Notes on STAT-433/STAT-833: Stochastic Processes II

Unversity of Waterloo

David Duan

Last Updated: January 2, 2022 (draft)

Contents

Remark. This document is slightly less polished than my other course notes, mostly because I didn't spend as much time rewriting the notes I took during the lectures (plus the assignments are much more time-consuming than expected). I will try to rewrite some of the sections when I find free time.

Chapter 1

Discrete Phase-Type Distribution

Section 1. DPH Definition and Probability Mass Function

1.1. Note: Let $\{X_n\}_{n=0,1,\ldots}$ be a DTMC with state space $S = \{0, 1, \ldots, N\}$, where

- states $0, 1, \ldots, M-1$ are transient,
- states $M, M+1, \ldots, N$ are absorbing.

Then the transition matrix of this DTMC can be expressed as

$$P = \begin{array}{c} 0 \\ \vdots \\ M-1 \\ \\ M \\ \vdots \\ N \end{array} \overset{\begin{array}{cc} 0 \cdots M-1 & M \cdots N \end{array}}{\begin{bmatrix} Q_{(M-1),(M-1)} & R_{(M-1),(N-M)} \\ \\ O_{(N-M),(M-1)} & I_{(N-M),(N-M)} \end{bmatrix}}$$

The bottom left submatrix is a zero matrix and the bottom right submatrix is identity because once you've entered an absorbing state, you will stay there forever.

1.2. Note: Define the **time until absorption** as the first time X travels to an absorbing state:

$$T := \min\{n \mid M \leq X_n \leq N\}$$

Denote the initial distribution as

$$\vec{\alpha}_0 = (\alpha_{0,0}, \alpha_{0,1}, \ldots, \alpha_{0,M-1}, \alpha_{0,M}, \ldots, \alpha_{0,N}),$$

where $\alpha_{0,i} = \Pr(X_0 = i) = \mu_0(i)$, i.e., $\alpha_{0,i}$ gives the probability that the chain starts at state i. Note the subvector consists of the first M entries

$$\vec{\alpha}_0' = (\alpha_{0,0}, \alpha_{0,1}, \ldots, \alpha_{0,M-1})$$

gives us the probability from starting at each transient state. We will see soon that we care about entries in $\vec{\alpha}_0'$ a lot while the other entries $(\alpha_{0,M}, \ldots, \alpha_{0,N})$ is not that important.[a]

[a] Starting at an absorbing state is quite boring because we will just stay there forever...

1.3. Note: Recall that we were focusing on the expected absorption time T (i.e., what is the expected time stamp such that the chain will enter an absorbing state?) in previous discussions. We are now interested on the *exact distribution* of T, i.e., calculating $\Pr(T = k)$ for each $k \geq 1$.

Let A, B be the set of transient and absorbing states, respectively. First, the probability that the chain gets absorbed at time 0 equals the sum of probabilities that it starts in an absorbing state:

$$\Pr(T = 0) = \Pr(X_0 \in B) = \sum_{i=M}^{N} \alpha_{0,i}.$$

For the chain to be absorbed at time $k > 0$, the chain must stays in the transient part of the DTMC in time $0, 1, \ldots, k-1$, then visits an absorbing state at time k, i.e.,

$$\begin{aligned}
\Pr(T = k) &= \Pr(X_0 \in A, X_1 \in A, \ldots, X_{k-1} \in A, X_k \in B) \\
&= \sum_{x_0, x_1, \ldots, x_{k-1} \in \{0,1,\ldots,M-1\}} \sum_{x_k=M}^{N} \Pr(X_0 = x_0, X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, X_k = x_k) \\
&= \sum_{x_0, x_1, \ldots, x_{k-1} \in \{0,1,\ldots,M-1\}} \sum_{x_k=M}^{N} \alpha_{0,x_0} P_{x_0,x_1} P_{x_1,x_2} \cdots P_{x_{k-1},x_k} \\
&= \sum_{x_0, x_1, \ldots, x_{k-1} \in \{0,1,\ldots,M-1\}} \alpha_{0,x_0} P_{x_0,x_1} P_{x_1,x_2} \cdots P_{x_{k-2},x_{k-1}} \sum_{x_k=M}^{N} P_{x_{k-1},x_k} \\
&=: \vec{\alpha}_0' Q^{k-1} \vec{q}
\end{aligned}$$

- Line 2: Enumerate all possible state traces for the chain from time 0 to time k.
- Line 3: Replace the probabilities with notations.
- Line 4: Move the inner summation to the only term that care about x_k.
- Line 5: We define

$$\vec{q} := \begin{bmatrix} \sum_{x_k=M}^{N} P_{0,x_k} \\ \sum_{x_k=M}^{N} P_{1,x_k} \\ \vdots \\ \sum_{x_k=M}^{N} P_{M-1,x_k} \end{bmatrix}$$

To summarize, the pmf of T is given by

$$f_T(k) := \Pr(T = k) = \begin{cases} \sum_{i=M}^{N} \alpha_{0,i} & k = 0 \\ \vec{\alpha}_0' Q^{k-1} \vec{q} & k > 0 \end{cases}$$

1.4. Definition: A distribution with a pmf

$$f_T(k) := \Pr(T = k) = \begin{cases} \vec{\alpha}_0' Q^{k-1} \vec{q} & k > 0 \\ \sum_{i=M}^{N} \alpha_{0,i} & k = 0 \end{cases}$$

is called a **discrete phase-type distribution** and is typically denoted as

$$T \sim \text{DPH}_M(\vec{\alpha}_0', Q),$$

where M is the dimension of the transient space (i.e., the number of transient states), Q is the transient part of the transition matrix, and $\vec{\alpha}_0'$ is the transient part of the initial distribution. [a]

[a] Intuitively, we only care about the transient part because the absorbing part is really not that interesting.

1.5. Remark: It seems that the value $\Pr(T=0)$ and the column vector $\hat{q}$ are missing in the parameterization. However, they can be derived as follows:

$$\Pr(T=0) = \alpha_{0,M} + \cdots + \alpha_{0,N} = 1 - \sum_{i=0}^{M-1} \alpha_{0,i} = 1 - \vec{\alpha}_0' \cdot \mathbf{1}.$$

$$q_i = \sum_{j=M}^{N} P_{i,j} = 1 - \sum_{=0}^{M-1} P_{ij} = 1 - \sum_{j=0}^{M-1} Q_{ij} \implies \hat{q} = (I - Q) \cdot \mathbf{1}.$$

Thus, $\hat{\alpha}_0'$ and Q completely determines the distribution $\mathrm{DPH}_M(\vec{\alpha}_0', Q)$.

1.6. Example: Consider the transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{bmatrix}$$

where $\alpha+\beta+\gamma = 1$. Given the initial distribution $\vec{\alpha}_0$, we want to get the distribution of absorption time to $\{0,2\}$. Rearrange the state (by putting the transient state 1 to top), we have

$$P' = \begin{bmatrix} \beta & \alpha & \gamma \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \implies \begin{bmatrix} \beta & 1-\beta \\ 0 & I \end{bmatrix}.$$

Let $\vec{\alpha}_0' = \alpha_{0,1} = p = \Pr(X_0 = 1)$, the probability starting in the transient state 1. Then

$$\alpha_{0,0} + \alpha_{0,2} = 1 - p.$$

Observe that $\vec{\alpha}_0' = p$, $Q = \beta^{k-1}$, and $\vec{q} = (1-\beta)$ (all of these are scalars in this case). Thus,

$$f_T(k) = \begin{cases} p\beta^{k-1}(1-\beta) & k = 1, 2, \ldots \\ 1-p & k = 0 \end{cases}$$

This is known as a **zero-modified geometric distribution**. In particular, if $p = \beta$, this becomes the geometric distribution $\mathrm{Geo}(1-p) = \mathrm{DPH}_1(p,p)$, which counts the number of failures until the first success.

Section 2. DPH Cumulative Distribution Function

2.1. Remark: From now on, let us denote the transient part of the initial distribution by $\vec{\alpha}_0$, i.e., omitting the "prime" and reassign $\vec{\alpha}_0 \leftarrow \vec{\alpha}_0'$. This makes sense because we won't be using the absorbing part of the vector anyway.

2.2. Note: Let $T \sim \text{DPH}_M(\alpha_0, Q)$. For $k = 0, 1, \ldots,$

$$\begin{aligned}
F_T(k) = \Pr(T \le k) &= 1 - \Pr(T > k) \\
&= 1 - \sum_{n=k+1}^{\infty} \Pr(T = n) \\
&= 1 - \sum_{n=k+1}^{\infty} \vec{\alpha}_0 Q^{n-1} \vec{q} && \text{See remark below.} \\
&= 1 - \vec{\alpha}_0 Q^k (I - Q)^{-1} (I - Q) && \vec{q} = (I - Q) \cdot \mathbf{1} \\
&= 1 - \vec{\alpha}_0 Q^k \cdot \mathbf{1}
\end{aligned}$$

- Line 3: Substitute in $\Pr(T = n)$ found in the previous section.
- Line 4: Define $S := Q^k + Q^{k+1} + \cdots$. Now

$$SQ = Q^{k+1} + Q^{k+2} + \cdots \implies S - SQ = S(I - Q) = Q^k \implies S = Q^k (I - Q)^{-1}.$$

To summarize, the CDF of DPH is given by $\boxed{F_T(k) = 1 - \vec{\alpha}_0 Q^k \cdot \mathbf{1}.}$

2.3. Remark: What does this result give us? We focus on this:

$$\Pr(T > k) = \vec{\alpha}_0 Q^k \cdot \mathbf{1}.$$

By definition, $\Pr(T > k)$ is the probability that the absorption hasn't happened at time k (and these probabilities are encoded in $\vec{q}$). The RHS of this essentially counts the total probability of all possible paths that does not get in absorbing states before time k.

- $\vec{\alpha}_0$ means we start from a transient state.
- Q^k means we stay in transient states until time k.
- $\mathbf{1}$ just adds up all these probabilities together at time k.

2.4. Remark: Compare the cdf and the pmf. For $k \ge 1$,

- $f_T(k) = \Pr(T = k) = \vec{\alpha}_0 Q^{k-1} \vec{q}$.
- $F_T(k) = \Pr(T \le k) = \vec{\alpha}_0 Q^{k-1} \cdot \mathbf{1}$.

Observe that for pmf, we need to make sure that the chain will go from a transient state to an absorbing state at time k, but we don't have the such restriction for cdf. It can get absorbed any time it wants, as long as it is after time k.

2.5. Note: Since T is a non-negative integer-valued random variable, we have

$$\begin{aligned}\mathbb{E}[T] &= \sum_{k=0}^{\infty} \Pr(T > k) \\ &= \sum_{k=0}^{\infty} \vec{\alpha}_0 Q^k \cdot \mathbf{1} \\ &= \vec{\alpha}_0 \sum_{k=0}^{\infty} Q^k \cdot \mathbf{1} \\ &= \vec{\alpha}_0 (I - Q)^{-1} \cdot \mathbf{1}\end{aligned}$$

- Line 4: Recall from the previous page that

  $$S := \sum_{i=k}^{\infty} Q^i = Q^k (I - Q)^{-1}.$$

  Then

  $$S/Q^k = \sum_{i=k}^{\infty} \frac{Q^i}{Q^k} = \sum_{i=0}^{\infty} Q_i = (I - Q)^{-1}.$$

Recall that

$$(I - Q)^{-1} \cdot \mathbf{1} = \vec{q} = \begin{bmatrix} \sum_{x_k=M}^{N} P_{0,x_k} \\ \sum_{x_k=M}^{N} P_{1,x_k} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbb{E}[T \mid X_0 = 0] \\ \mathbb{E}[T \mid X_0 = 1] \\ \vdots \end{bmatrix} = g'$$

This results agrees with what we got previously using first-step analysis.

Section 3. Properties of DPH

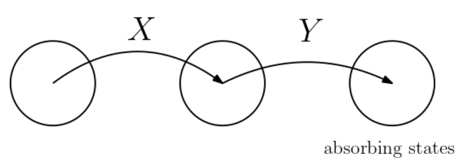
3.1. Motivation: In this section, we will show that the family of DPH is closed under several operations.

3.2. Proposition: *DPH is closed under independent sum.*

Proof. Let $X \sim \text{DPH}_m(\alpha_0, S)$ and $Y \sim \text{DPH}_n(\beta_0, T)$ with $X \perp\!\!\!\perp Y$. Define

$$s' = (I - S) \cdot \mathbf{1}$$
$$t' = (I - T) \cdot \mathbf{1}$$

We wish to show that $Z = X + Y$ is a DPH. By definition, it suffices to show that Z is the absorption time for some DTMC. Let us construct this DTMC with three parts. The time it takes to go from left to middle is X, the time it takes to go from middle to right is Y, and thus the total time it takes to get absorbed is $X + Y$.



Consider the following transition matrix:

$$P = \begin{array}{c} \\ X \\ Y \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{ccc} X & Y & \text{Abs} \end{array} \\ \begin{bmatrix} S & \vec{s}'\vec{\beta}_0 & \beta_{0,n}\vec{s}' \\ 0 & T & t' \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

- The three zero entries and the bottom-right identity are straightforward.
  - Starting from Y, we should not be able to visit X.
  - Starting from Abs, we should stay in Abs.
- [0][0]: S stores the probability of going from left to left. Correspond to Q in the definition.
- [1][1]: T stores the probability of going from middle to middle. Correspond to Q in the defn.
- [1][2]: Correspond to $\vec{q}$ in the definition.

It remains to justify the top center and top right entries.

- $\vec{s}'\vec{\beta}_0$: Since $\vec{s}' \in \mathbb{R}_{m\times 1}$ and $\beta_0 \in \mathbb{R}_{1\times n}$, the product is a $m \times n$ matrix. This can be thought of as the transient distribution of X then entering the initial transient distribution of Y.
- $\beta_{0,n} = 1 - \sum_{i=0}^{n-1} \beta_{0,i}$. This corresponds to the probability distribution of getting into the absorption state from X right away.

Intuitively, we can think of $\{Y, \text{Abs}\}$ as the absorbing states for X and Abs as the absorbing states for Y. The transient part of the new Markov Chain Z is

$$C = \begin{bmatrix} S & \vec{s}'\vec{\beta}_0 \\ 0 & T \end{bmatrix}.$$

The initial distribution of Z is

$$\gamma = (\alpha_0, \alpha_{0,m} \cdot \vec{\beta}_0).$$

where $\alpha \in \mathbb{R}_{1\times m}$ and $\alpha_{0,m} \cdot \vec{\beta}_0 \in \mathbb{R}_{1\times n}$. Together, they make up a $1 \times (m \times n)$ row vector.

We conclude that $Z \sim \mathrm{DPH}_{m+n}(\gamma, C)$. □

3.3. Proposition: *DPH is closed under mixture, i.e., given $X \sim \mathrm{DPH}_m(\vec{\alpha}_0, S), Y \sim \mathrm{DPH}_n(\vec{\beta}_0, T)$, and $X \perp\!\!\!\perp Y$, define*

$$Z = \begin{cases} X & \textit{with prob } p \\ Y & \textit{with prob } 1-p \end{cases}$$

and the choice between X and Y is independent of their values. Then Z is also a DPH.

Chapter 2

Continuous Time Markov Chain

> **3.4. Remark:** We will not do any review here. See my STAT-333 notes if you need a review on DTMCs.

Section 1. Introduction to CTMC

1.1. Motivation: We start with two questions:

- How long does this MC stay in state i?
- When it leaves the current state, how do we decide which state it will enter?

1.2. Note (Answer for 1): Let T_i be the random amount of time that the MC stays in state i (known as **sojourn time**). Then T_i follows an exponential distribution. The reason is that

$$\begin{aligned}\Pr(T_i > t+s \mid T_i > s) &= \Pr(\forall w \le u \le w+t+s : X(u) = i \mid \forall w \le u \le w+s : X(u) = i) \\ &= \Pr(\forall w+s < u \le w+t+s : X(u) = i \mid \forall w \le u \le w+s : X(u) = i) \\ &= \Pr(\forall w+s < u \le w+t+s : X(u) = i \mid X(w+s) = i; \forall w \le u < w+s : X(u) = i) \\ &= \Pr(\forall w+s < u \le w+t+s : X(u) = i \mid X(w+s) = i) \quad \text{Markov} \\ &= \Pr(\forall w < u \le w+t : X(u) = i \mid X(w) = i) \\ &= \Pr(T_i > t)\end{aligned}$$

This is the memoryless property. Hence, T_i follows an exponential distribution with parameter V_i:

$$T_i \sim \mathrm{Exp}(V_i).$$

In particular, each time the process enters a state i, the amount of time it spends there before going to another state is exponentially distributed with mean $1/V_i$; the mean sojourn time is $1/V_i$. Observe that the chain will stay longer at a state if V_i is small. In particular, if $V_i = 0$, then $\Pr(T_i > t) = 1$ for any t, so $T_i = \infty$, i.e., once you enter state i, you will never get out; state i is absorbing. In CTMC, saying $V_i = 0$ means state i is absorbing.

1.3. Note (Answer for 2): We claim that the transition probability, once the MC leaves i, does not depend on the sojourn time T_i. This follows from Markov's property, which tells us that the transition probability only depends on the current state. Thus, we can denote the transition probability that when the MC leaves i and goes to j by P_{ij}, which only depends on i and j. So how does a CTMC behave? To summarize, a CTMC stays in a state i for an exponential amount of time T_i, then jumps to another state j according to the probability P_{ij}, then stay in j for an exponential amount of time T_j, before jumping to another state, etc.

1.4. Note: $\{P_{ij}\}_{i,j\in S}$ is a transition matrix of a DTMC. However, note that we have one extra condition: $P_{i,i} = 0$. By definition,

$$\sum_{j\in S} P_{ij} = \sum_{j\ne i, j\in S} Pij = 1 \quad \forall i \in S.$$

The DTMC having $\{P_{ij}\}_{i,j\in S}$ as the transition matrix governs all the change of the states of the CTMC, but does not record sojourn times. It is called the **discrete skeleton** or **embedded DTMC** of the CTMC.

Section 2. Kolmogorov Backward and Forward Equations

2.1. Note: We are interested in how the function

$$\begin{aligned} P_{ij}(t) &= \Pr(X(t) = j \mid X(0) = i) \\ &= \Pr(X(t+s) = j \mid X(s) = i) \end{aligned}$$

changes as a function of t.

2.2. Note: Define $q_{ij} := V_i P_{ij}$, where V_i is the **intensity** going out of state i and P_{ij} is the probability of entering j from $i \neq j$. Note that

$$V_i = V_i \sum_{j \in S, j \neq i} P_{ij} = \sum_{j + i} q_{ij}.$$

For h small,

$$\begin{aligned} P_{ii}(h) &= \Pr(X(h) = i \mid X(0) = i) \\ &= \Pr(T_i > h) + \Pr(\text{at least 2 transitions happened between 0 and } h, X(h) = i \mid X(0) = i) \\ &= \Pr(T_i > h) + o(h) \\ &= e^{-V_i h} + o(h) \\ &= 1 - V_i h + o(h) + o(h) = 1 - V_i h + o(h) \\ &\implies \lim_{h \to 0} \frac{P_{ii}(h) - 1}{h} = -V_i. \end{aligned}$$

For the other case $i \neq j$,

$$\begin{aligned} P_{ij}(h) &= \Pr(X(h) = j \mid X(0) = i) \\ &= P_{ij}(\Pr(T_i \leq h) - \Pr(\text{at least two transitions})) + o(h) \\ &= P_{ij}(1 - e^{-V_i h} - o(h)) + o(h) \\ &= P_{ij}(V_i h - o(h) - o(h)) + o(h) \\ &= P_{ij} V_i h + o(h) = q_{ij} h + o(h) \\ &\implies \lim_{h \to 0} \frac{P_{ij}(h)}{h} = q_{ij}. \end{aligned}$$

2.3. Remark: We say $f(h)$ is a higher-order infinitesimal of h, denoted $f(h) = o(h)$, if

$$\frac{f(h)}{h} \to 0 \quad \text{as} \quad h \to 0.$$

The equality follows from A2 that

$$\Pr(T_i < h) = \Pr(\text{at least 1 transition happened}) = V_i h + o(h).$$

2.4. Note: Let $P(t) = \{P_{ij}(t)\}_{i,j \in S}$ be the transition matrix at time t. Then $P(0) = I$ since

$P_{ii}(0) = 1$ and $P_{ij}(0) = 0$ for $j \neq i$. Combining two equations from above, we have

$$P'(0) = \lim_{h \to 0} \frac{P(h) - P(0)}{h} = \begin{bmatrix} -V_0 & q_{01} & q_{02} \cdots & \\ q_{10} & -V_1 & q_{12} \cdots & \\ q_{20} & q_{21} & -V_2 \cdots & \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} =: R.$$

This matrix R is called the **(infinitesimal) generator** of the CTMC. It combines the information of sojourn times $\{V_i\}_{i \in S}$ and transition probabilities $\{P_{ij}\}_{i,j \in S}$. In particular, if we write $R = \{R_{ij}\}_{i,j \in S}$, we have

$$R_{ij} = \begin{cases} -V_i & i = j \\ q_{ij} = V_i P_{ij} & i \neq j \end{cases} \implies v_i = -R_{ii},\ Pij = -\frac{R_{ij}}{R_{ii}}.$$

As a remark, the row sum of R are always 0:

$$\sum_{j \neq i} V_i P_{ij} - V_i = V_i - V_i = 0.$$

2.5. Note: CK Equations still hold in continuous time.

$$\begin{aligned} P_{ij}(t+s) &= \Pr(X(t+s) = j \mid X(0) = i) \\ &= \sum_{k \in S} \Pr(X(t+s) = j \mid X(t) = k, X(0) = i) \cdot \Pr(X(t) = k \mid X(0) = i) \\ &= \sum_{k \in S} \Pr(X(t+s) = j \mid X(t) = k) \cdot \Pr(X(t) = k \mid X(0) = i) \\ &= \sum_{k \in S} P_{kj}(s) P_{ik}(t) \end{aligned}$$

In matrix notation, $P(t+s) = P(t)P(s)$.

In particular, $P(t+h) = P(h)P(t)$. Subtract $P(t)$ from both sides, $P(t+h) - P(t) = (P(h) - I)P(t)$,

$$\lim_{h \to 0} \frac{P(t+h) - P(t)}{h} = \lim_{h \to 0} \frac{P(h) - P(0)}{h} \cdot P(t) \implies P'(t) = P'(0)P(t) = RP(t).$$

This is called the Kolmogorov Backward Equation.

2.6. Note: Similarly, $P(t+h) = P(t)P(h)$.

$$\begin{aligned} \lim_{h \to 0} \frac{P(t+h) - P(t)}{h} &= P(t) \lim_{h \to 0} \frac{P(h) - P(0)}{h} \\ P'(t) &= P(t) \cdot P'(0) = P(t)R \end{aligned}$$

This is called the Kolmogorov Forward Equation. In entry-wise form,

- Backward: $P'_{ij}(t) = \sum_{k \in S} R_{ik} P_{kj}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - V_i P_{ij}(t)$.
- Forward: $P'_{ij}(t) = \sum_{k \neq j} P_{ik}(t) q_{kj} - V_j P_{ij}$.

2.7. Remark: Note that we interchanged the order of a limit and a summation in the derivation above, e.g.,

$$\lim_{h\to 0}\left[\frac{1}{h}(P(h)-I)P(t)\right] = \left[\lim_{h\to 0}\frac{P(h)-P(0)}{h}\right]P(t).$$

This can be justified if:

- The state space S is finite. This is clear since we are not dealing with limit of partial sums.
- For backward equation, this is always valid. Check bounded convergence theorem. (The forward equation holds only if "explosion" does not happen). In this sense, backward equation is more reliable and fundamental. However, forward equation is usually easier to solve.

2.8. Note: Now we know that $P(t)$ satisfies the matrix differential equation $P'(t) = RP(t)$ (backward) with initial condition $P(0) = I$. If everything were scalar, we should get e^{tR} as the solution. Now with everything being matrix, what should we get? We will still get $P(t) = e^{tR}$, defined as follows.

$$e^{tR} = I + tR + \frac{t^2R^2}{2} + \cdots = \sum_{n=0}^{\infty}\frac{t^n}{n!}R^n.$$

One can show that the sum above always converges. We now wish to show that such defined e^{tR} is the solution of the matrix differential equation $P'(t) = RP(t)$.

2.9. (Cont'd):

$$\begin{aligned}\frac{d}{dt}e^{tR} &= \frac{d}{dt}\left(I + \sum_{n=1}^{\infty}\frac{t^n}{n!}R^n\right)\\ &= \sum_{n=1}^{\infty}\frac{d}{dt}\frac{t^n}{n!}R^n\\ &= \sum_{n=1}^{\infty}\frac{t^{n-1}}{(n-1)!}R^n\\ &= R\sum_{n=1}^{\infty}\frac{t^{n-1}}{(n-1)!}R^{n-1}\\ &= Re^{tR}\end{aligned}$$

Also,

$$e^{tR}\Big|_{t=0} = e^0 = I.$$

Thus, $P(t) = e^{tR}$ solves the backward equation. Similarly, we can show that $P(t) = e^{tR}$ solves the forward equation.

2.10. (Cont'd): But how can we calculate this infinite matrix series? In general, $P(t) = e^{tR}$ is not easy to calculate. In particular, this is not an entry-wise computation:

$$e^{tR} \neq \begin{bmatrix} e^{tR_{00}} & e^{tR_{01}} & \cdots \\ e^{tR_{10}} & e^{tR_{11}} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

However, if the matrix R is diagonalizable, that is, there exists an invertible matrix B such that $R = BDB^{-1}$ with $D = \text{diag}(d_0, d_1, \ldots)$, then

$$\begin{aligned} e^{tR} &= \sum_{n=0}^{\infty} \frac{t^n}{n!} R^n \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \left(BDB^{-1}\right)^n \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} BD^n B^{-1} \\ &= B \left(\sum_{n=0}^{\infty} \frac{t^n}{n!} D^n \right) B^{-1} \\ &= Be^{tD} B^{-1} \end{aligned}$$

For diagonal D,

$$\begin{aligned} e^{tD} &= \sum_{n=0}^{\infty} \frac{t^n}{n!} D^n \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \begin{bmatrix} d_0^n & & \\ & d_1^n & \\ & & \ddots \end{bmatrix} \\ &= \begin{bmatrix} \sum_{n=0}^{\infty} \frac{t^n}{n!} d_0^n & & \\ & \sum_{n=0}^{\infty} \frac{t^n}{n!} d_1^n & \\ & & \ddots \end{bmatrix} \begin{bmatrix} e^{d_0 t} & & \\ & e^{d_1 t} & \\ & & \ddots \end{bmatrix} \end{aligned}$$

2.11. (Cont'd): To summarize, there are three general methods to get $P(t)$:

(1). Solve the backward equation.

(2). Solve the forward equation.

(3). Diagonalization of R: more than 2 states, this method is preferable.

Section 3. Birth and Death Processes

3.1. Motivation: Think $X(t)$ as the number of individuals (population) in a system at time t. At any time, the population can only (1) increase by 1, (2) decrease by 1, or (3) remain the same. This implies that if $|i-j|>1$, then $P_{ij=0}$ and $q_{ij}=0$. The generator R in this case takes the special tri-diagonal form as below. Since the row sums of R always equal to zero, we may introduce a different set of parameters λ and μ, where λ's are called the **birth rates** and μ's are called the **death rates**.

$$R = \begin{bmatrix} -V_0 & V_0 & 0 & 0 & \cdots \\ q_{10} & -V_1 & q_{12} & 0 & \cdots \\ 0 & q_{21} & -V_2 & q_{23} & \cdots \\ 0 & 0 & q_{32} & -V_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$
$$= \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

For $i=0$, $P_{0,1}=1$. For $i\geq 1$, we have

$$\begin{matrix} \lambda_i = q_{i,i+1} = V_i P_{i,i+1} \\ \mu_i = q_{i,i-1} = V_i P_{i,i-1} \end{matrix} \implies \begin{cases} P_{i,i+1} & = \frac{\lambda_i}{\lambda_i+\mu_i} \\ P_{i,i-1} & = \frac{\mu_i}{\lambda_i+\mu_i} \end{cases}$$

3.2. Intuition: When the system is in state i, the next (potential) birth happens after an exponential amount of time with intensity (rate) λ_i, and the next (potential) death happens after an exponential amount of time with intensity (rate) μ_i. The smaller one wins (i.e., happens), changing the system to the next state. Recall that if $T_i^{(b)} \sim \text{Exp}(\lambda_i), T_i^{(d)} \sim \text{Exp}(\mu_i)$ independent, then

(1). $\Pr(T_i^{(b)} < T_i^{(d)}) = \frac{\lambda_i}{\lambda_i+\mu_i} = P_{i,i+1}$.

(2). $\Pr(T_i^{(d)} < T_i^{(b)}) = \frac{\mu_i}{\lambda_i+\mu_i} = P_{i,i-1}$.

(3). $\min(T_i^{(b)}, T_i^{(d)})$, which is the time until the system state changes, follows $\text{Exp}(\lambda_i+\mu_i)$. This is why $v_i = \lambda_i+\mu_i$ for $i\geq 1$ and $v_0=\lambda_0$ since there is no μ_0.

Because of the special structure, for birthday and death processes, it is always easier to directly write R rather than writing $\{v_i\}$ and $\{P_{ij}\}$ first.

3.3. Note: Some special cases of birth and death processes:

- Pure death process. $\lambda_0=\lambda_1=\cdots=0$. No birth at all.
- Pure birth process. $\mu_1=\mu_2=0$. No death at all.

- Poisson process. $\mu_0 = 0$ and $\lambda_i = \lambda > 0$. The time between births are iid exponentials with the same intensity λ regardless of the state i. This is the counting process of exponential events, a Poisson process.
- Yule process. $\mu_i = 0$ and $\lambda_i = i\lambda$ for $i = 0, 1, \ldots$. The birth rate is proportional to the current population.
- Linear growth model. $\lambda_i = i\lambda$. $\mu_i = i\mu$. Individuals are independent and have the same birth rate λ and death rate μ.
- M/M/S Queue. See later.

3.4. Note: Consider the forward equations for birth and death processes $P'(t) = P(t)R$. The ith row is given by

$$\begin{aligned}
&(P'_{i0}(t), P'_{i1}(t), \cdots, P'_{i,j-1}(t), P'_{i,j}(t), P'_{i,j+1}(t), \ldots) \\
&= (P_{i0}(t), P_{i1}(t), \cdots, P_{i,j-1}(t), P_{i,j}(t), P_{i,j+1}(t), \cdots) \begin{bmatrix} -\lambda_0 & \lambda_0 & & & & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & & \\ & \mu_2 & -(\lambda_2+\mu_2) & \ddots & & \\ & & \mu_3 & \ddots & \lambda_{j-1} & \\ & & & \ddots & -(\lambda_j+\mu_i) & \ddots \\ & & & & \mu_{j+1} & \ddots \\ & & & & & \ddots \end{bmatrix}
\end{aligned}$$

Entry-wise, we have

$$\begin{cases} P'_{i0}(t) = -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t) \\ P'_{ij}(t) = \lambda_{j-1} P_{i,j-1}(t) - (\lambda_j + \mu_j) P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t), \quad j = 1, 2, \cdots \end{cases}$$

3.5. (Cont'd): For the special case of pure birth process, where $\mu_i = 0$ for $i = 1, 2, \ldots$, we have $P_{ij}(t) = 0$ for $j < i$. If $j = i$, we have

$$\begin{cases} P'_{00}(t) = -\lambda_0 P_{00}(t) \\ P'_{ii}(t) = \lambda_{i-1} \underbrace{P_{i,i-1}(t)}_{=0} - \lambda_i P_{ij}(t) = -\lambda_i P_{ii}(t), i = 1, 2, \cdots \end{cases}$$

The initial condition is $P_{ii}(0) = 1$. The solution of this equation is clearly

$$P_{ii}(t) = e^{-\lambda_i t}.$$

Intuitively, since the CTMC is a pure birth process,

$$\begin{aligned} P_{ii}(t) &= \Pr(X(t) = i \mid X(0) = i) \\ &= \Pr(T_i > t) \\ &= e^{-\lambda_i t} \end{aligned}$$

For $j > i$, this becomes

$$\begin{aligned} P'_{ij}(t) &= \lambda_{j-1} P_{i,j-1}(t) - \lambda_j P_{ij}(t) \\ P'_{ij}(t) + \lambda_j P_{ij}(t) &= \lambda_{j-1} P_{i,j-1}(t) \\ e^{\lambda_j t}\left(P'_{ij}(t) + \lambda_j P_{ij}(t)\right) &= e^{\lambda_j t} \lambda_{j-1} P_{i,j-1}(t) \\ \left(e^{\lambda_j t} P_{ij}(t)\right)' &= e^{\lambda_j t} \lambda_{j-1} P_{i,j-1}(t) \\ e^{\lambda_j t} P_{ij}(t) \Big|_{t=0}^{s} &= \int_0^s e^{\lambda_j t} \lambda_{j-1} P_{i,j-1}(t) dt \\ e^{\lambda_j s} P_{ij}(s) - P_{ij}(0) &= \lambda_{j-1} \int_0^s e^{\lambda_j t} P_{i,j-1}(t) dt \\ P_{ij}(s) &= \lambda_{j-1} e^{-\lambda_j s} \int_0^s e^{\lambda_j t} P_{i,j-1}(t) dt \end{aligned}$$

This yields a recursive formula for $P_{ij}(t)$ with $i < j$.

Section 4. Classification of States

4.1. Note: We have seen how to find $P(t) = \{P_{ij}(t)\}_{i,j\in S}$. As in the discrete case, once we have $P(t)$, it is easy to express the distribution of $X(t)$:

$$\begin{aligned}(\alpha_t)_j := \Pr(X(t) = j) &= \sum_{i\in S} \Pr(X(t) = j \mid X(0) = i) \cdot \Pr(X(0) = i) \\ &= \sum_{i\in S} \alpha_{0,i} P_{i,j}(t) \\ &= (\alpha_0 P(t))_j\end{aligned}$$

In words, the distribution of $X(t)$ is determined by the initial distribution α_0 and the transition matrix $P(t)$. Since $P(t)$ is determined by R, we conclude that the distribution of $X(t)$ depends on α_0 and R.

4.2. Note: R to $P(t)$ in CTMC is like P to $P^{(n)}$ in DTMC.

4.3. Note: Let f be a function defined on the state space $f : S \to \mathbb{R}$. Then

$$\begin{aligned}\mathbb{E}[f(X(t))] &= \sum_{j\in S} f(j) \Pr(X(t) = j) \\ &= \alpha_t f^T \\ &= \alpha_0 P(t) f^T\end{aligned}$$

where α_0 is a row vector and f^T is a column vector.

4.4. Note: Recall that if a CTMC $\{X(t)\}$ is only observed when the state changes, the resulting process is a DTMC, denoted as $\{X_n\}_{n=0,1,\ldots}$, known as the **discrete skeleton / embedded DTMC** of the CTMC $\{X(t)\}_{t\geq 0}$. The transition matrix of this DTMC is $\{P_{ij}\}_{i,j\in S}$, where $P_{ij} = \Pr(X(T_i) = j \mid X(0) = i)$. If $V_i = 0$, i is absorbing, then define $P_{ii} = 1$ and $P_{ij} = 0$ for all $j \neq i$. The question is, what do these concepts in discrete time become?

4.5. Definition: State j is said to be **accessbile** from i (or i is said to **communicate to** j) if

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

for some $t \geq 0$. Denote this as $i \to j$.

4.6. Definition: i and j are said to **communicate** with each other if $i \to j$ and $j \to i$. Denote it as $i \leftrightarrow j$.

4.7. Remark: By definition, a state i always communicates with itself, which is different from

DTMC setting (no self-loop). Why? Because when $V_i > 0$, the probability of the Sojourn time at state i being 0 is never 0 (tail probability of the exponential distribution).

4.8. It turns out that we have something simple for communication.

4.9. Proposition: *For $i \neq j$ in a CTMC, $i \to j$ iff $i \to j$ in the discrete skeleton.*

Proof. A3. □

4.10. Definition: A set $C \subseteq S$ is called a **communicating class** if for all $i, j \in C$, $i \leftrightarrow j$, and for all $i \in C$, $j \notin C$, $i \not\leftrightarrow j$.

4.11. Definition: A CTMC is called **irreducible** if all its states are in the same class.

4.12. Clearly, a CTMC is irreducible iff its discrete skeleton is irreducible by the preceding proposition.

4.13. Note: Let R_{ii} be the amount of (continuous) time until the CTMC revisits state i given $X(0) = i$. Define $R_{ii} = \infty$ if it never revisits i.

4.14. Definition: The state i is recurrent if either $\Pr(R_{ii} < \infty) = 1$ or i is absorbing. It is transient if $\Pr(R_{ii} = \infty) > 0$.

4.15. Remark: A CTMC revisits a (non-absorbing) state i iff its discrete skeleton revisits i. A state i is recurrent iff it is recurrent in the DS of the CTMC. As a result, recurrence/transience are class properties.

4.16. Note: An irreducible CTMC is recurrent/transient iff all its states are recurrent/transient. An irreducible CTMC is recurrent/transient iff its discrete skeleton is recurrent/transient.

4.17. Note: There is no concept of periodicity in continuous time. That's because if $\Pr(X(t) = j \mid X(0) = i) > 0$ for some $t \geq 0$, then $\Pr(X(t) = j \mid X(0) = i) > 0$ for all $t \geq 0$ because the sojourn times are exponential which can be arbitrarily large or small.

Section 5. Positive and Null Recurrence

5.1. Motivation: In the previous section, we have seen that *accessibility*, *communication*, *communicating class*, *recurrence*, and *transience* behave in the same way for CTMC as in the discrete skeleton, except for that each state communicates with itself in CTMC. As a result, *recurrence* and *transience* are still class properties. We also noted that there is no notion of *periodicity* for CTMC. You may ask, is there anything different between CTMC and its discrete skeleton? The only pair of concepts we haven't discussed so far is *positive* and *null recurrence*.

5.2. Definition: Let R_{ii} be the amount of (continuous) time until the MC revisits state i, given $X(0) = i$. A state i is called **positive recurrent** if $\mathbb{E}(R_{ii}) < \infty$ or state i is absorbing; i is called **null recurrent** if it is recurrent but $\mathbb{E}(R_{ii}) = \infty$.

5.3. Intuition: A state i is positive recurrent if the expected time it takes the chain to return to state i is finite; it is null recurrent if it is recurrent but the expected time it takes the chain to return to state i is infinite.

5.4. Is positive and null recurrence always the same for a CTMC and its discrete skeleton? Intuitively, recurrence and transience are the same for both a CTMC and its discrete skeleton because thse two concepts only concern whether a state will be revisited or not; the time factor isn't relevant here. However, positive and null recurrence take the time factor into consideration, and this *time* factor is something that discrete skeleton does not reflect. Again, the discrete skeleton contains only information about the state change and ignores the information about time. This hints that the positive and null recurrence criterion might be different between a CTMC and its discrete skeleton.

5.5. Example: Consider a DTMC with one-step transition matrix. Intuitively, at each step, you have q probability to go back to state 0 and p probability to move right one step, except at state 0 where you always go to state 1. In other words, as long as the chain is not in state 0, in each step, it has a fixed probability of returning to state 0. Thus, we can model the time it takes to go back to state 0 with a geometric random variable, because you just repeatedly try something (independent with the same probability) until a *success*.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ q & 0 & p & 0 & \cdots \\ q & 0 & 0 & p & \cdots \\ \cdots & 0 & 0 & 0 & \ddots \end{bmatrix}$$

Thus, 0 is recurrent and the amount of time it takes for the chain to leave and revisit state 0 is

$$R_{00} = 1 + \text{Geo}(q) \implies \mathbb{E}[R_{00}] = 1 + \frac{1}{q} < \infty.$$

Finally, since the chain is irreducible, the entire MC is positive recurrent.

5.6. Example: Now consider a CTMC whose discrete skeleton has the transition matrix above, because our goal is to compare and contrast how positive and null recurrence behave for a CTMC and its discrete skeleton.

Let $\{X(t)\}_{t\geq 0}$ be a CTMC taking the above DTMC as its discrete skeleton, i.e., with the same $P = \{P_{ij}\}_{i,j\in S}$ for its parameterization.

Let W be the number of transitions to return to 0. Note this W plays the role of R_{00} in the discrete setting, but it is no longer the R_{00} in the continuous setting.

Let T_i be the amount of (continuous) time the MC stays at state i (before visiting state $i+1$), i.e., the sojourn times of state i. For example, T_2 stores the amount of time the MC states at state 2 before visiting state 3. Define T_{W-1} to be the amount of time the MC stays at state $W-1$ before visiting state 0.

Then the total time for the MC to revisit state 0 is given by

$$R_{00} = T_0 + T_1 + \cdots + T_{W-1}.$$

We know that $T_i \sim \text{Exp}(v_i)$ independently. Then

$$\begin{aligned}
\mathbb{E}(R_{00}) &= \mathbb{E}\left(\sum_{i=0}^{W-1} T_i\right) \\
&= \mathbb{E}\left(\sum_{i=0}^{\infty} T_i \mathbf{1}_{\{W>i\}}\right) && \text{use indicator to replace upper bound} \\
&= \sum_{i=0}^{\infty} \mathbb{E}[T_i \mathbf{1}_{\{W>i\}}] && \text{monotone convergence theorem} \\
&= \sum_{i=0}^{\infty} \mathbb{E}[T_i]\mathbb{E}[\mathbf{1}_{\{W>i\}}] && X \perp\!\!\!\perp Y \implies \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \\
&= \sum_{i=0}^{\infty} \mathbb{E}[T_i]\Pr(W>i) && \text{defn of indicator variables}
\end{aligned}$$

Since $W = 1 + \text{Geo}(q)$,

$$\Pr(W > i) = (1-q)^{i-1} = p^{i-1}.$$

Combined with $\mathbb{E}[T_i] = 1/v_i$, we get

$$\mathbb{E}[R_{00}] = \sum_{i=0}^{\infty} \frac{1}{v_i} p^{i-1}.$$

By choosing a sequence of v_i which decreases to zero "fast enough", we can always make this expectation (the sum) to infinity. For example, taking $1/v_i = 1/p^{i-1}$, we get

$$\mathbb{E}[R_{00}] = \sum_{i=0}^{\infty} 1 = \infty.$$

5.7. Intuition: We conclude this example with some intuition. By letting the sojourn time increase fast enough as the CTMC goes far from zero, we can have different results on positive and null recurrent states between the discrete skeleton and the CTMC. In this example, the discrete skeleton is positive recurrent whereas the CTMC is null recurrent.

5.8. Remark: Positive and null recurrence are still class properties. As in the discrete case, we can prove that all the states in the same class must be positive recurrent/null recurrent/transient at the same time.

Section 6. Stationary Distribution

6.1. Definition: Let $\{X(t)\}_{t\geq 0}$ be a CTMC. A row vector $\pi = (\pi_0, \pi_1, \ldots)$ with $\pi_i \geq 0$ for all $i \in S$ is called a **stationary distribution** of $\{X(t)\}_{t\geq 0}$ if it satisfies the following two conditions:

- **Stationarity**: $\pi = \pi \cdot P(t)$ for all $t \geq 0$.
- **Normalization**: $\pi \cdot \mathbf{1} = \sum_{i\in S} \pi_i = 1$.

6.2. Intuition: Such a distribution is called **stationary** by the same reason as in the discrete case: If we start the CTMC from the initial distribution π (which is stationary), then the distribution of $X(t)$ will always be π for any $t \geq 0$. In other words, we have

$$\forall t \geq 0 : \Pr(X(t) = j) = \pi_j.$$

Indeed, $\alpha_t = \alpha_0 \cdot P(t)$. Taking $\alpha_0 = \pi$, we get $\pi = \pi \cdot P(t)$ which holds by definition.

6.3. In contrast to the discrete setting where we can find the stationary distribution by definition, working with definition in the continuous setting is much more complex, because $P(t) = e^{tR}$ is often messy to work with and we want this system of equations to hold for all $t \geq 0$. Is there a better way to find π? In particular, can we find π without messing with $P(t)$?

6.4. Note: Let $t \geq 0$ and let π be a stationary distribution, so $\pi = \pi \cdot P(t)$. Then

$$\begin{aligned}
\pi \cdot I &= \pi \cdot P(t) \\
\pi \cdot (P(t) - I) &= \mathbf{0} \\
\pi \frac{P(t) - I}{t} &= \mathbf{0} \\
\lim_{t\to 0} \pi \frac{P(t) - I}{t} &= \mathbf{0} \\
\pi \cdot \lim_{t\to 0} \frac{P(t) - I}{t} &= \mathbf{0} \\
\pi \cdot R &= \mathbf{0}
\end{aligned}$$

In words, any stationary distribution π must satisfy the equation $\pi \cdot R = \mathbf{0}$.

6.5. (Cont'd): On the other hand, assume the initial distribution is π satisfying $\pi \cdot R = 0$.

$$\begin{aligned}
(\alpha_t)' &= (\alpha_0 \cdot P(t))' \\
&= \alpha_0 \cdot (P(t))' \\
&= \alpha_0 \cdot R \cdot P(t) \\
&= \pi \cdot R \cdot P(t) && \alpha_0 = \pi \\
&= 0 \cdot P(t) = 0
\end{aligned}$$

Since the derivative of the distribution is zero, it is not changing over time. Thus, π is stationary.

6.6. (Cont'd): We conclude that π is stationary iff

- $\pi \cdot R = \mathbf{0}$
- $\pi \cdot \mathbf{1} = 1$

This is much easier to use compared to the one with $P(t)$.

6.7. Note: We are now interested in the relationships between π of a CTMC and the stationary distribution ψ for its discrete skeleton. We know the following two equations hold.

- Continuous: $\pi R = \mathbf{0}$
- Discrete: $\psi P = \psi$.

Let us expand the expression for the CTMC and consider the jth component of $\pi \cdot R$.

$$
\begin{aligned}
(\pi R)_j &= [\pi_0, \pi_1, \cdots] \begin{bmatrix} -v_0 & q_{01} & q_{02} & \cdots & q_{0j} & \cdots \\ q_{10} & -v_1 & q_{12} & \cdots & q_{1j} & \cdots \\ q_{20} & q_{21} & -v_3 & \cdots & q_{2j} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix} \\
&= \pi_0 q_{0j} + \pi_1 q_{1j} + \cdots + \pi_{j-1} q_{j-1,j} - \pi_j v_j + \pi_{j+1} q_{j+1,j} + \cdots
\end{aligned}
$$

Since $\pi R = \mathbf{0}$, the above expression equals 0, which allows us to write

$$
\begin{aligned}
\pi_j v_j &= \pi_0 q_{0j} + \pi_1 q_{1j} + \cdots + \pi_{j-1} q_{j-1,j} + \pi_{j+1} q_{j+1,j} + \cdots \\
&= \sum_{i \in S, i \neq j} \pi_i q_{ij} \\
\pi_j v_j &= \sum_{i \in S, i \neq j} \pi_i v_i P_{ij} \qquad\qquad q_{ij} = v_i P_{ij}
\end{aligned}
$$

Next, we look at the expression for the discrete skeleton. Since $P_{jj} = 0$ for all j, we have

$$
\begin{aligned}
\psi_j &= \sum_{i \in S} \psi_i P_{ij} \\
&= \sum_{i \in S, i \neq j} \psi_i P_{ij}.
\end{aligned}
$$

Observe that $\{\psi_j\}_{j \in S}$ satisfy the same system of equations as $\{\pi_j v_j\}_{j \in S}$! Thus,

- If π is a stationary distribution for the CTMC, then
$$\psi = \{\psi_j = \pi_j v_j\}_{j \in S}$$
satisfies the stationary condition for the discrete skeleton.
- If ψ is a stationary distribution for the discrete skeleton, then
$$\pi = \left\{ \pi_j = \frac{\psi_j}{v_j} \right\}_{j \in S}$$
satisfies the stationary condition for the CTMC.

Don't forget that we still need to satisfy the normalization condition! The stationary condition itself only gives us the proportion between probabilities and we need to normalize the result to obtain the final numbers. Starting with

$$\pi_j \propto \frac{\psi_j}{v_j}, \qquad \sum_{j\in S} \pi_j = 1,$$

we divide the numerator by the sum of all values to obtain the normalized result:

$$\pi_j = \frac{\psi_j/v_j}{\sum_{j\in S} \psi_j/v_j}$$

Similarly, starting with

$$\psi_j \propto \pi_j v_j, \qquad \sum_{j\in S} \psi_j = 1,$$

the normalized result is given by

$$\psi_j = \frac{\pi_j v_j}{\sum_{j\in S} \pi_j v_j}.$$

Of course, we are assuming both sums in the denominator are finite. From this derivation, we note that π and ψ are in general not the same, i.e., the CTMC and its discrete skeleton usually do not share the same stationary distribution; instead, there exists a relationship between between the stationary distributions. Below we explain why we have a factor of v_j and what role it plays here.

6.8. Intuition: Like in the discrete case, the stationary distribution/probability π_j is the *long-run fraction of time* that the CTMC stays in state j. On the other hand, ψ_j is the *long-run fraction of steps* that the MC spends in state j in the discrete skeleton. For the CTMC, each time the discrete skeleton visits state j, it will stay in j for an exponential amount of time with mean $1/v_j$. Thus, the long-run behaviour of the CTMC needs to be weighted by the mean sojourn time $1/v_j$. This explains why we have

$$\pi_j \propto \frac{\psi_j}{v_j}.$$

6.9. Remark:

- Recall if an irreducible DTMC has a stationary distribution, then the stationary distribution is unique. Combined with the relation derived here, we see that if the CTMC is irreducible and both π (stationary distribution for the CTMC) and ψ (stationary distribution for the discrete skeleton) exist, then the uniqueness of ψ implies the uniqueness of π. (A stronger result will come later.)
- It is possible that one of π, ψ exists while the other does not. For example, ψ exists but π does not if $\sum_{i\in S} \frac{\psi_i}{v_i}$ does not converge, which happens when v_i decreases to zero fast, e.g., take $v_i = \psi_i$ for an infinite state space S. With the denominator for normalization become infinity, we conclude that π does not exist.

6.10. Note: Let us rewrite the stationarity condition $\pi R = \mathbf{0}$. The jth component for this system of equations is

$$\begin{aligned} -v_j\pi_j + \sum_{i\neq j} q_{ij}\pi_i &= 0 \\ \implies \pi_j v_j &= \sum_{i\neq j} \pi_i q_{ij} \\ \pi_j v_j &= \sum_{i\neq j} \pi_i v_i P_{ij} \end{aligned}$$

Recall v_j is the parameter that governs the sojourn time; a large v_j implies we will leave state j fast and a small v_j implies we will stay at j relatively longer. Intuitively, you could view v_j as the speed at which we will leave state j, given that we are currently at state j. Combined with π_j which is the probability that we are currently in state j, we see the LHS represents the *rate* at which the CTMC leaves state j.

Now look at RHS. π_i is the probability that we are current at state i; q_{ij} is the speed where we go to state j from state i. Thus, the RHS is the total rate going from other states getting to state j.

In other words, LHS is the "probability flow" out of state j and the RHS is the "probability flow" into state j. Both are under the stationary distribution. Hence, $\pi R = \mathbf{0}$ simply says that the "probability flow" leaving state j should be the same as the "probability flow" entering state j.

6.11. Example: Consider an irreducible CTMC with generator

$$R = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

and stationary distribution $\pi = (\pi_0, \pi_1)$. This can be viewed as a generalization of A3Q4. To find π, we need to solve

$$\begin{aligned} \pi \cdot R &= \mathbf{0} \\ \pi \cdot \mathbf{1} &= 1 \end{aligned}$$

The first equation gives us the following system:

$$\begin{aligned} -\alpha\pi_0 + \beta\pi_1 &= 0 \\ \alpha\pi_0 - \beta\pi_1 &= 0 \end{aligned}$$

Note this is a linearly dependent system; one can be derived from the other. Thus, the stationary distribution gives us the proportion but not the final numbers:

$$\frac{\pi_0}{\pi_1} = \frac{\beta}{\alpha}.$$

Now use the normalization constraint, we get

$$\pi_0 = \frac{\beta}{\alpha+\beta}, \qquad \pi_1 = \frac{\alpha}{\alpha+\beta}.$$

We see that π exists and is unique.

Next, we try to find the limiting distribution. It is easy to check (by verifying the backward equation)

$$P(t) = \begin{bmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t} \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t} \end{bmatrix}$$

Taking $t \to \infty$, we get

$$\lim_{t\to\infty} P(t) = \begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix}$$

Note that the limiting distribution probability $\lim_{t\to\infty} P_{ij}(t)$ does not depend on i and is the same as the unique stationary distribution π. We can write

$$T_0 \sim \text{Exp}(\alpha), \quad T_1 \sim \text{Exp}(\beta), \quad R_{00} = T_0 + T_1.$$

Thus,

$$\frac{\mathbb{E}\left(T_0\right)}{\mathbb{E}\left(R_{00}\right)} = \frac{1/\alpha}{1/\alpha + 1/\beta} = \frac{\beta}{\alpha+\beta} = \pi_0$$

Similarly,

$$\pi_1 = \frac{\mathbb{E}(T_1)}{\mathbb{E}(R_{11})}.$$

Section 7. Birth and Death Processes Continued

7.1. Note: We look at classification of birth and death processes. Recall the generator matrix

$$R = \begin{bmatrix} -\lambda_0 & \lambda_0 & & & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

We assume the CTMC is irreducible, i.e., $\forall i : \lambda_i, \mu_i > 0$.

The balance equations are:

- State 0: $\pi_0\lambda_0 = \pi_1\mu_1$.
- State 1: $\pi_1(\lambda_1+\mu_1) = \pi_0\lambda_0 + \pi_2\mu_2$.
- State 2: $\pi_2(\lambda_2+\mu_2) = \pi_1\lambda_1 + \pi_3\mu_3$.
- $\cdots$
- State $n-1$: $\pi_{n-1}(\lambda_{n-1}+\mu_{n-1}) = \pi_{n-2}\lambda_{n-2} + \pi_n\mu_n$.

Equation 0 gives

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

(0) + (1) gives

$$\pi_1\lambda_1 = \pi_2\mu_2 \implies \pi_2 = \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_0\lambda_1}{\mu_1\mu 2}\pi_0.$$

Repeating this procedure, (0) + (1) + $\cdots$ + (n):

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

Consider state n (and $n+1, n+2, \cdots$). The only flow going into this part is $\pi_{n-1}\lambda_{n-1}$; the only flow going out of this part is $\pi_n\mu_n$. Thus, we need $\pi_{n-1}\lambda_{n-1} = \pi_n\mu_n$.

Balance equations give us proportions of π_i. Now using normalization:

$$\begin{aligned}\sum_{n=0}^{\infty}\pi_n =& 1 \implies \pi\left(1+\frac{\lambda_0}{\mu_1}+\cdots+\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}+\cdots\right) = 1\\ \pi_0 =& \frac{1}{1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}}\\ \pi_n =& \frac{\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}}{1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}}\end{aligned}$$

Thus, a stationary distribution exists iff

$$\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}=\sum_{n=1}^{\infty}\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}<\infty$$

To recapitulate, a necessary and sufficient condition for an irreducible birth and death process to be positive recurrent is

$$\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}=\sum_{n=1}^{\infty}\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}<\infty$$

7.2. Note: If the above condition does not hold, how can we know whether the CTMC is null recurrent or transient? Consider the discrete skeleton with the following transition matrix:

$$P=\begin{bmatrix} 0 & 1 & & & \\ \frac{\mu_1}{\lambda_1+\mu_1} & 0 & \frac{\lambda_1}{\lambda_1+\mu_1} & & \\ & \frac{\mu_2}{\lambda_2+\mu_2} & 0 & \frac{\lambda_2}{\lambda_2+\mu_2} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

Define $f_{n0}=\Pr(\text{MC ever (re)visits state } 0 \mid X(0)=n)$. Then $f_{00}=1$ iff 0s recurrent iff the MC is recurrent. Otherwise, it is transient. Using first-step analysis,

$$\begin{aligned} f_{00} &= 1f_{10}=f_{10} \\ f_{10} &= \frac{\mu_1}{\lambda_1+\mu_1}1+\frac{\lambda_1}{\lambda_1+\mu_1}f_{20} \\ (\lambda_1+\mu_1)\,f_{10} &= \mu_1+\lambda_1 f_{20} \\ f_{20}-f_{10} &= \frac{\mu_1}{\lambda_1}\left(f_{10}-1\right) \end{aligned}$$

In general,

$$\begin{aligned} f_{n0} &= \frac{\mu_n}{\lambda_n+\mu_n}f_{n-1,0}+\frac{\lambda_n}{\lambda_n+\mu_n}f_{n+1,0} \\ (\lambda_n+\mu_n)\,f_{n0} &= \mu_n f_{n-1,0}+\lambda_n f_{n+1,0} \\ f_{n+1,0}-f_{n0} &= \frac{\mu_n}{\lambda_n}\left(f_{n0}-f_{n-1,0}\right) \\ f_{n+1,0}-f_{n0} &= \frac{\mu_n}{\lambda_n}\left(f_{n0}-f_{n-1,0}\right)=\frac{\mu_n\cdots\mu_1}{\lambda_n\cdots\lambda_1}\left(f_{10}-1\right) \end{aligned}$$

Now

$$\begin{aligned} f_{n+1,0} &= f_{10}+\sum_{i=1}^{n}\left(f_{i+1,0}-f_{i,0}\right) \\ &= f_{10}+\sum_{i=1}^{n}\frac{\mu_i\cdots\mu_1}{\lambda_i\cdots\lambda_1}\left(f_{10}-1\right) \end{aligned}$$

If the MC is transient, $f_{00} = f_{10} < 1$ and $f_{10} - 1 < 0$. However, we need

$$\lim_{n\to\infty} f_{n+1,0} = f_{10} + \sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1} (f_{10} - 1) \geq 0$$

This means

$$\sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1} < \infty$$

We will need the other direction. For $n = 1, 2, \cdots$, define

$$f_{i0}^{(n+1)} = \Pr(\text{the MC visits 0 before n+1} \mid X(0) = i).$$

Then, by first-step analysis, we have the same equations as before, but now also with a boundary condition

$$f_{n+1,0}^{(n+1)} = 0.$$

Recall

$$\begin{aligned} 0 = f_{n+1,0}^{(n+1)} &= f_{10}^{(n+1)} + \sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1} \left(f_{10}^{(n+1)} - 1 \right) \\ -1 &= \left(f_{10}^{(n+1)} = 1 \right) \left(1 + \sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1} \right) \\ 1 - f_{10}^{(n+1)} &= \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1}} \end{aligned}$$

Take $n \to \infty$,

$$\begin{aligned} \lim_{n\to\infty} 1 - f_{10}^{(n+1)} &= \lim_{n\to\infty} \mathbf{P}(\text{ the MC visits } n+1 \text{ before } 0 \mid X(0) = i) \\ &= \mathbf{P}(\text{ the MC never visits } 0 \mid X(0) = i) \end{aligned}$$

If the MC is recurrent, then the probability is 0, and

$$\frac{1}{1 + \sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1}} = 0 \iff \sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1} = \infty$$

To conclude, a bnd process is transident iff

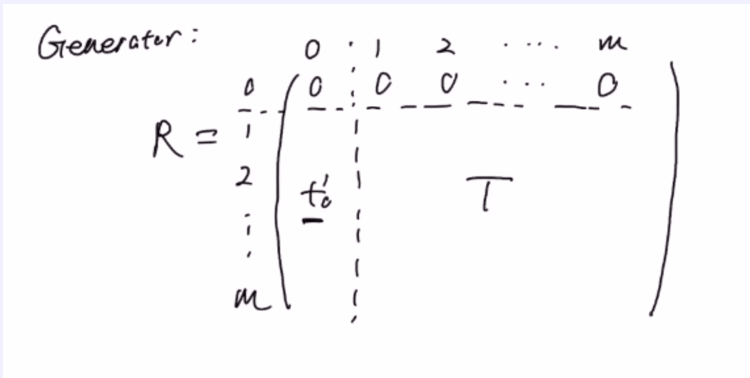
$$\sum_{i=1}^{\infty} \frac{\mu_i \cdots \mu_1}{\lambda_i \cdots \lambda_1} = \sum_{i=1}^{\infty} \prod_{i=1}^{n} \frac{\mu_i}{\lambda_i} < \infty$$

Chapter 3

Continuous Phase-Type Distribution

Section 1. Basic Setup

1.1. Definition: Let $\{X(t)\}_{t\geq 0}$ be a CTMC having m transient states $E = \{1, 2, \ldots, m\}$ and one absorbing state 0. Consider the generator



- The first row is a zero row because 0 is absorbing.
- $\vec{t_0}$ is a $m \times 1$ column vector that specifies the transition probabilities of going from a transient state to the absorbing state.
- T is a $m \times m$ matrix that specifies the transition probabilities of going from a transient state to another transient state.

Since the row sums of R are always 0, you have

$$\vec{t_0} + T \cdot \mathbf{1} = \vec{0}.$$

As a result, we have

$$\vec{t_0} = -T \cdot \mathbf{1}.$$

Define $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ to be the $1 \times m$ initial probability (of the transient part); that is, $\alpha_i = \Pr(X(0) = i)$ for $1 \leq i \leq m$. The probability of starting at state 0 is given by $1 - \sum_{i=1}^{m} \alpha_i = 1 - \alpha \cdot \mathbf{1}$.

Define Y to be the time until absorption

$$Y := \min\{t \geq 0 : X(t) = 0\}.$$

We say that such Y has a **continuous phase-type distribution** with representation

$$T \sim \text{CPH}_m(\alpha, T).$$

In words, a `CPH` is fully specified by

- m: the number of transient states.
- α: initial probabilities for the transient states.
- T: transient part of the generator.

Section 2. CDF of CPH

2.1. Note: We look for $\Pr(Y > y)$ and take the complement of this to obtain the CDF. For $y \geq 0$,

$$\begin{aligned}
\Pr(Y > y) &= \Pr(X(y) \in E) \\
&= \sum_{i=1}^{m} \Pr(X(y) \in E \mid X(0) = i) \cdot \Pr(X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \Pr(X(y) \in E \mid X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} \Pr(X(y) = j \mid X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} P_{ij}(y) \qquad (*)
\end{aligned}$$

The transition matrix at time y, $P(y)$, is given by

$$\begin{aligned}
P(y) &= e^{yR} \\
&= \sum_{n=0}^{\infty} \frac{y^n}{n!} R^n \\
&= I + \sum_{n=1}^{\infty} \frac{y^n}{n!} R^n \\
&= I + \sum_{n=1}^{\infty} \frac{y^n}{n!} \begin{bmatrix} 0 & 0 \\ X & T^n \end{bmatrix} && \text{not interested in } X \\
&= \begin{bmatrix} 1 & 0 \\ X & \sum_{n=1}^{\infty} \frac{y^n}{n!} T^n + I \end{bmatrix} && \text{not interested in } X \\
&= \begin{bmatrix} 1 & 0 \\ X & e^{yT} \end{bmatrix} && \text{not interested in } X
\end{aligned}$$

Note that X is not interesting to us, as we can compute it through

$$X = \mathbf{1} - e^{yT} \cdot \mathbf{1}.$$

Thus, to conclude, we have

$$P(y) = \begin{bmatrix} 1 & 0 \\ \mathbf{1} - e^{yT} \cdot \mathbf{1} & e^{yT} \end{bmatrix}$$

Go back to $(*)$, we have

$$\Pr(Y > y) = \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} P_{ij}(y)$$

$$= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} (e^{yT})_{ij}$$
$$= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} (e^{yT})_{ij} \mathbf{1}$$
$$= \alpha \cdot e^{yT} \cdot \mathbf{1}$$

Then the CDF of Y is given by

$$F_Y(y) = 1 - \Pr(Y > y) = 1 - \alpha \cdot e^{yT} \cdot \mathbf{1}.$$

Note that

$$\Pr(Y = 0) = F_Y(0) = 1 - \alpha \cdot I \cdot \mathbf{1} = 1 - \alpha \cdot \mathbf{1} = \alpha_0.$$

which makes sense.

Section 3. PHF of CPH

3.1. Note: For $y > 0$,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy}\left(1 - \vec{\alpha} e^{yT}\mathbf{1}\right) \\ &= -\vec{\alpha}\frac{d}{dy} e^{yT}\mathbf{1} \\ &= -\vec{\alpha} e^{yT} T\mathbf{1} \\ &= \vec{\alpha} e^{yT}\vec{t_0} \qquad\qquad \vec{t_0} = -T\mathbf{1} \end{aligned}$$

To summarize, Y is a random variable that has a discrete probability mass at 0 with probability $\alpha_0 = 1 - \alpha\mathbf{1}$ and a density for $y > 0$ given by

$$f_Y(y) = \alpha e^{yT}\vec{t_0}.$$

3.2. Example: Consider the exponential density

$$f(x) = \lambda e^{-\lambda x}.$$

The CPH representation is given by: generator:

$$R = \begin{bmatrix} 0 & 0 \\ \lambda & -\lambda \end{bmatrix}$$

- probability of going from transient 1 to absorbing 0 is λ.

Initial distribution: $(0, 1)$. The transient part: 1.
Number of transient states: 1.

Thus, this is a $\text{CPH}_1(1, -\lambda)$.

Section 4. Properties of CPH

4.1. Note (Independent Sum of CPH): Let $X \sim \text{CPH}_m(\alpha, T)$ and $Y \sim \text{CPH}_n(\beta, S)$ with $X \perp\!\!\!\perp Y$. Using the same reasoning as for DPH, we have

$$Z := X + Y \sim \text{CPH}_{m+n}(\delta, D)$$

where

$$\delta = (\alpha, \alpha_0 \beta)$$

$$D = \begin{array}{c c} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{bmatrix} T & \vec{t}_0 \beta \\ 0 & S \end{bmatrix} \end{array}$$

Note that $\vec{t}_0 \beta$ is an $m \times n$ matrix. This ensures the exit distribution for X is the initial distribution needed by Y.

4.2. Example (Gamma): Also known as Erlang distribution with parameters n, λ.

$$f(x) = \frac{\lambda^n e^{n-1} e^{-\lambda x}}{(n-1)!}, x > 0, \lambda > 0, n \in \mathbb{Z}^+$$

If $X \sim \text{Erlang}(n, \lambda)$, then $X \to_d \sum_{i=1}^n Y_i$ where $Y_i \overset{iid}{\sim} \text{Exp}(\lambda)$ (proof omitted). Thus, we have $X \sim \text{CPH}_n(\alpha, T)$ where

$$\alpha = (1, 0, 0, \ldots, 0)$$

$$T = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & \cdots & n \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} & \begin{bmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & \ddots & \ddots & \\ & & & -\lambda & \lambda \\ & & & & -\lambda \end{bmatrix} \end{array}$$

Note the same reasoning works for independent sums of exponential random variables with different parameters.

4.3. Note (Mixture of CPH): Let $X \sim \text{CPH}_m(\alpha, T)$ and $Y \sim \text{CPH}_n(\beta, S)$. Define

$$Z = \begin{cases} X & \text{with probability } p \\ Y & \text{with probability } 1 - p \end{cases}$$

the choice of going from X or Y is independent of X and Y. That is,

$$F_Z = pF_X + (1-p)F_Y.$$

As in the discrete case,

$$Z \sim \text{CPH}_{m+n}(gamma, G)$$

where $\gamma = (p\alpha, (1-p)\beta)$ and

$$G = \begin{array}{c c} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{bmatrix} T & 0 \\ 0 & S \end{bmatrix} \end{array}$$

As a result, the mixture of exponential or Erlang distribution are again CPH.

4.4. Note: All the moments of a CPH are finite.

Chapter 4

Queuing Theory.

Section 1. Setup

1.1. Definition ($A/S/m/c/p$ Queue): We characterize a **queue** using five parameters:

- A: **A**rrival process, often assume iid interarrival times (so this is a renewal process), e.g.,
  - M: exponential interarrival times, so Poisson arrival process.
  - G: general interarrival times.
  - D: deterministic (constant) interarrival times.
  - E_k: Erlang-k interarrival times.
  - PH: continuous phase-type interarrival times.
- S: **S**ervice process, using the same set of labels as A.
- m: Number of servers.
- c: **C**apacity, which equals waiting capacity + service capacity, often omitted when $c = \infty$.
- p: **P**opulation, the number of customers in total, often omitted when $p = \infty$.

1.2. Example ($M/M/1$ Queue): With $A = M$ (Poisson arrival times), $S = M$ (Poisson service times), $m = 1$ (one server), we are looking at a system that has one single server with Poisson service times and unlimited waiting places and customer population.

1.3. Example ($M/M/m/m$ Queue): Poisson arrival times and Poisson service times. However, we have m servers and m capacity. Since capacity equals service places, there is no waiting places, so any customer that cannot get served right away will leave the system right away.

1.4. Definition (Queue Length): Define **queue length** to be the total number of customers, including the ones waiting and the ones being served.

1.5. Definition (Service Discipline):

- FIFO/FCFS
- LIFO/LCFS
  - **Preemptive resume**: arriving customers will be immediately served; interrupted service *resumes* afterward.
  - **Preemptive restart**: arriving customers will be immediately served; interrupted service *restarts* afterward.
  - Non-preemptive: arriving customer waits until the ongoing services is finished.
- Scheduling of servers: services in rotating order, processor sharing.
- SIRO: service-in-random-order.
- SJF: shortest-job-first (unrealistically ideal case, minimize waiting time).

1.6. Remark: Do queuing systems always have the Markov property?

In general, the answer is No. Recall that the Markov property implies memoryless in time, which requires both interarrival times and services times to follow an exponential distribution. Thus, we see that only $M/M/s$ queues satisfy the Markov property and thus are directly CTMC.

However, there are cases where we can transform them into CTMC/DTMC.

(1). $G/M/s$ queue observed at each arrival is a DTMC, as this allows you to forget about interarrival times.

(2). $M/G/s$ queue observed at the beginning/end of each service is a DTMC, as this allows you to forget about service times.

Although Erlang-k and PH are not memoryless, due to their relation to the exponential distribution and CTMC, the systems with these distributions can be transformed into CTMC with additional parameters included in the states.

Section 2. Simple Case: $M/M/1$ Queue

2.1. Note: By definition, an $M/M/1$ queue means the system has one server and both the interarrival times and the service times follow a Poisson distribution. From the previous remark, we know that this is a CTMC, in particular a BnD process where each "birth" corresponds to the arrival of a new customer and each "death" corresponds to the completion of one customer's service. Since the arrival rate and service rate do not depend on the current number of customers in the system, we have $\lambda_i = \lambda$ (constant arrival rate) and $\mu_i = \mu$ (constant service rate, as there is only one server).

2.2. Definition: Define the **traffic intensity** for an $M/M/1$ queue by

$$\rho := \frac{\lambda}{\mu}.$$

2.3. Note: Recall that for a BnD process, a stationary distribution exists iff

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

In this case, the stationary distribution is given by

$$\begin{aligned} \pi_0 &= \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n}}, \\ \pi_i &= \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \pi_0. \end{aligned}$$

Since $\lambda_i = \lambda$ and $\mu_i = \mu$, we see that

$$\begin{aligned} \pi_i &= \left(\frac{\lambda}{\mu}\right)^i \pi_0, \\ \pi_0 &= \frac{1}{1 + \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n} = \frac{1}{\frac{1}{1-\lambda/\mu}} = 1 - \frac{\lambda}{\mu} \\ \implies \pi_i &= \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right)^i \end{aligned}$$

provided that $\lambda < \mu$. Using the newly-defined ρ, we can write

$$\pi_i = (1 - \rho)\rho^i.$$

In particular, a stationary distribution exists iff

$$\rho < 1 \iff \lambda < \mu.$$

2.4. Note: Some interesting quantities.

- Probability that the server is busy.

$$\Pr(\text{server is busy}) = 1 - \pi_0 = 1 - (1 - \rho) = \rho.$$

- Expected number of customers in the system.

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i\pi_i = 0\pi_0 + 1\pi_1 + \cdots = \frac{\rho}{1-\rho},$$

which is the mean of a geometric random variable $X \sim \text{Geo}(1-\rho)$ which starts counting from 0.

Section 3. Detailed Balance Condition

3.1. Definition: A distribution π satisfies the **detailed balance condition** for a CTMC with generate R if

$$\forall i, j \in S : \pi_i R_{ij} = \pi_j R_{ji}.$$

3.2. Intuition: $\pi_i R_{ij}$ is the total probability flow from i to j and $\pi_j R_{ji}$ is the total probability flow from j to i.

3.3. Theorem: *If a distribution π satisfies the detailed balance condition, then it is a stationary distribution.*

Proof. Note that

$$\begin{aligned}
\sum_{j\in S, j\neq i} \pi_i R_{ij} &= \sum_{j\in S, j\neq i} \pi_j R_{ji} \\
\pi_i \sum_{j\in S, j\neq i} R_{ij} &= \sum_{j\in S, j\neq i} \pi_j R_{ji} \\
-\pi_i R_{ii} &= \sum_{j\in S, j\neq i} \pi_j R_{ji} \\
0 &= \sum_{j\in S} \pi_j R_{ji} \\
\implies \forall i : (\pi R)_i &= 0
\end{aligned}$$

Moreover, π is a distribution. It follows that π is a stationary distribution. $\square$

3.4. Note: In general, a stationary distribution does not need to satisfy the detailed balance condition, so

$$\begin{aligned}
\text{detailed balance condition} &\implies \text{stationary distribution} \\
\text{stationary distribution} &\;\not\!\!\!\implies \text{detailed balance condition}
\end{aligned}$$

However, if the CTMC is a BnD process, then

$$\text{detailed balance condition} \iff \text{stationary distribution}.$$

Section 4. Queues with Infinity Population and Capacity: $M/M/c$ and $M/M/\infty$

4.1. Note: Recall the following result from STAT-330. Let $X_1, \dots, X_n$ be independent random variables where $X_i \sim \text{Exp}(\lambda_i)$. Then $\min\{X_1, \dots, X_n\} \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$. In particular, if $X_1, \dots, X_n$ are iid random variables following $\text{Exp}(\lambda)$, then

$$\min\{X_1, \dots, X_n\} \sim \text{Exp}(n\lambda).$$

4.2. Note: As before, an $M/M/c$ queue is a CTMC, and in particular, it is a BnD process. The birth rate is $\lambda_i = \lambda$ as the arrival rate is constant, but the death rate is more complicated.

$$\mu_i = \begin{cases} i\mu & i < c \\ c\mu & i \geq c \end{cases}$$

Suppose there are i customers in the system. If the number of customers is smaller than the number of servers in the system, i.e., $i < c$, then all customers will be served and i servers will be busy. Since the service times of these services are iid random variables following $\text{Exp}(\mu)$, we see that the next completion of any service is essentially the minimum of these i exponential random variables. Using the result given above, since each service time follows $\text{Exp}(\mu)$, the minimum of these i iid exponential random variables is $\mu_i = i\mu$ in this case.

Now suppose there are $i \geq c$ customers in the system, which implies that all servers will be busy. We can use the same argument to show that $\mu_c = c\mu$, because the next completion is the minimum of c iid exponential random variables with parameter μ. For $i > c$, since we are still waiting for the same number of servers to finish (which is c), the next completion remains the minimum of c iid exponential random variables. Hence, when $i \geq c$, the death rate is always $c\mu$.

To summarize, when the number of customers in the system is smaller than the number of servers, the death rate grows proportionally/linearly with the number of customers. However, when the number of customers reaches the number of servers, then the death rate will become a constant.

4.3. Note: We now derive the stationary distribution of an $M/M/c$ queue.

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{c-1} \frac{\lambda^i}{i!\mu^i} + \sum_{i=c}^{\infty} \frac{\lambda^i}{c!c^{i-c}\mu^i}}$$

$$\pi_i = \begin{cases} \dfrac{\lambda^i}{i!\mu^i}\pi_0 & 1 \leq i \leq c-1 \\ \dfrac{\lambda^i}{c!c^{i-c}\mu^i}\pi_0 & i \geq c \end{cases}$$

Note the stationary distribution exists iff the denominator of π_0 e, so let us look at the third term. Observe that

$$\sum_{i=c}^{\infty} \frac{\lambda^i}{c!c^{i-c}\mu^i} < \infty \iff \lambda < c\mu$$

4.4. Definition: Define the **traffic intensity** for an $M/M/c$ queue by

$$\rho := \frac{\lambda}{c\mu}.$$

4.5. Note: Continuing from Note 4.3, we see that a stationary distribution exists iff $\rho < 1$. In this case,

$$\begin{aligned}\pi_0 &= \left(1 + \sum_{i=1}^{c-1} \frac{(c\rho)^i}{i!} + \sum_{i=c}^{\infty} \frac{c^c}{c!}\rho^i\right)^{-1} \\ &= \left(1 + \sum_{i=1}^{c-1} \frac{(c\rho)^i}{i!} + \frac{c^c\rho^c}{c!(1-\rho)}\right)^{-1} \\ &= \left(1 + \sum_{i=1}^{c-1} \frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}\right)^{-1}\end{aligned}$$

and

$$\pi_i = \begin{cases} \dfrac{(c\rho)^i}{i!}\pi_0 & i \le c-1 \\ \dfrac{c^c\rho^i}{c!}\pi_0 & i \ge c \end{cases}$$

4.6. Note: Now consider the case where $c = \infty$. Using the same argument, this is a CTMC and a BnD process with $\lambda_i = \lambda$ and $\mu_i = i\mu$ for all $i \ge 0$.

$$\begin{aligned}\pi_0 &= \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda^i}{i!\mu^i}} = e^{-\lambda/\mu} < \infty \\ \pi_i &= \frac{(\lambda/\mu)^i e^{-(\lambda/\mu)}}{i!}\end{aligned}$$

Observe that this is exactly a Poisson distribution with parameter λ/μ! Moreover, this $M/M/\infty$ queue always has a stationary distribution as $\pi_0 < \infty$.

Section 5. Queues with Finite Capacity/Population: $M/M/1/c$ and $M/M/1/\infty/c$

5.1. Note ($M/M/1/c$): We now set a limit on the capacity, i.e., once the number of customers in the system reaches c, new customers will leave right away.

Birth rate and death rate:

$$\lambda_i = \begin{cases} \lambda & i \le c-1 \\ 0 & i \ge c \end{cases} \qquad \mu_i = \begin{cases} \mu & i \le c \\ \text{arbitrary}, 0 & i \ge c+1 \end{cases}$$

Stationary distribution:

$$\pi_i = \frac{\lambda_i \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i} \pi_0 = \left(\frac{\lambda}{\mu}\right)^i \pi_0$$

$$\pi_0 = \left(1 + \sum_{i=1}^{c} \left(\frac{\lambda}{\mu}\right)^i\right)^{-1} = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{c+1}}$$

$$\pi_i = \frac{\left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right)}{1 - \left(\frac{\lambda}{\mu}\right)^{c+1}}$$

5.2. Note ($M/M/1/\infty/c$): No waiting capacity, but the population is finite. Consider the following example. Suppose there are C machine in the system, each breaks according to $\text{Exp}(\lambda)$. Each server will repair the broken machines with service time $\text{Exp}(\mu)$.

Birth rate and death rate:

$$\lambda_i = \begin{cases} (c-i)\lambda & i \le c-1 \\ 0 & i \ge c \end{cases} \qquad \mu_i = \begin{cases} \mu & i \le c \\ 0 & i \ge c+1 \end{cases}$$

Stationary distribution:

$$\pi_i = \frac{\lambda_i \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i} \pi_0 = c! \lambda^i (c-i)! \mu^i \pi_0$$

$$\pi_0 = \left(\sum_{i=0}^{c} \frac{c!}{(c-i)!} \left(\frac{\lambda}{\mu}\right)^i\right)^{-1}$$

$$\pi_i = \frac{c! \left(\frac{\lambda}{\mu}\right)^i}{(c-i)! \sum_{j=0}^{c} \frac{c!}{(c-j)!} \left(\frac{\lambda}{\mu}\right)^j}$$

Section 6. Generating Function

6.1. Definition: The (probability) **generating function** of a distribution π on $\mathbb{Z}_{\geq 0}$ is

$$g(z) = \sum_{n=0}^{\infty} z^n \pi(n) = \mathbb{E}(z^X)$$

if $X \sim \pi$ for $z \in [0, 1]$.

6.2. Proposition: *The distribution π is recovered by taking derivatives of g at 0, i.e.,*

$$\pi_k = \pi(k) = \frac{g^{(k)}(0)}{k!}.$$

Proof. By absolute convergence, we interchange the series and derivative,

$$\begin{aligned} g^{(k)}(0) &= \sum_{n=0}^{\infty} \underbrace{\frac{d^k}{dz^k} z^n \pi(n)\bigg|_{z=0}}_{0 \text{ if } n<k \text{ or } n>k} \\ &= \frac{d^k}{dz^k} z^k \pi(k) = k!\pi(k) \\ \Longrightarrow \pi(k) &= \frac{g^{(k)}(0)}{k!} \end{aligned}$$

□

6.3. Proposition: $g(1) = \sum_{n=0}^{\infty} \pi(n) = 1$*, and the* k*th factorial moment satisfies*

$$\mathbb{E}[X(X-1)\cdots(X-k+1)] = g^{(k)}(1).$$

Proof. Using DCT,

$$\begin{aligned} \frac{d^k}{dz^k} \mathbb{E}\left(z^X\right) &= \mathbb{E}\left(\frac{d^k}{dz^k} z^X\right) \\ &= \mathbb{E}\left(X(X-1)\cdots(X-k+1)z^{x-k}\right) \\ \Longrightarrow \frac{d^k}{dz^k} \mathbb{E}\left(z^X\right)\bigg|_{z=1} &= \mathbb{E}(X(X-1)\cdots(X-k+1)) \end{aligned}$$

□

6.4. Proposition: *Let X, Y be independent random variables taking values in non-negative integers with generating functions g_X and g_Y. Then $g_{X+Y} = g_X g_Y$.*

Proof. Since X and Y are independent, $g_{X+Y}(z) = \mathbb{E}\left(z^{X+Y}\right) = \mathbb{E}\left(z^X\right)\mathbb{E}\left(z^Y\right) g_X(z) g_Y(z)$. □

Chapter 5

Renewal Theory

Highlight

- $\{N(t)\}_{t\geq 0}$: Number of events that happen by time t.
- X_i: Interarrival time before the ith arrival, i.e., time between the event $i-1$ and event i.
- S_i: Renewal time for the ith arrival, i.e., time of arrival of event i.
- μ: Mean interarrival time, $\mu := \mathbb{E}[X_i]$.

Section 1. Introduction to Renewal Processes

1.1. Definition: A **counting process** $\{N(t)\}_{t\geq 0}$ is a stochastic process that represents the number of events that happen by time t.

1.2. Note: Properties of counting processes:

- $N(0) = 0$
- $N(t) \in \mathbb{Z}_{\geq 0}$
- $s < t \implies N(s) \leq N(t)$

1.3. Definition: A **renewal process** is a counting process with iid interarrival times. The **renewal times** are the times that events happen.

1.4. Intuition: This type of process is called *renewal* because each time an event occurs, it is as if you restart/renew/reset the chain/process and you can think of the chain getting reset back to the starting point $t = 0$.

1.5. Note: Let X_i and S_i be the interarrival time and the renewal time for the ith arrival. Then

$$\begin{aligned} S_0 &= 0 \\ S_1 &= X_1 \\ &\cdots \\ S_i &= S_{i-1} + X_i = X_1 + X_2 + \cdots + X_i \end{aligned}$$

1.6. Example: The Poisson process is a renewal process with exponential interarrival times.

1.7. Remark: From now on, we assume the interarrival times follow a distribution with finite mean, i.e., $\mathbb{E}[X_1] < \infty$. In this case, by the Strong Law of Large Number, we have

$$\frac{S_n}{n} \to \mathbb{E}[X_1] =: \mu.$$

Section 2. Convolution

2.1. Motivation: In order to study the distribution of the sum of independent random variables, we introduce the technique of convolution.

2.2. Definition: Define

- F: cdf of a non-negative random variable X.
- g: non-negative function defined on $[0,\infty)$, locally bounded (bounded on all finite intervals).

The **convolution** of F and g is a function on $[0,\infty)$ defined by

$$\begin{aligned} F * g(t) &= \mathbb{E}\left[g(t-X)\mathbf{1}_{\{X\le t\}}\right] \\ &= \int_0^t g(t-x)dF(x) = \begin{cases} \displaystyle\int_0^t g(t-x)f(x)\,dx & \text{if } X \text{ has density } f \\ \displaystyle\sum_{0\le x\le t} g(t-x)\Pr(X=x) & \text{if } X \text{ is discrete} \end{cases} \end{aligned}$$

where $f(x) = \frac{d}{dx}F(x)$ almost everywhere (the undefined set has measure 0).

2.3. Note: Convolution is linear (inherited from integral). In particular,

- $F * (cg) = c(F * g)$,
- $F * (g_1 + g_2) = F * g_1 + F * g_2$,
- $\forall a \in [0,1] : (aF_1 + (1-a)F_2) * g = a(F_1 * g) + (1-a)(F_2 * g)$

2.4. Proposition: *If $g = G$ is the cdf of a non-negative random variable Y and $X \perp\!\!\!\perp Y$, then $F * G(t)$ is the cdf of $X + Y$.*

Proof. Algebraic proof: Assuming continuous case,

$$\begin{aligned} \Pr(X+Y\le t) &= \mathbb{E}[\Pr(X+Y\le t \mid X)] \\ &= \int_0^t \Pr(X+Y\le t \mid X=x)f(x)\,dx \\ &= \int_0^t \Pr(Y\le t-x)f(x)\,dx = \int_0^t G(t-x)f(x)\,dx = F*G(t) \end{aligned}$$

Probabilistic proof:

$$\begin{aligned} F*G(t) &= \mathbb{E}[G(t-X)\mathbf{1}_{\{X\le t\}}] \\ &= \mathbb{E}[\Pr(Y\le t-X \mid X)\mathbf{1}_{\{X\le t\}}] \\ &= \mathbb{E}[\Pr(Y\le t-X \mid X)] \\ &= \Pr(Y\le t-X) && \Pr(Y\le t-X \mid X=x>t)=0 \\ &= \Pr(X+Y\le t) \end{aligned}$$

□

2.5. Corollary: *If G is a cdf, then $F * G = G * F$.*

Proof. $X + Y$ and $Y + X$ have the same distribution. □

2.6. Corollary: *If X, Y are independent and have densities f, g, then $X + Y$ has density*

$$h(t) := \int_0^t g(t-x)f(x)\,dx.$$

Proof. Let F and G be cdfs of X and Y, respectively. Then the cdf of $X + Y$ is given by $F * G$.

$$\begin{aligned} h(t) &= \frac{d}{dt} F * G(t) \\ &= \frac{d}{dt}\int_0^t G(t-x)f(x)\,dx \\ &= \int_0^t g(t-x)f(x)\,dx. \end{aligned}$$

□

2.7. Proposition: *$F * g$ is also non-negative and bounded on finite intervals.*

Proof. Since g is non-negative, we have

$$F * g(t) = \mathbb{E}[g(t-X)\mathbf{1}_{\{X \le t\}}] \ge 0.$$

Moreover, if $g(s) \le M$ on $[0, t]$, then

$$F * g(s) = \mathbb{E}[g(s-X)\mathbf{1}_{\{X \le t\}}] \le M$$

so $F * g$ is bounded on finite intervals. □

2.8. Note: We now define what's known as *n-fold* convolution. Suppose $g = G$ is the cdf of a random variable Y. Then

$$F * (F * g) = F * F_{X+Y} = F_{X'+X+Y},$$

where X, X' have the same cdf F. Similarly,

$$(F * F) * g = F_{X+X'} * g = F_{X+X'+Y}.$$

But $X' + X + Y$ and $X + X' + Y$ have the same distribution!

2.9. Definition: Let F be the cdf of a non-negative random variable X. The **n-fold convolution** is defined by

$$F^n(t) = F * F * \cdots * F(t),$$

which is the cdf of the sum of n independent copies of X.

2.10. Note: Now consider the n-fold convolution in our renewal processes setting.

If F is the cdf of the interarrival times X_i, then F^n is the cdf of $S_n = X_1 + \cdots + X_n$, which is the renewal time of event n. Let's see what interesting quantities we can derive from this.

Since F^n is the cdf of S_n, by definition, we get

$$F^n(t) = \Pr(S_n \le t).$$

Equivalently, using the definition of counting process,

$$F^n(t) = \Pr(N(t) \ge n).$$

Can you see why these two probabilities are equal? [a]

Let us try to express $\Pr(N(t) = n)$ in terms of F^n. Observe that [b]

$$\Pr(N(t) = n) = \Pr(S_n \le t) - \Pr(S_{n+1} \le t) = F^n(t) - F^{n+1}(t).$$

This gives us one way of expressing $\Pr(N(t) = n)$.

Alternatively, let us define the **survival function** of X. [c]

$$\overline{F}(t) := 1 - F(t) = \Pr(X > t).$$

Then

$$\begin{aligned}
F^{n+1}(t) + F^n(t) * \overline{F}(t) &= F^n(t) * F(t) + F^n(t) * \overline{F}(t) \\
&= F^n(t) * F(t) + F^n(t) * (1 - F(t)) \\
&= F^n(t) * F(t) + F^n(t) * 1 - F^n(t) * F(t) \\
&= F^n * 1 \\
&= F^n(t) \\
\implies F^n(t) * \overline{F}(t) &= F^n(t) - F^{n+1}(t).
\end{aligned}$$

Since $\Pr(N(t) = n) = F^n(t) - F^{n+1}(t)$, we also have

$$\Pr(N(t) = n) = F^n * \overline{F}(t).$$

[a] Intuitively, $\Pr(S_n \le t)$ finds the probability that the event n occurs no later than time t, while $\Pr(N(t) \ge n)$ finds the probability that we have encountered at least n events in the time period $[0, t]$. Therefore, they are equivalently.

[b] The probability that we encounter example n events in the time period $[0, t]$ is equal to the probability that the event n occurs no later than time t minus the probability that the event $n + 1$ occurs no later than time t.

[c] Intuitively, if we view $F(t) = \Pr(X \le t)$ as a function modelling the probability that an animal dies before time t (i.e., its lifetime), then $\overline{F}(t)$ models the probability that an animal lives pass time t. Hence, it is called the *survival function*.

2.11. Remark: 1 serves as the identity element in convolution:

$$F * 1 = \mathbb{E}[1 \cdot \mathbf{1}_{\{X \le t\}}] = \Pr(X \le t) = F(t).$$

Section 3. Renewal Function

3.1. Motivation: From the previous section, we have $F^n(t) = \Pr(N(t) \geq n)$. Then the expected number of events in time period $[0, t]$ is given by

$$\mathbb{E}[N(t)] = \sum_{n=1}^{\infty} \Pr(N(t) \geq n) = \sum_{n=1}^{\infty} F^n(t).$$

3.2. Definition: The expected number of renewals by the time t is called the **renewal function** of the renewal process.

$$m(t) := \sum_{n=1}^{\infty} F^n(t) = \mathbb{E}[N(t)].$$

3.3. Note: The renewal function has the following properties (we will skip the proof):

- $m(t) \geq 0$ and $m(t)$ is non-decreasing.
- The renewal function completely determines the distribution of the interarrival time.

3.4. Proposition: *$m(t) < \infty$ for all $t \geq 0$.*

Proof. Fix $t > 0$.

Case 1. $F(t) < 1$. Then $a := \Pr(X_1 > t) > 0$. For any $n \in \mathbb{Z}_+$,

$$\begin{aligned} F^n(t) = \Pr(S_n \leq t) &= \Pr(X_1 + \cdots + X_n \leq t) \\ &\leq \Pr(X_1 \leq t, \cdots, X_n \leq t) \\ &= (\Pr(X_1 \leq t))^n && \text{iid} \\ &= (1-a)^n \end{aligned}$$

Since $1 - a < 1$, we have

$$m(t) = \sum_{n=1}^{\infty} F^n(t) \leq \sum_{n=1}^{\infty} (1-a)^n < \infty.$$

Case 2. $F(t) = 1$. Then there must exist $s > 0$ such that $\Pr(X_1 > s) > 0$. Let $k \in \mathbb{Z}_+$ be such that $ks > t$. As a result,

$$a := \Pr(X_1 + \cdots + X_k > t) \geq \Pr(X_1 > s, \ldots, X_k > s) = (\Pr(X_1 > s))^k > 0.$$

Define $Y_1 = S_k$, $Y_2 = S_{2k}$, ..., and let N_Y be the counting process of $\{Y_i\}$. Intuitively, we only count the kth, $2k$th, ... arrivals. Then

$$N_Y(t) = \left\lceil \frac{N(t)}{k} \right\rceil \implies N(t) < kN_Y(t) + k.$$

Thus, it suffices to prove that $m_Y(t) = \mathbb{E}[N_Y(t)] < \infty$. However, note that $Y_1, Y_2 - Y_1, Y_3 - Y_2, \ldots$ are iid and satisfy $a = \Pr(Y_1 > t) > 0$. This goes back to case 1. We are done. □

3.5. Corollary: *With probability 1, $N(t) < \infty$ for all $t < \infty$.*

3.6. Example: Consider a renewal process where the interarrival times are iid $\text{Exp}(\lambda)$, i.e., $\text{Poi}(\lambda t)$. Let us find the renewal function. Since sum of n iid $\text{Exp}(\lambda)$ follows $\text{Erlang}(n, \lambda)$ with cdf

$$F^n(t) = 1 - \sum_{i=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^i}{i!} \qquad t \geq 0.$$

Thus,

$$\begin{aligned} m(t) = \sum_{n=1}^{\infty} F^n(t) &= \sum_{n=1}^{\infty} \left(1 - \sum_{i=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^i}{i!}\right) \\ &= \sum_{n=1}^{\infty} \left(1 - \left(1 - \sum_{i=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^i}{i!}\right)\right) && \text{Poisson pmf} \\ &= \sum_{n=1}^{\infty} \sum_{i=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^i}{i!} \\ &= \sum_{i=1}^{\infty} \sum_{n=1}^{i} e^{-\lambda t} \frac{(\lambda t)^i}{i!} \\ &= \sum_{i=1}^{\infty} i e^{-\lambda t} \frac{(\lambda t)^i}{i!} \\ &= \lambda t && \text{expectation of } \text{Poi}(\lambda t) \end{aligned}$$

This is expected, because $N(t) \sim \text{Poi}(\lambda t)$.

Section 4. Renewal Equation

4.1. Note: Recall $m(t)$ denotes the expected number of renewals by time t. Let us condition on the first arrival/renewal.

$$\begin{aligned}
m(t) &= \mathbb{E}[N(t)] \\
&= \mathbb{E}[\mathbb{E}[N(t) \mid X_1]] \\
&= \int_0^\infty \mathbb{E}[N(t) \mid X_1 = x] \cdot f(x)\,dx \\
&= \int_0^t \mathbb{E}[N(t) \mid X_1 = x] \cdot f(x)\,dx && X_1 = x > t \implies N(t) = 0 \\
&= \int_0^t [\mathbb{E}[N(t) \mid X_1 = x] + \mathbb{E}[N(t) - N(x) \mid X_1 = x]] \cdot f(x)\,dx \\
&= \int_0^t (1 + \mathbb{E}[N(t-x)])f(x)\,dx \\
&= \int_0^t f(x)\,dx + \int_0^t m(t-x)f(x)\,dx \\
&= F(t) + F * m(t) && t \geq 0
\end{aligned}$$

4.2. Definition: The equation

$$m(t) = F(t) + F * m(t)$$

for $t \geq 0$ is referred to as the **renewal equation**.

4.3. Proposition: *The expected time of the next renewal is given by*

$$\mathbb{E}\left[S_{N(t)+1}\right] = \mu(m(t)+1), \quad t \geq 0,$$

where $\mu := \mathbb{E}[X_i]$.

Proof.

$$\begin{aligned}
\mathbb{E}\left[S_{N(t)+1}\right] &= \mathbb{E}\left[\sum_{j=1}^{N(t)+1} X_j\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{\infty} X_j \mathbf{1}_{\{N(t)+1 \geq j\}}\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{\infty} X_j \mathbf{1}_{\{N(t) \geq j-1\}}\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{\infty} X_j \mathbf{1}_{\{X_1 + \cdots + X_{j-1} \leq t\}}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{\infty} \mathbb{E}[X_j \mathbf{1}_{\{X_1+\cdots+X_{j-1}\le t\}}] \\
&= \sum_{j=1}^{\infty} \mathbb{E}[X_1] \Pr(S_{j-1} \le t) \\
&= \mu \sum_{j=1}^{\infty} F^{j-1}(t) \\
&= \mu \left(\sum_{n=1}^{\infty} F^n(t) + F^0(t) \right) \\
&= \mu(m(t)+1) \qquad t \ge 0
\end{aligned}
$$

□

4.4. Theorem: *For a renewal process whose interarrival times have pdf f and cdf F, if a function $Z(t)$ satisfies*

$$Z(t) = g(t) + \int_0^t Z(t-x) f(x)\, dx = g(t) + F * Z(t)$$

for $t \ge 0$ where g is a function that is bounded on finite intervals, then the unique solution for $Z(t)$ which is bounded on finite intervals is

$$Z(t) = g(t) + m * g(t) := g(t) + \sum_{n=1}^{\infty} F^n * g(t) \quad t \ge 0$$

Note we extended the definition of convolution to the case where g is not non-negative.

Proof. Later. □

4.5. Remark: The standard renewal equation for $m(t)$,

$$m(t) = F(t) + F * m(t)$$

satisfies the condition of the above equation, with $g(t) = F(t)$. Hence, by the theorem, the unique solution is

$$F(t) + \sum_{n=1}^{\infty} F^n * F(t) = \sum_{n=1}^{\infty} F^n(t),$$

which agrees with what we have known.

4.6. Definition: Define the following quantities regarding the current interarrival time (the one containing t).

- $\delta_t = t - S_{N(t)}$: current life time, or age.
- $\gamma_t = S_{N(t)+1} - t$: excess or residual life time.
- $\beta_t = \delta_t + \gamma_t$: total life time.

4.7. Note:

- Expected residual time: $\mathbb{E}[\gamma_t] = Z(t) = \mu(m(t)+1) - t$.
- Average rate of renewals: $\frac{N(t)}{t} \to \frac{1}{\mu}$ with probability 1.
- Elementary renewal theorem: $\lim_{t\to\infty} \frac{m(t)}{t} = \frac{1}{\mu}$.

4.8. Note: Let $\{N(t)\}_{t\geq 0}$ be renewal process with interarrival times $\{X_n\}_{n=1}^{\infty}$ with renewal times S_n. When event n occurs at S_n, collect a random reward R_n. Note that we do not assume R_n and X_n are independent. However, assume (X_n, R_n) are iid for $n = 1, 2, \ldots$ Then the total reward collected by time t is given by

$$R(t) = \sum_{n=1}^{N(t)} R_n, \quad R(t) = 0 \text{ if } N(t) = 0.$$

Here, $R(t)$ is said to be the renewal reward process.

4.9. Theorem: *If $\mathbb{E}[R_1] < \infty$ and $\mu = \mathbb{E}[X_1] < \infty$, then long run average renewal*

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R_1]}{\mu}.$$

4.10. Remark: We actually also have

$$\frac{\mathbb{E}[R(t)]}{t} \to \frac{\mathbb{E}[R_1]}{\mu}.$$

Note the result still holds if instead of collecting the reward at the end of each cycle, we continuously collect the reward in a cycle.

Section 5. Regenerative Process

5.1. Definition: A stochastic process $\{X(t)\}_{t\geq 0}$ is called a regenerative process, if it probabilistically restarts at renewal times $S_1, S2, \ldots$ That is, there exists a renewal process with renewal times $S_1, S_2, \ldots$ such that

$$\{X(t)\}_{t\geq 0} \stackrel{d}{=} \{X(t+S_n)\}_{t\geq 0}$$

for any n. The renewals are also called **regenerations** in this setting.

5.2. Example: Let $\{X(t)\}_{t\geq 0}$ be a recurrent CTMC with $X(0) = i$. Let the renewal time S_n be the nth return to state i. Then $\{X(t)\}_{t\geq 0}$ is a regenerative process by Markov property.

5.3. Note: Consider a regenerative process $\{X(t)\}_{t\geq 0}$. Like in CTMC, we still call the possible values of the process *states*. Imagine that we continuously collect a reward of 1 per unit of time only when the process is in state j. Then

- Reward in cycle n: $R_n = \int_{S_{n-1}}^{S_n} \mathbf{1}_{\{X(u)=j\}} du$
- Reward collected up to time t: $R_n = \int_0^t \mathbf{1}_{\{X(u)=j\}} du$

A derivation gives us

$$\lim_{t\to\infty} \frac{R(t)}{t} = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{X(u)=j\}} du = \frac{\mathbb{E}[R_1]}{\mu} = \frac{\mathbb{E}[\text{time in } j \text{ during one cycle}]}{\mathbb{E}[\text{cycle length}]}.$$

5.4. Theorem (Key Renewal Theorem): *For a renewal process where interarrival times have pdf f and expectation μ, if a function $Z(t)$ satisfies*

$$Z(t) = g(t) + \int_0^t Z(t-x) f(x)\, dx = g(t) + F * Z(t) \qquad t \geq 0,$$

where $g(t)$ is a directly Riemann integrable function that is bounded on finite intervals, then

$$\lim_{t\to\infty} Z(t) = \frac{1}{\mu} \int_0^\infty g(u)\, du.$$

5.5. Note: A function g defined on $[0,\infty)$ is called **directly Riemann integrable** if

$$\lim_{b\to 0} b \sum_{n=0}^{\infty} \inf\{g(t) : nb \leq t < nb+b\},$$
$$\lim_{b\to 0} b \sum_{n=0}^{\infty} \sup\{g(t) : nb \leq t < nb+b\}$$

both exist and are equal. This notion is stronger than "Riemann integrable on $[0,\infty)$". In particular, if a function g satisfies

- $g(t) \geq 0$ for all $t \geq 0$,
- $g(t)$ is non-increasing,
- $\int_0^\infty g(t)\,dt < \infty$,

then g is directly Riemann integrable.

5.6. Note: One can derive

$$\lim_{t\to\infty} \Pr(X(t) \in A) = \frac{\mathbb{E}[\text{time spent in } A \text{ in one cycle}]}{\mathbb{E}[\text{cycle length}]}$$

Take $A = \{j\}$, we have

$$\lim_{t\to\infty} \Pr(X(t) = j) = \lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R_1]}{\mu}.$$

That is, limiting probability equals the long-run fraction of the time spent in j equals

$$\frac{\mathbb{E}[\text{time spent in } j \text{ in one cycle}]}{\mathbb{E}[\text{cycle length}]}.$$