Markov Chains

February 28, 2023

Contents

0 Introduction **3**
0.1 Page-Rank algorithm 3

1 Markov Chains **5**

2 Simple random walks on $\mathbb{Z}^d$ **21**

Lectures

Lecture 1
Lecture 2
Lecture 3
Lecture 4
Lecture 5
Lecture 6
Lecture 7
Lecture 8
Lecture 9
Lecture 10
Lecture 11
Lecture 12

Start of lecture 1

0 Introduction

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

$$\text{past} \underset{\text{present}}{\perp} \text{future}$$

History

- Markov in 1906
- Poisson process, branching processes existed before.
  Motivation: Extend the law of large numbers to the non IID setting.
- Koluogorov in 1930: continuous time Markov processes.
- Brownian motion: fundamental object in modern probability theory.

Why Study Markov Chains?

Simplest mathematical models for random phenomena evolving in time.

- Simple: amenable to analysis - tools from probability, analysis, combinatorics.
- Applications: population growth, mathematical genetics, queuing networks, Monte Carlo simulation, ...

0.1 Page-Rank algorithm

This is an example of a simple algorithm which was previously used by search engines such as Google.

Model the web as a directed graph, $G : (V, E)$. V is the set of websites (the vertices), and $(i, j) \in E$ if and only if i contains a link to page j. Let $L(i)$ be the number of outgoing edges from i. Define

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

Now also define for $\alpha \in (0, 1)$,

$$p_{ij} = \alpha \hat{p}_{ij} + (1 - \alpha)\frac{1}{n}$$

A random surfer tosses a coin, with probability α and chooses to go to: either $\hat{p}$ or uniformly at random. We want to find the invariant distribution:

$$\pi = \pi p$$

where

$$\pi_i = \text{proportion of time spent at state } i \text{ by the surfer}$$

Once we solve for this, if $\pi_i > \pi_j$ then i is more important than j and Google ranks it higher.

1 Markov Chains

We will always denote state space by I, and it will always be finite or countable. The probability space will always be $(\Omega, \mathcal{F}, \mathbb{P})$. We will now more formally define a Markov Chain:

Definition (Markov Chain)**.** A stochastic process $(X_n)_{n\geq 0}$ is called a *Markov chain* (with values in I) if $\forall n \geq 0, \forall x_0, \ldots, x_{n+1} \in I$,

$$\mathbb{P}(\underbrace{X_{n+1} = x_{n+1}}_{\text{future}} \mid \underbrace{X_n = x_n}_{\text{present}}, \underbrace{\ldots, X_0 = x_0}_{\text{past}}) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

If $\mathbb{P}(X_{n=1} = y \mid X_n = x)$ is independent of n $\forall x, y$, then X is called *time-homogenous* (this is what we will focus on in this course). Otherwise *time-inhomogeneous.*

Define $P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x)$ for $x, y \in I$. P is called the transition matrix of the Markov chain.

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

P is called a *stochastic matrix.*

Definition. $(X_n)_{n\geq 0}$ 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, i.e.

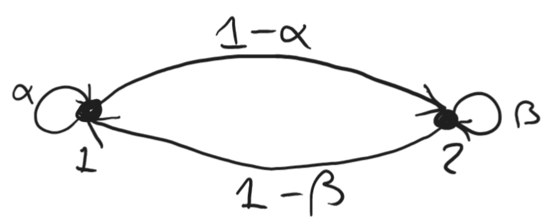
(1) $\mathbb{P}(X_0 = x) = \lambda(x)$ for all $x \in I$

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

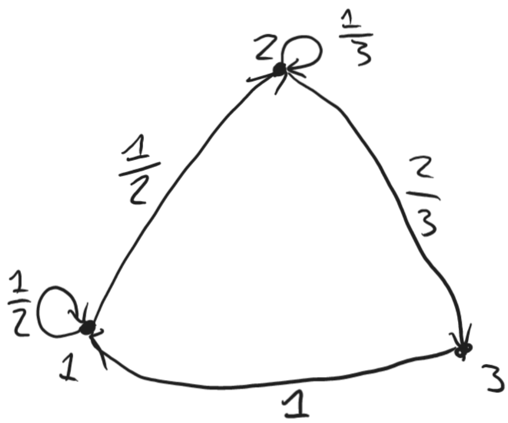
Notation. $P(x, y) = p_{xy} = p(x, y)$

Draw a diagram (directed graph), and put a directed edge between x and y $(x \to y)$ if $P(x, y) > 0$, and write the probability on top of these arrows.

- $$P = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} \qquad \alpha, \beta \in (0, 1)$$



- $$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & 0 \end{bmatrix}$$



Theorem. X is Markov(λ, P) if and only if for all $n \geq 0$ and $x_0, \dots, x_n \in I$,

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

Proof. $\Rightarrow$

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

$\Leftarrow$ for $n = 0$, $\mathbb{P}(X_0 = x_0) = \lambda(x_0)$

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

□

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

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

Definition. 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 set of random variables in I. They are independent if for all $i_1 < i_2 < \cdots < i_k$, for all k and for all $x_1, \dots, x_k$,

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

Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be 2 sequences. $X \perp Y$ if for all $k, m \in \mathbb{N}$, and for all $i_1 < \cdots < i_k$, $j_1 < \cdots < j_m$, $x_1, \dots, x_k, y_1, \dots, y_m$,

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

Start of lecture 2

Theorem (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 $x_0, x_1, \dots, x_n \in I$.

$$\begin{aligned} &\mathbb{P}(X_m = x_0, X_{m+1} = x_1, \dots, X_{m+n} = x_n \mid X_m = i) \\ &= \mathbb{1}_{i=x_0} \frac{\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n}{\mathbb{P}(X_m = i)} \end{aligned} \qquad (*)$$

$$\begin{aligned} &\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n) \\ &= \sum_{y_0, \dots, y_{m-1}} \mathbb{P}(X_0 = y_0, \dots, X_{m-1} = y_{m-1}, X_m = x_0, \dots, X_{m+n} = x_n) \\ &= \sum_{y_0, \dots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-2}, y_{m-1}) P(y_{m-1}, x_0) \cdots P(x_{n-1}, x_n) \\ &= p(x_0, x_1) \cdots P(x_{n-1}, x_n) \underbrace{\sum_{y_0, \dots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, \lambda_0)}_{= \mathbb{P}(X_m = i)} \end{aligned}$$

putting back into $(*)$ we get that

$$\mathbb{1}_{i=x_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n) \implies \text{Markov}(\delta_i, P)$$

($\mathbb{1}_{i=x_0}$ is another notation for δ_{ix_0}). Let $m \le i_1 < i_2 < \cdots < i_k$, $y_0 = i$. Then

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

□

$X \sim \text{Markov}(\lambda, P)$

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

By convention $P^0 = I$.

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

Conditional on $X_m = x$, $(X_{m+n})_{n \ge 0}$ is Markov(δ_x, P). So

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

We will write

$$p_{xy}(n) = (P^n)_{xy}$$

Let A be such an event. We will write

$$\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i)$$

Examples for P^n

Consider

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

$$P^{n+1} = P^n \cdot P = P \cdot P^n$$

So

$$p_{11}(n+1) = (1-\alpha)p_{11}(n) + p_{12}(n)$$

$$p_{11}(n) + p_{12}(n) = 1 \qquad p_{11}(0) = 1$$

$$p_{11}(n) = \begin{cases} \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \text{if } \alpha+\beta > 0 \\ 1 & \text{if } \alpha+\beta = 0 \end{cases}$$

In this simple case, it is easy to solve directly for P^n. However this is not generally the case for large matrices.

Finding eigenvalues of P is another useful method. Let P be a $k \times k$ stochastic matrix. Let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of P.

- If $\lambda_1, \dots, \lambda_k$ are all distinct, then P is diagonalisable.

$$P = UDU^{-1} \implies P^n = UD^nU^{-1}$$

$$p_{11}(n) = \alpha_1\lambda_1^n + \alpha_2\lambda_2^n + \cdots + \alpha_k\lambda_k^n$$

  $p_{11}(0) = 1$. Plug in small values of n, then solve the system to find $\alpha_1, \dots, \alpha_k$. If one of the eigenvalues is complex, say λ_{k-1}, then also its conjugate will be an eigenvalue say $\lambda_k = \overline{\lambda_{k-1}}$.

$$\lambda_{k-1} = re^{i\theta} = r\cos\theta + ir\sin\theta$$

$$\lambda_k = r\cos\theta - ir\sin\theta$$

  It becomes easier (calculations) to write the general form as

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

- If the eigenvalues are not all distinct then suppose λ appears with multiplicity 2. Then we also include the term $\alpha n + \beta)\lambda^n$ in the expression for $p_{11}(n)$. (Jordan normal form).

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

eigenvalues : $1, \frac{i}{2}, -\frac{i}{2}$.

$$p_{11}(n) = \alpha_1 + \alpha_2\left(\frac{1}{2}\right)^n \cos\left(\frac{n\pi}{2}\right) + \alpha_3\left(\frac{1}{2}\right)^n \sin\left(\frac{n\pi}{2}\right)$$

$$p_{11}(0) = 1 \qquad p_{11} = 0 \qquad p_{11}(2) = 0$$

$$p_{11}(n) = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left(\frac{4}{5}\cos\left(\frac{n\pi}{2}\right) - \frac{2}{5}\sin\left(\frac{n\pi}{2}\right)\right)$$

Start of lecture 3

Communicating classes

Definition (Communicating classes)**.** Let X be a Markov Chain with matrix P on I. Let $x, y \in I$. We say $x \to y$ (x leads to y) if

$$\mathbb{P}_x(X_m = y \text{ for some } m \geq 0) > 0.$$

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

Theorem. The following are equivalent:

(1) $x \to y$

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

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

(3) $\exists n \geq 0$ such that $p_{xy}(n) > 0$ (recall that $p_{xy}(n)$ is the (x, y) element of P^n, and is also equal to $\mathbb{P}_x(X_n = y)$)

Proof. First we prove (1) $\iff$ (3). We have:

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

If $x \to y$, then $\exists n \geq 0$ such that $\mathbb{P}_x(X_n = y) > 0$. From the definition of $\to$, we immediately have (3) $\implies$ (1).

Now we prove (2) $\iff$ (3):

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

so (2) $\iff$ (3). □

Corollary. $\leftrightarrow$ defines an equivalence relation on I.

Proof. $x \leftrightarrow x$, because $p_{xx}(0) = 1$. Transitivity: suppose $X \leftrightarrow y$ and $y \leftrightarrow z$. Then from (2), $x \leftrightarrow z$. □

Definition. 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 $y \in C$.

Definition. A matrix P (transition) is called *irreducible* if it has a single communicating class. In other words, $x \leftrightarrow y$ for all $x, y \in I$.

Definition. A state x is called *absorbing* if $\{x\}$ is a closed class. Equivalently if the Markov chain started from x then it stays at x forever.

Definition. $A \subseteq U$. $\tau_A : \Omega \to \mathbb{N} \cup \{\infty\}$

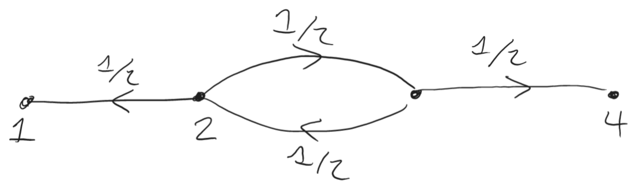
$$\tau_A = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

Convention: $\inf(\emptyset) = \infty$. τ_A is the first hitting time of A.

Denote $h_i^A = \mathbb{P}_i(\tau_A < \infty)$, $i \in I$.
$h^A : I \to [0,1]$, $(h_i^A : i \in I)$ is vector of hitting probability.

Also define $k^A : I \to \mathbb{R}_+ \cup \{\infty\}$, the mean hitting time. So

$$k_i^A = \mathbb{E}_i[\tau_A] = \sum_{n=1}^{\infty} n\mathbb{P}_i(\tau_A = n) + \underbrace{\infty \cdot \mathbb{P}_i(\tau_A = \infty)}_{0\cdot\infty=0}$$



$\mathbb{P}_2(\tau_4 < \infty) = h_2^{\{4\}}$. $\tau_4 = \tau_{\{4\}}$.

$$h_2 = \frac{1}{2}h_3 + \frac{1}{2}h_1$$
$$h_3 = \frac{1}{2} + \frac{1}{2}h_2$$
$$\implies h_2 = \frac{1}{3}$$

$(h_1 = 0,\ h_4 = 1)$ $k_2^{\{1,4\}} = \mathbb{E}_2[\tau_{\{1,4\}}]$

$$k_2 = 1 + \frac{1}{2} \cdot 0 + \frac{1}{2}k_3$$
$$k_3 = 1 + \frac{1}{2} \cdot 0 + \frac{1}{2}k_2$$
$$\implies k_2 = 2$$

$k_1 = k_4 = 0$.

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

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

The vector (h_i^A) is the minimal non-negative solution to this system.
A solution (h_i^A) is minimal if for any other non-negative solution (X_i), we have that $h_i^A \leq X_i \ \forall i$.

Proof. Clearly, if $i \in A$, then $h_I^A = 1$. Assume $i \notin A$.

$$h_i^A = \mathbb{P}_i(\tau_A < \infty)$$

$$\{\tau_A < \infty\} = \bigcup_{n=0}^{\infty} \{\tau_A = n\} = \bigcup_{n=0}^{\infty} \{X_0 \notin A, X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$$

$$\begin{aligned} \mathbb{P}_i(\tau_A < \infty) &= \sum_{n=1}^{\infty} \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) \\ &= \mathbb{P}_i(X_1 \in A) + \sum_{n=2}^{\infty} \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) \end{aligned}$$

Now compute:

$$\begin{aligned} \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) &= \sum_{j \notin A} \mathbb{P}_i(X_1 = j, X_2 \notin A, \dots, X_{n-1} \notin A, X_n \in A) \\ &= \sum_{j \notin A} \mathbb{P}_i(X_2 \notin A, \dots, X_{n-1} \notin A, X_n \in A \mid X_0 = i, X_1 = j) P(i,j) \\ &= \sum_{j \notin A} P(i,j) \mathbb{P}_j(X_1 \notin A, \dots, X_{n-2} \notin A, X_{n-1} \in A) \end{aligned}$$

Now plus back in:

$$\begin{aligned} h_i^A &= \mathbb{P}_i(X_1 \in A) + \sum_{n=1}^{\infty} \sum_{j \notin A} P(i,j) \underbrace{\mathbb{P}_j(X_1 \notin A, \dots, X_n \in A)}_{h_j^A} \\ &= \sum_{j \in A} P(i,j) \underbrace{h_j^A}_{=1} + \sum_{j \notin A} P(i,j) h_j^A \\ \implies h_i^A &= \sum_j P(i,j) h_j^A \end{aligned}$$

So h_i^A is a solution as claimed.

Now we prove minimality. Let (x_i) be another non-negative solution. Need to show that $h_i^A \le x_i$ for all i. If $i \notin A$, then

$$x_i = \sum_j P(i,j)x_j$$

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

$$x_i = \sum_{j \in A} P(i,j) + \sum_{j \notin A}\sum_{j \in A} P(i,j)P(j,k) + \sum_{j \notin A}\sum_{k \notin A} P(i,j)P(j,k)x_k$$
$$x_i = \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A}\sum_{k \notin A} P(i,j)P(j,k)x_k$$
$$x_i \ge \mathbb{P}_i(X_1 \in A) + \mathbb{P}_u(X_1 \notin A, X_2 \notin A) + \cdots + \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A)$$

(the inequality holds because the remaining terms are all non-negative since we assume that the x_i are non-negative). Rewriting:

$$x_i \ge \mathbb{P}_i(\tau_A \le n) \quad \forall n \in \mathbb{N}.$$

$$\{\tau_A \le n\} \nearrow \bigcup_n \{\tau_A \le n\} = \{\tau_A < \infty\}$$

so $\mathbb{P}_i(\tau_A \le n) \nearrow \mathbb{P}_i(\tau_A < \infty)$ hence $x_i \ge \mathbb{P}_i(\tau_A < \infty) = h_i^A$. □

Start of lecture 4

Examples

Simple random walk on $\mathbb{Z}_+$.

$$P(0,1) = 1$$

$$P(i,i+1) = p = 1 - P(i,i-1) \quad i \ge 1$$

Want to find $h_i = \mathbb{P}_i(T_0 < \infty)$.

$$h_0 = 1 \quad h_i = ph_{i+1} + qh_{i-1}$$

- $p \ne q$. Then
$$h_i = a + b\left(\frac{q}{p}\right)^i = a + (1-a)\left(\frac{q}{p}\right)^i$$
$i = 0$, $a + b = 1$. Assume $q > p$: to get non-negative and minimal solution need to take $a = 1$. So $h_i = 1$ for all $i \ge 1$. If instead we have $q < p$, then $a = 0$ implies $h_i = \left(\frac{q}{p}\right)^i$ for $i \ge 1$.

- If $p = q = \frac{1}{2}$. General solution $h_i = a + bi$, $h_0 = 1$ implies $a = 1$. For minimality need to take $b = 0$. So $h_i = 1$ for all $i \ge 1$.

Birth and death chains.

$$P(0,0)=1, \quad P(i,i+1)=p_i, \quad P(i,i-1)=q_i, \quad p_i+q_i=1$$

$$h_i=\mathbb{P}_i(T_0<\infty), \quad h_0=1$$

$$h_i=p_ih_{i+1}+q_ih_{i-1}$$

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

Set $u_i=h_i-h_{i-1}$.

$$u_{i+1}=\frac{q_i}{p_i}u_i=\cdots=\prod_{k=1}^{i}\frac{q_k}{p_k}u_1 \quad u_1=h_1-1$$

$$\begin{aligned} h_i &= \sum_{j=1}^{i}(h_j-h_{j-1})+1 \\ &= 1+\sum_{j=1}^{i}u_j \\ &= 1+u_1+\sum_{j=2}^{i}u_1\prod_{k=1}^{j-1}\frac{q_k}{p_j} \\ \implies h_i &= 1+(h_1-1)+(h_1-1)\sum_{j=2}^{i}\prod_{k=1}^{j-1}\frac{q_k}{p_k} \end{aligned}$$

Set $\gamma_j=\prod_{k=0}^{j}\frac{q_k}{p_k}$, $\gamma_0=1$. Then

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

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

$$(1-h_1)\le\frac{1}{\sum_{j=0}^{\infty}\gamma_j}$$

Minimality implies

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

- $\sum_{j=0}^{\infty}\gamma_j<\infty$, then
$$h_i=\frac{\sum_{j=1}^{\infty}\gamma_j}{\sum_{j=0}^{\infty}\gamma_j}$$
- $\sum_{j=0}^{\infty}\gamma_j=\infty$, then $h_i=1$ for all $i\ge 1$.

Mean hitting times

$A \subseteq I$, $\tau_A = \inf\{n \geq 0 : X_n \in A\}$. $k_i^A = \mathbb{E}_i[\tau_A]$.

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

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

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

$$\begin{aligned}
k_i^A &= \mathbb{E}_i[\tau_A] \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(\tau_A > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(X_0 \notin A, \ldots, X_n \notin A) \\
&= 1 + \sum_{n=1}^{\infty} \mathbb{P}_i(X_1 \notin A, \ldots, X_n \notin A) \\
&= 1 + \sum_{n=1}^{\infty} \sum_j \mathbb{P}_i(X_1 = j, X_2 \notin A, \ldots, X_n \notin A) \\
&= 1 + \sum_{n=1}^{\infty} \sum_j P(i,j) \mathbb{P}(X_1 \notin A, \ldots, X_n \notin A \mid \cancel{X_0 = i}, X_1 = j) \\
&= 1 + \sum_{n=1}^{\infty} \sum_j P(i,j) \mathbb{P}_j(X_0 \notin A, \ldots, X_{n-1} \notin A) \\
&= 1 + \sum_j P(i,j) \sum_{n=0}^{\infty} \mathbb{P}_j(X_0 \notin A, \ldots, X_n \notin A) \\
&= 1 + \sum_j P(i,j) \mathbb{E}_j[\tau_A] \\
&= 1 + \sum_j P(i,j) k_j^A \\
&= 1 + \sum_{j \notin A} P(i,j) k_j^A
\end{aligned}$$

Minimality: Let (x_i) be another non-negative solution. Then $x_i = 0$, $i \in A$. If $i \notin A$, then

$$\begin{aligned}
x_i &= 1 + \sum_{j \notin A} P(i,j) x_j \\
&= 1 + \sum_{j \notin A} P(i,j) + \sum_{j \notin A} \sum_{k \notin A} P(i,j) P(j,k) x_k \\
x_i &= 1 + \sum_{j_1 \notin A} P(i,j_1) + \cdots + \sum_{j_1, \dots, j_{n+1} \notin} P(i,j) \cdots P(j_{n-2}, j_{n-1}) + \sum \text{non-negative terms} \\
x_i &\ge 1 + \mathbb{P}_i(\tau_A > 1) + \mathbb{P}_i(\tau_A > 2) + \cdots + \mathbb{P}_i(\tau_n)
\end{aligned}$$

So $x_i \ge \sum_{k=0}^n \mathbb{P}_i(\tau_A > k)$ for all n. So

$$x_i \ge \sum_{k=0}^\infty \mathbb{P}_i(\tau_k) = \mathbb{E}_i[\tau_A] = k_i^A \qquad \square$$

Simple Markov Property

Recall that the Simple Markov property states that if $m \in \mathbb{N}$, $i \in I$, $X \sim \text{Markov}(\lambda, P)$ then conditional on $X_m = i$, $(X_{n+m})_{n \ge 0}$ is Markov(δ_i, P) and is independent of $X_0, \dots, X_m$. We would like to generalise this to a value of m that is randomly picked.

Definition. 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 $n \in \mathbb{N}$.

Example. $A \subseteq I$, $\tau_A = \inf\{n \ge 0 : X_n \in A\}$. Then $\{\tau_A = n\} = \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$ so first hitting times are always stopping times. What about last hitting time:

$$L_A = \sup\{n \le 10 : X_n \in A\}$$

Then L_A is *not* a stopping time, because for example $\{L_A = 5\}$ does not depend on $X_0, \dots, X_5$ only.

Start of lecture 5

Theorem (Strong Markov Property)**.** Let X be Markov(λ, P) and T be a stopping time. Conditional on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n \ge 0}$ is Markov(δ_i, P) and it is independent of $X_0, \dots, X_T$.

Proof. Let $x_0, \dots, x_n \in I$, $\omega \in \bigcup_k I^k$. Need to show

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

Let ω have length k. Then

$$\begin{aligned}
&\frac{\mathbb{P}(X_k = x_0, \dots, X_{k+n} = x_n, (x_0, \dots, X_k) = \omega, T = k \mid T < \infty, X_T = i)}{\mathbb{P}(T < \infty, X_T = i)} \\
&= \mathbb{P}(X_k = x_0, \dots, X_{k+n} = x_n \mid (X_0, \dots, X_k) = \omega, T = k, X_k = i) \\
&\times \frac{\mathbb{P}((X_0, \dots, X_k) = \omega, T = k, X_k = i)}{\mathbb{P}(T < \infty, X_T = i)} \qquad (*)
\end{aligned}$$

The event $\{T = k\}$ only depends on $X_0, \dots, X_k$ (T stopping time). So

$$\begin{aligned}
&\mathbb{P}(X_k = x_0, \dots, X_{k+n} \mid (X_0, \dots, X_k) = \omega, T = k, X_k = i) \\
&= \mathbb{P}(X_k = x_0, \dots, X_{k+n} = x_n \mid X_k = i)
\end{aligned}$$

by the Markov property. This is also equal to

$$\delta_{ix_0} p(x_0, x_1) \cdots p(x_{n-1}, x_n)$$

So the expression in $(*)$ is equal to

$$\mathbb{P}((X_0, \dots, X_k) = \omega, T = k \mid T < \infty, X_T = i) = \mathbb{P}((X_0, \dots, X_T) = \omega \mid T < \infty, X_T = i)$$

(ω has length k). □

Example

Consider a Markov chain with $P(0,1) = 1$, $P(i, i+1) = P(i, i-1) = \frac{1}{2}$. (like a random walk but restricted to $X \geq 0$). Let

$$T_0 = \inf\{n \geq 0 : X_n = 0\}$$

Let $h_i = \mathbb{P}_1(T_0 < \infty)$, $h_0 = 1$. What is h_1 equal to?

$$h_1 = \frac{1}{2} + \frac{1}{2} h_2$$

$$\begin{aligned}
h_2 &= \mathbb{P}_2(T_0 < \infty) \\
&= \mathbb{P}_2(T_1 < \infty, T_0 < \infty) \\
&= \mathbb{P}_2(T_0 < \infty \mid T_1 < \infty) \cdot \mathbb{P}_2(T_1 < \infty) \\
&= \mathbb{P}_2(T_0 < \infty \mid T_1 < \infty) h_1
\end{aligned}$$

Conditional on $T_1 < \infty$ ($X_{T_1} = 1$), by the Markov property $(X_{T_1+n})n \geq 0$ is Markov(δ_1, P). So (under the conditioning) 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 $\mathbb{P}_1$.

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

so $h_2 = h_1^2$. So

$$h_1 = \frac{1}{2} + \frac{1}{2} h_1^2 \implies h_1 = 1$$

Transience and recurrence

Definition. 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

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

Let

$$V_i = \sum_{l=0}^{\infty} \mathbb{1}(X_l = i) = \text{total number of visits to } i$$

We will calculate $\mathbb{P}_i(V_i > r)$ for some values of r. Let $T_i^(k)$ denote the k-th return time.

$$\mathbb{P}_(V_i > 0) = 1$$

$$\mathbb{P}_i(V_i > 1) = \mathbb{P}_i(T_i^{(1)} < \infty$$

$$\mathbb{P}_i(V_i > 2) = \mathbb{P}_i(T_i^(1) < \infty)^2$$

More formally: define $T_i^{(0)} = 0$ and for $k \geq 1$:

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

(k-th return time to i). Then

$$T_i^{(1)} = \inf\{n > 0 : X_n = i\}$$

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

Lemma. For all $r \in \mathbb{N}$, $\mathbb{P}_i(V_i > r) = f_i^r$. So V_i has a geometric distribution.

Proof. True for $r = 0$. Suppose it is true for $r \leq k$. We will prove it for $k + 1$.

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

The successive return times to i are stopping times, so conditional on $T_i^{(k)} < \infty$ (and hence $X_{T_i^{(k)}} = i$) $(X_{T_i^{(k)}+n})_{n\geq 0}$ is Markov(δ_i, P) is independent of $X_0, \dots, X_{T_i^{(k)}}$. So

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

□

Theorem. (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.

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

(a) If $f_i = 1$ then by the lemma, $\mathbb{P}_i(V_i = \infty) = 1$, so i is recurrent, so $\mathbb{E}_i[V_i] = \infty$ so $\sum_n p_{ii}(n) = \infty$.

(b) If $f_i < 1$ then by the lemma, $\mathbb{E}_i[V_i] = \frac{1}{1-f_i} < \infty$, so $\sum_n p_{ii}(n) < \infty$ so $\mathbb{P}_i(V_i < \infty) = 1$, so i is transient.

□

Theorem. Let x and y communicate. Then they are either both recurrent or both transient.

Proof. If x is recurrent, we will show y is also recurrent. $x \leftrightarrow y$ implies that there exists $m, r \geq 0$ such that $p_{xy}(m) > 0$, $p_{yx}(r) > 0$. Then

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

so

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

so y is also recurrent. □

Start of lecture 6

Corollary. All states in a communicating class are either all recurrent or all transient.

Theorem. 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 $x \to y$, $\exists m \geq 0$ such that $p_{xy}(m) < 0$, so

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

SO this shows that x is transient, contradiction. □

Theorem. A *finite* closed class is recurrent.

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

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

by the pigeonhole principle.

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

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

Theorem. Let P be irreducible and recurrent. Then for all x, y

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

Proof.

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

Suppose $\mathbb{P}_x(T_y < \infty) < 1$. Then $\mathbb{P}_x(T_y = \infty) > 0$. Pick $p_{yx}(m) > 0$, and define

$$\tilde{T}_y = \inf\{n \geq m : X_n = y\}$$

Then

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

so y is transient. □

2 Simple random walks on $\mathbb{Z}^d$

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

$$p(x, x+e_i) = p(x, x-e_i) = \frac{1}{2d} \quad \forall x \in \mathbb{Z}^d,\ \forall i = 1, \dots, d$$

where e_i is the standard basis of $\mathbb{R}^d$.

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

Proof. $d = 1$ Need to show 0 is recurrent, i.e. we want to show

$$\sum p_{00}(n) = \infty$$

$$p_{00}(n) = \mathbb{P}_0(X_n = 0)$$

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

Recall Stirling's formula: $n! \sim n^n e^{-n} \cdot e^{-n} \cdot \sqrt{2\pi n}$ so

$$\mathbb{P}_0(X_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$$

so $\sum_n p_{00}(2n) = \infty$. So simple random walk on $\mathbb{Z}$ is recurrent.

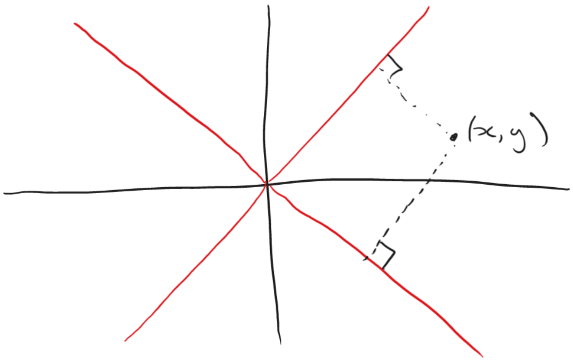
Now consider a random walk where we move right with probability p, and left with probability $q = 1 - p$, with $p \neq q$. Then

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

Since $p \neq q$, $4pq < 1$, so

$$\sum_n \frac{(4pq)^n}{\sqrt{\pi n}} < \infty$$

$d = 2$ Consider projecting the random walk as follows:



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

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

(X_n) simple random walk on $\mathbb{Z}^2$, $f(X_n) = (X_n^+, X_n^-)$. We claim that (X_n^+) and (X_n^-) are 2 independent simple random walks on $\frac{\mathbb{Z}}{2}$. Let (ξ_i) be an iid sequence

$$\mathbb{P}(\xi_i = (0,1)) = \mathbb{P}(\xi_2 = (1,0)) = \cdots = \frac{1}{4}$$

So let $X_n = \sum_{i=1}^n \xi_i$ and $\xi_i = (\xi_i^1, \xi_i^2)$.

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

So we want to show that $\xi_i^1 + \xi_i^2$ is independent of $\xi_i^1 - \xi_i^2$. This can be done by checking lots of calculations / cases. So (X_n^+) and (X_n^-) are independent. Now

$$\begin{aligned}\mathbb{P}_0(X_{2n} = 0) &= \mathbb{P}_0(X_{2n}^+ = 0, X_{2n}^- = 0) \\ &= \mathbb{P}_0(X_{2n}^+ = 0)\mathbb{P}_0(X_{2n}^- = 0) \\ &\sim \left(\frac{1}{\sqrt{n}}\right)^2 \\ &= \frac{1}{n}\end{aligned}$$

so $\sum_n \frac{1}{n} = \infty$ so recurrent.

Start of lecture 7

$d = 3$ We will prove that $\sum_n p_{00}(n) < \infty$, which will imply it is transient. Let's compute $p_{00}(2n)$. In order to be back at 0 after $2n$ steps it must make i steps to the right,

i to the left, j north and j south, k west and k east for some $i, j, k \geq 0$ and $i + j + k = n$. So

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

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

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

Let $n = 3m$. We claim that

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

To prove this, suppose the maximum over i, j, k is attained at some i, j, k with $i > j + 1$. Then

$$\binom{n}{i, j, k} < \binom{n}{i - 1, j + 1, k}$$

because $i!j! > (i - 1)!(j + 1)!$. So for $n = 3m$,

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

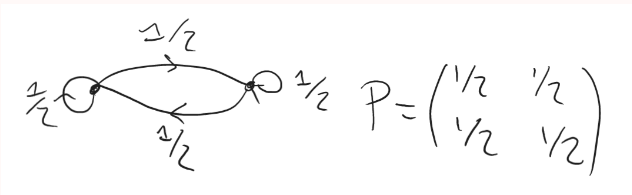
Stirling's formula gives

$$p_{00}(2n) \leq \frac{A}{n^{3/2}}$$

for some $A > 0$. So $\sum_m p_{00}(6m) < \infty$. But also $p_{00}(6m) \geq p_{00}(6m - 2) \left(\frac{1}{6}\right)^2$ and $p_{00}(6m) \geq p_{00}(6m - 4) \left(\frac{1}{6}\right)^4$. So $\sum_n p_{00}(2n) < \infty$ so it's transient. □

Invariant distribution

Definition. I discrete (countable / finite) set. $\lambda = (\lambda_i : i \in I)$ is a probability distribution if $\lambda_i \geq 0$ for all i and $\sum_{i \in I} \lambda_i = 1$.



Example.
Then $p_{11}(n) \to \frac{1}{2}$. So the Markov chain will converge to $(\frac{1}{2}, \frac{1}{2})$.

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

$$\begin{aligned}\mathbb{P}(X_1 = j) &= \sum_i \mathbb{P}(X_0 = i, X_1 = y) \\ &= \sum_i \mathbb{P}(X_1 = j \mid X_0 = i)\mathbb{P}(X_0 = 1) \\ &= \sum_i P(i,j)\pi(i)\end{aligned}$$

so $\pi(j) = \sum_i \pi(i)P(i,j)$ for all j. $\pi = \pi P$. (π as a row vector).

Definition. A probability distribution π is called *invariant* / *stationary* / *equilibrium* if $\pi = \pi P$.

Theorem. Let π be invariant and $X_0 \sim \pi$. Then $X_n \sim \pi$ for all n.

Proof. $n = 0$ is done.

$$\begin{aligned}\mathbb{P}(X_{n+1} = j) &= \sum_i \mathbb{P}(X_{n+1} = j, X_n = i) \\ &= \sum_i \mathbb{P}(X_{n+1} = j \mid X_n = i)\mathbb{P}(X_n = i) \\ &= \sum_i P(i,j)\pi(i) \\ &= \pi(j) && (\pi = \pi P)\end{aligned}$$ □

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

Proof.

$$\begin{aligned}\sum_{j\in I} \pi_j &= \sum_{j\in I} \lim_{n\to\infty} p_{ij}(n) \\ &= \lim_{n\to\infty} \sum_{j\in I} p_{ij}(n) \\ &= 1\end{aligned}$$

(we changed the order of sum and limit because the sum is finite). So π is a distribution.

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

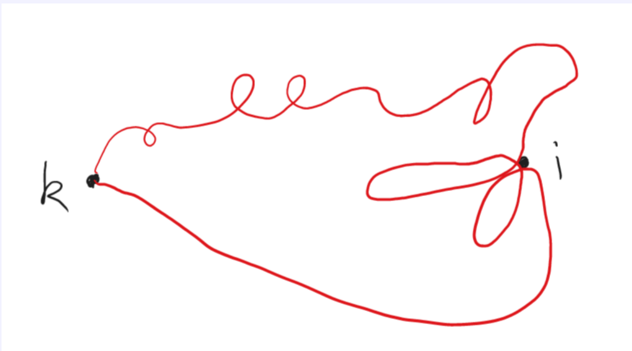
i.e. $\pi = \pi P$. □

Remark. I finite is essential: Consider a simple random walk on $\mathbb{Z}$. Then $p_{00}(2n) \sim \frac{A}{\sqrt{n}} \to 0$. Similarly $p_{0x}(n) \to \infty$ as $n \to \infty$.

Remark. P is a stochastic matrix, so 1 is always an eigenvalue. If P is irreducible, on a finite state space, then the Perron-Frobenius theorem from linear algebra ensures the existence of the invariant distribution.

Definition. $k \in I$, $T_k = \inf\{n \geq 1 : X_n = k\}$. First return time to k. $i \in I$

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



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

Start of lecture 8

Theorem. If P is irreducible and recurrent, then ν_k is an invariant measure ($\nu_k = \nu_k P$) satisfying

$$0 < \nu_k(i) < \infty \ \forall i$$

and

$$\nu_k(k) = 1$$

Proof. Obviously $\nu_k(k) = 1$. Let $i \in I$. We will prove

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

By recurrence we get $\tau_k < \infty$ with probability 1 and $X_{T_k} = k$. So

$$\begin{aligned}
\nu_k(i) &= \mathbb{E}_k\left[\sum_{l=1}^{T_k} \mathbb{1}(X_l = i)\right] \\
&= \mathbb{E}_k\left[\sum_{l=1}^{\infty} \mathbb{1}(X_l = i)\mathbb{1}(\tau_k \geq 1)\right] \\
&= \sum_{l=1}^{\infty} \mathbb{P}_k(X_l = i, \tau_k \geq l) \\
&= \sum_{l=1}^{\infty}\sum_j \mathbb{P}_k(X_l = i, X_{l-1} = j, \tau_k \geq l) \\
&= \sum_{l=1}^{\infty}\sum_j \mathbb{P}_k(X_l = i \mid X_{l-1} = j, \tau_k \geq l)\mathbb{P}_k(X_{l-1} > j, \tau_k \geq l)
\end{aligned}$$

$$\{\tau_k \geq l\} = \{T_k \leq l-1\}^c$$

By the Markov property

$$\begin{aligned}
\mathbb{P}_k(X_l = i \mid X_{l-1} = j, \tau_k \geq l) &= \mathbb{P}_k(\tau_l = i \mid X_{l-1} = j) \\
&= P(j,i)
\end{aligned}$$

so

$$\begin{aligned}
\nu_k(i) &= \sum_{l=1}^{\infty}\sum_j P(j,i)\mathbb{P}_k(X_{l-1} = j, \tau_k \geq l) \\
&= \sum_j P(j,i)\mathbb{E}_k\left[\sum_{l=1}^{\infty} \mathbb{1}(X_{l-1} = j, \tau_k \geq l)\right] \\
&= \sum_j P(j,i)\mathbb{E}_k\left[\sum_{l=0}^{\tau_k - 1} \mathbb{1}(X_l = j)\right] \\
\implies \nu_k(i) &= \sum_j \nu_k(j)P(j,i)
\end{aligned}$$

for all i. So since $\nu_k(i) > 0$:

$$\nu_k = \nu_k P^m$$

for all m

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

By irreducibility there exists m such that $p_{ki}(m) > 0$ so $\nu_k(i) > 0$.

$$\nu_k(k) = \sum_j \nu_k(j) p_{jk}(m)$$

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

Take n such that $p_{ik}(n) > 0$ (irreducibility of P) then we get

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

□

Theorem. If P is irreducible and λ is an invariant measure satisfying $\lambda_k = 1$, then

$$\lambda \geq \nu_k \quad (\forall i \ \lambda_I \geq \nu_k(i))$$

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

Proof. $\lambda_i \geq 0$ for all i.

$$\begin{aligned}
\lambda_i &= \sum_j \lambda_j P(j,i) \\
&= P(k,i) + \sum_{j_1 \neq k} P(j,i) \lambda_{j_1} \\
&= P(k,i) + \sum_{j_1 \neq k} P(k,j_1) P(j_1,1) + \sum_{\substack{j_1 \neq k \\ j_2 \neq k}} P(j_2,j_1) P(j_1,i) \lambda_{j_2} \\
&= P(k,i) + \sum_{j_1 \neq k} P(k,j_1) P(j_1,i) + \cdots + \sum_{j_1,\dots,j_{n-1} \neq q} P(k,j_{n-1}) \cdots P(j_1,i) \\
&\quad + \sum_{j_1,\dots,j_n \neq k} P(j_n,j_{n-1}) \cdots P(j_1,i) \lambda_{j_n}
\end{aligned}$$

So

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

so

$$
\begin{aligned}
\lambda_i &\geq \mathbb{E}_k\left[\sum_{l=1}^{n} \mathbb{1}(X_l = i, \tau_k \geq l+1)\right] \\
&= \mathbb{E} - k\left[\sum_{l=0}^{n} \mathbb{1}(X_l = i, l \leq \tau_k - 1)\right] \\
&= \sum_{l=0}^{n} \mathbb{E}_k[\mathbb{1}(X_l = i, l \leq \tau_k - 1)] \\
&\to \sum_{l=0}^{\infty} \mathbb{E}_k[\mathbb{1}(X_l = i, l \leq \tau_k - 1)] \\
&= \nu_k(i)
\end{aligned}
$$

$\lambda_i \geq \nu_k(i)$ for all i.
If P is recurrent, then ν_k is an invariant measure with $\nu_k(k) = 1$. So we also have that $\mu_i = \lambda_i - \nu_k(i)$ is an invariant measure, since we know $\lambda_i \geq \nu_k(i)$, hence $\mu_i \geq 0$. Need to show that $\mu_i = 0$ for all i. Let $i \in I$.

$$0 = \mu_k = \sum_j \mu_i P^m(j,k) \quad \forall m$$

$$\implies \mu_k \geq \mu_i P^m(i,k)$$

Take m such that $p_m(i,k) > 0$. Then $\mu_i = 0$. □

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

Question: When can we get an invariant distribution $\pi = \pi P$, $\sum \pi_i = 1$?

Let P be irreducible and recurrent. By the uniqueness (up to multiplication) we can get an invariant distribution (unique) if

$$\sum_{i \in I} \nu_k(i) < \infty$$

Then

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

$$\sum_{i \in I} \nu_k(i) = \sum_{i \in I} \mathbb{E}_k \left[\sum_{l=0}^{\tau_k - 1} \mathbb{1}(X_l = i) \right]$$
$$= \mathbb{E}_k \left[\sum_{l=0}^{\tau_k - 1} \sum_{i \in I} \mathbb{1}(X_l = i) \right]$$
$$= \mathbb{E}_k[\tau_k]$$

If $\mathbb{E}_k[\tau_k] < \infty$, then we can normalise.

Definition. Let P be irreducible and recurrent, $\tau_k = \inf\{n \geq 1 : X_n = k\}$, $\mathbb{P}_k(\tau_k < \infty) = 1$ for all k. We say k is *positive recurrent* if

$$\mathbb{E}_k[\tau_k] < \infty$$

We say k is *null recurrent* if

$$\mathbb{E}_k[\tau_k] = \infty$$

If k is positive recurrent, then

$$\pi_k = \frac{\nu_k(k)}{\mathbb{E}_k[\tau_k]} = \frac{1}{\mathbb{E}_k[\tau_k]}$$

Start of lecture 9

Recall that

$$\tau_k = \inf\{n \geq 1 : X_n = k\}$$

k is recurrent if and only if $\mathbb{P}_k(\tau_k < \infty) = 1$. k is positive recurrent if $\mathbb{E}_k[\tau_k] < \infty$. Otherwise k is null-recurrent.

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

(1) All states are positive recurrent

(2) Some state is positive recurrent

(3) There exists an invariant distribution π.

If any of the above holds, then

$$\pi_k = \frac{1}{\mathbb{E}_k[\tau_k]}$$

Proof.

(1) $\implies$ (2) Obvious.

(2) $\implies$ (3) Let k be the positive recurrent state.

$$\forall i \quad \nu_k(i) = \mathbb{E}_k \left[\sum_{l=0}^{\tau_k - 1} \mathbb{1}(X_l = i) \right]$$

implies k is also recurrent so by theorem from last time, $\nu_k P = \nu_k$. ν_k: invariant measure.

$$\sum_{i \in I} \nu_k(i) = \mathbb{E}_k \left[\sum_{l=0}^{\tau_k - 1} \sum_{i \in I} \mathbb{1}(X_l = i) \right] = \mathbb{E}_k[\tau_k]$$

Since k is positive recurrent, implies $\mathbb{E}_k[\tau_k] < \infty$. So we can define

$$\pi_i = \frac{\nu_k(i)}{\mathbb{E}_k[\tau_k]}$$

invariant distribution.

(3) $\implies$ (1) Let π be the invariant distribution. Let k be a state. Need to show k is positive recurrent. First show $\pi_k > 0$. Exists $i \in I$ such that $\pi_i > 0$. $\pi = \pi P = \pi P^n$ for all n.

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

Take n such that $P^n(i, k) > 0$ (irreducibility of P). Then

$$\pi_k \ge \pi_i P^n(i, k) > 0$$

Define $\lambda_i = \frac{\pi_i}{\pi_k}$: invariant measure, with $\lambda_k = 1$. So since P is irreducible $\lambda \ge \nu_k$, i.e. $\forall i \; \nu_k(i) \le \lambda_i$.

$$\mathbb{E}_K[\tau_k] = \sum_{i \in I} \nu_k(i) \le \sum_{i \in I} \lambda_i = \frac{1}{\pi_k}$$

So $\mathbb{E}_k[\tau_k] \le \frac{1}{\pi_k} < \infty$. So k is positive recurrent.

Suppose (1), (2), (3) hold. Let k be a state. Then k is positive recurrent. Define $\lambda_i = \frac{\pi_i}{\pi_k}$: invariant measure with $\lambda_k = 1$. Since P is recurrent, $\lambda = \nu_k$ (that is, $\lambda_i = \nu_k(i)$ for all i). So

$$\sum_{i \in I} \lambda_i = \sum_{i \in I} \nu_k(i)$$

so

$$\frac{1}{\pi_k} = \mathbb{E}_k[\tau_k]$$

$\square$

Corollary. P irreducible, π invariant distribution. Then for all x, y, $\nu_x(y) = \frac{\pi(y)}{\pi(x)}$.

Example. Simple random walk on $\mathbb{Z}$. $P(x, x+1) = P(x, x-1) = \frac{1}{2}$. P is recurrent ($d = 1$). Does there exist an invariant distribution? $\pi = \pi P$. Need

$$\pi_i = \frac{1}{2}\pi_{i-1} + \frac{1}{2}\pi_{i+1}$$

Then $\pi = 1$ for all i satisfies $\pi = \pi P$. Since P is recurrent, all invariant measures have to be multiples of $\pi_i = 1$ for all i. So there does not exist invariant distribution, so not positive recurrent.

Example. $\mathbb{Z}$, $P(x, x+1) = p$, $P(x, x-1) = q$, $p + q = 1$, $p > q$. Need $\pi = \pi P$, need

$$\pi_i = p\pi_{i-1} + q\pi_{i+1}$$

Solve to get

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

is an invariant measure for any choice of a, b. So no uniqueness up to multiplicative factors. Indeed, P is transient.

Example. Simple random walk on $\mathbb{Z}^3$ transient. $\pi_i = 1$ for all $i \in \mathbb{Z}^3$ invariant measure. This shows that the existence of an invariant *measure* does not imply recurrence.

Example. $\mathbb{Z}_+$, $P(x, x+1) = p$, $P(x, x-1) = q$, $p + q = 1$, $p < q$. $P(0, 1) = p$, $P(0, 0) = q$. Look for π such that $\pi = \pi P$.

$$\pi_i = p\pi_{i-1} + q\pi_{i+1} \quad i \geq 1$$

$$\pi_0 = q\pi_1 + \pi_0 q$$

$$\pi_1 = \pi_0\frac{p}{q}, \quad \pi_i = \left(\frac{p}{q}\right)^i \pi_0 \quad \forall i \geq 1$$

$p < q$, set $\pi_0 = 1 - \frac{p}{q}$, to get

$$\pi_i = \left(\frac{p}{q}\right)^i \left(1 - \frac{p}{q}\right), \quad i \geq 0$$

so there exists an invariant distribution, so positive recurrent.

Time Reversibility

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

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

and π is an invariant distribution. If P is irreducible, then $\hat{P}$ is too.

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 $y_0, \dots, y_N \in I$. Then

$$\begin{aligned}
\mathbb{P}(Y_0 = y_0, \dots, Y_N = y_N) &= \mathbb{P}(X_n = y_0, \dots, X_0 = y_N) \\
&= \mathbb{P}(X_0 = y_N, \dots, X_n = y_0) \\
&= \pi(y_N) P(y_N, y_{N-1}) \cdot P(y_1, y_0) \\
&= \pi(y_{N-1}) \hat{P}(y_{N-1}, y_N) P(y_{N-1}, y_{N-2}) \cdots P(y_1, y_0) \\
&= \cdots \\
&= \pi(y_0) \hat{P}(y_0, y_1) \cdots \hat{P}(y_{N-1}, y_N)
\end{aligned}$$

so Y is Markov$(\pi, \hat{P})$. Check $\hat{P}$ has invariant distribution π. Need to show $\pi \hat{P} = \pi$.

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

so $\pi \hat{P} = \pi$. □

Also, $\hat{P}$ is irreducible.

Start of lecture 10

Proof. Let $x, y \in I$. Need to show $\hat{P}^n(y, x) > 0$ for some n. P is irreducible, so there exists n and $x_0 = x, \dots, x_n = y$ such that

$$P(x_0, x_1) \cdots P(x_{n-1}, x_n) > 0$$

$$\begin{aligned}
\hat{P}(x_n, x_{n-1}) \cdots \hat{P}(x_1, x_0) &= \hat{P}(x_n, x_{n-1}) \cdots P(x_0, x_1) \frac{\pi(x_0)}{\pi(x_1)} \\
&= \cdots \\
&= P(x_0, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n) \frac{\pi(x_0)}{\pi(x_n)} > 0
\end{aligned}$$

□

Definition. A chain with matrix P and invariant distribution π is called *(time) reversible* if

$$\hat{P} = P$$

That is, for all x, y,

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

X is reversible if for all x, y,

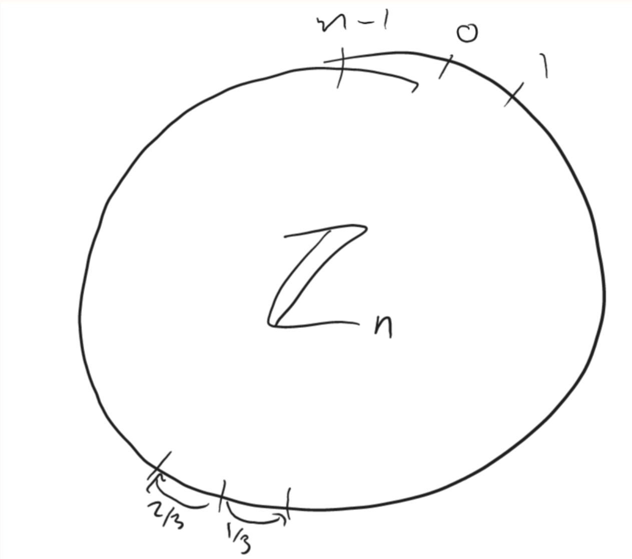
$$\boxed{\pi(x)P(x,y) = \pi(y)P(y,x)}$$

Detailed balance equation.

Equivalently, X is reversible if $\forall N \in \mathbb{N}$, when $X_0 \sim \pi$, then

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

$\mathbb{P}(X_0 = x, X_1 = y) = \pi(x)P(x,y)$.

Example. 

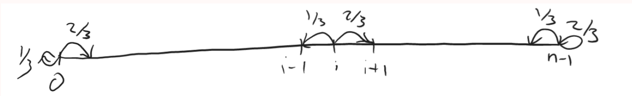
$$P(i, (i+1) \bmod n) = \frac{2}{3} \quad P(i, (i-1) \bmod n) = \frac{1}{3}$$

for all $i \in \mathbb{Z}_n$. Is P time reversible? Take $\pi_i = \frac{1}{n}$ for all i, then P is not reversible, because the Detailed Balance equation are not satisfied, for example

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

but

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

Example. 

Is this time reversible? $\lambda^i = 2^i$ invariant measure. $\pi(i) \propto 2^i$, $X_0 \sim \pi$. Check π satisfies Detailed Balance equation.

Lemma. Let μ be a distribution satisfying

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

Then μ is an invariant distribution.

Proof.

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

So $\mu = \mu P$. □

When looking for an invariant distribution, first we should look for a solution to the Detailed Balance equation. If a distribution that solves Detailed Balance equation exists, then it is an invariant distribution. If no solution to Detailed Balance equation exists, then if there exists an invariant distribution, it means the chain is not reversible.

Example. Simple random walk on a graph. $G = (V, E)$, E is edge set, V is vertex set. G is finite and connected.

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

$d(x)$ is the degree of x. Then since G is connected, P is irreducible. To find π, let's look at Detailed Balance equation

$$(x,y) \in E \quad \pi(x)\underbrace{P(x,y)}_{\frac{1}{d(x)}} = \pi(y)\underbrace{P(y,x)}_{\frac{1}{d(y)}}$$

so π must satisfy

$$\pi(x)\frac{1}{d(x)} = \pi(y)\frac{1}{d(y)} \quad \forall (x,y) \in E$$

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

$$\pi(x) = \frac{d(x)}{\sum_{y \in V} d(y)} = \frac{d(x)}{2|E|}$$

So simple random walk on G is *reversible*.

Convergence to equilibrium

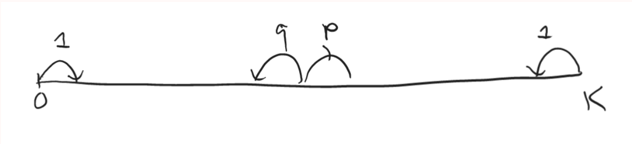
Theorem. *I finite, $i \in I$ such that $\forall j$*

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

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

P has invariant distribution π. Question: Under what conditions do we have convergence to π?

Example.



$p+q=1$, $p=\frac{2}{3}$, $q=\frac{1}{3}$ $P^n(0,0)$. If n is odd, then

$$P^n(0,0)=0$$

Definition. P transition matrix, $i \in I$. The period of i is defined

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

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

Lemma. 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. $\Leftarrow$ Obvious.

$\Rightarrow$ If $d_i = 1$ then want to show $P^n(i,i) > 0$ for all n large enough.

$$D(i) = \{n \geq 1 : P^n(i,i) > 0\}$$

Observation: if $n, m \in D(i)$ then $n + m \in D(i)$. So suffices to prove that $D(i)$ contains 2 consecutive integers. Say it contains $m, m+1$. Then by the observation it will also contain $am + b(m+1)$ for all $a, b \in \mathbb{N}$. One can check that

$$D(i) \supset \{n : n \geq m^2\}$$

Suppose $\min\{x - y : x > y, x, y \in D(i)\} = r \geq 2$. Let $n, m \in D(i)$ such that $n = m + r$. Then there exists $k = lr + s$ with $0 < s < r$, $l \in \mathbb{N}$ such that $k \in D(i)$. If there does not exist such k, then all elements of $D(i)$ would be multiples of r, and hence gcd $D(i)$ would be r, contradiction.
Let $a = (l+1)n$ and $b = (l+1)m + k$. By observation, $a, b \in D(i)$.

$$a - b = r - s < r$$

so $r = 1$, so $D(i)$ contains 2 consecutive integers.

□

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

Proof. Let $j \in I$. There exists $r, s \geq 0$ such that

$$P^r(i,j) > 0 \text{ and } P^s(j,i) > 0$$

(by irreducibility)

$$P^{n+r+s}(j,j) \geq P^s(j,i)P^n(i,i)P^r(i,j) > 0$$

for all n sufficiently large. □

Start of lecture 11

Theorem. Let P be irreducible and aperiodic with invariant distribution π. Let $X \sim \text{Markov}(\lambda, P)$. Then $\forall y$,

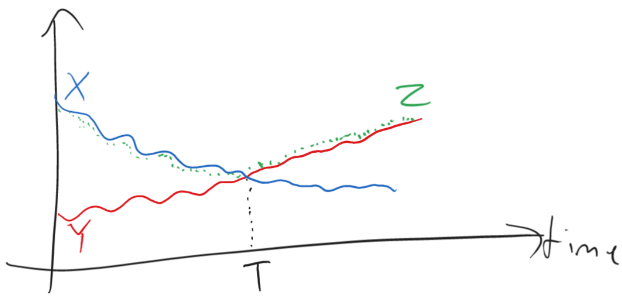
$$\mathbb{P}(X_n = y) \to \pi(y)$$

as $n \to \infty$. In particular, for all x and y,

$$P^n(x,y) \to \pi(y)$$

as $n \to \infty$. (Taking $\lambda = \delta_x$)

Proof. Coupling of Markov chains. Let $(Y_n)_{n\geq 0} \sim \text{Markov}(\pi, P)$ independent of X.



Consider $((X_n, Y_n))_{n\geq 0} \sim \text{Markov}(\lambda \times \pi, \tilde{P})$ where

$$\tilde{P}((x,y),(x',y')) = P(x,x')P(y,y')$$

We claim that $\tilde{P}$ is irreducible. So we want to show that there exist l, m such that

$$P^l(x,x') > 0 \quad \text{and} \quad P^m(y,y') > 0$$

Then note

$$P^n(x,x') \geq P^l(x,x')P^{n-l}(x',x') > 0$$

for all N sufficiently large by aperiodicity of P. Similarly

$$P^n(y,y') \geq P^m(y,y')P^{n-m}(y',y') > 0$$

for sufficiently large N. So

$$\tilde{P}^n((x,y),(x',y')) = P^n(x,x')P^n(y,y') > 0$$

for all n large enough. So $\tilde{P}$ is irreducible.
$\tilde{P}$ has invariant distribution $\tilde{\pi}(x,y) = \pi(x)\pi(y)$. So $\tilde{P}$ is positive recurrent. Fix $\alpha \in I$. Define $T = \inf\{n \geq 1 : (X_n, Y_n) = (a,a)\}$. T is a stopping time for (X,Y). $\tilde{P}$ is positive recurrent so $\mathbb{P}(T < \infty) = 1$. Define

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

Now claim that $Z \sim \text{Markov}(\lambda, P)$. Note that $Z_0 \sim \lambda$ because

$$\mathbb{P}(Z_0 = x) = \mathbb{P}(X_0 = x) = \lambda(x).$$

Let $A = \{Z_{n-1} = z_{n-1}, \ldots, Z_0 = z_0\}$. Need to show

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

So we calculate:

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

Now note that $\{T > n\} = \{T \leq n\}^c$, but $\{T \leq n\}^c$ only depends on $(X_0, Y_0), (X_n, Y_n)$ so by Strong Markov Property,

$$= P(x,y)\mathbb{P}(T > n \mid Z_n = x, A)$$

Similarly

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

So

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

So $Z \sim \text{Markov}(\lambda, P)$.
Now we want to show that $|\mathbb{P}(X_n = y) = \pi(y)| \to 0$ as $n \to \infty$. But since $Y \sim \text{Markov}(\pi, P)$ so $\mathbb{P}(Y_n = y) = \pi(y)$ for all n. So

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

Take $n \to \infty$ we get $\mathbb{P}(T > n) \to 0$ because $\mathbb{P}(T < \infty) = 1$. So $\mathbb{P}(X_n = y) \to \mathbb{P}(Y_n = y) = \pi(y)$. □

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

$$P^n(x,y) \to 0$$

as $n \to \infty$.

Proof. Consider $\tilde{P}((x,y),(x',y')) = P(x,x')P(y,y')$. As before, $\tilde{P}$ is irreducible.

- If $\tilde{P}$ transient, then

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

so $P^n(x,y) \to 0$ as $n \to \infty$.

- If $\tilde{P}$ is recurrent then

$$\nu_y(z) = \mathbb{E}_y\left[\sum_{i=0}^{\tau_y-1} \mathbb{1}(x_i = z)\right]$$

ν_y is invariant, i.e. $\nu_y P = \nu_y$. P is null-recurrent so $\mathbb{E}_y[\tau_y] = \infty$, 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

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

probability measure.

$$\begin{aligned} \mu P^n(z) &= \sum_x \mu(x) P^n(x,z) \\ &\le \sum_x \frac{\nu_y(x)}{\nu_y(A)} P^n(x,z) \\ &= \frac{\nu_y(x)}{\nu_y(A)} \end{aligned}$$

$(\nu_y P = \nu_y)$ So

$$\boxed{\mu P^n(z) \le \frac{\nu_y(z)}{\nu_y(z)}}$$

Consider $(X,Y) \sim \text{Markov}(\mu \times \delta_x, \tilde{P})$. Define $T = \inf\{n \ge 0 : (X_n, Y_n) = (x,x)\}$. $\mathbb{P}(T < \infty) = 1$, since $\tilde{P}$ is recurrent.

$$Z_n = \begin{cases} X_n & n < T \\ Y_n & n \ge T \end{cases}$$

Then $Z_n \sim \text{Markov}(\mu, P)$

$$\begin{aligned}\mathbb{P}(Z_n = y) &= \mu P^n(y) \\ &\le \frac{\nu_y(y)}{\nu_y(A)} \\ &= \frac{1}{\nu_y(A)} \\ &< \frac{1}{M}\end{aligned}$$

Need to show: $P^n(x, y) \to \infty$ as $n \to \infty$.

$$\begin{aligned}P^n(x, y) &= \mathbb{P}(Y_n = y) \\ &= \mathbb{P}(Y_n = y, n < T) + \mathbb{P}(Y_n = y, n \ge T) \\ &\le \mathbb{P}(T > n) + \mathbb{P}(Z_n = y)\end{aligned}$$

Let $n \to \infty$ then $\mathbb{P}(T > n) \to 0$ ($\mathbb{P}(T < \infty) = 1$). So

$$\lim_{n\to\infty} P^n(x, y) < \frac{1}{M}$$

Taking $M \to \infty$ finishes that proof. □

Start of lecture 12

Theorem (Ergodic theorem)**.** Let P be irreducible with an invariant distribution π. Suppose $X_0 \sim \lambda$. Then with probability 1 we have $\forall x \in I$,

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

Proof. Since P has an invariant distribution, it follows that it is recurrent and so $T_x < \infty$ with probability 1. By the strong Markov property,

$$(X_{T_x+n})_{n\ge 0} \sim \text{Markov}(\delta_x, P)$$

and is independent of $X_0, \dots, X_{T_x}$. But since $\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1}}{n}$ is not affected by changing the initial distribution, it suffices to consider $\lambda = \delta_x$. Write

$$v_n(x) = \sum_{i=0}^{n-1} \mathbb{1}(X_i = x) = \text{number of visits to } x \text{ by time } n - 1$$

Successive return times to x:

$$\begin{aligned}T_x^{(0)} &= 0 \\ T_x^{(k+1)} &= \inf\{t \ge T_x^{(k)} + 1 : X_t = x\}\end{aligned}$$

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, we see that $(S_x^{(k)})_k$ are independent identical distributions and have expectation

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

$T_x = T_x^{(1)}$.

$$T_x^{(v_n(x)-1)} \le n - 1 \tag{1}$$

$$T_x^{(v_n(x))} \ge n \tag{2}$$

(1) $\iff S_x^{(1)} + \cdots + S_x^{(v_n(x)-1)} \le n-1$ and (2) $\iff S_x^{(1)} + \cdots + S_x^{(v_n(x))} \ge n$. So

$$\boxed{S_x^{(1)} + \cdots + S_x^{(v_n(x)-1)} \le n \le S_x^{(1)} + \cdots + S_x^{(v_n(x))}} \tag{$*$}$$

Since $(S_x^{(k)})$ are IID and $\mathbb{E}[S_x^{(1)}] < \infty$ then by Strong Law of Large numbers:

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

as $k \to \infty$ with probability 1. By recurrence, $v_n(x) \to \infty$ as $n \to \infty$ so dividing ($*$) through by $v_n(x)$ we get both the LHS and RHS converge to

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

and hence

$$\lim_{n\to\infty} \frac{n}{v_n(x)} = \frac{1}{\pi(x)}$$

so

$$\lim_{n\to\infty} \frac{v_n(x)}{n} = \pi(x)$$ □

Continuous time Markov Chains (non-examinable)

We defined Markov chain as “the past and future are independent if we are given the present”. We only considered only discrete time Markov chains, but we could generalise:

- $(X_t)_{t\ge 0}$, $t \in \mathbb{R}^+$

- S_x = holding time at the state x

$$\begin{aligned}\mathbb{P}(S_x > t+s \mid S_x > s) &= \mathbb{P}(X_u = x, \forall u \in [0, t+s] \mid X_u = x, \forall u \in [0, s]) \\ &= \mathbb{P}(X_u = x, \forall u \in [s, t+s] \mid X_u = x, \forall u \in [0, s]) \\ &= \mathbb{P}(X_u = x, \forall u \in [s, t+s] \mid X_s = x) && \text{(Markov property)} \\ &= \mathbb{P}_x(X_u = x, \forall u \in [0, t]) && \text{(time-homogeneity)} \\ &= \mathbb{P}(S_x > t)\end{aligned}$$

  S_x has the property: $\mathbb{P}(S_x > t+s \mid S_x > s) = \mathbb{P}(S_x > t)$ for all s, t. So S_x has the memoryless property. Recall from IA probability that Memoryless property for a positive random variable S is equivalent to S having the exponential distribution with some parameter.

So the simplest example of a continuous time Markov chain is:
Poisson process:

$$S_1, S_2, \ldots$$

IID, $\sim \mathrm{Exp}(\lambda)$

$$J_i = \sum_{j=1}^{i} S_j$$

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