Continuous-Time Markov Chains

Cong Ma

University of Chicago, Winter 2026

Continuous-Time Markov Chains (CTMC)

Definitions. A stochastic process $\{X(t), t \geq 0\}$ with state space $\mathcal{X}$ is called a *continuous-time Markov chain* if for any two states i, $j \in \mathcal{X}$,

$$\begin{aligned} &\mathrm{P}(\underbrace{X(t+s) = j}_{\text{future}} \,|\, \underbrace{X(s) = i}_{\text{present}}, \underbrace{X(u) = x(u), \text{for } 0 \leq u < s}_{\text{past}}) \\ &= \mathrm{P}(\underbrace{X(t+s) = j}_{\text{future}} \,|\, \underbrace{X(s) = i}_{\text{present}}) \end{aligned}$$

If $\mathrm{P}(X(t+s) = j | X(s) = i)$ does not depend on s for all $i, j \in \mathcal{X}$, then it is denoted as

$$P_{ij}(t) = \mathrm{P}(X(t+s) = j | X(s) = i),$$

and we say the CTMC is *homogeneous* in time.

In STAT253/317, we focus on homogeneous CTMC only.

Exponential Waiting/Holding Time

Let $\{X(t), t \geq 0\}$ be a homogeneous continuous-time Markov chain. For $i \in \mathcal{X}$, let T_i denote the amount of time that $X(t)$ stays in state i before making a transition into a different state.

Claim: T_i has the *memoryless property*.

$$
\begin{aligned}
&\mathrm{P}(T_i \geq t + s | T_i \geq s) \\
&= \mathrm{P}(X(u) = i, \text{ for } s \leq u \leq s + t | X(u) = i, \text{ for } 0 \leq u \leq s) \\
&= \mathrm{P}(X(u) = i, \text{ for } s \leq u \leq s + t | X(s) = i) \quad \text{(Markov property)} \\
&= \mathrm{P}(X(u) = i, \text{ for } 0 \leq u \leq t | X(0) = i) \qquad \text{(time homogeneity)} \\
&= \mathrm{P}(T_i \geq t) \quad \Rightarrow \quad \text{So } T_i \text{ is memoryless.}
\end{aligned}
$$

Recall that the exponential distribution is the only continuous distribution having the memoryless property.
Thus $T_i \sim Exp(\nu_i)$ for some rate ν_i.

An Alternative Definition of CTMC

A stochastic process $\{X(t), t \geq 0\}$ with state space $\mathcal{X}$ is a *continuous-time Markov chain* if

- (exponential waiting time) when the chain reaches a state i, the time it stays at state $i \sim Exp(\nu_i)$, where ν_i is the transition rate at state i
- (embedded with a discrete time Markov chain) when the process leaves state i, it enters anther state j with probability P_{ij}, such that

$$P_{ii} = 0, \quad \sum_{j \in \mathcal{X}} P_{ij} = 1 \quad \text{for all } i, j \in \mathcal{X}.$$

Remark: The amount of time T_i the process spends in state i, and the next state visited, must be independent. For if the next state visited were dependent on T_i, then information as to how long the process has already been in state i would be relevant to the prediction of the next state—and this contradicts the Markovian assumption.

Example: Poisson Process

For a Poisson process with parameter λ, the holding-time parameters are constant:

$$q_i = \lambda, \qquad i = 0, 1, 2, \ldots$$

The process moves sequentially

$$0 \to 1 \to 2 \to \cdots$$

The transition matrix of the embedded chain is

$$\tilde{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

6.3 Birth and Death Processes

Let $X(t) =$ the number of people in the system at time t.
Suppose that whenever there are n people in the system, then

(i) new arrivals enter the system at an exponential rate λ_n, and

(ii) people leave the system at an exponential rate μ_n.

Such an $\{X(t), t \geq 0\}$ is called a *birth and death process*.

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3 \;\; \begin{matrix} \cdots \\ \cdots \\ \cdots \end{matrix} \;\; n-1 \underset{\mu_n}{\overset{\lambda_{n-1}}{\rightleftarrows}} n \underset{\mu_{n+1}}{\overset{\lambda_n}{\rightleftarrows}} n+1 \;\; \begin{matrix} \cdots \\ \cdots \\ \cdots \end{matrix}$$

Suppose the process is at state $i > 0$ at time t. Then

$$B_i = \text{waiting time until the next birth} \sim \text{Exp}(\lambda_i)$$
$$D_i = \text{waiting time until the next death} \sim \text{Exp}(\mu_i)$$

Hence, the waiting time until the next transition out of state i is $\min(B_i, D_i) \sim \text{Exp}(\lambda_i + \mu_i)$, from which we can get

$$\nu_i = \lambda_i + \mu_i, \text{ for } i > 0$$

6.3 Birth and Death Processes (Cont'd)

Moreover, given the process is at state $i > 0$ at time t, the probability that the next transition is a birth rather than a death is

$$P_{i,i+1} = P(B_i < D_i) = \frac{\lambda_i}{\lambda_i + \mu_i},$$

which implies $P_{i,i-1} = P(D_i < B_i) = \frac{\mu_i}{\lambda_i+\mu_i}$, for $i > 0$.

As only birth is possible at state 0, we know $\nu_0 = \lambda_0$ and $P_{01} = 1$.

To sum up, a birth and death process is a CTMC with state space $\mathcal{X} = \{0, 1, 2, \ldots\}$ such that

$$\nu_i = \lambda_i + \mu_i, i > 0, \quad \nu_0 = \lambda_0,$$

$$P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i},\ P_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}, i > 0$$

$$P_{01} = 1,\ P_{i,j} = 0 \quad \text{if } |i - j| > 1$$

The parameters $\{\lambda_n\}_{n=0}^{\infty}$ and $\{\mu_n\}_{n=1}^{\infty}$ are called, respectively, the arrival (or birth) and departure (or death) rates.

Examples of Birth and Death Processes

- Poisson Processes: $\mu_n = 0$, $\lambda_n = \lambda$ for all $n \geq 0$
- Pure Birth Process:

$$\mu_n = 0 \quad \Rightarrow \quad \nu_i = \lambda_i,\ P_{i,i+1} = 1,\ P_{i,i-1} = 0$$

- Yule Processes (Pure Birth Process with Linear Growth rate): If there are n people and each independently gives birth at at an exponential rate λ, then the total rate at which births occur is $n\lambda$.

$$\mu_n = 0, \quad \lambda_n = n\lambda$$

Examples of Birth and Death Processes

- Poisson Processes: $\mu_n = 0$, $\lambda_n = \lambda$ for all $n \geq 0$
- Pure Birth Process:

$$\mu_n = 0 \quad \Rightarrow \quad \nu_i = \lambda_i,\ P_{i,i+1} = 1,\ P_{i,i-1} = 0$$

- Yule Processes (Pure Birth Process with Linear Growth rate): If there are n people and each independently gives birth at at an exponential rate λ, then the total rate at which births occur is $n\lambda$.

$$\mu_n = 0, \quad \lambda_n = n\lambda$$

Reason: Let

$B_i =$ time until the ith individual give birth $\sim Exp(\lambda),\ i = 1, \ldots, n$

So the time until the next (first) birth when there are n individuals in the population is

$$\min(B_1, B_2, \ldots, B_n) \sim Exp(\lambda + \lambda + \cdots + \lambda) = Exp(n\lambda)$$

So the rate until the next birth is $\lambda_n = n\lambda$.

Example: Linear Growth Model with Immigration

- each individual independently gives birth at an exponential rate λ
- each individual independently die at at an exponential rate μ
- new immigrants come in at an exponential rate θ

Such a process is a birth-death process with birth and death rates

$$\mu_n = n\mu, \quad \lambda_n = n\lambda + \theta$$

Reason: Let

$$B_i = \text{time until the ith individual give birth} \sim Exp(\lambda),\ i = 1, \dots, n$$
$$T = \text{time until the next new immigrant comes in} \sim Exp(\theta)$$

So the time until the population size increase from n to $n+1$ is

$$\min(B_1, \dots, B_n, T) \sim Exp(\lambda + \cdots + \lambda + \theta) = Exp(n\lambda + \theta)$$

So the rate until the next birth is $\lambda_n = n\lambda + \theta$.

Similarly, one can show that the death rate is $\mu_n = n\mu$.

Example: $M/M/s$ Queueing Model

- s servers
- Poisson arrival of customers, rate $= \lambda$
- Exponential service time, rate $= \mu$

$\Rightarrow$ a birth and death process with constant birth rate $\lambda_n = \lambda$, and death (departure)rate $\mu_n = \min(n, s)\mu$.

Reason: Suppose, there are n customer in the system at time t. At most $\min(n, s)$ of them are being served. Let S_i be remaining service time of the ith server $\sim$ Exp(μ). Then, the waiting time until the next departure is

$$\min(S_1, \ldots, S_{\min(s,n)}) \sim \text{Exp}(\min(s, n)\mu).$$

6.4 The Transition Probability Function $P_{ij}(t)$

Recall the transition probability function $P_{ij}(t)$ of a CTMC $\{X(t), t \geq 0\}$ is

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

Example. (Poisson Processes with rate λ)

$$\begin{aligned} P_{ij}(t) &= \mathrm{P}(N(t+s) = j | N(s) = i) \\ &= \mathrm{P}(N(t+s) - N(s) = j - i) = \begin{cases} e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} & \text{if } j \geq i \\ 0 & \text{if } j < i \end{cases} \end{aligned}$$

Properties of Transition Probability Functions

- $P_{ij}(t) \geq 0$ for all $i, j \in \mathcal{X}$ and $t \geq 0$
- (Row sums are 1) $\sum_j P_{ij}(t) = 1$ for all $i \in \mathcal{X}$ and $t \geq 0$

Lemma 6.3 Chapman-Kolmogorov Equation

For all $i, j \in \mathcal{X}$ and $t \geq 0$,

$$P_{ij}(t+s) = \sum_{k \in \mathcal{X}} P_{ik}(t) P_{kj}(s)$$

Proof.

$$\begin{aligned}
&P_{ij}(t+s) \\
&= \mathrm{P}(X(t+s) = j | X(0) = i) \\
&= \sum_{k \in \mathcal{X}} \mathrm{P}(X(t+s) = j, X(t) = k | X(0) = i) \\
&= \sum_{k \in \mathcal{X}} \mathrm{P}(X(t+s) = j | X(t) = k, X(0) = i) \mathrm{P}(X(t) = k | X(0) = i) \\
&= \sum_{k \in \mathcal{X}} \mathrm{P}(X(t+s) = j | X(t) = k) \mathrm{P}(X(t) = k | X(0) = i) \text{ (Markov Property)} \\
&= \sum_{k \in \mathcal{X}} P_{kj}(s) P_{ik}(t)
\end{aligned}$$

The matrix notation

Let $\mathbf{P}(t) = [P_{ij}(t)]$ be the transition matrix at time t.
We have $\mathbf{P}(0) = \mathbf{I}$. And C-K equations read

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$$

One way to specify a CTMC is through $\{\mathbf{P}(t)\}_{t\geq 0}$. But this requires an infinite number of matrices. Can we simplify it?

Key: use derivatives $\mathbf{P}'(t)$

Transition rate matrix / infinitesimal generator Q

Assume that

$$\mathbf{P}'(0) = \lim_{h \to 0} \frac{\mathbf{P}(h) - \mathbf{P}(0)}{h} \quad \text{exists.}$$

In other words, for each i, j,

$$P'_{ij}(0) = \lim_{h \to 0} \frac{P_{ij}(h) - P_{ij}(0)}{h} \quad \text{exists.}$$

We will denote such limit as $\mathbf{Q} = [q_{ij}]$, the transition rate matrix.
What does this tell us about $\mathsf{P}'(t)$ for $t > 0$?

Kolmogorov's equations

By definition, one has

$$\mathsf{P}'(t) = \lim_{h\to 0} \frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \lim_{h\to 0} \frac{\mathbf{P}(t)\mathbf{P}(h) - \mathbf{P}(t)}{h}$$
$$= \lim_{h\to 0} \frac{\mathbf{P}(t)(\mathbf{P}(h) - \mathbf{I})}{h} = \mathbf{P}(t)\mathbf{Q}.$$

This is the so-called Kolmogorov's forward equations.

Similarly you can prove backward equations

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t).$$

These imply $\mathbf{P}(t) = \exp(t\mathbf{Q})$.

Transition rate matrix

How to compute $\mathbf{Q}$?

Lemma 6.2a

For any $i, j \in \mathcal{X}$, we have

$$q_{ii} = \lim_{h \to 0} \frac{P_{ii}(h) - 1}{h} = -\nu_i$$

Proof. Let T_i be the amount of time the process stays in state i before moving to other states.

$$\begin{aligned}
P_{ii}(h) &= \mathrm{P}(X(h) = i | X(0) = i) \\
&= \mathrm{P}(X(h) = i, \text{no transition in } (0,\text{h}] | X(0) = i) \\
&\quad + \mathrm{P}(X(h) = i, 2 \text{ or more transition in } (0,\text{h}] | X(0) = i) \\
&= \mathrm{P}(T_i > h) + o(h) \\
&= e^{-\nu_i h} + o(h) \\
&= 1 - \nu_i h + o(h)
\end{aligned}$$

Lemma 6.2b

For any $i \neq j \in \mathcal{X}$, we have

$$q_{ij} = \lim_{h \to 0} \frac{P_{ij}(h)}{h} = \nu_i P_{ij}$$

Proof.

$$\begin{aligned}
P_{ij}(h) &= \mathrm{P}(X(h) = j | X(0) = i) \\
&= \mathrm{P}(X(h) = j, \text{1 transition in (0,h]} | X(0) = i) \\
&\quad + \mathrm{P}(X(h) = j, \text{2 or more transition in (0,h]} | X(0) = i) \\
&= \mathrm{P}(T_i < h) P_{ij} + o(h) \\
&= (1 - e^{-\nu_i h}) P_{ij} + o(h) \\
&= \nu_i P_{ij} h + o(h)
\end{aligned}$$

For finite state space case $\mathcal{X} = \{1, 2, \ldots, m\}$, define the matrices

$$\mathbf{P}(t) = \begin{bmatrix} P_{11}(t) & \cdots & P_{1m}(t) \\ \vdots & & \vdots \\ P_{m1}(t) & \cdots & P_{mm}(t) \end{bmatrix}, \quad \mathbf{P}'(t) = \begin{bmatrix} P'_{11}(t) & \cdots & P'_{1m}(t) \\ \vdots & & \vdots \\ P'_{m1}(t) & \cdots & P'_{mm}(t) \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} q_{11} & \cdots & q_{1m} \\ \vdots & & \vdots \\ q_{m1} & \cdots & q_{mm} \end{bmatrix} = \begin{bmatrix} -\nu_1 & \nu_1 P_{12} & \cdots & \nu_1 P_{1m} \\ \nu_2 P_{21} & -\nu_2 & \cdots & \nu_2 P_{2m} \\ \vdots & \vdots & & \vdots \\ \nu_m P_{m1} & \nu_m P_{m2} & \cdots & -\nu_m \end{bmatrix}$$

In matrix notation,
Forward equation: $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$
Backward equation: $\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t)$

Alarm Clocks and Transition Rates

Idea. A homogeneous CTMC can be specified by transition rates

$$q_{ij} \geq 0, \qquad i \neq j.$$

Fix a state $i \in \mathcal{X}$. For each state j that can be reached from i:

- attach an independent alarm clock (i, j) that rings after

$$T_{ij} \sim \mathrm{Exp}(q_{ij});$$

- when the process hits i, all clocks $\{T_{ij}\}_{j \neq i}$ start simultaneously;
- the first clock to ring determines the next state visited.

If the (i, j) clock rings first, then the process jumps to j, and a fresh set of clocks $\{T_{jk}\}_{k \neq j}$ starts.

Holding Time Parameter q_i

Start in state i and start all clocks $\{T_{ij}\}_{j\neq i}$.

The holding time (time spent in state i) is the first alarm time:

$$T_i \;=\; \min_{k\neq i} T_{ik}, \qquad T_{ik} \sim \mathrm{Exp}(q_{ik}) \text{ independent.}$$

By the minimum-of-exponentials fact,

$$T_i \sim \mathrm{Exp}\Big(\sum\nolimits_{k\neq i} q_{ik}\Big).$$

Definition (total leaving rate).

$$q_i \;:=\; \sum\nolimits_{k\neq i} q_{ik}.$$

Interpretation

The rate at which the process leaves state i equals the sum of the rates to each possible next state.

Embedded Discrete-Time Chain: P_{ij}

Given the process is in state i, the next jump is to j iff clock (i, j) rings first.

Using the competing exponentials property,

$$P_{ij} \;:=\; \mathrm{P}(\text{next state is } j \mid X(0) = i) \;=\; \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} \;=\; \frac{q_{ij}}{q_i}, \qquad (j \neq i).$$

So a CTMC can be decomposed as:

- holding time at i: $\quad T_i \sim \mathrm{Exp}(q_i)$,
- next-state choice: $\quad \mathrm{P}(i \to j) = P_{ij} = q_{ij}/q_i$.

Consistency check. $\sum_{j \neq i} P_{ij} = 1$ and $P_{ii} = 0$.

Limiting Distribution

Definition. A probability distribution π on $\mathcal{X}$ is called the limiting distribution of a homogeneous CTMC if for all $i, j \in \mathcal{X}$,

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

- The limit does not depend on the initial state i.
- If such limits exist for all $j \in \mathcal{X}$, then $(\pi_j)_{j\in\mathcal{X}}$ is a distribution:

$$\pi_j \geq 0, \qquad \sum\nolimits_{j\in\mathcal{X}} \pi_j = 1.$$

Remark. If $\lim_{t\to\infty} P_{ij}(t)$ exists, we must have

$$\lim_{t\to\infty} P'_{ij}(t) = 0.$$

Stationary Distribution

Definition. A probability distribution π on $\mathcal{X}$ is called a stationary distribution if

$$\pi = \pi \mathbf{P}(t), \qquad t \geq 0.$$

Equivalently, for all states $j \in \mathcal{X}$,

$$\pi_j = \sum_{i \in \mathcal{X}} \pi_i P_{ij}(t), \qquad t \geq 0.$$

Interpretation

If the initial distribution is π, then the distribution of $X(t)$ remains π for all time.

Balanced equations

Recall the forward equations $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$

If you set $t \to \infty$, you have

$$0 = p^\top \mathbf{Q},$$

where $p = (P_1, P_2, \ldots)^\top$
This is the same as saying that

$$\nu_j P_j = \sum_{k \in \mathcal{X}, k \neq j} P_k q_{kj} \quad \text{for all } j \in \mathcal{X}$$

Interpretation of the Balanced Equations

$$\nu_j P_j = \sum_{k \in \mathcal{X}, k \neq j} P_k q_{kj}$$

$$\nu_j P_j = \text{rate at which the process } \textbf{leaves} \text{ state } j$$

$$\sum_{k \in \mathcal{X}, k \neq j} P_k q_{kj} = \text{rate at which the process } \textbf{enters} \text{ state } j$$

Balanced equations means that the rates at which the process enters and leaves state j are equal.

The limiting distribution $\{P_j\}_{j \in \mathcal{X}}$ can be obtained by solving the balanced equations along with the equation $\sum_{j \in \mathcal{X}} P_j = 1$.

Remarks. Just like discrete-time Markov chains, a sufficient condition for the existence of a limiting distribution is that the chain is irreducible and positive recurrent.

Examples

- **Poisson processes**: $\mu_n = 0$, $\lambda_n = \lambda$ for all $n \geq 0$

$$\nu_i = \lambda,\ P_{i,i+1} = 1, \quad q_{i,i+1} = \nu_i P_{i,i+1} = \lambda$$

  Balanced equations:

$$\nu_j P_j = P_{j-1} q_{j-1,j} \quad \Rightarrow \quad \lambda P_j = \lambda P_{j-1} \quad \Rightarrow \quad P_j = P_{j-1}$$

  No limiting distribution exists. The chain is not irreducible. All states are transient.

- **Pure birth processes with $\lambda_n > 0$ for all n**
  No limiting distribution exists. All states are transient.

Birth and Death Processes

For a birth and death process,

$$\begin{aligned} \nu_0 &= \lambda_0, \\ \nu_i &= \lambda_i + \mu_i,\ i > 0 \\ P_{01} &= 1, \\ P_{i,i+1} &= \tfrac{\lambda_i}{\lambda_i+\mu_i}, \quad\ i > 0 \\ P_{i,i-1} &= \tfrac{\mu_i}{\lambda_i+\mu_i}, \quad\ i > 0 \\ P_{i,j} &= 0 \qquad\qquad \text{if } |i-j| > 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} q_{i,i+1} = \nu_i P_{i,i+1} = \lambda_i,\ i \geq 0 \\ q_{i,i-1} = \nu_i P_{i,i-1} = \mu_i, i \geq 1 \end{aligned}$$

Balanced Equations for Birth and Death Processes

The balanced equations $\nu_j P_j = \sum_{k \in \mathcal{X}, k \neq j} P_k q_{kj}$ for a birth and death process are

$$\begin{aligned}
\lambda_0 P_0 &= \mu_1 P_1 \\
(\mu_1 + \lambda_1) P_1 &= \lambda_0 P_0 + \mu_2 P_2, \\
(\mu_2 + \lambda_2) P_2 &= \lambda_1 P_1 + \mu_3 P_3, \\
&\vdots \\
(\mu_{n-1} + \lambda_{n-1}) P_{n-1} &= \lambda_{n-2} P_{n-2} + \mu_n P_n \\
(\mu_n + \lambda_n) P_n &= \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1}
\end{aligned}$$

Adding up all the equations above and eliminating the common terms on both sides, we get

$$\lambda_n P_n = \mu_{n+1} P_{n+1} \quad n \geq 0,$$

We hence just need to solve $\lambda_n P_n = \mu_{n+1} P_{n+1}$ for the limiting distribution.

6.6. Time Reversibility

Definition. A continuous-time Markov chain with state space $\mathcal{X}$ is *time reversible* if

$$P_i q_{ij} = P_j q_{ji}, \quad \text{for all } i, j \in \mathcal{X} \quad \text{(detailed balanced equation)}$$

If a distribution $\{P_j\}$ on $\mathcal{X}$ satisfies the detailed balanced equation, then it is a stationary distribution for the process.

Example. We have just shown that for Birth and Death processes, the balanced equations would lead to the detailed balanced equations, which are

$$\lambda_n P_n = \mu_{n+1} P_{n+1}, \quad n \geq 0$$

Limiting Dist'n for Birth and Death Processes

Solving $\lambda_n P_n = \mu_{n+1} P_{n+1}$, $n \geq 0$ for the limiting distribution, we get

$$P_n = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1}}{\mu_n} \frac{\lambda_{n-2}}{\mu_{n-1}} P_{n-2} = \ldots = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1} P_0$$

To meet the requirement $\sum_{n=0}^{\infty} P_n = 1$, we need

$$\sum_{n=0}^{\infty} P_n = P_0 + P_0 \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1} = 1$$

In other words, to have a limiting distribution, it is necessary that

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

Limiting Dist'n for Birth and Death Processes (Cont'd)

If $\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}$ is finite, the limiting distribution is

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}$$

and

$$P_k = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1} / (\mu_1 \mu_2 \cdots \mu_k)}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}, \quad k \geq 1$$

Lemma: (Ratio Test) If $a_n \geq 0$ for all n, then

$$\sum_{n=1}^{\infty} a_n \begin{cases} < \infty & \text{if } \lim_{n\to\infty} a_n/a_{n-1} < 1 \\ = \infty & \text{if } \lim_{n\to\infty} a_n/a_{n-1} > 1 \end{cases}$$

For $a_n = \dfrac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}$, $a_n/a_{n-1} = \lambda_{n-1}/\mu_n$. By the ratio test, if

$$\lim_{n\to\infty} \frac{\lambda_{n-1}}{\mu_n} < 1,$$

then $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty$, the limiting distribution exists.

Example 6.4 Linear Growth Model with Immigration

$$\mu_n = n\mu, \quad \lambda_n = n\lambda + \theta$$

Using the Ratio Test,

$$\lim_{n\to\infty} \frac{\lambda_{n-1}}{\mu_n} = \lim_{n\to\infty} \frac{(n-1)\lambda + \theta}{n\mu} = \frac{\lambda}{\mu}$$

So the linear growth model has a limiting distribution if $\lambda < \mu$.

Example 5.5 (M/M/1 Queueing w/ Finite Capacity)

- single-server service station. Service times are i.i.d. $\sim Exp(\mu)$
- Poisson arrival of customers with rate λ
- Upon arrival, a customer would
  - go into service if the server is free (queue length $= 0$)
  - join the queue if 1 to $N-1$ customers in the station, or
  - walk away if N or more customers in the station

Q: What fraction of potential customers are lost?

Example 5.5 (M/M/1 Queueing w/ Finite Capacity)

- single-server service station. Service times are i.i.d. $\sim Exp(\mu)$
- Poisson arrival of customers with rate λ
- Upon arrival, a customer would
  - go into service if the server is free (queue length $= 0$)
  - join the queue if 1 to $N-1$ customers in the station, or
  - walk away if N or more customers in the station

Q: What fraction of potential customers are lost?

Let $X(t)$ be the number of customers in the station at time t.

$\{X(t),\ t \geq 0\}$ is a birth-death process with the birth and death rates below

$$\mu_n = \begin{cases} 0 & \text{if } n = 0 \\ \mu & \text{if } n \geq 1 \end{cases} \quad \text{and} \quad \lambda_n = \begin{cases} \lambda & \text{if } 0 \leq n < N \\ 0 & \text{if } n \geq N \end{cases}$$

Example 5.5 (M/M/1 Queueing w/ Finite Capacity)

Solving $\lambda_n P_n = \mu_{n+1} P_{n+1}$ for the limiting distribution

$$\begin{aligned}
P_1 &= (\lambda/\mu) P_0 \\
P_2 &= (\lambda/\mu) P_1 = (\lambda/\mu)^2 P_0 \\
&\vdots \\
P_i &= (\lambda/\mu)^i P_0, \qquad\qquad i = 1, 2, \ldots, N
\end{aligned}$$

Plugging $P_i = (\lambda/\mu)^i P_0$ into $\sum_{i=0}^N P_i = 1$, one can solve for P_0 and get

$$P_i = \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{N+1}} (\lambda/\mu)^i$$

Answer: The fraction of customers lost is $P_N = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}} (\lambda/\mu)^N$

Duration Times for Birth and Death Processes

Let

$$T_i = \text{time to move from state } i \text{ to state } i+1, \quad i = 0, 1, \ldots.$$

Suppose at some moment $X(t) = i$. Let

$$B_i = \text{time until the next birth} \sim Exp(\lambda_i)$$
$$D_i = \text{time until the next death} \sim Exp(\mu_i)$$

Then

$$T_i = \begin{cases} B_i & \text{if the next step is } i \to i+1, \text{ i.e., } B_i < D_i \\ D_i + T_{i-1} + T_i^* & \text{if the next step is } i \to i-1, \text{ i.e., } D_i < B_i \end{cases}$$
$$= \min(B_i, D_i) + \begin{cases} 0 & \text{if } B_i < D_i \text{, occur w/ prob. } \frac{\lambda_i}{\lambda_i+\mu_i} \\ T_{i-1} + T_i^* & \text{if } D_i < B_i \text{, occur w/ prob. } \frac{\mu_i}{\lambda_i+\mu_i} \end{cases}$$

Note

- T_i^* has the same distribution as T_i
- T_{i-1} and T_i^* are indep. of B_i and D_i because it's Markov

Duration Times for Birth and Death Processes

Taking expected value of

$$T_i = \min(B_i, D_i) + \begin{cases} 0 & \text{if } B_i < D_i\text{, occur w/ prob. } \frac{\lambda_i}{\lambda_i+\mu_i} \\ T_{i-1} + T_i^* & \text{if } D_i < B_i\text{, occur w/ prob. } \frac{\mu_i}{\lambda_i+\mu_i} \end{cases}$$

we get

$$\begin{aligned} \mathbb{E}[T_i] &= \mathbb{E}[\min(B_i, D_i)] + (\mathbb{E}[T_{i-1}] + \mathbb{E}[T_i]) \frac{\mu_i}{\lambda_i + \mu_i} \\ &= \frac{1}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i}(\mathbb{E}[T_{i-1}] + \mathbb{E}[T_i]) \end{aligned}$$

We obtain the recursive formula

$$\lambda_i \mathbb{E}[T_i] = 1 + \mu_i \mathbb{E}[T_{i-1}]$$

Duration Times for Birth and Death Processes (Cont'd)

Since $T_0 \sim Exp(\lambda_0)$, $\mathbb{E}[T_0] = 1/\lambda_0$.
Using the recursive formula $\lambda_i \mathbb{E}[T_i] = 1 + \mu_i \mathbb{E}[T_{i-1}]$, we have

$$
\begin{aligned}
\mathbb{E}[T_0] &= \frac{1}{\lambda_0} \\
\mathbb{E}[T_1] &= \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_1}\mathbb{E}[T_0] = \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_1\lambda_0} \\
\mathbb{E}[T_2] &= \frac{1}{\lambda_2} + \frac{\mu_2}{\lambda_2}\left(\frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_1\lambda_0}\right) = \frac{1}{\lambda_2} + \frac{\mu_2}{\lambda_2\lambda_1} + \frac{\mu_2\mu_1}{\lambda_2\lambda_1\lambda_0} \\
&\vdots \\
\mathbb{E}[T_i] &= \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}\mathbb{E}[T_{i-1}] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i\lambda_{i-1}} + \cdots + \frac{\mu_i\mu_{i-1}\cdots\mu_2\mu_1}{\lambda_i\lambda_{i-1}\cdots\lambda_2\lambda_1\lambda_0} \\
&= \frac{1}{\lambda_i}\left(1 + \sum_{k=1}^{i}\prod_{j=1}^{k}\frac{\mu_{i-j+1}}{\lambda_{i-j}}\right)
\end{aligned}
$$