *ergodicity*). As in the discrete time case, there is still the question of convergence. The following is proven in [11].

Theorem 6.4.12. *Let $X(t)$ be an ergodic continuous time Markov chain with unique invariant distribution η. Then, for all $i, j \in S$,*

$$\lim_{t \to \infty} P_{ij}(t) = \eta_j.$$

Example 6.4.13. Let $S = \{0, 1\}$ with transition rates $\lambda(0, 1) = 3$ and $\lambda(1, 0) = 1$. Then the generator matrix is

$$A = \begin{bmatrix} -3 & 3 \\ 1 & -1 \end{bmatrix}.$$

Solving directly for the left eigenvector of A with eigenvalue 0 yields

$$\pi = [1/4, 3/4],$$

which agrees with the result found in Example 6.3.9. □

As in the discrete time setting, we have an ergodic theorem, which we simply state. For a proof, see [13, Section 5.5].

Theorem 6.4.14. *Let $X(t)$ be an irreducible, positive recurrent continuous time Markov chain with unique stationary distribution η. Then, for any initial condition, and any $i \in S$,*

$$P\left(\frac{1}{t}\int_0^t 1_{\{X(s)=i\}} ds \to \eta_i, \quad \text{as } t \to \infty\right) = 1.$$

Moreover, for any bounded function $f : S \to \mathbb{R}$ we have

$$P\left(\frac{1}{t}\int_0^t f(X(s)) ds \to \bar{f}, \quad \text{as } t \to \infty\right) = 1,$$

where

$$\bar{f} = \sum_{j \in S} \eta_j f(j) = \mathbb{E}_\eta f(X_\infty),$$

where X_∞ has distribution η.

Thus, as in the discrete time setting, we see that η_i gives the proportion of time spent in state i over long periods of time. This gives us an algorithmic way to sample from the stationary distribution: simulate a single long trajectory and average over it.

6.5 The Generator, Revisited

Consider a function $f : S \to \mathbb{R}$. Noting that f is simply a mapping from S to $\mathbb{R}$, and that S is discrete, we can view f as a column vector whose ith component is equal to $f(i)$. For example, if $S = \{1, 2, 3\}$ and $f(1) = -2, f(2) = \pi$, and $f(3) = 100$, then we take

$$f = \begin{bmatrix} -2 \\ \pi \\ 100 \end{bmatrix}.$$

As A is a matrix, it therefore makes sense to discuss the well defined object Af, which is itself a column vector, and hence a function from S to $\mathbb{R}$.

Next, we note that if the initial distribution for our Markov chain is α, then for any f we have that

$$\begin{aligned} \mathbb{E}_\alpha f(X(t)) &= \sum_{j \in S} P_\alpha\{X(t) = j\} f(j) \\ &= \sum_{j \in S} \left(\sum_{i \in S} P\{X(t) = j \mid X(0) = i\} P\{X(0) = i\} \right) f(j) \\ &= \sum_{i \in S} \alpha_i \left(\sum_{j \in S} P_{ij}(t) f(j) \right) \\ &= \sum_{i \in S} \alpha_i (P(t)f)_i \\ &= \alpha P(t) f. \end{aligned} \tag{6.11}$$

Now recall that the forward equation stated that $P'(t) = P(t)A$. Integrating this equation yields

$$P(t) = I + \int_0^t P(s)A ds,$$

and multiplication on the right by f gives

$$P(t)f = f + \int_0^t P(s)Af ds. \tag{6.12}$$

Multiplying (6.12) on the left by α yields

$$\alpha P(t)f = \alpha f + \int_0^t \alpha P(s)(Af) ds,$$

which combined with (6.11) gives

$$\begin{aligned}\mathbb{E}_\alpha f(X(t)) &= \mathbb{E}_\alpha f(X(0)) + \int_0^t \mathbb{E}_\alpha\left(Af\right)\left(X(s)\right) ds \\ &= \mathbb{E}_\alpha f(X(0)) + \mathbb{E}_\alpha \int_0^t \left(Af\right)\left(X(s)\right) ds.\end{aligned}$$

This is a version of *Dynkin's formula*. For a more formal derivation in the Markov process setting, see [4, Section 1.1]. In the next section, we will use this formulation to calculate the mean and variance of a linear birth and death model.

Example 6.5.1. We will re-derive the mean and variance of a Poisson process using Dynkin's formula. Let $X(t)$ be a Poisson process with intensity $\lambda > 0$ defined on $S = \{0, 1, 2, \dots\}$. Then, for any function $f : S \to \mathbb{R}$

$$(Af) = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \ddots & & \ddots & \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ \vdots \end{bmatrix} = \begin{bmatrix} -\lambda f(0) + \lambda f(1) \\ -\lambda f(1) + \lambda f(2) \\ -\lambda f(2) + \lambda f(3) \\ \vdots \end{bmatrix},$$

and so, for any $i \ge 0$,

$$(Af)(i) = \lambda(f(i+1) - f(i)).$$

Letting $f(i) = i$, and taking $X(0) = 0$ with a probability of one, we therefore see that

$$\begin{aligned}\mathbb{E}f(X(t)) = \mathbb{E}X(t) &= 0 + \int_0^t \mathbb{E}(Af)(X(s))ds \\ &= \int_0^t \mathbb{E}\lambda\big(f(X(s)+1) - f(X(s))\big)ds \\ &= \lambda \int_0^t ds \\ &= \lambda t.\end{aligned}$$

Next, letting $g(i) = i^2$ (so as to find the second moment), we have

$$\begin{aligned}\mathbb{E}g(X(t)) = \mathbb{E}X(t)^2 &= 0 + \int_0^t \mathbb{E}(Af)(X(s))ds \\ &= \int_0^t \mathbb{E}\lambda\big(g(X(s)+1) - g(X(s))\big)ds \\ &= \lambda \int_0^t \mathbb{E}\big(X(s)^2 + 2X(s) + 1 - X(s)^2\big)ds \\ &= \lambda \int_0^t \mathbb{E}\big(2X(s) + 1\big)ds \\ &= \lambda \int_0^t (2\lambda s + 1)ds \\ &= \lambda^2 t^2 + \lambda t.\end{aligned}$$

Therefore, the variance is

$$\mathsf{Var}(X(t)) = \mathbb{E}X(t)^2 - (\mathbb{E}X(t))^2 = \lambda t.$$

$\square$

Of course, both the mean and variance of a Poisson process are well known. However, the above method is quite general and is useful in a myriad of applications.

Example 6.5.2. Consider a pure birth process with growth rate $\lambda(i) = bi$ for some $b > 0$. That is, the embedded chain is the deterministic monotone chain and the holding time in state i is bi. For an arbitrary function f, we have that

$$(Af)(i) = bi\big(f(i+1) - f(i)\big), \tag{6.13}$$

for all $i \geq 0$, where A is the generator for the continuous time chain. Assuming $X(0) = 1$, we will derive the mean of the process.

For $f(i) = i$, we have that

$$\begin{aligned}\mathbb{E}f(X(t)) = \mathbb{E}X(t) &= 1 + \int_0^t \mathbb{E}(Af)(X(s))ds \\ &= 1 + \int_0^t \mathbb{E}bX(s)\big(f(X(s)+1) - f(X(s))\big)ds \\ &= 1 + b\int_0^t \mathbb{E}X(s)ds.\end{aligned}$$

Therefore, defining $g(t) = \mathbb{E}X(t)$, we see that

$$g'(t) = bg(t), \quad g(0) = 1.$$

Thus,

$$g(t) = \mathbb{E}X(t) = e^{bt}.$$

This result should be compared with the solution to the ODE linear growth model $x'(t) = bx(t)$, which yields a similar solution. You will derive the variance for a homework exercise. $\square$

Now consider the (row) vector e_i, with a one in the ith component, and zeros everywhere else. Taking e_i as an initial distribution, we see from (6.11) that for all $t \geq 0$

$$e_i P(t) f = \mathbb{E}_i f(X(t)).$$

In words, the ith component of the vector $P(t)f$ gives $\mathbb{E}_i f(X(t))$. Next, note that

$$\begin{aligned}(Af)(i) = e_i(Af) = e_i(P'(0)f) &= e_i \lim_{h\to 0} \frac{1}{h}\left(P(h)f - P(0)f\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(e_i P(h) f - e_i f\right) \\ &= \lim_{h\to 0} \frac{\mathbb{E}_i f(X(h)) - f(i)}{h}.\end{aligned}$$

Further, taking $f(i) = 1_{\{i=j\}}$ for some j, we see that

$$(Af)(i) = \lim_{h\to 0} \frac{\mathbb{E}_i f(X(h)) - f(i)}{h}, \tag{6.14}$$

gives

$$A_{ij} = \lim_{h\to 0} \frac{1}{h}(P\{X(h) = j \mid X(0) = i\}) = \lambda(i,j),$$

when $i \neq j$, and

$$A_{jj} = \lim_{h\to 0} \frac{1}{h}(P\{X(h) = j \mid X(0) = j\} - 1) = -\lambda(j),$$

for the diagonal elements. Therefore, (6.14) could be taken as an alternative *definition* of the generator for a Markov process, though one which views the generator as an operator and not simply as a matrix that stores the transition intensities. In fact, in many ways this definition is *much* more useful than that of simply the matrix with transition rates.

Example 6.5.3. Consider a process with arrivals coming in at rate $\lambda > 0$ and departures taking place at rate $\mu X(t)$, where $X(t)$ is the number of items at time t. Then, for $i \geq 0$ we have

$$\begin{aligned}
(Af)(i) &= \lim_{h\to 0} \frac{\mathbb{E}_i f(X(h)) - f(i)}{h} \\
&= \lim_{h\to 0} \frac{1}{h}\Big[f(i+1)P_i\{X(h) = i+1\} + f(i-1)P_i\{X(h) = i-1\} \\
&\qquad\qquad + f(i)P_i\{X(h) = i\} - f(i) + o(h)\Big] \\
&= \lim_{h\to 0} \frac{1}{h}\Big[f(i+1)\lambda h + f(i-1)\mu i h + f(i)(1 - \lambda h - \mu i h) - f(i) + o(h)\Big] \\
&= \lambda(f(i+1) - f(i)) + \mu i (f(i-1) - f(i)).
\end{aligned}$$

So, for example, taking $f(y) = y$ to be the identity, and $X(0) = x$, we have that

$$\begin{aligned}
\mathbb{E} f(X(t)) = \mathbb{E} X(t) &= \mathbb{E} X(0) + \mathbb{E} \int_0^t (Af)(X(s))ds \\
&= x + \mathbb{E} \int_0^t \big(\lambda(X(s) + 1 - X(s))\big) + \mu X(s)\big(X(s) - 1 - X(s)\big)ds \\
&= x + \int_0^t (\lambda - \mu \mathbb{E} X(s))ds.
\end{aligned}$$

Setting $g(t) = \mathbb{E} X(t)$, we see that $g(0) = x$ and $g'(t) = \lambda - \mu g(t)$. Solving this initial value problem yields the solution

$$\mathbb{E} X(t) = x e^{-\mu t} + \frac{\lambda}{\mu}(1 - e^{-\mu t}).$$

The second moment, and hence the variance, of the process can be calculated in a similar manner. □

6.6 Continuous Time Birth and Death Processes

We again consider a Markovian birth and death process, though now in the continuous time setting. As in Section 4.2, our state space is $S = \{0, 1, 2, \dots\}$. The transition rates are

$$\begin{aligned} \lambda(n, n+1) &= b_n \\ \lambda(n, n-1) &= d_n \\ \lambda(n, j) &= 0, \qquad \text{if } |j - n| \geq 2, \end{aligned}$$

for some values $b_n, d_n \geq 0$, and $d_0 = 0$, yielding a tridiagonal generator matrix

$$A = \begin{bmatrix} -b_0 & b_1 & 0 & 0 & 0 & \cdots \\ d_1 & -(b_1 + d_1) & d_1 & 0 & 0 & \cdots \\ 0 & d_2 & -(b_2 + d_2) & b_2 & 0 & \cdots \\ 0 & 0 & d_3 & -(b_3 + d_3) & b_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \end{bmatrix}.$$

We begin with examples, many of which are analogous to those in the discrete time setting.

Example 6.6.1. The Poisson process is a birth-death process with $b_n \equiv \lambda$, for some $\lambda > 0$, and $d_n \equiv 0$. □

Example 6.6.2. A pure birth process with $b_n \geq 0$, and $d_n \equiv 0$ is an example of a birth and death process. □

Example 6.6.3 (Queueing Models)**.** We suppose that arrivals of customers are occurring at a constant rate of $\lambda > 0$. That is, we assume that $b_n \equiv \lambda$. However, departures occur when a customer has been served. There are a number of natural choices for the model of the service times.

(a) (Single server) If there is a single server, and that person always serves the first person in line, then we take $d_n = \mu > 0$, if $n \geq 1$, and $d_0 = 0$ (as there is no one to serve).

(b) (k servers) If there are $k \geq 1$ servers, and the first k people in line are always being served, then for some $\mu > 0$ we take

$$d_n = \begin{cases} n\mu, & \text{if } n \leq k \\ k\mu, & \text{if } n \geq k \end{cases}.$$

(c) (∞ servers) If we suppose that there are an infinite number of servers, then $d_n = n\mu$ for some $\mu > 0$.

□

Example 6.6.4 (Population Models)**.** Suppose that $X(t)$ represents the number of individuals in a certain population at time t. Assuming the rates of both reproduction and death are proportional to population size we have

$$b_n = \lambda n$$
$$d_n = \mu n,$$

for some $\lambda, \mu > 0$. □

Example 6.6.5 (Population with immigration)**.** Consider the previous system except $b_n = \lambda n + \nu$ for some $\nu > 0$, representing an inflow due to immigration. Now 0 is no longer an absorbing state. □

Example 6.6.6 (Fast growing population)**.** Consider a population that grows at a rate equal to the square of the number of individuals. Assuming no deaths, we have for some $\lambda > 0$ that

$$b_n = \lambda n^2, \quad \text{and} \quad \mu_n = 0.$$

We have seen that this population grows so fast that it reaches an infinite population in finite time with a probability of one. □

Example 6.6.7 (Chemistry 1)**.** Consider the chemical system $A \underset{k_2}{\overset{k_1}{\rightleftarrows}} B$ with $A(0) + B(0) = N$ and mass action kinetics. Then, $A(t)$, giving the number of A molecules at time t, is a birth and death process with state space $\{0, 1, \ldots, N\}$ and transition rates

$$b_n = k_2(N - n), \quad \text{and} \quad d_n = k_1 n.$$

□

Example 6.6.8 (Chemistry 2)**.** Consider the chemical system

$$\emptyset \underset{\mu}{\overset{\lambda}{\rightleftarrows}} A$$

and suppose $X(t)$ tracks the number of A molecules. Then this model is a birth and death process with the exact same transition rates as the infinite server queue of Example 6.6.3. □

Returning to a general system, consider the embedded discrete time Markov chain of a general continuous time birth and death process. The transition probabilities of this chain are

$$p_{n,n+1} = p_n \stackrel{\text{def}}{=} \frac{b_n}{b_n + d_n}$$
$$q_{n,n-1} = q_n \stackrel{\text{def}}{=} \frac{d_n}{b_n + d_n}.$$

Note that in this case we have $p_n + q_n = 1$ for all $n \geq 0$. We will first consider when these processes are recurrent and transient, and then consider positive recurrence. The following proposition follows directly from Proposition 4.2.6.

Proposition 6.6.9. *A continuous time birth and death process is transient if and only if*

$$\sum_{k=1}^{\infty} \frac{d_1 \cdots d_k}{b_1 \cdots b_k} < \infty.$$

Proof. From Proposition 4.2.6, the embedded chain, and hence the continuous time chain, is transient if and only if

$$\sum_{k=1}^{\infty} \frac{q_1 \cdots q_n}{p_1 \cdots p_k} < \infty.$$

Noting that

$$\sum_{k=1}^{\infty} \frac{q_1 \cdots q_n}{p_1 \cdots p_k} = \sum_{k=1}^{\infty} \frac{d_1 \cdots d_k}{b_1 \cdots b_k},$$

completes the proof. □

Similarly to the discrete time case, we can now conclude that the single server queue is transient if and only if $\mu < \lambda$, and that the k server queue is transient if and only if $k\mu < \lambda$. For the infinite server queue, and the analogous chemistry example in Example 6.6.8, we have

$$\sum_{k=1}^{\infty} \frac{d_1 \cdots d_k}{p_1 \cdots p_k} = \sum_{k=1}^{\infty} k! \left(\frac{\mu}{\lambda}\right)^k = \infty.$$

Thus, the infinite server queue is always recurrent.

We now turn to the question of positive recurrence and stationary distributions. We know that a stationary distribution η must satisfy $\eta A = 0$, which in component form is

$$\begin{aligned} \eta_0 b_0 &= \eta_1 d_1 \\ (b_k + d_k)\eta_k &= b_{k-1}\eta_{k-1} + d_{k+1}\eta_{k+1}, \quad \text{for} \quad k \geq 1. \end{aligned}$$

Noting that these are the same equations as (4.5) and (4.6), we can conclude that such an η exists and can be made into a probability vector if and only if

$$\sum_{k=1}^{\infty} \frac{b_0 b_1 \cdots b_{k-1}}{d_1 \cdots d_k} < \infty.$$

The following is analogous to Proposition 4.2.7.

Proposition 6.6.10. *There exists a stationary distribution for a continuous time birth and death chain if and only if*

$$\sum_{k=1}^{\infty} \frac{b_0 b_1 \cdots b_{k-1}}{d_1 \cdots d_k} < \infty.$$

In this case,

$$\eta_0 = \left(\sum_{k=0}^{\infty} \frac{b_0 b_1 \cdots b_{k-1}}{d_1 \cdots d_k} \right)^{-1},$$

where the $k = 0$ term in the above sum is taken to be equal to one, and for $k \geq 1$,

$$\eta_k = \frac{b_0 \cdots b_{k-1}}{d_1 \cdots d_k} \eta_0.$$

For example, for the single server queue we have

$$\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^k = \left(1 - \frac{\lambda}{\mu} \right)^{-1},$$

provided $\lambda < \mu$, and in this case

$$\eta_k = \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^k.$$

The expected length of the queue is

$$\sum_{k=0}^{\infty} k\eta_k = k \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^k = \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu} \right)^{-1} = \frac{\lambda}{\mu - \lambda},$$

which grows to infinity as λ approaches μ.

For the infinite server queue and the chemistry model of Example 6.6.8, we have

$$\sum_{k=0}^{\infty} \frac{b_0 \cdots b_{k-1}}{d_1 \cdots d_k} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k = e^{\lambda/\mu}.$$

Therefore, a stationary distribution exists, and since we already know the chain is recurrent we may conclude it is positive recurrent. Note that the stationary distribution is Poisson(λ/μ), and

$$\eta_k = e^{-\lambda/\mu} \frac{(\lambda/\mu)^k}{k!}, \quad \text{for } k \geq 0.$$

In the next chapter, we will see why many models from chemistry and biology have stationary distributions that are Poisson.

We close by considering the generator for a continuous time birth and death process. It is straightforward to show that it satisfies

$$(Af)(i) = b_i(f(i+1) - f(i)) + d_i(f(i-1) - f(i)),$$

for all $i \geq 0$. This fact can be used in the case of linear intensities to easily calculate the time-dependent moments.

Example 6.6.11. Consider linear birth and death process with transition rates

$$b_i = \lambda i$$
$$d_i = \mu i,$$

where $\lambda, \mu > 0$. The generator of the process satisfies

$$(Af)(i) = \lambda i(f(i+1) - f(i)) + \mu i(f(i-1) - f(i)),$$

for all $i \geq 0$. Taking $f(i) = i$ to be the identity, and $X(0) = x$, we have that

$$\begin{aligned}
\mathbb{E}f(X(t)) = \mathbb{E}X(t) &= \mathbb{E}X(0) + \mathbb{E}\int_0^t (Af)(X(s))ds \\
&= x + \mathbb{E}\int_0^t \Big[\lambda X(s)(X(s)+1-X(s)) + \mu X(s)\big(X(s)-1-X(s)\big)\Big]ds \\
&= x + (\lambda - \mu)\int_0^t \mathbb{E}X(s)ds.
\end{aligned}$$

Solving yields

$$\mathbb{E}X(t) = xe^{(\lambda-\mu)t}, \tag{6.15}$$

which, it is worth noting, is the solution to the ordinary differential equation $x'(t) = (\lambda - \mu)x(t)$ that is the standard *deterministic* model for this system. □

6.6.1 A brief look at parameter inference

While not a topic covered in this course to any great depth, we turn briefly to the question of parameter inference. We do so by considering the linear birth and death process of Example 6.6.11. Specifically, we suppose that we believe our system can be modeled as a linear birth and death system, however we do not know the key parameters, λ or μ.

We first note that we have multiple options for how to model the dynamics of the process with the two most common choices being (*i*) deterministic ODEs and (*ii*) the stochastic model considered in Example 6.6.11. If we choose to model using ordinary differential equations, then the time dependent solution to the process, equation (6.15), only depends upon the parameter $\lambda - \mu$, and *not* on the actual values of λ and μ. Therefore, there will not be a way to recover λ and μ from data, only their difference.

Perhaps surprisingly, more can be accomplished in the stochastic setting. While the mean value of $X(t)$ is a function of the single parameter $\lambda - \mu$ given in equation 6.15, we can also solve for the variance, which turns out to be (this is the subject of a homework exercise)

$$\mathsf{Var}(t) = X(0)\left(\frac{\lambda+\mu}{\lambda-\mu}\right)\left[e^{2(\lambda-\mu)t} - e^{(\lambda-\mu)t}\right]. \tag{6.16}$$

Note that this is a function of both the difference *and* the sum of λ and μ. Therefore, we may use the mean and variance of any data to approximate both λ and μ. In this way, we see that having noisy data actually *helps* us solve for parameters.

For example, suppose that we know that $X(0) = 60$ (perhaps because we begin each experiment with 60 bacteria), and after a number of experiments we found the mean of the process at time 1 is 22.108, and the variance is 67.692. Using the equations for the mean and variance, equations (6.15) and (6.16) respectively, this reduces to solving the system of equations

$$\begin{aligned}\lambda - \mu &= -.9984\\ \lambda + \mu &= 4.8406,\end{aligned}$$

yielding

$$\lambda = 1.9211 \quad \text{and} \quad \mu = 2.9195.$$

For comparison sake, the data reported above was actually generated from 1,000 samples of a process with actual values of $\lambda = 2$ and $\mu = 3$.

6.7 Exercises

1. Consider a continuous time Markov chain with state space $\{1, 2, 3, 4\}$ and generator matrix

$$A = \begin{bmatrix} -3 & 2 & 0 & 1 \\ 0 & -2 & 1/2 & 3/2 \\ 1 & 1 & -4 & 2 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

   Write a Matlab code that simulates a path of this chain. To do so, use the construction provided in the notes (i.e. simulate the embedded chain and holding times sequentially). Using this code and assuming that $X(0) = 1$, estimate $\mathbb{E}X(3)$ by averaging over 10,000 such paths. Note that you will need to make sure you break your "for" or "while" loop after you see that the time will go beyond $T = 3$, without updating the state for that step.

2. In Example 6.2.4, it was stated that if $X(0) = 0$, then the probability of an explosion was 1/3. Why is that?

3. For Example 6.5.2, verify that the generator of the process satisfies equation (6.13).

4. Using Dynkin's formula, calculate $\mathsf{Var}(X(t))$ of the linear birth process of Example 6.5.2.

5. Using Dynkin's formula, calculate $\mathsf{Var}(X(t))$ of the linear birth and death process of Example 6.6.11.

6. Consider the linear birth and death process of Example 6.16. Suppose that $X(0) = 100$, and at time $T = 2$ the mean of 100 experiments is 212, and the variance is 1,100. Estimate the parameters λ and μ.