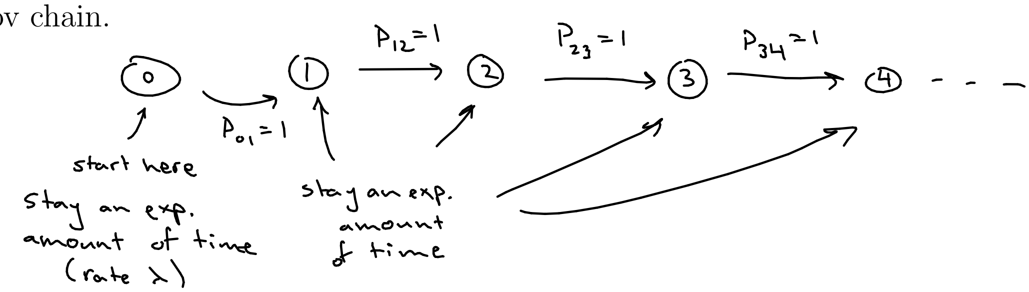
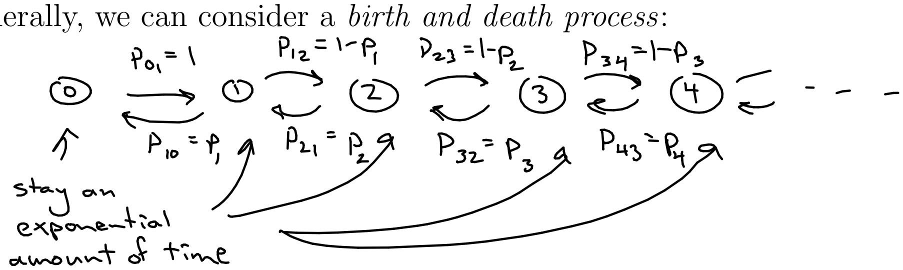
6.1 Introduction

We now study the continuous-time analogs of the Markov chains we studied in Chapter 4.

- Just like discrete-time Markov chains, continuous-time Markov chains are defined by the property that the future state, given the present state and all past states, is independent of the past, and depends only on the present.
- The difference is in the amount of time that the process spends in each state!
  - In a discrete-time Markov chain, the time steps are discrete. The length of time between successive transitions is constant, e.g., one step per day.
  - In a continuous-time Markov chain, we will see that the length of time between successive transitions is an exponential random variable.
- A Poisson process $\{N(t),\ t \geq 0\}$ with rate λ is one example of a continuous-time Markov chain.



- More generally, we can consider a *birth and death process*:



The properties of exponential random variables will continue to be extremely important in this chapter. Recall the following, in particular.

- Exponential random variables are the unique *memoryless* random variables.
- If $X_1, X_2, \ldots, X_n$ are independent exponential random variables with rates $\lambda_1, \lambda_2, \ldots, \lambda_n$, then
  - $\min\{X_1, X_2, \ldots, X_n\}$ is exponential with rate $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.
  - $P\{X_i = \min\{X_1, X_2, \ldots, X_n\}\} = \dfrac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$

6.2 Continuous-Time Markov Chains

Definition 6.2.1. A *continuous-time Markov chain* is a continuous-time stochastic process $\{X(t),\ t \geq 0\}$ with a countable state space that satisfies the so-called *Markov property*:

> The conditional distribution of the future state $X(t+s)$, given the present state $X(s)$ and past states $X(u),\ 0 \leq u < s$, depends only on the present state and is independent of the past.

In other words: $P\{X(t+s)=j \mid X(s)=i,\ X(u)=x(u) \text{ for all } 0 \leq u < s\}$

$$= P\{X(t+s)=j \mid X(s)=i\},$$

for all s,t, i,j, $x(u)$ for all $0 \leq u < s$.

If, in addition, the probability

$$P\{X(t+s)=j \mid X(s)=i\}$$

is independent of s, then the continuous-time Markov chain $\{X(t),\ t \geq 0\}$ is said to have

stationary transition probabilities.

We will assume that all Markov chains under consideration have stationary transition probabilities.

Observation 6.2.2. Suppose that a continuous-time Markov chain starts in state i, and let T_i be the time until the first transition out of state i. Then

$$P\{T_i > s+t \mid T_i > s\} = P\{T_i > t\} \quad \text{for all } s,t > 0.$$

- In other words, the random variable T_i is memoryless.
- By Proposition 5.2.6, this means that T_i is exponential.

Sketch of Proof. Suppose for some $s,t>0$ that $T_i > s$. So the process is in state i at all times $0 \leq u < s$, and at time s. By the Markov property, the probability that the process remains in state i at any time greater than s depends only on the assumption that $X(s)=i$, and not on the past. So the probability that $T_i > t+s$, given $T_i > s$, is just the unconditional probability that $T_i > t$. ∎

In fact, Observation 6.2.2 gives us another way of defining a continuous-time Markov chain. Namely, it is a stochastic process having the properties that each time it enters state i,

(i) the amount of time it spends in state i is

exponential with some rate, say v_i.

(ii) when the process leaves state i, it enters state j with

some fixed probability P_{ij}.

Note that the probabilities P_{ij} must satisfy

- $P_{ii} = 0$
- $\sum_{\text{all } j} P_{ij} = 1$

for all states i.

Thus, a continuous-time Markov chain is a stochastic process that moves from state to state in accordance with a discrete-time Markov chain. However, the amount of time that the process spends in each state, before transitioning to the next state, is exponentially distributed!

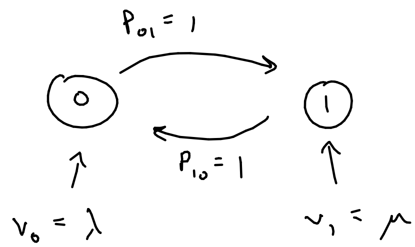
Observation 6.2.3. The amount of time T_i that a continuous-time Markov chain spends in state i before transitioning to another state, and the next state visited, are independent random variables.

Sketch of Proof. Suppose towards a contradiction that the next state visited depends on the length of time T_i spent in state i. Then information about how long the process has been in state i would be relevant to the probability that the process next enters state j, and this contradicts the Markov property. $\square$

Example 6.2.4. Lucas' friend Tom stands downtown and asks passersby to donate to a charity.

(a) Suppose that people stop to talk to Tom according to a Poisson process with rate λ. When someone stops, the amount of time that they spend talking to Tom is an exponential random variable with rate μ. Note that no one waits in line – if Tom is busy, everyone just passes by.

- For $t \geq 0$, let $X(t) = 1$ if Tom is talking to someone at time t, and 0 otherwise.
- Then $\{X(t),\ t \geq 0\}$ is a continuous-time Markov chain with...



Transition Rates: $v_0 = \lambda$ and $v_1 = \mu$

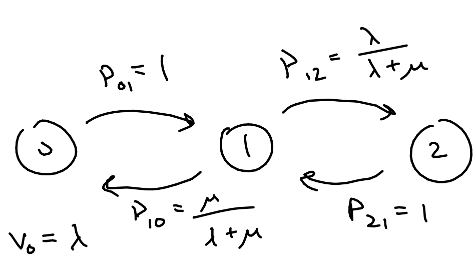
Transition Probabilities: $P_{01} = 1$ and $P_{10} = 1$

(b) Now suppose that Tom has a coworker named Lewis. People still stop according to a Poisson process with rate λ, and the time that they spend talking to either Tom or Lewis is exponential with rate μ. (Assume that everything is independent.)

- For $t \geq 0$, let $Y(t)$ be the number of people being served (by either Tom or Lewis) at time t.
- Then $\{Y(t),\ t \geq 0\}$ is a continuous-time Markov chain with...

Transition Rates: $v_0 = \lambda$, $v_1 = \lambda + \mu$, $v_2 = 2\mu$

Transition Probabilities: $P_{01} = 1$, $P_{12} = \dfrac{\lambda}{\lambda+\mu}$, $P_{10} = \dfrac{\mu}{\lambda+\mu}$, $P_{21} = 1$.



$v_2 = 2\mu$

minimum of the two service times

$v_1 = \lambda + \mu$

since this is the minimum of two independent exponentials

6.3 Birth and Death Processes

Consider a system whose state at any time is represented by the number of people in the system at that time. Suppose that whenever there are n people in the system,

- the length of time until the next person arrives in the system is

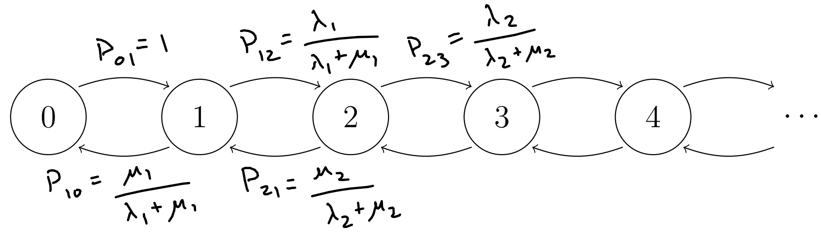
  exponential with rate λ_n.

- the length of time until the next person departs from the system is

  exponential with rate μ_n, independently of arrivals.

If $X(t)$ is the number of people in the system at time t, then $\{X(t), t \geq 0\}$ is called a *birth and death process*.

- The rates $\lambda_0, \lambda_1, \lambda_2, \ldots$ are called the birth rates
- The rates $\mu_1, \mu_2, \ldots$ are called the death rates.



$v_0 = \lambda_0 \quad v_1 = \lambda_1 + \mu_1 \quad v_2 = \lambda_2 + \mu_2$

Definition 6.3.1. A *birth and death process* with

- Birth rates: λ_n for $n \geq 0$
- Death rates: μ_n for $n \geq 1$

is a continuous-time Markov chain with states $\{0, 1, 2, \ldots\}$ for which the transition rates and probabilities are given by

$$v_0 = \lambda_0$$

$$v_i = \lambda_i + \mu_i \quad \text{for all } i \geq 1$$

$$P_{01} = 1$$

$$P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i} \quad \text{for all } i \geq 1$$

$$P_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i} \quad \text{for all } i \geq 1.$$

Example 6.3.2. Consider a birth and death process $\{N(t),\, t \geq 0\}$ having

- Birth rates: $\lambda_n = \lambda$ for all $n \geq 0$
- Death rates: $\mu_n = 0$ for all $n \geq 1$

In this process,

- the time between successive births (or arrivals) is exp. with rate λ.
- deaths (or departures) never occur!

So if $N(0) = 0$, then $\{N(t),\, t \geq 0\}$ is a Poisson process. :)

$$\textcircled{0} \xrightarrow{1} \textcircled{1} \xrightarrow{1} \textcircled{2} \xrightarrow{1} \textcircled{3} \longrightarrow \textcircled{4} \longrightarrow \cdots$$

$v_0 = \lambda \qquad v_1 = \lambda \qquad v_2 = \lambda$

Definition 6.3.3. A *pure birth process* is a birth and death process for which $\mu_n = 0$ for all $n \geq 1$.

Example 6.3.4. Consider a population of cells which can divide into two but cannot die. Suppose that each cell divides into two after an exponentially distributed amount of time with rate λ, independently of all other cells, and of the total number of cells. Let $X(t)$ be the number of cells present at time t.

Suppose that there are n cells present at some moment in time. What can you say about the time until the next division?

It is exponential with rate $\underbrace{\lambda + \lambda + \cdots + \lambda}_{n \text{ times}} = n\lambda$

So $\{X(t),\, t \geq 0\}$ is a pure birth process with $\lambda_n = n\lambda$

$$\textcircled{1} \xrightarrow{1} \textcircled{2} \xrightarrow{1} \textcircled{3} \xrightarrow{1} \textcircled{4} \longrightarrow \cdots$$

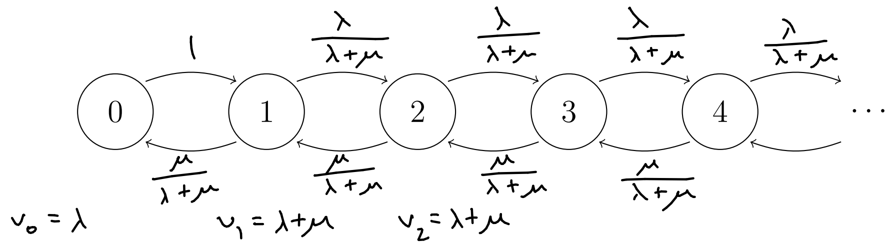
$v_1 = \lambda \qquad v_2 = 2\lambda \qquad v_3 = 3\lambda$

Example 6.3.5 (A single-server queuing system)**.** Suppose that customers arrive at a service station with a single server in accordance with a Poisson process with rate λ.

- Upon arrival, each customer goes directly into service if the server is free. Otherwise, they join the queue, i.e., they wait in line.
- When the server finishes serving a customer, the customer leaves the system and the next customer in line (if there is one) enters service.

Suppose that service times are independent exponential random variables with rate μ, which are also independent of the arrivals. For all $t \geq 0$, let $X(t)$ be the number of customers in the system at time t. Then $\{X(t), t \geq 0\}$ is a birth and death process with

- Birth rates: $\lambda_n = \lambda$ for all $n \geq 0$
- Death rates: $\mu_n = \mu$ for all $n \geq 1$.



This type of system is called an $M/M/1$ queuing system

arrivals are Markovian

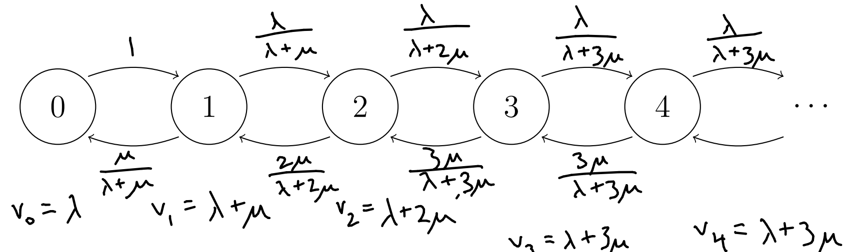
service times are Markovian

1 server

Example 6.3.6 (A multi-server queuing system)**.** Suppose now that everything is as in the previous example, but that we now have s servers for some $s \geq 1$. Suppose that service times at each server are exponential with common rate μ. If $X(t)$ is the number of customers in the system at time t, then $\{X(t), t \geq 0\}$ is a birth and death process with

- Birth rates: $\lambda_n = \lambda$ for $n \geq 0$.
- Death rates: $\mu_n = \begin{cases} s\mu & \text{for } n \geq s \\ n\mu & \text{for } n < s \end{cases}$

E.g., $s = 3$

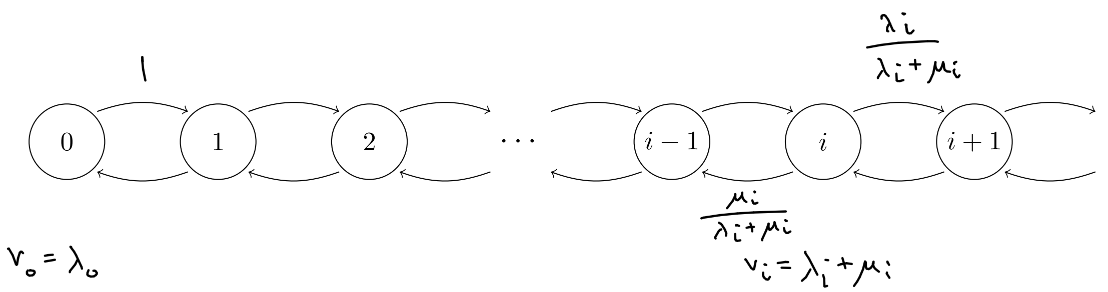


This type of system is called an

$M/M/s$ queuing system.

Now consider a general birth and death process with

- Birth rates: λ_n for $n \geq 0$
- Death rates: μ_n for $n \geq 1$



For $i \geq 0$, let T_i denote the time, starting from state i, until the process enters state $i+1$. We will develop a recursion for $E[T_i]$.

- Note that $E[T_0] = \frac{1}{\lambda_0}$
- For all $i \geq 1$, we now wish to express $E[T_i]$ in terms of $E[T_{i-1}]$.

Let S_i be the time until the first transition from state i, and let R_i be the remaining time, after the first transition from state i, until the process enters state $i+1$.

- Then $T_i = S_i + R_i$
- Note that $E[S_i] = \frac{1}{\lambda_i+\mu_i}$
- We compute $E[R_i]$ by conditioning on the first transition.

$I_i = \begin{cases} 1, & \text{if first transition is to } i-1 \\ 0, & \text{if first transition is to } i+1. \end{cases}$

Let I_i be an indicator variable for the event that the first transition from state i is to state $i-1$.

- Then $E[R_i] = E[R_i \mid I_i = 1] \cdot P\{I_i = 1\} + E[R_i \mid I_i = 0] \cdot P\{I_i = 0\}$

$$= \left(E[T_{i-1}] + E[T_i]\right) \frac{\mu_i}{\lambda_i+\mu_i} + 0 \cdot \frac{\lambda_i}{\lambda_i+\mu_i}$$

Given that the first transition was to state $i-1$, all times "reset" by the Markovian property, so the remaining time until we reach state $i+1$ is $T_{i-1} + T_i$.

We conclude that

$$E[T_i] = E[S_i] + E[R_i] = \frac{1}{\lambda_i+\mu_i} + \frac{\mu_i}{\lambda_i+\mu_i}\left(E[T_{i-1}] + E[T_i]\right)$$

which we can rearrange to find

$$E[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E[T_{i-1}], \quad \text{for all } i \geq 1.$$

We can also find a recursion for $\text{Var}(T_i)$.

- Note that $\text{Var}(T_0) = \dfrac{1}{\lambda_0^2}$
- Now for all $i \geq 1$, we wish to express $\text{Var}(T_i)$ in terms of $\text{Var}(T_{i-1})$.

To do this, we condition on I_i, and use the law of total variance. From our previous work, we have

$$E[T_i \mid I_i] = E[S_i \mid I_i] + E[R_i \mid I_i] = \frac{1}{\lambda_i + \mu_i} + I_i \cdot \left(E[T_{i-1}] + E[T_i]\right)$$

Thus we have

$$\text{Var}(E[T_i \mid I_i]) = \text{Var}\left(\frac{1}{\lambda_i + \mu_i} + I_i\left(E[T_{i-1}] + E[T_i]\right)\right)$$

constant

constant

$$= \left(E[T_{i-1}] + E[T_i]\right)^2 \cdot \text{Var}(I_i)$$

$$= \left(E[T_{i-1}] + E[T_i]\right)^2 \cdot \frac{\lambda_i \mu_i}{(\lambda_i + \mu_i)^2}$$

I_i is binomial with parameters $n = 1$ and $p = \dfrac{\mu_i}{\lambda_i + \mu_i}$.

So $\text{Var}(I_i) = np(1-p)$
$= p(1-p)$

We still need to compute $E[\text{Var}(T_i \mid I_i)]$.

- $\text{Var}(T_i \mid I_i = 0) = \text{Var}(S_i \mid I_i = 0) + \text{Var}(R_i \mid I_i = 0)$

$$= \frac{1}{(\lambda_i + \mu_i)^2} + 0$$

Since S_i and R_i are independent by the Markov property.

- $\text{Var}(T_i \mid I_i = 1) = \text{Var}(S_i \mid I_i = 1) + \text{Var}(R_i \mid I_i = 1)$

$$= \frac{1}{(\lambda_i + \mu_i)^2} + \text{Var}(T_{i-1} + T_i \mid I_i = 1)$$

$$= \frac{1}{(\lambda_i + \mu_i)^2} + \text{Var}(T_{i-1}) + \text{Var}(T_i)$$

- So $\text{Var}(T_i \mid I_i) = \dfrac{1}{(\lambda_i + \mu_i)^2} + I_i \cdot \left(\text{Var}(T_{i-1}) + \text{Var}(T_i)\right)$

Thus we have

$$E[\text{Var}(T_i \mid I_i)] = \frac{1}{(\lambda_i + \mu_i)^2} + \frac{\mu_i}{\lambda_i + \mu_i} \cdot \left(\text{Var}(T_{i-1}) + \text{Var}(T_i)\right)$$

We conclude that

$$\text{Var}(T_i) = \text{Var}\left(E[T_i \mid I_i]\right) + E\left[\text{Var}(T_i \mid I_i)\right]$$

which we can rearrange to find

$$\text{Var}(T_i) = \frac{1}{\lambda_i(\lambda_i + \mu_i)} + \frac{\mu_i}{\lambda_i}\text{Var}(T_{i-1}) + \frac{\mu_i}{\lambda_i + \mu_i}\left(E[T_{i-1}] + E[T_i]\right)^2.$$

To summarize, we have shown the following.

Proposition 6.3.7. *Consider a birth and death process with*

- *Birth rates:* λ_n *for* $n \geq 0$
- *Death rates:* μ_n *for* $n \geq 1$

and let T_i *be the time, starting from state* i*, until the process enters state* $i+1$*. Then we have the following recurrence relations.*

- $E[T_0] = \frac{1}{\lambda_0}$ *and*
$$E[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E[T_{i-1}] \quad \textit{for all } i \geq 1.$$
- $\mathrm{Var}(T_0) = \frac{1}{\lambda_0^2}$ *and*
$$\mathrm{Var}(T_i) = \frac{1}{\lambda_i(\lambda_i + \mu_i)} + \frac{\mu_i}{\lambda_i}\mathrm{Var}(T_{i-1}) + \frac{\mu_i}{\lambda_i + \mu_i}\left(E[T_{i-1}] + E[T_i]\right)^2 \quad \textit{for all } i \geq 1.$$

Given specific birth and death rates, we can compute $E[T_i]$ and $\mathrm{Var}(T_i)$ rather efficiently using these recurrence relations. In special cases, we can find explicit formulas for $E[T_i]$ and $\mathrm{Var}(T_i)$.

Example 6.3.8. Consider a birth and death process with

- Birth rates: $\lambda_n = \lambda$ for $n \geq 0$
- Death rates: $\mu_n = \mu$ for $n \geq 1$

Find an explicit formula for $E[T_i]$.

$$E[T_0] = \frac{1}{\lambda}$$

$$E[T_1] = \frac{1}{\lambda} + \frac{\mu}{\lambda} E[T_0] = \frac{1}{\lambda} + \frac{1}{\lambda}\cdot\frac{\mu}{\lambda} = \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda}\right)$$

$$E[T_2] = \frac{1}{\lambda} + \frac{\mu}{\lambda} E[T_1] = \frac{1}{\lambda} + \frac{\mu}{\lambda}\left(\frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda}\right)\right) = \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2\right)$$

$$E[T_3] = \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2 + \left(\frac{\mu}{\lambda}\right)^3\right)$$

Note: This can be proven by induction

In general, we have

$$E[T_i] = \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda} + \left(\frac{\mu}{\lambda}\right)^2 + \cdots + \left(\frac{\mu}{\lambda}\right)^i\right)$$

$$= \begin{cases} \frac{1}{\lambda}\cdot\left(\frac{1 - \left(\frac{\mu}{\lambda}\right)^{i+1}}{1 - \frac{\mu}{\lambda}}\right) & \text{if } \mu \neq \lambda \\ \frac{1}{\lambda}\cdot(i+1) & \text{if } \mu = \lambda \end{cases}$$

6.4 The Transition Probability Functions

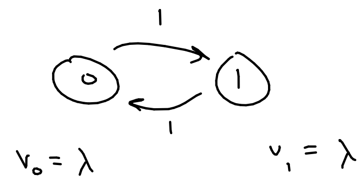
Recall that P_{ij} is the probability that when the process leaves state i, it enters state j. This constant probability is not to be confused with the probability function that we study in this section.

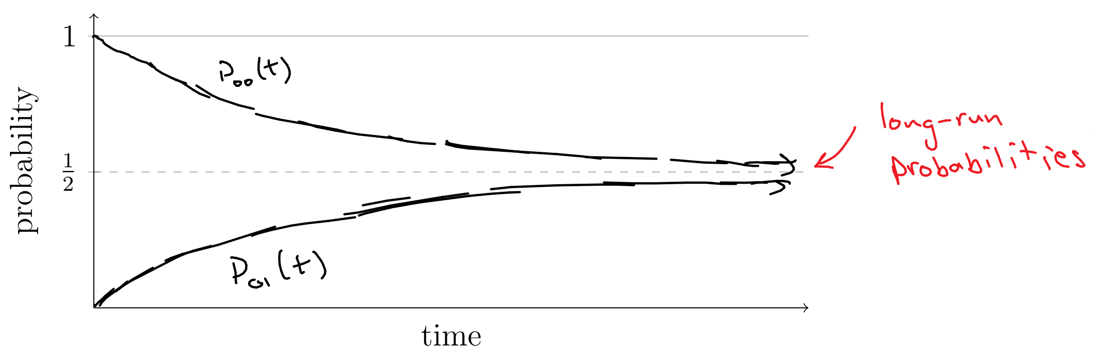
Definition 6.4.1. For a continuous-time Markov chain, $\{X(t), t \geq 0\}$ let

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

denote the probability that, starting in state i, the process will be in state j at time t. These functions $\{P_{ij}(t)\}$ are called the *transition probability functions.*

Example 6.4.2. Consider a continuous time Markov chain with state space $\{0, 1\}$ and transition rates $v_0 = v_1 = \lambda$. What do you think the graphs of the transition probability functions $P_{00}(t)$ and $P_{01}(t)$ look like?





The problem of finding the transition probability functions for a continuous-time Markov chain is a little bit like finding the n-step transition probabilities for a discrete-time Markov chain. Recall that for a discrete-time Markov chain, we had the Chapman-Kolmogorov equations, which said that

$$P_{ij}^{(n+m)} = \sum_{\text{all } k} P_{ik}^{(n)} P_{kj}^{(m)} \quad \text{for all } n, m \geq 0 \text{ and states } i, j.$$

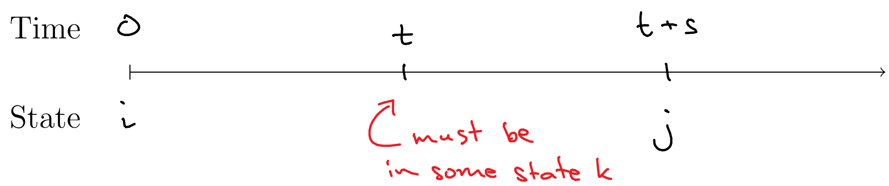
The following are called the Chapman-Kolmogorov equations for a continuous-time Markov chain:

$$P_{ij}(t+s) = \sum_{\text{all } k} P_{ik}(t) \cdot P_{kj}(s)$$

Just as the discrete-time Chapman-Kolmogorov equations were helpful in finding the n-step transition probabilities for a discrete-time Markov chain, the continuous-time Chapman-Kolmogorov equations will be useful for finding the transition probability functions of a continuous-time Markov chain.

Lemma 6.4.3 (The Chapman-Kolmogorov Equations). *For all times $s, t \geq 0$ and all states i and j of a continuous-time Markov chain, we have*

$$P_{ij}(t+s) = \sum_{\text{all } k} P_{ik}(t) \cdot P_{kj}(s)$$



Proof. Letting $\{X(t), t \geq 0\}$ denote our Markov chain, we have

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

since we must be in exactly one state at time t.

$$= \sum_{\text{all } k} P\{X(t+s) = j, X(t) = k \mid X(0) = i\}$$

$$= \sum_{\text{all } k} P\{X(t+s) = j \mid X(t) = k, X(0) = i\} \cdot P\{X(t) = k \mid X(0) = i\}$$

by the Markov property

$$= \sum_{\text{all } k} P\{X(t+s) = j \mid X(t) = k\} \cdot P\{X(t) = k \mid X(0) = i\}$$

$$= \sum_{\text{all } k} P\{X(s) = j \mid X(0) = k\} \cdot P\{X(t) = k \mid X(0) = i\}$$

$$= \sum_{\text{all } k} P_{kj}(s) \cdot P_{ik}(t) \qquad \square$$

Definition 6.4.4. For any pair of states i and j, let

$$q_{ij} = v_i P_{ij}$$

Warning: This is not $P_{ij}(t)$

rate at which we leave state i

probability that when we leave state i, we transition to state j.

- Essentially, q_{ij} is the rate at which the process makes a transition from state i into state j.
- The quantities $\{q_{ij}\}$ are called the *instantaneous transition rates*.

Lemma 6.4.5. *Let i and j be distinct states in a continuous-time Markov chain. Then the following hold.*

(a) $\lim_{h \to 0} \dfrac{1 - P_{ii}(h)}{h} = v_i$

(b) $\lim_{h \to 0} \dfrac{P_{ij}(h)}{h} = q_{ij}$

Proof. For (a), first note that since the amount of time spent in state i is exponential with rate λ_i, the probability of two or more transitions in a time h is $o(h)$. $\therefore$ $1 - P_{ii}(h)$, the probability that starting in state i, we are in a different state j after h units of time, is just the probability that a transition occurs by time h plus something small compared to h.

Therefore,

$$1 - P_{ii}(h) = v_i h + o(h).$$

Thus we conclude that $\lim_{h \to 0} \dfrac{1 - P_{ii}(h)}{h} = \lim_{h \to 0} \dfrac{v_i h + o(h)}{h} = v_i$.

For (b), note that $P_{ij}(h)$, the probability that, starting in state i, we are in state j at time h, is the probability that we make a transition from state i to state j by time h, plus something small compared to h.

Thus we have

$$P_{ij}(h) = v_i \, h \cdot P_{ij} + o(h)$$

rate at which we leave i

length of interval

prob. that we go to j when we leave i

$$= q_{ij} h + o(h)$$

$$\therefore \lim_{h \to 0} \frac{P_{ij}(h)}{h} = \lim_{h \to 0} \frac{q_{ij} h + o(h)}{h} = \lim_{h \to 0} q_{ij} + \lim_{h \to 0} \frac{o(h)}{h}$$

$$= q_{ij}.$$

$\square$

We have proven the following:

- $P_{ij}(s+t) = \sum_{\text{all } k} P_{ik}(s)\, P_{kj}(t)$ for all states i, j, all times $s, t \geq 0$.
- $\lim_{h\to 0} \frac{1 - P_{ii}(h)}{h} = v_i$ for all states i
- $\lim_{h\to 0} \frac{P_{ij}(h)}{h} = q_{ij}$ for all states $i \neq j$.

We use these facts to establish a set of differential equations that are satisfied by the transition probability functions.

Theorem 6.4.6 (Kolmogorov's Backward Equations)**.** *For all times $t \geq 0$ and all states i and j of a continuous-time Markov chain, we have*

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

Proof. Consider

$$P_{ij}(h+t) - P_{ij}(t) = \sum_{\text{all } k} P_{ik}(h) \cdot P_{kj}(t) - P_{ij}(t)$$

$$= \sum_{k \neq i} P_{ik}(h) \cdot P_{kj}(t) + P_{ii}(h) \cdot P_{ij}(t) - P_{ij}(t)$$

$$= \sum_{k \neq i} P_{ik}(h) \cdot P_{kj}(t) - P_{ij}(t) \left[1 - P_{ii}(h)\right]$$

Therefore,

$$P'_{ij}(t) = \lim_{h \to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h}$$

$$= \lim_{h\to 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} \cdot P_{kj}(t) - \lim_{h \to 0} P_{ij}(t) \cdot \frac{1 - P_{ii}(h)}{h}$$

this interchange of limit and sum can be justified (but we won't bother)

$$= \sum_{k \neq i} \lim_{h \to 0} \frac{P_{ik}(h)}{h} \cdot P_{kj}(t) - P_{ij}(t) \cdot \lim_{h \to 0} \frac{1 - P_{ii}(h)}{h}$$

$$= \sum_{k \neq i} q_{ik} P_{kj}(t) - P_{ij}(t) \cdot v_i,$$

as claimed in the theorem statement. $\square$

For a general continuous-time Markov chain, the backwards equations say that

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$$

For a pure birth process, the backwards equations become

$$P'_{ij}(t) = q_{i,i+1} \cdot P_{i+1,j}(t) - \lambda_i P_{ij}(t) = \lambda_i P_{i+1,j}(t) - \lambda_i P_{ij}(t)$$

For a birth and death process, they become

$$P'_{0j}(t) = q_{01} \cdot P_{1j}(t) - \lambda_0 P_{0j}(t) = \lambda_0 P_{1j}(t) - \lambda_0 P_{0j}(t)$$

and

$$P'_{ij}(t) = q_{i,i-1} \cdot P_{i-1,j}(t) + q_{i,i+1} \cdot P_{i+1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t)$$

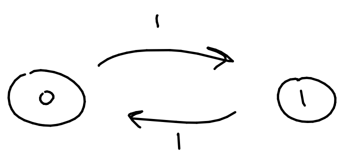
$q_{i,i-1}$

$= v_i \cdot P_{i,i-1}$

$$= (\lambda_i + \mu_i) \cdot \frac{\mu_i}{\lambda_i + \mu_i} \cdot P_{i-1,j}(t) + (\lambda_i + \mu_i) \cdot \frac{\lambda_i}{\lambda_i + \mu_i} \cdot P_{i+1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t)$$

$$= \mu_i P_{i-1,j}(t) + \lambda_i P_{i+1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) \quad \text{for } t \geq 1.$$

Example 6.4.7. Consider a machine that works for an exponential amount of time with rate λ before it breaks down. Suppose that it takes an exponential amount of time with rate rate μ to repair the machine. If the machine is in working condition at time 0, then what is the probability that it will be working at time t?



State 0 : working

State 1 : under repair

rate $v_0 = \lambda$ 　 rate $v_1 = \mu$

We want $P_{00}(t)$.

$q_{01} = \lambda$ 　 $q_{10} = \mu$

The Kolmogorov backward equations give

$$P'_{00}(t) = q_{01} \cdot P_{10}(t) - v_0 P_{00}(t) = \lambda [P_{10}(t) - P_{00}(t)]$$

$$P'_{10}(t) = q_{10} \cdot P_{00}(t) - v_1 P_{10}(t) = \mu [P_{00}(t) - P_{10}(t)]$$

So we have

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

Adding these two equations together gives

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

Integrating both sides gives

$$\mu P_{00}(t) + \lambda P_{10}(t) = C \quad \text{for some constant } C.$$

Since $P_{00}(0) = 1$ and $P_{10}(0) = 0$, we find $\mu = C$.

$\therefore$ we have

$$\mu P_{00}(t) + \lambda P_{10}(t) = \mu, \quad \text{or equiv.} \quad \lambda P_{10}(t) = \mu\left[1 - P_{00}(t)\right]$$

Substituting back into the equation $P_{00}'(t) = \lambda\left[P_{10}(t) - P_{00}(t)\right]$, we find

$$P_{00}'(t) = \mu\left[1 - P_{00}(t)\right] - \lambda P_{00}(t)$$

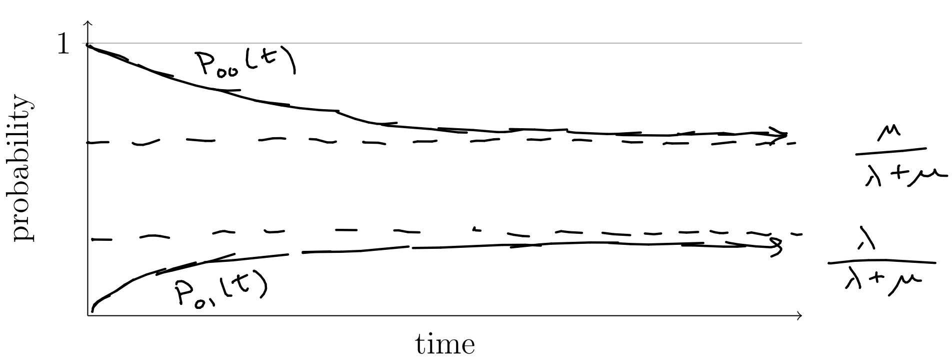
$$= \mu - (\lambda + \mu) P_{00}(t).$$

We solve this (on an extra page) to find

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

Then

$$P_{01}(t) = 1 - P_{00}(t)$$



Another set of differential equations, different from the backward equations, can also be derived in a similar manner.

Theorem 6.4.8 (Kolmogorov's forward equations)**.** *For all times $t \geq 0$ and all states i and j of a continuous-time Markov chain that satisfies certain regularity conditions, we have*

$$P'_{ij}(t) = \sum_{k \neq j} P_{ik}(t) \cdot q_{kj} - v_j P_{ij}(t)$$

Note: In particular, the "regularity conditions" are satisfied by all birth and death processes.

Proposition 6.4.9. *For a pure birth process, we have*

$$P_{ii}(t) = e^{-\lambda_i t} \qquad \text{for } i \geq 0$$

and

$$P_{ij}(t) = \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} P_{i,j-1}(s)\, ds \qquad \text{for } j > i.$$

<u>Proof:</u> One can use the forward equations to find these formulas – see the text.

We have $P_{00}'(t) = \mu - (\lambda+\mu)\, P_{00}(t)$

Letting $h(t) = P_{00}(t) - \frac{\mu}{\lambda+\mu}$

$$h'(t) = P_{00}'(t) = \mu - (\lambda+\mu) P_{00}(t) = \mu - (\lambda+\mu)\left[h(t) + \frac{\mu}{\lambda+\mu} \right]$$

$$= \mu - (\lambda+\mu)\, h(t) - \mu$$

$$= -(\lambda+\mu) h(t)$$

So $h'(t) = -(\lambda+\mu) h(t)$

$$\Longrightarrow \quad \frac{h'(t)}{h(t)} = -(\lambda+\mu)$$

Integrating both sides $\quad \ln(h(t)) = -(\lambda+\mu)t + C$

$$\Longrightarrow \quad h(t) = e^{-(\lambda+\mu)t + C}$$

$$\Longrightarrow \quad h(t) = Ke^{-(\lambda+\mu)t}, \quad \text{where } K = e^{C}.$$

Thus we have

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

Since $P_{00}(0) = 1$, we have $\quad 1 = K + \frac{\mu}{\lambda+\mu}$

$$\Longrightarrow \quad K = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}.$$

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

6.5 Limiting Probabilities

It can be a lot of work to determine the transition probability functions! Often, we are most interested in the *long-run behaviour* of these functions. We want to know the proportion of time that we spend in each state, in the long run. For example, if we have an $M/M/3$ queuing system, we may wish to answer the following questions.

- What long-run proportion of time are all three servers busy?
- What long-run proportion of arriving customers are served immediately?

As was the case with discrete-time Markov chains, the probability that a continuous-time Markov chain will be in state j at time t often converges to a limiting value that is independent of the initial state. If this is the case, then we let

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

As for discrete-time Markov chains, the quantity π_j represents the long-run proportion of time that the process spends in state j.

We now derive a set of equations that these limiting probabilities must satisfy (if they exist). We will make the following assumptions:

- The Kolmogorov forward equations hold.
- For all states i and j, we have $\lim_{t\to\infty} P_{ij}(t) = \pi_j$.
- We can interchange limit and summation wherever necessary.

From the Kolmogorov forward equations we have

$$P_{ij}'(t) = \sum_{k\neq j} P_{ik}(t)\cdot q_{kj} - v_j P_{ij}(t)$$

Taking limits on both sides, we have

$$\lim_{t\to\infty} P_{ij}'(t) = \sum_{k\neq j} \lim_{t\to\infty} P_{ik}(t)\cdot q_{kj} - \lim_{t\to\infty} v_j P_{ij}(t)$$

$$\Rightarrow \lim_{t\to\infty} P_{ij}'(t) = \sum_{k\neq j} \pi_k q_{kj} - v_j \pi_j$$

Since the right-hand side converges, so does the left-hand side.

Since $P_{ij}(t)$ is bounded between 0 and 1, we must have $\lim_{t\to\infty} P_{ij}'(t) = 0$.

Thus we have $$0 = \sum_{k\neq j} q_{kj}\pi_k - v_j\pi_j,$$

or $$v_j\pi_j = \sum_{k\neq j} q_{kj}\pi_k .$$

We made some assumptions on the previous page in order to simplify the derivation, but the following statement can be proven.

Theorem 6.5.1. *Consider a continuous-time Markov chain that satisfies the following conditions:*

- *All states communicate, i.e., for all states i and j, we have $P_{ij}(t) > 0$ for some $t > 0$.*
- *The Markov chain is positive recurrent, i.e., when we leave any state i, the expected time to return to state i is finite.*

Then the limiting probabilities π_j exist, and they satisfy the equations

$$\begin{cases} v_j \pi_j = \sum_{k \neq j} q_{kj} \pi_k & \text{for all states } j \\ \sum_{\text{all } j} \pi_j = 1 \end{cases}$$

These equations have a nice interpretation.

- In any interval $(0, t)$, the number of transitions into state j must be within 1 of the number of transitions out of state j.
- Hence, in the long run, the rate at which transitions into state j occur must equal the rate at which transitions out of state j occur.
- The rate at which the process leaves state j is

$$v_j \pi_j$$

rate at which we leave state j

long-run proportion of time that we are in state j.

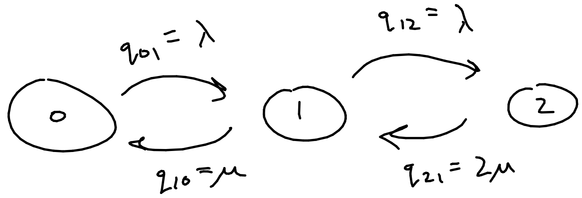
- The rate at which the process enters state j is

$$\sum_{k \neq j} q_{kj} \pi_k$$

- So the first set of equations above simply says that these rates are equal!
- For this reason, they are sometimes referred to as *balance equations*.

Example 6.5.2. Let us reconsider the example in which Tom and Lewis are asking passersby downtown to donate to charity. People that are willing to talk to them pass by according to a Poisson process with rate λ, but they only stop if at least one of Tom or Lewis is free. The time that they spend talking to Tom or Lewis is exponential with rate μ.

(a) Find the limiting probabilities for this process.



$v_0 = \lambda$ $\quad v_1 = \lambda + \mu$ $\quad v_2 = 2\mu$

The limiting probabilities $\pi_0, \pi_1,$ and π_2 satisfy

$$\pi_0 + \pi_1 + \pi_2 = 1 \qquad (*)$$

and the balance equations:

state	rate we leave = rate we enter	
0	$\lambda \pi_0 = \mu \pi_1$	(0)
1	$\lambda \pi_1 + \mu \pi_1 = \lambda \pi_0 + 2\mu \pi_2$	(1)
2	$2\mu \pi_2 = \lambda \pi_1$	(2)

From (0), $\pi_1 = \frac{\lambda}{\mu}\pi_0$. From (2), $\pi_2 = \frac{\lambda}{2\mu}\pi_1 = \frac{\lambda^2}{2\mu^2}\pi_0$

Now from $(*)$, $\pi_0 + \frac{\lambda}{\mu}\pi_0 + \frac{\lambda^2}{2\mu^2}\pi_0 = 1 \implies \pi_0\left[1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2}\right] = 1$

$$\implies \pi_0 = \frac{1}{1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2}}.$$

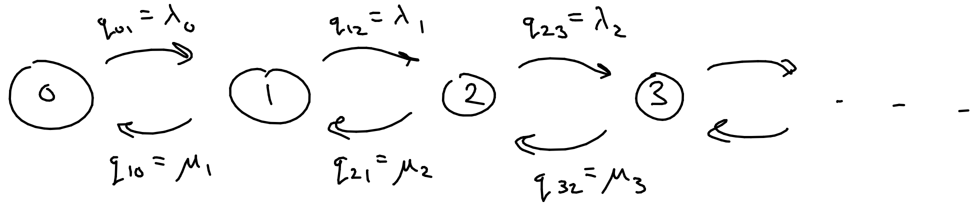
Now we can back substitute to find π_1 and π_2.

(b) If people that are willing to talk to Tom or Lewis arrive at a rate of 15 per hour, and the mean time that each person stops to talk is 5 minutes, find the long-run proportion of the people willing to talk that actually stop.

$\lambda = 15$ and the service times have mean $\frac{5}{60}$ or $\frac{1}{12}$ hours, so $\mu = 12$. The long-run proportion of time that Tom and Lewis are both busy is $\pi_2 = \frac{\lambda^2}{2\mu^2}\pi_0 = \frac{15^2}{2(12)^2} \cdot \frac{1}{3.03125} \approx 0.258$

So the long-run proportion of time that at least one of them is free is $1 - \pi_2 \approx 0.742$. This is also the long-run proportion of people willing to talk that actually stop.

Example 6.5.3. Find the limiting probabilities for a birth and death process with birth rates $\lambda_n > 0$ for all $n \geq 0$ and death rates $\mu_n > 0$ for all $n \geq 1$.



$v_0 = \lambda_0 \qquad v_1 = \lambda_1 + \mu_1 \qquad v_2 = \lambda_2 + \mu_2 \qquad v_3 = \lambda_3 + \mu_3$

To find the limiting probabilities $\pi_0, \pi_1, \pi_2, \ldots$, we solve

$$\sum_{n=0}^{\infty} \pi_n = 1 \quad (*)$$

and the balance equations:

state	rate we leave = rate we enter		
0	$\lambda_0 \pi_0 = \mu_1 \pi_1$	(0)	$\lambda_0 \pi_0 = \mu_1 \pi_1$
1	$\lambda_1 \pi_1 + \mu_1 \pi_1 = \lambda_0 \pi_0 + \mu_2 \pi_2$	(1)	$\lambda_1 \pi_1 = \mu_2 \pi_2$
2	$\lambda_2 \pi_2 + \mu_2 \pi_2 = \lambda_1 \pi_1 + \mu_3 \pi_3$	(2)	$\lambda_2 \pi_2 = \mu_3 \pi_3$
$\vdots$			$\vdots$
n	$\lambda_n \pi_n + \mu_n \pi_n = \lambda_{n-1} \pi_{n-1} + \mu_{n+1} \pi_{n+1}$	(n)	$\lambda_n \pi_n = \mu_{n+1} \pi_{n+1}$

add

get

add

get

We rearrange to find:

$$\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0$$

$$\pi_2 = \frac{\lambda_1}{\mu_2} \pi_1 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0$$

$$\pi_3 = \frac{\lambda_2}{\mu_3} \pi_2 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} \pi_0$$

$$\vdots$$

$$\pi_n = \frac{\lambda_{n-1}}{\mu_n} \pi_{n-1} = \frac{\lambda_0 \lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_n} \pi_0$$

From $(*)$, we have $\displaystyle\sum_{n=0}^{\infty} \pi_n = 1$

$$\Longrightarrow \pi_0 + \frac{\lambda_0}{\mu_1} \pi_0 + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0 + \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} \pi_0 + \cdots = 1$$

$$\Longrightarrow \pi_0 \left[1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \right] = 1 \quad \Longrightarrow \pi_0 = \frac{1}{1 + \Lambda}$$

Call this sum Λ

Our calculations on the previous page show that in order for the limiting probabilities to exist, it is necessary that

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty.$$

One can show that this condition is also sufficient. We summarize all of this in the next theorem statement.

Theorem 6.5.4. *In a birth and death process with birth rates $\lambda_n > 0$ for all $n \geq 0$ and death rates $\mu_n > 0$ for all $n \geq 1$, the limiting probabilities exist if and only if the following sum converges:*

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}$$

If this sum converges to Λ, then the limiting probabilities are given by

$$\pi_0 = \frac{1}{1+\Lambda}$$

and

$$\pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \cdot \frac{1}{1+\Lambda}$$

Example 6.5.5. Consider the $M/M/1$ queue, which has birth rates $\lambda_n = \lambda$ for all $n \geq 0$, and death rates $\mu_n = \mu$ for all $n \geq 1$.

(a) When do the limiting probabilities exist?

They exist when the sum

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=1}^{\infty} \frac{\lambda^n}{\mu^n} = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n \quad \text{converges.}$$

This is a geometric series with ratio $\frac{\lambda}{\mu}$, which converges when $\frac{\lambda}{\mu} < 1$, and diverges when $\frac{\lambda}{\mu} \geq 1$.

Note that this makes intuitive sense!

- If $\frac{\lambda}{\mu} > 1$, then people arrive faster than they can be served, and the queue will grow without bound, meaning the limiting prob.s will all be zero.
- If $\frac{\lambda}{\mu} < 1$, then people are served faster than they arrive, and we expect a small # of people in the queue most of the time.

Suppose now that the limiting probabilities exist, i.e., that $\frac{\lambda}{\mu} < 1$.

(b) What are the limiting probabilities?

$$\Lambda = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda}$$

(sum of geometric series is $\frac{a}{1-r}$, where a is the initial term and r is the ratio)

$$\pi_0 = \frac{1}{1+\Lambda} = \frac{1}{1 + \frac{\lambda}{\mu-\lambda}} = \frac{1}{\frac{\mu - \lambda + \lambda}{\mu - \lambda}} = \frac{\mu - \lambda}{\mu}$$

and $$\pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \pi_0 = \frac{\lambda^n}{\mu^n} \cdot \frac{\mu - \lambda}{\mu} = \left(\frac{\lambda}{\mu}\right)^n \cdot \left(1 - \frac{\lambda}{\mu}\right)$$

E.g., if $\lambda = 3$ and $\mu = 4$, then $\pi_n = \frac{3^n}{4^n} \cdot \frac{1}{4}$ for all $n \geq 0$.

$\pi_0 = \frac{1}{4}$, $\pi_1 = \frac{3}{16}$, $\pi_2 = \frac{9}{64}$, ...

(c) Find the long-run average number of customers in the system.

We find the expected value of the number of people in the system in the long run. This will be $\sum_{n=0}^{\infty} n \cdot \pi_n$, which can be calculated directly,

or we can use the fact that the number of people in the system in the long run is distributed as $X - 1$, where X is a geometric random variable with parameter $p = 1 - \frac{\lambda}{\mu}$. So the mean number of people is $E[X] - 1 = \frac{1}{1 - \frac{\lambda}{\mu}} - 1$.

E.g., if $\lambda = 3$ and $\mu = 4$, then

the average number of people in the system is

$$\frac{1}{1 - \frac{3}{4}} - 1 = 3$$