

# STATS 3U03

Sang Woo Park

January 24, 2017

## Course Outline

- Textbook: Introduction to stochastic processes
- Requirement: 5 assignments, 2 tests, and 1 final
- Test 1: Friday, February 10th
- Test 2: Friday, March 17th

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Review . . . . .	2
1.2	Stochastic processes . . . . .	2
<b>2</b>	<b>Markov chains (Discrete time Markov chains)</b>	<b>3</b>
2.1	Markov property . . . . .	3
2.2	Transition function and initial distribution . . . . .	5
2.3	Joint distribution . . . . .	8
2.4	Recurrence . . . . .	11
2.5	Absorption probabilities . . . . .	17
2.6	Birth-Death Markov Chain . . . . .	19

# 1 Introduction

## 1.1 Review

**Definition 1.1** (Independent random variables).  *$X$  and  $Y$  are independent iff for any  $a, b \in \mathbb{R}$ ,  $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$*

## 1.2 Stochastic processes

**Definition 1.2** (Stochastic process). *Let  $T$  be a subset of  $[0, +\infty]$ . For each  $t \in T$ , let  $X_t$  be a random variable. Then, the collection of  $\{X_t : t \in T\}$  is called a stochastic process. Simply put, a stochastic process is just a family of random variables.*

**Example 1.2.1.** Let  $T = \{0\}$ . Then,  $\{X_0\}$  is a stochastic process.

**Example 1.2.2.** Let  $T = \{1, 2, 3, \dots, m\}$  be a set of finite natural numbers. Then,  $\{X_1, X_2, X_3, \dots, X_m\}$  is a stochastic process.

**Example 1.2.3.** Let  $T = \{0, 1, 2, \dots\}$  be a set of all non-negative integers. Then,  $\{X_1, X_2, X_3, \dots\}$  is a stochastic process.

**Example 1.2.4.** Let  $T = [0, +\infty)$  be a set of all non-negative real numbers. Then,  $\{X_t : t \geq 0\}$  is a stochastic process.

**Definition 1.3** (Time index). *Let  $T$  be time index. If  $T = \{0, 1, 2, \dots\}$ , then the time is discrete. If  $T = [0, \infty)$ , then time is continuous.*

**Definition 1.4** (State Space). *State space,  $S$ , is the space where the random variable takes the values.*

Given a sample space,  $S$ , and time index  $t \in T$ , we can define  $X_t(w) \in S$ , to describe a stochastic process. Here,  $\{X_t : t \in T\}$  describes the dependence relation.

We can further categorize a stochastic process by considering the following two cases: countable and uncountable state space. Time index can also be categorized as follows: discrete and continuous time. Note that each stochastic process must belong to one of the four categories.

*Remark.* Every stochastic process can be described by the following three factors:

1. Time index
2. State space
3. Dependence relation

**Example 1.2.5.** Let  $S = \{0, 1\}$  and  $T = \{0, 1, 2, \dots\}$ . Given,

$$X_n = \begin{cases} 1 & \text{with probability of } 1/2 \\ 0 & \text{with probability of } 1/2 \end{cases}$$

$\{X_0, X_1, X_2, \dots\}$  is a stochastic process and is often noted as Bernoulli trials.

## 2 Markov chains (Discrete time Markov chains)

We will only be dealing with discrete time Markov chains in chapter 1 and 2. In other words,  $T = \{0, 1, 2, \dots\}$ . It follows that the state space,  $S$ , will be at most countable. Finally, Markov describes the dependence relation:  $X_0, X_1, X_2, \dots$ .

In example 1.2.5, every trial of the Bernoulli trials was independent. On the other hand, in a Markov model,  $X_{n+1}$  depends on  $X_n$  but not on any past states,  $X_1, X_2, \dots, X_{n-1}$ .

### 2.1 Markov property

**Definition 2.1.** Markov property can be expressed as follows:

$$\begin{aligned} P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\ = P(X_{n+1} = x_{n+1} | X_n = x_n) \end{aligned}$$

$P(X_{n+1} = y | X_n = x)$  is noted as the transition probability and it describes the one step transition from  $x$  to  $y$  starting at time  $n$ . If

$$P(X_{n+1} = y | X_n = x) = P(X_1 = y | X_0 = x),$$

then the Markov chain is called to have stationary transition, or homogeneous.

**Definition 2.2.** Let  $\{X_n : n = 0, 1, 2, \dots\}$  be a homogeneous Markov chain. Then,

$$P_{xy} = P(X_1 = y | X_0 = x) = P(X_{n+1} = y | X_n = x),$$

is the one-step transition probability.

**Definition 2.3.** Following the definition 2.2, we can now define one-step transition matrix:

$$\mathbb{P} = (P_{xy})_{x,y \in S}$$

*Remark.* Given,  $X_0$ ,  $\pi_0(x) = P(X_0 = x)$  is called the *initial distribution*.

Given a Markov chain, we wish to answer the following fundamental questions:

1. Distribution of  $X_n$  for any  $n \geq 1$ .
2. Joint distribution of  $X_{n_1}, \dots, X_{n_k}$  for any  $1 \leq n_1 < n_2 < \dots < n_k$   $k, \geq 2$ .
3. Long time behaviour, i.e.

$$\lim_{n \rightarrow \infty} P(X_n = x)$$

**Example 2.1.1.** We have the following Markov chain:  $\{X_n : n = 0, 1, 2, \dots\}$  where  $S = \{0, 1\}$ . For this model, its initial distribution can be described as follows:

$$\begin{cases} \pi_0(0) = P(X_0 = 0) = a \\ \pi_0(1) = 1 - a \end{cases}$$

Transition probabilities can be written in a similar fashion:

$$\begin{aligned} P(X_1 = 1|X_0 = 0) &= p, & P(X_1 = 0|X_0 = 0) &= 1 - p \\ P(X_1 = 0|X_0 = 1) &= q, & P(X_1 = 1|X_0 = 1) &= 1 - q \end{aligned}$$

where  $0 \leq p, q \leq 1$ . For this Markov chain, we can consider the following three cases:

**Case 1.**  $p = q = 0$ .

This case is trivial.

**Case 2.**  $p = q = 1$ .

This case is also trivial.

**Case 3.**  $0 \leq p + q \leq 2$ .

$$\begin{aligned} P(X_{n+1} = 0) &= P(X_{n+1} = 0 \cap X_n = 0) + P(X_{n+1} = 0 \cap X_n = 1) \\ &= P(X_n = 0)P(X_{n+1} = 0|X_n = 0) + P(X_n = 1)P(X_{n+1} = 0|X_n = 1) \\ &= P(X_n = 0)(1 - p) + P(X_n = 1)q \\ &= P(X_n = 0)(1 - p) + (1 - P(X_n = 0))q \\ &= (1 - p - q)P(X_n = 0) + q \end{aligned}$$

We can further expand this as follows:

$$\begin{aligned} P(X_{n+1} = 0) &= (1 - p - q)P(X_n = 0) + q \\ &= (1 - p - q)[(1 - p - q)P(X_{n-1} = 0) + q] + q \\ &= (1 - p - q)^n P(X_0 = 0) + q \sum_{j=0}^{n-1} (1 - p - q)^j \end{aligned}$$

Note that

$$\sum_{j=0}^{n-1} (1 - p - q)^j = \frac{(1 - p - q)^n - 1}{(1 - p - q) - 1}$$

Therefore, we have

$$\begin{aligned}
P(X_{n+1} = 0) &= (1 - p - q)^n a + q \frac{(1 - p - q)^n - 1}{(1 - p - q) - 1} \\
&= (1 - p - q)^n a - \frac{q}{p + q} ((1 - p - q)^n - 1)
\end{aligned}$$

For this Markov chain, we find that

$$\lim_{n \rightarrow \infty} P(X_n = 0) = \frac{q}{p + q}$$

## 2.2 Transition function and initial distribution

**Example 2.2.1.**

$$\begin{aligned}
P_{xy} &= P(X_{n+1} = y | X_n = x) \\
&= P(X_1 = y | X_0 = x)
\end{aligned}$$

**Definition 2.4.** Transition function,  $P(x, y) : S \times X \rightarrow [0, 1]$ , satisfies the following conditions:

1.  $P(x, y) \geq 0$ .
2.  $\sum_{y \in S} P(x, y) = 1$  for all  $x \in S$ .

**Definition 2.5.** Given a transition function,  $P(x, y)$ , a transition matrix is defined as follows:

$$\mathbb{P} = (P(x, y))_{x, y \in S}$$

**Example 2.2.2.**

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

**Example 2.2.3.**

$$\begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/8 & 1/4 & 5/8 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

**Definition 2.6.** Initial distribution is a probability mass function (pmf) that is defined as follows

$$\pi_0(x) = P(X_0 = x).$$

Note that it must satisfy the following conditions:

1.  $\pi_0(x) \geq 0$
2.  $\sum_{x \in S} \pi_0(x) = 1$

**Theorem 2.1.** Let  $\{x_N : n = 0, 1, 2, \dots\}$  be a Markov chain with initial distribution  $\pi_0(x)$ , and one-step transition matrix  $\mathbb{P} = (P(x, y))_{x, y \in S}$ . Then, the distribution of  $X_n$  is

$$\begin{aligned} P(X_n = x_n) &= \sum_{x_0 \in S} \sum_{x_1 \in S} \cdots \sum_{x_{n-1} \in S} \pi_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n) \\ &= \pi_0 \underbrace{\mathbb{P} \mathbb{P} \cdots \mathbb{P}}_n \end{aligned}$$

*Proof.* For any  $n \geq 1$ ,  $x_n \in S$

$$\begin{aligned} P(X_n = x_n) &= P(X_n = x_n, x_0 \in S, X_1 \in S, \dots, X_{n-1} \in S) \\ &= \sum_{x_0 \in S} \sum_{x_1 \in S} \cdots \sum_{x_{n-1} \in S} P(X_n = x_n, X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \end{aligned}$$

Note that

$$\begin{aligned} &P(X_n = x_n, X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &= P(X_0 = x_0) P(X_1 = x_1 | X_0 = x_0) P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \\ &\cdots P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \end{aligned}$$

Using the Markov property, it is evident that the equation above is equivalent to  $P(X_0 = x_0) P(X_1 = x_1 | X_0 = x_0) \cdots P(X_n = x_n | X_{n-1} = x_{n-1})$ .  $\square$

**Example 2.2.4.** Simple random walk is a Markov chain:

$$\begin{aligned} X_0 &= 0 \\ X_1 &= \begin{cases} 1 & p \\ -1 & q \end{cases} \\ X_2 &= \begin{cases} X_1 + 1 & p \\ X_1 - 1 & q \end{cases} \\ X_n &= \begin{cases} X_{n-1} + 1 & p \\ X_{n-1} - 1 & q \end{cases} \end{aligned}$$

where  $S = \{0, \pm 1, \pm 2, \dots\}$ .

**Example 2.2.5** (Ehrenfest chain). Suppose that we have a box and an invisible bar that divides the box into region I and II.  $d$  balls are placed in a box. Initially,  $n$  balls are distributed in region I and  $d - n$  balls are distributed in region II. You pick a ball at random. If it's from region I, you put it back in region II. If it's from region II, you put it back in region I.

First, note that this Markov chain has a state space of  $S = \{0, 1, 2, \dots, d\}$ . We observe that

$$P(0, y) = \begin{cases} 0 & y > 1 \\ 1 & y = 1 \end{cases}$$

$$P(1, y) = \begin{cases} 0 & y \neq 0, 2 \\ \frac{1}{d} & y = 0 \\ 1 - \frac{1}{d} & y = 2 \end{cases}$$

In general, we have

$$P(x, y) = \begin{cases} 0 & y \neq x \pm 1 \\ 1 - \frac{x}{d} & y = x + 1 \\ \frac{x}{d} & y = x - 1 \end{cases}$$

Combining these results, we have the following transition matrix:

$$\begin{pmatrix} 0 & 1 & 0 & \dots & \\ \frac{1}{d} & 0 & 1 - \frac{1}{d} & & \\ & & & \ddots & \\ & & & & \ddots \end{pmatrix}_{(d+1) \times (d+1)}$$

**Example 2.2.6** (Birth-Death Markov chain). At each time step, one person can die and a new person can be born:

$$X_{n+1} = \begin{cases} p_x & y = X_n + 1 \\ q_x & y = X_n - 1 \\ r_x & y = X_n \\ 0 & \text{else} \end{cases}$$

**Example 2.2.7** (Queuing chain). At each time step, one customer is served and new customers arrive:

$$X_{n+1} = \begin{cases} y_{n+1} & \text{if } X_n = 0 \\ X_n - 1 + y_{n+1} & \text{if } X_n \geq 1 \end{cases}$$

We introduce a new notation,  $x^+ = x \vee 0$ , which is essentially  $\max(x, 0)$ . Using this notation, we can rewrite the Markov chain as follows

$$X_{n+1} = (X_n - 1)^+ + y_n$$

**Example 2.2.8** (Branching Markov chain). If  $X_0 = 0$ , then  $X_n = 0$  for all  $n \geq 1$ . We call 0 an *absorbing state*.

Suppose  $X_0 \geq 1$ . An individual,  $i$ , will produce  $y_i$  number of offsprings at each generation. Then, we will have

$$X_1 = y_1^{(1)} + \dots + y_{X_0}^{(1)}$$

Each individual in generation 1 will also produce offsprings. Then,

$$X_2 = y_1^{(2)} + \cdots + y_{X_0}^{(2)}$$

We wish to understand how the population will evolve over time. To do so, we can look at the expected value. It's clear that the population will grow if  $E[y] > 1$ . On the other hand, if  $E[y] < 1$ , the population will eventually die out.

**Example 2.2.9** (Wright-Fisher Markov chain). For this Markov chain, we start by make the following assumptions:

1. The population size is fixed.
2. No generation overlap.

Within the population, there are  $N$  number of individuals of two types: I and II. Let  $X_0$  be number of type I individuals at time 0. Each individual in generation 1 pick its parent from generation 0 at random. This process is equivalent to repeating Bernoulli trials  $N$  times (also equivalent to binomial).

Therefore, we have

$$\begin{aligned} X_1 &\sim \text{Bin}(N, \frac{X_0}{N}) \\ X_2 &\sim \text{Bin}(N, \frac{X_1}{N}) \\ &\vdots \\ X_{n+1} &\sim \text{Bin}(N, \frac{X_n}{N}) \end{aligned}$$

### 2.3 Joint distribution

Given a Markov chain with  $\pi_0$  and  $\mathbb{P}$ , how do we find (1) the distribution of  $X_n$  and (2) the joint distribution of  $X_n$  and  $X_m$  where  $n < m$ ?

From the previous section, recall that  $\pi_n = \pi_0 \underbrace{\mathbb{P}\mathbb{P}\dots\mathbb{P}}_n$ .

**Example 2.3.1.** Consider the following transition matrix:

$$\mathbb{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Suppose that  $\pi_0 = (1, 0, 0, 0)$ . Then, we have

$$\pi_1 = (1, 0, 0, 0)\mathbb{P} = (1/2, 1/2, 0, 0)$$

$$\pi_2 = (1, 0, 0, 0)\mathbb{P}\mathbb{P} = (1/4, 1/2, 1/4, 0)$$

Eventually, all states will converge to the absorbing state and stay there.



To find the join distribution, we first note that

$$\begin{aligned} P(X_n = x, X_m = x_m) &= P(X_n = x_n)P(X_m = x_m|X_n = x_n) \\ &= P(X_m = x_m)P(X_{m-n} = x_m|X_0 = x_n) \end{aligned}$$

**Definition 2.7.** For any interger  $m$ ,  $m$ -step transition matrix is given by

$$P^m(x, y) = P(X_m = y|X_0 = x).$$

When  $m = 0$ , we have

$$P^0(x, y) = \begin{cases} 1 & y = x \\ 0 & y \neq x \end{cases}$$

We can decompose  $m$ -step transition matrix as follows:

$$\begin{aligned} P^m(x, y) &= P(X_m = y|X_0 = x) \\ &= P(X_m = y, x_1 \in S, \dots, X_{m-1} \in S|X_0 = x) \\ &= \sum_{x_1 \in S} \sum_{x_2 \in S} \cdots \sum_{x_{m-1} \in S} P(x_0, x_1) \cdots P(x_{m-1}, y) \end{aligned}$$

Then, we have

$$\begin{aligned} P(X_m = y) &= P(X_m = y, X_m \in S) \\ &= \sum_{x_0 \in S} P(X_m = y, X_0 = x_0) \\ &= \sum_{x_0 \in S} P(X_0 = x_0)P^m(X_0, y) \\ &= \sum_{x_0 \in S} \pi_0(x_0)P^m(X_0, y) \end{aligned}$$

Therefore, we have

$$(P(X_m = x_m))_{x_m \in S} = \pi_0 \mathbb{P}^m$$

**Definition 2.8** (Hitting time). Given  $A \subset S$ , hitting time  $T_A$  is defined as follows:

$$T_a = \min\{n \geq 1 : X_n \in A\}$$

If  $A = \{x\}$ , then we have  $T_x = T_x$ . Note that

- $T_A \geq 1$
- If  $x_n \notin A$  for all  $n \geq 1$ . we have  $T_A = +\infty$

Now, we wish to understand the distribution of  $T_y$  given that  $X_0 = x$ . First, note that we have

$$P_x(T_y) = 1 = P(T_y = 1 | X_0 = x) = P(x, y)$$

Similarly, we have

$$\begin{aligned} P_x(T_y = 2) &= P_x(x_1 \neq y, x_2 = y) \\ &= \sum_{w \neq y} P(x, w) P(w, y) \end{aligned}$$

Generally, we have

$$\begin{aligned} P_x(T_y = n + 1) &= P_x(x_1 \neq y, \dots, x_n \neq y, x_{n+1} = y) \\ &= \frac{P(X_0 = x, X_1 \neq y \dots X_n \neq y, X_{n+1} = y)}{P(X_0 = x)} \\ &= \frac{P(X_0 = x, X_1 \neq y)}{P(X_0 = x)} \frac{P(X_0 = x, X_1 \neq y \dots X_n \neq y, X_{n+1} = y)}{P(X_0 = x, X_1 \neq y)} \\ &= \sum_{x_1 \neq y} P(x, x_1) P_{x_1}(T_y = n) \end{aligned}$$

Note that the last result follows from the Markov property.

**Lemma 2.1.**

$$P^m(x, y) = \sum_{k=1}^m P_x(T_y = k) P^{m-k}(y, y)$$

*Proof.*

$$\begin{aligned} P^m(x, y) &= P(X_m = y | X_0 = x) \\ &= P(X_m = y, T_y \leq m | X_0 = x) \\ &= \sum_{k=1}^m P(X_m = y, T_y = k | X_0 = x) \\ &= \sum_{k=1}^m \frac{P(X_0 = x, T_y = k, X_m = y)}{P(X_0 = x)} \\ &= \sum_{k=1}^m \frac{P(X_0 = x, T_y = k)}{P(X_0 = x)} \frac{P(X_0 = x, T_y = k, X_m = y)}{P(X_0 = x, T_y = k)} \\ &= \sum_{k=1}^m P_x(T_y = k) P(X_m = y | X_0 = x, x \neq y, x_k = y) \\ &= \sum_{k=1}^m P_x(T_y = k) P(X_m = y | x_k = y) \end{aligned}$$

□

## 2.4 Recurrence

Before we define recurrent and transient states, we introduce the following notation:

$$\rho_{xy} = P_x(T_y \leq \infty) = \sum_k P_x(T_y = k).$$

**Definition 2.9** (Recurrent and Transient states). *A state  $x$  is called recurrent if  $P_{xx} = 1$ . Otherwise, it is called transient.*

We introduce more notations:

- $I_x(y) = \begin{cases} 1 & y = x \\ 0 & \text{else} \end{cases}$  (indicator function of  $x$ ).
- $N(y) = \sum_{n=1}^{\infty} I_y(X_n)$

**Theorem 2.2.**

1.  $P_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1}$
2.  $P_x(N(y) = m) = \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy})$
3.  $P_x(N(y) = 0) = 1 - \rho_{xy}$

*Proof.* First, assume that theorem 1 is true. Then, we have

$$\begin{aligned} P_x(N(y) = m) &= P_x(N(y) \geq m) - P_x(N(y) \geq m+1) \\ &= \rho_{xy}\rho_{yy}^{m-1} - \rho_{xy}\rho_{yy}^m \\ &= \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy}) \end{aligned}$$

Now, we want to prove theorem 3:

$$\begin{aligned} P_x(N(y) = 0) &= 1 - P_x(N(y) \geq 1) \\ &= 1 - \rho_{xy} \end{aligned}$$

Finally, we just have to prove theorem 1:

$$\begin{aligned} P_x(N(y) \geq m) &= P_x(\text{The Markov chain visits state } y \text{ at least } m \text{ times}) \\ &= \sum_{n_1 \geq 1} \cdots \sum_{n_m \geq 1} P_x(T_y = n_1)P_y(T_y = n_2) \cdots P_y(T_y = n_m) \\ &= \sum_{n_1 \geq 1} P_x(T_y = n_1) \sum_{n_2 \geq 1} P_y(T_y = n_2) \cdots \sum_{n_m \geq 1} P_y(T_y = n_m) \\ &= P_x(T_y < \infty)P_y(T_y < \infty) \cdots P_y(T_y < \infty) \\ &= \rho_{xy}\rho_{yy}^{m-1} \end{aligned}$$

□

Before looking at the next theorem, we introduce another notation:  $E_x[\cdot]$  is the expectation given the initial state of  $x$ . Then, we have

$$\begin{aligned} E_x[I_y(X_n)] &= P_x(I_y(X_n) = 1) \\ &= P_x(X_n = y) \\ &= P^n(x, y) \end{aligned}$$

Furthermore, we introduce the notation,  $G$ :

$$\begin{aligned} G(x, y) &= E_x[N(y)] \\ &= E_x \left[ \sum_y I_y(x_n) \right] \\ &= \sum_y E_x[I_y(x_n)] \\ &= \sum_y P^n(x, y) \end{aligned}$$

**Theorem 2.3.**

1. If  $y$  is transient, then for any  $x \in S$ ,  $P_x(N(y) < \infty) = 1$  and  $G(x, y) = \frac{P_{xy}}{1 - P_{yy}} < \infty$ .
2. If  $y$  is recurrent, then for any  $x \in S$ ,  $P_x(N(y) = \infty) = 1$  and  $G(y, y) = \infty$ . Furthermore, we have  $P_x(N(y) = \infty) = \rho_{xy}$  and

$$G(x, y) = \begin{cases} \infty & \text{if } \rho_{xy} > 0 \\ 0 & \text{if } \rho_{xy} = 0 \end{cases}$$

*Proof.* Suppose  $y$  is transient. Then, we have  $\rho < 1$ . For any  $x$ , we have

$$\begin{aligned} P_x(N(y) = \infty) &= P_x \left( \bigcap_{m=1}^{\infty} \{N(y) \geq m\} \right) \\ &= \lim_{m \rightarrow \infty} P_x(N(y) \geq m) \\ &= \lim_{m \rightarrow \infty} \rho_{xy} \rho_{yy}^{m-1} \\ &= 0 \end{aligned}$$

Therefore, we have

$$P_x(N(y) < \infty) = 1 - P_x(N(y) = \infty) = 1 - 0 = 1$$

Furthermore,

$$\begin{aligned}
G(x, y) &= \sum_{m=1}^{\infty} m P_x(N(y) = m) \\
&= \sum_{m=1}^{\infty} m \rho_{xy} \rho_{yy}^{m-1} (1 - \rho_{yy}) \\
&= \rho_{xy} (1 - \rho_{yy}) \sum_{m=1}^{\infty} m \rho_{yy}^{m-1} \\
&= \rho_{xy} (1 - \rho_{yy}) \left( \sum_{m=1}^{\infty} \rho_{yy}^m \right)' \\
&= \rho_{xy} (1 - \rho_{yy}) \frac{1}{(1 - \rho_{yy})^2}
\end{aligned}$$

Let's prove the second statement. If  $y$  is recurrent, then  $\rho_{yy} = 1$ . For any  $x$ , we have

$$\begin{aligned}
P_x(N(y) = \infty) &= P_x\left(\bigcap \{N(y) \geq m\}\right) \\
&= \lim_{m \rightarrow \infty} P_x(N(y) \geq m) \\
&= \lim_{m \rightarrow \infty} \rho_{xy} \rho_{yy}^{m-1} \\
&= \rho_{xy}
\end{aligned}$$

Then, we have

$$\begin{aligned}
G(x, y) &= \sum m \rho_{xy} \rho_{yy}^{m-1} \\
&= \sum m \rho_{xy} \\
&= \rho_{xy} \sum m \\
&= \infty
\end{aligned}$$

□

**Example 2.4.1.** Let  $y$  be a transient state. Find

$$\lim_{n \rightarrow \infty} P^n(x, y).$$

Recall that  $G(x, y) = \sum_{n=1}^{\infty} P^n(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty$ . Since the series converges, it is easy to see that  $\lim_{n \rightarrow \infty} P^n(x, y) = 0$

**Example 2.4.2.** Let  $\{X_n, n \geq 0\}$  be a two state Markov  $S = \{0, 1\}$ . Can both be transient?

We start by noting that  $P_x(X_n \in S) = 1$ . If both are transient, we have

$$\lim_{n \rightarrow \infty} P_x(X_n \in S) = \lim_{n \rightarrow \infty} P^n(x, 0) + P^n(x, 1) = 0,$$

yielding a contradiction.

**Definition 2.10.** A Markov chain is recurrent if all states are recurrent, and the chain is transient if all states are transient.

**Definition 2.11.** A state  $x$  leads to state  $y$  if  $\rho_{xy} > 0$  denoted  $x \rightarrow y$ .

*Remark.* It is possible that  $x \not\rightarrow x$ .

**Lemma 2.2.**

1.  $x \rightarrow y$  iff there exists  $n \geq 1$  such that  $P^n(x, y) > 0$ .
2. If  $x \rightarrow y$  and  $y \rightarrow z$ , then  $x \rightarrow z$ .

*Proof.* By definition,  $x \rightarrow y$  iff  $\rho_{xy} > 0$ . In other words,  $P_x(T_y < \infty) > 0$ . Then,

$$0 < P_x(T_y < \infty) = \sum_{n=1}^{\infty} P_x(T_y = n) \leq \sum_{n=1}^{\infty} P^n(x, y)$$

Therefore,  $P^n(x, y) > 0$  for some  $n$ . Conversely, if  $P^n(x, y) > 0$  for some  $n \geq 1$ , we can define

$$n_0 = \min\{n \geq 1, P^n(x, y) > 0\}.$$

Clearly,  $0 < P^n(x, y) \leq P_x(T_y = n_0) < \rho_{xy}$ .

To prove the second statement, note that  $x \rightarrow y$  iff  $\exists n_1 \geq 1$  such that  $P^{n_1}(x, y) > 0$ . Similarly,  $y \rightarrow z$  iff  $\exists n_2 \geq 1$  such that  $P^{n_2}(y, z) > 0$ . Then,

$$P^{n_1+n_2}(x, z) \geq P^{n_1}(x, y)P^{n_2}(y, z) > 0$$

Therefore,

$$x \rightarrow z$$

□

**Theorem 2.4.** If  $x$  is recurrent and  $x \rightarrow y$ , then  $y$  is recurrent and  $\rho_{xy} = \rho_{yx} = 1$ .

*Proof.* To yield contradiction, suppose  $\rho_{yx} \neq 1$ . Then,

$$1 - \rho_{yx} > 0.$$

Furthermore, if  $x \rightarrow y$ , there exists  $n_1$  such that  $P^{n_1}(x, y) > 0$ . This implies that

$$P^{n_1}(x, y)(1 - \rho_{yx}) > 0$$

The first part is the probability that  $x$  reaches  $y$  in  $n_1$  steps. However, the second part says that  $y$  never goes to  $x$ , contradicting the assumption that  $x$  is recurrent. Therefore,

$$\rho_{yx} = 1.$$

To prove that  $y$  is recurrent, we first note that if  $x \rightarrow y$ , there exists  $n_1$  such that  $P^{n_1}(x, y) > 0$ . Similarly, if  $\rho_{yx} = 1$ ,  $y \rightarrow x$  and there exists  $n_2$  such that  $P^{n_2}(y, x) > 0$ . Then,

$$\begin{aligned} G(y, y) &= \sum_{n=1}^{\infty} P^n(y, y) \\ &\geq \sum_{m=1}^{\infty} P^{n_1+n_2+m}(y, y) \\ &\geq \sum_{m=1}^{\infty} P^{n_2}(y, x) P^m(x, x) P^{n_1}(x, y) \\ &= P^{n_2}(y, x) \underbrace{\left( \sum_{m=1}^{\infty} P^m(x, x) \right)}_{G(x, x)=\infty} P^{n_1}(x, y) \end{aligned}$$

Finally, to prove that  $\rho_{xy} = 1$ , we note that  $y$  is recurrent. Then, by following the proof of the first statement, we can prove that  $\rho_{xy} = 1$ .  $\square$

**Definition 2.12.** If  $x \rightarrow y$  and  $y \rightarrow x$ , we write

$$x \leftrightarrow y$$

and say that  $x$  communicates with  $y$

**Definition 2.13.** A subset  $C$  is closed if for any  $x \in C$  and  $y \notin C$ ,  $x \not\rightarrow y$  ( $\rho_{xy} \leq 0$ ).

**Definition 2.14.** A closed subset  $C$  is irreducible if every  $x, y \in C$  communicate with each other.

We can further define *closed and irreducible set* where (1)  $x, y \in C$ ,  $x \leftrightarrow y$ , and (2)  $x \in C$ ,  $z \notin C$ ,  $\rho_{xz} = 0$ . *Closed, irreducible, and recurrent set* is then defined as (1)  $x, y \in C$ ,  $x \leftrightarrow y$ ,  $\rho_{xy} = \rho_{yx} = 1$ , and (2)  $x \in C$ ,  $z \notin C$ ,  $\rho_{xz} = 0$ .

Then, we can decompose a state space,  $S$ , into a set of recurrent and transient state:

$$S = C_R \cup C_T$$

**Theorem 2.5.** If for  $x, y \in C_R$ ,  $C_x \cap C_y \neq \emptyset$ . Then,  $C_x = C_y$ .

*Proof.* Let  $w \in C_x \cap C_y$ . Then,  $w \leftrightarrow x$  and  $w \leftrightarrow y$ . For any  $z \in C_x$ , we have

$$z \leftrightarrow x \leftrightarrow w \leftrightarrow y$$

and

$$z \in C_y,$$

implying that  $C_x \subset C_y$ . By symmetry,  $C_y \subset C_x$ . Therefore,  $C_x = C_y$ .  $\square$

**Theorem 2.6.** *The state space  $S$  of a Markov chain can be decomposed as two union of  $C_R$  and  $C_T$ . Furthermore,  $C_R$  can be decomposed into the union of at most countable number of closed, irreducible, recurrent sets.*

Note that you have to stay in a recurrent set if you start from a recurrent set. On the other hand, if you start from a transient set, you have to move to a recurrent state if the set contains finite elements. If the set contains infinite number of elements, it is possible to stay in the transient set forever.

**Example 2.4.3.** Let  $\{X_n, n = 0, 1, 2, \dots\}$  be a Markov Chain with  $S = \{0, 1, 2, 3, 4, 5\}$  and the following one step transient matrix:

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\ 0 & 1/5 & 2/5 & 1/5 & 0 & 1/5 \\ 0 & 0 & 0 & 1/6 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/3 \\ 0 & 0 & 0 & 1/4 & 0 & 3/4 \end{pmatrix}$$

(a) Find  $C_R$  and  $C_T$

Note that 0 is an absorbing state. If you start from state 1 or 2, you have a positive probability of going to state 0. Therefore, state 1 and 2 are transient. On the other hand, we have  $3 \rightarrow 4 \rightarrow 5 \rightarrow 3$ , implying that

$$3 \leftrightarrow 4 \leftrightarrow 5.$$

Then,  $\{3, 4, 5\}$  form a closed, irreducible, and recurrent state. Therefore,

$$C_R = \{0, 3, 4, 5\}$$

$$C_T = \{1, 2\}$$

(b) Decompose  $C_R$

Clearly,  $C_R = \{0\} \cup \{3, 4, 5\}$  and two subsets are irreducible.

*Remark.* All closed, irreducible, finite set are recurrent set.

**Example 2.4.4.** Let  $S = \{0, 1, 2, 3, \dots\}$ . Given the following transition matrix,

$$\mathbb{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \\ 0 & 0 & 1 & 0 & \\ \vdots & & & & \end{pmatrix}$$



Then, each state is an absorbing state and we have

$$C_R = \bigcap_{i=0}^{\infty} \{i\}.$$

**Example 2.4.5.** Consider the following transtion matrix:

$$\mathbb{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Then, since  $1 \rightarrow 2 \rightarrow 3$ , we have

$$C_R = \{1, 2, 3\}$$

**Example 2.4.6.** Consider the following transition matrix:

$$\mathbb{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Then, we have  $C_T = \{1, 2\}$  and  $C_R = \{3\}$ .

**Example 2.4.7.** Consider the following transition matrix:

$$\mathbb{R} = \begin{pmatrix} 0 & 1 & 0 \\ a & 0 & 1-a \\ 0 & 0 & 1 \end{pmatrix}$$

For all  $0 \leq a < 1$ , decomposition of the state space does not change. Higher  $a$  only implies that it will take longer to get to the absorbing state.

## 2.5 Absortion probabilities

**Definition 2.15** (Absortion probabilities). *Let  $C$  be a recurrent, irreducible, closed set. For  $x \in C_T$ , probability of  $x$  being abosrbed by  $C$  is given by*

$$\rho_C(x) = \rho_x(T_c < \infty)$$

To calculate the absortion state, we must solve

$$\rho_C(x) = \sum_{y \in C} P(x, y) + \sum_{y \in C_T} P(x, y) \rho_C(y).$$

This is in fact a system of linear equations. We are interested the uniqueness of the solution.

**Theorem 2.7.** *If  $C_T$  is finite, then the system*

$$w_x = \sum_{y \in C} P(x, y) + \sum_{y \in C_T} w_y$$

*has a unique solution  $w_x = \rho_C(x)$ .*

*Proof.* Let  $\{w_x : x \in C_T\}$  be any solution. Then,

$$\begin{aligned}
w_x &= \sum_{y \in C} P(x, y) + \sum_{y \in C_T} P(x, y)w_y \\
&= \sum_{y \in C} P(x, y) + \sum_{y \in C_T} P(x, y) \left[ \sum_{z \in C} P(y, z) + \sum_{y \in C_T} P(y, z)w_z \right] \\
&= \sum_{y \in C} P(x, y) + \sum_{y \in C_T} \sum_{z \in C_T} P(x, y)P(y, z)w_z + \sum_{y \in C_T} \sum_{z \in C} P(x, y)P(y, z) \\
&= \sum_{y \in C} P(x, y) + \sum_{z \in C_T} P^2(x, z)w_z + \sum_{y \in C_T} \sum_{z \in C} P(x, y)P(y, z) \\
&= P_x(T_C \leq 2) + \sum_{z \in C_T} P^2(x, z)w_z \\
&= \dots \\
&= P_x(T_c \leq n) + \sum_{z \in C_T} P^n(x, z)w_z
\end{aligned}$$

Now, we can take the limit as  $n$  goes to infinity:

$$\begin{aligned}
w_x &= \lim_{n \rightarrow \infty} \left( P_x(T_c \leq n) + \sum_{z \in C_T} P^n(x, z)w_z \right) \\
&= P_x(T_c \leq \infty) + \sum_{z \in C_T} \lim_{n \rightarrow \infty} P^n(x, z)w_z
\end{aligned}$$

Since  $C_T$  is finite,  $\lim_{n \rightarrow \infty} P^n(x, z) = 0$ , and therefore,  $w_x = P_x(T_c \leq \infty)$ .  $\square$

**Example 2.5.1.** Let  $\{X_n, n = 0, 1, 2, \dots\}$  be a Markov chain with  $S = \{1, 2, 3, 4\}$  and

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Let  $C = \{1\}$ . Find  $\rho_C(2), \rho_C(3)$

First, note that we can decompose the set as follows:

$$C_R = \{1, 4\}, C_T = \{2, 3\}$$

Since  $C_T$  is finite, we have

$$w_x = \sum_{y \in C} P(x, y) + \sum_{y \in C_T} P(x, y)w_y$$

Then, we have

$$\begin{aligned}
w_x &= P(3, 1) + P(2, 2)w_2 + P(2, 3)w_3 \\
&= \frac{1}{4} + \frac{3}{4}w_3
\end{aligned}$$

Similarly, we have

$$w_3 = \frac{1}{3}w_2 + \frac{1}{3}w_3$$

Therefore, we have

$$w_3 = \frac{1}{5}, w_2 = \frac{2}{5}$$

**Theorem 2.8.** *If for any  $x, y \in S$ ,  $x \leftrightarrow y$ , then the chain is irreducible. Then, it follows that a finite state of irreducible Markov Chain is recurrent.*

*Remark.* Infinite, irreducible Markov chain can be transient. Irreducibility doesn't imply recurrence.

Then, when will an infinite state, irreducible Markov chain be recurrent? We look at the birth-death Markov chain to understand this idea.

## 2.6 Birth-Death Markov Chain

**Definition 2.16.** *A Markov Chain  $\{X_n, n = 1, 2, \dots\}$  is called a birth-death Markov chain if*

1.  $S = \{0, 1, 2, \dots, d\}$  where  $d$  can be either finite or infinite. When  $d = \infty$ ,  $S = \{0, 1, 2, \dots\}$ .

$$2. P(x, y) = \begin{cases} p_x & y = x + 1 \\ q_x & y = x - 1 \\ r_x & y = x \\ 0 & \text{else} \end{cases}.$$

Note that if  $p_x > 0, q_x > 0$ , for  $1 \leq x \leq d - 1$  and  $p_0 > 0, q_d > 0$ , then the chain is irreducible. If the chain is irreducible and  $d < \infty$ , then the birth-death chain is recurrent.

**Theorem 2.9.** *For any  $a, b \in S$  and  $a < b$ . Let  $u(x) = P_x(T_a < T_b)$  for  $a \leq x \leq b$ ,  $u(a) = 1, u(b) = 0$ . Also, define*

$$\Gamma_0 = 1, \Gamma_k = \frac{q_1 q_2 \cdots q_k}{p_1 p_2 \cdots p_k}, k \geq 1$$

Then,

$$u(x) = \frac{\sum_{r=x}^{b-1} \Gamma_r}{\sum_{r=1}^{b-1} \Gamma_r}$$