Module 6

Continuous-time Markov chain

Topics: Definitions and equivalence, Kolmogorov equations, Long-run behavior

Definitions and equivalence

Markov property in continuous-time: A discrete-state stochastic process $\{X_n: n = 0,1,2, \dots\}$ is a Markov chain if $P(X_{n+1} = j | X_n = i, X_{n-1} = i_1, X_{n-2} = i_2, \dots, X_0 = i_n) = P(X_{n+1} = j | \mathrm{X_n} = i)$ for all $j, i, i_1, i_2, \dots, i_n$ and n. Here, time is measured in discrete steps, and sometimes, the time gaps between successive steps can be different. So, time is not captured in its true sense in Markov chain. Continuous-time Markov chain (CTMC) bridges this gap.

CTMCs are continuous-time discrete-state stochastic processes having the Markov property, i.e., the latest known state alone matters in determining future. Formally, $\{X(t): t \geq 0\}$ with non-negative integer-valued $X(t)$ is a CTMC if $P(X(t) = j | X(s) = i, X(s_1) = i_1, X(s_2) = i_2, \dots) = P(X(t) = j | X(s) = i)$ for all $t > s > s_1 > s_2 > \cdots$ and $j, i, i_1, i_2, \dots$. Note that the state space has been restricted to non-negative integers. This is not really a restriction, as we can always establish 1-1 correspondence between arbitrary discrete states and non-negative integer. Working with the later facilitates the discussion.

In the discrete-time case, we need to specify $p_{ij}^{n,n+1} = P(X_{\mathrm{n+1}} = j | X_{\mathrm{n}} = i)$ for all $i, j \in \Omega$ and $n = 0,1,2, \dots$. However, with stationarity of the transition probabilities, specifying $p_{ij} = P(X_1 = j | X_0 = i)$ for all $i, j \in \Omega$ alone is sufficient. For CTMC, in the general case, we need to specify $p_{ij}(\mathrm{s,t}) \coloneqq P(X(\mathrm{t}) = j | X(\mathrm{s}) = i)$ for all $i, j \in \Omega$ and $\mathrm{s} < \mathrm{t}$. With stationarity of the transition probabilities, specifying $p_{ij}(t) \coloneqq P(X(t) = j | X(0) = i)$ for all $i, j \in \Omega$ and $t > 0$ is sufficient. With stationarity, $p_{ij}(\mathrm{s,t}) = P(X(t) = j | X(\mathrm{s}) = i) = P(X(t - s) = j | X(0) = i) = p_{ij}(\mathrm{t - s})$ for all $i, j \in \Omega$ and $s < \mathrm{t}$. Unlike the discrete-time case where 1 is the least denomination of time, we do not have such a thing in CTMCs. That is why, we must specify $p_{ij}(t)$ for all $t > 0$. Let us consider some examples.

A trivial example of a CTMC is the Poisson process. For $t > s > s_1 > s_2 > \cdots$ and $j \geq i \geq i_1 \geq i_2 \geq \cdots$ (as we are talking about a counting process),

$$\begin{aligned}
P(X(t) = j | X(s) = i, X(s_1) = i_1, X(s_2) = i_2, \dots) \\
&= P(X(t) - X(s) = j - i | X(s) = i, X(s_1) = i_1, X(s_2) = i_2, \dots) \\
&= P(X(t) - X(s) = j - i | X(s) = i), \text{by independent increment property} \\
&= P(X(t) = j | X(s) = i), \text{as required.}
\end{aligned}$$

Clearly, the Markov property holds. So, the Poisson process in a CTMC. Let us find its transition probabilities. For $s < \mathrm{t}$ and $i \leq j$,

$$\begin{aligned}
p_{ij}(s, \mathrm{t}) &= P(X(\mathrm{t}) = j | X(\mathrm{s}) = i) = P(X(\mathrm{t}) - X(\mathrm{s}) = j - i | X(\mathrm{s}) = i) \\
&= P(X(\mathrm{t}) - X(\mathrm{s}) = j - i), \text{by independent increment property}
\end{aligned}$$

$$= P(X(t-s) = j-i), \text{by stationary increment property}$$
$$= e^{-\lambda(t-s)} \frac{\{\lambda(t-s)\}^{j-i}}{(j-i)!},$$

and for $i > j$, $p_{ij}(s,\mathrm{t}) = 0$. Observe that the transition probabilities are stationary, because $p_{ij}(s,\mathrm{t}) = p_{ij}(t-s)$ for all $i, j \in \Omega$ and $s < \mathrm{t}$.

Due to independent increment property, non-homogeneous Poisson process, too, possesses Markov property. Thus, the non-homogeneous Poisson process is a CTMC. Let us find its transition probabilities. For $s < \mathrm{t}$ and $i \leq j$,

$$p_{ij}(s,\mathrm{t}) = P(X(\mathrm{t}) = j | X(\mathrm{s}) = i) = P(X(\mathrm{t}) - X(\mathrm{s}) = j - i | X(\mathrm{s}) = i)$$
$$= P(X(\mathrm{t}) - X(\mathrm{s}) = j-i), \text{by independent increment property}$$
$$= e^{-\int_s^t \lambda(\tau)d\tau} \frac{\left\{\int_s^t \lambda(\tau)d\tau\right\}^{j-i}}{(j-i)!},$$

and for $i > j$, $p_{ij}(s,\mathrm{t}) = 0$. Here, transition probabilities are generally not stationary because $p_{ij}(t-s) = P(X(\mathrm{t}-\mathrm{s}) = j-i) = e^{-\int_0^{t-s}\lambda(\tau)d\tau}\left(\int_0^{t-s}\lambda(\tau)d\tau\right)^{j-i}/(j-i)!$ is different from $p_{ij}(s,\mathrm{t})$, except when $\lambda(\tau)$ is a constant function.

Let us consider a renewal process with $U(5,10)$ inter-event times. Here, $P(N(11) = 2| N(10) = 1, N(9) = 0) = 0$, as it requires $T_2 < 2$, which is impossible. However, $P(N(11) = 2|N(10) = 1, N(9) = 1) > 0$, as $\{N(11) = 2|N(10) = 1, N(9) = 1\} \equiv \{T_1 \leq 6, T_2 \leq 11 - T_1\}$ is possible. The knowledge of $N(9)$ is not irrelevant in the presence of $N(10)$, which goes against the Markov property. Thus, renewal process is generally not a CTMC.

In this course, we will consider CTMCs with stationary transition probabilities. Such CTMCs will be specified by $p_{ij}(t)$ for all $i, j \in \Omega$ and all $t > 0$. We will see that obtaining $p_{ij}(t)$ is no easy task, except for some trivial CTMCs such as the Poisson process. There is a simpler way. We can specify a CTMC having stationary transition probabilities by $p_{ij}(h)$ for all $i, j \in \Omega$, where $p_{ij}(h)$ may involve $o(h)$ terms and h is an arbitrarily small positive quantity. We used this approach in the 3rd condition in the definition of Poisson process. Note the similarity between $p_{ij}(h)$ for CTMC and p_{ij} for the discrete-time case. In CTMC, $h \to 0^+$ assumes the role of the minimum denomination of time. Later, we will develop Kolmogorov equations for calculating $p_{ij}(t)$ for $t > 0$ using $p_{ij}(h)$. In the discrete-time case, we obtained $p_{ij}^{(n)}$ through the Chapman-Kolmogorov equations using p_{ij}. Let us specify the Poisson process, which is a CTMC having stationary transition probabilities, using $p_{ij}(h)$.

$$p_{ij}(h) = \begin{cases} 0 \ \text{ if } j < i \\ P(N(h) = 0) = 1 - \lambda h + o(h) \ \text{ if } j = i \\ P(N(h) = 1) = \lambda h + o(h) \ \text{ if } j = i+1 \\ o(h) \ \text{ if } j \geq i+2 \end{cases}$$

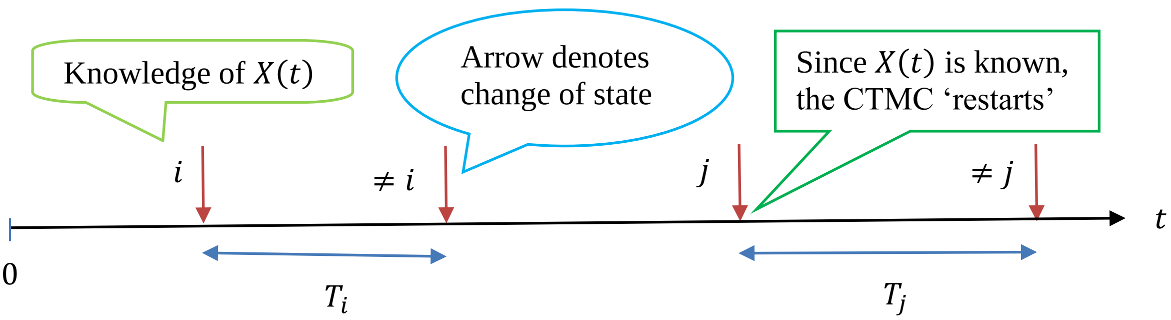
Holding times: So far, we have considered only one CTMC, i.e., the Poisson process. This is because verifying whether a continuous-time discrete-state stochastic process possesses the Markov property or not is a difficult task. Here, we develop an alternate characterization of CTMC that will allow us to easily check if a process is CTMC.

Let $T_i \coloneqq \min\{t: X(t) \neq i | X(0) = i\}$ denote the time CTMC spends in state-i (before going to some other state) after starting in state-i. We refer to T_i as the holding time of state-i. Note that whenever the CTMC reaches state-i, it spends T_i amount of time before going to some other state, due to Markov property and stationary transition probabilities. For $u, v > 0$,

$$
\begin{aligned}
P(T_i > u + v | T_i > u) &= P(X(t) = i \; \forall t \le u + v | X(s) = i \; \forall s \le u) \\
&= P\left(\underbrace{X(t) = i \; \forall t \in (u, u+v]}_{A}, \underbrace{X(s) = i \; \forall s \in [0, u]}_{B} \,|\, \underbrace{X(s) = i \; \forall s \in [0, u]}_{B}\right) \\
&= P(AB|B) = P(AB)/P(B) = P(A|B) \\
&= P(X(t) = i \; \forall t \in (u, u+v] | X(s) = i \; \forall s \in [0, u]) \\
&= P(X(t) = i \; \forall t \in (u, u+v] | X(u) = i), \text{by Markov property} \\
&= P(X(t) = i \; \forall t \in (0, v] | X(0) = i), \text{by stationary transition probabilities} \\
&= P(T_i > v)
\end{aligned}
$$

T_i is memoryless, and there is only one continuous random variable having the memoryless property. So, T_i is exponentially distributed, and this is true for all $i \in \Omega$. Hence, the holding times in CTMCs with stationary transition probabilities are exponential random variables. Following the above derivation, we can see that CTMSs having non-stationary transition probabilities do not have memoryless (equivalently, exponential) holding times.

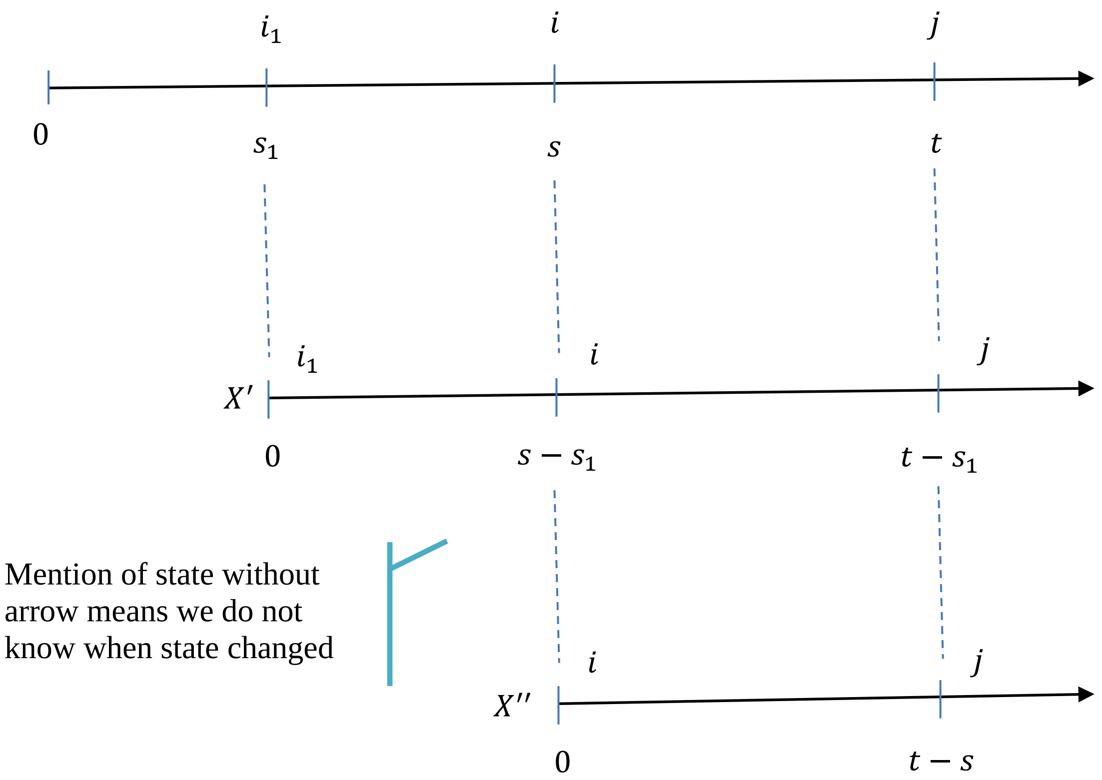
For $i \neq j$, T_i and T_j both are exponentially distributed. Let us examine the relation between these random variables. Due to the Markov property, from the following diagram, T_i and T_j are independent. So, holding times in different states of a CTMC are independent. If $j = i$, due to the Markov property, $T_i^{(1)}$ and $T_j = T_i^{(2)}$ are independent too. Thus, all holding times encountered in a CTMC, irrespective of the states, are independent.



Alternate characterization: At the end of the holding time in a state, say state-i, the CTMC moves to a different state. There is randomness in this transition, which we can capture by defining transition probabilities at the point of state change $p_{ij} \coloneqq P(X(T_i) = j)$ for all $j \neq i$. Due to the Markov property and stationary transition probabilities, these p_{ij}'s are fixed for given $i \neq j$, i.e., the same transition probabilities govern the CTMC at all times. If we write p_{ij}'s in the matrix form, like the transition probability matrix of a Markov chain, then the row sums are 1, because $\sum_{j \in \Omega} p_{ij} = \sum_{j \in \Omega} P(X(T_i) = j) = P\left(\cup_{j \in \Omega}\{X(T_i) = j\}\right) = P(X(T_i) \in \Omega)$ $= 1$. Also, the principal diagonal of the matrix consists of zeros, because $X(T_i) = i$ is an impossibility. The diagonal elements are the only differentiating factor between this matrix and the transition probability matrix of a Markov chain.

A continuous-time discrete-state stochastic process that has independent exponential holding times and fixed transition probabilities at the points of state change seems to satisfy Markov property, which says that the latest known state matters. Let us propose a second definition of CTMC: A continuous-time discrete-state stochastic process $\{X(t): t \geq 0\}$ is a CTMC with stationary transition probabilities if its holding times are independent exponential random variables and the transition probabilities at the points of state change are fixed quantities. We refer to this definition as D2, and the original one as D1.

In order to show validity of D2, we need to establish its equivalence with D1. D2 obviously follows from D1, as shown above. We need to show that D1 follows from D2, i.e., we need to show that D2 implies Markov property and stationarity of transition probabilities.



Consider a stochastic process defined through D2. For $t > s > s_1$, consider three observers: one observing the process from the beginning, another observing from time s_1, and the last one observing from time s, as shown in the above diagram. Due to the memoryless property and independence of the holding times and fixed transition probabilities at the points of state change, we can say that $P(X(t) = j | X(s) = i, X(s_1) = i_1) = P(X'(t - s_1) = j | X'(s - s_1) =$

$i, X'(0) = i_1) = P(X''(t-s) = j | X''(0) = i)$, which is same as $P(X(t) = j | X(s) = i)$ as well as $P(X(t-s) = j | X(0) = i)$ for all j, i, i_1.

Following the above procedure, we can show that $P(X(t) = j | X(s) = i, X(s_1) = i_1, X(s_2) = i_2, \ldots) = P(X(t) = j | X(s) = i) = P(X(t-s) = j | X(0) = i)$ for all $t > s > s_1 > s_2 > \cdots$ and $j, i, i_1, i_2, \ldots$. So, the process possesses Markov property, and the transition probabilities are stationary. Thus, D1 follows from D2, as required.

With the alternate characterization, let us consider the M/M/1 queue, where

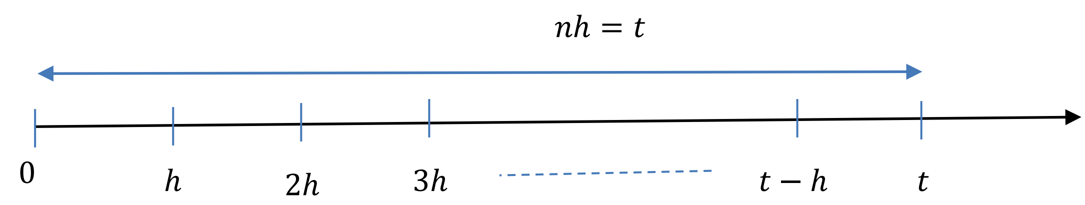
(a) customers arrive for availing service in accordance with a Poisson process with rate λ,
(b) there is a single server that takes $Exp(\mu)$ time for serving a customer,
(c) customers are served in the first-come first-served mode,
(d) a customer always waits if the server is busy.

Let $X(t)$ denote the total number of customers in the system, i.e., in the queue and in the server, at time-t. $\{X(t): t \geq 0\}$ a continuous-time non-negative integer-valued process. If the process is in state-i, say $i \geq 1$, then the state changes either when a new customer arrives or when the customer in-service complete its service, whichever is earlier. Due to memoryless property, time till these events are $Exp(\lambda)$ and $Exp(\mu)$ respectively. Considering these to be independent, holding time in state-i, $T_i = \min\big(Exp(\lambda), Exp(\mu)\big) \sim Exp(\lambda + \mu)$. If $i = 0$, then an arrival is the only way for state change. So, $T_0 \sim Exp(\lambda)$. Thus, the holding times are exponentially distributed. Due to memoryless property and independence of arrival process and service times, holding times are independent as well.

Considering the current state $i \geq 1$, the next state is $i+1$ if $Exp(\lambda) < Exp(\mu)$ and it is $i-1$ if $Exp(\mu) < Exp(\lambda)$. So, $p_{i,i+1} = \lambda/(\lambda+\mu)$ and $p_{i,i-1} = \mu/(\lambda+\mu)$ for $i \geq 1$. For $i = 0$, the next state is always 1, i.e., $p_{0,1} = 1$. Clearly, the transition probabilities at the points of state change are fixed quantities. Hence, $\{X(t): t \geq 0\}$ is a CTMC. It is not easy to show that the M/M/1 queue is a CTMC using the original definition.

<u>Conversion of specifications</u>: For a CTMC, consider that we have the knowledge of $p_{ij}(h)$ for all $i, j \in \Omega$. We want to determine θ_i for all $i \in \Omega$ and p_{ij} for all $i \neq j$, i.e., we want to convert one kind of specification into the other. Let us define $\lambda_i \coloneqq \lim_{h\to 0} \big(1 - p_{ii}(h)\big)/h$ for all $i \in \Omega$. Then we can write $p_{ii}(h) = 1 - \lambda_i h + o(h)$. With this, $\lim_{h\to 0} \big(1 - p_{ii}(h)\big)/h$ $= \lambda_i + \lim_{h\to 0} o(h)/h = \lambda_i$. Let us obtain $P(T_i > t) = e^{-\theta_i t}$ in terms of λ_i.

$P(T_i > t) = P(\text{No state change in } (0,t] | X(0) = i)$
$= P(\text{No state change in any of the } n \text{ subintervals in the following diagram} | X(0) = i\,)$
$= P(E_n E_{n-1} \cdots E_1 | X(0) = i)$, where $E_k = \{\text{No state change in } k^{th} \text{ subinterval}\}$
$= P(E_n | E_{n-1} \cdots E_1, X(0) = i) \cdot P(E_{n-1} | E_{n-2} \cdots E_1, X(0) = i) \cdots P(E_1 | X(0) = i)$
$= P^n(E_1 | X(0) = i)$, due to Markov property and stationarity



Now, we need to obtain $P(E_1|X(0)=i) = P(\text{No state change in } (0,h]|X(0)=i)$. Note that $p_{ii}(h) = P(X(h)=i|X(0)=i) \geq P(E_1|X(0)=i)$, because $p_{ii}(h)$ considers the possibility of state change, which $E_1|X(0)=i$ does not. However, if h is decreased, the chances of state change in $(0,h]$ becomes increasingly rare, and then the gap between these two quantities reduces. As $h \to 0$, the gap vanishes, i.e., $\lim_{h\to 0} P(E_1|X(0)=i) = \lim_{h\to 0} p_{ii}(h)$. Then

$$P(T_i > t) = \lim_{h\to 0} p_{ii}^n(h) = \lim_{h\to 0}\left(1 - \lambda_i h + o(h)\right)^n = \lim_{n\to\infty}(1-\lambda_i t/n)^n = e^{-\lambda_i t}$$

Clearly, the holding time parameter of state-i, $\theta_i = \lambda_i = \lim_{h\to 0}\left(1-p_{ii}(h)\right)/h$. Next, we determine p_{ij} for all $j \neq i$ and $i \in \Omega$.

$$p_{ij} = P(X(T_i)=j) = \int_0^\infty P(X(T_i)=j|T_i=t)P(T_i=t)$$

$$= \int_0^\infty P(X(t)=j|X(s)=i\ \forall s<t, X(t)\neq i) f_{T_i}(t)dt$$

$$= \int_0^\infty \lim_{h\to 0} P(X(t)=j|X(t-h)=i, X(t)\neq i)\, f_{T_i}(t)dt\text{, due to Markov property}$$

$$= \int_0^\infty \lim_{h\to 0} P\left(\underbrace{X(h)=j}_{C}\,|\,\underbrace{X(0)=i}_{A}, \underbrace{X(h)\neq i}_{B}\right) f_{T_i}(t)dt\text{, due to stationarity}$$

$P(C|AB) = P(ABC)/P(AB) = P(BC|A)/P(B|A)$. Note that $BC = C$, implying $P(C|AB) = P(C|A)/P(B|A) = P(X(h)=j|X(0)=i)/P(X(h)\neq i|X(0)=i) = p_{ij}(h)/\left(1-p_{ii}(h)\right)$.

$$\Rightarrow p_{ij} = \int_0^\infty \lim_{h\to 0}\frac{p_{ij}(h)}{1-p_{ii}(h)} f_{T_i}(t)dt = \lim_{h\to 0}\frac{p_{ij}(h)}{1-p_{ii}(h)}$$

So, we obtained $\theta_i = \lim_{h\to 0}\left(1-p_{ii}(h)\right)/h$ and $p_{ij} = \lim_{h\to 0} p_{ij}(h)/\left(1-p_{ii}(h)\right)$. Let us use these conversion formulas to obtain θ_i and p_{ij} for Poisson process, where

$$p_{ij}(h) = \begin{cases} 0 \ \text{ if } j < i \\ 1-\lambda h + o(h) \ \text{ if } j = i \\ \lambda h + o(h) \ \text{ if } j = i+1 \\ o(h) \ \text{ if } j \geq i+2 \end{cases}$$

Then $\theta_i = \lim_{h\to 0}\left(\lambda h + o(h)\right)/h = \lambda$ for all i and

$$p_{ij} = \lim_{h\to 0}\frac{p_{ij}(h)}{\lambda h + o(h)} = \begin{cases} 1 \ \text{ if } j = i+1 \\ 0 \text{ otherwise} \end{cases} \text{ for all } j \neq i$$

Now, we go the other way, i.e., we obtain $p_{ij}(h)$ for all $i, j \in \Omega$ from θ_i and p_{ij}'s.

$p_{ii}(h) = P(X(h) = i|X(0) = i)$
$= P(\text{No state change during } (0, h]|X(0) = i) + P(i \to i \text{ during } [0, h] \text{ with state change})$
$= P(T_i > h) + o(h) = e^{-\theta_i h} + o(h) = 1 - \theta_i h + o(h)$

The event that $i \to i$ during $[0, h]$ with state change requires $T_i + T_j \leq h$ for $j \neq i$. Probability of this event is no more than $P\big(T_i + T_j \leq h\big)$, where $j \neq i$ is chosen so that the probability is highest. $P\big(T_i + T_j \leq h\big) \leq P(T_i \leq h)P\big(T_j \leq h\big) = \big(1 - e^{-\theta_i h}\big)\big(1 - e^{-\theta_j h}\big) = \big(\theta_i h + o(h)\big) \times \big(\theta_j h + o(h)\big) = o(h)$, which we used in the above derivation of $p_{ii}(h)$.

For $j \neq i$, $p_{ij}(h) = P(X(h) = j|X(0) = i) = P(i \to j \text{ during } [0, h] \text{ with one state change})$ $+P(i \to j \text{ during } [0, h] \text{ with at least two state changes})$. From the above arguments,

$$p_{ij}(h) = P(i \to j \text{ during } [0, h] \text{ with one state change}) + o(h)$$
$$= \int_{s=0}^{h} P\Big(\underbrace{T_i = s}_{A}, \underbrace{X(s) = j}_{B}, \underbrace{T_j > h - s}_{C}\Big) + o(h)$$

$$\text{Now,} P(ABC) = P(C|AB)P(B|A)P(A) = P\big(T_j > h - s\big)P(X(T_i) = j)P(T_i = s)$$
$$= e^{-\theta_j(h-s)} \cdot p_{ij} \cdot \big(\theta_i e^{-\theta_i s} ds\big) = \theta_i p_{ij} e^{-\theta_j h} e^{(\theta_j - \theta_i)s} ds$$

$$\Rightarrow p_{ij}(h) = \theta_i p_{ij} e^{-\theta_j h} \int_0^h e^{(\theta_j - \theta_i)s} ds + o(h)$$

If $\theta_i = \theta_j$, then

$$p_{ij}(h) = \theta_i p_{ij} e^{-\theta_j h} \cdot h + o(h) = \theta_i p_{ij}\{1 - \theta_j h + o(h)\}h + o(h) = \theta_i p_{ij} h + o(h)$$

If $\theta_i \neq \theta_j$, then

$$p_{ij}(h) = \theta_i p_{ij} e^{-\theta_j h} \cdot \frac{e^{(\theta_j - \theta_i)h} - 1}{\theta_j - \theta_i} + o(h) = \theta_i p_{ij} \frac{e^{-\theta_i h} - e^{-\theta_j h}}{\theta_j - \theta_i} + o(h)$$
$$= \theta_i p_{ij} \frac{\{1 - \theta_i h + o(h)\} - \{1 - \theta_j h + o(h)\}}{\theta_j - \theta_i} + o(h) = \theta_i p_{ij} h + o(h)$$

So, we found $p_{ij}(h) = \theta_i p_{ij} h + o(h)$ if $j \neq i$ and $p_{ii}(h) = 1 - \theta_i h + o(h)$. Let us use these conversion formulas to obtain $p_{ij}(h)$ for the M/M/1 queue. For $i \geq 1$,

$$p_{ij}(h) = \begin{cases} \theta_i p_{ij} h + o(h) \text{ for } j \neq i \\ 1 - \theta_i h + o(h) \text{ for } j = i \end{cases} = \begin{cases} \mu h + o(h) & \text{if } j = i - 1 \\ \lambda h + o(h) & \text{if } j = i + 1 \\ 1 - (\lambda + \mu)h + o(h) & \text{if } j = i \\ o(h) & \text{otherwise} \end{cases}$$

Similarly, $p_{00}(h) = 1 - \lambda h + o(h)$, $p_{01}(h) = \lambda h + o(h)$, and $p_{0j}(h) = o(h)$ for $j \geq 2$.

Kolmogorov equations

<u>Transition rates</u>: We have two definitions for CTMC, and accordingly, there are two ways for specifying a CTMC. Here, we learn about a third way of specifying CTMC. We introduce a new quantity called the transition rate. It is the rate of change of $p_{ij}(h)$ with h.

$$q_{ij} \coloneqq p'_{ij}(h) = \lim_{h\to 0}\frac{p_{ij}(h)-p_{ij}(0)}{h} = \begin{cases} \lim_{h\to 0}\frac{p_{ii}(h)-1}{h} & \text{if } j=i \\ \lim_{h\to 0}\frac{p_{ij}(h)-0}{h} & \text{if } j\neq i \end{cases}$$

Note that $p_{ij}(0) = P(X(0)=j|X(0)=i)$ is either 1 (when $j=i$) or 0 (when $j\neq i$). From the conversion formulas, $p_{ii}(h) = 1-\theta_i h + o(h)$ and $p_{ij}(h) = \theta_i p_{ij} h + o(h)$ for $j\neq i$.

$$\Rightarrow q_{ij} = \begin{cases} \lim_{h\to 0}\frac{-\theta_i h + o(h)}{h} = -\theta_i & \text{if } j=i \\ \lim_{h\to 0}\frac{\theta_i p_{ij} h + o(h)}{h} = \theta_i p_{ij} & \text{if } j\neq i \end{cases}$$

If q_{ij} are known, then $\theta_i = -q_{ii}$ and $p_{ij} = -\,q_{ij}/q_{ii}$ for $j\neq i$. Transition rate from a state to itself is negative, signifying that the probability of remaining in the same state decreases with time. If $q_{ii}=0$, then state change from i is impossible, in other words, i is an absorbing state. Transition rates from a state to other states are non-negative. In fact, $\sum_{j\in\Omega} q_{ij} = \sum_{j\neq i} q_{ij} + q_{ii} = \sum_{j\neq i}\theta_i p_{ij} - \theta_i = \theta_i - \theta_i = 0$, as $\sum_{j\neq i} p_{ij} = 1$. So, the net transition rate from a state to all states is zero. If we represent the transition rates in the matrix form, then the entries in the principal diagonal are negative and the row sums are zero.

Let us calculate transition rates for M/M/1 queue, where $\theta_0 = \lambda$, $\theta_i = \lambda+\mu$ for $i\geq 1$, $p_{0,1}=1$, $p_{i,i-1} = \mu/(\lambda+\mu)$ and $p_{i,i+1} = \lambda/(\lambda+\mu)$ for $i\geq 1$. Then the transition rate matrix is as follows:

q_{ij}	0	1	2	3	$\cdots$
0	$-\lambda$	λ	0	0	$\cdots$
1	μ	$-(\lambda+\mu)$	λ	0	$\cdots$
2	0	μ	$-(\lambda+\mu)$	λ	$\cdots$
3	0	0	μ	$-(\lambda+\mu)$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

<u>Birth and death process</u>: Birth and death processes are CTMCs where state change happens via birth and death, i.e., if the current state is i, then the next state (after the holding time in state-i is over) is either $i+1$ (corresponding to a birth) or $i-1$ (corresponding to a death). In a birth and death process, $p_{i,i-1}+p_{i,i+1} = 1$ for all i. In terms of transition rates, $q_{ij}=0$ for all $j\leq i-2$ and $j\geq i+2$. For state-i, $q_{i,i+1}$ is referred to as the birth rate and $q_{i,i-1}$ is referred to as the death rate. The rate of 'no change' $q_{ii} = -\left(q_{i,i+1}+q_{i,i-1}\right)$ is negative of the sum of birth and death rates. Note that a birth and death process can be specified by birth and death rates. M/M/1 queue is a birth and death process with birth rate $q_{i,i+1}=\lambda$ for all i and death rate $q_{i,i-1}=\mu$ for $i\geq 1$ and 0 for $i=1$.

Markovian queueing models are most prominent examples of birth and death process. We have studied the M/M/1 queue. Let us study some more Markovian queues. Consider the M/M/2 queue, which is same as the M/M/1 queue except that there are two identical servers. Here, $T_0 \sim Exp(\lambda)$, $T_1 = \min\big(Exp(\lambda), Exp(\mu)\big) \sim Exp(\lambda+\mu)$, $T_2 = \min(Exp(\lambda), Exp(\mu), Exp(\mu)\big) \sim Exp(\lambda + 2\mu)$, as there are two in-service customers, and for the same reason, $T_i \sim Exp(\lambda + 2\mu)$ for $i \geq 2$. Also, $p_{0,1} = 1$, $p_{1,0} = \mathrm{P}\big(Exp(\mu) < Exp(\lambda)\big) = \mu/(\lambda+\mu)$, $p_{1,2} = \mathrm{P}\big(Exp(\lambda) < Exp(\mu)\big) = \lambda/(\lambda+\mu)$, and for $i \geq 2$, $p_{i,i-1} = \mathrm{P}\big(\min\{Exp(\mu), Exp(\mu)\} < Exp(\lambda)\big) = 2\mu/(\lambda+2\mu)$ and $p_{i,i+1} = \mathrm{P}(Exp(\lambda) < \min\{Exp(\mu), Exp(\mu)\}) = \lambda/(\lambda+2\mu)$. M/M/2 queue is a CTMC and a birth and death process having birth rate $q_{i,i+1} = \lambda$ for all i and death rate $q_{i,i-1} = 2\mu$ for $i \geq 2$, μ for $i = 1$, and 0 for $i = 1$.

We can generalize M/M/1 and M/M/2 queue to M/M/c queue that has $c \geq 1$ identical servers. It is a CTMC as well as a birth and death process. Its holding times $T_i \sim Exp(\lambda + i\mu)$ for $i \leq c$ and $T_i \sim Exp(\lambda + c\mu)$ for $i > c$. Transition probabilities are: $p_{i,i+1} = \lambda/(\lambda + i\mu)$, $p_{i,i-1} = i\mu/(\lambda + i\mu)$ for $i \leq c$ and $p_{i,i+1} = \lambda/(\lambda + c\mu)$, $p_{i,i-1} = c\mu/(\lambda + c\mu)$ for $i > c$. Its birth and death rates are: $q_{i,i+1} = \lambda$ for all i, $q_{i,i-1} = i\mu$ for $i \leq c$ and $c\mu$ for $i > c$.

A special case of M/M/c queue arises when c is large enough so that no customer ever waits, or a server can be made available whenever all existing servers are busy. Since no customer waits in this setup, we can model this situation as M/M/∞ queue, even though the number of servers may not be infinity. In this case, $T_i \sim Exp(\lambda + i\mu)$, $p_{i,i+1} = \lambda/(\lambda + i\mu)$ and $p_{i,i-1} = i\mu/(\lambda + i\mu)$ for all i. Also, $q_{i,i+1} = \lambda$ and $q_{i,i-1} = i\mu$ for all i.

At times there can be physical restriction to queue capacity in M/M/c queues. If the system capacity (i.e., queue capacity plus the number of servers) is K, then a customer arriving at time-t with $X(t) = K$ cannot join the queue. This changes system behavior in state-K. Then the state can change from K to $K-1$ only via a service completion. Then $p_{K,K-1} = 1$ and $q_{K,K+1} = 0$. Everything else remains the same.

In queueing theory, we are interested in finding the expected length of the queue, which then helps us obtain other quantities interest such as the expected waiting time of a customer. The results we are going to obtain in CTMC will enable us to calculate these quantities.

Kolmogorov backward equations: Kolmogorov equations, both backward and forward, are tools for obtaining $p_{ij}(t)$, which is essential for understanding short-run behavior of CTMCs. For the discrete-time case, this is equivalent to obtaining $p_{ij}^{(n)} = P(X_n = j | X_0 = i)$. There we used Chapman-Kolmogorov equation and obtained $p_{ij}^{(n)} = \sum_{k\in\Omega} p_{ik} p_{kj}^{(n-1)}$ in a recursive manner. For CTMCs, the equivalence of Chapman-Kolmogorov equation is:

$$p_{ij}(t) = P(X(t) = j | X(0) = i) = \sum_{k\in\Omega} P(X(t) = j, X(s) = k | X(0) = i) \text{ for some } s < t$$

$$
\begin{aligned}
&= \sum_{k\in\Omega} P(X(t)=j|X(s)=k, X(0)=i)P(X(s)=k|X(0)=i) \\
&= \sum_{k\in\Omega} P(X(t)=j|X(s)=k)p_{ik}(s) \text{ by Markov property} \\
&= \sum_{k\in\Omega} p_{ik}(s)P(X(t-s)=j|X(0)=k) \text{ due to stationarity} \\
&= \sum_{k\in\Omega} p_{ik}(s)p_{kj}(t-s)
\end{aligned}
$$

We need to set s to the smallest possible time unit for which $p_{ik}(s)$ is known. So, we set $s = h$ with $h \to 0$, but then to reach time t we need $t/h \to \infty$ number of recursive steps. Clearly, this mechanism of obtaining $p_{ij}(t)$ from $p_{ik}(h)$ is impossible. This problem did not arise in the discrete case because the smallest denominator of time was 1 (not $h \to 0^+$). Kolmogorov equations, which are differential equations, provide us the way out.

$$
\begin{aligned}
&p_{ij}(t+h) = \sum_{k\in\Omega} p_{ik}(h)p_{kj}(t) \text{ by Chapman Kolmogorov equation for CTMCs} \\
&\Rightarrow p_{ij}(t+h) - p_{ij}(t) = \sum_{k\neq i} p_{ik}(h)p_{kj}(t) + (p_{ii}(h)-1)p_{ij}(t) \\
&\Rightarrow \frac{p_{ij}(t+h)-p_{ij}(t)}{h} = \sum_{k\neq i} \frac{p_{ik}(h)}{h}p_{kj}(t) - \frac{1-p_{ii}(h)}{h}p_{ij}(t) \\
&\Rightarrow \lim_{h\to 0}\frac{p_{ij}(t+h)-p_{ij}(t)}{h} = \lim_{h\to 0}\sum_{k\neq i} \frac{\theta_i p_{ik}h + o(h)}{h}p_{kj}(t) - \lim_{h\to 0}\frac{\theta_i h + o(h)}{h}p_{ij}(t) \\
&\Rightarrow \frac{\mathrm{d}p_{ij}(t)}{\mathrm{dt}} = \sum_{k\neq i}\lim_{h\to 0}\left(\theta_i p_{ik} + \frac{o(h)}{h}\right)p_{kj}(t) - \theta_i p_{ij}(t), \text{exchanging limit and sum} \\
&\Rightarrow p'_{ij}(t) = \sum_{k\neq i}\theta_i p_{ik}p_{kj}(t) + q_{ii}p_{ij}(t) = \sum_{k\in\Omega} q_{ik}p_{kj}(t)
\end{aligned}
$$

The exchange of limit and sum is possible for the finite-state CTMCs. For the infinite-state CTMCs, it is not obvious, even with the convergence theorems. Yet we can show validity of the above result. A proof is presented in Appendix A (optional).

Solving the equations: Let us solve the backward equations and obtain $p_{ij}(t)$ for a simple CTMC. A machine works for $Exp(\lambda)$ amount of time and then fails, where $\lambda = 1/3$ per day. Repair takes take $Exp(\mu)$ amount of time, where $\mu = 1$ per day. On an average, the machine works for three days before failing and its repair takes one day. After the repair, the machine resumes working in a regular manner. We need to determine the probability that the machine is in working state at the end of the second day, given that it started in the working condition. Note that the question does not ask about the machine to be in the working state for two days, which is simply $P(Exp(1/3) > 2) = e^{-2/3} = 0.51$. In the question the machine can break down in-between, which makes the question difficult.

Let 1 denote the working state and 0 denote the breakdown state of the machine. Let $X(t)$ denote state of the machine at time t. It is obvious that $\{X(t): t \geq 0\}$ is a CTMC with $\theta_0 = \mu$, $\theta_1 = \lambda$ and $p_{0,1} = p_{1,0} = 1$. Also, $q_{0,0} = -\mu$, $q_{0,1} = \mu$ and $q_{1,0} = \lambda$, $q_{1,1} = -\lambda$. We need to determine $p_{11}(2)$, which we already know to be more than 0.51. Kolmogorov backward equations, i.e., $p'_{ij}(t) = \sum_{k\in\Omega} q_{ik}p_{kj}(t)$, for this CTMC are as follows:

$$p'_{11}(t) = q_{10}p_{01}(t) + q_{11}p_{11}(t) = \lambda p_{01}(t) - \lambda p_{11}(t) \quad \cdots\cdots (1)$$
$$p'_{10}(t) = q_{10}p_{00}(t) + q_{11}p_{10}(t) = \lambda p_{00}(t) - \lambda p_{10}(t) \quad \cdots\cdots (2)$$
$$p'_{01}(t) = q_{00}p_{01}(t) + q_{01}p_{11}(t) = -\mu p_{01}(t) + \mu p_{11}(t) \quad \cdots\cdots (3)$$
$$p'_{00}(t) = q_{00}p_{00}(t) + q_{01}p_{10}(t) = -\mu p_{00}(t) + \mu p_{10}(t) \quad \cdots\cdots (4)$$

Now, $(1) \times \mu + (3) \times \lambda \Rightarrow \mu p'_{11}(t) + \lambda p'_{01}(t) = 0 \Rightarrow \mu p_{11}(t) + \lambda p_{01}(t) = c$, a constant. At $t = 0$, $c = \mu p_{11}(0) + \lambda p_{01}(0) = \mu \times 1 + \lambda \times 0 = \mu$. Therefore, $\mu p_{11}(t) + \lambda p_{01}(t) = \mu \Rightarrow \lambda p_{01}(t) = \mu - \mu p_{11}(t)$. Using this in (1), we get the following equation in $p_{11}(t)$ alone.

$$p'_{11}(t) = \mu - (\lambda + \mu)p_{11}(t) \Rightarrow \frac{p'_{11}(t)}{\lambda + \mu} + p_{11}(t) - \frac{\mu}{\lambda + \mu} = 0$$

Let $h(t) = p_{11}(t) - \mu/(\lambda + \mu)$. Then $h'(t) = p'_{11}(t)$, and the above equation becomes

$$\frac{h'(t)}{\lambda + \mu} + h(t) = 0 \Rightarrow \frac{h'(t)}{h(t)} = -(\lambda + \mu) \Rightarrow h(t) = ke^{-(\lambda+\mu)t}, \text{where } k \text{ is a constant}$$

At $t = 0, h(0) = p_{11}(0) - \dfrac{\mu}{\lambda + \mu} = \dfrac{\lambda}{\lambda + \mu}$ equals $ke^0 \Rightarrow k = \dfrac{\lambda}{\lambda + \mu}$

$$\Rightarrow h(t) = \frac{\lambda}{\lambda + \mu}e^{-(\lambda+\mu)t} \Rightarrow p_{11}(t) = h(t) + \frac{\mu}{\lambda + \mu} = \frac{\lambda e^{-(\lambda+\mu)t} + \mu}{\lambda + \mu}$$

With $\lambda = 1/3$ and $\mu = 1$, $p_{11}(2) = 0.77 > 0.51$, as expected. We considered one of the simplest CTMCs, and yet obtaining a closed form expression for $p_{ij}(t)$ was difficult. Thus, we need numerical methods for real-life CTMCs. Let us write down the differential equations in the matrix form. Then we get

$$\begin{bmatrix} p'_{00}(t) & p'_{01}(t) & p'_{02}(t) & \cdots \\ p'_{10}(t) & p'_{11}(t) & p'_{12}(t) & \cdots \\ p'_{20}(t) & p'_{21}(t) & p'_{22}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} q_{00} & q_{01} & q_{02} & \cdots \\ q_{10} & q_{11} & q_{12} & \cdots \\ q_{20} & q_{21} & q_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \times \begin{bmatrix} p_{00}(t) & p_{01}(t) & p_{02}(t) & \cdots \\ p_{10}(t) & p_{11}(t) & p_{12}(t) & \cdots \\ p_{20}(t) & p_{21}(t) & p_{22}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$\equiv \boldsymbol{P}'(t) = \boldsymbol{Q} \times \boldsymbol{P}(t)$, in short.

We have a system of homogeneous linear differential equations. In the matrix form, it has a closed-form solution given below.

$$\boldsymbol{P}(t) = e^{\boldsymbol{Q}t} = \sum_{n=0}^{\infty} \frac{(\boldsymbol{Q}t)^n}{n!} = \lim_{n\to\infty}\left(\boldsymbol{I} + \frac{\boldsymbol{Q}t}{n}\right)^n, \text{where } \boldsymbol{I} \text{ is the identity matrix}$$

Getting closed-form expression for an individual term in $\boldsymbol{P}(t)$ is not possible from the above solution. However, we can evaluate numerically. We need to consider sufficiently large n in the above sum or limit so that the quantity of interest converges (with the required precision). For the above example, we see convergence of $p_{11}(2)$ to 0.77 by $n = 10$.

<u>Kolmogorov forward equations</u>: While obtaining the backward equations, we started with $p_{ij}(t+h) = \sum_{k\in\Omega} p_{ik}(h)p_{kj}(t)$. If we start with $p_{ij}(t+h) = \sum_{k\in\Omega} p_{ik}(t)p_{kj}(h)$, then the resultant equations are known as the Kolmogorov forward equations. It is so named because $p_{..}(h)$ appears in the 'front', while in the backward equations, $p_{..}(h)$ appears in the 'back'.

$$p_{ij}(t+h) = \sum_{k\in\Omega} p_{ik}(t)p_{kj}(h) \text{ by Chapman Kolmogorov equation for CTMCs}$$
$$\Rightarrow \frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_{k\neq j} p_{ik}(t)\frac{p_{kj}(h)}{h} + p_{ij}(t)\frac{p_{jj}(h) - 1}{h}$$
$$\Rightarrow \lim_{h\to 0}\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \lim_{h\to 0}\sum_{k\neq j} p_{ik}(t)\frac{\theta_k p_{kj}h + o(h)}{h} - \lim_{h\to 0} p_{ij}(t)\frac{1 - p_{jj}(h)}{h}$$
$$\Rightarrow \frac{\mathrm{d}p_{ij}(t)}{\mathrm{dt}} = \sum_{k\neq j} p_{ik}(t)\lim_{h\to 0}\left(\theta_k p_{kj} + \frac{o(h)}{h}\right) - p_{ij}(t)\theta_j, \text{exchanging limit and sum}$$
$$\Rightarrow p'_{ij}(t) = \sum_{k\neq j} p_{ik}(t)\theta_k p_{kj} + p_{ij}(t)q_{jj} = \sum_{k\in\Omega} p_{ik}(t)q_{kj}$$

Exchange of limit and sum is possible for finite-state CTMCs. However, for the infinite-state CTMCs, it is not always possible. Unlike the backward equations, here the limit-independent part adds up to $\sum_{k\neq j} p_{ik}(t) = 1 - p_{ij}(t) \leq 1$, but the problem here is with the convergence of the limit-dependent part. For a specific k, $\lim_{h\to\infty}\left(\theta_k p_{kj} + o(h)/h\right) = \theta_k p_{kj} \leq \theta_k$ exists, but we cannot say the same for arbitrary $k \neq j$. This situation did not arise in the backward equation, where we needed $\lim_{h\to\infty}(\theta_i p_{ik} + o(h)/h)$ to exist for a specific i.

As it turns out, the limit exists, i.e., θ_k is finite for all k, only for a class of CTMCs known as the non-explosive CTMCs. The proof of this fact is complicated and is skipped. However, we discuss non-explosive vs. explosive CTMCs, as the restriction of non-explosiveness will be required when we study the long-run behavior of CTMCs.

A CTMC is explosive if there exists $t < \infty$ such that the expected number of state change in $(0, t]$ is infinity. Consider a CTMC with $\theta_i = 2^i$ and $p_{i,i+1} = 1$. This CTMC, after reaching state i, spends $T_i \sim Exp(2^i)$ amount of time and then moves to state $i+1$. With $X(0) = 0$, the expected time to reach state n is: $E[\sum_{i=0}^{n-1} T_i] = \sum_{i=0}^{n-1} E[T_i] = \sum_{i=0}^{n-1} 2^{-i} = 2(1 - 2^{-n})$. Note that $\lim_{n\to\infty} E[\sum_{i=0}^{n-1} T_i] = 2$. So, the expected number of state change during $(0,2]$ is infinity. Thus, we have an explosive CTMC. Observe that this feature does not arise in the discrete case, where a state change takes one unit of time.

One may think that if θ_i is unbounded, then the CTMC would be explosive. This is not true, though its converse, i.e., CTMCs with bounded θ_i are non-explosive, is true. Let $\theta_i \leq M < \infty$ for all i. Then expected holding time in state-i, $E[T_i] = 1/\theta_i \geq 1/M$ for all i. Then the expected number of state change in $(0, t]$ can not exceed $t/(1/M) = tM < \infty$. Thus, all CTMCs with bounded θ_i are non-explosive. Finite-state CTMC are always non-explosive. Since there are finite number of states, then $\max\{\theta_i: i \in \Omega\}$ exists and we have a bound. For the infinite-state CTMCs, we may or may not have a bound. Consider the Poisson process where $\Omega = \{0,1,2,\dots\}$ but $\theta_i = \lambda\ \forall i$ are bounded. If θ_i are unbounded for an infinite-state CTMC, then it may or may not be explosive.

We started with an example where the infinite-state CTMC with unbounded θ_i is explosive. Now we will see a CTMC with similar features, but it is non-explosive. Consider the M/M/∞ queue, where $\Omega = \{0,1,2,\dots\}$ and $\theta_i = \lambda + i\mu$ is unbounded. However, the expected number of arrivals in $(0, t]$ is $t/(1/\lambda) = t\lambda < \infty$, as the arrival follows Poisson process with rate λ, and then the expected number of departures in $(0, t]$ is at the most $t\lambda$. Since a state change is equivalent to an arrival or a departure, expected number of state change in $(0, t]$ is bounded by $2\lambda t < \infty$. Hence, this CTMC is non-explosive.

Finding solution of the forward equations have the same difficulty as that with the backward equations. In the matrix form, forward equations can be represented as $\boldsymbol{P}'(t) = \boldsymbol{P}(t) \times \boldsymbol{Q}$, and its solution is same as that of the backward equations, i.e., $\boldsymbol{P}(t) = e^{\boldsymbol{Q}t}$.

Long-run behavior

<u>Absorbing CTMC</u>: We begin our study of long-run behavior of CTMCs with the absorbing chains (to be defined). We restrict ourselves to non-explosive CTMCs (having stationary transition probabilities). We exclude explosive CTMC, because such chains make infinite number of state change in finite duration, we do not have to study their long-run behavior. In the short-run itself, states of explosive CTMCs diverge to infinity.

In the discrete case, i is an absorbing state if $p_{ii} = 1$. In CTMCs, $p_{ii} = 0\ \forall i$ by definition. Absorbing states in CTMCs are captured by θ_i's. Holding time is state-i, $T_i \sim Exp(\theta_i) \Rightarrow P(T_i \leq t) = 1 - e^{-\theta_i t}$. If $\theta_i = 0$, then $P(T_i \leq t) = 0$ for all finite t, signifying that the holding time in state-i is never-ending. Such a state, if ever entered, can never be exited. Therefore, state-i of a CTMC is absorbing if $\theta_i = 0$. A CTMC is absorbing if it has at least one absorbing state, and from every non-absorbing state some absorbing state can be reached in finite time with positive probability.

Like discrete-time absorbing MCs, absorption is guaranteed in finite-state absorbing CTMCs as well. If it is not true, then the CTMC must remain among the non-absorbing states forever. Since the holding times cannot be infinity, then the CTMC must make infinite state changes among the non-absorbing states. We can construct a repeated coin tossing experiment where getting a head is equivalent to moving to an absorbing state. If probability of getting a head is positive, which is the case with finite-state CTMCs, then we eventually get head if we keep tossing, i.e., absorption is guaranteed. For the infinite-state case, absorption may or may not be guaranteed, just like its discrete-time counterpart.

The similarity between Markov chain and CTMC is not limited to the above observation. The only significant difference between these two is that the unit times in Markov chains become independent exponential random variables in CTMC. Since these random variables cannot be infinity (except for absorbing states) and they cannot be vanishingly small either (as we have restricted ourselves to non-explosive CTMCs), the long-run behaviors are hardly different. At least, the core ideas and basic patterns are the same.

Let us develop the continuous-time version of the first-step analysis for answering the key questions associated with an absorbing CTMC. Let $\tau \coloneqq \min\{t: X(t) \in A\}$ denote the time till absorption, where A denotes the set of absorbing states. We are interested in the probability of absorption into an absorbing state $u_{ij} = P(X(\tau) = j | X(0) = i)$ for $j \in A$ and the expected time till absorption $v_i = E[\tau | X(0) = i]$. Note that for $i \in A$, $u_{ij} = \mathbb{1}(j = i)$ and $v_i = 0$. So, we need to calculate these quantities for $i \in A^c$.

$$
\begin{aligned}
u_{ij} = P(X(\tau) = j | X(0) = i) &= \sum_{k \in \Omega} P(X(\tau) = j, X(T_i) = k | X(0) = i) \\
&= \sum_{k \in \Omega} P(X(\tau) = j | X(T_i) = k, X(0) = i)\, P(X(T_i) = k | X(0) = i) \\
&= \sum_{k \in \Omega} P(X(\tau) = j | X(T_i) = k)\, p_{ik}, \text{due to Markov property} \\
&= \sum_{k \in \Omega} p_{ik} P(X(\tau) = j | X(0) = k)\,, \text{due to stationarity} \\
&= \sum_{k \in \Omega} p_{ik} u_{kj} = \sum_{k \in A} p_{ik} u_{kj} + \sum_{k \in A^c} p_{ik} u_{kj} = p_{ij} + \sum_{k \in A^c} p_{ik} u_{kj}
\end{aligned}
$$

If we construct an absorbing Markov chain from the CTMC having absorbing states A, non-absorbing states A^c, and transition probabilities p_{ij} (with suitable changes in p_{ii} values for $i \in A$), then the absorption probabilities in the Markov chain are same as those in the CTMC. However, the time till absorption will be different. Intuition suggests that the unit times in the Markov chain would be expected holding time in CTMC.

$$
\begin{aligned}
v_i = E[\tau | X(0) = i] &= E_{X(T_i)}\big[E[\tau | X(T_i), X(0) = i]\big] \\
&= \sum_{k \in \Omega} E[\tau | X(T_i) = k, X(0) = i] P(X(T_i) = k | X(0) = i) \\
&= \sum_{k \in \Omega} E[T_i + \tau | X(0) = k]\, p_{ik}, \text{due to Markov property} \\
&= \sum_{k \in \Omega} p_{ik} (E[T_i] + E[\tau | X(0) = k]) = \sum_{k \in \Omega} p_{ik} (1/\theta_i + v_k)
\end{aligned}
$$

$$= 1/\theta_i + \sum_{k \in A} p_{ik} v_k + \sum_{k \in A^c} p_{ik} v_k = 1/\theta_i + \sum_{k \in A^c} p_{ik} v_k$$

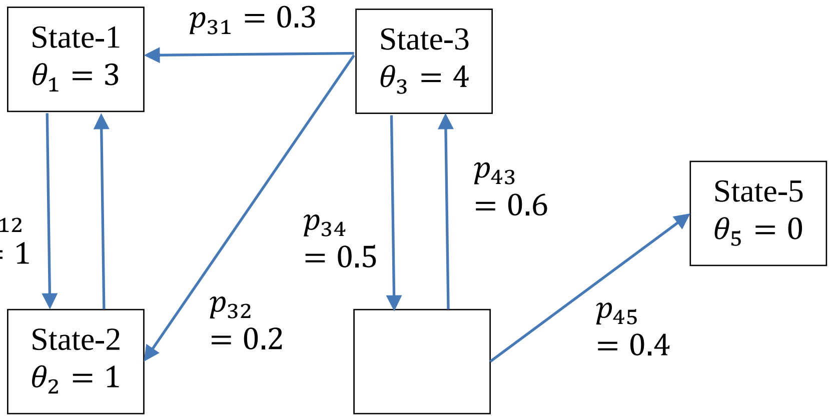
Communication and decomposition: Now, we begin long-run analysis of non-absorbing CTMCs. Like the case of Markov chains, we start with communication between states and decomposition of state-space. We say that state-i leads to state-j, denoted by $i \to j$, if $\exists t < \infty$ such that $p_{ij}(t) > 0$. States i and j are said to communicate, denoted by $i \leftrightarrow j$, if $i \to j$ and $j \to i$. A non-empty subset of the state-space $C \subseteq \Omega$ is called communicating block if $i \leftrightarrow j$ for all $i, j \in C$ and $i \nrightarrow j$ for all $i \in C$ and $j \in \Omega \setminus C$. We can uniquely decompose the state-space Ω into communicating blocks $C_1, C_2, \dots, C_k$ for some $k \geq 0$ and the set of remaining states $R = \Omega \setminus (C_1 \cup C_2 \cup \dots \cup C_\text{k})$.

Consider the CTMC in the figure (not the table) below. $1 \to 2$ as

$$p_{12}(t) = P(X(t) = 2 | X(0) = 1) \geq P(T_1 \leq t, X(T_1) = 2, T_1 + T_2 > t | X(0) = 1)$$
$$= P(T_1 \leq t | X(0) = 1) \cdot P(X(T_1) = 2 | T_1 \leq t, X(0) = 1)$$
$$\cdot P(T_1 + T_2 > t | X(T_1) = 2, T_1 \leq t, X(0) = 1)$$
$$= P(T_1 \leq t) P(X(T_1) = 2) P(T_1 + T_2 > t | T_1 \leq t), \text{by Markov property and independence}$$
$$= \left(1 - e^{-\theta_1 t}\right) p_{12} \int_0^t P(T_2 > t - s | T_1 = s) P(T_1 = s) = (1 - e^{-3t}) \times 1 \times \int_0^t e^{-\theta_2 (t-s)} \theta_1 e^{-\theta_1 s} ds$$
$$= (1 - e^{-3t}) \theta_1 e^{-\theta_2 t} \int_0^t e^{(\theta_2 - \theta_1)s} ds = (1 - e^{-3t}) 3 e^{-t} \frac{e^{(\theta_2 - \theta_1)t} - 1}{\theta_2 - \theta_1}$$
$$= 3(e^{-t} - e^{-4t}) \frac{1 - e^{-2t}}{2} > 0 \; \forall t > 0$$

In a similar manner, one can show that $2 \to 1$, $3 \to 1,2,4,5$ and $4 \to 1,2,3,5$. Then $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$. Furthermore, $\{1,2\}$ is a communicating block and so is $\{5\}$, an absorbing state, but not $\{3,4\}$, which is the set of remaining states.

p'_{ij}	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	0	0
3	0.3	0.2	0	0.5	0
4	0	0	0.6	0	0.4
5	0	0	0	0	1



We have an easier way of identifying communication and communicating blocks. Imagine a Markov chain with the state-space of the CTMC and transition probabilities $p'_{ij} = p_{ij}$ (of the CTMC) for all i, j, except for p'_{ii} for absorbing i (p'_{55} in the above example), which is $p'_{ii} = 1$. We call this chain as the embedded Markov chain (EMC) of the CTMC. The above table shows the transition probabilities of the EMC of the CTMC given in the figure. Observe that communication and decomposition are identical in the CTMC and its EMC.

If $i \to j$ in a CTMC, then $p_{ij}(t) > 0$ for some $t < 0$. Since t is finite and the CTMC is non-explosive, going from i to j involves finite number of state changes. One state change in the CTMC is equivalent to one step in the EMC. Then the EMC can go from i to j in finite steps with positive probability, i.e., $i \to j$ in the EMC. On the other hand, if $i \to j$ in the EMC, then there exists a path $i, i_1, i_2, \dots, i_{n-1}, j$ of length $n < \infty$ with positive probability. Holding times in states $i, i_1, i_2, \dots, i_{n-1}, j$ can be adjusted to fit into any finite t implying that $p_{ij}(t) > 0$, i.e., $i \to j$ in the CTMC. Due to this equivalence, as was seen in the above example, we can find communications and communicating blocks in a CTMC by studying its EMC.

After the decomposition of a CTMC into communicating blocks and remaining states, we can construct an absorbing CTMC by replacing communicating blocks with absorbing states and modifying transition probabilities (at the time of state change) as: $p_{ij_C} = \sum_{j \in C} p_{ij}$ for all $i \notin C$, where C denotes a communicating block and j_C is its replacement absorbing state. First-step analysis of this absorbing CTMC tells us about the absorption probabilities into different communicating blocks. If we know what happens inside a communicating block, then by weighing with the absorption probabilities, we can tell about the long-run behavior of the CTMC. So, we focus on the communicating blocks, which individually are CTMCs. These are non-explosive and have stationary transition probabilities. All states in these CTMCs communicate with one another. We call these irreducible CTMCs.

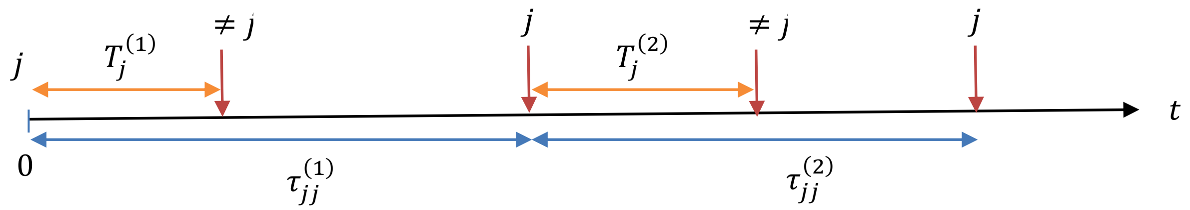
In order to study long-run behavior of irreducible CTMCs, we need to classify states into recurrent and transient. Let $\tau_{ij} \coloneqq \min\{t > 0 : X(t) = j | X(0) = i\}$ denote the time of first visit to state-j from state-i. If $j = i$, then τ_{ii} is the time of first return to state-i. Note that $\tau_{ij} > T_i$. State-i is recurrent if $P(\tau_{ii} < \infty) = 1$, i.e., the return is guaranteed, otherwise i is transient. Again, EMC is useful in identifying nature of states in a CTMC.

Let us consider an arbitrary path from i to i via other states: $i, i_1, i_2, \dots, i_n, i$ where $1 \leq n < \infty$ and $i_1, i_2, \dots, i_n \neq i$. Probabilities of all such paths contribute to $P(\tau_{ii} < \infty)$ both for the case of CTMC and its EMC. Paths of other forms do not contribute to $P(\tau_{ii} < \infty)$. If probabilities of the path $i, i_1, i_2, \dots, i_n, i$ are same in the CTMC and in its EMC, then $P(\tau_{ii} < \infty)$ is same for both the chains, and then the nature of state-i would be the same. Probability of the path in the EMC is $p_{ii_1} p_{i_1 i_2} \cdots p_{i_n i}$. In the CTMC we need to consider the role of holding times in states $i, i_1, i_2, \dots, i_n$. Since there is no restriction on the total time, all possible combinations of the holding times can be considered. Then the holding times do not influence the probability, i.e., probability of the path in the CTMC is also $p_{ii_1} p_{i_1 i_2} \cdots p_{i_n i}$. So, we have established that the nature of a state is same in a CTMC and in its EMC.

With the above observations, all the results associated with communication and classification of states that we encountered in Markov chain are available for CTMCs via its EMC. These are: (a) $i \leftrightarrow j \Rightarrow$ nature of i and j are the same, (b) a state not in any of the communicating blocks is transient, (c) nature of all states in an irreducible CTMC are the same, and (d) states of a finite-state irreducible CTMC are recurrent. Note that we cannot comment on positive vs. null recurrence and periodic vs. aperiodic at this point.

Long-run fractions: Let $\pi_{ij} := \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}(X(s) = j|X(0) = i)ds$ for all $i, j \in \Omega$. It denotes the long-run fraction of time spent by a CTMC in state-j after stating in state-i. In absorbing CTMCs, $\pi_{ij} = u_{ij}$ (absorption probability) if j is an absorbing state, else $\pi_{ij} = 0$. This is because absorption happens in finite time and π_{ij} is about infinite time horizon. In non-absorbing CTMCs, if j is not in any of the communicating blocks, then j is transient, and then $\pi_{ij} = 0$. If j is in communicating block C, then $\pi_{ij} = u_{iC}\pi'_j$ where u_{iC} is the absorption probability into communicating block C and π'_j is the long-run fraction for state-j in the irreducible CTMC "C". If C is transient, then $\pi'_j = 0$ for all $j \in C$, and then $\pi_{ij} = u_{iC}\pi'_j = 0$. If "$C$" is recurrent, we need to calculate π'_j, which we study next.

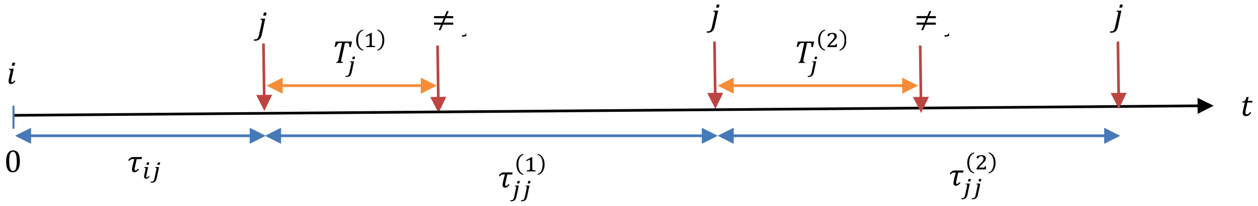
Consider an irreducible recurrent CTMC with starting state-j and we want to determine π_{jj}. Since the CTMC starts in state-j, it spends T_j amount of time in state-j and then enters into a different state. Since return to state-j is guaranteed due to recurrence, the same sequence of events repeats after the return at time $\tau_{jj} > T_j$, as depicted in the following diagram.



$$\text{Then } \pi_{jj} = \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}(X(s) = j|X(0) = j)ds$$
$$= \lim_{t\to\infty} \frac{1}{t}\left[\int_0^{\tau_{jj}^{(1)}} \mathbb{1}(X(s) = j)ds + \int_{\tau_{jj}^{(1)}}^{\tau_{jj}^{(1)}+\tau_{jj}^{(2)}} \mathbb{1}(X(s) = j)ds + \cdots\right]$$
$$= \lim_{t\to\infty} \frac{1}{t}\left[T_j^{(1)} + T_j^{(2)} + \cdots\right], \text{as } t = \tau_{jj}^{(1)} + \tau_{jj}^{(2)} + \cdots \text{ when } t \to \infty$$

Due to the Markov property, $\tau_{jj}^{(1)}, \tau_{jj}^{(2)}, \tau_{jj}^{(3)}, \ldots$ are *iid* random variables, and so are $T_j^{(1)}, T_j^{(2)}, T_j^{(3)}, \ldots$. Then we can consider a renewal-reward process taking place in the CTMC with $\tau_{jj}^{(1)}, \tau_{jj}^{(2)}, \tau_{jj}^{(3)}, \ldots$ as the inter-event times and $T_j^{(1)}, T_j^{(2)}, T_j^{(3)}, \ldots$ as the corresponding rewards. Then the quantity $\lim_{t\to\infty}\left(T_j^{(1)} + T_j^{(2)} + \cdots\right)/t$ represents the long-run reward rate, and it must be same as $E[T_j]/E[\tau_{jj}]$ by the renewal-reward theorem. Hence, $\pi_{jj} = E[T_j]/E[\tau_{jj}]$.

Now, let us consider the case when the irreducible recurrent CTMC starts in state $i \neq j$, and we want to determine π_{ij}. This situation is depicted in the following diagram.



Note that the CTMC does not visit state-j during $(0, \tau_{ij}]$ and after that the proceedings are same as before. If we can show that τ_{ij} is finite, i.e., $P(\tau_{ij} < \infty) = 1$, then π_{ij} is same as before, as π_{ij} is about infinite time horizon. Consider the CTMC to start in state-j. Since j is recurrent and $j \to i$, state-i is visited infinitely many times. If $P(\tau_{ij} < \infty) < 1$, then from one such visit to i it is possible that the CTMC never returns to j, but this contradicts with the fact that j is recurrent. Hence, τ_{ij} is finite and $\pi_{ij} = E[T_j]/E[\tau_{jj}]$, irrespective of i.

Let us consider the example mentioned previously with $\Omega = \{1,2,3,4,5\}$ and communicating blocks $\{1,2\}$ and $\{5\}$. Let us consider the communicating block $\{1,2\}$ as a CTMC. It is an irreducible and recurrent CTMC. Then $\pi_j = E[T_j]/E[\tau_{jj}]$ for $j = 1,2$. Since $p_{12} = p_{21} = 1$, $\tau_{11} = \tau_{22} = T_1 + T_2$. Then $\pi_1 = E[T_1]/(E[T_1] + E[T_2]) = \theta_1^{-1}/(\theta_1^{-1} + \theta_2^{-1}) = \theta_2/(\theta_1 + \theta_2)$ $= 0.25$ and $\pi_2 = \theta_1/(\theta_1 + \theta_2) = 0.75$.

Now, consider the whole CTMC with $X(0) = 3$. By the first-step analysis, probability of absorption into $\{1,2\}$ from 3 is $5/7$. Then $\pi_{31} = (5/7) \times 0.25 = 5/28$ and $\pi_{32} = 5/7 \times 0.75 = 15/28$. Also, $\pi_{35} = (2/7) \times 1 = 2/7$ and $\pi_{33} = \pi_{34} = 0$. Note that the stating state matters now as the whole CTMC is not irreducible.

While obtaining $\pi_j = E[T_j]/E[\tau_{jj}]$ for irreducible recurrent CTMCs, we implicitly assumed that $E[\tau_{jj}] < \infty$. It is possible that $P(\tau_{jj} < \infty) = 1$ (as j is recurrent), but $E[\tau_{jj}] = \infty$. Such a recurrent state is called null recurrent, and then $\pi_j = 0$. Instead, if $E[\tau_{jj}] < \infty$, then we call j to be positive recurrent. Positive and null recurrence are class properties, i.e., all states in a recurrent CTMC are either positive recurrent or null recurrent. However, the association of CTMC with its EMC does not work here. It is possible that a CTMC is positive recurrent, but its EMC is null recurrent and vice versa.

<u>Stationary distribution</u>: We need to calculate $E[\tau_{jj}]$, which can be very difficult at times, for obtaining π_j in an irreducible positive recurrent CTMCs. For null-recurrent CTMCs, $\pi_j = 0$ for all j, but we still need to calculate $E[\tau_{jj}]$ for at least one j for verifying null recurrence. Stationary distribution (SD) and its connection with the long-run fractions helps us in this. SD eliminates the need to calculate τ_{jj} altogether.

$\{\alpha_j : j \in \Omega\}$ with $\alpha_j \geq 0 \; \forall j$ and $\sum_{j \in \Omega} \alpha_j = 1$ is a SD for a CTMC if $\sum_{i \in \Omega} \alpha_i p_{ij}(t) = \alpha_j$ for all $j \in \Omega$ and all $t > 0$. Earlier, while discussing Kolmogorov backward equations, we obtained the following probability for the single-machine workshop example.

$p_{11}(t) = \dfrac{\lambda e^{-(\lambda+\mu)t} + \mu}{\lambda + \mu}$ and $p_{10}(t) = 1 - p_{11}(t) = \dfrac{\lambda - \lambda e^{-(\lambda+\mu)t}}{\lambda + \mu}$ $\forall t > 0$

By symmetry, $p_{00}(t) = \dfrac{\mu e^{-(\lambda+\mu)t} + \lambda}{\lambda + \mu}$ and $p_{01}(t) = \dfrac{\mu - \mu e^{-(\lambda+\mu)t}}{\lambda + \mu}$ $\forall t > 0$

Consider $\alpha_0 = \lambda/(\lambda + \mu)$ and $\alpha_1 = \mu/(\lambda + \mu)$. Then for all $t > 0$,

$$\sum_{i\in\Omega} \alpha_i p_{i0}(t) = \alpha_0 p_{00}(t) + \alpha_1 p_{10}(t) = \frac{\lambda}{\lambda+\mu}\frac{\mu e^{-(\lambda+\mu)t} + \lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}\frac{\lambda - \lambda e^{-(\lambda+\mu)t}}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu} = \alpha_0$$
$$\sum_{i\in\Omega} \alpha_i p_{i1}(t) = \alpha_0 p_{01}(t) + \alpha_1 p_{11}(t) = \frac{\lambda}{\lambda+\mu}\frac{\mu - \mu e^{-(\lambda+\mu)t}}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}\frac{\lambda e^{-(\lambda+\mu)t} + \mu}{\lambda+\mu} = \frac{\mu}{\lambda+\mu} = \alpha_1$$

Clearly, $\{\alpha_0, \alpha_1\} = \{\lambda/(\lambda + \mu), \mu/(\lambda + \mu)\}$ is a SD for the CTMC.

In the above example, we verified whether a given distribution is SD for a CTMC. We could have obtained the same distribution as SD by solving the equations $\sum_{i\in\Omega} \alpha_i p_{ij}(t) = \alpha_j$ for all $j \in \Omega$ and $\sum_{j\in\Omega} \alpha_j = 1$. In both the cases, we need the expressions for $p_{ij}(t)$ for all i, j and t. These expressions are almost impossible to obtain for a general CTMC. However, we do not need anything other than $p_{ij}(h)$ for all i, j, which we know for a CTMC, in order to verify or obtain SD. This is due to the stationarity of the transition probabilities. If $\{\alpha_j : j \in \Omega\}$ qualifies as SD for an arbitrarily small $h > 0$, then

$$\sum_{i\in\Omega} \alpha_i p_{ij}(2h) = \sum_{i\in\Omega} \alpha_i \sum_{k\in\Omega} p_{ik}(h) p_{kj}(h) = \sum_{k\in\Omega}\sum_{i\in\Omega} \alpha_i p_{ik}(h) p_{kj}(h) = \sum_{k\in\Omega} \alpha_k p_{kj}(h) = \alpha_j \ \forall j.$$

Proceeding recursively, $\sum_{i\in\Omega} \alpha_i p_{ij}(nh) = \alpha_j \ \forall j \in \Omega$ and all $n \geq 1$. Now for any finite $t > 0$,

$$\sum_{i\in\Omega} \alpha_i p_{ij}(t) = \sum_{i\in\Omega} \alpha_i \cdot \lim_{\substack{n\to\infty \\ nh=t}} p_{ij}(nh) = \lim_{\substack{n\to\infty \\ nh=t}} \sum_{i\in\Omega} \alpha_i p_{ij}(nh) = \lim_{\substack{n\to\infty \\ nh=t}} \alpha_j = \alpha_j \ \forall j.$$

The exchange of limit and sum is possible due to the bounded convergence theorem.

<u>Long-run fractions and SD</u>: Much like the discrete-time case, there is a close connection between SD and long-run fractions, as stated in the following result. It simplifies calculation of long-run fractions for positive recurrent CTMCs. It also helps us verifying if an irreducible CTMC is positive recurrent or not.

Res: An irreducible positive recurrent CTMC has a unique SD given by its long-run fractions.

Proof: Long-run fractions are defined as: $\pi_{ij} = \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}(X(s) = j | X(0) = i) ds$.

Taking expectation, $E[\pi_{ij}] = E\left[\lim_{t\to\infty} \dfrac{1}{t}\displaystyle\int_0^t \mathbb{1}_{X(s)=j|X(0)=i} ds\right]$

$$\Rightarrow \pi_{ij} = \lim_{t\to\infty} \frac{1}{t}\int_0^t E[\mathbb{1}(X(s) = j | X(0) = i)] ds = \lim_{t\to\infty} \frac{1}{t}\int_0^t p_{ij}(s) ds$$

$E[\pi_{ij}] = \pi_{ij}$ because π_{ij} is constant, and exchange of expectation and limit is possible due to convergence theorem. Using this alternate measure of long-run fractions, first we show that $\{\pi_j : j \in \Omega\}$ is a SD for irreducible positive-recurrent CTMC. Next, we show its uniqueness. Note that we dropped i from π_{ij}, as it does not depend on i in recurrent CTMC.

From the definition it is clear that $\pi_j \geq 0$ for all $j \in \Omega$. Let us check if these add up to 1.

$$\sum_{j\in\Omega} p_{ij}(s) = 1 \;\forall s > 0 \Rightarrow \frac{1}{t}\int_0^t \sum_{j\in\Omega} p_{ij}(s)\,ds = \sum_{j\in\Omega} \frac{1}{t}\int_0^t p_{ij}(s)ds = 1 \;\forall t > 0$$
$$\Rightarrow \lim_{t\to\infty} \sum_{j\in\Omega} \frac{1}{t}\int_0^t p_{ij}(s)ds = 1 \Rightarrow \sum_{j\in\Omega} \lim_{t\to\infty} \frac{1}{t}\int_0^t p_{ij}(s)ds = 1 \Rightarrow \sum_{j\in\Omega} \pi_j = 1 \text{ (as required)}$$

Exchange of limit and sum is not obvious for the infinite-state case. However, it is possible. See notes on Markov chain for the reasoning. So $\{\pi_j : j \in \Omega\}$ is a valid distribution over Ω. Let us check if it satisfies $\sum_{i\in\Omega} \pi_i p_{ij}(t) = \pi_j$ for all $j \in \Omega$ and $t > 0$.

$$p_{kj}(s+t) = \sum_{i\in\Omega} p_{ki}(s) p_{ij}(t) \;\;\forall j \in \Omega \text{ and } s,t > 0;\; k \text{ is arbitrary starting state}$$
$$\Rightarrow \frac{1}{u}\int_0^u p_{kj}(s+t)ds = \frac{1}{u}\int_0^u \sum_{i\in\Omega} p_{ki}(s)p_{ij}(t)\,ds = \sum_{i\in\Omega}\left(\frac{1}{u}\int_0^u p_{ki}(s)ds\right)p_{ij}(t) \;\;\forall j,u,t$$
$$\Rightarrow \lim_{u\to\infty}\frac{1}{u}\int_0^u p_{kj}(s+t)ds = \lim_{u\to\infty}\sum_{i\in\Omega}\left(\frac{1}{u}\int_0^u p_{ki}(s)ds\right)p_{ij}(t) \;\;\forall j,t$$
$$\Rightarrow \lim_{u\to\infty}\frac{1}{u}\left[\int_0^u p_{kj}(s)ds + \int_u^{u+t} p_{kj}(s)ds - \int_0^t p_{kj}(s)ds\right] = \sum_{i\in\Omega}\left(\lim_{u\to\infty}\frac{1}{u}\int_0^u p_{ki}(s)ds\right)p_{ij}(t) \;\;\forall j,t$$
$$\Rightarrow \pi_j = \sum_{i\in\Omega} \pi_i p_{ij}(t) \;\;\forall j,t \text{ (as required)}$$

Again, the exchange of limit and sum is not obvious for the infinite-state case. However, it is possible for the reasons mentioned in the notes on Markov chain.

With $\{\pi_j : j \in \Omega\}$ established as a SD for irreducible positive recurrent CTMC, we establish its uniqueness by showing that an arbitrary SD $\{\alpha_j : j \in \Omega\}$ is same as $\{\pi_j : j \in \Omega\}$.

$$\sum_{i\in\Omega} \alpha_i p_{ij}(s) = \alpha_j \;\forall j \in \Omega, s \Rightarrow \frac{1}{t}\int_0^t \sum_{i\in\Omega} \alpha_i p_{ij}(s)\,ds = \sum_{i\in\Omega} \alpha_i \frac{1}{t}\int_0^t p_{ij}(s)ds = \alpha_j \;\forall j,t$$
$$\Rightarrow \lim_{t\to\infty}\sum_{i\in\Omega} \alpha_i \frac{1}{t}\int_0^t p_{ij}(s)ds = \sum_{i\in\Omega} \alpha_i \lim_{t\to\infty}\frac{1}{t}\int_0^t p_{ij}(s)ds = \sum_{i\in\Omega} \alpha_i \pi_j = \pi_j = \alpha_j \;\forall j, \text{as required}$$

The exchange of limit and sum is possible due to bounded convergence theorem. □

Let us consider the single-machine workshop example introduced earlier. We specified the two-state CTMC with $\theta_0 = \mu$, $\theta_1 = \lambda$, $p_{0,1} = p_{1,0} = 1$. Then $p_{00}(h) = 1 - \mu h + o(h)$, $p_{01}(h) = \mu h + o(h)$, $p_{10}(h) = \lambda h + o(h)$, $p_{11}(h) = 1 - \lambda h + o(h)$. Let us determine the unique SD for this positive recurrent CTMC.

We need to solve for $\pi_j = \sum_{i \in \Omega} \pi_i p_{ij}(h)$ for $j \in \{0,1\}$ and $\pi_0 + \pi_1 = 1$. For $j = 0$,

$$\pi_0 = \pi_0 p_{00}(h) + \pi_1 p_{10}(h) = \pi_0\{1 - \mu h + o(h)\} + (1 - \pi_0)\{\lambda h + o(h)\}$$
$$\Rightarrow \pi_0\{(\lambda + \mu)h + o(h)\} = \lambda h + o(h) \Rightarrow \lim_{h \to 0} \pi_0 \left\{(\lambda + \mu) + \frac{o(h)}{h}\right\} = \lim_{h \to 0} \left\{\lambda + \frac{o(h)}{h}\right\}$$
$$\Rightarrow \pi_0 = \frac{\lambda}{\lambda + \mu} \text{ and } \pi_1 = 1 - \pi_0 = \frac{\mu}{\lambda + \mu} \text{ (as expected)}$$

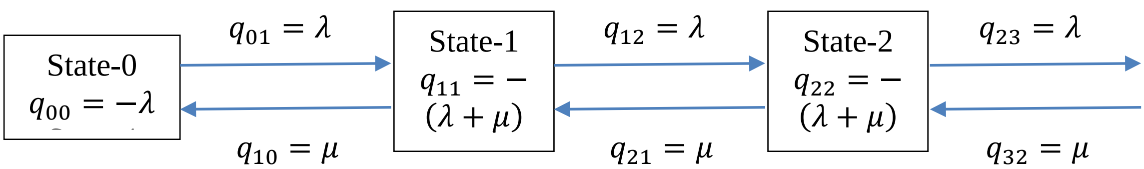
The second part of the above proof, where we showed that "if $\{\alpha_j : j \in \Omega\}$ is a SD then it must be same as $\{\pi_j : j \in \Omega\}$" is valid for all irreducible CTMCs. This implies that SD cannot exist for transient and null recurrent CTMCs. If SD exists for such chains, then long-run fractions cannot be zero, as concluded earlier. This observation allows us to check if an irreducible CTMC is positive recurrent or not. If we get non-negative solution of $\sum_{i \in \Omega} \pi_i p_{ij}(h) = \pi_j$ $\forall j \in \Omega$ and $\sum_{j \in \Omega} \pi_j = 1$, then the CTMC is positive recurrent, else it is not.

<u>Rate balance equations</u>: For positive recurrent CTMCs, we can obtain long-run fractions by solving $\pi_j = \sum_{i \in \Omega} \pi_i p_{ij}(h)$ $\forall j \in \Omega$ and $\sum_{j \in \Omega} \pi_j = 1$. We need to set $h \to 0$ for determining π_j. We can set the limit before using the expressions for $p_{ij}(h)$. The resultant equations are known as the rate balance equations, which also gives us long-run fractions.

$$\pi_j = \sum_{i \in \Omega} \pi_i p_{ij}(h) \Rightarrow \pi_j \left(1 - p_{jj}(h)\right) = \sum_{i \neq j} \pi_i p_{ij}(h)$$
$$\Rightarrow \pi_j \left(\theta_j h + o(h)\right) = \sum_{i \neq j} \pi_i \left(\theta_i p_{ij} h + o(h)\right) \ \forall j \in \Omega$$
$$\Rightarrow \lim_{h \to 0} \pi_j \left(\theta_j + \frac{o(h)}{h}\right) = \lim_{h \to 0} \sum_{i \neq j} \pi_i \left(\theta_i p_{ij} + \frac{o(h)}{h}\right) = \sum_{i \neq j} \pi_i \left(\theta_i p_{ij} + \lim_{h \to 0} \frac{o(h)}{h}\right)$$
$$\Rightarrow \pi_j(-q_{jj}) = \sum_{i \neq j} \pi_i q_{ij} \Rightarrow \sum_{i \in \Omega} \pi_i q_{ij} = 0 \ \forall j \in \Omega$$

Exchange of limit and sum is possible due to bounded convergence theorem. Now, we can solve $\sum_{i \in \Omega} \pi_i q_{ij} = 0$ $\forall j \in \Omega$ and $\sum_{j \in \Omega} \pi_j = 1$ to get the long-run fractions.

Let us use the rate balance equations to study M/M/1 queue, which is depicted below.



Earlier, we have noted that M/M/1 queue is a CTMC with stationary transition probabilities. It is non-explosive. It is irreducible as its EMC is irreducible. Its infinite-state, so we are not sure about positive-recurrence. If it has SD, then it is positive-recurrent. Let us try to solve the rate balance equations $\sum_{i\in\Omega}\pi_i q_{ij} = 0\ \forall j \in \Omega$.

For $j = 0$, $\sum_{i\in\Omega}\pi_i q_{i0} = \pi_0 q_{00} + \pi_1 q_{10} = -\lambda\pi_0 + \mu\pi_1 = 0 \Rightarrow \pi_1 = (\lambda/\mu)\pi_0$.
For $j = 1$, $\sum_{i\in\Omega}\pi_i q_{i1} = \pi_0 q_{01} + \pi_1 q_{11} + \pi_2 q_{21} = \lambda\pi_0 - (\lambda+\mu)\pi_1 + \mu\pi_2 = 0$
$\Rightarrow \mu\pi_2 = (\lambda+\mu)(\lambda/\mu)\pi_0 - \lambda\pi_0 \Rightarrow \pi_2 = (\lambda/\mu)^2\pi_0$.
For $j = 2$, $\sum_{i\in\Omega}\pi_i q_{i2} = \pi_1 q_{12} + \pi_2 q_{22} + \pi_3 q_{32} = \lambda\pi_1 - (\lambda+\mu)\pi_2 + \mu\pi_3 = 0$
$\Rightarrow \mu\pi_3 = (\lambda+\mu)(\lambda/\mu)\pi_1 - \lambda\pi_1 \Rightarrow \pi_3 = (\lambda/\mu)^2\pi_1 = (\lambda/\mu)^3\pi_0$.

The next equations are similar to the above, and in general $\pi_j = (\lambda/\mu)^j\pi_0$ for $n = 0,1,2,\ldots$. Now, $\sum_{j\in\Omega}\pi_j = \sum_{j=0}^{\infty}(\lambda/\mu)^j\pi_0 = \pi_0(1-\lambda/\mu)^{-1}$, provided $\lambda < \mu$, else $\sum_{j\in\Omega}\pi_j = \infty$ and then we cannot equate it to1. Thus, SD exists, i.e., M/M/1 queue is positive recurrent, *iff* $\lambda < \mu$, and then $\sum_{j\in\Omega}\pi_j = 1 \Rightarrow \pi_0 = 1-\lambda/\mu \Rightarrow \pi_j = (\lambda/\mu)^j(1-\lambda/\mu)$ for all j. The condition that $\lambda < \mu$ is equivalent to server utilization ratio $\rho \coloneqq \lambda/\mu < 1$. For any Markovian queue to be positive recurrent, its utilization ratio must be strictly less than 1. Note that the expression for ρ is different for different queues.

<u>Limiting distribution</u>: $\beta_{ij} \coloneqq \lim_{t\to\infty} p_{ij}(t)$, provided limit exists, is the limiting probability of state-j in a CTMC with starting state i. If β_{ij} exists for all $j \in \Omega$ and $\sum_{j\in\Omega}\beta_{ij} = 1$, then $\{\beta_{ij}: j \in \Omega\}$ is called the limiting distribution (LD) of the CTMC with starting state i. First, we discuss the case of irreducible positive recurrent CTMC, which has unique stationary distribution given by the long-run fractions.

In Markov chains, we have the concept of periodicity which leads to further classification of irreducible positive recurrent Markov chains into ergodic and non-ergodic chains, and it is the ergodic chains, that have LD. We also showed that the LD is a SD whenever it exists. This implies that LD and SD are the same for ergodic Markov chains. It also implies that transient and null recurrent Markov chains do not have LD.

In CTMCs, the concept of periodicity is irrelevant. If $p_{ij}(t) > 0$ for some $t < \infty$, then we can always adjust the holding times to ensure that $p_{ij}(t) > 0$ for all $t > 0$. Here LD exists for all irreducible positive recurrent CTMCs. Existence of LD for such chains can be established following the arguments presented for the case of Markov chains. Equality of LD and SD for irreducible positive recurrent CTMC follows from the result below, which also implies that transient and null recurrent CTMCs do not have LD. For a general CTMC, LD exists if all its communicating blocks are positive recurrent. We need to weigh the individual LDs with the corresponding absorption probabilities to obtain LD for the whole CTMC. The remaining states, being transient, do not feature in the LD.

Res: Limiting distribution, if exists, is a stationary distribution for a CTMC.

Proof: Let $\{\beta_{ij}: j \in \Omega\}$ be the LD, with the starting state being i. Then $\beta_{ij} = \lim_{t\to\infty} p_{ij}(t)$ for all $j \in \Omega$ and $\sum_{j\in\Omega} \beta_{ij} = 1$. Now, we check for the condition for SD.

$$p_{ij}(s+t) = \sum_{k\in\Omega} p_{ik}(s)p_{kj}(t)\ \forall j \in \Omega, s, t \Rightarrow \lim_{s\to\infty} p_{ij}(s+t) = \lim_{s\to\infty} \sum_{k\in\Omega} p_{ik}(s)p_{kj}(t)$$
$$= \sum_{k\in\Omega} \left(\lim_{s\to\infty} p_{ik}(s)\right) p_{kj}(t)\ \forall j, t \Rightarrow \beta_{ij} = \sum_{k\in\Omega} \beta_{ik}p_{kj}(t)\ \forall j, t$$

Exchange of limit and sum is not obvious for the infinite-state case. However, it is possible. See the notes on Markov chain for the reasoning. So, $\{\beta_{ij}: j \in \Omega\}$ is a SD. □

In the M/M/1 queue with $\rho = \lambda/\mu < 1$, the irreducible CTMC is positive recurrent, and with the above discussion, LD exists for it and $\beta_{ij} = \pi_j = (\lambda/\mu)^j(1-\lambda/\mu) = \rho^j(1-\rho)$ for $j = 0,1,2,\ldots$. We say that a queue is in steady state when LD exists and the gaps between $p_{ij}(t)$ and β_{ij} are within a tight error bound and it decreases with t. In the steady state, an M/M/1 queue on an average has $\sum_{j=1}^{\infty} j\beta_{ij} = \sum_{j=1}^{\infty} j\rho^j(1-\rho) = \rho(1-\rho)\frac{d}{d\rho}\left(\sum_{j=1}^{\infty}\rho^j\right) = \rho(1-\rho) \cdot \frac{d}{d\rho}\left(\frac{\rho}{1-\rho}\right) = \rho(1-\rho)\cdot\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho}$ number of customers in the system. Next, we can use the *Little's Law* to obtain the expected time spent by a customer in the system in steady state. From these, we can obtain more quantities of interest.

Practice problems

Book-1: Introduction to Probability Models by Sheldon Ross [10th edition]

Modeling aspects
Book-1, Chapter-6, Exercise No. 4, 5, 9, 12a

Long-run behavior
Book-1, Chapter-6, Exercise No. 12b, 15, 20, 24

Appendix A

We need to show validity of the Kolmogorov backward equations for infinite-state CTMCs. Following derivation of the equations, we just need to show that

$$\lim_{h\to 0}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t)\ \text{ is same as }\ \sum_{k\neq i}\lim_{h\to 0}\frac{p_{ik}(h)}{h}p_{kj}(t) = \sum_{k\neq i} q_{ik}p_{kj}(t)$$

We cannot exchange limit and sum because we have infinite sum and convergence theorems are not useful here. If we restrict the sum over k to some finite n, then we can exchange limit and sum, and obtain a lower bound. We also get an upper bound. Then we show that the bounds converge to $\sum_{k\neq i} q_{ik}p_{kj}(t)$ as $n \to \infty$, implying the desired result.

$$\lim_{h\to 0}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t) \geq \lim_{h\to 0}\sum_{\substack{k\neq i\\ k\leq n}}\frac{p_{ik}(h)}{h}p_{kj}(t) = \sum_{\substack{k\neq i\\ k\leq n}}\lim_{h\to 0}\frac{p_{ik}(h)}{h}p_{kj}(t) = \sum_{\substack{k\neq i\\ k\leq n}} q_{ik}p_{kj}(t)\ \forall n$$

$$\Rightarrow \lim_{h\to 0}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t) \geq \lim_{n\to\infty}\sum_{k\neq i,k\leq n} q_{ik}p_{kj}(t) = \sum_{k\neq i} q_{ik}p_{kj}(t)\text{, a lower bound}$$

For the upper bound, we split the infinite sum at a finite $n > i$. Then

$$\begin{aligned}\lim_{h\to 0}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t) &= \lim_{h\to 0}\sum_{k\neq i,k\leq n}\frac{p_{ik}(h)}{h}p_{kj}(t) + \lim_{h\to 0}\sum_{k>n}\frac{p_{ik}(h)}{h}p_{kj}(t)\\ &\leq \sum_{k\neq i,k\leq n} q_{ik}p_{kj}(t) + \lim_{h\to 0}\sum_{k>n}\frac{p_{ik}(h)}{h}\ \forall n > i\end{aligned}$$

Now, $\lim_{h\to 0}\sum_{k>n}\frac{p_{ik}(h)}{h} = \lim_{h\to 0}\frac{1-\sum_{k\leq n}p_{ik}(h)}{h} = \lim_{h\to 0}\frac{1-p_{ii}(h)}{h} - \lim_{h\to 0}\sum_{k\neq i,k\leq n}\frac{p_{ik}(h)}{h}$

$$= -q_{ii} - \sum_{k\neq i,k\leq n}\lim_{h\to 0}\frac{p_{ik}(h)}{h} = -\sum_{k\leq n} q_{ik}\ \forall n > i$$

$$\Rightarrow \lim_{h\to 0}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t) \leq \sum_{k\neq i,k\leq n} q_{ik}p_{kj}(t) - \sum_{k\leq n} q_{ik}\ \forall n > i$$

$$\Rightarrow \lim_{h\to 0}\sum_{k\neq i}\frac{p_{ik}(h)}{h}p_{kj}(t) \leq \lim_{n\to\infty}\sum_{k\neq i,k\leq n} q_{ik}p_{kj}(t) - \lim_{n\to\infty}\sum_{k\leq n} q_{ik} = \sum_{k\neq i} q_{ik}p_{kj}(t) - \sum_{k\in\Omega} q_{ik}$$

$$= \sum_{k\neq i} q_{ik}p_{kj}(t)\text{, an upper bound}$$

Note that the lower and upper bounds are the same, as required.