IB Markov Chains

Ishan Nath, Michaelmas 2022

Based on Lectures by Dr. Perla Sousi

October 29, 2022

Contents

0 Introduction

Markov chains are random processes (sequence of random variables) that retain no memory of the past.

0.1 History

These were first studied by Markov in 1906. Before Markov, Poisson processes and branching processes were studied. The motivation was to extend the law of large numbers to a non-iid setting.

After Markov, Kolmogorov began studying continuous time Markov chains, also known as Markov processes. An important example is Brownian motion, which is a fundamental object in modern probability theory.

Markov chains are the simplest mathematical models for random phenomena evolving in time. They are **simple** in the sense they are amenable to tools from probability, analysis and combinatorics.

Applications of Markov chains include population growth, mathematical genetics, queueing networks and Monte Carlo simulation.

0.2 PageRank Algorithm

This is an algorithm used by Google Search to rank web pages. We model the web as a directed graph, $G = (V, E)$. Here, V is the set of vertices, which are associated to the website, and $(i, j) \in E$ if i contains a link to page j.

Let $L(i)$ be the number of outgoing edges from i, i.e. the outdegree, and let $|V| = n$. Then we define a set of probabilities

$$\hat{p}_{ij} = \begin{cases} \frac{1}{L(i)} & \text{if } L(i) > 0, (i,j) \in E, \\ \frac{1}{n} & \text{otherwise.} \end{cases}$$

Take $\alpha \in (0, 1)$, then we define $p_{ij} = \alpha \hat{p}_{ij} + (1 - \alpha)\frac{1}{n}$. Consider a random surfer, who tosses a coin with bias α, and either goes to $\hat{p}$, or chooses a website uniform at random.

We wish to find an invariant distribution $\pi = \pi P$. Then π_i is the proportion of time spent at webpage i by the surfer. We can then rank the pages by the values of π_i.

1 Formal Setup

We begin with a state space I, which is either finite or countable, and a σ-algebra $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1. A stochastic process $(X_n)_{n\geq 0}$ is called a **Markov chain** if for all $n \geq 0$, and $x_0, x_1, \ldots, x_{n+1} \in I$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

If $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ is independent of n for all x, y, then X is called **time-homogeneous**. Otherwise, it is **time-inhomogeneous**.

For a time-homogeneous Markov chain, define $P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x)$. P is called the **transition matrix** of the Markov chain. We have

$$\sum_{y\in I} P(x, y) = \sum_{y\in I} \mathbb{P}(X_1 = y \mid X_0 = x) = 1.$$

Such a matrix is called a **stochastic matrix**.

Definition 1.2. $(X_n)_{n\geq 0}$ with values in I is called $\text{Markov}(\lambda, P)$ if $X_0 \sim \lambda$ and $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix P.

There are several equivalent definitions for Markov chains.

Theorem 1.1. *X is $\text{Markov}(\lambda, P)$ if for all $n \geq 0$, $x_0, x_1, \ldots, x_n \in I$,*

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \lambda(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Proof: If X is $\text{Markov}(\lambda, P)$, then

$$\begin{aligned}\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) &= \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \times \cdots \\ &\quad \times \mathbb{P}(X_1 = x_1 \mid X_0 = x_0)\mathbb{P}(X_0 = x_0) \\ &= \lambda(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).\end{aligned}$$

For the other direction, note for $n = 0$, we have $\mathbb{P}(X_0 = x_0) = \lambda(x_0)$, so $X_0 \sim \lambda$, and

$$\begin{aligned}\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) &= \frac{\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0)}{\mathbb{P}(X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)} \\ &= P(x_{n-1}, x_n).\end{aligned}$$

Definition 1.3. Let $i \in I$. The δ_i-mass at i is defined by

$$\delta_{ij} = \mathbb{1}(i = j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.4. Let $X_1, \dots, X_n$ be discrete random variables with values in I. They are independent if for all $x_1, \dots, x_n \in I$,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

Let $(X_n)_{n \geq 0}$ be a sequence of random variables in I.They are independent if for all $i_1 < i_2 < \dots < i_k$, and for all $x_1, \dots, x_k \in I$,

$$\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = i_k) = \prod_{j=1}^{k} \mathbb{P}(X_{i_j} = x_j).$$

Let $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ be two sequences. We say $X \perp Y$, or X independent to Y, if for all $k, m \in \mathbb{N}$, $i_1 < \dots < i_k$, $j_1 < \dots < j_m$, $x_1, \dots, x_k, y_1, \dots, y_m$,

$$\begin{aligned} &\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k, Y_{j_1} = y_1, \dots, Y_{j_m} = y_m) \\ &\quad = \mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k)\mathbb{P}(Y_{j_1} = y_1, \dots, Y_{j_m} = y_m). \end{aligned}$$

Theorem 1.2 (Simple Markov Property)**.** *Suppose X is* Markov(λ, P) *with values in I. Let $m \in \mathbb{N}$ and $i \in I$. Then conditional on $X_m = i$, the process $(X_{m+n})_{n \geq 0}$ is* Markov(δ_i, P) *and it is independent of $X_0, \dots, X_m$.*

Proof: Let $x_0, \dots, x_n \in I$. Then

$$\begin{aligned} &\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n \mid X_m = i) \\ &\qquad = \delta_{ix_0} \frac{\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n)}{\mathbb{P}(X_m = i)}, \\ &\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n) \\ &\qquad = \sum_{y_0, \dots, y_{m-1}} \mathbb{P}(X_0 = y_0, \dots, X_m = x_0, \dots, X_{m+n} = x_0) \\ &\qquad = \sum_{y_0, \dots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, x_0) \cdots P(x_{n-1}, x_n) \\ &\qquad = P(x_0, x_1) \cdots P(x_{n-1}, x_n) \sum_{y_0, \dots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, x_0), \\ &\mathbb{P}(X_m = i) = \sum_{y_0, \dots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, i). \end{aligned}$$

Putting this together, we get the probability is

$$\delta_{ix_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n) \implies \text{Markov}(\delta_i, P).$$

Now we show independence. Let $m \leq i_1 < \cdots < i_k$. Then,

$$\begin{aligned}
&\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k, X_0 = y_0, \dots, X_m = y_m \mid X_m = i) \\
&\quad = \frac{\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k, X_0 = y_0, \dots, X_m = y_m)}{\mathbb{P}(X_m = i)} \\
&\quad = \frac{\lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, y_m)}{\mathbb{P}(X_m = i)} \mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k \mid X_m = i) \\
&\quad = \mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k \mid X_m = i) \mathbb{P}(X_0 = y_0, \dots \mid X_m = i).
\end{aligned}$$

Let $X \sim \text{Markov}(\lambda, P)$. How can we find $\mathbb{P}(X_n = x)$? Evaluating,

$$\begin{aligned}
\mathbb{P}(X_n = x) &= \sum_{x_0, \dots, x_{n-1}} \mathbb{P}(X_0 = x_0, \dots, X_n = x) \\
&= \sum_{x_0, \dots, x_{n-1}} \lambda(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x) = (\lambda P^n) x.
\end{aligned}$$

Here, λ is a row vector, and P^n is the n'th power of the transition matrix. By convention, $P^0 = I$.

Consider the related problem of finding $\mathbb{P}(X_{n+m} = y \mid X_m = x)$. From the simple Markov property, $(X_{m+n})_{n \geq 0}$ is $\text{Markov}(\delta_x, P)$. So

$$\mathbb{P}(X_{n+m} = y \mid X_m = x) = (\delta_x P^n) y = (P^n) xy.$$

Example 1.1. Take the transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Then $p_{11}(n+1) = (1-\alpha) p_{11}(n) + \beta p_{12}(n)$. Since $p_{11}(n) + p_{12}(n) = 1$, we get the general form

$$p_{11}(n) = \begin{cases} \dfrac{\alpha}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \alpha+\beta > 0, \\ 1 & \alpha+\beta = 0. \end{cases}$$

Suppose P is $k \times k$ stochastic, and let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of P.

If $\lambda_1, \dots, \lambda_k$ are all distinct, then P is diagonalisable, so we can write

$$P = U \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix} U^{-1}.$$

Then we get

$$P^n = U \begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^n \end{pmatrix}.$$

Hence $p_{11}(n) = \alpha_1 \lambda_1^n + \cdots + \alpha_k \lambda_k^n$.

If one of the eigenvalues is complex, say λ_{k-1}, then also its conjugate is an eigenvalue. Say $\lambda_k = \overline{\lambda_{k-1}}$. If $\lambda_{k-1} = re^{i\theta} = r\cos\theta + ir\sin\theta$, $\lambda_k = r\cos\theta - ir\sin\theta$, then we can write the general form as

$$p_{11}(n) = \alpha_1 \lambda_1^n + \cdots + \alpha_{k-2}\lambda_{k-2}^n + \alpha_{k-1} r^n \cos n\theta + \alpha_k r^n \sin n\theta.$$

If an eigenvalue λ has multiplicity r, then we must include the term $(a_{r-1}n^{r-1} + \cdots + a_1 n + a_0)\lambda^n$, by Jordan Normal Form.

1.1 Communicating Classes

Definition 1.5. X is a Markov chain with matrix P on I. Let $x, y \in I$. We say $x \to y$ (x leads to y) if

$$\mathbb{P}_x(X_n = y \text{ for some } n \geq 0\} > 0.$$

We say that x and y communicate and $x \leftrightarrow y$ if both $x \to y$ and $y \to x$.

Theorem 1.3. *The following are equivalent:*

(i) $x \to y$;

(ii) *There exists a sequence* $x = x_0, x_1, \dots, x_k = y$ *such that*

$$P(x_0, x_1), \dots, P(x_{k-1}, x_k) > 0;$$

(iii) *There exists* $n \geq 0$ *such that* $P_{xy}(n) > 0$.

Proof:

We show (i) if and only if (iii). Note

$$\{X_n = y \text{ for some } n \geq 0\} = \bigcup_{n \geq 0} \{X_n = y\}.$$

If $x \to y$, then there exists $n \geq 0$ such that $\mathbb{P}_x(X_n = y) > 0$, so $P_{xy}(n) > 0$. If there exists $n \geq 0$ such that $P_x(X_n = y) > 0$, then $x \to y$.

Now we note (ii) iff (iii) as

$$\mathbb{P}_x(X_n = y) = \sum_{x_1, \ldots, x_{n-1}} P(x, x_1) \cdots P(x_{n-1}, y).$$

Corollary 1.1. *$\leftrightarrow$ defines an equivalence class on I.*

Proof: $x \leftrightarrow x$ as $P_{xx}(0) = 1$, and $x \leftrightarrow y \iff y \leftrightarrow x$.

Then transitivity follows from property (ii).

Definition 1.6. The equivalence classes induced by $\leftrightarrow$ on I are called communicating classes. We say that a class C is closed if whenever $x \in C$ and $x \to y$, then $y \in C$.

A matrix P is called irreducible if it has a single communicating class.

A state x is called absorbing if $\{x\}$ is a closed class. Equivalently, if the Markov chain started from x, it would remain at x forever.

Definition 1.7. For $A \subset I$, we let $T_A : \Omega \to \mathbb{N} \cup \{\infty\}$. Then we define

$$T_A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}.$$

By convention, we take $\inf(\emptyset) = \infty$. Then T_A is the first hitting time of A.

Denote $h_i^A = \mathbb{P}_i(T_A < \infty)$. Then $h^A : I \to [0, 1]$ is a vector of hitting probabilities. We can also define

$$k^A : I \to \mathbb{R}_+ \cup \{\infty\},$$

as the mean hitting time. Then

$$k_i^A = \mathbb{E}_i[T_A] = \sum_{n=1}^{\infty} n\mathbb{P}_i(T_A = n),$$

if $\mathbb{P}_i(T_A = \infty) = 0$.

Theorem 1.4. *Let $A \subset I$. The vector $(h_i^A : i \in A)$ is a solution to the linear system*

$$h_i^A = \begin{cases} 1 & i \in A, \\ \sum_j P(i,j) h_j^A & i \notin A. \end{cases}$$

The vector (h_i^A) is the minimal non-negative solution to this solution.

Proof: Clearly, if $i \in A$, then $h_i^A = 1$. So assume $i \notin A$. Then

$$\begin{aligned} h_i^A &= \mathbb{P}_i(T_A < \infty) = \sum_{n=1}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \\ &= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_n \in A) \\ &= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \sum_{j \notin A} \mathbb{P}_i(X_1 = j, X_2 \notin A, \ldots, X_n \in A) \\ &= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \sum_{j \notin A} \mathbb{P}_i(X_2 \notin A, \ldots, X_n \in A \mid X_0 = i, X_1 = j) P(i,j) \\ &= \mathbb{P}_i(X_1 = A) + \sum_{n=1}^{\infty} \sum_{j \notin A} P(i,j) \mathbb{P}_j(X_1 \notin A, \ldots, X_n \in A) \\ &= \sum_{j \in A} P(i,j) h_j^A + \sum_{j \notin A} P(i,j) h_j^A \\ &= \sum_j P(i,j) h_j^A. \end{aligned}$$

Now we prove minimality. Let (x_i) be another non-negative solution. We are required to show $h_i^A \le x_i$. Assume $i \notin A$. Then we know

$$\begin{aligned}
x_i &= \sum_j P(i,j) x_j \\
x_i &= \sum_{j \in A} P(i,j) + \sum_{j \notin A} P(i,j) x_j \\
&= \sum_{j \in A} P(i,j) + \sum_{j \notin A} \sum_{k \in A} P(i,j)P(j,k) + \sum_{j \in A} \sum_{k \notin A} P(i,j)P(j,k)x_k \\
&= \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A} \sum_{k \in A} P(i,j)P(j,k)x_k \\
&\ge \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \cdots + \mathbb{P}_i(X_1 \notin A, \ldots, X_n \in A).
\end{aligned}$$

Hence $x_i \ge \mathbb{P}_i(T_A \le n)$, and since the union of these events is $\{T_A < \infty\}$, we get

$$x_i \ge \mathbb{P}_i(T_A < \infty) = h_i^A.$$

Example 1.2. Consider a simple random walk on $\mathbb{Z}_+$. We have a transition matrix $P(0,1) = 1$, and $P(i, i+1) = p = 1 - P(i, i-1)$. Then we wish to find $h_i = \mathbb{P}_i(T_0 < \infty)$. We know $h_0 = 1$, and $h_i = p \cdot h_{i+1} + q h_{i-1}$. This gives

$$h_i = a + b \left(\frac{q}{p}\right)^i = a + (1-a)\left(\frac{q}{p}\right)^i.$$

We assume $q > p$: to get the non-negative and minimal solution we need to take $a = 1$. Then $h_i = 1$ for all $i \ge 1$.

Now assume $q < p$: We get $a = 0$, s $h_i = (q/p)^i$.

If $p = q = 1/2$, we get $h_i = a + bi$, so by boundedness, $a = 1$, and $b = 0$. So $h_i = 1$ for all $i \ge 1$.

1.2 Birth and Death chains

The above example is almost a birth and death chain, with equal probability at each i. Here, we have $P(0,0) = 1$, $P(i, i+1) = p_i$, $P(i, i-1) = q_i$.

We have $h_i = \mathbb{P}_i(T_0 < \infty)$, $h_0 = 1$. Then

$$h_i = p_i h_{i+1} + q_i h_{i-1}.$$

This gives

$$p_i(h_{i+1} - h_i) = q_i(h_i - h_{i-1}).$$

We set $u_i = h_i - h_{i-1}$. Then

$$u_{i+1} = \frac{q_i}{p_i} u_i = \cdots = \prod_{k=1}^{i} \frac{q_k}{p_k},$$

with $u_1 = h_1 - 1$. Moreover, we have

$$h_i = \sum_{j=1}^{i} (h_j - h_{j-1}) + 1 = 1 + \sum_{j=1}^{i} u_j = 1 + u_1 + \sum_{j=2}^{i} u_1 \prod_{k=1}^{j-1} \frac{q_k}{p_k}.$$

Hence

$$h_i = 1 + (h_1 - 1) + (h_1 - 1) \sum_{j=2}^{i} \prod_{k=1}^{j-1} \frac{q_k}{p_k}.$$

We will let

$$\lambda_j = \sum_{k=0}^{j} \frac{q_k}{p_k},$$

where $\lambda_0 = 1$. Then

$$h_i = 1 - (1 - h_1) \sum_{j=0}^{i-1} \lambda_j.$$

We want (h_i) to be the minimal non-negative solution, so

$$(1 - h_1) \leq \frac{1}{\sum_{j=0}^{\infty} \lambda_j},$$

and

$$h_1 = 1 - \frac{1}{\sum_{j=0}^{\infty} \lambda_j},$$

by minimality. Thus if $\sum \lambda_j < \infty$, then

$$h_i = \frac{\sum_{j=i}^{\infty} \lambda_j}{\sum_{j=0}^{\infty} \lambda_j}.$$

If the sum $\sum \lambda_j = \infty$, then $h_i = 1$.

1.3 Mean Hitting Times

For $A \subset I$, $T_A = \inf\{n \geq 0 \mid X_n \in A\}$. Then $k_i^A = \mathbb{E}_i[T_A]$.

Theorem 1.5. *The vector $(k_i^A \mid i \in I)$ is the minimal non-negative solution to the system*

$$k_i^A = \begin{cases} 0 & i \in A, \\ 1 + \sum_{j \notin A} P(i,j) k_j^A & i \in A \end{cases}$$

Proof: If $i \in A$, then $k_i^A = 0$. Assume $i \notin A$. Then

$$\begin{aligned} k_i^A = \mathbb{E}_i[T_A] &= \sum_{n=0}^{\infty} \mathbb{P}_i(T_A > n) = \sum_{n=0}^{\infty} \mathbb{P}_i(X_0 \notin A, \ldots, X_n \not\ni A) \\ &= 1 + \sum_{n=1}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_n \notin A) \\ &= 1 + \sum_{n=1}^{\infty} \sum_j \mathbb{P}_i(X_1 = j, X_1 \notin A, \ldots, X_n \notin A) \\ &= 1 + \sum_{n=1}^{\infty} \sum_j P(i,j) \mathbb{P}(X_1 \notin A, \ldots, X_n \notin A \mid X_1 = j) \\ &= 1 + \sum_{n=1}^{\infty} \sum_j P(i,j) \mathbb{P}_j(X_0 \notin A, \ldots, X_{n-1} \notin A) \\ &= 1 + \sum_j P(i,j) \sum_{n=0}^{\infty} \mathbb{P}_j(X_0 \notin A, \ldots, X_n \notin A) \\ &= 1 + \sum_j P(i,j) k_j^A = 1 + \sum_{j \notin A} P(i,j) k_j^A. \end{aligned}$$

Now we show minimality. Let (x_i) be another non-negative solution. The $x_i = 0$, $i \in A$. If $i \notin A$,

$$\begin{aligned}
x_i &= 1 + \sum_{j \notin A} P(i,j)x_j = 1 + \sum_{j \notin A} P(i,j) + \sum_{j \notin A}\sum_{k \notin A} P(i,j)P(j,k)x_k \\
&= 1 + \sum_{j_1 \notin A} P(i,j_1) + \cdots + \sum_{j_1,\dots,j_{n-1} \notin A} P(i,j_1)\cdots P(j_{n-2},j_{n-1}) + \cdots \\
&\geq 1 + \mathbb{P}_i(T_A > 1) + \cdots + \mathbb{P}_i(T_A > n).
\end{aligned}$$

So $x_i \geq \mathbb{E}_i[T_A] = k_i^A$.

1.4 Strong Markov Property

We have proven that the past and the future are independent, as the simple Markov property. This says, for $m \in \mathbb{N}$, $i \in I$, and $X \sim \text{Markov}(\lambda, P0$, that conditional on $X_m = i$, $(X_{n+m})_{n \geq 0}$ is $\text{Markov}(\delta_i, P)$, and is independent of $X_0, \dots, X_m$.

We look to replace the constant m with a random variable.

1.5 Stopping Times

Definition 1.8. A random variable $T : \Omega \to \{0, 1, \dots\} \cup \{\infty\}$ is called a stopping time if the event $\{T = n\}$ depends on $X_0, \dots, X_n$, for all $n \in \mathbb{N}$.

For example, let $T_A = \inf\{n \geq 0 \mid X_n \in A\}$. Then $\{T_A = n\} = \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$. So the first hitting times are always stopping times.

However, $L_A = \sup\{n \leq 10 \mid X_n \in A\}$ is not a stopping time.

The strong Markov property is as follows:

Proposition 1.1. *Let X be* $\text{Markov}(\lambda, P)$ *and let T be a stopping time. Conditioning on $T < \infty$ and $X_T = i$, then $(X_{T+n})_{n \geq 0}$ is* $\text{Markov}(\delta_i, P)$*, and it is independent of $X_0, \dots, X_T$.*

Proof: Take $x_0, \dots, x_n \in I$, and $w \in \bigcup I^k$. Then we will show

$$\begin{aligned}
&\mathbb{P}(X_T = x_0, \dots, X_{T+n} = x_n, (X_0, \dots, X_T) = \omega \mid T < \infty, X_T = i) \\
&\qquad = \delta_{ix_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n) \mathbb{P}((X_0, \dots, X_T) = \omega \mid T < \infty, X_T = i).
\end{aligned}$$

This proves both statements. Set ω to have length k. Then

$$
\begin{aligned}
\mathbb{P}(X_k =& x_0, \ldots, X_{k+n} = x_n, (X_0, \ldots, X_k) = \omega, T = k \mid T < \infty, X_T = i) \\
=& \frac{\mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n, (X_0, \ldots, X_k) = \omega, T = k, X_k = i)}{\mathbb{P}(T < \infty, X_T = i)} \\
=& \mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n \mid (X_0, \ldots, X_k) = \omega, T = k, X_k = i) \\
& \times \frac{\mathbb{P}((X_0, \ldots, X_k) = \omega, T = k, X_k = i)}{\mathbb{P}(T < \infty, X_T = i)} \\
=& \mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n \mid X_k = i) \\
& \times \mathbb{P}((X_0, \ldots, X_k) = \omega, T = k \mid T < \infty, X_T = i) \\
=& \delta_{ix_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n) \mathbb{P}((X_0, \ldots, X_T) = \omega \mid T < \infty, X_T = i).
\end{aligned}
$$

This proves the Strong Markov property.

Example 1.3. Consider a Markov chain with $P(0,1) = 1$, and $P(i, i+1) = P(i, i-1) = 1/2$. Let $T_0 = \inf\{n \geq 0 \mid X_n = 0\}$, and let $h_i = \mathbb{P}_i(T_0 < \infty)$.

We know that $h_0 = 1$. We can condition to find $h_1 = 1/2 + h_2/2$.

We can use more knowledge of the Markov chain to find h_2. Notice that

$$h_2 = \mathbb{P}_2(T_0 < \infty) = \mathbb{P}_2(T_1 < \infty, T_0 < \infty) = \mathbb{P}_2(T_0 < \infty \mid T_1 < \infty)\mathbb{P}_2(T_1 < \infty).$$

Then conditional on $T_1 < \infty$, by the strong Markov property, X_{T_1+n} is Markov(δ_1, P). So we can express $T_0 = T_1 + \tilde{T}_0$, where $\tilde{T}_0$ is independent of T_1 and has the same law as T_0 under $\mathbb{P}_1$. Hence

$$\mathbb{P}_2(T_0 < \infty \mid T_1 < \infty) = \mathbb{P}_2(\tilde{T}_0 + T_1 < \infty \mid T_1 < \infty) = \mathbb{P}_1(T_0 < \infty) = h_1.$$

This gives $h_2 = h_1^2$, so $h_1 = 1/2 + h_1^2/2$ implies $h_1 = 1$.

2 Transience and Recurrence

Definition 2.1. A state i is called recurrent if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$

A state i is called transient if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 0.$$

Define the total number of visits V_i as

$$V_i = \sum_{l=0}^{\infty} \mathbb{1}(X_l = i).$$

Then by definition, we known $\mathbb{P}_i(V_i > 0) = 1$. Moreover, we can write $\mathbb{P}_i(V_i > 1) = \mathbb{P}_i(T_i^{(1)} < \infty)$, where $T_i^{(1)}$ is the first return time.

Using the strong Markov property, we can then write $\mathbb{P}_i(V_i > 2) = \mathbb{P}_i(T_i^{(1)} < \infty)^2$. Indeed, let $T_i^{(0)} = 0$, and for $k \geq 1$, define

$$T_i^{(k)} = \inf\{n > T_i^{(k-1)} \mid X_n = i\}.$$

This is the k-th return time to i. Define

$$f_i = \mathbb{P}_i(T_i^{(1)} < \infty).$$

Lemma 2.1. *For all $r \in \mathbb{N}$,*

$$\mathbb{P}_i(V_i > r) = f_i^r.$$

Thus V_i has geometric distribution.

Proof: $r = 0$ is trivially true. Suppose it is true for $r \leq k$. We will prove it for $k + 1$. Indeed,

$$\begin{aligned}\mathbb{P}_i(V_i > k+1) = \mathbb{P}_i(T_i^{(k+1)} < \infty) &= \mathbb{P}_i(T_i^{(k+1)} < \infty, T_i^{(k)} < \infty) \\ &= \mathbb{P}_i(T_i^{(k+1)} < \infty \mid T_i^{(k)} < \infty)\mathbb{P}_i(T_i^{(k)} < \infty).\end{aligned}$$

By induction, we have $\mathbb{P}_i(T_i^{(k)} < \infty) = \mathbb{P}_i(V_i > k) = f_i^k$.

Moreover, the successive return times to i are stopping times, so conditional on $T_i^{(k)} < \infty$, we have $(X_{T_i^{(k)}+n})$ is Markov(δ_i, P), and is independent of $X_0, \ldots, X_{T_i^{(k)}}$ by the strong Markov property. Hence,

$$\mathbb{P}_i(T_i^{(k+1)} < \infty \mid T_i^{(k)} < \infty) = \mathbb{P}_i(T_i^{(1)} < \infty) = f_i.$$

This finishes the proof.

Theorem 2.1.

(a) *If $f_i = 1$, then i is recurrent and*

$$\sum_{n \geq 0} p_{ii}(n) = \infty.$$

(b) *If $f_i < 1$, then i is transient and*

$$\sum_{n \geq 0} p_{ii}(n) < \infty.$$

Proof: Note that

$$\mathbb{E}_i[V_i] = \mathbb{E}_i\left[\sum_{l=0}^{\infty} \mathbb{1}(X_l = i)\right] = \sum_{l=0}^{\infty} p_{ii}(l).$$

(a) If $f_i = 1$, then by our lemma, $\mathbb{P}_i(V_i = \infty) = 1$. Hence i is recurrent, so

$$\mathbb{E}_i[V_i] = \infty \implies \sum_{n \geq 0} p_{ii}(n) = \infty.$$

(b) If $f_i < 1$, from our lemma,

$$\mathbb{E}_i[V_i] = \frac{1}{1 - f_i} < \infty \implies \sum_{n \geq 0} p_{ii}(n) < \infty,$$

which implies that $\mathbb{P}_i(V_i < \infty) = 1$, so i is transient.

Theorem 2.2. *If $x \leftrightarrow y$, then x and y are either both recurrent or both transient.*

Proof: We will show if x is recurrent, then y is as well. Note if $x \leftrightarrow y$, then there exist $m, r > 0$ such that $p_{xy}(m) > 0$, and $p_{yx}(r) > 0$. Hence

$$p_{yy}(n + m + r) \geq p_{yx}(r) p_{xx}(n) p_{xy}(m),$$

so

$$\sum_{n \geq 0} p_{yy}(n + m + r) \geq p_{yx}(r) p_{xy}(m) \sum_{n \geq 0} p_{xx}(n) = \infty,$$

as x is recurrent. Hence y is also recurrent.

Corollary 2.1. *All sates in a communicating class are either all recurrent or all transient.*

Theorem 2.3. *If C is a recurrent communicating class, then C is closed.*

Proof: Let $x \in C$, and $x \to y$ but $y \notin C$. Since $x \to y$, there exist some $m > 0$ such that $p_{xy}(m) > 0$. But then

$$\mathbb{P}_x(V_x < \infty) \geq p_{xy}(m) > 0.$$

So this shows that x is transient.

Theorem 2.4. *A finite closed class is recurrent.*

Proof: Let $x \in C$. Since C is finite, there exists $y \in C$ such that

$$\mathbb{P}_x(X_n = y \text{ for infinitely many } n) > 0$$

by the pigeonhole principle. Hence

$$\begin{aligned}
&\mathbb{P}_y(X_n = y \text{ for infinitely many } n) \\
&\qquad \geq \mathbb{P}_y(X_m = x, X_n = y \text{ for infinitely many } n \geq m) \\
&\qquad = \mathbb{P}_y(X_n = y \text{ for infinitely many } n \geq m \mid X_m = x)\mathbb{P}_y(X_m = x) \\
&\qquad = \mathbb{P}_x(X_n = y \text{ for infinitely many } n)p_{yx}(m) > 0.
\end{aligned}$$

So $\mathbb{P}_y(X_n = y \text{ for infinitely many } n) > 0$, so y is recurrent.

Theorem 2.5. *If P is irreducible and recurrent, then for all x, y,*

$$\mathbb{P}_x(T_y < \infty) = 1.$$

Proof:

$$\begin{aligned}
\mathbb{P}_x(X_n = y \text{ infinitely many times}) \\
&= \mathbb{P}_x(T_y < \infty, X_n = y \text{ for infinitely many } n \geq T_y) \\
&= \mathbb{P}_x(X_n = y \text{ for infinitely many } n \geq T_y \mid T_y < \infty)\mathbb{P}_x(T_y < \infty) \\
&= \mathbb{P}_y(X_n = y \text{ for infinitely many } n)\mathbb{P}_x(T_y < \infty) = \mathbb{P}_x(T_y < \infty).
\end{aligned}$$

Suppose that $\mathbb{P}_x(T_y < \infty) < 1$, so $\mathbb{P}_x(T_y = \infty) > 0$. Then

$$\begin{aligned}
\mathbb{P}_y(V_y < \infty) &\geq \mathbb{P}_y(X_m = x, \tilde{T}_y = \infty) \\
&= \mathbb{P}_y(\tilde{T}_y = \infty \mid X_m = x)\mathbb{P}_y(X_m = x) \\
&= \mathbb{P}_x(T_y = \infty)p_{yx}(m) > 0.
\end{aligned}$$

This implies y is transient, a contradiction.

3 Simple Random Walks on $\mathbb{Z}^d$

Definition 3.1. A simple random walk on $\mathbb{Z}^d$ is a Markov chain with transition matrix

$$P(x, x + e_i) = P(x, x - e_i) = \frac{1}{2d}.$$

Theorem 3.1 (Pólya)**.** *A simple random walk is recurrent when $d \leq 2$ and it is transient when $d \geq 3$.*

Proof: We first prove recurrence when $d = 1$. This is a simple random walk on $\mathbb{Z}$. Since this is irreducible, it is enough to show 0 is recurrent, that is,

$$\sum_{n \geq 0} p_{00}(n) = \infty.$$

Note $p_{00}(n) = \mathbb{P}_0(X_n = 0)$. Note n must be even for this to be possible. We can calculate

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n! \cdot n!} \cdot \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}},$$

where we use Stirling's formula. Hence, since this sum diverges, the simple random walk on $\mathbb{Z}$ is recurrent.

Consider when $P(i, i+1) = p$, $P(i, i-1) = q$, where $p \neq q$, $p + q = 1$. We show this is not recurrent. Indeed,

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n} p^n q^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}.$$

If $p \neq q$, then $4pq < 1$, so this sum converges as it is exponential, implying that the walk is transient.

For $d = 2$, we project the random walk onto the lines $y = x$ and $y = -x$. Then we prove that these are independent random walks on $\sqrt{2}\mathbb{Z}$.

Define a function $f : \mathbb{Z}^2 \to \mathbb{R}^2$ by

$$f(x, y) = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}} \right).$$

Then for a simple random walk (X_n) on $\mathbb{Z}^2$, let

$$f(X_n) = (X_n^+, X_n^-).$$

Theorem 3.2. $(X_n^+), (X_n^-)$ *are two independent simple random walks on* $\mathbb{Z}/\sqrt{2}$.

Indeed, let (ξ_i) be an iid sequence, where

$$\mathbb{P}(\xi_i = (0,1)) = \mathbb{P}(\xi_i = (1,0)) = \mathbb{P}(\xi_i = (0,-1)) = \mathbb{P}(\xi_i = (-1,0)) = \frac{1}{4}.$$

Let $\xi_i = (\xi_i^1, \xi_i^2)$, and define X_n as the partial sums of the ξ_i. Then

$$f(X_n) = \left(\sum_{i=1}^n \frac{(\xi_i^1 + \xi_i^2)}{\sqrt{2}}, \sum_{i=1}^n \frac{(\xi_i^1 - \xi_i^2)}{\sqrt{2}} \right) = (X_n^+, X_n^-).$$

It is easy to see that $(X_n^+), (X_n^-)$ are simple random walks on $\mathbb{Z}/\sqrt{2}$. To show they are independent, it is enough to check that $\xi_i^1 + \xi_i^2$ is independent of $\xi_i^1 - \xi_i^2$. This can be done by calculating all possible calculations.

Hence, applying this, we get

$$\mathbb{P}_0(X_{2n} = 0) = \mathbb{P}_0(X_{2n}^+ = 0, X_{2n}^- = 0) \sim \frac{1}{\pi n}.$$

This sum diverges, hence $\mathbb{Z}^2$ is recurrent.

Index