STAT 516: Stochastic Modeling of Scientific Data **Autumn 2024**

Lecture 3: Discrete-Time Markov Chain – Part I

Instructor: Yen-Chi Chen

These notes are partially based on those of Mathias Drton.

3.1 Introduction

Before introducing Markov chain, we first talk about stochastic processes. A stochastic process is a family of RVs X_n that is indexed by n, where $n \in \mathcal{T}$. Note that sometimes people write X_t with $t \in \mathcal{T}$. The set $\mathcal{T}$ is called the index set. There are two common types of stochastic processes:

- $\mathcal{T}$ is discrete. For instance, $\mathcal{T} = \{0, 1, 2, 3, \cdots\}$. Then X_n is called a discrete-time stochastic process.
- $\mathcal{T}$ is continuous. For instance $\mathcal{T} = [0, \infty)$. Then X_n is called a continuous-time stochastic process.

Each random variable X_n can have a discrete, continuous, or mixed distribution. For example, in a queue X_n could represent the time that the n-th customer waits after arrival before receiving service, with a distribution that has an atom at zero but is otherwise continuous. Usually, each X_n takes its values in the same set, which is called the *state space* and denoted S. Therefore, $X_n \in S$.

We will focus on discrete time stochastic processes with a discrete state space in this course.

Let $\{X_n : n = 0, 1, 2, ...\}$ be a discrete time stochastic process and the state space S is discrete. The probability model of this process is determined by the joint distribution function:

$$P(X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n)$$

for all $n = 0, 1, ...$ and $x_0, x_1, ... \in S$.

In general, this joint distribution function can be arbitrary so it is very complex. We need some addition modeling on the joint distribution function to make it simpler enough that we can analyze. One particular example is the IID assumptions – $X_0, X_1, \cdots$ are IID. In this case, the join distribution function can be factorized into products of identical functions.

However, IID assumptions are often too strong for some scenarios. For instance, when modeling the genetic drift, the IID assumption on each generation is not a good model. Here are two examples where the IID assumptions do not work.

Example 1: Excess number of heads over tails in tossing a coin. Assume we toss a coin n times and let X_n denote the excess number of heads over tails. Clearly, $X_n \in \{\cdots, -2, -1, 0, 1, 2, \cdots\}$. Assume that each coin is tossed independently, then the process $\{X_n\}$ has a one-step memory such that

$$P(x_0, \cdots, x_n) = P(x_n|x_{n-1}) \times P(x_{n-1}|x_{n-2}) \times P(x_1|x_0) \times P(x_0),$$

where $P(x_0, \cdots, x_n) = P(X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n)$. Clearly, $P(x_0, x_1) \neq P(x_0)P(x_1)$ so the process is not independent although $P(x_0, x_2|x_1) = P(x_0|x_1)P(x_2|x_1)$.

Example 2: Precipitation level. Let the precipitation level of day n be X_n where $X_n = 0$ (dry) or 1 (wet). Assume that the precipitation level of a day given all the past history only depends on the precipitation level of the last two days. Namely,

$$P(x_n|x_{n-1}, x_{n-2}, \cdots, x_0) = P(x_n|x_{n-1}, x_{n-2}).$$

Then the joint distribution function is

$$P(x_0, \cdots, x_n) = P(x_n|x_{n-1}, x_{n-2}) \times P(x_{n-1}|x_{n-2}, x_{n-3}) \times P(x_2|x_1, x_0) \times P(x_1|x_0) \times P(x_0).$$

Again, it is clear that the RVs $X_0, X_1, \cdots, X_n$ are not IID. Although this model shows that the process has a two-step memory, we can reformulate it to a one-step memory process by defining a sequence of random vectors $Y_0, Y_1, \cdots, Y_n, \cdots$ such that $Y_n = (X_n, X_{n+1})$. Then

$$P(y_n|y_{n-1}, y_{n-2}, \cdots, y_0) = P(y_n|y_{n-1})$$

so the process $\{Y_n\}$ has a one-step memory.

The above two examples motivates us to study the process with a one-step memory. And such a stochastic process is known as the **Markov chain** .

3.2 Markov Chain

A discrete time stochastic process $\{X_n : n = 0, 1, 2, \cdots\}$ is called a *Markov chain* if for every $x_0, x_1, \cdots, x_{n-2}, i, j \in S$ and $n \geq 0$,

$$P(X_n = i|X_{n-1} = j, \cdots, X_0 = x_0) = P(X_n = i|X_{n-1} = j)$$

whenever both sides are well-defined. The Markov chain has a one-step memory.

If the distribution function $P(X_n = x_n|X_{n-1} = x_{n-1}) = p_{ij}$ is independent of n, we called $\{X_n\}$ a **homogeneous Markov chain**. Otherwise we called it an inhomogeneous Markov chain. For a homogeneous Markov chain,

$$\sum_{j \in S} p_{ij} = 1, \quad p_{ij} \geq 0$$

for every i, j. Note that sometimes people write $p_{ij} = p_{i \to j}$, where $p_{i \to j}$ stands for that the probability moving from state i to state j.

Because S is a discrete set, we often label it as $S = \{1, 2, 3, \cdots, s\}$ and the elements $\{p_{ij} : i, j = 1, \cdots, s\}$ forms an $s \times s$ matrix $\mathbf{P} = \{p_{ij}\}$. $\mathbf{P}$ is called the **transition (probability) matrix (t.p.m)**. The property of homogeneous Markov chain implies that

$$\mathbf{P} \geq 0, \quad \mathbf{P}\mathbf{1}_s = \mathbf{1}_s, \tag{3.1}$$

where $\mathbf{1}_s = (1, 1, 1, 1, \cdots, 1)^T$ is the vector of 1's. Note that any matrix satisfying equation (3.1) is called a stochastic matrix.

Example 3: SIS (Susceptible-Infected-Susceptible) model.

Suppose we observe an individual over a sequence of days $n = 1, 2, \ldots$ and classify this individual each day as

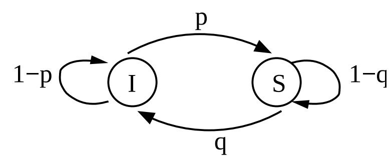
$$X_n = \begin{cases} I & \text{if infected} \\ S & \text{if susceptible.} \end{cases}$$

We would like to construct a stochastic model for the sequence $\{X_n : n = 1, 2, ...\}$. One possibility is to assume that the X_n's are independent with $P(X_n = I) = 1 - P(X_n = S) = \alpha$. However, this model is not very realistic since we know from experience that the individual is more likely to stay infected if he or she is already infected.

Since Markov chains are the simplest models that allow us to relax independence, we proceed by defining a transition probability matrix:

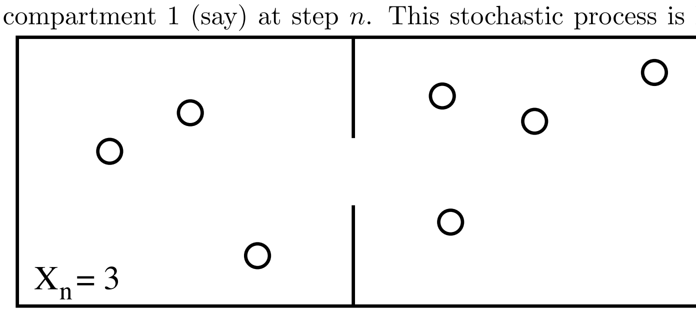
$$\mathbf{P} = \begin{array}{cc} \nearrow & \begin{array}{cc} I & S \end{array} \\ \begin{array}{c} I \\ S \end{array} & \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \end{array}$$

It can be helpful to visualize the transitions that are possible (have positive probability) by a *transition diagram*:



Example 4: Example: Ehrenfest Model of Diffusion.

We start with N particles in a closed box, divided into two compartments that are in contact with each other so that particles may move between compartments. At each time epoch, one particle is chosen uniformly at random and moved from its current compartment to the other compartment. Let X_n be the number of particles in compartment 1 (say) at step n. This stochastic process is Markov by construction.



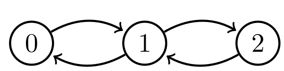
Transition probabilities of the Markov chain are:

$$p_{ij} = \begin{cases} \frac{i}{N}. & \text{for } j = i - 1, \\ 1 - \frac{i}{N}, & \text{for } j = i + 1, \\ 0, & \text{otherwise.} \end{cases}$$

The probability of transfer depends on the number of particles in each compartment. For $N = 2$ we have states 0, 1, 2 and t.p.m.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

and the transition diagrm



Example 5: Snoqualmie Falls Precipitation.

There is a data on the precipitation (in inches), recorded by the UW Weather Service, at Sonqualmie Falls in the years 1948–1983. We examine the data for January only and consider dry (=0) and wet (=1) only. If we condition on the state on January 1st we obtain the frequencies of the four different transitions as:

	0	1	Total
0	186	123	309
1	128	643	771
Total	(314)	(766)	1080

For example, there were 123 occasions on which a wet day followed a dry day.

From the table of frequencies we can compute the relative frequencies of transitions:

$$\widehat{\mathbf{P}} = \begin{bmatrix} 0.602 & 0.398 \\ 0.166 & 0.834 \end{bmatrix}$$

This is an estimate (hence the hat!) of the t.p.m.

3.3 Property of Markov chain

Here are some important properties of a Markov chain.

Property: joint probability. Suppose we observe a finite realization of the discrete Markov chain and want to compute the probability of this random event:

$$\begin{aligned}
&P(X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) \\
&= P(X_n = i_n | X_{n-1} = i_{n-1}, \cdots, X_0 = i_0)\, P(X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) \\
&= p_{i_{n-1}, i_n} \times P(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \times P(X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \\
&= \cdots \\
&= p_{0 i_0} p_{i_0, i_1} p_{i_1, i_2} \cdots p_{i_{n-2}, i_{n-1}} p_{i_{n-1}, i_n}
\end{aligned}$$

where $p_0 = (p_{01}, p_{02}, \ldots)^T$ is the distribution of X_0, called the *initial distribution* of $\{X_n\}$. Thus, every Markov chain is fully specified by its transition probability matrix $\mathbf{P}$ and initial distribution p_0.

Property: Markov property. The Markov chain has a powerful property called the *Markov property* – the distribution of X_{m+n} given a set of previous states depends only on the latest available state. Assume that we observe a Markov chain from $n = 0, 1, \cdots, n$ and we are analyzing the distribution of X_{m+n}. Then

$$P(X_{m+n} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{m+n} = j | X_n = i). \tag{3.2}$$

To give an intuition about how we obtain the Markov property, consider a simple case where $n = 1$ and $m = 2$.

$$\begin{aligned}
P(X_3 = i_3 | X_1 = i_1, X_0 = i_0) &= \sum_{i_2} P(X_3 = i_3, X_2 = i_2 | X_1 = i_1, X_0 = i_0) \\
&= \sum_{i_2} P(X_3 = i_3 | X_2 = i_2, X_1 = i_1, X_0 = i_0) P(X_2 = i_2 | X_1 = i_1, X_0 = i_0) \\
&\overset{!}{=} \sum_{i_2} P(X_3 = i_3 | X_2 = i_2, X_1 = i_1) P(X_2 = i_2 | X_1 = i_1) \\
&= \sum_{i_2} P(X_3 = i_3, X_2 = i_2 | X_1 = i_1) \\
&= P(X_3 = i_3 | X_1 = i_1).
\end{aligned}$$

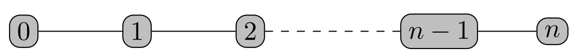
To argue the equality '$\stackrel{!}{=}$', observe that

$$P(X_3 = i_3 | X_2 = i_2, X_1 = i_1, X_0 = i_0) = P(X_3 = i_3 | X_2 = i_2).$$

But we also have that

$$\begin{aligned} P(X_3 = i_3 | X_2 = i_1, X_1 = i_1) &= \sum_{i_0} P(X_3 = i_3, X_0 = i_0 | X_2 = i_2, X_1 = i_1) \\ &= \sum_{i_0} P(X_3 = i_3 | X_2 = i_2, X_1 = i_1, X_0 = i_0) P(X_0 = i_0 | X_2 = i_2, X_1 = i_1) \\ &= P(X_3 = i_3 | X_2 = i_2) \sum_{i_0} P(X_0 = i_0 | X_2 = i_2, X_1 = i_1) \\ &= P(X_3 = i_3 | X_2 = i_2). \end{aligned}$$

Property: conditional independence. We can represent the Markov chain using a simple graphical model:



The claim of the Markov property is now obvious from the theorem on conditional independence and graphical factorization. Indeed, the latest available state serves as a separating set.

Using the graph representation, we obtain an interesting property about a Markov chain: *the past and the future are independent given the present.*

To see this, again we consider a simple case where $n = 2$ and we have X_0, X_1, X_2. Here X_0 denotes the past, X_1 denotes the present, and X_2 denotes the future. Then

$$\begin{aligned} P(X_0 = i_0, X_2 = i_2 | X_1 = i_1) &= \frac{P(X_0 = i_0, X_1 = i_1, X_2 = i_2)}{P(X_1 = i_1)} \\ &= \frac{P(X_2 = i_2 | X_1 = i_1, X_0 = i_0) P(X_1 = i_1, X_0 = i_0)}{P(X_1 = i_1)} \\ &= P(X_2 = i_2 | X_1 = i_1) \frac{P(X_1 = i_1, X_0 = i_0)}{P(X_1 = i_1)} \\ &= P(X_2 = i_2 | X_1 = i_1) P(X_0 = i_0 | X_1 = i_1) \end{aligned}$$

for any $i_0, i_1, i_2 \in S$. Namely, X_0 and X_2 are conditional independent given X_1.

3.4 n-step Transition Probability and Chapman-Kolmogorov Equation

For a Markov chain, we define the *n-step transition probability* as

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

The n-step transition probability is time invariant.

Lemma 3.1 *Let $\{X_n\}$ be a homogeneous Markov chain and let $p_{ij}^{(n)}$ be the n-step transition probability. Then for any $k = 0, 1, 2, \cdots$,*

$$P(X_{n+k} = j | X_k = i) = p_{ij}^{(n)}.$$

Proof:

$$\begin{aligned} P(X_{n+k} = j | X_k = i) &= \sum_{i_{k+1}, i_{k+2}, \cdots, i_{n+k-1}} P(X_{n+k} = j | X_{n+k-1} = i_{n+k-1}) \times \cdots \times P(X_{k+1} = i_{k+1} | X_k = i) \\ &= \sum_{i_{k+1}, i_{k+2}, \cdots, i_{n+k-1}} P(X_n = j | X_{n-1} = i_{n+k-1}) \times \cdots \times P(X_1 = i_{k+1} | X_0 = i) \\ &= P(X_n = j | X_0 = i) = p_{ij}^{(n)}. \end{aligned}$$

■

The n-step transition probabilities are related to each other via the famous *Chapman-Kolmogorov Equation.*

Lemma 3.2 *Let $\{X_n\}$ be a homogeneous Markov chain and let $p_{ij}^{(n)}$ be the n-step transition probability. Then for any $n, m = 0, 1, 2, \cdots$*

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}. \tag{3.3}$$

Proof:

$$\begin{aligned} p_{ij}^{(n+m)} &= P(X_{n+m} = j | X_0 = i) \\ &= \sum_{k \in S} P(X_{n+m} = j, X_m = k | X_0 = i) \\ &= \sum_{k \in S} P(X_{n+m} = j | X_m = k, X_0 = i) P(X_m = k | X_0 = i) \\ &= \sum_{k \in S} P(X_{n+m} = j | X_m = k) P(X_m = k | X_0 = i) \quad \text{(Markov property)} \\ &= \sum_{k \in S} P(X_n = j | X_0 = k) P(X_m = k | X_0 = i) \quad \text{(time-invariant property)} \\ &= \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}. \end{aligned}$$

■

The Chapman-Kolmogorov Equation (equation (3.3)) also implies

$$\begin{aligned} \textit{Forward equation} : p_{ij}^{(n+1)} &= \sum_k p_{ik}^{(n)} p_{kj}, \text{ for } n = 1, 2, \ldots \text{ and} \\ \textit{Backward equation} : p_{ij}^{(n+1)} &= \sum_k p_{ik} p_{kj}^{(n)}, \text{ for } n = 1, 2, \ldots. \end{aligned}$$

The forward equation singles out the final step and has the initial state i fixed. The equation is most useful when interest centers on the $p_{ij}^{(n)}$'s for a particular i but all values of j. Conversely, the backward equation singles out the change from the initial state i and has the final state j fixed. This equation is useful when interest is in the $p_{ij}^{(n)}$'s for a particular j but all values of i. The backward equation can be interesting, in particular, when there is an absorbing state j from which there is no escape ($p_{jj} = 1$).

If we collect the n-step transition probabilities into the matrix $\mathbf{P}^{(n)} = \{p_{ij}^{(n)}\}$, then Kolmogorov's forward and backward equations can be rewritten in matrix form as

$$\mathbf{P}^{(n+1)} = \mathbf{P}^{(n)}\mathbf{P} = \mathbf{P}\mathbf{P}^{(n)},$$

where $\mathbf{P}^{(1)} = \mathbf{P}$. Therefore, $\mathbf{P}^{(n)} = \mathbf{P}^n$.

This matrix form also implies a cool property about the marginal distribution of X_n. Assume that X_0 has a distribution $p_0 = (p_{01}, p_{02}, \cdots, p_{0s})^T$. Let $p_n = (p_{n1}, \cdots, p_{ns})^T$ be the marginal distribution of X_n, i.e., $p_{nj} = P(X_n = j)$. Then

$$\begin{aligned} p_{nj} = P(X_n = j) &= \sum_i P(X_n = j, X_0 = i) \\ &= \sum_i P(X_n = j | X_0 = i) P(X_0 = i) \\ &= \sum_i p_{0i} p_{ij}^{(n)}. \end{aligned}$$

Using the matrix form, we obtain

$$p_n^T = p_0^T \mathbf{P}^n.$$

Example 3: SIS model (revisited).

Recalled that SIS model has a transition probability

$$\mathbf{P} = \begin{array}{c} \nearrow \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \end{array}$$

Note that we use $\{0, 1\}$ to denote the state I and S in the SIS model.

Assume that the initial distribution $p_0 = (1-\alpha, \alpha)$, i.e., $P(X_0 = 0) = 1-\alpha$. Moreover, assume that $\beta = 1-\alpha$ so the distribution of X_1 will be

$$\begin{aligned} p_1^T &= p_0^T \mathbf{P} = (1-\alpha, \alpha) \begin{bmatrix} 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{bmatrix} \\ &= [(1-\alpha)^2 + \alpha(1-\alpha), \alpha(1-\alpha) + \alpha^2] = (1-\alpha, \alpha) = p_0^T. \end{aligned}$$

What will be the distribution of X_n? Using the matrix form, we know that

$$p_n^T = p_0^T \mathbf{P}^n = p_1^T \mathbf{P}^{n-1} = p_0^T \mathbf{P}^{n-1} = \cdots = p_0^T.$$

Therefore, $P(X_n = 0) = 1-\alpha$ and $P(X_n = 1) = \alpha$ for all $n = 1, 2, 3, \cdots$.

A more interesting fact is the joint distribution of $X_0, X_1, \cdots, X_n$:

$$\begin{aligned} P(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n) &= p_{0i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \\ &= \alpha^{i_0}(1-\alpha)^{1-i_0} \alpha^{i_1}(1-\alpha)^{1-i_1} \alpha^{i_2}(1-\alpha)^{1-i_2} \cdots \alpha^{i_n}(1-\alpha)^{1-i_n}, \end{aligned}$$

which is the joint PMF of IID random Bernoulli variables with parameter α. Therefore, under this special case, the Markov chain reduces to IID Bernoulli RVs.

Note that in general, when the rows of t.p.m are the same, the corresponding Markov chain is a sequence of IID RVs whose distribution is given by the first/any row of the t.p.m.

Example 5: Snoqualmie Falls Precipitation (revisited).

In the Snoqualmie Falls Precipitation problem, we have a t.p.m

$$\widehat{\mathbf{P}} = \begin{bmatrix} 0.602 & 0.398 \\ 0.166 & 0.834 \end{bmatrix}.$$

If we consider the 2-step transition probability,

$$\widehat{\mathbf{P}}^2 = \begin{bmatrix} 0.428 & 0.572 \\ 0.238 & 0.762 \end{bmatrix}.$$

When we consider n-step transition probability with n large (in this case, $n \geq 10$), it turns out that the n-step transition probability matrix becomes

$$\widehat{\mathbf{P}}^n = \begin{bmatrix} 0.294 & 0.706 \\ 0.294 & 0.706 \end{bmatrix}.$$

This implies that the initial distribution is uninformative – whether it is dry or wet on Jan 18th tells us little about Jan 27th.

Recall that in the previous example, we saw that when the rows of a t.p.m. are the same, the corresponding random variables are IID. Therefore, if we consider another sequence of RVs $\{Y_0(n) = X_k, Y_1(n) = X_{k+n}, Y_2(n) = X_{k+2n}, \cdots\}$, then $Y_0(n), Y_1(n), Y_2(n), \cdots$ are IID when $n \to \infty$.

After seeing this example, one may conjecture that if the limit $\mathbf{P}_\infty = \lim_{n\to\infty} \mathbf{P}^n$ will have equal rows. However, this is not always true. A counterexample is

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

3.5 First Step Analysis of Markov Chain

First-step analysis is a general strategy for solving many Markov chain problems by conditioning on the first step of the Markov chain. We demonstrate this technique on a simple example – the Gambler's ruin problem.

Gambler's ruin problem: Two players bet one dollar in each round. Player 1 wins with probability α and loses with probability $\beta = 1 - \alpha$. We assume that player 1 starts with a dollars and player 2 starts with b dollars. Let X_n be the fortune of player 1 after n rounds. X_n can take values from 0 to $a + b$:

$$p_{ij} = P(X_{n+1} = j | X_n = i) = \begin{cases} \alpha & \text{if } j = i + 1, \\ \beta & \text{if } j = i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let T be the time of rounds when one of the players loses all his/her money. Because of the randomness of this model, T is also a random variable. We are interested in the probability that player 1 wins the game, which occurs when $X_T = a + b$. Apparently, this probability depends on the initial amount of money that player 1 has so we will denote it as

$$u(a) = P(X_T = a + b | X_0 = a).$$

Note that $u(0) = 0$ and $u(a+b) = 1$. We may view $u(a)$ as the probability that the chain is absorbed into the state $a+b$ at the hitting time T when the chain starts at $X_0 = a$.

First step analysis proceeds as follows:

$$\begin{aligned}
u(a) &= P(X_T = a+b | X_0 = a) \\
&= \sum_{j=1}^{a+b} P(X_T = a+b, X_1 = j | X_0 = a) \\
&= \sum_{j=1}^{a+b} P(X_T = a+b | X_1 = j, X_0 = a) P(X_1 = j | X_0 = a) \\
&= \sum_{j=1}^{a+b} P(X_T = a+b | X_1 = j) P(X_1 = j | X_0 = a) \quad \text{(Markov property)} \\
&= P(X_T = a+b | X_1 = a+1) P(X_1 = a+1 | X_0 = a) + \\
&\quad\; P(X_T = a+b | X_1 = a-1) P(X_1 = a-1 | X_0 = a) \\
&= u(a+1)\alpha + u(a-1)\beta.
\end{aligned}$$

Therefore, we have $u(a) = u(a+1)\alpha + u(a-1)\beta$ with two boundary conditions: $u(0) = 0, u(a+b) = 1$. Because $\alpha + \beta = 1$,

$$(\alpha + \beta) u(a) = u(a) = u(a+1)\alpha + u(a-1)\beta$$

which implies

$$\alpha(u(a+1) - u(a)) = \beta(u(a) - u(a-1)).$$

Define $v(a) = u(a) - u(a-1)$. Then the above leads to

$$\alpha v(a+1) = \beta v(a) \Rightarrow v(a+1) = \frac{\beta}{\alpha} v(a).$$

By telescoping,

$$v(a+1) = \frac{\beta}{\alpha} v(a) = \left(\frac{\beta}{\alpha}\right)^2 v(a-1) = \cdots = \left(\frac{\beta}{\alpha}\right)^a v(1).$$

Using the boundary condition $u(0) = 0$,

$$\begin{aligned}
u(a) &= u(a) - u(0) \\
&= \sum_{j=1}^{a} [u(j) - u(j-1)] \\
&= \sum_{j=1}^{a} v(j) \\
&= v(1) \sum_{j=0}^{a-1} \left(\frac{\beta}{\alpha}\right)^j \\
&= \begin{cases} v(1) \times a & \text{if } \alpha = \beta \\ v(1) \times \left(\frac{1 - \left(\frac{\beta}{\alpha}\right)^a}{1 - \left(\frac{\beta}{\alpha}\right)} \right) & \text{if } \alpha \neq \beta \end{cases}.
\end{aligned}$$

To find out $v(1)$, we use the other boundary condition

$$
\begin{aligned}
1 = u(a+b) &= u(a+b) - u(0) \\
&= \sum_{j=1}^{a+b} [u(j) - u(j-1)] \\
&= \sum_{j=1}^{a+b} v(j) \\
&= \begin{cases} v(1) \times (a+b) & \text{if } \alpha = \beta \\ v(1) \times \left(\frac{1-\left(\frac{\beta}{\alpha}\right)^{a+b}}{1-\left(\frac{\beta}{\alpha}\right)} \right) & \text{if } \alpha \neq \beta \end{cases}.
\end{aligned}
$$

Therefore,

$$
v(1) = \begin{cases} \frac{1}{a+b} & \text{if } \alpha = \beta \\ \frac{1-\frac{\beta}{\alpha}}{1-\left(\frac{\beta}{\alpha}\right)^{a+b}} & \text{if } \alpha \neq \beta. \end{cases}
$$

and

$$
u(a) = \begin{cases} \frac{a}{a+b} & \text{if } \alpha = \beta \\ \frac{1-\left(\frac{\beta}{\alpha}\right)^{a}}{1-\left(\frac{\beta}{\alpha}\right)^{a+b}} & \text{if } \alpha \neq \beta. \end{cases}
$$

3.6 Classification of States

We now turn to a classification of the states of a Markov chain that is crucial to understanding the behavior of Markov chains.

An *equivalence relation* "$\sim$" is a binary relation between elements of a set satisfying

1. Reflexivity: $i \sim i$ for all i
2. Symmetry: $i \sim j \Rightarrow j \sim i$
3. Transitivity: $i \sim j$, $j \sim k \Rightarrow i \sim k$.

For a set $\mathcal{S}$ and $a \in \mathcal{S}$, $\{s \in \mathcal{S} : s \sim a\}$ is called an *equivalence class*. Equivalence relations will allow us to split Markov chain state spaces into equivalence classes.

State j is *accessible* from state i ($i \to j$) if there exists $m \geq 0$ such that $p_{ij}^{(m)} > 0$. We say that i *communicates* with j ($i \leftrightarrow j$) if j is accessible from i and i is accessible from j. A set of states $\mathcal{C}$ is a **communicating class** if every pair of states in $\mathcal{C}$ communicates with each other, and no state in $\mathcal{C}$ communicates with any state not in $\mathcal{C}$.

Proposition 3.3 *Communication of states is an equivalence relation.*

Proof: Reflexivity and symmetry are clear. To prove transitivity, let $i \leftrightarrow j$ and $j \leftrightarrow k$. We then want to show that $i \leftrightarrow k$.

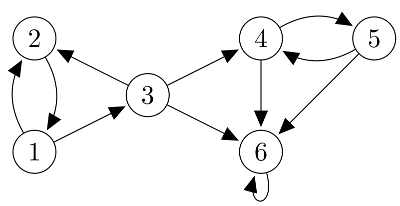
Note that $i \to j$ if and only if the transition diagram contains a path from i to j. So there is a path from i to j and from j to k. Concatenate the two to obtain a path from i to k, which testifies to the fact that $i \to k$.

Analogously, we have $k \to i$ and conclude that $i \leftrightarrow k$.

■

A set of states $\mathcal{C}$ is *closed* if $\sum_{j \in \mathcal{C}} p_{ij} = 1$ for all $i \in \mathcal{C}$.

Example 6. Consider a Markov chain with the following transition diagram:



Then $\{1,2,3\}$, $\{4,5\}$, and $\{6\}$ are the communication classes.

A Markov chain $\{X_n\}$ is called **irreducible** if it has only one communication class, i.e., for all i and j , $i \leftrightarrow j$. For state i, $d_i = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$ is called its **period**, where gcd = greatest common divisor and $d_i = +\infty$ if $p_{ii}^{(n)} = 0$ for all $n \geq 1$.

Example 7. Consider the example with state space $S = \{0, 1, 2, ...\}$ and X_n such that

$$P(X_{n+1} = i | X_n = 0) = \begin{cases} p & \text{if } i = 1, \\ 1-p & \text{if } i = 0, \\ 0 & \text{otherwise.} \end{cases}$$

and for $j \neq 0$

$$P(X_{n+1} = i | X_n = j) = \begin{cases} p & \text{if } i = j+1, \\ 1-p & \text{if } i = j-1, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $d_2 = \gcd\{2, 4, 5, 6, ...\} = 1$ though 1 is not in the list (think about why $p_{22}^{(5)} > 0$).

Example 8: Simple (1-D) Random Walk on the Integers. Consider another example with state space $\mathbb{Z}$. Let X_n be the position at time n. Then

$$P(X_{n+1} = i - 1 | X_n = i) = q \text{ and } P(X_{n+1} = i + 1 | X_n = i) = p$$

with $p = 1 - q$. Suppose we start at 0, then it is clear that we cannot return to 0 after an odd number of steps, so $p_{00}^{(2n+1)} = 0$ for all n, i.e.

$$d_0 = \gcd\{n \geq 1 : p_{00}^{(n)} > 0\} = \gcd\{2, 4, 6, \dots\} = 2.$$

Proposition 3.4 *Period is a communication class property. Namely, $i \leftrightarrow j \Rightarrow d_i = d_j$.*

Proof: $i \leftrightarrow j \Rightarrow$ there exists n_1, n_2 such that $p_{ij}^{(n_1)} > 0$ and $p_{ji}^{(n_2)} > 0$. Then, by Chapman-Kolmogorov:

$$p_{ii}^{(n_1+n_2)} = \sum_k p_{ik}^{(n_1)} p_{ki}^{(n_2)} \geq p_{ij}^{(n_1)} p_{ji}^{(n_2)} > 0.$$

Consequently we know that $d_i | n_1 + n_2$. For example, suppose $n_1 = 3$ and $n_2 = 5$, then $n_1 + n_2 = 3 + 5 = 8$. Therefore we know that $d_i \leq 8$, i.e. we could have $d_i = 8, 4, 2, 1$; we know we could return after 8 time steps, but it could be less.

Note: $a|b$ means a divides b, i.e. there is an integer c s.t. $b = ac$.

Now, take any n such that $p_{jj}^{(n)} > 0$. Then

$$\begin{aligned} p_{ii}^{(n_1+n_2+n)} &= \sum_k p_{ik}^{(n+n_1)} p_{ki}^{(n_2)} \geq p_{ij}^{(n+n_1)} p_{ji}^{(n_2)} \\ &= \left[\sum_k p_{ik}^{(n_1)} p_{kj}^{(n)} \right] p_{ji}^{(n_2)} \geq p_{ij}^{(n_1)} p_{jj}^{(n)} p_{ji}^{(n_2)} > 0. \end{aligned}$$

Hence, $d_i | n_1 + n_2 + n$.

Together, $n_1 + n_2 = c_1 d_i$ and $n_1 + n_2 + n = c_2 d_i$ imply that $n = (c_2 - c_1) d_i$ and as a result, $d_i | n$ for all n such that $p_{jj}^{(n)} > 0$.

Since d_i is a divisor of the set $\{n : p_{jj}^{(n)} > 0\}$ and d_j is the greatest common divisor of the same set (by definition of period), $d_i \leq d_j$.

By symmetry, $d_j \leq d_i \Rightarrow d_i = d_j$.

■

In a communication class, all states have the same period. Since all states communicate in an irreducible Markov chains, it makes sense to define the **period** of such a Markov chain. If $d_i = 1$, state i is called aperiodic. An irreducible Markov chain with period 1 is also called **aperiodic**.

Theorem 3.5 (Lattice Theorem (Brémaud p.75)) *Suppose d is the period of an irreducible Markov chain. Then for all states i, j there exists $m \in \{0, \ldots, d-1\}$ and $k \geq 0$ such that*

$$p_{ij}^{(m+nd)} > 0, \quad \forall n \geq k.$$

Theorem 3.6 (Cyclic Classes) *For any irreducible Markov chain one can find a unique partition of S into d classes $C_0, C_1, ..., C_{d-1}$ such that for all k, and for $i \in C_k$,*

$$\sum_{j \in C_{k+1}} p_{ij} = 1,$$

where, by convention $C_d = C_0$ and where d is maximal (that is, there is no other such partition $C'_0, C'_1, ..., C'_{d'-1}$ with $d' > d$).

Proof: Fix a state i and classify states j by the value of m in Lattice Theorem. ■

The number $d \geq 1$ is the *period* of the chain. The classes $C_0, C_1, ..., C_{d-1}$ are called the *cyclic classes.*

Example 8: Simple (1-D) Random Walk on the Integers (revisited). Random walk on $S = \mathbb{Z} = C_0 + C_1$ where C_0 and C_1 are the sets of even and odd integers.

3.7 Strong Markov Property

The Markov property states that the random variable at time $n+m$ conditional on its behavior at time n is independent of the at time prior to n. However, what if the time n is random?

Say we are interested in the behavior of X_{T+m} given X_T, where T is the first time that the Markov chain hits the state 0. Do we still have the Markov property?

Some random time does not have the Markov property. Recall that the Markov property states that for any $m < n < k$, $X_m \perp X_k | X_n$ for *non-random* m, n, k. Let $\{X_n\}$ be a Markov chain with state space $S = \{1, 2, 3\}$ and consider a random time

$$T = \inf\{n \geq 1 : (X_{n-1}, X_n, X_{n+1}) = (2, 1, 3) \text{ or } (3, 1, 2)\}.$$

To see why Markov property does not work for random m, n, k, consider $m = T-1, n = T, k = T+1$. Then the probability

$$P(X_k = 3|X_n = 1, X_m = 2) = P(X_{T+1} = 3|X_T = 1, X_{T-1} = 2) = 1 \neq P(X_k = 3|X_n = 1)$$

because when $X_m = X_{n-1} = 3$, $P(X_k = 2|X_n = 1, X_m = 3) = P(X_{T+1} = 2|X_T = 1, X_{T-1} = 3) = 1$ so $P(X_k = 3|X_n = 1, X_m = 3) = 0$. Thus, the conditional probability of X_k given X_n depends on X_m, which is a violation of Markov property.

Therefore, it is crucial to identify a class of random time such that the Markov property holds. It turns out that there is a simple class of random times that has the Markov property. This class is called the stopping time.

A random variable $\tau \in \{1, 2, 3, \cdots\} \cup \{\infty\}$ is called a **stopping time** if the event $\{\tau = m\}$ can be expressed in terms of $X_0, X_1, \cdots, X_m$. Intuitively, a stopping time is a random time such that *we can observe it when the time arrives.*

Examples 9: Stopping times.

- Return time. Let $T_i = \inf\{n \geq 1 : X_n = i\}$ is a stopping time because $\{T_i = m\} = \{X_1 \neq i, \cdots, X_{m-1} \neq i, X_m = i\}$. T_i is interpreted as the first time the chain returns to state i.
- Successive Returns. Let τ_k be the time of the k-th return to state i (note that $\tau_1 = T_i$). Then τ_k is a stopping time because
$$\{\tau_k = m\} = \left\{\sum_{n=1}^{m} I(X_n = i) = k, X_m = i\right\}.$$
- Counterexample – non-stopping time: Let $\tau = \inf\{n \geq 1 : X_{n+1} = i\}$ is not a stopping time because when the time arrives at m, $\{\tau = m\} = \{X_1 \neq i, \cdots, X_m \neq i, X_{m+1} = i\}$ depends on X_{m+1}.

Stopping time is a very important class of random variable in statistics. Many statistical procedure involves a stopping time. For instance, if we are performing a sequence of experiments and we will stop when we observe certain behavior such as a high signal or enough anomaly. Then the time (of related to the number of sample) is a stopping time. If we want to use data from this sequence of experiments, then we need to use theorems of stopping time (such as optional sampling theorem).

Theorem 3.7 (Strong Markov Property) *Let $\{X_n\}$ be a homogeneous Markov chain with a transition probability matrix $\mathbf{P} = \{p_{ij}\}$ and let τ be a stopping time with respect to $\{X_n\}$. Then for any integer k,*

$$P(X_{\tau+k} = j|X_\tau = i, X_\ell = i_\ell, 0 \leq \ell < \tau) = P(X_k = j|X_0 = i) = p_{ij}^{(k)}$$

and

$$P(X_{\tau+k} = j | X_\tau = i) = P(X_k = j | X_0 = i) = p_{ij}^{(k)}.$$

Proof: We first prove the first equality.

$$\begin{aligned} P(X_{\tau+k} = j | X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell) &= \frac{P(X_{\tau+k} = j, X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)}{P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)} \\ &= \frac{\sum_{r=1}^{\infty} P(X_{\tau+k} = j, X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r)}{P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)}. \end{aligned} \tag{3.4}$$

Now, because τ is a stopping time, the event $\{\tau = r\}$ can be expressed as a function of $X_0, \cdots, X_r$ so the Markov property implies

$$P(X_{\tau+k} = j | X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r) = P(X_{r+k} = j | X_r = i) = p_{ij}^{(k)}.$$

Therefore, equation (3.4) becomes

$$\begin{aligned} &P(X_{\tau+k} = j | X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell) \\ &\quad = \frac{\sum_{r=1}^{\infty} P(X_{\tau+k} = j, X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r)}{P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)} \\ &\quad = \frac{\sum_{r=1}^{\infty} P(X_{\tau+k} = j | X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r) P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r)}{P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)} \\ &\quad = \frac{\sum_{r=1}^{\infty} p_{ij}^{(k)} P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r)}{P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)} \\ &\quad = p_{ij}^{(k)} \frac{\sum_{r=1}^{\infty} P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell, \tau = r)}{P(X_\tau = i, 0 \le \ell < \tau, X_\ell = i_\ell)} \\ &\quad = p_{ij}^{(k)}. \end{aligned}$$

The second equality follows simply from the first equality:

$$\begin{aligned} P(X_{\tau+k} = j | X_\tau = i) &= \frac{\sum_{r=1}^{\infty} P(X_{\tau+k} = j | X_r = i, \tau = r) P(X_r = i, \tau = r)}{P(X_\tau = i)} \\ &= p_{ij}^{(k)} \frac{\sum_{r=1}^{\infty} P(X_r = i, \tau = r)}{P(X_\tau = i)} \\ &= p_{ij}^{(k)}. \end{aligned}$$

■

3.8 Stationary distribution

It is often of great interest to study the limiting behavior of a Markov chain X_n when $n \to \infty$. Here, for simplicity, we assume that our Markov chain is homogeneous. A property of limiting behavior is that X_n and X_{n+1} should have the same distribution when n is large. So we are interested in understanding if a Markov chain will eventually converge to a 'stable' distribution (formally, we will call it a *stationary distribution*). In particular, we would like to know *given a Markov chain,*

- does this chain has a stationary distribution?

- if so, what is the stationary distribution?
- and does this stationary distribution unique?

It turns out that to answer these questions, we will use concepts related to return time. Thus, we start with understanding properties about return time.

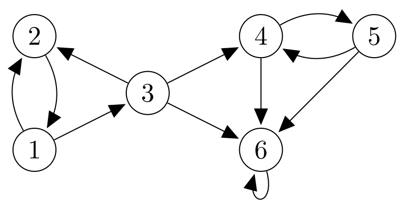
3.8.1 Return Times

Let $N_i = \sum_{n=1}^{\infty} I(X_n = i)$ denotes the number of visits of $\{X_n\}$ to state i not counting the initial state. We also define the following notations:

$$P(\cdot|X_0 = i) = P_i(\cdot), \quad \mathbb{E}(\cdot|X_0 = i) = \mathbb{E}_i(\cdot).$$

Note that the quantity N_i may equal to ∞. It is a finite number with a non-zero probability if there are some states such that when the chain enters one of them, the chain never go back to state i. Later we will describe this phenomena using the concept of transient states and recurrent states.

Example 6 (revisited). Consider a Markov chain with the following transition diagram:



As can be seen easily, when the Markov chain enters states $\{4, 5, 6\}$, it never comes back to any of $\{1, 2, 3\}$. Thus, N_1 takes a non-trivial probability to be a finite number.

Let $T_i = \inf\{n \geq 1 : X_n = i\}$ be the return time. Then the following events can be defined using either T_i or N_i:

$$\{T_i = \infty\} = \{N_i = 0\}, \quad \{T_i < \infty\} = \{N_i > 0\}.$$

These are useful later.

We then define $f_{ji} = P_j(T_i < \infty) = P_j(N_i > 0)$ to be the probability of reaching state i in a finite number of time when the chain starts at state j. Note that because $P_j(T_i = \infty) + P_j(T_i < \infty) = 1$, we have $f_{ii} = P_i(T_i < \infty)$ and $P_i(T_i = \infty) = 1 - f_{ii}$.

Proposition 3.8

$$P_j(N_i = r) = \begin{cases} f_{ji} f_{ii}^{r-1}(1 - f_{ii}) & \textit{if } r \geq 1 \\ 1 - f_{ji} & \textit{if } r = 0 \end{cases}.$$

Proof: The case $r = 0$ is very simple because $\{N_i = 0\} = \{T_i = \infty\}$. Thus, $P_j(N_i = 0) = P_j(T_i = \infty) = 1 - P_j(T_i < \infty) = 1 - f_{ji}$.

For the rest of cases, we will do a proof by induction. Before doing that, we first investigate the case $P_j(N_i = r)$ for $r > 0$. Let τ_r be the r-th return time. Note that the event $\{X_{\tau_r} = i\} = \{N_i \geq r\}$.

Then

$$
\begin{aligned}
P_j(N_i = r) &= P_j(N_i = r, X_{\tau_r} = i) \\
&= P_j(N_i = r | X_{\tau_r} = i) P_j(X_{\tau_r} = i) \\
&= P_j\left(\sum_{t=\tau_r+1}^{\infty} I(X_t = i) = 0 \mid X_{\tau_r} = i\right) P_j(X_{\tau_r} = i) \\
&= P_i\left(\sum_{t=1}^{\infty} I(X_t = i) = 0 \mid X_0 = i\right) P_j(X_{\tau_r} = i) \quad \text{(Strong Markov property)} \\
&= P_i(N_i = 0) P_j(N_i \geq r) \\
&= P_i(T_i = \infty) P_j(N_i \geq r).
\end{aligned}
$$

Therefore, we conclude

$$P_j(N_i = r) = P_i(T_i = \infty) P_j(N_i \geq r) = (1 - f_{ii}) P_j(N_i \geq r).$$

To start with the proof by induction, consider $r = 1$. $P_j(N_i \geq 1) = 1 - P_j(N_i = 0) = f_{ji}$ so $P_j(N_i = 1) = (1 - f_{ii}) f_{ji}$, which agrees with what we need for $r = 1$.

Assume that it works for $r \leq k$. Now we show that it works for $r = k + 1$. Note that this means that we have

$$P_j(N_i = r) = \begin{cases} f_{ji} f_{ii}^{r-1}(1 - f_{ii}) & \text{if } r = 1, \cdots k \\ 1 - f_{ji} & \text{if } r = 0 \end{cases}.$$

For the case of $r = k + 1$, we use the fact that

$$P_j(N_i = k + 1) = (1 - f_{ii}) P_j(N_i \geq k + 1),$$

so all we need is the probability $P_j(N_i \geq k + 1)$.

This quantity can be easily calculated via

$$
\begin{aligned}
P_j(N_i \geq k + 1) = 1 - P_j(N_i \leq k)& \\
&= 1 - (1 - f_{ji}) - \sum_{r=1}^{k} f_{ji} f_{ii}^{r-1}(1 - f_{ii}) \\
&= f_{ji} - f_{ji}(1 - f_{ii})(1 + f_{ii} + f_{ii}^2 + \cdots + f_{ii}^{k-1}) \\
&= f_{ji} - f_{ji}(1 - f_{ii}) \frac{1 - f_{ii}^k}{1 - f_{ii}} \\
&= f_{ji} f_{ii}^k.
\end{aligned}
$$

Thus,

$$P_j(N_i = k + 1) = (1 - f_{ii}) P_j(N_i \geq k + 1 = f_{ji} f_{ii}^k (1 - fii)$$

which is the formula for $r = k + 1$. Thus, by induction, the result holds.

■

The above formula also gives an interesting result on the case of 'starting from state i, returning to state i' when we set $j = i$:

$$P_i(N_i = r) = f_{ii}^r (1 - f_{ii}), \quad P_i(N_r > r) = f_{ii}^{r+1},$$

where $f_{ii} = P_i(T_i < \infty)$.

We have seen many situations that T_i and N_i are closely related. Here is another result about their relationship.

Corollary 3.9

$$P_i(N_i = \infty) = 1 \Leftrightarrow P_i(T_i < \infty) = 1$$

and

$$P_i(T_i < \infty) < 1 \Leftrightarrow P_i(N_i = \infty) = 0 \Leftrightarrow \mathbb{E}_i(N_i) < \infty.$$

Corollary 3.9 links the finiteness of T_i and N_i and also relates it to the expectation. With the following formula of expectation, Corollary 3.9 will be very useful:

$$\mathbb{E}(X) = \sum_{t=1}^{\infty} P(X \geq t), \tag{3.5}$$

when X is a random variable taking integer values.

3.8.2 Recurrence and Transience

Based on the return time property, we classify a state i as

$$\begin{cases} \textbf{recurrent/persistent}, & \text{if } P_i(T_i < \infty) = f_{ii} = 1 \\ \textbf{transient}, & \text{otherwise.} \end{cases}$$

Furthermore, a recurrent state is called

$$\begin{cases} \textbf{positive recurrent}, & \text{if } \mathbb{E}_i(T_i) < \infty \\ \textbf{null recurrent}, & \text{otherwise.} \end{cases}$$

Note that: either $P_i(N_i = \infty) = 0$ or $P_i(N_i = \infty) = 1$, with nothing in between (if $f_{ii} < 1$, then $P_i(N_i = \infty) = 0$; if $f_{ii} = 1$, then $P_i(N_i = \infty) = 1$). This, together with Corollary 3.9, implies that $\mathbb{E}_i(N_i) = \infty \iff P_i(N_i = \infty) = 1$.

Note that:

$$f_{ii} = P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1.$$

In other words, if a Markov chain returns to state i in finite time, then the chain visits this state infinitely often.

Proposition 3.10 *State i is recurrent* $\iff \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

Proof: State i is recurrent $\iff P_i(T_i < \infty) = f_{ii} = 1 \iff P_i(N_i = \infty) = 1$ by Corollary 3.9.

It is easy to see that $P_i(N_i = \infty) = 1 \iff \mathbb{E}_i(N_i) = \infty$.

Using equation (3.5), $\mathbb{E}_i(N_i) = \sum_{n=1}^{\infty} p_{ii}^{(n)}$ and the result follows. ■

Proposition 3.11 *Recurrence is a communication class property, i.e. if $i \leftrightarrow j$ and i is recurrent, then j is recurrent.*

Proof: Homework. ■

Example: Gambler's Ruin+. Recall that in Gambler's ruin, the game ends when X_n hits 0 or $a+b$. Now we extend the problem in the sense that the game does not ends when a player loses/takes all money but the value of X_n stays the same once it hits 0 or $a+b$. Namely, $X_n = 0 \Rightarrow X_{n+1} = 0$ and $X_n = a+b \Rightarrow X_{n+1} = a+b$. In this case $p_{00}^{(k)} = p_{a+b,a+b}^{(k)} = 1$ for all $k = 1, 2, \cdots$. Therefore, $\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} p_{a+b,a+b}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty$. Hence, 0 and $a+b$ are recurrent states. Once they are reached we stay there forever. Let q be the probability that player 1 loses. Consider state 1. If the next round the player 1 loses, the chain stuck at 0 forever. Namely, $T_1 = \infty$ because we can never come back. So $P_1(T_1 = \infty) \geq q$, which implies

$$P_1(T_1 < \infty) = 1 - P_1(T_1 = \infty) \leq 1 - q < 1 \text{ if } q \in (0, 1).$$

Note that the inequality in $P_1(T_1 = \infty) \geq q$ is due to the fact that even if player 1 wins, the game may end at $a+b$, so the return time to state 1 may still be infinite. Therefore, by definition, 1 is a transient state. Since states $\{1, \ldots, a+b-1\}$ form a communication class, all states in this class are also transient. These states are transient because they occur a finite number of times before absorption into states 0 or $a+b$.

Example 8: 1-D Random Walk (revisited). Let X_n be a random walk on the set of all integers $\mathbb{Z}$ such that

$$p_{ij} = \begin{cases} p & \text{if } j = i+1 \\ q := 1-p & \text{if } j = i-1. \end{cases}$$

Let's study recurrence of state 0. We know that $p_{00}^{(2n+1)} = 0$ for all $n \geq 0$ and that, conditional on $X_0 = 0$, $X_{2n} =_d \xi_1 + \cdots + \xi_{2n}$, where $\xi_1, \ldots, \xi_n$ are i.i.d. with $P(\xi_i = 1) = 1 - P(\xi = -1) = p$. Hence,

$$p_{00}^{(2n)} = P(X_{2n} = 0 | X_0 = 0) = \binom{2n}{n} p^n q^n.$$

Recall that Stirlings formula says that $n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}$, meaning that

$$\lim_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}} = 1.$$

Therefore,

$$\begin{aligned} p_{00}^{(2n)} &= \frac{(2n)!}{n!n!} p^n q^n \\ &\sim \frac{(2n)^{2n+\frac{1}{2}} e^{-2n} \sqrt{2\pi}}{n^{2n+1} e^{-2n} 2\pi} (pq)^n \\ &= \frac{2^{2n+\frac{1}{2}} n^{2n+\frac{1}{2}}}{n^{2n+1} 2^{\frac{1}{2}} \sqrt{\pi}} (pq)^n = \frac{(pq)^n 2^{2n}}{\sqrt{\pi n}} = \frac{(4pq)^n}{\sqrt{\pi n}}. \end{aligned}$$

We deduce that

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} p_{00}^{(2n)} = \infty \quad \Leftrightarrow \quad 4pq \geq 1 \quad \Leftrightarrow \quad p = q = \frac{1}{2}.$$

(Ratio Test: Let $\sum_{n=1}^{\infty} a_n$ be a series which satisfies $\lim_{n\to\infty} |\frac{a_{n+1}}{a_n}| = k$. If $k > 1$ the series diverges, if $k < 1$ the series converges.) Conclusion: Only the *symmetric* random walk is recurrent on $\mathbb{Z}$. Interestingly, the symmetric random walk on $\mathbb{Z}^2$ is also recurrent, but it is transient on $\mathbb{Z}^n$ for $n \geq 3$, See Brémaud (1999, p. 98).

3.8.3 Invariant Measures

With the knowledge about recurrence, we are able to talk about the invariant measures and stationary distribution of a stochastic matrix.

A vector $x \neq 0$ is called an **invariant measure** of a stochastic matrix $\mathbf{P}$ if

- $\infty > x_i \geq 0$ for each i, and
- $x^T \mathbf{P} = x^T$, i.e., $x_i = \sum_j x_j p_{ji}$ for each i.

A probability vector π on a Markov chain state space is called a **stationary distribution** of a stochastic matrix $\mathbf{P}$ if $\pi^T \mathbf{P} = \pi^T$, i.e., $\pi_i = \sum_j \pi_j p_{ji}$ for each i.

The equation $x^T \mathbf{P} = x^T$ or $\pi^T \mathbf{P} = \pi^T$ is also called the *global balance equaitons* – the probability flow in equals the flow out. Note that for an invariance measure x such that $c = \sum_i x_i < \infty$, $c^{-1}x$ is a stationary distribution. But it may happen that $c = \infty$ for some invariant measure so one cannot always normalize it.

Example 9: Two-State Markov Chain. Consider a Markov chain with two states and a transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \quad 0 < p < 1, \quad 0 < q < 1.$$

The global balance equations:

$$\begin{bmatrix} \pi_0, \pi_1 \end{bmatrix} \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} = \begin{bmatrix} \pi_0, \pi_1 \end{bmatrix} \text{ or}$$

$$\begin{cases} (1-p)\pi_0 + q\pi_1 & = \pi_0 \\ p\pi_0 + (1-q)\pi_1 & = \pi_1 \end{cases} \Rightarrow \quad p\pi_0 = q\pi_1 \quad \Rightarrow \quad \pi_0 = \frac{q}{p}\pi_1.$$

Using that $\pi_0 + \pi_1 = 1$, we obtain

$$\frac{q}{p}\pi_1 + \pi_1 = 1 \quad \Rightarrow \quad \pi_1 = \frac{p}{p+q}$$

and deduce that the global balance equations have the unique solution

$$\pi^T = \left[\frac{q}{p+q}, \frac{p}{p+q} \right],$$

which is the stationary distribution.

Example: Gambler's Ruin+ (simple version). Let the total fortune of both players be $a+b = 4$. Then

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

By inspection, vectors $\pi_\alpha^T = [\alpha, 0, 0, 0, 1-\alpha]$ satisfy global balance equations: $\pi_\alpha^T \mathbf{P} = \pi_\alpha^T$ for any $\alpha \in (0, 1)$. So the Gambler's ruin chain has an uncountable number of stationary distributions.

Here, we see the case where a Markov chain may have infinite number of stationary distribution. And in some cases it may not even have a stationary distribution! So returning to our original questions, we would like to know (i) when will a Markov chain has a stationary distribution? and (ii) how to find a stationary distribution? and (ii) when the stationary distribution will be unique?

The following proposition partially answer the first question. Note that a Markov chain is recurrent if all its states are recurrent.

Proposition 3.12 *Let $\{X_n\}$ be an irreducible, recurrent, homogeneous Markov chain with transition probability matrix* $\mathbf{P}$. *For each $i \in S$ define*

$$y_i = \mathbb{E}_0\left[\sum_{n=1}^{\infty} I(X_n = i) I(n \leq T_0)\right],$$

where 0 is an arbitrary reference state and $T_0 = \inf\{n \geq 1 : X_n = 0\}$ is the first return time to 0. Then $y_i \in (0, \infty)$ for all $i \in S$, and $\mathbf{y}^T = [y_0, y_1, \dots]$ is an invariant measure of $\mathbf{P}$.

Note: For $i \neq 0$, y_i is the expected number of visits to state i before returning to 0.

Before starting the proof, we note the following three properties.

(P1) When $i = 0$,

$$y_0 = \mathbb{E}_0\left[\sum_{n=1}^{\infty} I(X_n = 0) I(n \leq T_0)\right] = 1$$

because for $n \geq 1$, $X_n = 0$ if and only if $n = T_0$.

(P2)

$$\begin{aligned}
\sum_{i \in S} y_i &= \sum_{i \in S} \mathbb{E}_0\left[\sum_{n=1}^{\infty} I(X_n = i) I(n \leq T_0)\right] \\
&= \mathbb{E}_0\left[\sum_{n=1}^{\infty} \sum_{i \in S} I(X_n = i) I(n \leq T_0)\right] \\
&= \mathbb{E}_0\left[\sum_{n=1}^{\infty} I(n \leq T_0)\right] \\
&= \mathbb{E}_0(T_0).
\end{aligned}$$

(P3) For any $i \in S$, we define

$$q_{0i}^{(n)} \equiv \mathbb{E}_0(I(X_n = i) I(n \leq T_0)) = P_0(X_1 \neq 0, X_2 \neq 0, \cdots, X_{n-1} \neq 0, X_n = i)$$

to be the probability of visiting state i at time point n before returning to state 0. Thus,

$$y_i = \sum_{n=1}^{\infty} \mathbb{E}_0(I(X_n = i) I(n \leq T_0)) = \sum_{n=1}^{\infty} q_{0i}^{(n)}$$

and $q_{0i}^{(1)} = \mathbb{E}_0(I(X_1 = i) I(1 \leq T_0)) = p_{0i}$.

Proof: This proof consists of two parts. In the first part, we prove that each y_i satisfies $y_i = \sum_{j \in S} y_j p_{ji}$. In the second part, we will show that $0 < y_i < \infty$ for every $i \in S$.

Part 1. To show that $y_i = \sum_{j\in S} y_j p_{ji}$, we analyze $q_{0i}^{(n)}$ defined in property (P3):

$$\begin{aligned}
q_{0i}^{(n)} &= P_0(X_1 \neq 0, X_2 \neq 0, \cdots, X_{n-1} \neq 0, X_n = i) \\
&= \sum_{j\neq 0} P_0(X_1 \neq 0, X_2 \neq 0, \cdots, X_{n-1} = j, X_n = i) \\
&= \sum_{j\neq 0} P_0(X_n = i \mid X_1 \neq 0, X_2 \neq 0, \cdots, X_{n-1} = j) \underbrace{P_0(X_1 \neq 0, X_2 \neq 0, \cdots, X_{n-1} = j)}_{=q_{0j}^{(n-1)}} \\
&= \sum_{j\neq 0} P(X_n = i \mid X_{n-1} = j) q_{0j}^{(n-1)} \quad \text{(Markov property)} \\
&= \sum_{j\neq 0} q_{0j}^{(n-1)} p_{ji}.
\end{aligned}$$

Thus,

$$\begin{aligned}
y_i &= \sum_{n=1}^{\infty} q_{0i}^{(n)} \\
&= p_{0i} + \sum_{n=2}^{\infty} q_{0i}^{(n)} \\
&= p_{0i} + \sum_{n=2}^{\infty} \sum_{j\neq 0} q_{0j}^{(n-1)} p_{ji} \\
&= p_{0i} + \sum_{n=1}^{\infty} \sum_{j\neq 0} q_{0j}^{(n)} p_{ji} \\
&= p_{0i} + \sum_{j\neq 0} \underbrace{\left(\sum_{n=1}^{\infty} q_{0j}^{(n)} \right)}_{=y_j} p_{ji} \\
&= \underbrace{y_0}_{=1} p_{0i} + \sum_{j\neq 0} y_j p_{ji} \\
&= \sum_{j\in S} y_j p_{ji}.
\end{aligned}$$

Part 2. Now we show that $0 < y_i < \infty$. First note that $y_0 = 1$ so we only need to focus on cases $y \neq 0$.

Because the Markov chain is irreducible, for each state i there exists a number $n(i) \geq 1$ such that $p_{0i}^{(n(i))} > 0$. Then using the fact that $y^T = y^T \mathbf{P}$ implying $y^T = y^T \mathbf{P}^{(n(i))}$,

$$y_i = \sum_{j\in S} y_j p_{ji}^{(n(i))} = \underbrace{y_0 p_{0i}^{(n(i))}}_{>0} + \sum_{j\neq 0} y_j p_{ji}^{(n(i))} > 0.$$

To show that $y_i < \infty$, we prove by contradiction. Assume that $y_i = \infty$. Because the Markov chain is irreducible, there exists a constant $k(i)$ such that $p_{i0}^{(k(i))} > 0$. Then

$$y_0 = \sum_{j\in S} y_j p_{j0}^{(k(i))} = \underbrace{y_i p_{i0}^{(k(i))}}_{=\infty} + \sum_{j\neq i} y_j p_{j0}^{(k(i))} = \infty,$$

a contradiction. Thus, $y_i < \infty$.

■

Proposition 3.13 *The invariant measure of an irreducible and recurrent chain is unique up to a multiplicative factor.*

Proof: See Brémaud (1999, p. 102). ■

Proposition 3.14 *An irreducible, recurrent and homogeneous Markov chain is positive recurrent ⇔ all of its invariant measures y satisfy $\sum_{i \in S} y_i < \infty$.*

Proof: By Proposition 3.12, there is an invariant measure y with

$$y_i = \mathbb{E}_0 \left[I(X_n = i) I(n \le T_0) \right].$$

Moreover, by Proposition 3.13, this invariant measure is unique up to a multiplicative factor. So what remains to prove is to show that $\sum_{i \in S} y_i < \infty$.

Using property (P2),

$$\sum_{i \in S} y_i = \mathbb{E}_0(T_0).$$

Therefore, positive recurrent $\iff \mathbb{E}_0(T_0) < \infty \iff \sum_{i \in S} y_i < \infty$.

■

To see why positive recurrent is important, consider the following example about a $1-D$ random walk on all integers $\mathbb{Z}$ with $p \neq q$ is transient and recurrent if $p = q = 0.5$. This Markov chain has an invariant measure $\mathbf{y}^T = [1, 1, \dots]$ for any p and q since

$$\mathbf{P} = \begin{bmatrix} \ddots & \ddots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & q & 0 & p & \cdots & \cdots & \cdots \\ \cdots & \cdots & q & 0 & p & \cdots & \cdots \\ \cdots & \cdots & \cdots & q & 0 & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots \end{bmatrix}$$

Since this measure is not normalizable (the state space is $\mathbb{Z}$), the 1-D random walk can not be positive recurrent. Thus, we see that an irreducible homogeneous Markov chain can have an invariant measure and still be transient or null recurrent.

Lemma 3.15 *Let $\{X_n\}$ be a homogeneous Markov chain with state space S and n-step transition probability matrix $\mathbf{P}^n = \{p_{ij}^{(n)}\}$. If $i \in S$ is a transient state, then $\lim_{n \to \infty} p_{ji}^{(n)} = 0$ for all $j \in S$.*

Proof: This proof relies a trick – if $\sum_{n=1}^{\infty} p_{ji}^{(n)} < \infty$, then $\lim_{n \to \infty} p_{ji}^{(n)} = 0$. Thus, we only need to show that $\sum_{n=1}^{\infty} p_{ji}^{(n)} < \infty$ when $i \in S$ is a transient state.

By definition,

$$\sum_{n=1}^{\infty} p_{ji}^{(n)} = \sum_{n=1}^{\infty} P_j(X_n = i) = \sum_{n=1}^{\infty} \mathbb{E}_j(I(X_n = i)) = \mathbb{E}_j \left(\sum_{n=1}^{\infty} I(X_n = i) \right) = \mathbb{E}_j(N_i).$$

So we can switch our goal to $\mathbb{E}_j(N_i)$.

Because N_i is a RV taking integer values, we can rewrite its expectation as

$$\mathbb{E}_j(N_i) = \sum_{k=1}^{\infty} P_j(N_i \geq k) = \sum_{k=0}^{\infty} P_j(N_i \geq k+1).$$

So we only need to compute each of $P_j(N_i > k)$. Now, recalled from the proof of Proposition 3.8,

$$P_j(N_i \geq k+1) = f_{ji} f_{ii}^k.$$

We obtain

$$\mathbb{E}_j(N_i) = \sum_{k=0}^{\infty} P_j(N_i \geq k+1) = f_{ji} \sum_{k=0}^{\infty} f_{ii}^k.$$

Because the state i is transient, $f_{ii} < 1$ so the above summation becomes

$$\mathbb{E}_j(N_i) = f_{ji} \sum_{k=0}^{\infty} f_{ii}^k = \frac{f_{ji}}{1 - f_{ii}} < \infty,$$

which is the desired result.

■

Finally, we obtain the criterion for stationary distribution.

Theorem 3.16 (Stationary Distribution Criterion) *An irreducible homogeneous Markov chain is positive recurrent if and only if it has a stationary distribution. Moreover, if the stationary distribution* $\pi^T = [\pi_1, \pi_2, \dots]$ *exists, it is unique and* $\pi_i > 0$ *for all* $i \in S$.

Proof: $\Rightarrow$:
By Propositions 3.12 and 3.14, the vector y defined in Proposition 3.12 is an invariant measure with $\sum_{i \in S} y_i < \infty$. Thus, the probability vector $\pi = y / \sum_{i \in S} y_i$ is the stationary distribution.

The uniqueness follows from Proposition 3.13.

$\Leftarrow$:
To prove this direction, we use proof by contradiction. Because recurrence is a communication class property (Proposition 3.11) and the Markov chain is irreducible, the fact that a state i is transient implies every state is transient. Let π be a stationary distribution and we assume that the Markov chain is transient.

By Lemma 3.15, $\lim_{n \to \infty} p_{ji}^{(n)} = 0$ for any state $j \in S$. Since π is a stationary distribution, $\pi^T = \pi^T \mathbf{P}^n$.

Using the dominated convergence theorem (we can exchange summation and limit)[1],

$$\pi_i = \lim_{n \to \infty} \pi_i = \lim_{n \to \infty} \sum_{j \in S} \pi_j p_{ji}^{(n)} = \sum_{j \in S} \pi_j \lim_{n \to \infty} p_{ji}^{(n)} = \sum_{j \in S} \pi_j \times 0 = 0$$

for every state $i \in S$.

Then we conclude $\sum_{i \in S} \pi_i = 0 \neq 1$, a contradiction to the definition of stationary distribution. Thus, the Markov chain is recurrent then by Proposition 3.14, the Markov chain is positive recurrent.

■

[1] This works because $\pi_j p_{ji}^{(n)} \leq \pi_j$ for every n and $\sum_{j \in S} \pi_j = 1$.

In the above case, we are working on a state space S that may possibly contain infinite number of states. In many realistic scenarios the number of states is finite. Does the finiteness of state number gives us any benefits? The answer is yes – and it gives us a huge benefit.

Theorem 3.17 *An irreducible homogeneous Markov chain on a finite state space is positive recurrent. Therefore, it always has a stationary distribution.*

Proof:

We first prove that the chain is recurrent. We proceed by proof by contradiction. Assume that the chain is transient. In the proof of Lemma 3.15, we have shown that if a state i is transient, then

$$\sum_{n=1}^{\infty} p_{ji}^{(n)} = \mathbb{E}_j(N_i) = f_{ji} \sum_{k=0}^{\infty} f_{ii}^k = \frac{f_{ji}}{1 - f_{ii}} < \infty.$$

Because the number of state space is finite,

$$\sum_{i \in S} \sum_{n=1}^{\infty} p_{ji}^{(n)} < \infty.$$

However, if we exchange the summations,

$$\sum_{i \in S} \sum_{n=1}^{\infty} p_{ji}^{(n)} = \sum_{n=1}^{\infty} \sum_{i=1}^{s} p_{ji}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty,$$

which is a contradiction. So we conclude that the chain is recurrent.

To see if the chain is positive recurrent, note that Proposition 3.12 shows that there exists an invariant measure y. Because the number of state space is finite, $\sum_{i \in S} y_i < \infty$ so by Proposition 3.14, the chain is positive recurrent.

■

Finally, we end this lecture on the relation between the return time and the stationary distribution.

Theorem 3.18 *Let $\{X_n\}$ be an irreducible homogeneous positive recurrent Markov chain. Then*

$$\pi_i = \frac{1}{\mathbb{E}_i(T_i)},$$

where $\pi = (\pi_1, \cdots, \pi_s)$ is the stationary distribution of $\{X_n\}$ and $T_i = \inf\{n \geq 1 : X_n = i\}$ is the return time to state i.

Proof: Define a vector y such that $y_i = \mathbb{E}_0\left(\sum_{n=1}^{\infty} I(X_n = i) I(n \leq T_0)\right)$. We already know that such a vector describes an invariant measure and $\pi_i = \frac{y_i}{\sum_{j \in S} y_j}$.

Now we consider the case $i = 0$. Then $y_0 = \mathbb{E}_0\left(\sum_{n=1}^{\infty} I(X_n = 0) I(n \leq T_0)\right) = 1$ by property (P1). Moreover, $\sum_{i \in S} y_i = \mathbb{E}_0(T_0)$ due to property (P2). Thus, $\pi_0 = \frac{y_0}{\sum_{i \in S} y_i} = \frac{1}{\mathbb{E}_0(T_0)}$.

Because state 0 is just a reference state, we can apply the same argument to any other state. Thus, we conclude that $\pi_i = \frac{1}{\mathbb{E}_i(T_i)}$ for each $i \in S$.

■

Here is a short summary about what we have learned so far:

1. Irreducibility + recurrence $\Rightarrow$ There exists an invariant measure that is unique up to a proportionality constant.
2. Irreducibility + *positive* recurrence $\Leftrightarrow$ Irreducibility + there exists a stationary distribution π and it is unique. Moreover, when π exists, $\pi_i > 0$ and $\pi_i = 1/\mathbb{E}_i[T_i]$.
3. Irreducibility + finite state-space $\Rightarrow$ Irreducibility + positive recurrence.

Here is a summary about the classification of states. Recall $f_{ii} = P_i(T_i < \infty)$ is the probability of return to i given we start at i and $\mathbb{E}_i(T_i)$ is the expected return time. State i is called:

1. *Recurrent* if $f_{ii} = 1$.
2. *Transient* if $f_{ii} < 1$.
3. *Positive Recurrent* if $f_{ii} = 1$ and $\mathbb{E}_i(T_i) < \infty$.
4. *Null Recurrent* if $f_{ii} = 1$ and $\mathbb{E}_i(T_i) = \infty$.
5. *Periodic* with period d_i if $p_{ii}^{(n)} = 0$ for all n not divisible by d_i, and d_i (> 1) is the greatest such integer.
6. *Aperiodic* if $d_i = 1$.
7. *Ergodic* if 3. and 6. apply.
8. *Absorbing* if $p_{ii} = 1$.