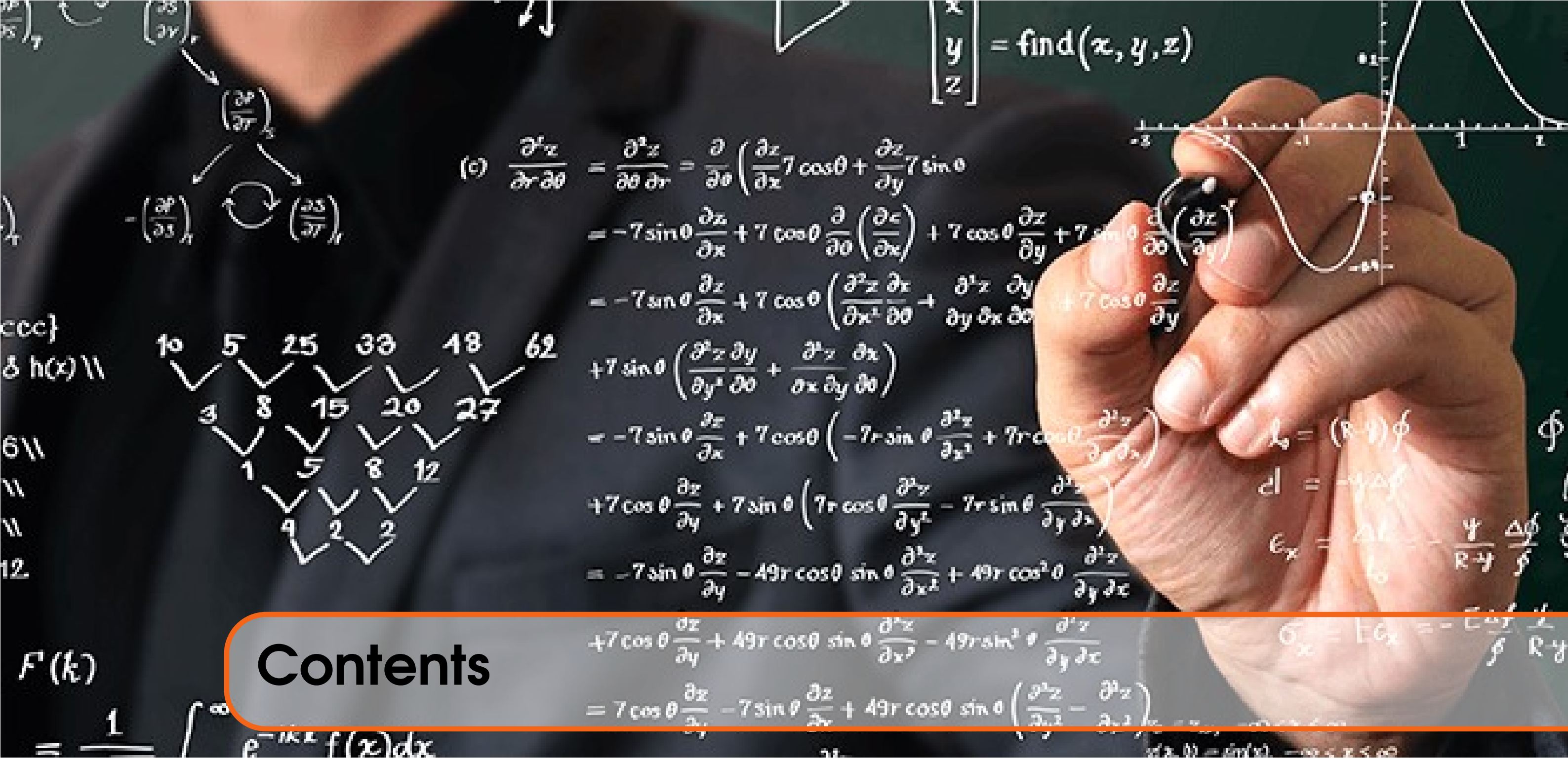
STAT 433 Course Notes

University of Waterloo

The One And Only
Waterloo 76er
Bill Zhuo



Contents

I Review of STAT333

Review of STAT333

1. Preparation

1.1 Probability Space and Random Variable

Definition 1.1.1 — Probability Space. A probability space consists of a triplet $(\Omega, \mathscr{E}, \mathbf{P})$, where

1. Ω is the sample space, the collection of all the possible outcomes of a random experiment.

   R Quick examples could be rolling a 6-face die, which has $\Omega = \{1,2,3,4,5,6\}$ or flipping a coin, which has $\Omega = \{H,T\}$. Or if you want to be fancy, we can also consider the sequence of coin tosses, which has $\{HHT, HTH, \cdots\}$. It could even be different kinds of weather for tomorrow. Very flexible definition.

2. $\mathscr{E}$ is the $\sigma-$algebra, a collection of events, subsets of Ω. An event E is a subset of Ω for which we can talk about probability.

   R A simple $\mathscr{E}$ could be the power set of Ω. That is a big one. Usually the $\mathscr{E}$ that we work with are quite smaller sets.

3. $\mathbf{P}$ is the probability measure. A set function that maps from $\mathscr{E}$ to $[0,1]$.

$$\mathbf{P} : \mathscr{E} \to [0,1]$$

   A probability needs to satisfy the probability axioms:
   (a) $\mathbf{P}(E) \in [0,1], \forall E \in \mathscr{E}$
   (b) $\mathbf{P}(\Omega) = 1$
   (c) **Countable additivity**: for $\{E_n\}_{n\in\mathbb{N}}$ disjoint event, we have

$$\mathbf{P}\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \mathbf{P}(E_n)$$

You see the sample space of a random experiment is quite flexible and therefore less rigorous. We clarifies things by labelling. This labelling can be considered as a function or mapping. That's essentially a random variable.

Definition 1.1.2 — Random Variable. A random variable (r.v.) is a mapping $X : \Omega \to \mathbb{R}$ where $\omega \mapsto X(\omega)$.

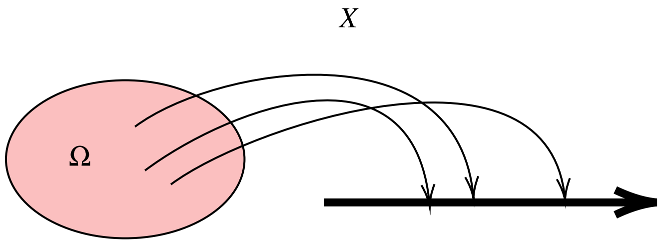


Figure 1.1.1: *How random variable works*

1.2 Stochastic Processes

Intuitively, a process is something that evolves/changes over time. Stochastic is essentially a fancier name for random. Why is this used?

Etymology of "Stochastic"

In the ancient Greek, it means "aim at" or "guess". Really self-explanatory. Then, this word was brought up again by the Russian mathematician, Khinchin (ngl, his name is like kimchi). He knew what he was doing is essentially random process. But the commies in Soviet Russia at that time might not want something that related to randomness (dialectical materialism). So, to dodge the bullet, literally, he chose to use this ancient word, "stochastic".

> *"Nobody understands ancient Greeks anyway."*
>
> — Prof.Shen

Later on, two American mathematicians, Feller and Doob, translated the Russian random process as "stochastic process". So, basically, the takeaway here is the following picture.

Suppose we have a number, deterministic for now. We add two secret ingredients.

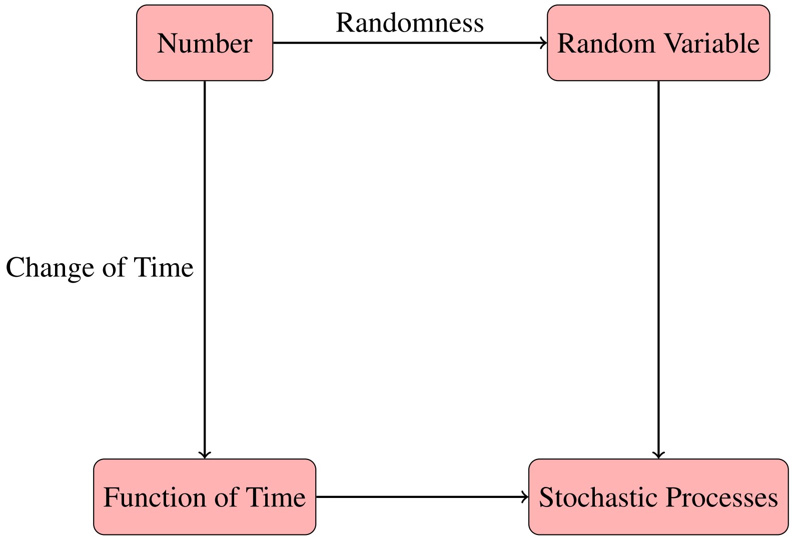


Figure 1.2.1: *Formulation of Stochastic Process Definition*

1. A sequence/family of random variables (simpler, so we take this as the definition)
2. A random function (hard to formulate)

Definition 1.2.1 — Stochastic Process. A stochastic process $\{X_t\}_{t\in T}$ is a collection of random variables defined on the common probability space, where T is an index set. In most cases, T corresponds to “time", which can be either discrete or continuous.
1. In discrete case, we rather write $\{X_n\}_{n=0,1,\ldots}$
2. In continuous case, we typically consider $t \in [0,\infty)$

Definition 1.2.2 — States. The possible values of $X_t, t \in T$ are called the states of the process. There collection is called the **state space**, often denoted by S. The state space can either be discrete or continuous. The famous example of continuous state space is the **Brownian motion**. Suppose you use Brownian motion to model a future index. It would be continuous state space of the future index values (well, more or less). But in this course, we will be start with the discrete state space so that we can relabel the states in S into the so-called **standardized state space**

$$S^* = \{0,1,2,\ldots\} \longleftarrow \textbf{Countable State Space}$$

or

$$S^* = \{0,1,2,\ldots,n\} \longleftarrow \textbf{Finite State Space}$$

1.3 Review on DTMC

1.3.1 Review on Conditional Probability

Definition 1.3.1 — Conditional Probability. The conditional probability of an event given an event A with $\mathbf{P}(A) > 0$ is given by

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(B\cap A)}{\mathbf{P}(A)}$$

Theorem 1.3.1 Let $A_1, A_2, \ldots$ be disjoint events such that

$$\bigcup_i A_i = \Omega$$

then we have

1. **Law of Total Probability**:

$$\mathbf{P}(B) = \sum_{i=1}^{\infty} \mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i)$$

2. **Bayes' Rule**:

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i)}{\sum_{j=1}^{\infty} \mathbf{P}(B|A_j) \cdot \mathbf{P}(A_j)}$$

You see conditionally probabilities are sometimes (to me, most of the time) hassles. So what's good?

Definition 1.3.2 — Independence. Two events A, B are called independent if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$$

we denote this as $A \perp\!\!\!\perp B$.

Exercise 1.1 Show that if $\mathbf{P}(A) > 0$, then $A \perp\!\!\!\perp B$ if and only if $\mathbf{P}(B|A) = \mathbf{P}(B)$. ■

1.4 Basic Properties of DTMC

Definition 1.4.1 — Discrete time markov chain (DTMC). $\{X_n\}_n = 0, 1, \cdots$ is called a DTMC with transition matrix $P = \{P_{ij}\}_{i,j \in S}$, if for any time n, and any state $j, i, i_{n-1}, \cdots, i_0 \in S$, we have

$$\mathbf{P}(\underbrace{X_{n+1} = j}_{\text{future}} | \underbrace{X_n = i}_{\text{present}}, \underbrace{X_{n-1} = i_{n-1}, \cdots, X_0 = i_0}_{\text{past/history}}) = \mathbf{P}(X_{n+1} = j | X_n = i) = P_{ij}$$

this is the **Markov property**

The Markov property indicates that the future and the past are conditionally independent for a DTMC. In other words, the past will only have impact on the future through the present (state). Moreover, note that the quantity $\mathbf{P}(X_{n+1} = j | X_n = i)$ by all means not necessarily constant over time. But in this course, we assume this is also independent of time, which makes the notation of P_{ij} sensible. This assumption is called **time homogeneity** that the transitional property of this Markov chain is independent of time. Note that P_{ij} is indexed by states from S and putting things together gives us the transition matrix P. It needs to satisfy the following properties:

1. **Total probability:** $P_{ij} \geq 0$, $\sum_{j \in S} P_{ij} = 1, \forall i \in S$, this is a row sum
2. $n-$**step transition probabilities**: P is a 1-step transition matrix. What about n-step? Consider the $n-$step transition matrix $P^{(n)} = \{P^{(n)_{ij}}\}_{i,j \in S}$ where

$$P_{ij}^{(n)} := \mathbf{P}(X_{m+n} = j | X_m = i) \overset{\text{time homogeneity}}{=} \mathbf{P}(X_n = j | X_0 = i)$$

By Chapman-Kolmogorov Equation (C-K Equation), we know that

$$P^{(n)} = P^n$$

these two sides have very different meanings. The right hand side is multiplying the 1-step transition matrix n times. More generally, we have

$$P^{(n+m)} = P^{(m)}P^{(n)}$$
$$P_{ij}^{(n+m)} = \sum_{k\in S} P_{ik}^{(n)}P_{kj}^{(m)}$$

In this course, we let the initial distribution be the following row vector,

$$\begin{aligned}\mu &= (\mu(0), \mu(1), \cdots) \\ &= (\mathbf{P}(X_0 = 0), \mathbf{P}(X_0 = 1), \cdots)\end{aligned}$$

similarly, the distribution of X_n is given by

$$\mu_n = (\mathbf{P}(X_n = 0), \mathbf{P}(X_n = 1), \cdots) = \mu P^{(n)} = \mu P^n$$

we note that μ and P fully characterizes the distribution of a DTMC.

1.4.1 Conditional Expectation and $\mathbb{E}(f(X_n))$

In this part, we want to discuss how to compute expected payoff of a DTMC based on some reward function f. To do this, we first consider conditional expectation.

$$\mathbb{E}(g(X)|Y = y) = \begin{cases} \sum_x g(x)\mathbf{P}(X = x|Y = y) & \text{discrete case} \\ \int g(x) f_{X|Y}(x|y)dx & \text{continuous} \end{cases}$$

note that this expression of conditional expectation is a number when $Y = y$ is given information. More generally, without specifying a value for Y, the conditional expectation becomes a function of y. Therefore, a random variable that maps $Y(\omega)$ to $\mathbb{R}$.

$$\mathbb{E}(g(X)|Y)_{(\omega)} = \mathbb{E}(g(X)|Y = Y(\omega))$$

why do we want this?

Theorem 1.4.1 — Law of Iterated Expectation.

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$$

Now, we are ready to have

$$\mathbb{E}(f(X_n)) = \mu P^n f', \quad f' = \begin{bmatrix} f(0) \\ f(1) \\ \vdots \end{bmatrix}$$

where f' is a column vector. In this course, we use row vector to denote the distribution of states and column vector to denote a function on the process. How do we interpret this result?

1. **Weighted average:** if we use C-K equation result first, we can think of the final dot product as a weighted average of the function value on the $n-$step process.

$$\mu P^n f' = \mu_n f'$$

2. **Conditional function:** since matrix multiplication is associative, we can certainly calculate the product of the last two components first.

$$\mu f^{(n)'}, \ f^{(n)'} = \underbrace{\begin{bmatrix} \mathbb{E}(f(X_n)|X_0 = 0) \\ \mathbb{E}(f(X_n)|X_0 = 1) \\ \vdots \end{bmatrix} = P^n f'}_{\text{why this works?}}$$

We can think of $f^{(n)'}$ as the conditional function value, say a reward that is conditioned to the initial distribution.

1.4.2 Classification and Class Properties

Definition 1.4.2 — Communication. x communicates to y denoted by $x \to y$ if

$$\rho_{xy} := \mathbf{P}_x(T_y < \infty) > 0$$

where $\mathbf{P}_x(\cdot) := \mathbf{P}(\cdot|X_0 = x)$ and $T_y := \min\{n \geq 1 : X_n = y\}$. This T_y can be interpreted as the first visiting time (or re-visit time if the chain starts with y).

There is an equivalent definition. The intuition is simple. Say $x \to y$, which means eventually the chain can go from x to y. In other words, it is possible for you to go from x to y in some number of steps. Therefore equivalently, there exists $n \geq 1$ such that $P^n_{xy} > 0$. In particular, this relation defined by $\to$ is **transitive**. Namely, if $x \to y, y \to z$, then $x \to z$. This motivates the following definition.

Definition 1.4.3 — Communicating Class. A subset $C \subseteq S$ is a communicating class, if

1. $\forall i, j \in C, i \leftrightarrow j \; (i \to j, j \to i)$
2. $\forall i \in C, j \notin C, i \not\leftrightarrow j \; (i \not\to j$ or $j \not\to i)$

What if you only have one class?

Definition 1.4.4 — Irreducible. A DTMC is called irreducible if all the states are in the same class.

Find Classes on Graph: "Find the Loops"

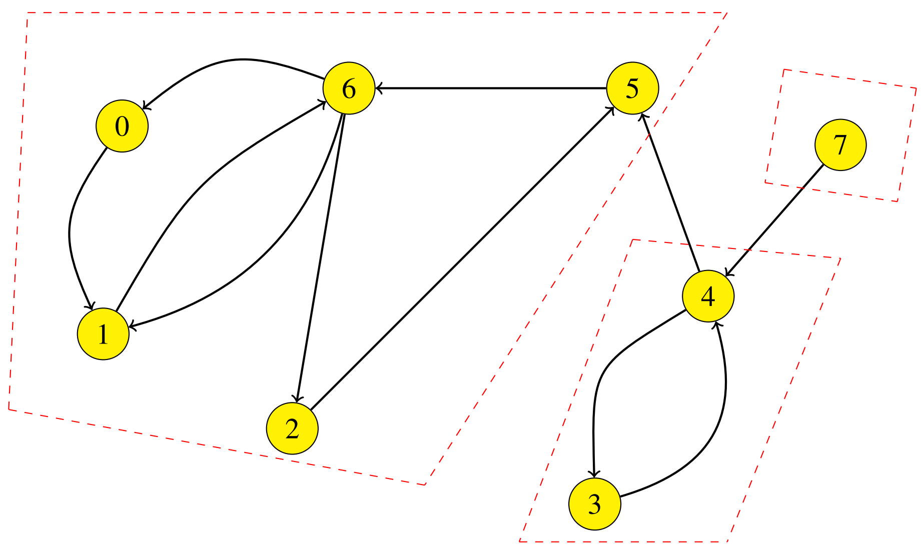


Figure 1.4.1: *Each red box indicates a communicating class*

1.4.3 Recurrence and Transience

Definition 1.4.5 — Recurrent & Transient State. A state $y \in S$ is called recurrent, if $\rho_{yy} = 1$, which means always return to y.
It is called transient if $\rho_{yy} < 1$, which means there is a pass probability that the chain never visits y again.

What are the criteria?

Recurrence	Transience
$\rho_{yy} = \mathbf{P}_y(T_y < \infty) = 1$	$\rho_{yy} < 1$
$\mathbf{P}_y(N(y) = \infty) = 1$	$\mathbf{P}_y(N(y) < \infty) = 1$
$\mathbb{E}_y(N(y)) = \infty$	$\mathbb{E}_y(N(y)) < \infty$
$\sum_{n=1}^{\infty} P_{yy}^n = \infty$	$\sum_{n=1}^{\infty} P_{yy}^n < \infty$

Exercise 1.2
1. Construct a random variable X such that $\mathbf{P}(X < \infty) = 1$ but $\mathbb{E}(X) = \infty$.
   Hint: St.Petersburg Paradox
2. Prove that $\mathbb{E}_y(N(y)) = \sum_{n=1}^{\infty} P_{yy}^n$.
   Hint: check notes

■

Recurrence/transience are class properties. Thus, we can classify communicating classes as recurrent or transient.

Other criteria and properties for recurrence/transience

1. If $\rho_{xy} > 0, \rho_{yx} < 1$, then x is transient.

   Corollary 1.4.2 If x is recurrent and $\rho_{xy} > 0$, then $\rho_{yx} = 1$.

2. Consider the following definition.

   Definition 1.4.6 — Closed Set. A set A is called closed if $i \in A, j \notin A$ implies $P_{ij} = 0$ (not from i to j in one step). Equivalently, this is to say, if $i \in A$ and $j \notin A$, then $i \not\to j$ (not from i to j in no-matter-how-many steps).

   "Cannot get out once the chain goes into A."

   With this definition, we have an intuitive result.

   Lemma 1.5 In a finite closed set, there has to be at least one recurrent state.

   Corollary 1.5.1 A **finite** closed class must be recurrent. In particular, an irreducible Markov Chain with finite state space must be recurrent.

   It turns out we can always decompose the state space as follow.

   Theorem 1.5.2 — Decomposition of the State Space. The state space S can be written as a disjoint union
   $$S = T \cup R_1 \cup R_2 \cup \ldots$$
   where T is the set of all transient states (T is not necessarily one class), and $R_i, i = 1, 2, \ldots$ are closed recurrent classes. Equivalently, in other textbook, we can say irreducible sets of recurrent states.

3. Finally, we give a formula to compute

   Proposition 1.5.3 — Starting from x, how many visits to y on average?.

   $$\mathbb{E}_x(N(y)) = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

1.5.1 Periodicity

In this part, we will define another class property, the periodicity. First, consider the **strong Markov property**. Recall that $T_y = \min\{n \geq 1 : X_n = y\}$.

Theorem 1.5.4 — Strong Markov Property for (time-homogeneous) MC. The process $\{X_{T_y+k}\}_{k=0,1,2,...}$ behaves like the MC with initial state y. (forget about the history and restart at state y)

> **R** Why is this stronger? The Markov property that we have encountered uses a deterministic time but here T_y itself is a random variable.

Definition 1.5.1 — Period of a State. The period of a state x is defined as

$$d(x) := \gcd\{n \geq 1 : P^n_{xx} > 0\}$$

> **R** Note that we are taking the gcd of the steps when the probability of state x going back to x is not 0. There is no guarantee for going-back.

Definition 1.5.2 — Aperiodic. If x has period 1, then we say x is **aperiodic**. If all states in a MC is **aperiodic**, then we call this MC is **aperiodic**.
If $P_{xx} > 0$, then x is obviously aperiodic. But the converse is not true, x is aperiodic does not imply $P_{xx} > 0$.

If x is **aperiodic**, we do not necessarily have $P_{xx} > 0$.

Lemma 1.6 — "Period is a class property".

$$x \to y, y \to x \implies d(x) = d(y)$$

1.6.1 Stationary Distribution and Limiting Behaviour

Definition 1.6.1 — Stationary Distribution. A probability distribution $\pi = (\pi_0, \pi_1, \dots)$ is called a stationary distribution (invariant distribution) of the DTMC $\{X_n\}_{n=0,1,...}$ with transition matrix P if

1. $\pi = \pi P$ as a system of equations
2. $\sum_{i \in S} \pi_i = 1$ by the definition of probability distribution

From this point on, we define the following short-hand notations.
Overall Conditions:

1. I: The MC is irreducible (one and only one class, everything communicates with everything)
2. A: The MC is aperiodic (i.e, all the states have period 1)
3. R: All the states are recurrent
4. S: There exists a stationary distribution π.

Definition 1.6.2 — Stationary Measure (Invariant Measure). Let a row vector

$$\mu^* = (\mu^*(0), \mu^*(1), \dots, \mu^*(i), \dots)$$

is called a stationary measure (invariant measure), if $\mu^*(i)^* \geq 0, \forall i \in S$ and $\mu^* P = \mu^*$.

A stationary measure is a stationary distribution without normalization. If $\sum_i \mu^*(i) < \infty$, then it can be normalized to get a stationary distribution.

Theorem 1.6.1 Let $\{X_n\}_{n=0,1,2,...}$ be an irreducible and recurrent DTMC with transition matrix P. Let $x \in S$ and $Tx := \min\{n \geq 1 : X_n = x\}$, then

$$\mu_x(y) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x > n), y \in S$$

defines a stationary measure with $0 < \mu_x(y) < \infty, \forall y \in S$.

$\mu_x(y)$ is also the expected number of visits to y before returning to x given that it starts with x. It is immediate that μ can be normalized to a stationary distribution if and only if

$$\sum_{y \in S} \pi_x(y) = \mathbb{E}_x(T_x) < \infty$$

(this is in fact the condition for *positive recurrence*)

1.7 Main Theorems

Theorem 1.7.1 — Convergence Theorem. Suppose I,A,S. Then,

$$P^n_{xy} \longrightarrow_{n\to\infty} \pi(y), \forall x, y \in S$$

no matter where you starts, only depends on the target states.

> "The limiting transition probability, hence also the limiting distribution, does not depend on where we start. (Under the conditions of I,A,S)"

Or we can write

$$\lim_{n\to\infty} P^n_{xy} = \pi(y), \forall x, y \in S \Longrightarrow \lim_{n\to\infty} \mathbf{P}(X_n = y) = \pi(y)$$

Corollary 1.7.2 If y is transient, then $\pi(y) = 0$ for any stationary distribution π.

Corollary 1.7.3

$$\text{I and S} \Longrightarrow \text{R}$$

Theorem 1.7.4 — Long-Run Frequency. Suppose I, R, if $N_n(y)$ is the **number of visits to** y **up to time** n, then

$$\frac{N_n(y)}{n} \underset{n\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}_y(T_y)}$$

where

$$T_y = \min\{n \geq 1 : X_n = y\}$$

we consider $\frac{N_n(y)}{n}$ as the fraction of time spent in y (up to time n).

"Long run fraction of time spent in y is $\frac{1}{\mathbb{E}_y(T_y)}$"

where $\mathbb{E}_y(T_y)$ is the expected revisit time to y given that we start with y, which is also the "expected cycle length".

Theorem 1.7.5 — How to find a stationary distribution?. Suppose I and S, then,

$$\pi(y) = \frac{1}{\mathbb{E}_y(T_y)}$$

In particular, the stationary distribution is unique.

Corollary 1.7.6 — Nicest Case. Suppose I, A, S, (R), then

$$\pi(y) = \lim_{n\to\infty} P^n_{xy} = \lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{\mathbb{E}_y(T_y)}$$

$$\begin{aligned}\textbf{Stationary Distribution} &= \textbf{Limiting transition probability}\\ &= \textbf{Long-run fraction of time}\\ &= \frac{1}{\textbf{Expected revisit time}}\end{aligned}$$

Theorem 1.7.7 — Long-run Average. Suppose I, S, and $\sum_x |f(x)|\pi(x) < \infty$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_x f(x)\pi(x) = \pi f'$$

1.7.1 Special Cases

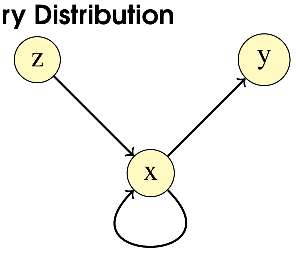
Detailed Balance Condition

Definition 1.7.1 — Detailed Balance Condition. A distribution $\pi = \{\pi(x)\}_{x\in S}$ is said to satisfy the detailed balance condition if

$$\pi(x)P_{xy} = \pi(y)P_{yx}, \forall x, y \in S$$

Proposition 1.7.8 — Detailed Balance Condition $\Longrightarrow$ Stationary Distribution. If a distribution π satisfies the detailed balance condition, then π is a stationary distribution.

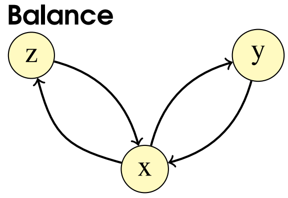
Stationary Distribution



Total probability "flow" entering x should be equal to the total probability flow leaving x

$$\sum_z \pi(z)P_{zx} = (\pi P)_x = \pi(x)$$

Detailed Balance



Flow $x \to y$ should be the flow $y \to x$

$$\pi(x)P_{xy} = \pi(y)P_{yx}$$

R The converse is not generally true as we can see the graph on the left has a stationary distribution but it is not detailed balanced. In STAT333, if P is **tri-diagonal**, the converse is true. (tri-diagonal P means it looks like something shown below)

$$\begin{bmatrix} \backslash & \backslash & 0 \\ \backslash & \backslash & \backslash \\ 0 & \backslash & \backslash \end{bmatrix}$$

Time Reversibility

Start with a DTMC $\{X_m\}_{m=0,1,\ldots}$. Fix n, then $\{Y_m\}_{m=0,1,\ldots,n}$ given by $Y_m = X_{n-m}$ is called the reversed process of $\{X_m\}$.

Theorem 1.7.9 If $\{X_m\}$ starts from a stationary distribution π satisfying $\pi(i) > 0$ for any $i \in S$, then its reversed process $\{Y_m\}$ is a DTMC with transition matrix given by

$$\begin{aligned} \hat{P}_{ij} &= \mathbf{P}(Y_{m+1} = j | Y_m = i) \\ &= \frac{\pi(j) P_{ji}}{\pi(i)} \end{aligned}$$

Definition 1.7.2 — Time-Reversable DTMC. A DTMC $\{X_m\}_{m=0,1,\ldots}$ is called time-reversable, of its reversed chain $\{Y_m\}_{m=0}^n$ has the same distribution as $\{X_m\}_{m=0}^n$ for all n.

R This is much stronger than reversability, not all reversable DTMC is time-reversable. But the other way around is clearly true. This is somehow related to the detailed balance condition in an intuitive way. Since the detail balance condition provides a two-way transition for each state, which provides the "time-reversability".

Proposition 1.7.10 A DTMC $\{X_m\}_{m=0,1,\ldots}$ is time-reversable **if and only if** it satisfies the detailed balance condition.

1.7.2 Infinite State Space

All the results covered in the previous parts hold for both finite and infinite state spaces (unless otherwise specified).
There is one distribution (one pair of two notions) which only makes sense in finite state space.

Definition 1.7.3 — Positive Recurrent and Null Recurrent. A state x is called positive recurrent if $\mathbb{E}_x(T_x) < \infty$ (recall $T_x = \min\{n \geq 1 : X_n = X\}$).
A recurrent state x is called null recurrent, if $\mathbb{E}_x(T_x) = \infty$.

R Recall that recurrence means $\mathbf{P}(T_x < \infty) = 1$ and transient means $\mathbf{P}(T_x = \infty) > 0$. The classification should be as follow:

Category	Subcategory
Recurrent	Positive Recurrent
Recurrent	Null Recurrent
Transient	Transient

All of these are still class properties.

We summarize all introduced result through $L_j = \lim_n P_{ij}^{(n)}$.

1. **Existence:** if j is aperiodic. When it exists, it also gives us the long-run fraction of time spends at state j
2. **Uniqueness:** does not depend on i if and only if there exists at most one positive recurrent class
3. **Value:**

$$L_j = \begin{cases} 0 & j \text{ is transient/null recurrent} \\ 0 & i \text{ is recurrent, } j \text{ is positive recurrent, } i, j \text{ not in same class} \\ > 0 & i, j \text{ in same class and positive recurrent} \\ \text{Determined by exit probability} & i \text{ transient, } j\text{positive recurrent} \end{cases}$$

This set of $\{L_j\}_{j \in S}$ will give a stationary distribution if it forms a distribution (can be added up to 1).

More on stationary distribution:

1. **Existence:** stationary distribution exists if and only if there exists at least one positive recurrent class.
2. **Uniqueness:** stationary distribution is unique if and only if there exists only one positive recurrent class.
3. $\pi(j) = 0, \forall \pi$ if and only if j is transient/null recurrent

Proposition 1.7.11 A DTMC with finite state space must have at least one positive recurrent class and no null recurrent class/state.

1.8 Simple Random Walk and Branching Process

Recall that

Example 1.1 — Simple Random Walk. Let $X_1, X_2, \dots$ be i.i.d r.v.s. For each one of them

$$\begin{cases} \mathbf{P}(X_i = 1) = p \\ \mathbf{P}(X_i = -1) = 1 - p \end{cases}$$

define $S_0 = 0$, for the other S_i, we have

$$S_n = \sum_{i=1}^{n} X_i$$

Then, $\{S_n\}_{n=0,1,2,\dots}$ is a stochastic process. With the state space,

$$\mathbf{S} = \mathbb{Z}$$

This $\{S_n\}_{n=0,1,2,\dots}$ is called a “simple random walk". Note that

$$S_n = \begin{cases} S_{n-1} + 1 & \text{with probability } p \\ S_{n+1} - 1 & \text{with probability } 1 - p \end{cases}$$

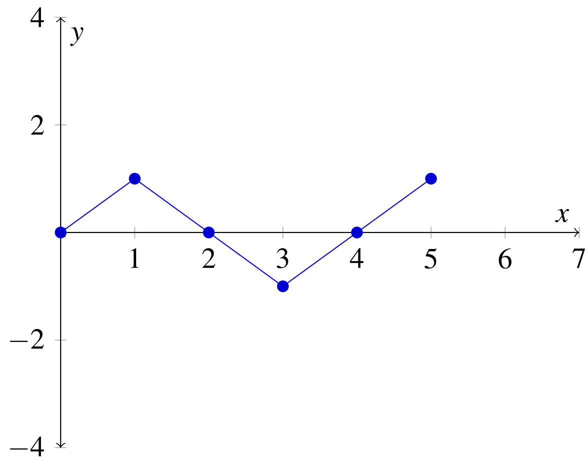


Figure 1.8.1: *Simple Random Walk*

R

1. $p \neq \frac{1}{2} \implies$ transient
2. $p = \frac{1}{2} \implies$ null recurrent

■

1.8.1 Branching Process

Consider a population. Each organism, at the end of its life, produces a random number Y of offsprings. The distribution of Y is denoted as

$$\mathbf{P}(Y = k) = P_k, P_k \geq 0, k = 0, 1, \ldots$$

and $\sum_{k=0}^{\infty} P_k = 1$.
Start from one common ancestor, $X_0 = 1$. **The number of offsprings of different individuals are independent**.
Let

$$X_n := \text{the number of individual (in the population) in the } n\text{-th generation.}$$

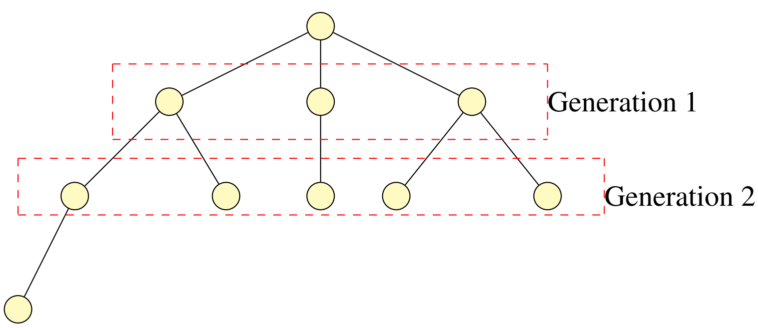


Figure 1.8.2: *Branching Process*

Then,

$$X_{n+1} = Y_1^n + Y_2^{(n)} + \cdots + Y_{X_n}^{(n)}$$

where $Y_1^n, Y_2^{(n)}, \dots, Y_{X_n}^{(n)}$ are **independent** copies of Y and $Y_i^{(n)}$ is the ubmer of offsprings of the i-th individual in the n-th generation.
One thing we care about is the expectation of the number of offsprings of the n-th generation:

$$\mathbb{E}(X_n) = ?$$

Assuming $\mathbb{E}(Y) = \mu$.

Solution:

$$\begin{aligned}\mathbb{E}(X_{n+1}) &= \mathbb{E}\left(Y_1^n + Y_2^{(n)} + \dots + Y_{X_n}^{(n)}\right)\\ &= \mathbb{E}\left(\mathbb{E}\left(Y_1^n + Y_2^{(n)} + \dots + Y_{X_n}^{(n)} | X_n\right)\right)\\ &= \mathbb{E}(\mu X_n) = \mu\mathbb{E}(X_n)\end{aligned}$$

R This result is known as **Wald's Identity** in statistics

$$\mathbb{E}(X_{n+1}) = \mu\mathbb{E}(X_n)$$

We can continue this inductively to see that

$$\mathbb{E}(X_n) = \mu^n \mathbb{E}(X_0) = \mu^n, n = 0, 1, \dots$$

1.8.2 Extinction Probability

As long as $P_0 > 0$, state 0 is absorbing and all the other states are transient.
But it does not mean that the population will extinct for sure.
If the population on average keeps growing and tends to infinity with positive probability, then we probability of extension is smaller than 1.
To find the extinction probability, we introduce the mathematical tool, **Generating Functions**.

Definition 1.8.1 — Generating Function. Let $P = \{P_0, P_1, \dots\}$ be a distribution on on $\{0, 1, \dots\}$.
Let η be a random variable following distribution P. That is

$$\mathbf{P}(\eta = i) = P_i$$

The **generating function** of η or of P is defined by

$$\begin{aligned}\varphi(s) &= \mathbb{E}(s^\eta)\\ &= \sum_{k=0}^{\infty} P_k s^k \quad 0 \le s \le 1\end{aligned}$$

Proposition 1.8.1 — Properties of Generating Functions. Let $\varphi(s)$ be a generating function

1. $\varphi(0) = P_0$, $\varphi(1) = \sum_{k=0}^{\infty} P_k = 1$
2. Generating function determines the distribution.

$$P_k = \frac{1}{k!}\frac{d^k\varphi(s)}{ds^k}\bigg|_{s=0}$$

Reason (This is immediate by Taylor Expansion, a short illustration is included here):

$$\varphi(s) = P_0 + P_1 s^1 + \dots + P_{k-1}s^{k-1} + P_k s^k + P_{k+1}s^{k+1} + \dots$$

then,

$$\frac{d^k\varphi(s)}{ds^k} = k!P_k + (\cdots)s + (\cdots)s^2 + \dots$$

Then,

$$\left.\frac{d^k\varphi(s)}{ds^k}\right|_{s=0} = k!P_k \Longrightarrow P_k = \frac{1}{k!}\left.\frac{d^k\varphi(s)}{ds^k}\right|_{s=0}$$

In particular, given $P_1, P_2, \dots \geq 0$, this implies $\varphi(s)$ is **increasing** and **convex** (all of its derivatives will be positive).

3. Let $\eta_1, \dots, \eta_n$ be independent random variables with generating functions $\varphi_1, \dots, \varphi_n$, then

$$X = \eta_1 + \cdots + \eta_n$$

have generating function

$$\varphi_X(s) = \varphi_1(s)\dots\varphi_n(s)$$

Proof.

$$\begin{aligned}
\varphi_x(s) &= \mathbb{E}\left(s^X\right) \\
&= \mathbb{E}(s^{\eta_1}\dots s^{\eta_n}) \\
&= \mathbb{E}(s^{\eta_1})\dots\mathbb{E}(s^{\eta_n}) \quad \textbf{independence} \\
&= \varphi_1(s)\dots\varphi_n(s)
\end{aligned}$$

■

4. **Pseudo Moments:** very useful!

$$\left.\frac{d^k\varphi(s)}{ds^k}\right|_{s=1} = \left.\frac{d^k\mathbb{E}(s^\eta)}{ds^k}\right|_{s=1} = \mathbb{E}\left.\left(\frac{d^k s^\eta}{ds^k}\right)\right|_{s=1} = \mathbb{E}\left.\Big(\eta(\eta-1)\dots(\eta-k+1)s^{\eta-k}\Big)\right|_{s=1}$$
$$= \mathbb{E}(\eta(\eta-1)\dots(\eta-k+1))$$

In particular,

$$\mathbb{E}(\eta) = \varphi'(1)$$

and

$$\mathbf{Var}(\eta) = \varphi''(1) + \varphi'(1) - \left(\varphi'(1)\right)^2$$

The Graph of a Generating Function

Back to extinction probability. Define

$$N = \min\{n \geq 0 : X_n = 0\}$$

to be the **extinction time** and

$$u_n = \mathbf{P}(N \leq n) = \mathbf{P}(X_n = 0)$$

where $\mathbf{P}(N \leq n)$ is the extinction happens before or at time n. Note that $\{u_n\}$ is an increasing sequence and bounded above, by **Monotone Convergence Theorem**, it is well-defined to define

$$u := \lim_{n\to\infty} u_n = \mathbf{P}(N < \infty) = \mathbf{P}(\text{the population eventually die out}) = \text{extinction probability}$$

Our goal is to find u

(meiyouzhuya)

Note that we have the following relation between u_n and u_{n-1}. Then,

$$u_n = \sum_{k=0}^{\infty} P_k(u_{n-1})^k = \varphi(u_{n-1})$$

where φ is the generating fucntion of Y.

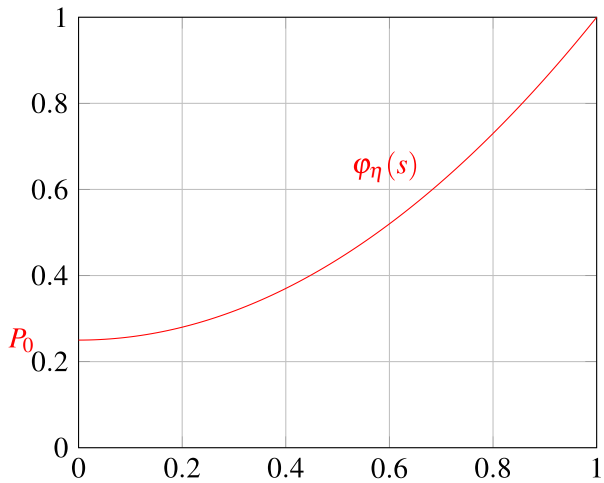


Figure 1.8.3: *The Graph of a Generating Function*

Reason:

Note that each sub-population has the same distribution as the whole population. The whole population dies out in n steps if and only if each sub-population initiated by an individual in generating 1 dies out in $n-1$ steps.

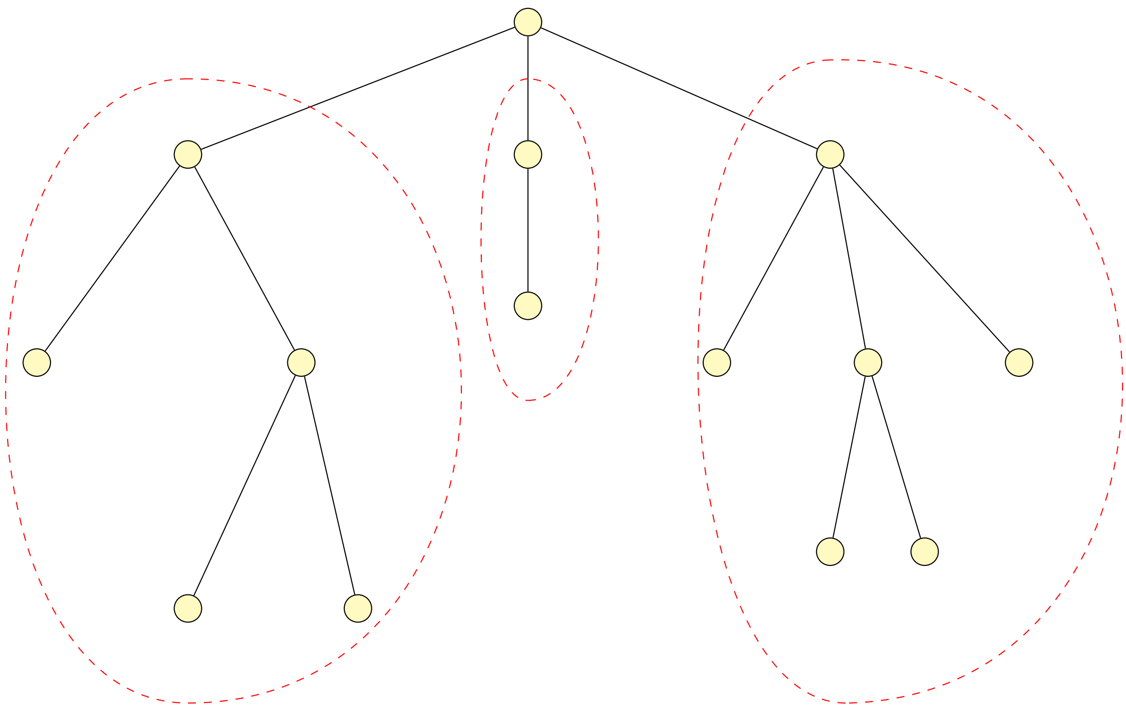


Figure 1.8.4: *We can divide the whole population into sub-populations*

Then,

$$\begin{aligned} u_n &= \mathbf{P}(N \le n) \\ &= \sum_k \mathbf{P}(N \le n | X_1 = k)\mathbf{P}(X_1 = k) \\ &= \sum_k \mathbf{P}(N_1 \le n-1, \dots, N_k \le n-1 | X_1 = k) P_k \\ &= \sum_k P_k u_{n-1}^k = \varphi(u_{n-1}) \end{aligned}$$

where N_m is the number of steps for the sub-population to die out. And we can also write $u_{n-1}^k = \mathbb{E}(u_{n-1}^Y)$.
Thus, the problem becomes:

With an initial value $u_0 = 0$ since $(X_0 = 1)$ and the relation

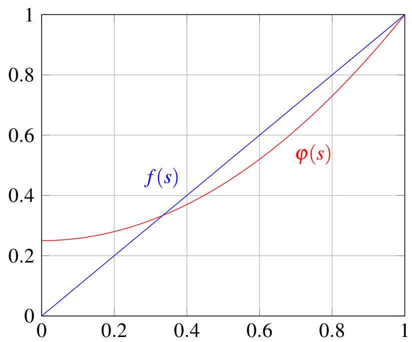
$$u_n = \varphi(u_{n-1})$$

what is $\lim_{n\to\infty} u_n = u$?

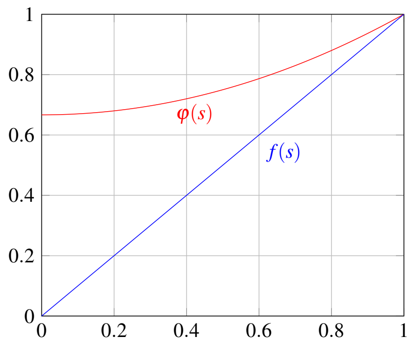
Recall that

1. $\varphi(0) = P_0 > 0$
2. $\varphi(1) = 1$
3. $\varphi(s)$ is increasing
4. $\varphi(s)$ is convex

Draw $\varphi(s)$ and the function $f(s) =$ between 0 and 1. We have two possibilities.



Case 1: $u < 1$



Case 2: $u = 1$ **extinction happens for sure**

Theorem 1.8.2 The extinction probability u will be the smallest intersection of $\varphi(s)$ and $f(s)$. Equivalently, it is the smallest solution of $\varphi(s) = s$ between 0 and 1.

Reason:
see the dynamics of the graph
This dynamic process verifies the results for **case 1** and **case 2**.

Q: How to tell whether we are in Case 1 or Case 2?

A: We can check the derivative at $s = 1$!

Note that $\varphi'(1) = \mathbb{E}(Y)$ and

$$\begin{aligned} \varphi'(1) > 1 &\Longrightarrow \textbf{Case 1} \\ \varphi'(1) \le 1 &\Longrightarrow \textbf{Case 2} \end{aligned}$$

Thus, we conclude that

$\mathbb{E}(Y) > 1 \Longrightarrow$ **Extinction with certain porbability less than 1 and u is a unique solution**

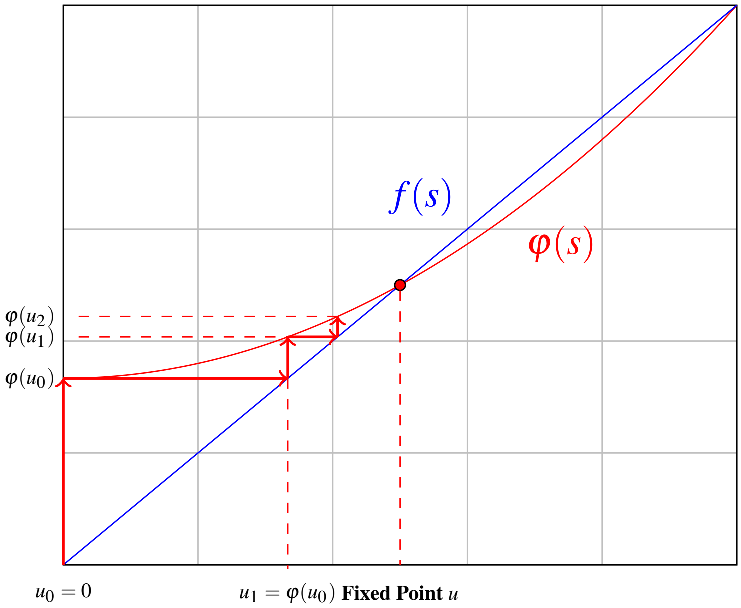


Figure 1.8.5: *Fixed Point Iteration*

also, we can think of $\mathbb{E}(Y) > 1$ as **on average, there are more than 1 offspring**, so the population will probably explode, which diminish the chance to wipe out the whole population. (*Thanos has left the chat.*)

$$\mathbb{E}(Y) \leq 1 \implies \textbf{Extinction happes for sure (with prob. 1)}$$

we can think of $\mathbb{E}(Y) \leq 1$ as **on average, there is less than or equal to 1 offspring**, so there is always a risk to have the population to die out.

1.9 Absorption Probability and Absorption Time

1.9.1 Basic Setting

The subsets A, B of the state space. $C = S - (A \cup B)$ is finite. The question is

> "Starting in a state in C, what is the probability that the chain exits C by entering A or B."

Mathematical Formulation

We define $V_A = \min\{n \geq 0 : X_n \in A\}$ and $V_B = \min\{n \geq 0, X_n \in B\}$. Then, what is $\mathbf{P}_x(V_A < V_B)$?

■ **Example 1.2**

$$P = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0.25 & 0.6 & 0 & 0.15 \\ 0 & 0.2 & 0.7 & 0.1 \\ & & 1 & \\ & & & 1 \end{pmatrix} \end{array}$$

then, let's say

$$C = \{1, 2\} \qquad A = \{3\} \qquad B = \{4\}$$

We think the entry $P_{33} = 1$ and $P_{44} = 1$ are not that important since we only care about the chain before going to 3 or 4. ■

Let $h(1) = \mathbf{P}_1(V_3 < V_4), h(2) = \mathbf{P}_2(V_3 < V_4)$. Discuss the first-step

$$\begin{aligned} h(1) &= \mathbf{P}_1(V_3 < V_4) \\ &= \sum_{x=1}^{4} \mathbf{P}(V_3 < V_4 | X_1 = x, X_0 = 1)\mathbf{P}(X_1 = x | X_0 = 1) \end{aligned}$$

$$\mathbf{P}(V_3 < V_4 | X_1 = x, X_0 = 1) = \begin{cases} \mathbf{P}_1(V_3 < V_4) = h(1) & x = 1 \\ \mathbf{P}_2(V_3 < V_4) = h(2) & x = 2 \\ 1 & x = 3 \\ 0 & x = 4 \end{cases}$$

$$\implies h(1) = 0.25h(1) + 0.6h(2)$$

similarly,

$$h(2) = 0.2h(2) + 0.7$$

Solve this system of equations, we have

$$\begin{cases} h(1) = 0.7 \\ h(2) = \frac{7}{8} \end{cases}$$

R To solve this question, we need to introduce the idea of First-Step Analysis.

1.9.2 General Result

Suppose $S = A \cup B \cup C$ where C is finite. Starting from any state in C, we are interested in the probability that the chain gets absorbed to set A rather than B ($V_A < V_B$), assuming this probability is positive. Given this setup, we have the following result.

Theorem 1.9.1 Let $S = A \cup B \cup C$, where A, B, C are disjoint sets, and C is finite. If $\mathbf{P}_x(V_A \wedge V_B < \infty) > 0$, for all $x \in C$. Then,

$$h(x) := \mathbf{P}_x(V_A < V_B)$$

is the unique solution of the system of equations

$$h(x) = \sum_y P_{xy} h(y), x \in C$$

with boundary conditions

$$h(a) = 1, a \in A \qquad\qquad h(b) = 0, b \in B$$

Proof. By first-step analysis,

$$\begin{aligned} h(x) &= \mathbf{P}(V_A < V_B | X_0 = x) \\ &= \sum_{y \in S} \mathbf{P}(V_A < V_B | X_0 = x, X_1 = y) \cdot \mathbf{P}(X_0 = x | X_1 = y) \\ &= \sum_{y \in S} P_{xy} h(y) = \sum_{y \in C} P_{xy} h(y) + \sum_{y \in A} P_{xy} \end{aligned}$$

boundary conditions hold trivially.

Hence, we only need to look at the uniqueness. Note that the system of equations can be written as

$$h' = Qh' + R'_A$$

where $h = (h(x_1), h(x_2), \dots)$ for $x_1, x_2, \dots \in C$. Note that

$$\begin{array}{c c} & \begin{array}{ccc} C & & A \quad B \end{array} \\ \begin{array}{c} C \\ \\ A \\ B \end{array} & \begin{pmatrix} Q & | & R \\ -- & -- & -- \\ & & \\ & & \end{pmatrix} \end{array} \qquad\qquad R'_A = \begin{pmatrix} \sum_{y\in A} P_{x_1,y} \\ \sum_{y\in A} P_{x_2,y} \\ \vdots \end{pmatrix}$$

The reason is that

$$\begin{aligned} h(x) &= \sum_{y\in S} P_{xy} h(y) \\ &= \underbrace{\sum_{y\in C} P_{xy} h(y)}_{=(Qh)(x)} + \underbrace{\sum_{y\in A} P_{xy}}_{=(R'_A)(x)} \end{aligned}$$

then,

$$\begin{aligned} I \cdot h' &= Qh' + R'_A \\ (I-Q)h' &= R'_A \end{aligned} \Longrightarrow h' = (I-Q)^{-1} R'_A$$

is unique as long as $I - Q$ is invertible. Note that

$$\begin{array}{c c} & \begin{array}{ccc} C & & A \quad B \end{array} \\ \begin{array}{c} C \\ \\ A \\ B \end{array} & \begin{pmatrix} Q & | & R \\ -- & -- & -- \\ & & I \quad \\ 0 & | & \quad I \end{pmatrix} \end{array}$$

Since for $\mathbf{P}_X(V_A < V_B)$, we only need to observe the chain before it hits A or B, the change of the rows in P corresponding to A and B will not change the result of this problem.
By doing this change, A and B are now **absorbing**, and all the states in C becomes transient! (since $\mathbf{P}_x(V_A \wedge V_B < \infty) > 0$). Now, we are working with a modified transition matrix P' and, therefore, a modified DTMC $\{X'_n\}_n$ To show $I - Q$ is invertible, note that since the states in C are transient (in P') and C is finite

$$\begin{aligned} 0 &= \lim_{n\to\infty} \mathbf{P}_x(X'_n \in C) \\ &= \lim_{n\to\infty} \sum_{y\in C} \left(P'^n\right)_{xy} \\ &= \lim_{n\to\infty} \sum_{y\in C} (Q^n)_{xy} \end{aligned}$$

The last equality holds because of the block structure of P', recall that

$$P' = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \Longrightarrow P'^n = \begin{pmatrix} Q^n & \cdots \\ 0 & I \end{pmatrix}$$

R There is another linear algebra way of looking at this (brute force).

$$\begin{aligned} Q^n_{xy} &= \sum_{x_1,\cdots,x_{n-1}\in C} Q_{xx_1}Q_{x_1x_2}\cdots Q_{x_{n-1}y} \\ &= \sum_{x_1,\cdots,x_{n-1}\in C} \mathbf{P}(X'_1 = x_1, X'_2 = x_2, \cdots, X'_n = y | X'_0 = x) \end{aligned}$$

then,

$$\begin{aligned} \sum_{y\in C} Q^n_{xy} &= \sum_{x_1,\cdots,x_{n-1},y\in C} \mathbf{P}(X'_1 = x_1, X'_2 = x_2, \cdots, X'_n = y | X'_0 = x) \\ &= \mathbf{P}(X'_1 \in C, X'_2 \in C, \cdots, X'_n \in C | X'_0 = x) \\ &= \mathbf{P}_x(\text{no absorption until time } n) \\ &= \mathbf{P}_x(X'_n \in C) \end{aligned}$$

This corresponds to the fact that in order to have $X'_n \in C$, we must have $X'_0, \ldots, X'_{n-1} \in C$ which implies that

$$\lim_{n\to\infty} Q^n = 0$$

as the zero matrix. Then, all the eigenvalues of Q have norm smaller than 1. Thus, there does not exist a non-zero f' such that

$$I \cdot f' = f' = Qf' \iff (I-Q)f' = 0$$

Thus, $I-Q$ is invertible. We are done! ■

1.9.3 Absorption Time

Similar to the absorption distribution part, but now we are interested in the **expected time** that the chain exits a part of the state space.
More precisely, let $S = A \cup C$, A, C disjoint and C is finite. Define

$$V_A := \min\{n \geq 0 : X_n \in A\}$$

which is the **first time the chain exits** C, which is also the **first time the chain hits/visits** A.
We want to know

$$g(x) := \mathbb{E}_x(V_A) = \mathbb{E}(V_A | X_0 = x), x \in C$$

Then, $g(a) = 0, \forall a \in A$. For $x \in C$, by first-step analysis, we have

$$\begin{aligned} g(x) &= \mathbb{E}_x(V_A) \\ &= \sum_{y\in S} \mathbb{E}(V_A | X_1 = y, X_0 = x)\mathbf{P}(X_1 = y | X_0 = x) \quad = \sum_{y\in S} P_{xy} \begin{cases} g(y)+1 & y \in C \\ 1 & y \in A \end{cases} \\ &= \sum_{y\in S} P_{xy} g(y) + 1 = \sum_{y\in C} \underbrace{P_{xy}}_{=Q_{xy}} g(y) + 1 \end{aligned}$$

Using matrix notation, this is

$$\begin{bmatrix} g(x_1) \\ \vdots \end{bmatrix} =: g' = Qg' + \mathbf{1}' \implies g' = (I-Q)^{-1}\mathbf{1}'$$

The matrix $I-Q$ is exactly the same as we have already seen in the part of absorption probability. We know it is invertible, hence, the solution is unique. In conclusion, we have g' as the unique solution of the system of equations $g' = Qg' + \mathbf{1}'$ given by $g' = (I-Q)^{-1}\mathbf{1}'$.

2. Discrete Phase-Type Distribution

2.1 General Setting

Let $\{X_n\}_n$ be a DTMC with finite states

$$S := \left\{ \underbrace{0, 1, \cdots, M-1,}_{\text{transient states } A} \; \underbrace{M, \cdots, N}_{\text{Absorbing states } B} \right\}$$

Then, without loss of generality, we can partition the state space into transient states and absorbing states, and rearrange them in the preceding order. This DTMC has the corresponding transition matrix.

$$P = \begin{array}{c} \\ 0 \\ \vdots \\ M-1 \\ M \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cc} 0 \cdots M-1 & M \cdots N \end{array} \\ \begin{pmatrix} Q & R \\ & \\ 0 & I \end{pmatrix} \end{array}$$

We define the **time until absorption** as

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

Assume we start from a general initial distribution

$$\vec{\alpha}_0 = (\underbrace{\alpha_{0,0}, \alpha_{0,1}, \cdots, \alpha_{0,M-1}}_{=:\vec{\alpha}_0'}, \alpha_{0,M}, \cdots, \alpha_{0,N})$$

where $\alpha_{0,i} = \mathbf{P}(X_0 = i)$.

R The first part, $\vec{\alpha}_0'$, the initial distribution of the transient states, turns out to be important.

We are interested in the exact distribution of T. Basically, calculate $\mathbf{P}(T=k), k \geq 1$.

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

where

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

For the edge case,

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

In summary, the probability mass function (pmf) of T is as follow,

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

Definition 2.1.1 — Discrete Phase-type Distribution (DPH). Such a distribution derived above is called a **discrete phase-type distribution**, and it is typically denoted as

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

where M is the dimension of the transient space, the number of transient states, and the dimension of Q, the transient part of the transition matrix. $\vec{\alpha}_0'$ is the transient part of the initial distribution.

R

1. We note that $\mathbf{P}(T=0)$'s value seems not included in the parametrization. But note that

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

which can be derived from $\vec{\alpha}_0'$ sufficiently.

2. The vector $\vec{q}$ is also not included in the parametrization since it can be computed as follow:

$$q_i = \sum_{j=M}^{N} P_{i,j} = 1 - \sum_{j=0}^{M-1} P_{ij} = 1 - \sum_{j=0}^{M-1} Q_{ij} \Longrightarrow \vec{q} = (I - Q)\mathbf{1}'$$

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

■ **Example 2.1 — Zero-Modified Geometric Distribution.** Consider the transition matrix

$$P = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{pmatrix} 1 & & \\ \alpha & \beta & \gamma \\ & & 1 \end{pmatrix} \end{array}$$

where $\alpha+\beta+\gamma = 1$. We want to get the distribution of absorption time to $\{0,2\}$ given the initial distribution $\vec{\alpha}_0$. Rearrange the state:

$$P' = \begin{array}{c c} & \begin{array}{ccc} 1 & 0 & 2 \end{array} \\ \begin{array}{c} 1 \\ 0 \\ 2 \end{array} & \begin{pmatrix} \beta & \alpha & \gamma \\ & 1 & \\ & & 1 \end{pmatrix} \end{array} \Longrightarrow \begin{pmatrix} \beta & 1-\beta \\ 0 & I \end{pmatrix}$$

Let $\vec{\alpha}'_0 = p = \mathbf{P}(X_0 = 1)$. Then, $\alpha_{0,0} + \alpha_{0,2} = 1 - p$. Then,

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

This is a **zero-modified** geometric distribution. In particular, if $p = \beta$, it is the geometric distribution $\text{geo}(1-p) = \mathbf{DPH}_1(p,p)$ (this counts the number of failure). ■

2.1.1 CDF of DPH

Let $T \sim \mathbf{DPH}_M(\vec{\alpha}_0, Q)$ (we omit the prime from now). For $k = 0,1,\cdots$, we have the CDF as

$$\begin{aligned} F_T(k) &= \mathbf{P}(T \le k) = 1 - \mathbf{P}(T > k) \\ &= 1 - \sum_{n=k+1}^{\infty} \mathbf{P}(T = n) \\ &= 1 - \sum_{n=k+1}^{\infty} \vec{\alpha}_0 Q^{n-1} \vec{q} \end{aligned}$$

we need to find (use magic)

$$\begin{aligned} S &= \sum_{n=k+1}^{\infty} Q^{n-1} = Q^k + Q^{k+1} + \cdots \\ SQ &= Q^{k+1} + Q^{k+2} + \cdots \\ S(I-Q) &= Q^k \\ S &= Q^k (I-Q)^{-1} \end{aligned}$$

R *Why the hell you can do this with a infinite series?* Think about all the eigenvalues of Q.

Thus,

$$F_T(k) = 1 - \vec{\alpha}_0 Q^k (I-Q)^{-1} \vec{q} = 1 - \vec{\alpha}_0 Q^k (I-Q)^{-1} (I-Q)\mathbf{1}' = 1 - \vec{\alpha}_0 Q^k \mathbf{1}'$$

And the survival function is $\mathbf{P}(T > k) = \vec{\alpha}_0 Q^k \mathbf{1}'$. This seems simple, huh? There is an intuitive interpretation. $\mathbf{P}(T > k)$ is essentially the probability of all the possible path that does not get in absorbing states before time k.

1. $\vec{\alpha}_0$ means we start from a transient state
2. Q^k means we stay in transient states until time k
3. $\mathbf{1}$ just adding all these path probabilities together at time k
4. Recall the $\vec{q}$ in pmf. It means we need to make sure at time k, the chain will go from transient to an absorbing state

Since T is a non-negative integer-valued random variable, we have

$$\mathbb{E}(T) = \sum_{k=0}^{\infty} \mathbf{P}(T > k)$$

Proof. Well, I will do it.

$$\begin{aligned}
\mathbb{E}(T) &= \sum_{k=0}^{\infty} k\mathbf{P}(T = k) = \sum_{k=1}^{\infty} k\vec{\alpha}_0 Q^{k-1}\vec{q} = \sum_{k=1}^{\infty} k\vec{\alpha}_0 Q^{k-1}(I-Q)\mathbf{1}' \\
&= \vec{\alpha}_0 \left(\sum_{k=1}^{\infty} kQ^{k-1}\right)(I-Q)\mathbf{1}' \\
&= \vec{\alpha}_0(I-Q)^{-2}(I-Q)\mathbf{1}' \qquad = \vec{\alpha}_0(I-Q)^{-1}\mathbf{1}'
\end{aligned}$$

note that

$$f(x) = \frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \Longrightarrow f'(x) = \frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1} = \sum_{k=1}^{\infty} kx^{k-1}$$

For the RHS just check the deduction below, really using the same series trick. ■

Then,

$$\begin{aligned}
\mathbb{E}(T) &= \sum_{k=0}^{\infty} \mathbf{P}(T > k) \\
&= \sum_{k=0}^{\infty} \vec{\alpha}_0 Q^k \mathbf{1}' \\
&= \vec{\alpha}_0 \sum_{k=0}^{\infty} Q^k \mathbf{1} \\
&= \vec{\alpha}_0 (I-Q)^{-1}\mathbf{1}'
\end{aligned}$$

The last step is intuitive but not trivial (why it converges?). We note that

$$(I-Q)^{-1}\mathbf{1}' = \begin{bmatrix} \mathbb{E}(T|X_0 = 0) \\ \mathbb{E}(T|X_0 = 1) \\ \vdots \end{bmatrix} = g'$$

This result, therefore, agrees with what we got previously using first-step analysis.

2.2 Properties of DPH

Proposition 2.2.1 The class of DPH is closed under several operations.

1. Sum of two independent DPHs is a DPH.

   Proof. Let $X \sim \mathbf{DPH}_m(\vec{\alpha}_0, S), Y \sim \mathbf{DPH}_n(\vec{\beta}_0, T)$ and $X \perp\!\!\!\perp Y$. Define $\vec{s}' = (I-S)\mathbf{1}'$ and $\vec{t}' = (I-T)\mathbf{1}'$. Let $Z = X + Y$. We need to construct explicit DTMC with a transition matrix to check DPH from first principle. Consider the following picture.

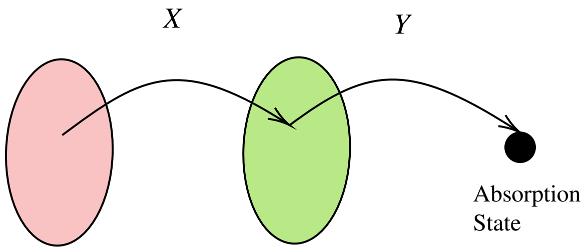


Figure 2.2.1: *How to construct Z?*

Let

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

where

$$\underbrace{\vec{s}'}_{m\times 1}\underbrace{\vec{\beta}_0}_{1\times n}$$

which is a $m \times n$ matrix. And $\beta_{0,n} = 1 - \sum_{i=0}^{n-1}\beta_{0,i}$. Why we gonna have these two weird terms here?

(a) We think of $\vec{s}'\vec{\beta}_0$ as the transient distribution of X then entering the initial transient distribution of Y.

(b) We think of $\beta_{0,n}\vec{s}'$ as the probability distribution of getting into the absorption state from X right away

R Moreover, we can think of $[Y, \text{abs}]$ as the absorbing states for X and abs as the absorbing states for Y. The block matrix

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

is the transient part of the new Markov chain Z as we will use this to characterize Z.

We also need initial distribution for Z to characterize a **DPH**. Let

$$\vec{\gamma}_0 = [\vec{\alpha}_0, \alpha_{0,m}\vec{\beta}_0]$$

be a $1 \times (m+n)$ row vector where $\alpha_{0,m} = 1 - \sum_{i=0}^{m-1}\alpha_{0,i}$. $\vec{\alpha}_0$ is the transient part of X and $\alpha_{0,m}\vec{\beta}_0$ is the tranisent part of Y given that we get into Y from X. Now, all dimensions make sense. Then, we have

$$Z \sim \textbf{DPH}_{m+n}(\vec{\gamma}_0, C)$$

■

2. The mixture of two DPHs is a DPH. Let $X \sim \textbf{DPH}_m(\vec{\alpha}_0, S), Y \sim \textbf{DPH}_n(\vec{\beta}_0, T)$ and $X \perp\!\!\!\perp Y$. Let

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

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

Proof. Assignment problem. *Hint:* construct a DTMC like in 1.

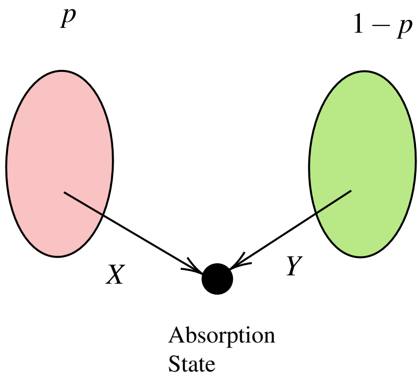


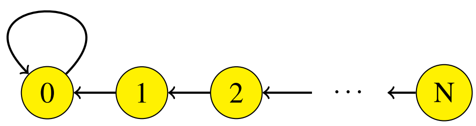
Figure 2.2.2: *How to construct Z?*

■

3. Any discrete distribution on the non-negative integers with finite support is a DPH. ("with finite support" mean $\{n : \mathbf{P}(X = n) > 0\}$ is finite)

■ **Example 2.2** Binomial and uniform distributions have finite support while geometric and Poisson have countably infinite support. ■

Proof. Let the distribution be $\{b_n\}_{n=0}^{N}$, this means $\mathbf{P}(T = n) = b_n$. We again need to construct a DTMC in a very deterministic way.



with

$$P = \begin{array}{c c} & \begin{array}{cccccc} N & N-1 & \cdots & 1 & 0 \end{array} \\ \begin{array}{c} N \\ N-1 \\ \vdots \\ 1 \\ 0 \end{array} & \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \end{array}$$

where 0 is the absorbing state and the transient part of the transition matrix is the sub-diagonal nilpotent matrix S. Note that with this chain, the absorption time is determined by X_0. $T = n$ if and only if $X_0 = n$. Hence we only need to make the initial distribution to be

$$(b_N, b_{N-1}, \cdots, b_1, b_0)$$

with $\vec{\alpha}_0 = [b_N, b_{N-1}, \cdots, b_1]$ being the transient part of the distribution. Therefore, we characterize

$$\{b_n\} \sim \mathbf{DPH}_N(\vec{\alpha}_0, S)$$

As a result, one can approximate any distribution on non-negative integers arbitrarily closely by DPH. Assume we want to approximate $\{b_n\}$ with countably infinite support by a DPH such thht for T following this DPH,

$$\sum_n |\mathbf{P}(T = n) - b_n| < \varepsilon$$

the error is small. We can do this since $\sum_n b_n = 1$, there exists N such that the tail sum $\sum_{n=N}^{\infty} b_n < \frac{\varepsilon}{2}$. Then, we can simply set

$$T \sim \left[b_0, b_1, \cdots, b_{N-2}, b_{N-1} + \sum_{n=N}^{\infty} b_n \right]$$

alternatively, we can define

$$T \sim c[b_0, b_1, \cdots, b_{N-2}, b_{N-1}]$$

by some scaling factor to make this distribution valid $c = \frac{1}{\sum_{n=0}^{N-1} b_n}$. Both constructions are DPHs as they have finite supports now. ■

> **R** Well, as you can see here. Our construction seems very simple and intuitive. We can even do this with some deterministic DTMC. This should tell you that our construction is not the unique one. Consider the following example.

■ **Example 2.3 — DPH representation is indeed not unique..** Let $A \sim \mathbf{DPH}_1(\beta, \beta)$ which is a geometric distribution with $1-\beta$ as the parameter. Let

$$B \sim \mathbf{DPH}_2\left(\begin{bmatrix} \beta \\ 0 \end{bmatrix}, \begin{bmatrix} \beta-\alpha & \alpha \\ 0 & \beta \end{bmatrix} \right)$$

where $0 < \alpha < \beta < 1$. We want to show B also characterize a geometric distribution. One can find the pmf/cdf of this DPH using the result we derived in previous sections. Intuitively, since

$$P = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & \text{Abs} \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \text{Abs} \end{array} & \begin{pmatrix} \beta-\alpha & \alpha & 1-\beta \\ 0 & \beta & 1-\beta \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

we note that there are two entries in the absorption states are the same. This mean there is always a fixed probability $1-\beta$ of getting absorbed given that it has not been absorbed earlier. Isn't this just $\text{Geo}(1-\beta)$? The initial distributions also match. Thus, the second DPH is also a $\text{Geo}(1-\beta)$.
Which one to choose then? In general, when we have different ways to represent a distribution as DPHs, we would prefer a representation with the smallest dimension. ■

3. Review of Poisson Processes

3.1 Exponential and Poisson Distributions

Consider the exponential distribution. The readers are expected to know its pdf, cdf, expectation, variance, mgf, and cf.

Memoryless property

For $X \sim \mathbf{Exp}(\lambda)$, then

$$\mathbf{P}(X > t+s | X > t) = \mathbf{P}(X > s),\ t, s \geq 0$$

this means whether we have waited for t units of time does not matter, we still need to wait for s units as we just start waiting.

Min of independent exponentials

Let $X_1, X_2, \cdots, X_n$ be a family of independent $\mathbf{Exp}(\lambda_i)$ r.v.s. Then,

$$\min\{X_i\}_i \sim \mathbf{Exp}\left(\sum_i \lambda_i\right)$$

moreover,

$$\mathbf{P}(\min\{X_i\}_i = X_k) = \frac{\lambda_k}{\sum_i \lambda_i}$$

Consider Poisson distribution. The readers are expected to know the pmf, expectation, variance, mgf, and cf.

Independent Poisson sum

Let $X_1, X_2, \cdots, X_n$ be a family of independent $\mathbf{Poi}(\lambda_i)$ r.v.s, then

$$\sum_{i=1}^{n} X_i \sim \mathbf{Poi}\left(\sum_i \lambda_i\right)$$

3.2 Poisson Process and Properties

3.2.1 Time-Homogeneous Poisson Process

Definition 3.2.1 — Time-homogeneous Poisson process. A time-homogeneous Poisson process is a renewal process with exponential interarrival time.

Basic properties

1. **Continuous-time Markov property:** this means

$$\mathbf{P}(\underbrace{N(t_m)=j}_{\text{future}}|\underbrace{N(t_{m-1})=i}_{\text{present}},\underbrace{N(t_{m-2})=i_{m-2},\cdots,N(t_i)=i_1}_{\text{past}})$$
$$=\mathbf{P}(N(t_m)=j|N(t_{m-1})=i)$$

   for any m, $t_1<t_2<\cdots<t_m$, and $i_1,\cdots,i_{m-1},i,j\in S$.
2. **Independent increments:** $N(t_1)-N(t_0),N(t_2)-N(t_1),\cdots,N(t_n)-N(t_{n-1})$ are all independent as long as $\{[t_i,t_{i+1}]\}_{i=0}^{n-1}$ intervals do not have overlaps.
3. **Poisson increment:**

$$N(t_2)-N(t_1)\sim\mathbf{Poi}(\lambda(t_2-t_1))$$

   where λ is the intensity/rate. In particular, this implies that $N(t)\sim\mathbf{Poi}(\lambda t)$.
4. **Zero start**: $N(0)=0$

Theorem 3.2.1 — Characterization of Poisson process. If $\{N(t)\}$ satisfies independent increments, Poisson increment, and zero start, then $\{N(t)\}$ is a time-homogeneous Poisson process.

Combining and splitting

Proposition 3.2.2 — Combining. Let $\{N_1(t)\}\sim\mathbf{Poi}(\lambda_1 t)$ and $\{N_2(t)\}\sim\mathbf{Poi}(\lambda_2 t)$ be independent. Then, $\{N(t)=N_1(t)+N_2(t)\}\sim\mathbf{Poi}((\lambda_1+\lambda_2)t)$.

Splitting is like the reverse of combining.

Proposition 3.2.3 — Splitting. Let $\{N(t)\}\sim\mathbf{Poi}(\lambda t)$ where

$$N(t)=\begin{cases}N_1(t) & \text{with probability } p\\ N_2(t) & \text{with probability } 1-p\end{cases}$$

this relation is independent of everything else. Then, $\{N_1(t)\}\sim\mathbf{Poi}(p\lambda t)$ and $\{N_2(t)\sim\mathbf{Poi}((1-p)\lambda t)\}$. Moreover, these two processes are also independent.

Order statistics property

Proposition 3.2.4 — Order statistics property. Let S_i be the occurrence time of the ith event in a Poisson process with a fixed intensity. Then,

$$S_1,\cdots,S_n|N(t)=n\overset{d}{=}(U_{(1)},\cdots,U_{(n)})$$

where $U_{(1)},\cdots,U_{(n)}$ are order statistics of $U_1,\cdots,U_n\overset{\text{i.i.d}}{\sim}U([0,t])$. How do we interpret this? This means given that at time t, we have n occurrences, to obtain the distribution of the occurrence time in order $1,\cdots,n$ is no different than roll n continuous uniform random variable on $[0,t]$ and take their order statistics.

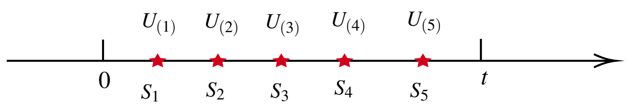


Figure 3.2.1: *Order statistics property*

3.2.2 Non-homogeneous Poisson Process

Definition 3.2.2 — Non-homogeneous Poisson process. We say $\{N(t)\}$ is a non-homogeneous Poisson process if it satisfies the following:

1. $N(0) = 0$
2. $N(t) - N(s) \sim \mathbf{Poi}\left(\int_s^t \lambda(r)dr\right)$ where λ is a rate function.

   R We note that when the rate function is constant, it goes back the homogeneous definition.

3. independent increments

Properties

1. no longer a renewal processs
2. **Continuous-time Markov property** is still true
3. **Independent increments** is still true
4. **Poisson increments** still holds
5. Combining and splitting still hold but we have different rate functions
   (a) For combining, we now have
   $$\lambda_1(r), \lambda_2(r) \Longrightarrow \lambda(r) = \lambda_1(r) + \lambda_2(r)$$
   (b) For splitting, we now have
   $$N(t) = \begin{cases} N_1(t) & \text{with probability } p(t) \\ N_2(t) & \text{with probability} 1 - p(t) \end{cases}$$
   and the new rate functions are $\lambda_1(r)p(r)$ and $\lambda_2(r)(1-p(r))$.
6. **Order statistics property** still holds but modified. The occurrence times are no longer have the same joint distribution as order statistics of uniform r.v.s but i.i.d r.v.s with density function
   $$f(s) = \frac{\lambda(s)}{\int_0^t \lambda(s)ds}, \ s \in [0,t]$$
   this means the occurrence time should distribution with proportion to their rate function. For example, in the function below, the rate function might be larger in the blue interval, therefore, there could be more occurrences during that period.

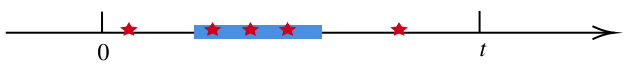


Figure 3.2.2: *Non-uniform density*

3.2.3 Compound Poisson process

A **compound Poisson process** $S(t)$ is

$$S(t) = Y_1 + \cdots + Y_{N(t)}$$

where $N(t)$ is some Poisson process and Y_is are i.i.d.

Proposition 3.2.5 — Decomposition of Variance (EVVE Law). Let X, Y be random variables. Then,

$$\mathbf{Var}(X) = \mathbb{E}\left(\mathbf{Var}(X|Y)\right) + \mathbf{Var}(\mathbb{E}(X|Y))$$

For a Poisson process with rate λ, $S(t)$ has a mean $\lambda + \mu$ where $\mu = \mathbb{E}(Y_1)$ and a variance $\lambda t \mathbb{E}(Y_1^2) = (\mu^2 + \sigma^2)\lambda t$ where $\sigma^2 = \mathbf{Var}(Y_1)$.

II

Continuous-Time Markov Chain

4. Continuous-Time Markov Chain (CTMC)

4.1 Basic Setup

Definition 4.1.1 — Continuous-Time Markov Chain (CTMC). A stochastic process $\{X_t\}_{t\geq 0}$ is called a continuous-time Markov chain (CTMC), if

1. For $t \geq 0$, $X(t)$ takes possible values in an at most countable set S. Typically, $S = \{0, 1, 2, \cdots\}$ can be standardized.
2. **Markov property:** for $s, t \geq 0$, $i, j \in S$, then

$$\mathbf{P}(X(t+s) = j | X(s) = i, X(u) = x_u, 0 \leq u \leq s) = \mathbf{P}(X(t+s) = j | X(s) = i)$$

R The first part of the definition of a CTMC requires S to be at most countable set. So you are looking at the naturals, integers, and the rationals. But this is only for simplification. Consider a stock price process which could take on any positive real value (another possible reason why Brownian motion is desirable for modeling financial asset prices). This version will have some gap in S and the process is called **jump processes**.

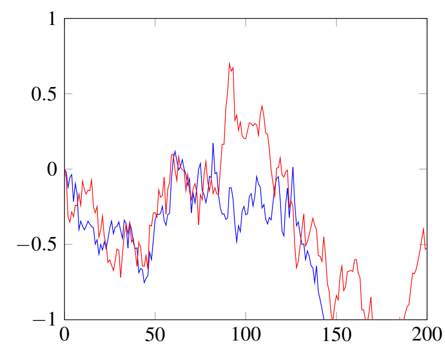


Figure 4.1.1: *Brownian motion is a CTMC with uncountable state space*

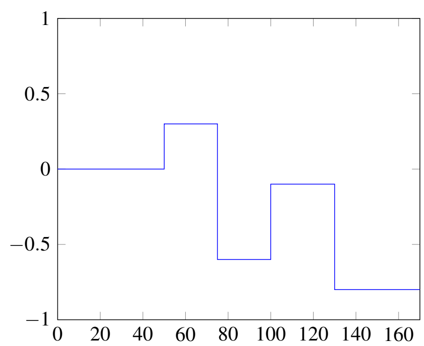


Figure 4.1.2: *CTMC with countable state space-jump processes*

The second part is analogous to Markov property in discrete time. The intuition is, given the present state, the probabilistic properties of the future only depends on the current state, not the past.

We call the Markov chain **(time)-homogeneous** if $\mathbf{P}(X(t+s)=j|X(s)=i)$ does not depend on time s. We will take this as a default setting. In this case, we define

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

Visualization of CTMC

Our previous graph of DTMC might no longer be a good idea for visualization since we cannot see how long the CTMC stays at a particular state. In particular, this random amount of time spent at certain state is called **Sojourn time**. Instead, we consider it as a function of time.

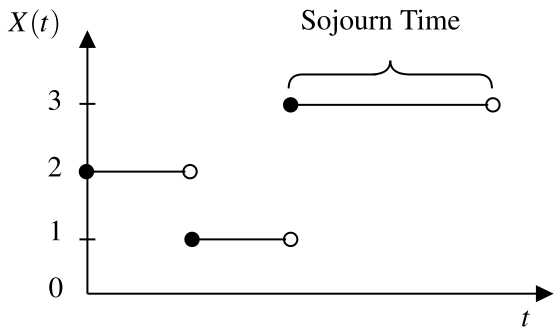


Figure 4.1.3: *Visualization of a CTMC $X(t)$*

We note that, to characerize a CTMC, essentially, we need to answer the following two questions:

1. How long does this CTMC stay in a state $i \in S$?
2. Once the CTMC leaves the current state, how to decide which state it will enter?

4.1.1 Sojourn Time

The answer to the first question in the characterization problem for CTMC is as follow.

Let T_i be the random amount of time that the CTMC stays in state i (the Sojourn time at i). Then T_i follows an exponential distribution.

Why? (Proof Sketch)

Consider the scenario where we enter the state i at time w and it leaves state i later than time $w+t+s$.

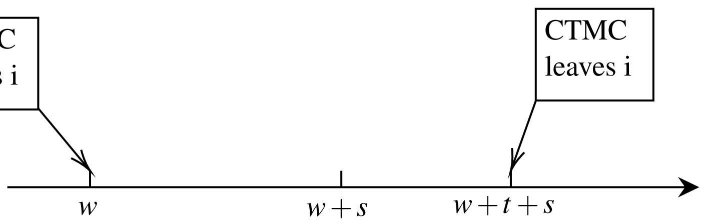


Figure 4.1.4: *Why T_i is an exponential?*

Then,

$$\begin{aligned}
\mathbf{P}(T_i > t+s | T_i > s) &= \mathbf{P}(X(u) = i, \forall u \in [w, w+t+s] | X(u) = i, \forall u \in [w, w+s]) \\
&= \mathbf{P}(X(u) = i, \forall u \in (w+s, w+t+s] | X(u) = i, \forall u \in [w, w+s]) \\
&= \mathbf{P}(X(u) = i, \forall u \in (w+s, w+t+s] | X(w+s) = i, X(u) = i, \forall u \in [w, w+s)) \\
&\overset{\text{Markov Prop.}}{=} \mathbf{P}(X(u) = i, \forall u \in (w+s, w+t+s] | X(w+s) = i) \\
&\overset{\text{Time Homo.}}{=} \mathbf{P}(X(u) = i, \forall u \in (w, w+t] | X(w) = i) \\
&= \mathbf{P}(T_i \geq t)
\end{aligned}$$

This is indeed the memoryless property. And we know that exponentials are the only family of distribution that has the thsi property in continuous random variables. Let the parameter be V_i. Then, $T_i \sim \mathbf{Exp}(V_i)$. 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 $\frac{1}{V_i}$. Thus, the mean sojourn time is $\frac{1}{V_i}$. If $V_i = 0$, then $\mathbf{P}(T > t) = 1$ for any $t \geq 0$. This means $T_i = \infty$ $\mathbf{P}-$a.s. and state i is **absorbing**.

4.1.2 Discrete Skeleton

Now, we answer the second question regarding the characterization of a CTMC. In short, it can be summarized as

> How long we stay in a state is **independent** of where we are going once we decide to leave the state.

The transition probability, once the CTMC leaves i, does not depend on the sojourn time T_i.

Why? (Proof Sketch)

By Markov property, the transition probability only depends on the current state, but the current state does not provide information about sojourn time.
Thus, we can denote the transition probability taht when the CTMC leaves i, it goes to j, by P_{ij}, which only depends on i and j.

Summary: how does a CTMC behave?

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 exponential amount of time T_j, then "jumps"...
This $\{P_{ij}\}_{i,j \in S}$ is in fact a transition matrix of a DTMC. However, note that we have one extra condition: $P_{ii} = 0$. by definition. Thus,

$$\sum_{j \in S} P_{ij} = \sum_{j \in S, i \neq j} P_{ij} = 1, \forall i \in S$$

The DTMC corresponding to $\{P_{ij}\}_{i,j \in S}$ governs all the "jumps" of the CTMC, but does not record sojourn times. This is called the **discrete skeleton/embedded DTMC** of the CTMC (time is flesh...eww). This can be interpreted as only taking photos of the CTMC when it jumps, thus gives a discrete change of state when you look at the photos. (Or should I say X-ray?)

4.2 Kolmogorov Backward and Forward Equations

How does $P_{ij}(t) = \mathbf{P}(X(t) = j | X(0) = i) = \mathbf{P}(X(t+s) = j | X(s) = i)$ change as a function of t? To derive this dynamics, we first define

$$q_{ij} := V_i P_{ij}$$

where V_i is the intensity (like current) going out of state i and P_{ij} is the probability of entering j from $i \neq j$. This q_{ij} can be considered as the "probability flow" going from i to j. Note that

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

for h small,

$$\begin{aligned} P_{ii}(h) &= \mathbf{P}(X(h) = i | X(0) = i) \\ &= \mathbf{P}(T_i > h) + \mathbf{P}(\text{at least 2 transitions happen between 0 and h}, X(h) = i | X(0) = i) \\ &= \mathbf{P}(T_i > h) + o(h) \end{aligned}$$

Digression: higher order infinitesimal of h ($o(h)$)

We say $f(h) = o(h)$ if $\lim_{h\to 0} \frac{f(h)}{h} = 0$. So why we have the last equality? In A2, we shall see a alternative version of the memoryless property

$$\mathbf{P}(T_i < h) = \mathbf{P}(\text{at least 1 transition}) = V_i h + o(h)$$

Then, since $T_i \sim \mathbf{Exp}(V_i)$, we have

$$P_{ii}(h) = e^{-V_i h} + o(h) = 1 - V_i h + o(h) + o(h) = 1 - V_i h + o(h)$$

This means

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

Now, for $i \neq j$,

$$\begin{aligned} P_{ij}(h) &= \mathbf{P}(X(h) = j | X(0) = i) \\ &= P_{ij}(\mathbf{P}(T_i \leq h) - \mathbf{P}(\text{at least 2 transitions})) + o(h) \\ &= P_{ij}(1 - e^{-V_i h} - o(h)) + o(h) \\ &= P_{ij}(V_i h - o(h) - o(h)) + o(h) \\ &= P_{ij} V_i h + o(h) = q_{ij} h + o(h) \end{aligned}$$

thus,

$$\lim_{h\to 0} \frac{P_{ij}(h)}{h} = q_{ij} \quad (2)$$

Now we have two equations. Let $P(t) = \{P_{ij}(t)\}_{i,j\in S}$ be the transition matrix at time t. Then, $P(0) = I$ since $P_{ii}(0) = 1, P_{ij}(0) = 0, j \neq i$. Combining (1) and (2), we have

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

Note that the LHS side is a Newton quotient of matrix valued function, which resembles a derivative (even though we are dealing with a matrix). Thus, this matrix R is called the infinitesimal generator of this CTMC. It combines the information of sojourn times $\{V_i\}_{i\in S}$ and transitions $\{P_{ij}\}_{i,j\in S}$. In particular, if we write $R = \{R_{ij}\}_{i,j\in S}$, note that

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

R The row sum of R are always 0.

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

CTMC C-K Equation

C-K equation still holds in continuous time.

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

In matrix notation,

$$P(t+s) = P(t)P(s)$$

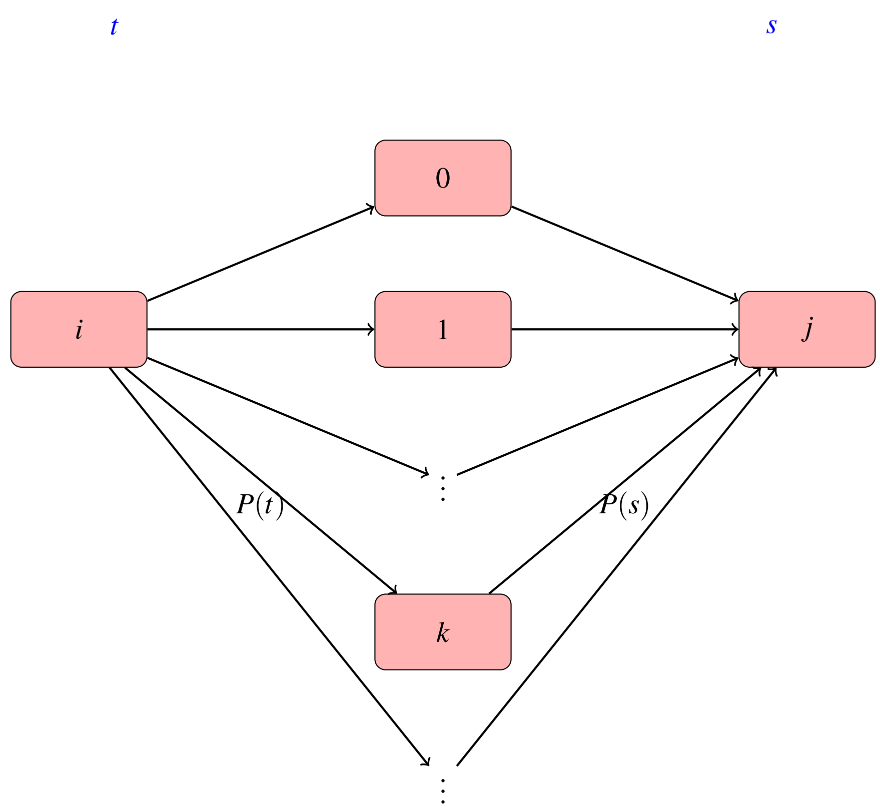


Figure 4.2.1: *Intuition of C-K Equation*

Thus, we have $P(t+h) = P(h)P(t)$

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

$P'(t) = RP(t)$ is called the **Kolmogorov Backward Equation**.
Similarly, $P(t+h) = P(t)P(h)$,

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

this called the **Kolmogorov Forward Equation**.

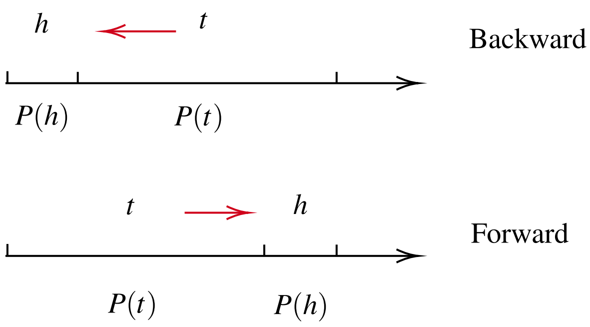


Figure 4.2.2: *Visualization of the backward and forward equations: which direction to change the length of the time interval?*

In entry-wise-form:

1. **Backward equation:**

$$\begin{aligned} P'_{ij}(t) &= \sum_{k\in S} R_{ik}P_{kj}(t) \\ &= \sum_{k\neq i} q_{ik}P_{kj}(t) - V_i P_{ij}(t) \end{aligned}$$

2. **Forward equation:**

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

R In the derivation we actually interchanged the order of a limit and a summation. For example,

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

these two lines are actually different where the first sequence is actually a sequence of summations. But the second line is the summation of the sequence limit. How is this valid? This can be justified if

1. The state space S is finite; this is clear since we are not dealing with limit of partial sums.
2. For backward equation, this is always valid. Check bounded convergence condition.

In this sense, backward equation is more reliable and fundamental. However, forward equation is usually easier to solve.

Now, we know that $P(t)$ satisfies the matrix differential equations

$$P'(t) = RP(t)$$

with initial condition $P(0) = I$. If everything were scalar, we should get e^{tR} as the solution.

Question:

now with everything being matrix, whaty should we get?
The answer is, we still get $P(t) = e^{tR}$, defined in the following way:

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

One can show that the sum above always converges.
Show that such defined e^{tR} is the solution of the matrix differential equation

$$P'(t) = RP(t)$$

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

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.
But how can we calculate this infinite matrix series? In general, $P(t) = e^{tR}$ is not easy to calculate, in particular, it is warned that

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

this is not entry-wise computation. However, if the matrix R is diagonalizable, that is there exists invertible matrix B, s.t., $R = BDB^{-1}$, where D is diagonal

$$D = \begin{bmatrix} d_0 & & \\ & d_1 & \\ & & \ddots \end{bmatrix}$$

then,

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

for diagonal matrix D,

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

Thus, 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, probably use this method.

4.3 Birth and Death Processes

Think $X(t)$ as the number of individuals (population) in a system at time t. At any time, the population can only increase by 1 (birth), decrease by 1 (death), or remain the same (undying). That is,

$$P_{ij} = 0, |i - j| > 1$$

this is equivalent to

$$q_{ij} = 0, |i - j| > 1$$

Note that this is tri-diagonal.

Tus, the generater R in this case takes the form

$$R = \begin{bmatrix} -v_0 & v_0 & & & \\ q_{10} & -v_1 & q_{12} & & \\ & q_{21} & -v_2 & q_{23} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} =: \begin{bmatrix} -\lambda_0 & \lambda_0 & & & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

we call $\{\lambda_i\}_{i=0,1,\cdots}$ are called **birth rates** and $\{\mu_i\}_{i=1,2,\cdots}$ are called **death rates**.
For $i \geq 1$, $\lambda_i = q_{i,i+1} = v_i P_{i,i+1}$ and $\mu_i = q_{i,i-1} = v_i P_{i,i-1}$. Then,

$$\begin{cases} P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i} \\ P_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i} \end{cases}, \quad v_i = \lambda_i + \mu_i$$

for $i = 0$, $P_{0,1} = 1$, $v_0 = \lambda_0$.

Intuition

When the system is in state i, the next (potential) birth happens after an exponential amount of time with intensity (rate) λ_i. The next (potential) death happens after an exponential amount of time with intensity (rate) μ_i. These two times will "compete" with each other. The smaller one "wins" (i.e., happens), changing the system to the next state. Recall that if $T_i^{(b)} \sim \mathbf{Exp}(\lambda_i), T_i^{(a)} \sim \mathbf{Exp}(\mu_i)$ independently, then

1. $\mathbf{P}(T_i^{(b)} < T_i^{(a)}) = \frac{\lambda_i}{\lambda_i + \mu_i}$ and $\mathbf{P}(T_i^{(a)} < T_i^{(b)}) = \frac{\mu_i}{\lambda_i + \mu_i}$

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

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

4.3.1 Special Cases of Birth and Death Processes

Definition 4.3.1 — Pure Death Process. When $\lambda_i = 0, i = 0, 1, 2, \cdots$. No birth, this is called the **pure death process**.

Definition 4.3.2 — Pure birth process. When $\mu_i = 0, i = 1, 2, \cdots$. No death, this is called the **pure birth process**.

Definition 4.3.3 — Poisson process. When $\mu_i = 0$ and $\lambda_i = \lambda > 0$. The time between births are iid exponentials with the same intensity λ regardless the state i. This is the counting process of exponential events, a Poisson process.

Definition 4.3.4 — Yule process. When $\mu_i = 0$ and $\lambda_i = i\lambda, i = 0, 1, \cdots$. The birth rate is proportional to the current population.

R This is intuitive. Each individual has the same constant birth rate λ, and the individuals are independent. Thus, the time until the next birth, given current population i, is the smallest among i iid exponential r.v.s with parameter λ each. Then, this birth time follows $\mathbf{Exp}(i\lambda)$.

Definition 4.3.5 — Linear growth model. When $\lambda_i = i\lambda, \mu_i = i\mu$. Individuals are independent and have the same birth rate λ and death rate μ.

Definition 4.3.6 — M/M/S Quene. Detailed discussion in "quening theory" part.

4.3.2 Forward Equation for Pure Birth Processes

Consider the forward equations for birth and death processes:

$$P'(t) = P(t)R$$

the $i-$th row:

$$(P'_{i0}(t), P'_{i1}(t), \cdots, P'_{i,j-1}(t), P'_{i,j}(t), P'_{i,j+1}(t), \cdots)$$

$$=(P_{i0}(t), P_{i1}(t), \cdots, P_{i,j-1}(t), P_{i,j}(t), P_{i,j+1}(t), \cdots)\begin{bmatrix} -\lambda_0 & \lambda_0 & & & & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & & \\ & \mu_2 & -(\lambda_2+\mu_2) & \ddots & & \\ & & \mu_3 & \ddots & \lambda_{j-1} & \\ & & & \ddots & -(\lambda_j+\mu_i) & \ddots \\ & & & & \mu_{j+1} & \ddots \\ & & & & & \ddots \end{bmatrix}$$

that is, entry-wise,

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

For the special case of **pure birth process**, where $\mu_i = 0, i = 1, 2, \cdots$, we have

$$P_{ij}(t) = 0, \ j < i$$

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 (1) is clearly

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

Intuitively, sincwe the CTMC is a pure birth process,

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

For $j > i$, (2) becomes

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

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

4.3.3 Forward Equation for Poisson Process

Now, let us look at an even more special case where $\lambda_i = \lambda$ and $\mu_i = 0$. We can see that

$$P_{i,i+k}(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k = 0, 1, 2, \cdots$$

we shall prove this is the solution of the recursive formula presented above inductively.

1. **Base case:** $P_{ii}(t) = e^{-\lambda t}$ by previous result ($k = 0$)
2. **Inductive step:** assume that

$$P_{i,i+k}(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

now, by the recursive formula, we get

$$
\begin{aligned}
P_{i,i+k+1}(t) &= \lambda_{i+k} e^{-\lambda_{i+k+1}t} \int_0^t e^{\lambda_{i+k+1}s} P_{i,i+k}(s) ds \\
&= \lambda e^{-\lambda t} \int_0^t e^{\lambda s} e^{-\lambda s} \frac{(\lambda s)^k}{k!} ds \\
&= \lambda e^{-\lambda t} \int_0^t \frac{(\lambda s)^k}{k!} ds \\
&= \lambda e^{-\lambda t} \left[\frac{1}{\lambda} \frac{(\lambda s)^{k+1}}{k!(k+1)} \right]_{s=0}^t \\
&= e^{-\lambda t} \frac{(\lambda t)^{k+1}}{(k+1)!}
\end{aligned}
$$

and we are done.

4.4 Classification of States

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

$$
\begin{aligned}
(\vec{\alpha}_t)_j := \mathbf{P}(X(t) = j) &= \sum_{i \in S} \mathbf{P}(X(t) = j | X(0) = i) \mathbf{P}(X(0) = i) \\
&= \sum_{i \in S} \alpha_{0,i} P_{ij}(t) \qquad\qquad \Longrightarrow \vec{\alpha}_t = \vec{\alpha}_0 P(t) \\
&= (\vec{\alpha}_0 P(t))_j
\end{aligned}
$$

where $\vec{\alpha}_t$ is the distribution of $X(t)$ and $\vec{\alpha}_0$ is the initial distribution. Thus, the distribution of $X(t)$ is determined by the initial distribution and the transition matrix, where $P(t)$ is determined by the infinitestimal generator R.
Let f be a function defined on the state space, $f : S \to \mathbb{R}$. Then,

$$
\begin{aligned}
\mathbb{E}(f(X(t))) &= \sum_{j \in S} f(j) \mathbf{P}(X(t) = j) \\
&= \vec{\alpha}_t \vec{f} \\
&= \vec{\alpha}_0 P(t) \vec{f}'
\end{aligned}
$$

where $\vec{f}'$ is the column vector of outputs of f on S.
Recall that if a CTMC $\{X(t)\}$ is only observed when the state changes, the resulting process is a DTMC, denoted as $\{X_n\}_{n=0,1,\cdots}$. This is called 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 $P_{ij} = \mathbf{P}(X(T_i) = j | X(0) = i)$. If $V_i = 0$, i is absorbing, then define $P_{ii} = 1$ and $P_{ij} = 0$ for $j \neq i$. The question is, what do these concepts in discrete time become?

4.4.1 Accessibility and Communication

Definition 4.4.1 — Accessible. State j is said to be accessible from state i (or i is said to communicate to j), if

$$\mathbf{P}(X(t) = j | X(0) = i) = P_{ij}(t) > 0$$

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

Definition 4.4.2 — Communicate with each other. If i, j are said to communicate with each other, if $i \to j, j \to i$. Denote it as $i \leftrightarrow j$.

> **R** Note that 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).

It turns out that we have something simple for communication.

Proposition 4.4.1 For $i \neq j$ in a CTMC, $i \to j$ if and only if $i \to j$ in the discrete skeleton.

Proof. Assignment 3. ■

Definition 4.4.3 — Communicating class. A set of states $C \subseteq S$ is called a communicating class, if for any $i, j \in S$, $i \leftrightarrow j$, and for any $i \in C, j \notin C$, $i \nleftrightarrow j$ (one direction is fine).

Definition 4.4.4 — Irreducible CTMC. A CTMC is called irreducible, if all its states are in the same class. In other words, all its states communicate with each other.

Clearly, a CTMC is irreducible if and only if its discrete skeleton is irreducible by the preceding proposition.

4.4.2 Recurrence and Transience

Let R_{ii} be the amount of (continuous) time until the CTMC revisits state i given $X(0) = i$.

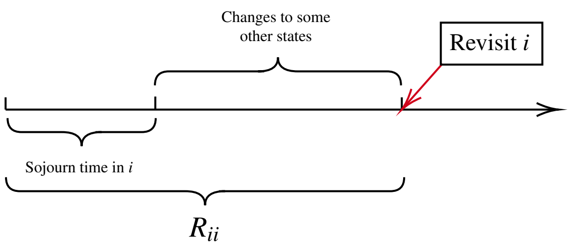


Figure 4.4.1: *Definition of R_{ii}*

Define $R_{ii} = \infty$ if the chain never revisits i.

Definition 4.4.5 — Recurrent/transient state. The state i is called recurrent, if either

$$\mathbf{P}(R_{ii} < \infty) = 1$$

or i is absorbing (This means the CTMC will revisit state i for sure).
The state i is called transient, if

$$\mathbf{P}(R_{ii} = \infty) > 0$$

Note that a CTMC revisits a (non-absorbing) state i if and only if its discrete skeleton revisits state i. A state i is recurrent if and only if it is recurrent in the discrete skeleton of the CTMC.
As a result, recurrence/transience are class properties.

Definition 4.4.6 — Irreducible recurrent/transient CTMC. An irreducible CTMC is called recurrent/transient if and only if all its states are recurrent/transient.

> **R** This is equivalent to one of its states being recurrent/transient.

An irreducible CTMC is recurrent/transient if and only if its discrete skeleton is recurrent/transient.

Difference: periodicity

There is no concept of periodicity in continuous time.
Why? If $\mathbf{P}(X(t) = j|X(0) = i) > 0$ for some $t \geq 0$, then $\mathbf{P}(X(t) = j|X(0) = i) > 0$ for all $t \geq 0$ because the sojourn times are exponential which can be arbitrarily large or small.

4.4.3 Positive/Null Recurrence

Definition 4.4.7 — Positive/null recurrent. A state i is called positive recurrent if $\mathbb{E}(R_{ii}) < \infty$ or i is absorbing (expected time it takes for the chain to return to state i is finite).
i is called null recurrent if it is recurrent but $\mathbb{E}(R_{ii}) = \infty$.

The next question is: **is positive/null recurrence always the same for a CTMC and its discrete skeleton?**
First, why recurrence/transience is the same for both CTMC and discrete skeleton? This is intuitive since these two definition only concerns whether a state will be revisted or not. They do not consider the time factor. This **foreshadows** that positive/null recurrence criterion might be difference between the discrete skeleton and CTMC.

■ **Example 4.1** Consider a DTMC with one-step transition matrix

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

which can be represented by the following graph

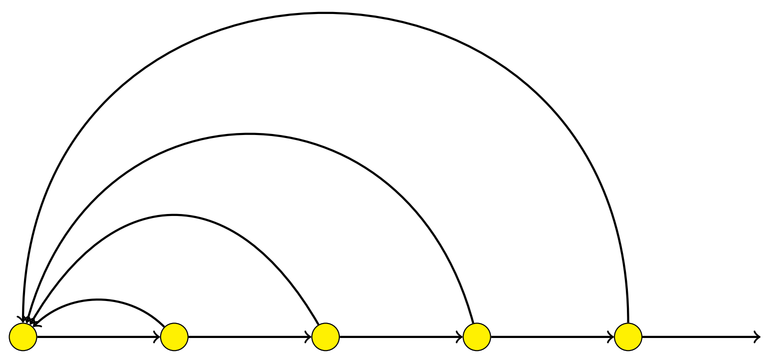


Figure 4.4.2: $p \in (0,1), q = 1-p$

Note that this chain is irreducible. As long as the chain is not in state 0, in each step, it has a fixed probability p or returning to state 0. Thus, 0 is a recurrent state and

$$R_{00} = 1 + \text{Geo}(q) \Longrightarrow \mathbb{E}(R_{00}) = 1 + \frac{1}{q} < \infty$$

Thus, state 0 is positive recurrent. Hence, the whole DTMC is positive recurrent.
On the other hand, let $\{X(t)\}_{t\geq 0}$ be a CTMC taking this DTMC as a discrete skeleton. That is, they share the same P. Let W be the number of transitions to return to 0.

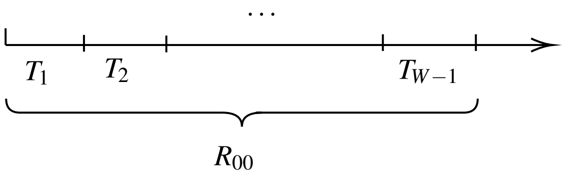


Figure 4.4.3: *What is R_{00}?*

Then,

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

where $T_i \sim \mathbf{Exp}(v_i)$ independently. Then,

$$\begin{aligned}
\mathbb{E}(R_{00}) &= \mathbb{E}(\sum_{i=0}^{W-1} T_i) \\
&= \mathbb{E}(\sum_{i=0}^{\infty} T_i \mathbf{1}_{\{W>i\}}) \\
&= \sum_{i=0}^{\infty} \mathbb{E}(T_i \mathbf{1}_{\{W>i\}}) \qquad \text{MCT} \\
&= \sum_{i=0}^{\infty} \mathbb{E}(T_i)\mathbb{E}(\mathbf{1}_{\{W>i\}}) \\
&= \sum_{i=0}^{\infty} \mathbb{E}(T_i)\mathbf{P}(W>i) \\
&= \sum_{i=0}^{\infty} \frac{1}{v_i} p^{i-1}
\end{aligned}$$

since $W = 1 + \text{Geo}(q)$. By choosing a sequence of $\{v_i\}$ which decreases to 0 fast enough, we can always make $\mathbb{E}(R_{00}) = \infty$. For example, take $\frac{1}{v_i} = \frac{1}{p^{i-1}}$, we will have

$$\mathbb{E}(R_{00}) = \sum_{i=0}^{\infty} 1 = \infty$$

Intuition

Let the sojourn time increase fast enough as the CTMC goes far from 0. We conclude that we can have different results on positive/null recurrent states between the discrete skeleton and the CTMC. However, positive/null recurrence remain 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. ■

4.5 Stationary Distribution

Definition 4.5.1 Let $\{X(t)\}_{t\geq 0}$ be a CTMC. A row vector $\pi = (\pi_0, \pi_1, \cdots)$ with $\pi_i \geq 0, \forall i \in S$ is called a stationary distribution of $\{X(t)\}_{t\geq 0}$, if

1. **Stationarity condition:** $\pi = \pi P(t), \forall t \geq 0$
2. **Normalization condition:** $\pi \mathbf{1}' = \sum_{i\in S} \pi_i = 1$

R Such a distribution is called stationary by the same reason as in the discrete case:

1. if we starts the CTMC from the initial distribution π, then the distribution of $X(t)$ will always be π for any $t > 0$. That is,

$$\mathbf{P}(X(t) = j) = \pi_j, \forall t \geq 0$$

indeed,

$$\vec{\alpha}_t = \vec{\alpha}_0 P(t) = \pi$$

How to find π?

$$\begin{aligned}
\pi =& \pi P(t), \quad \forall t \geq 0 \\
\pi I =& \pi P(t) \\
\pi(P(t) - I) =& 0 \\
\pi \frac{P(t) - I}{t} =& 0 \;\; \forall t \geq 0
\end{aligned}$$

Take $t \to 0$,

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

The conclusion is, a stationary distribution π must satisfy the equation $\pi R = 0$.
On the other hand, assume the initial distribution is π, and it satisfies $\pi R = 0$, then

$$\begin{aligned}
(\vec{\alpha}_t)' &= (\vec{\alpha}_0 P(t))' \\
&= \vec{\alpha}_0 (P(t))' \\
&= \vec{\alpha}_0 R P(t) \\
&= \pi R P(t) \\
&= 0 P(t) = 0
\end{aligned}$$

Thus, the distribution is not changing over time and π is a stationary distribution.
Thus, π is stationary if and only if

1. $\pi R = 0$
2. $\pi \mathbf{1}' = 1$

This is a much easier set of equations to solve compared to the original definition.

Relation between π and stationary distribution ψ for the discrete skeleton

For continuous time, $\pi R = 0$. For discrete case, $\psi P = \psi$. The $j-$th component of πR is

$$[\pi_0, \pi_1, \cdots] \begin{bmatrix} -v_0 & q_{01} & q_{02} & \cdots & q_{0j} & \cdots \\ q_{10} & -v_1 & q_{12} & \cdots & q_{1j} & \cdots \\ q_{20} & q_{21} & -v_3 & \cdots & q_{2j} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

$$= \pi_0 q_{0j} + \pi_1 q_{1j} + \cdots + \pi_{j-1} q_{j-1,j} - \pi_j v_j + \pi_{j+1} q_{j+1,j} + \cdots = 0$$

then,

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

For the discrete skeleton,

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

Note that $\{\psi_i\}$ satisfy as the same system of equations as $\{\pi_j v_j\}$! 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. Conversely, 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 renormalize

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

thus,

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

Conversely,

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

assuming these sums are finite. Note that π, ψ are in general not the same.

Intuition

Like in the discrete case, the stationary distribution/probability π_j is the long-run fraction of time that the CTMC stays at 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. Each time when the discrete skeleton visits state j, for the CTMC, it will stays in j for an exponential amount of time with mean $\frac{1}{v_j}$.

Thus, for CTMC, the long-run behaviour needs to by weighted by the mean sojourn time $\frac{1}{v_j}$. This is why we have

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

R

1. If the MC is irreducible, and both ψ, π exist, then the uniqueness of ψ implies the uniqueness of π.
2. It is possible that one exists while the other is not. For example, ψ exists, but π does not where

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

might not have a convergent denominator. We can choose v_i decreases to 0 fast. For example, $v_i = \psi_i$, and the state space is infinite. Then, π does not exist.

Let us rewrite the stationary conditions $\pi R = 0$. The $j-$th component is

$$\pi_j v_j = \sum_{i\in S, i\neq j}\pi_i q_{ij} = \sum_{i\in S, i\neq j}\pi_i v_i P_{ij}$$

the LHS, probability flow, is the rate leaving state j. The RHS is the total flow from other states getting to state j. Both are under the stationary distribution. Hence, $\pi R = 0$ simply says that the probability flow leaving state j should be the same as the probability flow entering state j for any j.

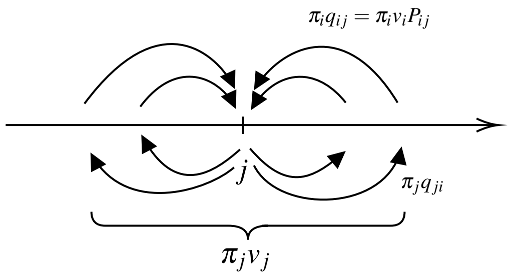


Figure 4.5.1: *Flow analogy*

Due to this interpretation, the stationary condition is also called **balanced equations**.

4.6 Limiting Behaviour

We want to consider the limiting transition probabilities

$$\lim_{t\to\infty} P_{ij}(t) = \lim_{t\to\infty} \mathbf{P}(X(t) = j|X(0) = i)$$

Theorem 4.6.1 — Basic Limit Theorem for CTMC. In a CTMC which is irreducible and recurrent,

$$\lim_{t\to\infty} P_{jj}(t) = \lim_{t\to\infty} P_{ij}(t) =: P_j = \frac{\mathbb{E}(T_j)}{\mathbb{E}(R_{jj})} = \frac{1/v_j}{\mathbb{E}(R_{jj})}, \forall i,j \in S$$

regardless the initial state. In addition, if the MC has a stationary distribution π, then it is unique, and $\pi = P = (P_0, P_1, \cdots)$ and the CTMC is positive recurrent.

Proof. Consider the process $\{X_n^h := X(nh)\}_{n=0,1,\cdots}$ for any $h > 0$. This is a discretization of the CTMC (not the discrete skeleton!) It is easy to show that $\{X_n^h\}_n$ is a DTMC with transition matrix $P(h)$. Moreover, since $\{X(t)\}_{t\geq 0}$ is irreducible and recurrent, $\{X_n^h\}$ is also irreducible and recurrent (assignment, practice question), and aperiodic. Hence, we have

$$\lim_{n\to\infty} P_{ij}(nh) =: \pi_j^h$$

which is the long-run fraction of time $\{X_n^h\}$ spends in j.
When $nh \leq t \leq (n+1)h$,

$$\begin{aligned} P_{ij}(t) &\geq P_{ij}(nh)\mathbf{P}(\text{stay in } j \text{ from } nh \text{ to time } t) \\ &= P_{ij}(nh)e^{-\lambda_j(t-nh)} \\ &> P_{ij}(nh)e^{-\lambda_j h} \end{aligned}$$

Take $t \to \infty$, we have

$$\liminf_{t\to\infty} P_{ij}(t) \geq e^{-\lambda_j h}\pi_j^h$$

take $h \to 0$, we have

$$\liminf_{t\to\infty} P_{ij}(t) \geq \limsup_{h\to 0} e^{-\lambda_j h}\pi_j^h = \limsup_{h\to 0} \pi_j^h$$

On the other hand, for $nh \leq t < (n+1)h$, we also have

$$P_{ij}((n+1)h) \geq P_{ij}(t)e^{-\lambda_j((n+1)h-t)} \geq P_{ij}(t)e^{-\lambda_j h}$$

then, $P_{ij}(t) \le e^{\lambda_j h} P_{ij}((n+1)h)$. Take $t \to \infty$,

$$\limsup_{t\to\infty} P_{ij}(t) \le e^{\lambda_j h}\pi_j^h$$

take $h \to 0$,

$$\limsup_{t\to\infty} P_{ij}(t) \le \liminf_{h\to 0} \pi_j^h$$

Combine the above results,

$$\limsup_{h\to 0} \pi_j^h \le \liminf_{t\to\infty} P_{ij}(t) \limsup_{t\to\infty} P_{ij}(t) \le \liminf_{h\to 0} \pi_j^h$$

but $\limsup_{h\to 0} \pi_j^h \ge \liminf_{h\to 0} \pi_j^h$. Thus, all the inequalities must be equalities. In particular,

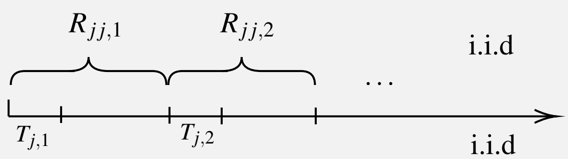
$$\lim_{t\to\infty} P_{ij}(t) =: P_j$$

exists is equal to $\lim_{h\to 0} \pi_j^h$.
Recall that π_j^h gives the long-run fraction of time that $\{X_n^h\}$ spends in state j. If we take a limit of this as $h \to 0$, we will get the long-run fraction of time that $\{X(t)\}$ spends in j, mathematically, it is

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{X(s)=j\}} ds$$

By a Law of Large Number argument, this can also be written as

$$\frac{\mathbb{E}(T_j)}{\mathbb{E}(R_{jj})} = \frac{1/v_j}{\mathbb{E}(R_{jj})}$$



Total time in j is $T_{i,1} + T_{i,2} + \cdots \approx \mathbb{E}(T_j) \times$number of cycles The total time is $R_{jj,1} + R_{jj,2} + \cdots \approx \mathbb{E}(R_{jj}) \times$numbeer of cycles. Thus, the fraction of time spent in j will converge to $\frac{\mathbb{E}(T_j)}{\mathbb{E}(R_{jj})}$.
If the CTMC has a stationary distribution π, then $\pi P(t) = \pi$ for any $t \ge 0$. In particular, $\pi P(h) = \pi$. Thus, π is the stationary distribution of $\{X_n^h\}$ (unique since the discretized DTMC is irreducible). By the basic limit theorem of DTMC, $\pi_j = \pi_j^h$ for any component $j \in S, h > 0$. Thus,

$$\pi_j = \lim_{h\to 0} \pi_j^h =: P_j, j \in S$$

Hence, $\pi = \pi$ is unique. Since $\{X_n^h\}$ has stationary distribution π, it is positive recurrent, and $\pi_j = \pi_j^h > 0$ for all $j \in S$. Then,

$$P_j = \frac{1/v_j}{\mathbb{E}(R_{jj})} > 0 \Longrightarrow \mathbb{E}(R_{jj}) < \infty$$

Thus, j is positive recurrent in the CTMC. ■

Corollary 4.6.2 Positive recurrence is a class property (for CTMC).

Proof. Consider the (recurrent) class containing a positive recurrent state as a separate irreducible recurrent CTMC. ∎

R

1. Actually, we have that the existence of stationary distribution if and only if the MC is positive is recurrent.
2. When the CTMC is null recurrent, all the terms in the formula above will be 0.

Corollary 4.6.3 A CTMC with finite state space must have at least one positive recurrent state/class.

Proof. We know that recurrence/transience is the same for the CTMC and its discretized process, and a DTMC having a finite state space must have at least one recurrent class. Thus, the CTMC also has at least one recurrent class. Focus on that class, which can be considered as a separate irreducible recurrent CTMC. For this CTMC,

$$1 = \sum_{j \in S} \lim_{t \to \infty} P_{ij}(t) = \sum_{j \in S} P_j \Longrightarrow P_j > 0$$

for some j. Thus, $\mathbb{E}(R_{jj}) < \infty$ and j is positive recurrent. ∎

Corollary 4.6.4 An irreducible CTMC with finite state space is positive recurrent.

■ **Example 4.2 — Two states system.** Consider a CTMC with

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

This is a generalization of Q4 in assignment 3. This is irreducible with a stationary distribution π that solves

$$\begin{cases} \pi R = 0 \\ \pi \mathbf{1} = 1 \end{cases}$$

The first equation gives

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

these two are linearly dependent, this is analogous behaviour as in DTMC. We get the ratio

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

We need to use the second equation $\pi_0 + \pi_1 = 1$ and we get

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

We can see that π exists and is unique.

Limiting distribution

It is easy to check (by verifying the backward equaion)

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

Take $t \to \infty$, 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 transition 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

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

thus,

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

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

4.7 Birth and Death Processes (Continued)

4.7.1 Classification of Birth and Death Processes

Recall the generator of birth and death processes is

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

We can assume this CTMC being irreducible, namely, $\lambda_i, \mu_i > 0, \forall i$. The balance equations are:

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

$$(0) \implies \pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$$

$$(0)+(1) \implies \pi_1\lambda_1 = \pi_2\mu_2 \implies \pi_2 = \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_0\lambda_1}{\mu_2\mu_1}\pi_0$$

$$\vdots$$

$$(0)+\cdots+(n-1) \implies \pi_{n-1}\lambda_{n-1} = \pi_n\mu_n \implies \pi_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\pi_0$$

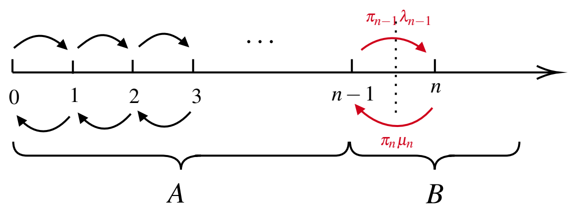


Figure 4.7.1: *Graphical interpretation*

As a result, we have all the proportions of π_i. By normalization condition,

$$\sum_{n=0}^{\infty} \pi_n = 1 \Longrightarrow \pi\left(1+\frac{\lambda_0}{\mu_1}+\cdots+\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}+\cdots\right)=1$$

$$\pi_0 = \frac{1}{1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}}$$

$$\pi_n = \frac{\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}}{1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}}$$

Thus, a stationary distribution exists if and only if

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

In this case, π is also the limiting distribution.
To recapitulate, a necessary and sufficient condition for an (irreducible) birth and death process to be positive recurrent is

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

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 $f_{n0} := \mathbf{P}(\text{MC ever (re)visits state } 0|X(0)=n)$. Then, $f_{00}=1$ if and only if the whole chain is recurrent. Otherwise, transient. By first-step analysis,

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

In general,

$$f_{n0} = \frac{\mu_n}{\lambda_n+\mu_n}f_{n-1,0}+\frac{\lambda_n}{\lambda_n+\mu_n}f_{n+1,0}$$

$$(\lambda_n+\mu_n)f_{n0} = \mu_n f_{n-1,0}+\lambda_n f_{n+1,0}$$

$$f_{n+1,0}-f_{n0} = \frac{\mu_n}{\lambda_n}(f_{n0}-f_{n-1,0})$$

Thus,

$$f_{n+1,0}-f_{n0} = \frac{\mu_n}{\lambda_n}(f_{n0}-f_{n-1,0}) = \frac{\mu_n\cdots\mu_1}{\lambda_n\cdots\lambda_1}(f_{10}-1)$$

Then,

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

If the MC is transient, $f_0 0 = f_{10} < 1$ 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 still need the other direction. For $n = 1, 2, \cdots$, define

$$f_{i0}^{(n+1)} = \mathbf{P}(\text{the MC visits 0 before } n+1 | X(0) = i)$$

Then, by the 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,

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

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

Take $n \to \infty$,

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

If the MC is recurrent, then this probability is 0. Then,

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

To conclude, a birth and death process is transient if and only if

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

5. Continuous Phase-Type Distribution

5.1 Basic setup

Let $\{X(t)\}_{t\geq 0}$ be a CTMC having m transient states and one absorbing state. Let

$$E = \{1, \cdots, m\}$$

be the transient states and 0 as the absorbing state. Consider the generator

$$R = \begin{array}{c|c:cccc} & 0 & 1 & 2 & \cdots & m \\ \hline 0 & 0 & 0 & 0 & \cdots & 0 \\ \hdashline 1 & & & & & \\ 2 & \vec{t}_0 & & & T & \\ \vdots & & & & & \\ m & & & & & \end{array}$$

where

1. the first row of 0 is due to the absorbing state
2. $\vec{t}_0$ is a $m \times 1$ column vector that accounts for the transitions from transient states to the absorbing state 0
3. T is a $m \times m$ transient part of the generator

Because of row sums of R are 0, we have

$$\vec{t}_0 + T\mathbf{1} = 0 \Longrightarrow \vec{t}_0 = -T\mathbf{1}$$

Define the initial distribution (of the transient part) to be

$$\vec{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)$$

to be the $1 \times m$ row vector. That is

$$\alpha_i = \mathbf{P}(X(0) = i),\ i = 1, \cdots, m$$

and

$$\alpha_0 := \mathbf{P}(X(0) = 0) = 1 - \sum_{i=1}^{m} \alpha_i = 1 - \vec{\alpha}\mathbf{1}$$

Definition 5.1.1 — CPH. Define Y to be the time until obsorption.

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

We say that such Y has a continuous phase-type distribution with representation $\vec{\alpha}$ and T, denoted as

$$Y \sim \mathbf{CPH}_m(\vec{\alpha}, T)$$

R This is analogous to **DPH**. We sometimes just write PH for CPH.

5.1.1 CDF of Y

For $y \geq 0$, consider

$$\begin{aligned}
\mathbf{P}(Y > y) = \mathbf{P}(X(y) \in E) &= \sum_{i=1}^{m} \mathbf{P}(X(y) \in E | X(0) = i)\mathbf{P}(X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \mathbf{P}(X(y) \in E | X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} \mathbf{P}(X(y) = j | X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} P_{ij}(y) \quad (*)
\end{aligned}$$

Recall that the transition matrix at time y is given by

$$\begin{aligned}
P(y) &= e^{yR} \\
&= \sum_{n=0}^{\infty} \frac{y^n}{n!} R^n = I + \sum_{n=1}^{\infty} \frac{y^n}{n!} R^n \\
&= \sum_{n=1}^{\infty} \frac{y^n}{n!} \begin{bmatrix} 0 & 0 \\ ? & T^n \end{bmatrix} + I \\
&= \begin{bmatrix} 1 & 0 \\ ? & \sum_{n=1}^{\infty} \frac{y^n}{n!} T^n + I \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ ? & e^{yT} \end{bmatrix}
\end{aligned}$$

why we don't need to know this "?"? (HAHA) Since we can compute it through

$$? = \mathbf{1} - e^{yT}\mathbf{1}$$

Thus,

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

Back to $(*)$, we have

$$\begin{aligned}\mathbf{P}(Y > y) &= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} P_{ij}(y) \\ &= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} (e^{yT})_{ij} \\ &= \vec{\alpha} e^{yT} \mathbf{1}\end{aligned}$$

Then, the CDF of Y is

$$F_Y(y) = 1 - \vec{\alpha} e^{yT} \mathbf{1}$$

Note that

$$\mathbf{P}(Y = 0) = F_Y(0) = 1 - \vec{\alpha} I \mathbf{1} = \alpha_0$$

which makes intuitive sense.

> **R** Note that when $\alpha_0 > 0$, there is a point mass at $y = 0$. This means we don't have a pdf for all $y \geq 0$ but rather a mixed type.

5.1.2 PDF of CPH

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 $\alpha_0 = 1 - \vec{\alpha}\mathbf{1}$, and a density for $y > 0$, given by

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

■ **Example 5.1 — Simple example of CPH.** Consider the exponential density

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

The CPH representation is given by

$$R = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0 & 0 \\ \lambda & -\lambda \end{pmatrix} \end{array}, \quad \vec{\alpha} = 1, \quad m = 1$$

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

5.2 Properties of CPH

1. Let $X \sim \mathbf{CPH}_m(\vec{\alpha}, T)$ and $X \sim \mathbf{CPH}_n(\vec{\beta}, S)$ where $X \perp\!\!\!\perp Y$. Using the same reasoning as for DPH, we have

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

   where

   $$\vec{\delta} = (\vec{\alpha}, \alpha_0 \vec{\beta})$$

and

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

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

R As a result, some common distributions are memebers of CPH family.

Example 5.2 — Gamma distribution. Consider a Gamma pdf

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

also known as Erlang distribution with parameters n, λ. If $X \sim \text{Erlang}(n, \lambda)$, then $X \xrightarrow{d} \sum_{i=1}^{n} Y_i$, where $Y_i \overset{\text{i.i.d}}{\sim} \mathbf{Exp}(\lambda)$. (The proof requires characteristic function/MGF or convolution) Thus, $X \sim \mathbf{CPH}_n(\vec{\alpha}, T)$ where

$$\vec{\alpha} = (1, 0, \cdots, 0)$$

with

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

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

2. **Mixture:** let $X \sim \mathbf{CPH}_m(\vec{\alpha}, T)$ and $Y \sim \mathbf{CPH}_n(\vec{\beta}, S)$. Define

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

the choice is independent of X and Y. That is

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

As in the discrete case,

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

where

$$\vec{\gamma} = (p\vec{\alpha}, (1-p)\beta)$$

and

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

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

3. All the moments of a CPH are finite.

Proof. To see this, look at the Laplace transform of the random variable X which follows a $\mathbf{CPH}_m(\vec{\alpha}, T)$. Then,

$$\tilde{f}(s) = \mathbb{E}(e^{-sX})$$

always well-defined for positive random variable X.

$$\begin{aligned}
\frac{d^n}{ds^n}\tilde{f}(s)\bigg|_{s=0} &= \mathbb{E}\left(\frac{d^n}{ds^n}e^{-sX}\right)\bigg|_{s=0} \qquad \text{DCT} \\
&= \mathbb{E}\left((-X)^n e^{-sX}\bigg|_{s=0}\right) \\
&= (-1)^n \mathbb{E}(X^n) \\
\implies \mathbb{E}(X^n) &= (-1)^n \frac{d^n}{ds^n}\tilde{f}(s)\bigg|_{s=0}
\end{aligned}$$

For $X \sim \mathbf{CPH}_m(\vec{\alpha}, T)$,

$$\begin{aligned}
\tilde{f}(s) = \mathbb{E}(e^{-sX}) &= \alpha_0 e^{-s\times 0} + \int_0^\infty e^{-sx}\vec{\alpha}e^{Tx}\vec{t_0'}dx \\
&= \alpha_0 + \vec{\alpha}\left(\int_0^\infty e^{-sx}e^{Tx}dx\right)\vec{t_0'} \\
&= \alpha_0 + \vec{\alpha}\left(\int_0^\infty e^{-sx}Ie^{Tx}dx\right)\vec{t_0'} \\
&= \alpha_0 + \vec{\alpha}\left(\int_0^\infty e^{-sxI}e^{Tx}dx\right)\vec{t_0'} \\
&= \alpha_0 + \vec{\alpha}\left(\int_0^\infty e^{(T-sI)x}dx\right)\vec{t_0'} \\
&= \alpha_0 + \vec{\alpha}\left(e^{(T-sI)x}(T-sI)^{-1}\right)\bigg|_0^\infty \vec{t_0'} \qquad \text{can show existence of } (T-sI)^{-1}, s \geq 0 \\
&= \alpha_0 + \vec{\alpha}\left(\lim_{x\to\infty} e^{(T-sI)x} - I\right)(T-sI)^{-1}\vec{t_0'}
\end{aligned}$$

Note that $T - sI$ is a possible transient part of a generator matrix. That is, there exists a generator

$$R' = \begin{bmatrix} 0 & 0 \\ ? & T-sI \end{bmatrix}$$

Then,

$$\left(e^{(T-sI)x}\right)_{ij} = P'_{ij}(x)$$

where $P'(t)$ is the transition matrix of the CTMC corresponding to R'. Since j is transient, $P_{ij}(x) \to 0$ as $x \to \infty$. Thus,

$$\lim_{x\to\infty} e^{(T-sI)x} = \mathbf{0}$$

Thus,

$$\tilde{f}(s) = \alpha_0 + \vec{\alpha}(sI-T)^{-1}\vec{t_0'},\ s \geq 0$$

Then,

$$\begin{aligned}
\mathbb{E}(X) &= -\tilde{f}'(s)\bigg|_{s=0} \\
&= -(-1)\vec{\alpha}(sT-I)^{-2}\vec{t_0'}\bigg|_{s=0} \\
&= \vec{\alpha}(-T)^{-2}\vec{t_0'}
\end{aligned}$$

Recall that $T\mathbf{1} = -t_0'$. Then,

$$\mathbb{E}(X) = \vec{\alpha}(-T)^{-1}\mathbf{1}$$

In general,

$$\mathbb{E}(X^n) = n!\vec{\alpha}(-T)^{-n}\mathbf{1},\ n = 1, 2, \cdots$$

$\blacksquare$

III

Queuing Theory

6. Queuing Theory

Consider you are running a shop and there are multiple/single queue(s) of customers arriving at different times and you have server(s) to provide services at different times.

6.1 Basic Setup

6.1.1 Terminology

Definition 6.1.1 — A/S/m/c/p Queue.

1. A: arrival process. We often assume the inter-arrival times are i.i.d. This means we have a renewal process. Here are some possible acronyms:
   (a) M: (memoryless/Markov) exponential inter-arrival times. This would imply a Poisson arrival process.
   (b) G: general inter-arrival times
   (c) D: deterministic/constant inter-arrival times
   (d) E_k: Erlang-k inter-arrival times
   (e) PH: continuous phase-type inter-arrival times
2. S: service process. Same set of acronyms as in A.
3. m: number of servers
4. c: number of systems places; capacity: the sum of waiting places and the service places. This is often omitted when $c = \infty$.
5. p: customer population. This often omitted when $p = \infty$.

■ **Example 6.1**

1. $M/M/1$ queue: we have one queue and the customer arrives follow a Poisson process. There is one server that has exponential service time. Unlimited waiting places and infinite customer population.

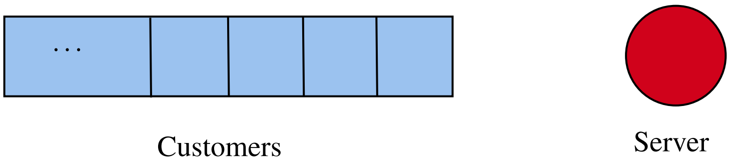


Figure 6.1.1: *$M/M/1$ queue*

2. $M/M/m/m$ queue: we have m queues and the customer arrives follow a Poisson process. There are m server that has exponential service time. There are M system places and unlimited customer population.

   R This could be a model for a call centre without waiting option. (Always more than 20-minute...)

■

Definition 6.1.2 — Queue length. This typically refers to the total number of customers, including waiting customers and those being served.

Definition 6.1.3 — Service discipline.
1. FCFS/FIFO: first-come-first-serve or first-in-first-out
2. LCFS/LIFO: last-come-first-serve or last-in-first-out (stack)
   (a) **Preemptive resume:** arriving customer will be immediately served. Interrupted service resumes when the new service is done.
   (b) **Preemptive restart:** arriving customer will be immediately served. Interrupted service will restart when the new service is done.
   (c) **Non-preemptive:** arriving customer waits until the ongoing services is finished.
3. Scheduling of servers: services in rotating order, processor sharing
4. SIRO: service-in-random-order
5. SJF: shortest job first

6.1.2 Queuing System and Markov Property

Do queuing systems always have the Markov property?
In general, the answer is **NO**. The Markov property means memoryless in time. This would require exponential distribution for both arrival and service times. Thus, only $M/M/s$ queues are directly CTMC.
However, there are cases where we can “turn" them into CTMC or DTMC.
1. $G/M/s$ queue: observed at each arrival is a DTMC since, for each arrival, we do not need to keep track on the waiting timme for the next arrival.
2. $M/G/s$ queue: observed whenever a service is done is a DTMC.

Although, Erlang-k and PH are note 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.

6.2 $M/M/1$ Queue

This is a CTMC. In particular, a birth and death process, with $\lambda_i = \lambda$ (constant arrival rate) and $\mu_i = \mu$ (constant service rate) as there is only 1 server.
Recall that for a birth and death process, a stationary distribution exists if and only if

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

In this case,

$$\pi_0 = \frac{1}{1+\sum_{n=1}^{\infty}\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}}, \quad \pi_i = \frac{\lambda_0\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i}\pi_0 = \frac{\frac{\lambda_0\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i}}{1+\sum_{n=1}^{\infty}\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}}$$

since $\lambda_i = \lambda, \mu_i = \mu$, then

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \pi_0, \; \pi_0 = \frac{1}{1+\sum_{n=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^n} = \frac{1}{\frac{1}{1-\frac{\lambda}{\mu}}} = 1-\frac{\lambda}{\mu}$$

and

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

when $\lambda < \mu$. Define the traffic intensity for an $M/M/1$ queue is given by $\rho := \frac{\lambda}{\mu}$. Thus, $\pi_i = (1-\rho)\rho^i$. A stationary distribution exists if and only if $\rho < 1 \iff \lambda < \mu$.
Let's derive some interesting quantities.

$$\mathbf{P}(\text{server is busy}) = 1-\pi_0 = 1-(1-\rho) = \rho$$

The **mean number of customers** in the system is given by

$$0\pi_0 + 1\pi_1 + \cdots = \mathbb{E}(X) = \frac{\rho}{1-\rho}$$

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

6.3 Detailed Balanced Condition

Definition 6.3.1 — Detailed balanced. We say a distribution π satisfies the detailed balanced condition for a CTMC with generator R if

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

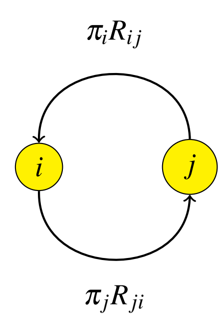


Figure 6.3.1: *Pairwise equivalent probability flow*

Theorem 6.3.1 If a distribution π satisfies the detailed balanced condition, then it is a stationary distribution.

Proof. Note that

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

Thus, $\pi R = 0$. Moreover, π is a distribution. Thus, π is a stationary distribution. ■

In general, a stationary distribution does not need to satisfy the detailed balanced condition. Thus, the converse of the above theorem is not always valid. However, just as in the discrete-time case, if the CTMC is a birth and death process (with a tri-diagonal R), then the converse is true.

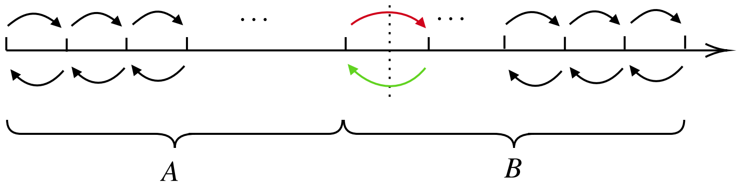


Figure 6.3.2: *Birth and death process makes the converse work*

6.4 $M/M/c$ Queue

This is another birth and death process with $\lambda_i = \lambda$ and

$$\mu_i = \begin{cases} i\mu & i \leq c-1 \\ c\mu & i \geq c \end{cases}$$

For $i = 1, 2, \cdots, c-1$,

$$\pi_i = \frac{\lambda_0 \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i} = \frac{\lambda^i}{i!\mu^i}\pi_0$$

For $i = c, c+1, \cdots,$

$$\pi_i = \frac{\lambda_0 \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i} = \frac{\lambda^i}{c!c^{i-c}\mu^i}\pi_0$$

then,

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

Note that

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

Define $\rho = \frac{\lambda}{c\mu}$ to be the traffic intensity of an $M/M/c$ queue, then a stationary distribution exists if and only if $\rho < 1$. 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

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

6.4.1 $M/M/\infty$ Queue

We now have infinite servers! This seems bizarre in first glance. But there can be some applications, for examples,

1. People who got a flu
2. People asking for tourism guide

Again, we are dealing with a birth and death process with $\lambda_i = \lambda$ and $\mu_i = i\mu$ for all $i \ge 0$. Then,

$$\pi_0 = \frac{1}{1+\sum_{i=1}^{\infty}\frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}} = \frac{1}{1+\sum_{i=1}^{\infty}\frac{\lambda^i}{i!\mu^i}}$$

note that

$$1+\sum_{i=1}^{\infty}\frac{\lambda^i}{i!\mu^i} = e^{\frac{\lambda}{\mu}}$$

Therefore,

$$\pi_0 = e^{-\frac{\lambda}{\mu}} < \infty,\ \pi_i = \frac{(\lambda/\mu)^i e^{-(\lambda/\mu)}}{i!}$$

This is exactly a $\mathbf{Poi}\left(\frac{\lambda}{\mu}\right)$! Moreover, this $M/M/\infty$ queue always has a stationary distribution.

6.4.2 $M/M/1/c$ Queue

Now we introduce a capacity. This is still a birth and death process with

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

Then,

$$\pi_i = \frac{\lambda_i\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i}\pi_0 = \left(\frac{\lambda}{\mu}\right)^i\pi_0 \quad ,i=1,2,\cdots,c$$

$\pi_0 = 0$ for $i \ge c+1$. Then,

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

and

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

6.4.3 $M/M/1/\infty/c$ Queue

Now, we set a limit on population but no waiting capacity.

■ **Example 6.2** Consider c machines in the system. Each breaks according to $\mathbf{Exp}(\lambda)$. A single server repairs the broken machines with service time $\mathbf{Exp}(\mu)$.

$$\lambda_i = \begin{cases} (c-i)\lambda & i \leq c-1 \\ 0 & i \geq c \end{cases}, \quad \mu_i = \begin{cases} \mu & i \leq c \\ 0 & i \geq c+1 \end{cases}$$

Then,

$$\pi_i = \frac{c!\lambda^i}{(c-i)!\mu^i}\pi_0, \; i = 1, \cdots, c$$

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

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

■

6.5 Generating Function Method

Definition 6.5.1 — Probability generating function. The (probability) generating function of a distribution π on $\{0, 1, \cdots\}$ is defined as

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

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

Proposition 6.5.1 — Properties of generating functions.

1. The distribution π is recovered by taking derivatives of g at 0,

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

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

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

■

2. $g(1) = \sum_{n=0}^{\infty} \pi(n) = 1$ and the **k-th** factorial moment

$$\mathbb{E}\left(X(X-1)\cdots(X-k+1)\right) = g^{(k)}(1)$$

Proof. By DCT,

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

■

3. Let X, Y be independent random variables taking values in the non-negative integers with g.f.s g_X, g_Y then,

$$g_{X+Y} = g_X g_Y$$

Proof. By independence of functions of independent random variables,

$$g_{X+Y}(z) = \mathbb{E}\left(z^{X+Y}\right) = \mathbb{E}(z^X)\mathbb{E}(z^Y)g_X(z)g_Y(z)$$

■

6.5.1 Erlang Models

Recall the pdf of Erlang-r

$$f(x) = \frac{\alpha^r}{(r-1)!}x^{r-1}e^{-\alpha x},\ x > 0$$

Erlang-r is the sum of r i.i.d $\mathbf{Exp}(\alpha)$ random variables. It has mean $\frac{r}{\alpha}$ and variance $\frac{r}{\alpha^2}$.

6.5.2 $M/E_r/1$ Queue

The service time follows an Erlang-r with mean $\frac{1}{\mu}$ (service rate, this means the mean for each exponential is $\frac{1}{\mu r}$). We can define a CTMC out of this. Let's consider r "stages" for each service job, we can recover the Markov property as in the exponential case. Consider the state, total number of stages needed to be completed (for all customers) in the system. If there are k customers, and the customer being served already completed i stages, then the state $j = (k-1)r + (r-i) = kr - i$. Conversely,

$$k = \left\lceil \frac{j}{r} \right\rceil$$

and the number of stages completed for the current customer is

$$\begin{cases} r - (j \mod r) & r \nmid j \\ 0 & r \mid j \end{cases}$$

Let $\pi(i)$ be the limiting/stationary probability that there are i customers and let $P(j)$ be the limiting/stationary probability that the system is in state j. Then,

$$\pi(i) = P(ir) + P(ir-1) + \cdots + P(r(i-1)+1) = \sum_{j=r(i-1)+1}^{ir} P(j)$$

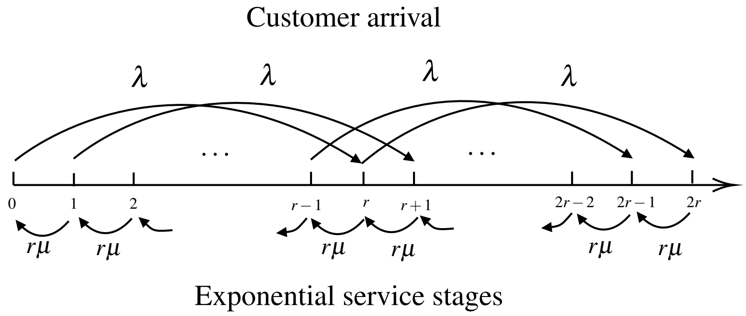


Figure 6.5.1: *Visualization of the change of states*

One thing we can see from this is that $M/E_r/1$ queue is no longer a birth and death process as states can jump to non-adjacent states. We can no longer use the detailed balanced conditions to find a stationary distribution but resort to balanced equations.

Balanced equations

$$\lambda P(0) = r\mu P(1)$$
$$(\lambda + r\mu)P(j) = \lambda P(j-r) + r\mu P(j+1),\ j = 1,2,3,\cdots \quad (*)$$

if we define $P(-1) = P(-2) = \cdots = 0$. Multiply $(*)$ by z^j and sum over j from 1 to ∞:

$$\begin{aligned}
(\lambda + r\mu)\sum_{j=1}^{\infty} P(j)z^j &= \lambda \sum_{j=1}^{\infty} P(j-r)z^j + r\mu \sum_{j=1}^{\infty} P(j-1)z^j \\
(\lambda + r\mu)(g(z) - P(0)) &= \lambda z^r \sum_{j=1}^{\infty} P(j-r)z^{j-r} + \frac{r\mu}{z}\sum_{j=1}^{\infty} P(j+1)z^{j+1} \\
(\lambda + r\mu)(g(z) - P(0)) &= \lambda z^r g(z) + \frac{r\mu}{z}(g(z) - P(0) - P(1)z) \\
(\lambda + r\mu)z(g(z) - P(0)) &= \lambda z^{r+1} g(z) + r\mu(g(z) - P(0) - \frac{\lambda}{r\mu}P(0)z)
\end{aligned}$$

Thus,

$$g(z) = \frac{r\mu(z-1)P(0)}{\lambda z + r\mu z - \lambda z^{r+1} - r\mu} = \frac{(\lambda + r\mu)(g(z) - P(0))}{(z-1)(r\mu - \lambda z^r - \lambda z^{r-1} - \cdots - \lambda z} = \frac{r\mu P(0)}{r\mu - \lambda z^r - \cdots - \lambda z}$$

Recall that

$$1 = g(1) = \frac{r\mu P(0)}{r\mu - r\lambda} = \frac{\mu}{\mu - \lambda}P(0) \Longrightarrow P(0) = 1 - \frac{\lambda}{\mu}$$

Thus,

$$g(z) = \frac{r(\mu - \lambda)}{-\lambda(z^r + \cdots + z) + r\mu}$$

Note that when $r = 1$, we have

$$g(z) = \frac{\mu - \lambda}{\mu - \lambda z} = \frac{1-\rho}{1-\rho z} = (1-\rho)\sum_{i=0}^{\infty} \rho^i z^i = \sum_{i=0}^{\infty} P(i)z^i$$

then, $P(i) = \pi(i) = (1-\rho)\rho^i, i = 0,1,\cdots$ which agrees with what we got for $M/M/1$ queue.

■ **Example 6.3** Consider an $M/E_3/1$ queue with $\lambda = 4$ and $\mu = 6$. In this case

$$\begin{aligned} g(z) &= \frac{3(6-4)}{-4(z^3+z^2+z)+3\times 6} = \frac{3}{9-2z-2z^2-2z^3} = \frac{1}{3}\frac{1}{1-\frac{2}{9}(z+z^2+z^3)} \\ &= \frac{1}{3}\left(1+\frac{2}{9}(z+z^2+z^3)+\left(\frac{2}{9}(z+z^2+z^3)\right)^2+\cdots\right) \\ &= \underbrace{\frac{1}{3}}_{P(0)} + \underbrace{\frac{2}{27}}_{P(1)}z + \underbrace{\frac{22}{3^5}}_{P(2)}z^2 + \underbrace{\frac{242}{3^7}}_{P(3)}z^3 + \underbrace{\frac{1204}{3^9}}_{P(4)}z^4 + \underbrace{\frac{10328}{3^{11}}}_{P(5)}z^5 + \underbrace{\frac{81532}{3^{13}}}_{P(6)}z^6 + \cdots \end{aligned}$$

Thus,

$$\begin{aligned} \pi(0) &= P(0) = \frac{1}{3} \\ \pi(1) &= P(1)+P(2)+P(3) = \frac{602}{2187} \approx 0.275 \\ \pi(2) &= P(4)+P(5)+P(6) = \frac{272008}{1594323} \approx 0.171 \\ &\vdots \end{aligned}$$

Note that $P(0) = 1-\frac{\lambda}{\mu} > 0$ if and only if $\lambda < \mu$. This agrees with the queuing systems that we have seen previously and intuitively makes sense. ■

6.5.3 Bulk Arrivals

Now, we consider interarrival times that follows $\mathbf{Exp}(\lambda)$ but customers arrive in groups of size r. The service time follows $\mathbf{Exp}(\mu)$. We model this as a CTMC with the following transition diagram.

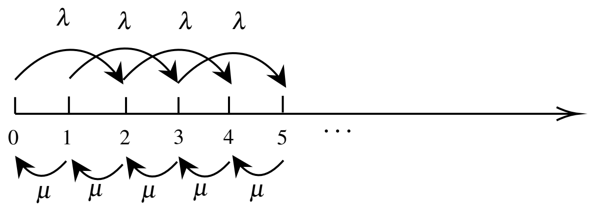


Figure 6.5.2: *Bulk arrival transition diagram*

This is the same diagram as an $M/E_r/1$ queue with $r\mu$ replaced by μ. Thus,

$$g(z) = \frac{\mu - r\lambda}{-\lambda(z^r+z^{r-1}+\cdots+z)+\mu}$$

What if the bulk size is random?
Assume bulk size are i.i.d random variable following distribution

$$\vec{a} = (a(0), a(1), \cdots)$$

where $a(0) = 0$. Let $h(z) = \sum_{k=1}^{\infty} a(k)z^k$ be the gf of $\vec{a}$. We can model this with the following transition diagram.

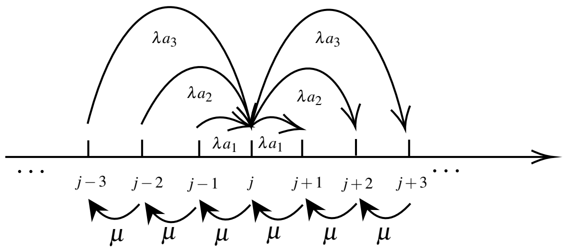


Figure 6.5.3: *Random bulk arrival size*

Now, the balance equations look like:

$$\lambda\pi(0) = \mu\pi(1) \tag{1}$$

$$(\lambda+\mu)\pi(j) = \lambda\sum_{k=1}^{j} a(k)\pi(j-k) + \mu\pi(j+1) \tag{2}$$

Multiply (2) by z^j and sum j from 1 to ∞,

$$\begin{aligned}
(\lambda+\mu)\sum_{j=1}^{\infty}\pi(j)z^j &= \lambda\sum_{j=1}^{\infty}\sum_{k=1}^{j} a(k)\pi(j-k)z^j + \mu\sum_{j=1}^{\infty}\pi(j+1)z^j \\
(\lambda+\mu)(g(z)-\pi(0)) &= \frac{\mu}{z}(g(z)-\pi(0)-\pi(1)z) + \lambda\sum_{k=1}^{\infty}\sum_{j=k}^{\infty} a(k)z^k\pi(j-k)z^{j-k} \\
&= \frac{\mu}{z}(g(z)-\pi(0)-\pi(1)z) + \lambda\sum_{k=1}^{\infty} a(k)z^k\sum_{l=0}^{\infty}\pi(l)z^l \\
&= \frac{\mu}{z}(g(z)-\pi(0)-\pi(1)z) + \lambda h(z)g(z)
\end{aligned}$$

Thus,

$$g(z) = \frac{\mu\pi(0)(z-1)}{(\lambda+\mu)z-\mu-\lambda zh(z)}$$

Then, to see the normalizing condition of a stationary distribution,

$$1 = \lim_{z\uparrow 1} g(z) \overset{\text{L'Hopital}}{=} \left.\frac{\mu\pi(0)}{(\lambda+\mu)-\lambda(h(z)+zh'(z))}\right|_{z=1} = \frac{\mu\pi(0)}{\mu-\lambda h'(1)}$$

This means

$$\pi(0) = \frac{\mu-\lambda h'(1)}{\mu} = 1-\underbrace{\frac{\lambda h'(1)}{\mu}}_{=:\rho}$$

Note that

$$h'(1) = \mathbb{E}(\text{Bulk size}) \Longrightarrow \lambda h'(1) = \mathbb{E}(\text{\# of customers arriving per unit of time})$$

This justifies the definition of ρ as the traffic intensity. Now,

$$g(z) = \frac{\mu(1-\rho)(z-1)}{(\lambda+\mu)z-\mu-\lambda zh(z)}$$

R The gf method is developed to deal with queuing systems with infinite state space. When the capacity (waiting room) is finite, we have a finite dimensional system of equations and we can directly solve it to get the stationary distribution.

In general, we can introduce $E_k/E_r/s$ queues. We need to introduce stages for both the arrival process and the service process. A state becomes a triplet

$$(\text{\# of customers}, \text{stage of arrival}, \text{stage of service})$$

Same principle, just more complicated.

6.6 Quasi-Birth-and-Death Process (QBD)

Definition 6.6.1 — QBD. A QBD is a 2-dimensional CTMC with generator having a tridiagonal block structure.

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} B_{0,0} & B_{0,1} & & & \\ B_{1,0} & B_{1,1} & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & A_0 \\ & & \ddots & \ddots & \ddots \end{pmatrix} \end{array}$$

where entries are block matrices and these labeling numbers are **levels**. The intra-block structure is consisted of phases. We define the following notation

$$A_0/A_1/A_2(i,l)$$

where i is the current phase and l is the target phase.

A simple result with such a generator Q would be levels $1,2,3,\cdots$ have the same number of phases. To make the notation clear, for example,

$$A_0(i,l) = Q((1,i),(2,l)) = Q((2,i),(3,l)) = \cdots$$

Stability of QBDs

If we ignore the level but only look at the phase, then we get a new CTMC with generator

$$A = A_0 + A_1 + A_2$$

and stationary distribution $\nu = \{\nu_i\}_i$. If the QBD admits a stationary distribution $\pi = \{\pi(k,i)\}$, then

$$\nu_i = \sum_k \pi(k,i)$$

since the transition to the next level only depends on the current phase, the probability flow of going to the next higher level is

$$\sum_i \nu_i \sum_l A_0(i,l) = \nu A_0 \mathbf{1}'$$

in particular $\nu_i A_0(i,l)$ means the probability flow from phase i in the current level to phase l in the next level.

Similarly, the probability flow going to the next lower level is

$$\sum_i \nu_i \sum_l A_2(i,l) = \nu A_2 \mathbf{1}'$$

In order to have the system to be **stable**, the necessary and sufficient condition is

$$\nu A_0 \mathbf{1}' < \nu A_2 \mathbf{1}'$$

Intuitively, the probability flow going up needs to be less than the probability flow going down.

■ **Example 6.4 — $E_2/E_3/1$ queue.** The level is the queue length, the number of customers in the system. There are 2 phases for arrival and 3 for service. Thus, in total, we have 6 combinations of phases $(i,j), i = 1,2, j = 1,2,3$. Then,

$$A_0 = \begin{array}{c} \\ (1,1) \\ (1,2) \\ (1,3) \\ (2,1) \\ (2,2) \\ (2,3) \end{array} \begin{array}{c} \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (2,1) & (2,2) & (2,3) \end{array} \\ \begin{pmatrix} & & & & & \\ & & & & & \\ & & & & & \\ 2\lambda & & & & & \\ & 2\lambda & & & & \\ & & 2\lambda & & & \end{pmatrix} \end{array}$$

The total arrival rate is

$$\lambda = 1/\mathbb{E}(\text{interarrival time}) = 1/2\mathbb{E}(\text{Exponential time for one phase})$$

Thus,

$$1/\mathbb{E}(\text{Exponential time for one phase}) = 2\lambda$$

If we only look at the arrival phase, we have the following picture.

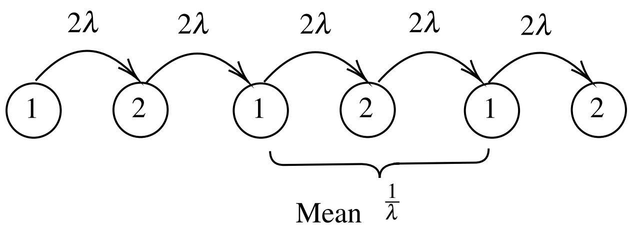


Figure 6.6.1: *Arrival phase changes as level increases*

Next,

$$A_2 = \begin{array}{c} \\ (1,1) \\ (1,2) \\ (1,3) \\ (2,1) \\ (2,2) \\ (2,3) \end{array} \begin{array}{c} \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (2,1) & (2,2) & (2,3) \end{array} \\ \begin{pmatrix} & & & & & \\ & & & & & \\ 3\mu & & & & & \\ & & & & & \\ & & & & & \\ & & & 3\mu & & \end{pmatrix} \end{array}$$

$$A_1 = \begin{array}{c} \\ (1,1) \\ (1,2) \\ (1,3) \\ (2,1) \\ (2,2) \\ (2,3) \end{array} \begin{array}{c} \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (2,1) & (2,2) & (2,3) \end{array} \\ \begin{pmatrix} -2\lambda-3\mu & 3\mu & & 2\lambda & & \\ & -2\lambda-3\mu & 3\mu & & 2\lambda & \\ & & -2\lambda-3\mu & & & 2\lambda \\ & & & -2\lambda-3\mu & 3\mu & \\ & & & & -2\lambda-3\mu & 3\mu \\ & & & & & -2\lambda-3\mu \end{pmatrix} \end{array}$$

Then,

$$A = A_0 + A_1 + A_2 = \begin{array}{cc} & \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (2,1) & (2,2) & (2,3) \end{array} \\ \begin{array}{c} (1,1) \\ (1,2) \\ (1,3) \\ (2,1) \\ (2,2) \\ (2,3) \end{array} & \begin{pmatrix} -2\lambda-3\mu & 3\mu & & 2\lambda & & \\ & -2\lambda-3\mu & 3\mu & & 2\lambda & \\ 3\mu & & -2\lambda-3\mu & & & 2\lambda \\ 2\lambda & & & -2\lambda-3\mu & 3\mu & \\ & 2\lambda & & & -2\lambda-3\mu & 3\mu \\ & & 2\lambda & 3\mu & & -2\lambda-3\mu \end{pmatrix} \end{array}$$

■

R Observe that the column sums of A are all 0 as well as its row sums. This is a doubly stochastic generator. Hence, the stationary distribution ν is uniform.

$$\nu = (\frac{1}{6}, \cdots, \frac{1}{6})$$

(Why?) Then,

$$\nu A_0 \mathbf{1}' = \lambda, \quad \nu A_2 \mathbf{1} = \mu$$

Thus, the stability condition is $\lambda < \mu$, which makes a lot of sense intuitively.

6.6.1 Limiting Probabilities of QBD

Let $\pi = (\pi_0, \pi_1, \cdots)$ be the limiting/stationary distribution of the QBD where π_i is the distribution for level i, which is still a row vector. Thus, π solves the system

$$\begin{cases} \pi Q = 0 \\ \pi \mathbf{1} = 1 \end{cases}$$

Then,

$$0 = \pi Q = (\pi_0, \pi_0, \cdots) \begin{array}{cc} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} & \begin{pmatrix} B_{0,0} & B_{0,1} & & & \\ B_{1,0} & B_{1,1} & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & A_0 \\ & & \ddots & \ddots & \ddots \end{pmatrix} \end{array}$$

$$\begin{aligned} \Longrightarrow & \pi_0 B_{00} + \pi_1 B_{10} = \mathbf{0} \\ & \pi_0 B_{01} + \pi_1 B_{11} + \pi_2 A_2 = \mathbf{0} \\ & \pi_j A_0 + \pi_{j+1} A_1 + \pi_{j+2} A_2 = \mathbf{0} \quad j = 1, 2, \cdots \quad (*) \end{aligned}$$

If A_0, A_1, A_2 were numbers, the general solution of $(*)$ would be a geometric series. So here, we guess that the solution will have a "matrix geometric" form. That is

$$\pi_j = \pi_{j-1} R = \pi_1 R^{j-1}, \quad j = 1, 2, \cdots$$

for some fixed matrix R. Assuming this is true, then $(*)$ becomes

$$\begin{aligned} & \pi_1 R^{j-1} A_0 + \pi_1 R^j A_1 + \pi_1 R^{j+1} A_2 = \mathbf{0} \\ \Longrightarrow & \pi_1 R^{j-1} (A_0 + R A_1 + R^2 A_2) = \mathbf{0} \end{aligned}$$

Since this holds for all $j = 1, 2, \cdots$ and $\pi_1 \neq 0, R \neq 0$, we must have

$$A_0 + RA_1 + R^2A_2 = \mathbf{0} \quad (**)$$

Indeed, one can prove that R is the entry-wise smallest non-negative solution of $(**)$.
How to find R? A numerical method:

$$\begin{aligned}(**) \Longrightarrow RA_1 &= -A_0 - R^2R_2 \\ R &= -A_0A_1^{-1} - R^2A_2A_1^{-1}\end{aligned}$$

Iteratively,

$$\begin{aligned}R(0) &:= 0 \\ R(i) &= -A_0A_1^{-1} - R(i-1)^2A_2A_1^{-1}, \quad i = 1, 2, \cdots\end{aligned}$$

It can be shown that $\{R(i)\}_{i=1}^{\infty}$ is entry-wise non-decreasing, and $R(i) \uparrow R$ converges.
Once we find R, the only thing left is to find π_0, π_1. Using the boundary level equations,

$$\begin{aligned}\pi_0B_{00} + \pi_1B_{10} &= \mathbf{0} \\ \pi_0B_{01} + \pi_1B_{11} + \pi_2A_2 &= \mathbf{0} \\ \Longrightarrow \pi_0B_{01} + \pi_1(B_{11} + RA_2) &= \mathbf{0} \\ \Longrightarrow (\pi_0, \pi_0)\begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} + RA_2 \end{bmatrix} &= \mathbf{0}\end{aligned}$$

Then, use the normalization condition

$$\begin{aligned}1 = \pi\mathbf{1}' &= \pi_0\mathbf{1}' + \pi_1\mathbf{1}' + \cdots \\ &= \pi_0\mathbf{1}' + \sum_{j=1}^{\infty} \pi_j\mathbf{1}' \\ &= \pi_0\mathbf{1}' + \pi_1\sum_{j=1}^{\infty} R^{j-1}\mathbf{1}' \\ &= \pi_0\mathbf{1}' + \pi_1(I-R)^{-1}\mathbf{1}'\end{aligned}$$

6.7 Little's Formula

We define L to be the random length of the queue in a steady state (under a stationary distribution). Let W be the system time, including waiting time and service time) of a randomly arriving customer in steady state. λ is the arrival rate, which is the expected number of customers entering the system in 1 unit of time in steady state.

Theorem 6.7.1 — Little's Formula. For a queuing system in steady state, we have

$$\mathbb{E}(L) = \lambda\mathbb{E}(W)$$

the average queue length = arrival rate × average system time.

Proof. Define

$$\begin{aligned}\alpha(t) &:= \text{nubmer of customers arriving in } (0,t) \\ \beta(0,t) &:= \text{number of customers leaving in } (0,t) \\ L(t) &:= \text{nubmer of customers in the system at time } t\end{aligned}$$

Thus, $L(t) = \alpha(t) - \beta(t)$. Define $\gamma(t)$ as the total time spent in the system by time t, which is

$$\gamma(t) = \int_0^t L(s)ds$$

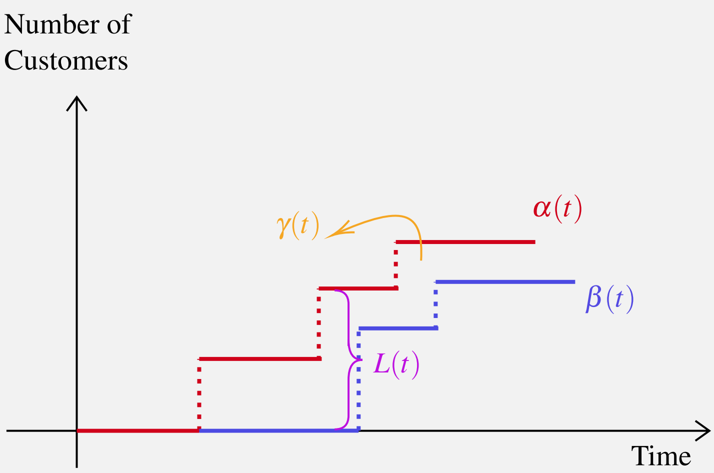


Figure 6.7.1: α, β, L, γ

Then define

$$\begin{aligned}\lambda(t) &:= \text{average arrival rate in } (0,t)\\ &= \frac{\alpha(t)}{t} \iff t = \frac{\alpha(t)}{\lambda(t)}\\ \mathbb{E}(W(t)) &= \text{average system time during } (0,t)\\ &= \frac{\gamma(t)}{\alpha(t)} \iff \gamma(t) = \mathbb{E}(W(t))\alpha(t)\\ \mathbb{E}(L(t)) &= \text{average queue length during } (0,t)\\ &= \frac{1}{t}\int_0^t L(s)ds = \frac{\gamma(t)}{t}\end{aligned}$$

Thus,

$$\mathbb{E}(L(t)) = \frac{\gamma(t)}{t} = \frac{\mathbb{E}(W(t))\alpha(t)}{\alpha(t)/\lambda(t)} = \lambda(t)\mathbb{E}(W(t))$$

Take $t \to \infty$, we get

$$\mathbb{E}(L) = \lambda\mathbb{E}(W)$$

■

R The above proof is not rigorous in two ways:

1. We did not show the existence of the limits as $t \to \infty$.
2. If we understand $\mathbb{E}(L)$ and $\mathbb{E}(W)$ as long-run averages, then the proof is rigorous after solving 1. However, if we treat $\mathbb{E}(L)$ and $\mathbb{E}(W)$ as expectations, then one should also show that the expectation, the average realizations, is the same as the average along one realization as $t \to \infty$. This is the ergodicity property of the queuing system. True, but non-trivial.

Intuitions of Little's Formula

1. $\frac{1}{\mathbb{E}(W)}$ is how fast an average customer leaves the system. Then, $\frac{\mathbb{E}(L)}{\mathbb{E}(W)}$ is the total departure rate. In steady state, we have the departure rate being equal to the arrival rate. Thus,

$$\frac{\mathbb{E}(L)}{\mathbb{E}(W)} = \lambda \iff \mathbb{E}(L) = \lambda \mathbb{E}(W)$$

2. An average customer spends $\mathbb{E}(W)$ in the system. Right before she leaves the system, she/he looks back to see the queue length. It will be simply the number of customers who arrived during her system time. That is the average queue length $= \lambda \mathbb{E}(W)$.

R Typically, we know λ from modeling, and L from solving for steady state/stationary distribution, so Little's formula is often used to derive the average system time $\mathbb{E}(W)$.

IV Renewal Theory

7. Renewal Theory

7.1 Basic Setup

Definition 7.1.1 — Counting process. A counting process $\{N(t)\}_{t\geq 0}$ is a stochastic process that represents the number of events that happen by time t.

Proposition 7.1.1 — Properties of counting processes.

1. $N(0) = 0$ actually an assumption
2. $N(t) \in \{0, 1, 2, \cdots\}$
3. $s < t \Longrightarrow N(s) \leq N(t)$

Definition 7.1.2 — Renewal process. A counting process with i.i.d. interarrival times is called a renewal process. The renewal times are the times that events happen.

$$S_0 = 0, S_1 = X_1, \cdots, S_n = S_{n-1} + X_n = \sum_{i=1}^{n} X_i$$

where $X_1, X_2, \cdots$ are i.i.d. interarrival times.

■ **Example 7.1 — Poisson process.** The Poisson process is a renewal process with exponential interarrival times. ■

R From now on, we always assume $\mathbb{E}(X_1) < \infty$. In this case, by the strong law of large number (SLLN), we have

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}(X_1) =: \mu$$

7.2 Convolution

In order to study the distribution of the sum of independent r.v.s, we introduce the technique of convolution. Let F be a CDF of a non-negative r.v. X. Let g be a non-negative function defined on

$[0,\infty)$, bounded on finite intervals. The convolution of F and g is a function on $[0,\infty)$ defined as

$$\begin{aligned}(F*g)(t) &= \mathbb{E}(g(t-X)\mathbf{1}_{\{X\le t\}})\\ &= \int_0^t g(t-x)dF(x) \qquad \text{Lebesgue-Stieltjes integral}\\ &= \begin{cases}\int_0^t g(t-x)f(x)dx & \text{continuous } X\\ \sum_{0\le x\le t} g(t-x)\mathbb{P}(X=x) & \text{discrete } X\end{cases}\end{aligned}$$

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

Proposition 7.2.1 — Properties of Convolution. 1. **Linearity:**

$$\begin{aligned}F*(cg) &= c(F*g)\\ F*(g_1+g_2) &= F*g_1+F*g_2\\ (aF_1+(1-a)F_2)*g &= a(F_1*g)+(1-a)(F_2*g) \quad a\in[0,1]\end{aligned}$$

2. If $g=G$ is the CDF of a non-negative r.v. Y and $X \perp\!\!\!\perp Y$, then $F*G(t)$ is the CDF of $X+Y$.

Proof. (a) **Algebraic proof:** assuming continuous case,

$$\begin{aligned}\mathbf{P}(X+Y\le t) &= \mathbb{E}(\mathbf{P}(X+Y\le t|X))\\ &= \int_0^t \mathbf{P}(X+Y\le t|X=x)f(x)dx\\ &= \int_0^t \mathbf{P}(Y\le t-x)f(x)dx\\ &= \int_0^t G(t-x)f(x)dx\\ &= F*G(t)\end{aligned}$$

(b) **Probabilitic proof:**

$$\begin{aligned}F*G(t) &= \mathbb{E}(G(t-X)\mathbf{1}_{\{X\le t\}})\\ &= \mathbb{E}(\mathbf{P}(Y\le t-X|X)\mathbf{1}_{X\le t})\\ &= \mathbb{E}(\mathbf{P}(Y\le t-X|X))\\ &= \mathbf{P}(Y\le t-X)\\ &= \mathbf{P}(X+Y\le t)\end{aligned}$$

■

Corollary 7.2.2 (a) If G is a CDF, then $F*G=G*F$.
(b) If X,Y has densities f,g respectively, $X \perp\!\!\!\perp Y$, then $X+Y$ has density

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

Proof. let F,G be the CDFs of X,Y respectively, then the CDF of $X+Y$ is given by

> $F * G$,
> $$\begin{aligned} h(t) &= \frac{d}{dt}(F*G)(t) \\ &= \frac{d}{dt}\int_0^t G(t-x)f(x)dx \\ &= \int_0^t g(t-x)f(x)dx + G(t-t)f(t) \quad G(0)=0 \\ &= \int_0^t g(t-x)f(x)dx \end{aligned}$$
> ■

3. $F * g$ is also non-negative and bounded on finite intervals.

   Proof. Note that
   $$F * g(t) = \mathbb{E}(\underbrace{g(t-X)}_{\geq 0}\mathbf{1}_{\{X \leq t\}}) \geq 0$$
   Moreover, if $g(s) \leq M$ on $[0,t]$, then
   $$F * g(s) = \mathbb{E}(\underbrace{g(s-X)}_{\leq M}\mathbf{1}_{\{X \leq s\}}) \leq M, \;\; s \in [0,t]$$
   Thus, $F * g$ is bounded on finite intervals.
   When $g = G$ is the CDF of a r.v. Y, then
   $$F * (F * g) = F * \underbrace{F * g}_{\text{CDF of } X+Y} = F_{X'+X+Y}$$
   where $X \overset{d}{=} X'$ are independent but having the same CDF F. Now,
   $$(F * F) * g = F_{X+X'} * G = F_{X+X'+Y}$$
   ■

As a result, we can define the $n-$**fold convolution of** F
$$F^n(t) = \underbrace{F * F * F * \cdots * F}_{n \text{ terms}}(t)$$
F^n is the CDF of th sum of n independent copies of $X, X_1, \cdots, X_n$.
Thus, if F is the CDF of the interarrival time, then F^n is the CDF of S_n, i.e,
$$F^n(t) = \mathbf{P}(S_n \leq t) = \mathbf{P}(N(t) \geq n)$$
We can express $\mathbf{P}(N(t) = n)$ in terms of F^n:
$$\begin{aligned} \mathbf{P}(N(t)=n) &= \mathbf{P}(S_n \leq t) - \mathbf{P}(S_{n+1} \leq t) \\ &= F^n(t) - F^{n+1}(t) \end{aligned}$$
Alternatively, define the survival function $\bar{F}(t) = 1 - F(t) = \mathbf{P}(X > t)$. Then,
$$\underbrace{F^n * F}_{F^{n+1}}(t) + F^n * \bar{F}(t) = F^n * 1 = F^n$$
Then,
$$\mathbf{P}(N(t)=n) = F^n(t) - F^{n+1}(t) = F^n * \bar{F}(t)$$

Now, we focus on

$$\mathbb{E}(N(t)) = \sum_{n=1}^{\infty} \mathbf{P}(N(t) \geq n)$$
$$= \sum_{n=1}^{\infty} F^n(t)$$

Definition 7.2.1 — Renewal function. The expected number of renewals by time t is called the renewal function of the renewal process

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

Proposition 7.2.3 — Properties of renewal function.

1. $m(t) \geq 0$ and $m(t)$ is non-decreasing
2. The renewal function completely determines the distribution of the interarrival time (use the properties of Laplace transform or Fourier transform).
3. $m(t) < \infty$ for all $t \geq 0$.

Proof. Fix $t > 0$.

(a) **Case 1:** $F(t) < 1$. Then $a := \mathbf{P}(X_1 > t) > 0$. For $n = 1, 2, \cdots$,

$$\begin{aligned} F^n(t) &= \mathbf{P}(S_n \leq t) \\ &= \mathbf{P}(X_1 + \cdots + X_n \leq t) \\ &\leq \mathbf{P}(X_1 \leq t, X_2 \leq t, \cdots, X_n \leq t) \\ &\overset{iid}{=} (F(t))^n = (1-a)^n \end{aligned}$$

Then,

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

(b) **Case 2:** If $F(t) = 1$, there must exist $s > 0$ such that $\mathbf{P}(X_1 > s) > 0$. Let $k \in \mathbb{Z}^+$ be such that $ks > t$. As a result,

$$\begin{aligned} a := \mathbf{P}(X_1 + \cdots + X_n > t) &\geq \mathbf{P}(X_1 > s, \cdots, X_k > s) \\ &> 0 \end{aligned}$$

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

$$N_Y(t) = \left\lceil \frac{N(t)}{k} \right\rceil$$

Then, $N(t) < kN_Y(t) + k$. Thus, it suffices to prove that $m_Y(t) = \mathbb{E}(N_Y(t)) < \infty$. However, note that $Y_1, Y_2 - Y_1, \cdots$ are i.i.d. and satisfy $\mathbf{P}(Y_1 > t) > 0$, which goes back to case Y. Hence,

$$m_Y(t) < \infty \Longrightarrow m(t) < \infty$$

■

Corollary 7.2.4 With probability 1, $N(t) < \infty$ for all $t < \infty$.

■ **Example 7.2 — Poisson process.** Consider a renewal process where the interarrival times are i.i.d exponential $\mathbf{Exp}(\lambda)$. This is just a Poisson process $\mathbf{Poi}(\lambda t)$. Sum of n i.i.d will follow $Erlang(n, \lambda)$ with CDF

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

Thus,

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

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

Renewal Equation

Condition on the first arrival/renewal:

$$\begin{aligned}
m(t) &= \mathbb{E}(N(t)) \\
&= \mathbb{E}(\mathbb{E}(N(t)|X_1)) \\
&= \int_0^{\infty} \mathbb{E}(N(t)|X_1 = x) f(x) dx \\
&= \int_0^{t} \mathbb{E}(N(t)|X_1 = x) f(x) dx && N(t) = 0 \Longleftarrow X_1 = x > t \\
&= \int_0^{t} (\mathbb{E}(N(x)|X_1 = x) + \mathbb{E}(N(t) - N(x)|X_1 = x)) f(x) dx \\
&= \int_0^{t} (1 + \mathbb{E}(N(t-x))) f(x) dx \\
&= \int_0^{t} f(x) dx + \int_0^{t} m(t-x) f(x) dx \\
&= F(t) + F * m(t) \quad t \geq 0
\end{aligned}$$

Definition 7.2.2 — Renewal equation. The equation $m(t) = F(t) + F * m(t)$ for $t \geq 0$ is referred as the renewal equation.

Expected time of the next renewal

$$
\begin{aligned}
\mathbb{E}(S_{N(t)+1}) &= \mathbb{E}\left(\sum_{j=1}^{N(t)+1} X_j\right) = \mathbb{E}\left(\sum_{j=1}^{\infty} X_j \mathbf{1}_{\{N(t)+1\geq j\}}\right) \\
&= \mathbb{E}\left(\sum_{j=1}^{\infty} X_j \mathbf{1}_{\{N(t)\geq j-1\}}\right) \\
&= \mathbb{E}\left(\sum_{j=1}^{\infty} X_j \mathbf{1}_{\{X_1+\cdots+X_{j-1}\geq t\}}\right) \\
&\overset{\text{MCT}}{=} \sum_{j=1}^{\infty} \mathbb{E}\left(X_j \mathbf{1}_{\{X_1+\cdots+X_{j-1}\leq t\}}\right) && \text{Filtration idea} \\
&= \sum_{j=1}^{\infty} \mathbb{E}(X_1)\mathbf{P}(S_{j-1}\leq t) \\
&= \mu \sum_{j=1}^{\infty} F^{j-1}(t) \\
&= \mu(\sum_{n=1}^{\infty} F^n(t) + F^0(t)) = \mu(m(t)+1)
\end{aligned}
$$

Theorem 7.2.5 For a renewal process whose interarrival times have pdf f and CDF F, if a funtion $z(t)$ satisfies

$$
\begin{aligned}
z(t) &= g(t) + \int_0^t z(t-x)f(x)dx \\
&= g(t) + F * z(t), \quad t \geq 0
\end{aligned}
$$

where g is a function that is bounded in finite intervals, then the unique solution for $z(t)$ which is bounded on finite intervals is

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

(This extends the definition of convolution to the case where g is not non-negative)

Proof. We first check that $z(t) = g(t) + m * g(t)$ is a solution of the equation

$$
z(t) = g(t) + F * z(t)
$$

Note that

$$
\begin{aligned}
F * z(t) &= F * (g(t) + m * g(t)) \\
&= F * g(t) + F * m * g(t) = F * g(t) + F * \left(\sum_{n=1}^{\infty} F^n * g(t)\right) \\
&= F * g(t) + \sum_{n=1}^{\infty} F^{n+1} * g(t) \\
&= \sum_{n=1}^{\infty} F^n * g(t) = m * g(t)
\end{aligned}
$$

Thus,

$$
\text{LHS} = z(t) = g(t) + m * g(t)
$$

and

$$\text{RHS} = g(t) + F * z(t) = g(t) + m * g(t)$$

For uniqueness, let z_1, z_2 be two solutions of the equation $z(t) = g(t) + F * z(t)$ which are bounded on finite intervals. Then, let $h(t) = z_1(t) - z_2(t)$ and note that $h(t) = F * h(t)$. Note that h is invariant of the convolution with F. Thus, $h(t) = F^n * h(t)$ for any $n \geq 1$. Then,

$$h(t) = \lim_{n\to\infty} F^n * h(t)$$

For each $n \geq 1$, we have

$$F^n * h(t) = \mathbb{E}\big(h(t - S_n)\mathbf{1}_{\{S_n \leq t\}}\big)$$

Since z_1, z_2 are bounded on finite intervals $[0,t]$, so is h. Take M such that $|h(s)| \leq M$ for any $s \in [0,t]$. Then,

$$\begin{aligned} |F^n * h(t)| &= \big|\mathbb{E}\big(h(t - S_n)\mathbf{1}_{\{S_n \leq t\}}\big)\big| \\ &\leq \mathbb{E}\big(|h(t - S_n)|\mathbf{1}_{\{S_n \leq t\}}\big) \qquad \text{Jensen's Inequality} \\ &= \mathbb{E}(M\mathbf{1}_{S_n \leq t}) \\ &= M\mathbf{P}(S_n \leq t) = MF^n(t) \end{aligned}$$

Since $m(t) = \sum_{n=1}^{\infty} F^n(t) < \infty$, then $\lim_{n\to\infty} F^n(t) \to 0$ point-wise. Then,

$$\lim_{n\to\infty} F^n * h(t) = h(t) = 0$$

As this folds for all t, we conclude that $z_1 = z_2$. ■

R The standard renewal equation for $m(t)$

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

satisfies the conditiions of the above theorem 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.