IB Markov Chains

Ishan Nath, Michaelmas 2022

Based on Lectures by Prof. Perla Sousi

November 23, 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 Markov(λ, 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 Markov(λ, 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 Markov(λ, 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 $X_1, \ldots, X_n$ be discrete random variables with values in I. They are independent if for all $x_1, \ldots, x_n \in I$,

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

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

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

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 $k, m \in \mathbb{N}$, $i_1 < \cdots < i_k$, $j_1 < \cdots < j_m$, $x_1, \ldots, x_k, y_1, \ldots, y_m$,

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

1.1 Simple Markov Property

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 $X_0, \ldots, X_m$.*

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

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

$$\mathbb{P}(X_m = i) = \sum_{y_0, \ldots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, i).$$

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 $m \le i_1 < \cdots < i_k$. Then,

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

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

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

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 Markov(δ_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 $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of P.

If $\lambda_1, \ldots, \lambda_k$ 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.

1.2   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 $x = x_0, x_1, \ldots, x_k = y$ such that*

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

(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.

1.3   Hitting Probabilities

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*

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

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

$$h_i^A = \mathbb{P}_i(T_A < \infty) = \sum_{n=1}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A)$$

$$= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_n \in A)$$

$$= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \sum_{j \notin A} \mathbb{P}_i(X_1 = j, X_2 \notin A, \ldots, X_n \in A)$$

$$= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \sum_{j \notin A} \mathbb{P}_i(X_2 \notin A, \ldots, X_n \in A \mid X_0 = i, X_1 = j)P(i,j)$$

$$= \mathbb{P}_i(X_1 = A) + \sum_{n=1}^{\infty} \sum_{j \notin A} P(i,j)\mathbb{P}_j(X_1 \notin A, \ldots, X_n \in A)$$

$$= \sum_{j \in A} P(i,j)h_j^A + \sum_{j \notin A} P(i,j)h_j^A$$

$$= \sum_{j} P(i,j)h_j^A.$$

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

$$x_i = \sum_{j} P(i,j)x_j$$

$$x_i = \sum_{j \in A} P(i,j) + \sum_{j \notin A} P(i,j)x_j$$

$$= \sum_{j \in A} P(i,j) + \sum_{j \notin A} \sum_{k \in A} P(i,j)P(j,k) + \sum_{j \in A} \sum_{k \notin A} P(i,j)P(j,k)x_k$$

$$= \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A} \sum_{k \in A} P(i,j)P(j,k)x_k$$

$$\geq \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \cdots + \mathbb{P}_i(X_1 \notin A, \ldots, X_n \in A).$$

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

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

> **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 $i \geq 1$.
>
> 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 $i \geq 1$.

1.4 Birth and Death chains

The above example is almost a birth and death chain, with equal probability at each i. A *birth and death chain* is a random walk on $\mathbb{Z}_{\geq 0}$, with transition probabilities $P(0,0) = 1$, $P(i, i+1) = p_i$, $P(i, i-1) = q_i$.

We look at the hitting probabilities for a birth and death chain. Let $h_i = \mathbb{P}_i(T_0 < \infty)$, $h_0 = 1$. Then,

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

This gives

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

If we define $u_i = h_i - h_{i-1}$, 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

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

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

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

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

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 $h_i = 1$.

1.5   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

$$k_i^A = \mathbb{E}_i[T_A] = \sum_{n=0}^{\infty} \mathbb{P}_i(T_A > n) = \sum_{n=0}^{\infty} \mathbb{P}_i(X_0 \notin A, \ldots, X_n \not\ni A)$$

$$= 1 + \sum_{n=1}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_n \notin A)$$

$$= 1 + \sum_{n=1}^{\infty} \sum_j \mathbb{P}_i(X_1 = j, X_1 \notin A, \ldots, X_n \notin A)$$

$$= 1 + \sum_{n=1}^{\infty} \sum_j P(i,j) \mathbb{P}(X_1 \notin A, \ldots, X_n \notin A \mid X_1 = j)$$

$$= 1 + \sum_{n=1}^{\infty} \sum_j P(i,j) \mathbb{P}_j(X_0 \notin A, \ldots, X_{n-1} \notin A)$$

$$= 1 + \sum_j P(i,j) \sum_{n=0}^{\infty} \mathbb{P}_j(X_0 \notin A, \ldots, X_n \notin A)$$

$$= 1 + \sum_j P(i,j) k_j^A = 1 + \sum_{j \notin A} P(i,j) k_j^A.$$

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

$$x_i = 1 + \sum_{j \notin A} P(i,j) x_j = 1 + \sum_{j \notin A} P(i,j) + \sum_{j \notin A} \sum_{k \notin A} P(i,j) P(j,k) x_k$$

$$= 1 + \sum_{j_1 \notin A} P(i,j_1) + \cdots + \sum_{j_1, \ldots, j_{n-1} \notin A} P(i,j_1) \cdots P(j_{n-2}, j_{n-1}) + \cdots$$

$$\geq 1 + \mathbb{P}_i(T_A > 1) + \cdots + \mathbb{P}_i(T_A > n).$$

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

1.6   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 $X \sim \text{Markov}(\lambda, P)$, that conditional on $X_m = i$, $(X_{n+m})_{n \geq 0}$ is $\text{Markov}(\delta_i, P)$, and is independent of $X_0, \ldots, X_m$.

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

1.7   Stopping Times

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

For example, let $T_A = \inf\{n \geq 0 \mid X_n \in A\}$. Then $\{T_A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}$. 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* Markov(λ, P) *and let T be a stopping time. Conditioning on $T < \infty$ and $X_T = i$, then $(X_{T+n})_{n \geq 0}$ is* Markov(δ_i, P), *and it is independent of $X_0, \ldots, X_T$.*

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

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

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

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

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

2.1   Transient and Recurrent Classes

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 states 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

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

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:

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

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

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

This implies y is transient, a contradiction.

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

Definition 2.2. 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 2.6 (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 2.7. $(X_n^+), (X_n^-)$ *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 possiblities.

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.

Finally, we prove $\mathbb{Z}^3$ is transient by showing $\sum_n p_{00}(n) < \infty$. In order for the walk to be back at 0 after $2n$ steps is it to make:

- i steps to the right, i steps to the left,

- j steps north, j steps south,

- k steps east, k steps west,

for $i + j + k = n$. Hence we can write

$$p_{00}(2n) = \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{2n}{i,i,j,j,k,k} \left(\frac{1}{6}\right)^{2n}$$

$$= \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i,j,k}^2 \left(\frac{1}{3}\right)^{2n}.$$

By a simple combinatorial argument, we get

$$\sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i,j,k} \left(\frac{1}{3}\right)^n = 1.$$

We now prove a lemma. Let $n = 3m$. Then our claim is

$$\binom{n}{i,j,k} \leq \binom{n}{m,m,m}.$$

Indeed, if the maximum is attained at i, j, k, then $i \leq j + 1$, as

$$\binom{n}{i,j,k} \geq \binom{n}{i-1,j+1,k} \implies i!j! \leq (i-1)!(j+1)! \implies i \leq j+1.$$

Since this holds for any pair in (i, j, k), we have $(i, j, k) = (m, m, m)$. Hence, for $n = 3m$, we can bound

$$p_{00}(2n) \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\left(\frac{1}{3}\right)^n \binom{n}{m,m,m} \leq \frac{A}{n^{3/2}},$$

for some A, by Stirling's formula. So $\sum_m p_{00}(6m) < \infty$. But now

$$p_{00}(6m) \geq p_{00}(6m-2)\left(\frac{1}{6}\right)^2, \quad p_{00}(6m) \geq p_{00}(6m-4)\left(\frac{1}{6}\right)^4.$$

So $\sum_n p_{00}(2n) < \infty$, giving transience.

3   Invariant Distributions

3.1   Probability Distributions

Definition 3.1. Let I be a discrete set. Then $\lambda = (\lambda_i \mid i \in I)$ is a probability distribution if $\lambda_i \geq 0$ and $\sum_{i \in I} \lambda_i = 1$.

Example 3.1.

Consider the Markov chain with transition probability

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

Then by looking at P^n, we can show that $p_{11}(n) \to 1/2$ as $n \to \infty$. So the Markov chain will converge to $(1/2, 1/2)$.

We want to find a probability distribution π such that if $X_0 \sim \pi$, then $X_n \sim \pi$ for all n. Then

$$\pi(j) = \mathbb{P}(X_1 = j) = \sum_i \mathbb{P}(X_0 = i, X_1 = j)$$

$$= \sum_i \mathbb{P}(X_1 = j \mid X_0 = i)\mathbb{P}(X_0 = i) = \sum_i P(i,j)\pi(i).$$

Hence $\pi(j) = \sum_i \pi(i)P(i,j)$, or $\pi = \pi P$, where π is a row vector.

Definition 3.2. A probability distribution π is called invariant (or stationary or equilibrium) of $\pi = \pi P$.

Theorem 3.1. *Let π be invariant and $X_0 \sim \pi$. Then $X_n \sim \pi$.*

Proof: We proceed by induction. Our base case $n = 0$ is given to us. Now assume $X_n \sim \pi$. Then

$$\mathbb{P}(X_{n+1} = j) = \sum_i \mathbb{P}(X_{n+1} = j, X_n = i)$$

$$= \sum_i \mathbb{P}(X_{n+1} = j \mid X_n = i)\mathbb{P}(X_n = i)$$

$$= \sum_i P(i,j)\pi(i) = \pi(j).$$

Theorem 3.2. *Let I be a finite set and suppose that there exists $i \in I$ such that $p_{ij}(n) \to \pi(j)$ as $n \to \infty$. Then $\pi = (\pi(i) \mid i \in I)$ is an invariant distribution.*

Proof:
$$\sum_{j \in I} \pi_j = \sum_{j \in I} \lim_{n \to \infty} p_{ij}(n) = \lim_{n \to \infty} \sum_{j \in I} p_{ij}(n) = 1,$$

so π is a probability distribution. Now note

$$\pi_j = \lim_{n \to \infty} p_{ij}(n) = \lim_{n \to \infty} \sum_{k \in I} p_{ik}(n-1)P(k,j))$$

$$= \sum_{k \in I} \lim_{n \to \infty} p_{ik}(n-1)P(k,j) = \sum_{k \in I} \pi_k P(k,j),$$

so $\pi = \pi P$.

Remark. I is finite is essential. Consider a walk on $\mathbb{Z}$, with $P(i, i+1) = 1$. Then $p_{0i} \to 0$ for all i.

Since P is a stochastic matrix, 1 is always an eigenvalue. If P is irreducible on a finite state sequence, then the Perron-Frobenius theorem ensures the existence of an invariant distribution.

3.2   Invariant Measures

Definition 3.3. For $k \in I$, define $T_k = \inf\{n \geq 1 \mid X_n = k\}$ be the first return time to k. Then for every state i, define

$$\nu_k(i) = \mathbb{E}_k\left[\sum_{\ell=0}^{T_k - i} \mathbb{1}(X_\ell = i)\right],$$

so $\nu_k(i)$ is the expected number of visits to i during an excursion from k. Thus ν_k is a measure on I.

Theorem 3.3. *If P is irreducible and recurrent, then ν_k is an invariant measure satisfying $0 < \nu_k(i) < \infty$, and $\nu_k(k) = 1$.*

Proof: Clearly $\nu_k(k) = 1$. Let $i \in I$. We will prove invariance, so $\nu_k(i) = \sum P(j,i)\nu_k(j)$. By recurrence, we get that $T_k < \infty$ with probability 1, and

$X_{T_k} = k$. So,

$$\nu_k(i) = \mathbb{E}_k\left[\sum_{\ell=1}^{T_k} \mathbb{1}(X_\ell = i)\right] = \mathbb{E}_k\left[\sum_{\ell 1}^{\infty} \mathbb{1}(X_\ell = 1)\mathbb{1}(T_k \geq \ell)\right]$$

$$= \sum_{\ell=1}^{\infty} \mathbb{P}_k(X_\ell = i, T_k \geq l) = \sum_{\ell=1}^{\infty}\sum_j \mathbb{P}_k(X_\ell = i, X_{\ell-1} = j, T_k \geq \ell)$$

$$= \sum_{\ell=1}^{\infty}\sum_j \mathbb{P}_k(X_\ell = i \mid X_{\ell-1} = j, T_k \geq l)\mathbb{P}_k(X_{\ell-1} = j \mid T_k \geq \ell).$$

Applying the Strong Markov property,

$$\mathbb{P}_k(X_\ell = i \mid X_{\ell-1} = j, T_k \geq l) = \mathbb{P}_k(X_\ell = i \mid X_{\ell-1} = k) = P(j, i).$$

So

$$\nu_k(i) = \sum_{\ell=1}^{\infty}\sum_j P(j, i)\mathbb{P}_k(X_{\ell-1} = j, T_k \geq \ell)$$

$$= \sum_j P(j, i)\mathbb{E}_k\left[\sum_{\ell=1}^{\infty} \mathbb{1}(X_{\ell-1} = j, T_k \geq \ell)\right]$$

$$= \sum_j P(j, i)\nu_k(j).$$

Hence ν_k is invariant. Now we show $\nu_k(i) > 0$. Since $\nu_k = \nu_k P^m$, we have

$$\nu_k(i) \geq \nu_k(k)p_{ki}(m) = p_{ki}(m).$$

By irreducibility, there exists m such that $p_{ki}(m) > 0$, so $\nu_k(i) > 0$. Now we show $\nu_k(i) < \infty$. Indeed,

$$1 = \nu_k(k) \geq \nu_k(i)p_{ik}(m).$$

Take n such that $p_{ik}(n) > 0$. Then we get

$$\nu_k(i) \leq \frac{1}{p_{ik}(n)} < \infty.$$

Theorem 3.4. *If P is irreducible and λ is an invariant measure satisfying $\lambda_k = 1$, then $\lambda \geq \nu_k$.*

If P is also recurrent, then $\lambda = \nu_k$.

Proof: We have $\lambda_i \geq 0$ for all i. Then

$$\lambda_i = \sum_j \lambda_j P(j, i) = P(k, i) + \sum_{j_1 \neq k} P(j_1, i) \lambda_{j_1}$$

$$= P(k, 1) + \sum_{j_1 \neq k} P(k, j_1) P(j_1, i) + \sum_{\substack{j_1 \neq k, \\ j_2 \neq k}} P(j_2, j_1) P(j_1, i) \lambda_{j_2}$$

$$= P(k, i) + \sum_{k_1 \neq k} P(k, j_1) P(j_1, i) + \cdots$$

$$+ \sum_{j_1, \ldots, j_{n-1} \neq k} P(k, j_{n-1} \cdots P(j_1, i) + \sum_{j_1, \ldots, j_n \neq k} P(j_n, j_{n-1}) \cdots P(j_1, i) \lambda_{j_n}.$$

Using the fact $\lambda_{j_n} \geq 0$, we have

$$\lambda_i \geq \mathbb{P}_k(X_1 = i, T_k \geq 2) + \mathbb{P}_k(X_2 = i, T_k \geq 3)$$
$$+ \cdots + \mathbb{P}_k(X_n = i, T_k \geq n + 1),$$

$$\lambda_i \geq \mathbb{E}_k \left[\sum_{\ell=1}^{n} \mathbb{1}(X_\ell = i, T_k \geq \ell + 1) \right]$$

$$= \sum_{\ell=0}^{n} \mathbb{E}_k[\mathbb{1}(X_\ell = i, \ell \leq T_k - 1)].$$

Taking $n \to \infty$, we get that $\lambda_i \geq \nu_k(i)$.

If P is recurrent, then ν_k is an invariant measure with $\nu_k(k) = 1$. Let $\mu_i = \lambda_i - \nu_k(i)$, an invariant measure with $\mu_i \geq 0$ and $\mu_k = 0$. Then we want to show $\mu_i = 0$ for all i. Let $i \in I$. Then,

$$0 = \mu_k = \sum_j \mu_j P^m(j, k) \implies \mu_k \geq \mu_i P^m(i, k).$$

Taking m such that $P_m(i, k) > 0$, we have $\mu_i = 0$.

Remark. If P is irreducible and recurrent, then all invariant measures are unique up to multiplicative factors.

Now we want to find out when we can get an invariant distribution $\pi = \pi P$ with $\sum \pi_i = 1$.

For P irreducible and recurrent, by uniqueness, we can get an invariant distribution

which is unique. If $\sum \nu_k(i) < \infty$, then we can find an invariant distribution

$$\pi_i = \frac{\nu_k(i)}{\sum_j \nu_k(j)}.$$

Thus we are interested in when this sum is finite. Note

$$\sum_{i \in I} \nu_k(i) = \sum_{i \in I} \mathbb{E}_k \left[\sum_{\ell=0}^{T_k-1} \mathbb{1}(X_\ell = i) \right] = \mathbb{E}_k \left[\sum_{\ell=0}^{T_k-1} \sum_{i \in I} \mathbb{1}(X_{\ell=i}) \right]$$
$$= \mathbb{E}_k[T_k].$$

So if $\mathbb{E}_k[T_k] < \infty$, then we can normalise.

3.3   Positive and Null Recurrence

Definition 3.4. For P irreducible and recurrent, let $T_k = \inf\{n \geq 1 \mid X_n = k\}$. Then P recurrent implies $\mathbb{P}_k(T_k < \infty) = 1$.

We say k is *positive recurrent* if $\mathbb{E}_k[T_k] < \infty$. Otherwise, we say k is *null recurrent*.

If k is positive recurrent, then

$$\pi_j = \frac{\nu_k(k)}{\mathbb{E}_k[T_k]} = \frac{1}{\mathbb{E}_k[T_k]}.$$

Theorem 3.5. *Let P be an irreducible matrix. Then the following are equivalent:*

(i) *All states are positive recurrent;*

(ii) *Some state is positive recurrent;*

(iii) *There exists an invariant distribution π_i.*

If any of the above holds, then $\pi_k = (\mathbb{E}_k[T_k])^{-1}$ exists and is an invariant distribution.

Proof: Clearly (i) $\implies$ (ii). Now we show (ii) $\implies$ (iii). Let k be the positive recurrent state. Then for all i,

$$\nu_k(i) = \mathbb{E}_k \left[\sum_{\ell=0}^{T_k-1} \mathbb{1}(X_\ell = i) \right].$$

Since k is recurrent, we have $\nu_k P = \nu_k$, so ν_k is an invariant measure. Now,

remember
$$\sum_{i\in I}\nu_k(i) = \mathbb{E}_k\left[\sum_{\ell=0}^{T_k-1}\sum_{i\in I}\mathbb{1}(X_\ell = i)\right] = \mathbb{E}_k[T_k].$$

Since k is positive recurrent, $\mathbb{E}_k[T_k] < \infty$. So we can define $\pi_i = \nu_k(i)/\mathbb{E}_k[T_k]$, as an invariant distribution.

Finally, we show (iii) $\implies$ (i). Let π be the invariant distribution. Let k be a state, then we want to show k is positive recurrent. First, we show $\pi_k > 0$. Note there exists $i \in I$ such that $\pi_i > 0$. Then since $\pi = \pi P = \pi P^n$ for all n,

$$\pi_k = \sum_j \pi_j P^n(j,k).$$

Take n such that $P^n(i,k) > 0$ by irreducibility, then $\pi_k \geq \pi_i P^n(i,k) > 0$. Define $\lambda_i = \pi_i/\pi_k$. Then this is an invariant measure on which $\lambda_k = 1$. So since P is irreducible, $\lambda \geq \nu_k$. Hence,

$$\mathbb{E}_k[T_k] = \sum_{i\in I}\nu_k(i) \leq \sum_{i\in I}\lambda_i = \frac{1}{\pi_k} < \infty,$$

giving k positive recurrent.

Now suppose (i), (ii) and (iii) holds, and let k be a state. We know k is positive recurrent. Define $\lambda_i = \pi_i/\pi_k$, which is an invariant measure with $\lambda_k = 1$. Then P recurrent gives $\lambda = \nu_k$, so from the same calculation as above,
$$\mathbb{E}_k[T_k] = \frac{1}{\pi_k}.$$

Corollary 3.1. *If P is irreducible, and π is the invariant distribution, then for all x, y,*
$$\nu_x(y) = \frac{\pi_y}{\pi_x}.$$

Example 3.2.

Take a symmetric simple random walk on $\mathbb{Z}$. Then P is recurrent. Does there exist an invariant distribution? We have
$$\pi_i = \frac{1}{2}\pi_{i-1} + \frac{1}{2}\pi_{i_1}.$$

Then $\pi_i \equiv 1$ is an invariant measure. Since P is recurrent, all invariant

measures have to be multiples of $\pi_i \equiv 1$. So there does not exist an invariant distribution, and so this walk in not positive recurrent.

Take an asymmetric simple random walk on $\mathbb{Z}$, with $P(x, x+1) = p$, $P(x, x-1) = q = 1 - p$ with $p > q$. Then $\pi_i = p\pi_{i-1} + q\pi_{i+1}$ gives

$$\pi_i = a + b\left(\frac{p}{q}\right)^i$$

as an invariant measure for any choice of a, b. So, there is no uniqueness up to a multiplicative factor, and indeed, this is valid under the observation that P is transient.

Again, we can take a simple random walk on $\mathbb{Z}^3$, which is transient. Then $\pi_i \equiv 1$ is an invariant measure, so the existence of an invariant measure does not imply recurrence.

Now take the asymmetric simple random walk $\mathbb{Z}_+$ with $P(x, x+1) = p$, $P(x, x-1) = q$, where $p < q$. We look for invariant measures $\pi = \pi P$. Similar to the walk on $\mathbb{Z}$,

$$\pi_i = \left(\frac{p}{q}\right)^i \pi_0$$

is an invariant measure. Setting $\pi_0 = 1 - p/q$, we can get the invariant distribution

$$\pi_i = \left(\frac{p}{q}\right)^i \left(1 - \frac{p}{q}\right).$$

Hence there exists an invariant distribution, so this walk is positive recurrent.

4   Time Reversibility

Proposition 4.1. *Let P be irreducible, and π an invariant distribution. Fix $N \in \mathbb{N}$ and $X_0 \sim \pi$. Define $Y_n = X_{N-n}$, for $0 \leq n \leq N$. Then, (Y_n) is a Markov chain with transition matrix*

$$\hat{P}(x,y) = \frac{\pi(y)}{\pi(x)}P(y,x)$$

and invariant distribution π.

Proof: $\hat{P}$ is a transition matrix, since

$$\sum_y \hat{P}(x,y) = \sum_y \frac{\pi(y)}{\pi(x)}P(y,x) = \frac{\pi(x)}{\pi(x)} = 1.$$

Let $y_0, \ldots, y_n \in I$. Then,

$$\begin{aligned}
\mathbb{P}(Y_0 = y_0, \ldots, Y_n = y_0) &= \mathbb{P}(X_n = y_0, \ldots, X_0 = y_N) \\
&= \mathbb{P}(X_0 = y_N, \ldots, X_N = y_0) = \pi(y_N)P(y_N, y_{N-1}) \cdots P(y_1, y_0) \\
&= \pi(y_{N-1})\hat{P}(y_{N_1}, Y_N)P(y_{N-1}, y_{N_2}) \cdots P(y_1, y_0) \\
&= \cdots = \pi(y_0)\hat{P}(y_0, y_1) \cdots \hat{P}(y_{N-1}, Y_n).
\end{aligned}$$

So Y is Markov$(\pi, \hat{P})$. Now we check $\hat{P}$ has invariant distribution π. Indeed,

$$\sum_x \pi(x)\hat{P}(x,y) = \sum_x \pi(x)\frac{\pi(y)}{\pi(x)}P(y,x) = \pi(y).$$

So $\pi\hat{P} = \pi$, and $\hat{P}$ is irreducible. Indeed, let $x, y \in I$. Then as P is irreducible, there exists $x_0 = x, \ldots, x_n = y$ with $P(x_0, x_1) \cdots P(x_{n-1}, x_n) > 0$. So,

$$\hat{P}(x_n, x_{n-1}) \cdots \hat{P}(x_1, x_0) = P(x_0, x_1) \cdots P(x_{n-1}, x_n)\frac{\pi(x_0)}{\pi(x_n)} > 0.$$

4.1   Detailed Balance Equation

Definition 4.1. A Markov chain with matrix P and invariant distribution π is called *(time) reversible* if $\hat{P} = P$, i.e for all $x, y \in I$,

$$\pi(x)P(x,y) = \pi(y)P(y,x),$$

which is the detailed balance equation.

Equivalently X is reversible of for all $N \in \mathbb{N}$, when $X_0 \sim \pi$, then

$$(X_0, \ldots, X_N) \sim (X_N, \ldots, X_0).$$

Example 4.1.

Take a walk on $\mathbb{Z}_n$, where $P(i, i+1) = \frac{2}{3}$ and $P(i, i-1) = \frac{1}{3}$. Then the invariant distribution is $\pi_i = \frac{1}{n}$, and so P is not reversible, because the

$$\pi(i)P(i, i+1) = \frac{2}{3n} \neq \frac{1}{3n} = \pi(i+1)P(i+1, i),$$

so the detailed balance equations are not satisfied.

Now take the same walk, except on $\{0, 1, \ldots, n-1\}$, with $P(0, 0) = \frac{1}{3}$ and $P(n-1, n-1) = \frac{2}{3}$. Then $\lambda_i = 2^i$ is an invariant distribution, so $\pi(i) \propto 2^i$.

We can then check π satisfies the detailed balance equations, so P is time reversible.

Lemma 4.1. *Let μ be a distribution satisfying*

$$\mu(x)P(x, y) = \mu(y)P(y, x),$$

for all $x, y \in I$. Then μ is an invariant distribution.

Proof:
$$\sum_x \mu(x)P(x, y) = \sum_x \mu(y)P(y, x) = \mu(y).$$

So $\mu = \mu P$.

Thus, when we look for an invariant distribution, we should look for a solution to the detailed balance equations. If a distribution that solves the detailed balance equation exists, then it is an invariant distribution.

On the other hand, if no solution to the detailed balance equation exists, and there is an invariant distribution, then the chain is not reversible.

Example 4.2.

Take a simple random walk on a graph $G = (V, E)$, where we stipulate that

G is finite and connected. The probabilities are

$$P(x, y) = \begin{cases} \frac{1}{d(x)} & (x, y) \in E, \\ 0 & \text{else.} \end{cases}$$

Since G is connected, P is irreducible. To find an invariant distribution, we look at the detailed balance equation:

$$\pi(x) \frac{1}{d(x)} = \pi(x) P(x, y) = \pi(y) P(y, x) = \pi(y) \frac{1}{d(y)}.$$

Taking $\nu(x) = d(x)$, then ν is an invariant measure, so

$$\pi(x) = \frac{d(x)}{\sum d(y)} = \frac{d(x)}{2|E|}$$

is an invariant distribution. Hence a simple random walk on G is reversible.

5   Convergence to Equilibrium

Theorem 5.1. *Let I be finite, and suppose there is $i \in I$ such that for all j,*

$$p_{ij}(n) \to \pi(j),$$

as $n \to \infty$. Then π is invariant.

Hence, if this holds, then we can show P has an invariant distribution π. Thus we want to find under what conditions do we have convergence to π.

> **Example 5.1.**
>
> Take a walk on $\{0, 1, \ldots, K\}$. Then, no matter the probabilities, $P^n(0,0) = 0$, hence we cannot find an invariant distribution in this way.

Definition 5.1. Let P be a transition matrix, and let $i \in I$. The period of i is defined as

$$d_i = \gcd\{n \geq 1 \mid P^n(i,i) > 0\}.$$

i is called *aperiodic* if $d_i = 1$.

Lemma 5.1. *Let P be a transition matrix and $i \in I$. Then $d_i = 1$ if and only if $P^n(i,i) > 0$ for all n sufficiently large.*

> **Proof:** It suffices to show that if $d_1 = 1$, then $P^n(i,i) > 0$ for all n large enough. Let $D(i) = \{n \geq 1 \mid P^n(1,1) > 0\}$.
>
> Observe if $n, m \in D(i)$, then $n + m \in D(i)$. Now it suffices to prove that $D(i)$ contains two consecutive integers: if it contains m and $m + 1$, then it will contain $am + b(m + 1)$ for all $a, b \in \mathbb{N}$. In particular, this hits all integers greater than m^2.
>
> Let $n, m \in D(i)$ such that $n = m + r$, with r minimal. Then there exists $k = \ell r + s \in D(i)$ with $0 < s < r$. Otherwise, all elements of $D(i)$ would be multiples of r.
>
> Let $a = (\ell + 1)n$, and $b = (\ell + 1)m + k$. Then $a, b \in D(i)$. Notice that $a - b = r - s < r$, contradiction.
>
> So $r = 1$, and $D(i)$ contains 2 consecutive integers.

Lemma 5.2. *If P is irreducible and $i \in I$ is aperiodic, then all states are aperiodic.*

Proof: Let $j \in I$, such that $P^r(i,j) > 0$ and $P^s(j,i) > 0$. Then for all large n,
$$P^{n+r+s}(j,j) \geq P^s(j,i)P^n(i,i)P^r(i,j) > 0.$$
So j is aperiodic.

5.1   Convergence to Equilibrium Theorem

Theorem 5.2. *Let P be irreducible and aperiodic with invariant distribution π. Let $X \sim \mathrm{Markov}(\lambda, P)$. Then for all y,*
$$\mathbb{P}(X_n = y) \to \pi(y) \ \text{as } n \to \infty.$$
In particular, for all x and y,
$$P^n(x,y) \to \pi(y) \ \text{as } n \to \infty,$$
by taking $\lambda = \delta_x$.

Proof: We couple Markov chains. Let $(Y_n) \sim \mathrm{Markov}(\pi, P)$, independently of X.

Consider $((X_n, Y_n)) \sim \mathrm{Markov}(\lambda\pi, \tilde{P})$, where
$$\tilde{P}((x,y),(x',y')) = P(x,x')P(y,y').$$
Then we claim $\tilde{P}$ is irreducible. Indeed, P is irreducible, so there exist ℓ, m such that
$$P^\ell(x,x') > 0, \qquad P^m(y,y') > 0.$$
Then take n sufficiently large such that $P^{n-\ell}(x',x') > 0$ and $P^{n-m}(y',y') > 0$ (by aperiodicity). Then
$$\tilde{P}^n((x,y),(x',y')) = P^n(x,x')P^n(y,y') > 0,$$
for all n large enough.

Notice that $\tilde{P}$ has invariant distribution $\tilde{\pi}(x,y) = \pi(x)\pi(y)$. Hence $\tilde{P}$ is positive recurrent. For $a \in I$, we can define
$$T = \inf\{n \geq 1 \mid (X_n, Y_n) = (a,a)\}.$$
Then T is a stopping time, and since $\tilde{P}$ is positive recurrent, $\mathbb{P}(T < \infty) = 1$. Now define
$$Z_n = \begin{cases} X_n & n < T, \\ Y_n & n \geq T \end{cases}.$$

(This is the coupled Markov chain). Then we claim $Z \sim \text{Markov}(\lambda, P)$. Indeed, $\mathbb{P}(Z_0 = x) = \mathbb{P}(X_0 = x) = \lambda(x)$, since $T \neq 0$, and if we let $A = \{Z_{n+1} = z_{n+1}, \ldots, Z_0 = z_0\}$, then we can write

$$\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) = \mathbb{P}(Z_{n+1} = y, T > n \mid Z_n = x, A)$$
$$+ \mathbb{P}(Z_{n+1} = y, T \leq n \mid Z_n = x, A).$$

Now notice that

$$\mathbb{P}(Z_{n+1} = y, T > n \mid Z_n = x, A)$$
$$= \mathbb{P}(X_{n+1} = y \mid T > n, X_n = x, A)\mathbb{P}(T > n \mid Z_n = x, A)$$
$$= P(x, y)\mathbb{P}(T > n \mid Z_n = x, A),$$

by the Strong Markov property. Similarly, we can compute

$$\mathbb{P}(Z_{n+1} = y, T \leq n \mid Z_n = x, A)$$
$$= P(x, y)\mathbb{P}(T \leq n, Z_n = x, A).$$

Adding these up, we get

$$\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) = P(x, y),$$

so $(Z_n) \sim (X_n)$. Now we want to show that $|\mathbb{P}(X_n = y) - \pi(y)| \to 0$ as $n \to \infty$.

But Y is stationary, so $\mathbb{P}(Y_n = y) = \pi(y)$, and

$$|\mathbb{P}(X_n = y) - \mathbb{P}(Y_n = y)| = |\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)|$$
$$= |\mathbb{P}(Z_n = y, n < T) + \mathbb{P}(Z_n = y, n \geq T) - \mathbb{P}(Y_n = y)|$$
$$= |\mathbb{P}(X_n = y, n < T) + \mathbb{P}(Y_n = y, n \geq T) - \mathbb{P}(Y_n = y)|$$
$$= |\mathbb{P}(X_n = y, n < T) - \mathbb{P}(Y_n = y, n < T)| \leq \mathbb{P}(T > n).$$

By taking $n \to \infty$, we get $\mathbb{P}(T > n) \to 0$, as $\mathbb{P}(T < \infty) = 1$. This gives the required result.

Theorem 5.3. *Let P be irreducible and aperiodic. Suppose P is null-recurrent. Then for all x, y,*

$$P^n(x, y) \to 0 \ \text{as } n \to \infty.$$

Proof: Again, consider $\tilde{\big(}P(x,y),(x',y')\big) = P(x,x')P(y,y')$. Then we have shown $\tilde{P}$ is irreducible.

We will further show that $\tilde{P}$ is transient. Indeed,

$$\sum_n \tilde{P}^n((x,x),(y,y)) = \sum_n (P^n(x,y))^2 < \infty.$$

Moreover $\tilde{P}$ is recurrent, as before. Now, since P is recurrent,

$$\nu_y(z) = \mathbb{E}_y\left[\sum_{i=1}^{T_y-1} \mathbb{1}(X_i = z)\right]$$

is an invariant measure. But as P is null-recurrent, $\mathbb{E}_y[T_y] = \infty$, so $\nu_y(I) = \infty$.

Fix $M \in \mathbb{N}$. Since $\nu_y(I) = \infty$, we can find a finite set A such that $\nu_y(A) > M$. Define

$$\mu(x) = \frac{\nu_y(x)}{\nu_y(A)}\mathbb{1}(x \in A).$$

This is a probability measure which satisfies

$$\mu P^n(z) = \sum_x \mu(x)P^n(x,z) \leq \sum_x \frac{\nu_y(x)}{\nu_y(A)}P^n(x,z) = \frac{\nu_y(z)}{\nu_y(A)}.$$

Consider $(X,Y) \sim \mu \times \delta_x$ under $\tilde{P}$, and $T = \inf\{n \geq 0 \mid (X_n,Y_n) = (x,x)\}$.

Since $\tilde{P}$ is recurrent, $\mathbb{P}(T < \infty) = 1$, so again we define

$$Z_n = \begin{cases} X_n & n < T, \\ Y_n & n \geq T. \end{cases}$$

Then $Z_n \sim \text{Markov}(\mu, P)$, similar to before. We can then see that

$$\mathbb{P}(Z_n = y) = \mu P^n(y) \leq \frac{\nu_y(y)}{\nu_y(A)} = \frac{1}{\nu_y(A)} < \frac{1}{M}.$$

Therefore, we can bound

$$P^n(x,y) = \mathbb{P}(Y_n = y) = \mathbb{P}(Y_n = y, n < T) + \mathbb{P}(Y_n = y, n \geq T)$$

$$\leq \mathbb{P}(T > n) + \mathbb{P}(Z_n = y) < \mathbb{P}(T > n) + \frac{1}{M}$$

Taking $n \to \infty$, $\mathbb{P}(T > n) \to 0$, so $\lim P^n(x,y) < 1/M$. Then taking $M \to 0$, we get the desired result.

5.2   Ergodic Theorem

Theorem 5.4. *Let P be irreducible with an invariant distribution π. Suppose $X_0 \sim \lambda$. Then with probability 1, we have, for all $x \in I$,*

$$\lim_{n \to \infty} \frac{\sum_{i=0}^{n-1} \mathbb{1}(X_i = x)}{n} \to \pi(x).$$

Proof: Since P has an invariant distribution, it is recurrent and so $T_x < \infty$.

By the strong Markov property, $(X_{T_x + n}) \sim \mathrm{Markov}(\delta_x, P)$, independent of $X_0, \ldots, X_{T_x}$. Since the limit is not affected by changing the initial distribution, it suffices to consider $\lambda = \delta_x$.

Write $\nu_n(x) = \sum_{i=0}^{n-1} \mathbb{1}(X_i = x)$, the number of visits to x by time $n-1$, and define the successive return times to x by $T_x^{(0)} = 0$, and

$$T_x^{(k+1)} = \inf\{t \geq T_x^{(k)} + 1 \mid X_t = x\}.$$

Now, these are stopping times. Define

$$S_x^{(k)} = \begin{cases} T_x^{(k)} - T_x^{(k-1)} & \text{if } T_x^{(k-1)} < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

By the strong Markov property, $(S_x^{(k)})$ are iid and have expectation

$$\mathbb{E}[S_x^{(1)}] = \mathbb{E}_x[T_x] = \frac{1}{\pi(x)}.$$

By definition, we have $T_x^{(\nu_n(x)-1)} \leq n-1$, and $T_x^{(\nu_n(x))} \geq n$. Using these two inequalities,

$$S_x^{(1)} + \cdots + S_x^{(\nu_n(x)-1)} \leq n-1,$$
$$S_x^{(1)} + \cdots + S_x^{(\nu_n(x))} \geq n,$$
$$\implies S_x^{(1)} + \cdots + S_x^{(\nu_n(x)-1)} < n \leq S_x^{(1)} + \cdots + S_x^{(\nu_n(x))}. \qquad (*)$$

Since $(S_x^{(k)})$ are iid and have finite expectation, by the strong law of large numbers,

$$\frac{S_x^{(1)} + \cdots + S_x^{(k)}}{k} \to \mathbb{E}[S_x^{(1)}],$$

with probability 1. By recurrence, $\nu_n(x) \to \infty$ as $n \to \infty$. So, dividing through $(*)$ by $(\nu_n(x))$, both the left hand side and right hand side converge to $\mathbb{E}[S_x^{(1)}] = \frac{1}{\pi(x)}$, and hence

$$\lim_{n \to \infty} \frac{n}{\nu_n(x)} \to \frac{1}{\pi(x)}, \implies \lim_{n \to \infty} \frac{\nu_n(x)}{n} = \pi(x).$$

6   Continuous Time Markov Chains

We look at Markov chains (X_t), for $t \in \mathbb{R}_+$. Let S_x be the holding time at the state x. Then, using the Markov property,

$$
\begin{aligned}
\mathbb{P}(S_x > t + s \mid S_x > s) &= \mathbb{P}(X_u = x \ \forall u \in [0, t + s] \mid X_u = x \ \forall u \in [o, s]) \\
&= \mathbb{P}(X_u = x \ \forall u \in [s, t + s] \mid X_u = x \ \forall u \in [0, s]) \\
&= \mathbb{P}(X_u = x \ \forall u \in [s, t + s] \mid X_s = x) \\
&= \mathbb{P}_x(X_u = x \ \forall u \in [0, t]) = \mathbb{P}(S_x > t).
\end{aligned}
$$

So S_x has the probability $\mathbb{P}(S_x > t + s \mid S_x > s) = \mathbb{P}(S_x > t)$, for all s, t. So S_x has the memoryless property, which implies S has the exponential distribution with some parameter.

Hence we can think of a continuous time Markov chain as a discrete Markov chain, with exponential holding time.

6.1   Poisson process

Suppose $S_1, S_2, \ldots$ are iid with distribution $\text{Exp}(\lambda)$. Then, let $J_i = \sum_{j=1}^{i} S_j$. We can then define

$$
X_t = i \text{ if } J_i \le t < J_{i+1}.
$$

We can think of the Poisson process as $X = (J_i)_i$.

Then, in any interval, the number of arrivals follows a Poisson distribution, hence the name.

Index