Markov Chains

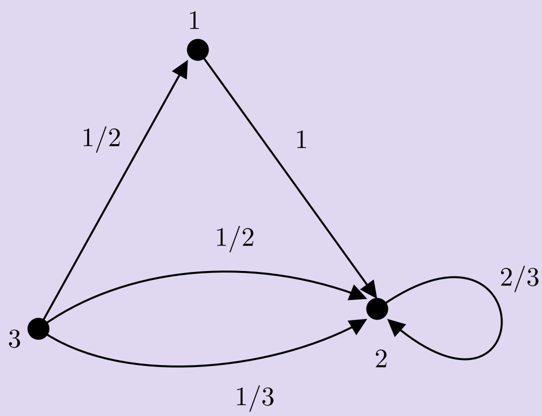
Hasan Baig

Michaelmas 2021

Contents

0 Overview

Example. $I = \{1, 2, 3\}$



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

We call P the 'transition matrix'.

1 Definitions and Basic Properties

Note. We will make the following standing assumptions:

- I is a countable set, the state space; $I = \{1, 2, \dots\}$.
- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space on which all relevant random vairables are defined.

Definition. A sequence of random vairables $(X_n)_{n=0,1,\dots}$ is a **Markov Chain** if, for $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

$$\mathbb{P}[X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n] = \mathbb{P}[X_{n+1} = i_{n+1} | X_n = i_n]$$

(conditioning if the event $X_0 = i_0, \dots, X_n = i_n$ has positive probability)
It is **homogeneous** if, for all $i, j \in I$:

$$\mathbb{P}[X_{n+1} = j | X_n = i] = \mathbb{P}[X_1 = j | X_0 = i]$$

Note. From now on, all Markov Chains are assumed homogeneous.

Definition. A Markov Chain is characterised by:

(i) the **intitial distribution**: $\lambda = (\lambda_i)_{i \in I}$ given by $\lambda_i = \mathbb{P}[X_0 = i]$
(ii) the **transition matrix**: $P = (p_{ij})_{i,j \in I}$ given by $\mathbb{P}[X_1 = j | X_0 = i]$

Remarks.

- λ is a distribution, i.e. $\lambda_i \geq 0$ for all $i \in I$ and $\sum_{i \in I} \lambda_i = 1$
- P is a stochastic matrix, i.e., $(p_{ij})_j$ is a distribution for every $i \in I$

Definition. (X_n) is a Markov Chain with initial distribution λ and transition matrix P, or (X_n) is Markov(λ, P), if (i) and (ii) hold.

Theorem. (X_n) is Markov(λ, P) iff for all $n \geq 0,\ i_0, \dots, i_n$ with $n \in I$,

$$\mathbb{P}[X_0 = i_0, \dots, X_n = i_n] = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} \tag{*}$$

Proof. Suppose (X_n) is Markov(λ, P). Then

$$\begin{aligned}\mathbb{P}[X_0 = i_0, \dots, X_n = i_n] &= \mathbb{P}[X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}] \cdot \mathbb{P}[X_0 = i_0, \dots, X_{n-1} = i_{n-1}] \\ &= p_{i_{n-1} i_n} \cdot \mathbb{P}[X_0 = i_0, \dots, X_{n-1} = i_{n-1}] \text{ by the Markov property} \\ &= p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \dots p_{i_0 i_1} \mathbb{P}[X_0 = i_0] \text{ by induction} \\ &= p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \dots p_{i_0 i_1} \lambda_{i_0}\end{aligned}$$

Conversely assume (*) holds for all n and $i_0, \dots, i_n$. For $n = 0$, $\mathbb{P}[X_0 = i_0] = \lambda_{i_0}$.
Also, by (*)

$$\begin{aligned}\mathbb{P}[X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}] &= \frac{\mathbb{P}[X_0 = i_0, \dots, X_n = i_n]}{\mathbb{P}[X_0 = i_0, \dots, X_{n-1} = i_{n-1}]} \\ &= p_{i_{n-1} i_n}\end{aligned}$$

Thus (i) and (ii) hold, i.e. (X_n) is Markov(λ, P).

Notation. Let $\delta_i = (\delta_{ij} : j = I)$ be the unit mass at $i \in I$:

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

Theorem. Let (X_n) be Markov(λ, P). Then conditional on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov (δ_i, P) and is independent of $X_0, \dots, X_m$.

Proof. It suffices to show:

(i)
$$\mathbb{P}[X_m = i_m, \dots, X_{m+n} = i_{m+n} | X_m = i] = \delta_{ii_m} p_{i_m i_{m+1}} \dots p_{i_{n+m-1} i_{n+m}}$$

(ii) For every event A determined by $X_1, \dots, X_m$ and every event B determined by $X_m, X_{m+1}, \dots$
$$\mathbb{P}[A \cap B | X_m = i] = \mathbb{P}[A | X_m = i] \cdot \mathbb{P}[B | X_m = i]$$

The previous theorem implies both for the elements:

$$A = \{X_0 = i_0, \dots, X_m = i_m\}$$

$$B = \{X_m = i_m, \dots, X_{n+m} = I_{n+m}\}$$

Indeed, after multiplying by $\mathbb{P}[X_m = i]$ the claim is

$$\mathbb{P}[X_m = i_m, \dots, X_{m+n} = i_{m+n}] = \delta_{ii_m} p_{i_m i_{m+1}} \dots p_{i_{n+m=1} i_{n+m}} \mathbb{P}[X_m = i]$$

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B | X_m = i] = \delta_{ii_m} \mathbb{P}[A] \mathbb{P}[B]$$

Now, any A and B in (i) and (ii) can be written as a countable union of elementary A and B, and hence the general claim follows by summing over the identities for elementary A and B

Notation. We regard distrbiutions and measures $(\lambda_i)_{i\in I}$ as row vectors.
Matrix multiplication:
$$(\lambda P)_j = \sum_{i\in I} \lambda_i p_{ij}$$
$$(P^2)_{ij} = \sum_{k\in I} p_{ik} p_{kj} = p_{ij}^{(2)}, \dots$$
with $P_0 = 1$ the $I \times I$ identity matrix $1_{ij} = \delta_{ij}$.
When $\lambda_i > 0$, write $\mathbb{P}_i[A] = \mathbb{P}[A | X_0 = i]$

Remark. By the Markov property, $(X_n)_{n\geq 0}$ is Markov(δ_i, P) under $\mathbb{P}_i$. (So the behaviour of (X_n) under $\mathbb{P}_i$ does not depend on λ)

Theorem. Let (X_n) be Markov(λ, P). Then for all $n, m \geq 0$:

(i)

$$\mathbb{P}[X_n = j] = (\lambda P^n)_j$$

(ii)

$$\mathbb{P}_i[X_n = j] = p_{ij}^{(n)}$$

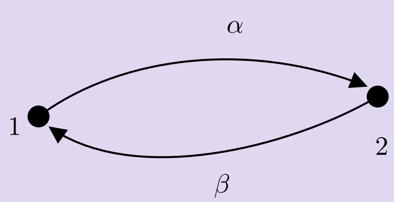
Proof.

(i)

$$\begin{aligned}\mathbb{P}[X_n = j] &= \sum_{i_0,\dots,i_{n-1}\in I} \mathbb{P}[X_0 = i_0, \dots, X_n = i_n] \\ &= \sum_{i_0,\dots,i_{n-1}\in I} \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j} \\ &= (\lambda P^n)_j\end{aligned}$$

(ii) Use the Markov property and $\lambda = \delta_i$ and (i)

Example. The general two state Markov Chain is:



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

some $\alpha, \beta \in [0, 1]$

$$P^{n+1} = P^n \cdot P \implies p_{11}^{(n+1)} = p_{12}^{(n)} \beta + p_{11}^{(n)} (1-\alpha)$$

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

Since $p_{11}^{(0)}$, this recursion relation has unique solution:

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

Method. General method to find $p_{ij}^{(n)}$ for an N state Markov Chain

- Find the eigenvalues $\lambda_1, \dots \lambda_N$ of P, i.e., roots of $\det(\lambda - P) = 0$
- If all eigenvalues are distinct, then $p_{ij}^{(n)}$ has the form:

$$p_{ij}^{(n)} = a_1 \lambda_1^n + \cdots + a_N \lambda_N^n \text{ where the } a_i \text{ are constants}$$

  If an eigenvalue λ is repeated once then the general form includes a term $(a + bn)\lambda^n$. Similar formulas hold for eigenvalues with higher multiplicities.
- As roots of a polynomial with real coefficients, any complex eigenvalues come in conjugate pairs. These are oftenbest written in terms of sin and cos

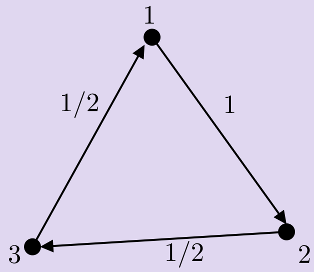
Justification: If P has distinct eigenvalues, then it can be diagonalised as

$$P = U \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} U^{-1} \implies P^n = U \begin{bmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_N^n \end{bmatrix} U^{-1}$$

$\implies p_{ij}^{(n)}$ is of the desired form.

If P has repeated eigenvalues, the more general claim can be seen from the Jordan normal form

Example.



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

What is $p_{11}^{(n)}$?
Eigenvalues:

$$0 = \det(\lambda - P) = \lambda(\lambda - \frac{1}{2})^2 - \frac{1}{4} = \frac{1}{4}(\lambda - 1)(4\lambda^2 + 1)$$

$$\implies \lambda = 1.\frac{i}{2}, -\frac{i}{2}$$

$$\implies p_{11}^{(n)} = a + b\left(\frac{i}{2}\right)^n + c\left(-\frac{i}{2}\right)^n$$

for some constant a, b, c

$$\left(\pm\frac{i}{2}\right)^n = \left(\frac{1}{2}\right)^n e^{\pm i\pi n/2} = \left(\frac{1}{2}\right)^n \left(\cos\left(\frac{1}{2}\pi n\right) \pm i\sin\left(\frac{1}{2}\pi n\right)\right)$$

$$\implies p_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left[\beta\cos\left(\frac{1}{2}\pi n\right) + \gamma\sin\left(\frac{1}{2}\pi n\right)\right]$$

for some ocnstant α, β, γ.
Note:

$$1 = p_{11}^{(0)} = \alpha + \beta$$
$$0 = p_{11}^{(1)} = \alpha + \frac{1}{2}\beta$$
$$0 = p_{11}^{(2)} = \alpha + \frac{1}{4}\beta$$

and so $\alpha = \frac{1}{5}$, $\beta = \frac{4}{5}$, $\gamma = -\frac{2}{5}$

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

2 Class Structure

Definition. For $i, j \in I$,
- i **leads to** j ($i \to j$) if $\mathbb{P}_i[X_n = j \text{ for some } n] > 0$
- i **communicates with** j ($i \leftrightarrow j$) if $i \to j$ and $j \to i$

Theorem. For $i \neq j$ the following are equivalent:
(i) $i \to j$
(ii) $p_{i_1 i_2} \dots p_{i_{n-1} i_n} > 0$ for some $i_1, \dots, i_n$ with $i_1 = i,\ i_n = j$
(iii) $p_{ij}^{(n)} > 0$ for some n

Proof. Equivalence of (i) and (iii) follows from

$$p_{ij}^{(n)} = \mathbb{P}_i[X_n = j] \leq \mathbb{P}_i[X_k = j \text{ for some } k] \leq \sum_{k=0}^{\infty} p_{ij}^{(k)}$$

Equivalence of (ii) and (iii) follows from

$$p_{ij}^{(n)} = \sum_{i_2, \dots, i_{n-1}} p_{ii_2} \dots p_{i_{n-1} j}$$

Prop. The relation is $i \leftrightarrow j$ is an equivalence relation

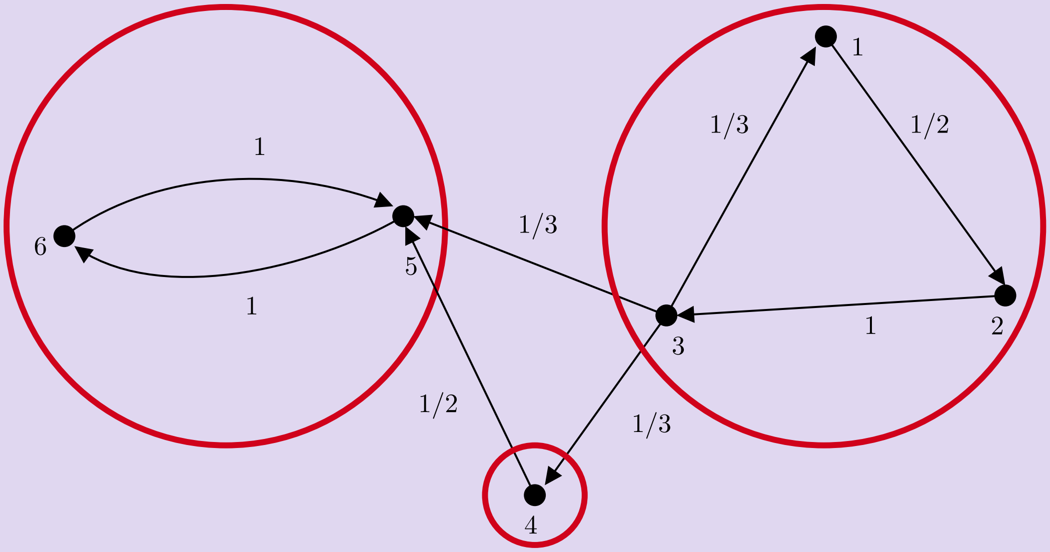
Proof. We must show that $i \leftrightarrow j$ is reflexive, symmetric and transitive. That $\leftrightarrow$ is reflexive ($i \leftrightarrow i$) and symmetric ($i \leftrightarrow j$ implies $j \leftrightarrow i$) are clear from the definition. That $\leftrightarrow$ is transitive ($i \leftrightarrow j$) and $j \leftrightarrow k$ implies $i \leftrightarrow k$) follows from (ii) of the theorem.

Definition. The equivalence classes of $\leftrightarrow$ are called **communicating classes**. The chain is irreducibleif there is only a single communicating class, i.e., $i \leftrightarrow j$ for all $i, j \in I$

Definition. A subset $C \subseteq I$ is **closed** if $i \in C,\ i \to j \implies j \in C$.
A state $i \in I$ is **absorbing** if $\{i\}$ is closed.

Example.

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



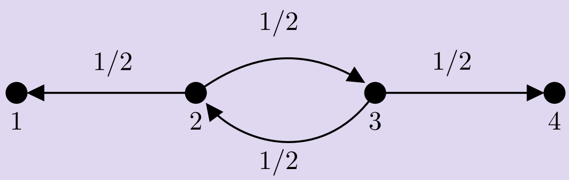
The communicating classes are $\{1,2,3\}$, $\{4\}$, $\{5,6\}$.
Only $\{5,6\}$ is closed.

3 Hitting and Absorption Probabilities

Definition. Let (X_n) be a Markov Chain.

- The **hitting time** of a set $A \subseteq I$ is the random variable $H^A : \Omega \to \{0, 1, 2, \dots\} \cup \{+\infty\}$ given by
$$H^A(\omega) = \inf\{n \geq : X_n(\omega) \in A\}, \ \inf \varnothing = +\infty$$
- The **hitting probability** of A is
-
$$h_i^A = \mathbb{P}_i[H^A < \infty] = \mathbb{P}_i[\text{hit } A]$$
If A is a closed class, h_i^A is called the **absorption probability**.
- The **mean hitting time** is the expected time to reach A.
$$k_i^A = \mathbb{E}_i[H^A] = \mathbb{E}_i[\text{time to hit } A]$$

Example.



Starting from 2, what is the probability of absorption in 4? And how long does it take until the chain is absorbed in 1 or 4?
Let $h_i = \mathbb{P}_i[\text{hit } 4]$ and $k_i = \mathbb{E}_i[\text{time to hit } 1 \text{ or } 4]$.
Note that $h_1 = 0$, $h_4 = 1$.
$$h_2 = \frac{1}{2}h_1 + \frac{1}{2}h_3$$
$$h_4 = \frac{1}{2}h_2 + \frac{1}{2}h_4$$
$k_1 = 0$, $k_4 = 0$
$$k_2 = 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3$$
$$k_3 = 1 + \frac{1}{2}k_2 + \frac{1}{2}k_4$$
$$\implies h_2 = \frac{1}{2}\left(\frac{1}{2}h_2 + \frac{1}{2}\right) = \frac{1}{4}h_2 + \frac{1}{4} = \frac{1}{3}$$
$$k_2 = 1 + \frac{1}{2}\left(1 + \frac{1}{2}k_2\right) = \frac{3}{2} + \frac{1}{4}k_2 = 2$$

Theorem. The vector of hitting probabilities $h^A = (h_i)^A_{i\in I}$ is the minimal nonnegative solution to

$$(*)\begin{cases} h_i^A = 1 & (i \in A) \\ h_i^A = \sum_{j\in I} p_{ij} h_j^A & (i \notin A) \end{cases}$$

Minimal means that if $x = (x_i)_{i\in A}$ is another solution with $x_i \geq 0$ for all $i \in I$ then $h_i^A \geq x_i$ for all $i \in I$.

Proof.

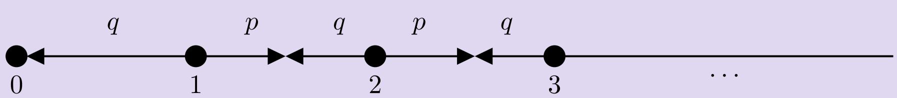
- Step 1: h^A is a solution to (*).
  If $X_0 = i \in A$ then clearly $H^A = 0$, so $h_i^A = 1$.
  If $X_0 = i \in A$, then by the Markov property,
  $$\mathbb{P}_i[H^A < \infty | X_1 = j] = \mathbb{P}_j[H^A < \infty] = h_j^A$$
  $$\begin{aligned} \implies h_i^A = \mathbb{P}_i[H^A < \infty] &= \sum_{j\in I} \mathbb{P}_i[H^A < \infty, X_1 = j] \\ &= \sum_{j\in I} \mathbb{P}_i[H^A < \infty | X_1 = j]\mathbb{P}_i[X_1 = j] \\ &= \sum_j h_j^A p_{ij} \end{aligned}$$
  $$\implies h^A \text{ is a solution to } (*)$$
- Step 2: h^A is minimal.
  Let x be any nonnegative solution to (*). If $i \in A$, clearly $h_i^A = 1 = x_i$. So suppose $i \notin A$. Then
  $$\begin{aligned} x_i = \sum_{j\in I} p_{ij} x_j = \sum_{j\in A} p_{ij} x_j + \sum_{j\notin A} p_{ij} x_j \\ = \sum_{j\in A} p_{ij} + \sum_{j\notin A} p_{ij} \left(\sum_{k\in A} p_{jk} + \sum_{k\notin A} p_{jk} x_k \right) \\ = \mathbb{P}_i[X_i \in A] + \mathbb{P}_i[X_1 \notin A, X_2 \in A] + \sum_{j\notin A}\sum_{k\notin A} p_{ij} p_{jk} x_k \end{aligned}$$
  By repeated substitution,
  $$\begin{aligned} x_i =& \mathbb{P}_i[X_1 \in A] + \mathbb{P}_i[X_1 \notin A, X_2 \in A] + \mathbb{P}[X_1 \notin A, X_2 \notin A, X_3 \in A] + \\ & \cdots + \mathbb{P}[X_1 \notin A, \ldots, X_n \in A] + \underbrace{\sum_{j_1\notin A} \cdots \sum_{j_n\notin A} p_{ij_1} p_{j_1 j_2} \ldots p_{j_{n-1} j_n} x_{j_n}}_{\geq 0 \text{ as } x \text{ non-neg.}} \end{aligned}$$
  $$\implies x_i \geq \mathbb{P}_i[H^A \leq n] \text{ for all } n$$
  $$\implies x_i \geq \lim_{n\to\infty} \mathbb{P}_i[H^A \leq n] = \mathbb{P}_i[H^A < \infty] = h_i^A$$
  $$\implies h^A \text{ is minimal}$$

Example. (continued from previous one) Recall that $h = h^A$

$$(*) \begin{cases} h_1 = h_1 \\ h_4 = 1 \\ h_2 = \frac{1}{2}h_1 + \frac{1}{2}h_3 \\ h_3 = \frac{1}{2}h_2 + \frac{1}{2}h_4 \end{cases}$$

The system (*) does not determine h_1 but by the minimality condition, we must choose $h_1 = 0$. So we find the same solution

Example (Gambler's Ruin)**.**



$p_{00} = 1$ $\qquad$ $0 < p = 1 - q < 1$

Starting with a fortune of i£, what is the probability of leaving broke? I.e., what is $h_i = \mathbb{P}_i[\text{hit } 0]$
By the theorem,

$$\begin{cases} h_0 = 1 \\ h_i = ph_{i+1} + qh_{i-1} \ (i = 1, 2, 3, \dots) \end{cases}$$

Assume $p \neq q$. The general solution to the recursion is

$$h_i = A + B\left(\frac{q}{p}\right)^i$$

If $p < q$ (most casinos): $0 \leq h_i \leq 1$ for all $i \implies B = 0,\ A = 1$, and so $h_i = 1$ for all i.
If $p > q$:

$$h_0 = 1 : h_0 = 1 \implies B = 1 - A \implies h_i = \left(\frac{q}{p}\right)^i + A\left(1 - \left(\frac{q}{p}\right)^i\right)$$

$h_i \geq 0$ for all $i \implies A \geq 0$. And minimality implies

$$A = 0 \implies h_i = \left(\frac{q}{p}\right)^i$$

If $p = q$ (fair casino), the general solution to the recursion is

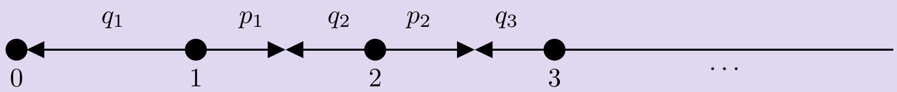
$$h_i = A + Bi$$

$$0 \leq h_i \leq 1 \implies B = 0$$

$$h_0 = 1 \implies A = 1$$

and so $h_i = 1$ for all i

Example (Birth and death chain)**.**



$h_i = \mathbb{P}_i[\text{hit } 0]$ is the extinction probability from i

$$(*) \begin{cases} h_0 = 1 \\ h_i = p_i h_{i+1} + q_i h_{i-1} \ (i = 1, 2, \dots) \end{cases}$$

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

$$\begin{aligned} p_i u_{i+1} + q_i u_i &= p_i h_i - h_{i+1} - q_i h_{i-1} + q_i h_i \\ &= (p_i + q_i - 1) h_i = 0 \end{aligned}$$

$$\implies u_{i+1} = \frac{q_i}{p_i} u_i = \underbrace{\left(\frac{q_i q_{i-1} \dots q_1}{p_i p_{i-1} \dots p_1} \right)}_{\gamma_i} = \gamma_i u_i$$

$$\implies h_i = 1 - \underbrace{(h_0 - h_i)}_{u_1 + \dots + u_i} = 1 - A(\gamma_0 + \dots + \gamma_{i-1})$$

with $A = u_1$ unknown.
If $\sum_{i=0}^{\infty} \gamma_i = \infty$: $0 \leq h_i \leq 1 \implies A = 0 \implies h_i = 1$ for all i
If $\sum_{i=0}^{\infty} \gamma_i < \infty$: minimal solution is $A = \left(\sum_{i=0}^{\infty} \gamma_i\right)^{-1}$

$$\implies h_i = \frac{\sum_{j=1}^{\infty} \gamma_j}{\sum_{j=0}^{\infty} \gamma_j}$$

Since for any i, we have $h_i < q$, the population survives with positive probability.

Theorem. The vector of mean hitting times $k^A = (k_i^A)_{i \in I}$ is the minimal solution to

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

Proof.

- Step 1: k^A satisfies $(\dagger)$.
  If $X_0 = i \in A$, then $H^A = 0$ so clearly $k_i^A = \mathbb{E}_i[H^A] = 0$
  If $X_0 = i \notin A$, then $H^A \geq 1$, so by the Markov prop.,

$$\mathbb{E}[H^A | X_1 = j] = 1 + \mathbb{E}_j[H^A] = 1 + k_j^A$$

$$k_i^A = \mathbb{E}_i[H^A] = \sum_{j \in I} \mathbb{E}_i[H^A | X_1 = j] \underbrace{\mathbb{P}_i[X_1 = j]}_{p_{ij}} = 1 + \sum_{j \notin A} p_{ij} k_j^A$$

- Step 2: k^A is minimal.
  Suppose x is any nonnegative solution to $(\dagger)$. Then $x_i = k_i^A = 0$ for all $i \in A$. For $i \notin A$,

$$\begin{aligned} x_i = 1 + \sum_{j \notin A} p_{ij} x_j &= 1 + \sum_{j \notin A} p_{ij} \left(1 + \sum_{k \notin A} p_{jk} x_k \right) \\ &= \mathbb{P}_i[H^A \geq 1] + \mathbb{P}_i[H^A \geq 2] + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} x_k \end{aligned}$$

  Again, by repeated substitution, for any n,

$$x_i = \mathbb{P}_i[H^A \geq 1] + \cdots + \mathbb{P}_i[H^A \geq n] + \underbrace{\sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} \ldots p_{j_{n-1} j_n} x_{j_n}}_{\geq 0}$$

$$\implies x_i \geq \sum_{n=1}^{\infty} \mathbb{P}_i[H^A \geq n] = \mathbb{E}_i[H^A] = k_i^A$$

  Thus k^A is the minimal solution.

4 Strong Markov Property

Definition. A random variable $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{+\infty\}$ is a **stopping time** if the event $\{T = n\}$ only depends on $X_0, \ldots, X_n$ for $n = 0, 1, 2, \ldots$

Examples.

(i) The **first passage time**

$$T_j = \inf\{n \geq 1 : X_n = j\}$$

is a stopping time since $\{T_j = n\} = \{X_1 \neq j, \ldots, X_{n-1} \neq j\}$

(ii) The hitting time H^A of a set A is a stopping time

$$\{H^A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}$$

(iii) The last exit time of a set A

$$L^A = \sup\{n \geq 0 : X_n \in A\}$$

is in general not a stopping time because $\{L^A = n\}$ depends on whether $(X_{n+m})_{m \geq 1}$ visits A or not.

Theorem (Strong Markov Property)**.** Let $(X_n)_{n \geq 0}$ be Markov(λ, P), and let T be a stopping time for (X_n). Then conditional on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n \geq 0}$ is Markov(δ_i, P) and independent of $X_1, \ldots, X_T$

Proof. Let B be an event determined by $X_0, \ldots, X_T$. Then $X \cap \{T = m\}$ is determined by $X_0, \ldots, X_m$. So by (usual) Markov property

$$\begin{aligned}&\mathbb{P}[\{X_T = j_0, \ldots, X_{T+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}] \\ &= \mathbb{P}[X_0 = j_0, \ldots, X_n = j_n]\mathbb{P}[B \cap \{T = m\} \cap \{X_T = i\}]\end{aligned}$$

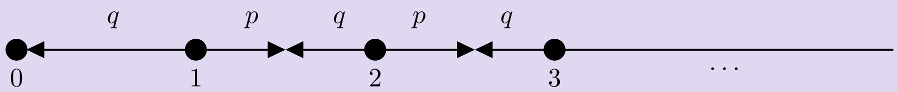
Summing over m gives

$$\begin{aligned}&\mathbb{P}[\{X_T = j_0, \ldots, X_{T+n} = j_n\} \cap B \cap \{T < \infty\} \cap \{X_T = i\}] \\ &= \mathbb{P}[X_0 = j_0, \ldots, X_n = j_n]\mathbb{P}[B \cap \{T < \infty\} \cap \{X_T = i\}]\end{aligned}$$

Dividing by $\mathbb{P}[T < \infty, X_T = i]$ (if it is positive) gives

$$\begin{aligned}&\mathbb{P}[\{X_T = j_0, \ldots, X_{T+n} = j_n\} \cap B | T < \infty, X_T = i] \\ &= \mathbb{P}[X_0 = j_0, \ldots, X_n = j_n]\mathbb{P}[B | T = m, X_T = i]\end{aligned}$$

Example (Gambler's ruin continued)**.**



$p_{00} = 1$ $\qquad$ $0 < p = 1 - q < 1$

We have previously found $\mathbb{P}_i[\text{hit } 0]$. We now find the distribution of time to hit 0 starting from 1. Let

$$H_j = \inf\{n \geq 0 : X_n = j\}$$

$$\begin{aligned} \phi(s) = \mathbb{E}_1[s^{H_0}] &= \mathbb{E}_1[s^{H_0} 1_{H_0 < \infty}] \\ &= \sum_{n=0}^{\infty} s^n \mathbb{P}[H_0 = n] \end{aligned}$$

Clain: $\mathbb{E}_2[s^{H_0}] = \phi(s)^2$

Conditional on $H_1 < \infty$ under $\mathbb{P}_2$, we can write $H_0 = H_1 + \tilde{H}_0$ where $\tilde{H}_0$ is the time it takes after H_1 to reach state 0. Since H_1 is a stopping time, by the strong Markov property at H_1, $\tilde{H}_0$ is independent of H_1 (as it only depends on $(X_{H_1+n})_{n\geq 0}$)

$$\begin{aligned} \implies \mathbb{E}_2[s^{H_0}] &= \mathbb{E}_2[s^{H_1} | H_1 < \infty] \mathbb{E}_2[s^{\tilde{H}_0} | H_1 < \infty] \mathbb{P}[H_1 < \infty] \\ &= \underbrace{\mathbb{E}_2[s^{H_1} 1_{H_1 < \infty}]}_{\mathbb{E}_1[s^{H_0}]} \underbrace{\mathbb{E}_2[s^{\tilde{H}_0} | H_1 < \infty]}_{\mathbb{E}_1[s^{H_0}]} = \phi(s)^2 \end{aligned}$$

as $\tilde{H}_0$ is conditionally independent from H_1

Example (continued)**.** Claim:

$$ps\phi(s)^2 - \phi(s) + qs = 0$$

Conditional on $X_1 = 2$, we have $H_0 = 1 + \bar{H}_0$ where $\bar{H}_0$ is the time it takes after 1 step to reach 0. By Markov property, $\bar{H}_0$ under $\mathbb{P}[\,\cdot\,|X_2 = 2]$ has the same distribution as H_0 under $\mathbb{P}_2$.

$$\begin{aligned}
\implies \phi(s) = \mathbb{E}_1[s^{H_0}] &= p\mathbb{E}_1[s^{H_0}|X_1 = 2] + q\mathbb{E}_1[s^{H_0}|X_0 = 0] \\
&= p\mathbb{E}[s^{1+\bar{H}_0}|X_1 = 2] + qs \\
&= ps\underbrace{\mathbb{E}_1[s^{\bar{H}_0}|X_1 = 2]}_{\mathbb{P}_2[s^{H_0}] = \phi(s)^2} + qs \\
&= ps\phi(s)^2 + qs
\end{aligned}$$

$$\implies \phi(0) = 0 \text{ and } \phi(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps} \text{ for } s > 0$$

Since $\phi(s) \leq 1$ and since $\phi(s)$ is continuous, only then negative root is possiblefor all $s \in [0, 1)]$

$$\begin{aligned}
\implies \phi(s) &= \frac{1 - \sqrt{1 - 4pqs^2}}{2ps} \\
&= \frac{1}{2ps}\left[1 - \left(1 + \frac{1}{2}(-4pqs^2) - \frac{1}{8}(4pqs^2)^2 + \dots\right)\right] \\
&= qs + pq^2s^3 + \dots \\
&= s\mathbb{P}[H_0 = 1] + s^2\mathbb{P}[H_0 = 2] + s^2\mathbb{P}[H_0 = 3] + \dots
\end{aligned}$$

$$\mathbb{P}[H_0 = 1] = q$$

$$\mathbb{P}[H_0 = 2] = 0$$

etc. As $s \to 1$ from below, we have $\phi(s) \to \mathbb{P}_1[H_0 < \infty]$

$$\implies \mathbb{P}_1[H_0 < \infty] = \frac{1 - \sqrt{1 - 4pq}}{2p} = \begin{cases} 1 & \text{if } p \leq q \\ \frac{q}{p} & \text{if p>q} \end{cases}$$

Also, if $p \leq q$,

$$\mathbb{E}_1[H_0] = \mathbb{E}_1[H_0 1_{H_0 < \infty}] = \lim_{\delta \uparrow 1} \phi'(s)$$

Differentiating the quadratic equation gives

$$2ps\phi(s)\phi'(s) + p\phi(s)^2 + \phi'(s) + 1 = 0$$

$$\implies \phi'(s) = \frac{p\phi(s)^2 + q}{1 - 2ps\phi(s)} \to \frac{1}{1 - 2p} = \frac{1}{q - p}$$

$$\mathbb{E}_1[H_0] = \frac{1}{q - p}$$

as $s \uparrow 1$

5 Recurrence and transcience

Definition. Let (X_n) be a Markov Chain. A state $i \in I$ is

- **recurrent** if $\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = 1$
- **transient** if $\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = 0$

First passage time to j : $T_j = \inf\{n \geq 1 : X_n = j\}$

Theorem. The following dichotomy holds:

(i) If $\mathbb{P}_i[T_i < \infty] = 1$ then i is recurrent and

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

(ii) If $\mathbb{P}_i[T_i < \infty] < 1$ then i is transcient and

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$$

In particular, every state is either recurrent or transient.

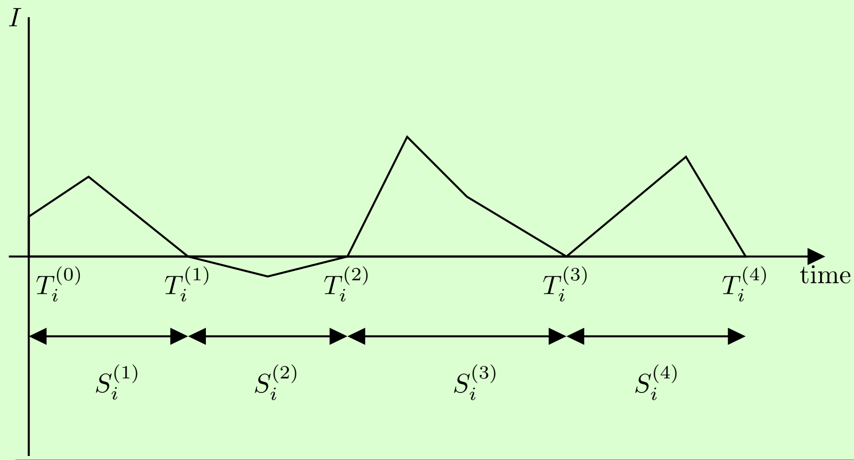
Proof.

- Step 1: Inductively, define the r-th passage time to j:

$$T_j^{(0)} = 0,\ T_j^{(1)} = T_j,\ T_j^{(r+1)} = \inf\{n \geq T_j^{(r)} + 1 : X_n = j\}$$

The length of the r-th excursion is defined by

$$S_i^r = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if } T_i^{(r-1)} < \infty \\ 0 & \text{otherwise} \end{cases}$$



Lemma. For $r = 2, 3, \ldots$, conditional on $T^{(r-1)} < \infty$, the length of the r-th excursion $S_i^{(r)}$ is independent of $\{X_m : m < T_i^{(r-1)})\}$ and

$$\mathbb{P}[S_i^{(r)} = n | T_i^{(r-1)} < \infty] = \mathbb{P}_i[T_i = n]$$

Proof.
By the strong Markov property, conditional on $T_i^{(r-1)} < \infty$, $(X_{T_i^{(r-1)}+n})_{n \geq 0}$ is Markov(δ_i, P) and is independent of $X_0, \ldots, X_{T_i^{(r-1)}}$. Now

$$S_i^{(r)} = \inf\{n \geq 1 : X_{T_i^{(r-1)}+n} = i\}$$

is the first passage time of $(X_{T_i^{(r-1)}+n})_{n \geq 0}$ to state i.

Theorem.

Proof.

- Step 2: Let V_i denote the number of visits to i:

$$V_i = \sum_{n=0}^{\infty} 1_{X_n=i}$$

Then

$$\mathbb{E}_i[V_i] = \mathbb{E}[\sum_{n=0}^{\infty} 1_{X_n=i}] = \sum_{n=0}^{\infty} \mathbb{P}_i[X_n = i] = \sum_{n=0}^{\infty} p_{ii}^n$$

Let f_i be the return probability to i:

$$f_i = \mathbb{P}_i[T_i < \infty]$$

Lemma. For $r = 0, 1, 2, \ldots$, we have $\mathbb{P}_i[V_i > r] = f_i^r$

Proof. Note that $\{V_i > r\} = \{T_i^{(r)} < \infty\}$ if $X_0 = i$.
Also note that $\mathbb{P}_i[V_i > 0] = 1$. By induction,

$$\begin{aligned}\mathbb{P}_i[V_i > r+1] &= \mathbb{P}_i[T_i^{(r+1)} < \infty] \\ &= \mathbb{P}[T_i^{(r)} < \infty, S_i^{(r+1)} < \infty] \\ &= \underbrace{\mathbb{P}_i[T_i^{(r)} < \infty]}_{f_i^r} \underbrace{\mathbb{P}[S_i^{(r+1)} < \infty | T_i^{(r) < \infty}]}_{f_i} = f_i^{r+1}\end{aligned}$$

(i) If $\mathbb{P}_i[T_i < \infty] = 1$, then by the last lemma,

$$\mathbb{P}_i[V_i = \infty] = \lim_{r \to \infty} \mathbb{P}_i[V_i > r] = 1$$

So i is recurrent and

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

(ii) If $\mathbb{P}_i[T_i < \infty] < 1$, then

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbb{E}_i[V_i] = \sum_{r=0}^{\infty} \mathbb{P}_i[V_i > r] = \sum_{r=0}^{\infty} f_i^r = \frac{1}{1-f_i} < \infty$$

So $\mathbb{P}_i[V_i = \infty] = 0$, so i is transient.

Theorem. Recurrence and transience are **class properties**: for any communicating class, either all states $i \in C$ are recurrent or all are transient

Proof. Let $i, j \in C$ and assume that i is transient. Since i and j communicatem there exist n, m s.t.

$$p_{ij}^{(n)} > 0 \text{ and } p_{ji}^{(m)} > 0$$

For all $r \geq 0$, then

$$p_{ii}^{(n+m+r)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)}$$

$$\implies \sum_{r=0}^{\infty} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{ii}^{(n+m+r)} < \infty$$

So j is transcient aswell.

Theorem. Every recurrent class is closed.

Proof. Let C be a class that is not closed, i.e., there is $i \in C, j \notin C$ and $m \geq 1$ s.t.

$$\mathbb{P}_i[X_m = j] > 0$$

Since C is a communicating class and $j \notin C$,

$$\mathbb{P}_i[\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}] = 0$$

$$\begin{aligned} \implies \mathbb{P}_i[X_n = i \text{ for infinitely many } n] &= \sum_{j \in I} \mathbb{P}_i[X_n = i \text{ for infinitely many } n, X_m = j] \\ &< \sum_{j \in I} \mathbb{P}_i[X_m = j] = 1 \end{aligned}$$

Thus i is not recurrent and since recurrence is a class property, this means that C is not recurrent (i.e. transient).

Theorem. Every finite closed class is recurrent.

Warning. Infinite closed classes may be transient

Proof. Let C be a finite closed class and suppose $X_0 \in C$

$$\begin{aligned} \implies 0 &< \mathbb{P}[X_n = i \text{ for infinitely many } n] \text{ for some } i \in C \\ &= \mathbb{P}[X_n = i \text{ for some } n] \mathbb{P}_i[X_n = i \text{ for infinitely many } i] \text{ by the strong Markov prop.} \end{aligned}$$

$$\implies \mathbb{P}_i[X_n = i \text{ for infinitely many } n] > 0$$

$$\implies i \text{ is not transient} \implies i \text{ is recurrent}$$

Corollary. Finite classes are recurrent iff closed.

Theorem. Suppose P is irreducible and recurrent. Then for all $j \in I$,

$$\mathbb{P}[T_j < \infty] = 1$$

Proof. It suffices to show that $\mathbb{P}_i[T_j < \infty] = 1$ for all $i \in I$ since then

$$\mathbb{P}[T_j < \infty] = \sum_i \mathbb{P}[X_0 = i]\mathbb{P}_i[T_j < \infty] = 1$$

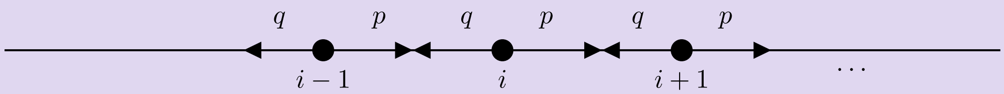
Since P is irreducible, there is m s.t. $p_{ji}^{(m)} > 0$.
Since P is recurrent,

$$\begin{aligned}
1 &= \mathbb{P}_j[X_n = j \text{ for infinitely many } n] \\
&= \mathbb{P}_j[X_n = j \text{ for some } n \geq m+1] \\
&= \sum_{k \in I} \mathbb{P}_j[X_n = j \text{ for some } n \geq m+1 | X_m = k]\mathbb{P}_j[X_m = k] \\
&= \sum_{k \in I} \mathbb{P}_k[X_n = j \text{ for some } n \geq 1] p_{jk}^{(m)} \\
&= \sum_{k \in I} \mathbb{P}_k[T_j < \infty] p_{jk}^{(m)}
\end{aligned}$$

$$\implies \mathbb{P}_i[T_j < \infty] = 1 \text{ since } \sum_k p_{jk}^{(m)} = 1 \text{ and } p_{ji}^{(m)} > 0$$

6 Recurrence and Transience of Random Walks

Example (Simple Random Walk on $\mathbb{Z}$)**.**



$p_{00}^{(2n+1)} = 0$ since the walk cannot return to 0 after an odd number of steps.

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n = \frac{(2n)!}{(n!)^2} p^n q^n$$

Stirling's formula: $n! \sim \sqrt{2\pi n} e^{-n} n^n$ where $A_n \sim B_n$ if $\lim_{n\to\infty} \frac{A_n}{B_n} = 1$

$$\implies p_{00}^{(2n)} \sim \frac{\sqrt{4\pi n}}{2\pi n} \frac{(2n)^{2n}}{n^{2n}} (pq)^n = \frac{C}{\sqrt{n}} (4pq)^n$$

Case $p = q = \frac{1}{2}$:

$$p_{00}^{(2n)} \sim \frac{C}{\sqrt{n}} \implies p_{00}^{(2n)} \geq \frac{C}{2\sqrt{n}} \text{ for } n \geq n_0$$

$$\implies \sum_{n=0}^{\infty} p_{00}^{(n)} \geq \sum_{n=n_0}^{\infty} p_{00}^{(2n)} \geq \frac{C}{2} \sum_{n=n_0}^{\infty} n^{-1/2} = \infty$$

$$\implies \text{ Random walk is recurrent}$$

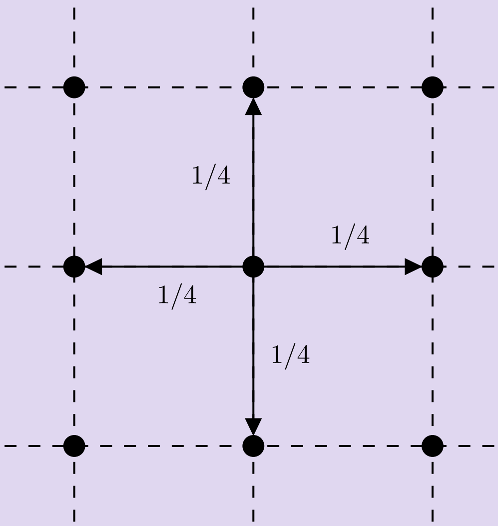
Case $p \neq q$:

$$r = 4pq < 1 \implies p_{00}^{(2n)} \leq r^n \text{ for } n \geq n_0$$

$$\implies \sum_{n=0}^{\infty} p_{00}^{(2n)} < \infty$$

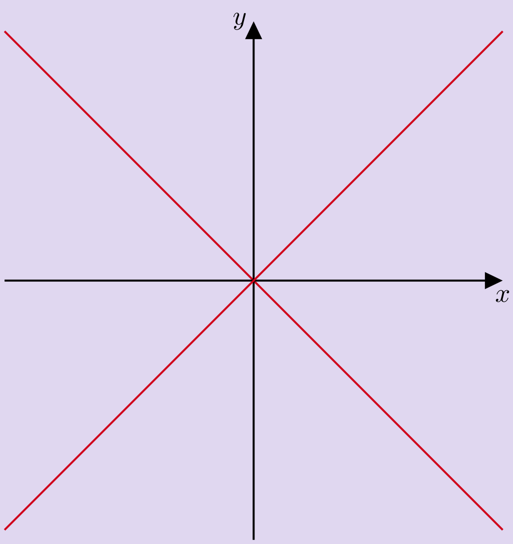
Random walk is transcient

Example (Simple Random Walk on $\mathbb{Z}^2$)**.**



$$p_{ij} = \begin{cases} \frac{1}{4} & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose $X_0 = 0$ and write $X_n^{\pm}$ for the orthogonal projections onto the lnes $y = \pm x$



Observation: $X_n^{\pm}$ are independent simple symmetric random walks on $\frac{1}{\sqrt{2}}\mathbb{Z}$ and $X_0 = 0$ iff $X_0^{\pm} = 0$

$$\implies p_{00}^{(2n)} = \left(\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right)^2 \sim \frac{C}{n}$$

since both X^+ and X^- must take $2n$ steps if X does and ust return to 0

$$\implies \sum_{n=0}^{\infty} p_{00}^{(2n)} = \infty \implies \text{ The random walk is recurrent}$$

Example (Simple Random Walk on $\mathbb{Z}^3$).

$$p_{ij} = \begin{cases} \frac{1}{6} & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

We will show the random walk is transient.
Again $p_{00}^{(2n+1)} = 0$.
All walks from 0 to 0 must take the same number of steps in direction (1,0,0) as in direction (-1,0,0), and analogously for the other two coordinates.

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

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

Fact 1. If $n = 3m$ then $\binom{n}{i\ j\ k} \leq \binom{n}{m\ m\ m}$ for i, j, k.
(suppose the maximal $\binom{n}{i\ j\ k}$ has $i > j+1$ t. Then $i!j! > (i-1)!(j+1)!$ thus $\binom{n}{i\ j\ k} < \binom{n}{i-1\ j+1\ k}$ so $\binom{n}{i\ j\ k}$ wasn't max.)
Fact 2.

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

(The LHS is the total prob. of distribution of three balls in three bins.)

$$\implies p_{00}^{(2n)} \leq \binom{2n}{n} \binom{3m}{m\ m\ m} \left(\frac{1}{3}\right)^{3m} \sim C\frac{\sqrt{n}}{\sqrt{n}^2} \cdot \frac{\sqrt{n}}{\sqrt{n}^3} = Cn^{-3/2}$$

Since $p_{00}^{(2n)} \geq \left(\frac{1}{6}\right)^2 p_{00}^{(2n-2)}$ up to changing C,

$$p_{00}^{(2n)} \leq Cn^{-3/2} \text{ for all } n$$

$$\implies \sum_n p_{00}^{(2n)} \leq C \sum_n n^{-3/2} < \infty$$

$$\implies \text{ The random walk is transient}$$

7 Invariant Measures

Definition. A measure $\lambda = (\lambda_i)_{i\in I)}$ with $\lambda_i \geq 0$ for al $i \in I$ is **invariant** (or **stationary** or **in equilibrium**) if

$$\lambda P = \lambda$$

Theorem. Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and suppose that λ is invariant for P. Then $(X_{n+m})_{n\geq 0}$ is also Markov(λ, P).

Proof.

$$\mathbb{P}[X_m = i] = (\lambda P^m)_i = \lambda_i \text{ for all } i \in I$$

so the intitial distribution of $(X_{n+m})_{n\geq 0}$ is λ
Also, conditional on $X_{n+m} = i$, by the Markov property for (X_n), X_{n+m+1} is independent $X_m,\ X_{m+1}, \dots,\ X_{n+m}$ and it has distrbution $(p_{ij})_{j\in I}$.

Theorem. Suppose I is finite. For some $i \in I$, suppose $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$, for all $j \in I$. Then $(\pi_j)_j$ is an invariant distribution

Proof. (π) is a distribution:

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

noting we can swap sum and limit as I finite.
(π) is invariant:

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_{k\in I} p_{ik}^{(n)} p_{kj} = \sum_{k\in I} \lim_{n\to\infty} p_{ik}^{(n)} p_{kj} = (\pi P)_j$$

Remark. For the simple symmetric random walk on $\mathbb{Z}^d$, we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$, for all $i, j \in \mathbb{Z}^d$. The limit 0 is invariant, but not a distribution.

Example.

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

We found earlier that

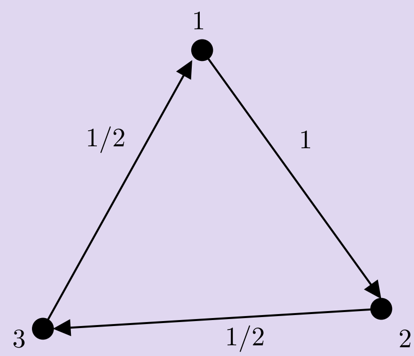
$$p_{11}^{(n)} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \text{if } \alpha+\beta > 0 \\ 1 & \text{otherwise} \end{cases}$$

So if $\alpha + \beta \notin \{0, 1\}$, we have $p_{11}^{(n)} \to \frac{\beta}{\alpha+\beta}$. Similarly,

$$P^n \to \begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\beta}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\beta}{\alpha+\beta} \end{bmatrix}$$

So by the theorem, $(\beta/(\alpha+\beta), \alpha/(\alpha+\beta))$ is an invariant distribution.

Example.



$$\pi P = \pi \iff \begin{cases} \pi_1 = \frac{1}{2}\pi_3 \\ \pi_2 = \pi_1 + \frac{1}{2}\pi_2 \\ \pi_3 = \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3 \end{cases}$$

$$\pi_1 + \pi_2 + \pi_3 = 1 \implies \pi_3 = \frac{2}{5},\ \pi_1 = \frac{1}{5},\ \pi_2 = \frac{2}{5}$$

Definition. For each state $k \in I$, let γ_i^k be the expected time spent in the state i between two visits to k:

$$\begin{aligned} \gamma_i^k &= \mathbb{E}_k \sum_{n=0}^{T_k - 1} 1_{X_n = i} \\ &= \mathbb{E}_k \sum_{n=0}^{T_k} 1_{X_n = i} \text{ if } k \neq i \end{aligned}$$

Theorem. Let P be irreducible and recurrent. Then

(i) $\gamma_k^k = 1$

(ii) $\gamma^k = (\gamma_i^k)_{i \in I}$ is an invariant measure

(iii)
$$\gamma^k P = \gamma^k$$

(iv) $0 < \gamma_i^k < \infty$ for all $i \in I$

Proof.

(i) obvious from definition.

(ii) Since P is recurrent,

(iii)
$$\mathbb{P}_k[T_k < \infty, X_0 = X_{T_k} = k] = 1$$

$$\begin{aligned}
\gamma_j^k &= \mathbb{E}_k \sum_{n=1}^{T_k} 1_{X_n = j} = j \\
&= \mathbb{E}_k \sum_{n=1}^{\infty} 1_{X_n = j \text{ and } n \leq T_k} \\
&= \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, n \leq T_k] \\
&= \sum_{i \in I} \sum_{n=1}^{\infty} \underbrace{\mathbb{P}_k[X_{n-1} = i, X_n = j, n \leq T_k]}_{\mathbb{P}_k[X_{n-1}=i, n \leq T_k]\mathbb{P}[X_n = j | X_{n-1} = i]} \\
&= \sum_{i \in I} p_{ij} \sum_{n-1}^{\infty} \mathbb{P}_k[X_{n-1} = i, n \leq T_k] \\
&= \sum_{i \in I} p_{ij} \underbrace{\mathbb{E}_k[\sum_{n-1}^{T_k - 1} 1_{X_n = i}]}_{\gamma_i^k} = \sum_{i \in I} p_{ij} \gamma_i^k = (\gamma^k P)_j
\end{aligned}$$

(iv) P irreducible $\implies \exists n, m \geq 0$ s.t. $p_{ik}^{(n)} > 0, \; p_{ki}^{(m)} > 0$

$$\implies \gamma_i^k \geq \gamma_k^k p_{ki}^{(m)} = p_{ki}^{(m)} > 0$$

$$1 = \gamma_k^k \geq \gamma_i^k p_{ik}^{(n)} \implies \gamma_i^k \leq \frac{1}{p_{ik}^{(n)}} < \infty$$

Theorem. Let P be irreducible and λ be an invariant measure for P with $\lambda_k = 1$. Then $\lambda_i \geq \gamma_i^k$ for all i. If in addition P is recurrent, then $\lambda = \gamma^k$

Proof. Since λ is invariant,

$$\begin{aligned}
\lambda_j = \sum_{i_1 \in I} \lambda_{i_1} p_{i_1 j} &= \sum_{i_1 \neq k} \lambda_{i_1} p_{i_1 j} + p_{kj} \\
&= \sum_{i_1 \neq k} \left(\sum_{i_2 \neq k} \lambda_{i_2} p_{i_2 i_1} + p_{k i_1} \right) p_{i_1 j} + p_{kj} \\
&= \dots \\
&= \underbrace{\sum_{i_1, \dots, i_n \neq k} \lambda_{i_n} p_{i_n i_{n-1}} \dots p_{i_1 j}}_{\geq 0} \\
&\quad + \left(p_{kj} + \sum_{i_1 \neq k} p_{k i_1} p_{i_1 k} + \dots + \sum_{i_1, \dots, i_{n-1} \neq k} p_{k i_{n-1}} \dots p_{i_2 i_1} p_{i_1 j} \right)
\end{aligned}$$

$\implies$ for $j \neq k$,

$$\begin{aligned}
\lambda_j &\geq \mathbb{P}_k[X_1 = j, T_k \geq 1] + \mathbb{P}_k[X_2 = j, T_k \geq 2] + \dots + \mathbb{P}_k[X_n = j, T_k \geq n] \\
&= \mathbb{E}_k \left[\sum_{m=1}^{\min(n, T_k)} 1_{X_m = j} \right] = \mathbb{E}_k \left[\sum_{m=0}^{\min(n, T_k - 1)} 1_{X_m = j} \right] \\
&\to \gamma_j^k \text{ as } n \to \infty
\end{aligned}$$

$$\implies \lambda_j \geq \gamma_j^k$$

If P is recurrent, γ^k is invariant, so $\mu = \lambda - \gamma^k \geq 0$ is invariant.
P is irreductible $\implies$ $\forall i \; \exists n$ s.t. $p_{ik}^{(n)} > 0$.

$$\implies 0 = \mu_k = \sum_{j \in I} \mu_j p_{jk}^{(n)} \geq \mu_i p_{ik}^{(n)} \implies \mu_i = 0$$

$$\implies \mu = 0 \implies \lambda = \gamma^k$$

Example. The simple symmetric random walk on $\mathbb{Z}$ is irreducible and we have also seen that it is recurrent. The measure $\pi = (\pi_i)$ where $\pi_i = 1$ for all $i \in \mathbb{Z}$ is invariant:

$$\pi = \pi P \iff \pi_i = \frac{1}{2} \pi_{i-1} + \frac{1}{2} \pi_{i+1} \checkmark$$

By the theorem, every invariant is a multiple of of π. Since $\sum_{i \in \mathbb{Z}} \pi_i = +\infty$, there is no invariant distribution.

Example. The simple symmetric random walk on $\mathbb{Z}^3$ has an invariant measure, but it is not recurrent.

Note. Recall that i is recurrent if $\mathbb{P}_i[X_n = i \text{ for inf. many } n] = 1$, or equivalently $\mathbb{P}_i[T_i < \infty] = 1$ This does not imply that the expected return time $m_i = \mathbb{E}_i[T_i]$ is finite.

Definition.
- i is **positive recurrent** if $m_i < \infty$
- i is **null recurrent** if i is recurrent but $m_i = \infty$

Theorem. Let P be irreducible. Then the following are equivalent:
(i) Every state is positive recurrent
(ii) Some state is positive recurrent
(iii) P has an invariant distribution
Moreover, when (iii) holds, then $m_i = 1/\pi_i$

Proof. (i) $\implies$ (ii): clear.
(ii) $\implies$ (iii): If i is positive recurrent, it is recurrent in particular. Therefore γ^i is invariant. Since

$$\sum_{j \in I} \gamma_j^i = m_i < \infty]$$

Thus $\pi_j = \frac{\gamma_j^i}{m_i}$ defines an invariant distribution. (iii) $\implies$ (i): first note that, for every $k \in I, \pi_k > 0$. Indeed, since π is invariant and P irreducible,

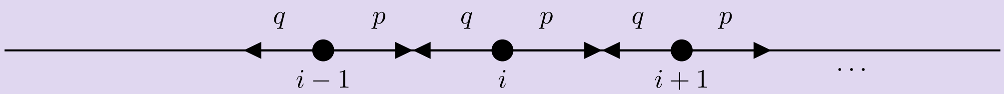
$$\pi_k = \sum_{i \in I} \pi_i p_{ik}^{(n)} > 0 \text{ for some } n$$

Now set $\lambda_i = \frac{\pi_i}{\pi_k}$. Then λ is an invariant measure with $\lambda_k = 1$. Therefore $\lambda \geq \gamma^k$.

$$\implies m_k = \sum_{i \in I} \gamma_i^k \leq \sum_{i \in I} \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty \qquad (*)$$

Thus k is positive recurrent.
Finally, knowing that P is recurrent, we have previously seen that every invariant measure λ with $\lambda_k = 1$ must satisfy $\lambda = \gamma^k$. Thus, we have equiality in (*)

Example.



$q < p$

Invariant measure equation:

$$\begin{aligned}\pi_i &= \sum_j \pi_j p_{ji} \\ &= \pi_{i-1} p + \pi_{i+1} q\end{aligned}$$

This recurrence relation has the following general solution:

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

So there is a two-parameter family of invariant measures. Uniqueness up to multiples does not hold.

8 Convergence to Equilibrium

Example.

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

$$\implies P^2 = I \implies P^{2n} = I \text{ and } P^{2n+1} = P$$

$$\implies P^n \text{ does not converge}$$

But note that P has invariant distribution $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$

Definition. A state $i \in I$ is **aperiodic** if $p_{ii}^{(n)} > 0$ for n sufficiently large. P is aperiodic if all states are aperiodic.

Lemma. Let P be irreducible and have an aperiodic state i. Then for all $j, k \in I$,

$$p_{jk}^{(n)} > 0 \text{ for } n \text{ sufficiently large}$$

In particular, all states are aperiodic.

Proof. P irreducible $\implies \exists r, s$ s.t. $p_{ji}^{(r)},\ p_{ik}^{(s)} > 0/$

$$\implies p_{jk}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ik}^{(s)} > 0 \text{ for } n \text{ sufficiently large}$$

since i is aperiodic.

Theorem. Let P be irreducible and aperiodic and suppose π is an invariant distribution for P. Let λ be any distribution, and suppose that (X_n) is Markov(λ, P). Then for all $j \in I$,

$$\mathbb{P}[X_n = j] \to \pi_j \text{ as } n \to \infty$$

In particular,

$$p_{ij}^{(n)} \to \pi_j \text{ as } n \to \infty \text{ for all } i, j$$

Proof. The proof is by **coupling**.
Let (Y_n) be Markov(π, P) and independent of (X_n). Fix a reference state $b \in I$ and set

$$T = \inf\{n > 1 : X_n = Y_n = b\}$$

Claim: $\mathbb{P}[T < \infty] = 1$.
$W_n = (X_n, Y_n)$ is a Markov Chain on state space $I \times I$ and

- transition probabilities $\tilde{p}_{(i,k)(j,l)} = p_{ij}p_{kl}$
- initial distribution $\tilde{\lambda}_{(i,k)} = \lambda_i \pi_k$

Since P is aperiodic, the lemma implies that for all $i, j, k, l \in I$,

$$\tilde{p}_{(i,k)(j,l)}^{(n)} > 0 \text{ for } n \text{ sufficiently large}$$

$$\implies \tilde{P} \text{ is irreducible}$$

$\tilde{P}$ has invariant distribution $\tilde{\pi}_{(i,k)} = \pi_i \pi_k$

$$\implies \tilde{P} \text{ is positive recurrent}$$

T is the first passage time of (W_n) to (b, b).
Since P is irreductible and recurrent,

$$\mathbb{P}[T < \infty] = 1.$$

From the claim, it follows that

$$\begin{aligned}
\mathbb{P}[X_n = i] &= \mathbb{P}[X_n = i, n < T] + \mathbb{P}[X_n = i, n \geq T] \\
&= \mathbb{P}[X_n = i, n < T] + \mathbb{P}[Y_n = i, n \geq T] \text{ by strong Markov property} \\
&= \mathbb{P}[X_n = i, n < T] + \underbrace{\mathbb{P}[Y_n = i]}_{\pi_i} - \mathbb{P}[Y_n = i, n < T]
\end{aligned}$$

$$\implies |\mathbb{P}[X_n = i] - \pi_i| = |\mathbb{P}[X_n = i, n < T] - \mathbb{P}[Y_n = i, n < T]| \leq \mathbb{P}[n < T] \to 0$$

Example (continued)**.**

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \pi = \left(\frac{1}{2}, \frac{1}{2}\right)$$

If X is Markov(δ_0, P) and Y is Markov (π, P) then with probability $\frac{1}{2}$ one has $Y_0 = 1$ but $X_0 = 0$ and X and Y will never meet

Remark. What happens when (X_n) is periodic?

Lemma. Let P be irreducible. There exists an integer $d \geq 1$ (the period) and a partition

$$I = C_0 \cup \cdots \cup C_{d-1}$$

such that, setting $C_{nd+r} = C_r$,

(i) $p_{ij}^{(n)} > 0$ only if $i \in C_r$ and $j \in C_{r+n}$ for some r

(ii) $p_{ij}^{(nd)} > 0$ for sufficiently large n, for all $i, j \in C_r$, for all r.

Proof. (In Norris' book)

Theorem. Let P be irreducible of period d with the corresponding $C_0, \ldots, C_{d-1}$ as in the lemma. Let λ be a distribution with $\sum_{i \in C_0} \lambda_i = 1$. Suppose (X_n) is Markov(λ, P).
Then for $r = 0, \ldots, d-1,\ j \in C_r$,

$$\mathbb{P}[X_{nd+r} = j] \to \frac{d}{m_j}\ (n \to \infty)$$

where m_j is the expected return time to j

Proof. (In Norris' book)

9 Time Reversal

Theorem. Let P be irreducible and have invariant distribution π. Suppose $(X_n)_{0\leq n\leq N}$ is Markov(π, P), and set $Y_n = X_{N-n}$. Then $(Y_n)_{0\leq n\leq N}$ is Markov$(\pi, \hat{P})$ where

$$\pi_j \hat{p}_{ji} = \pi_i p_{ij} \tag{*}$$

and $\hat{P}$ is irreducible with invariant distribution π

Proof. $\hat{P}$ is well-defined by (*) and is a stochastic matrix since

$$\sum_{i\in I} \hat{p}_{ji} = \frac{1}{\pi_j} \sum_{i\in I} \pi_i p_{ij} = \frac{\pi_j}{\pi_j} = 1$$

(have $\pi_j > 0$ since P is irreducible and π invariant). π is invariant for $\hat{P}$:

$$\sum_{j\in I} \pi_j \hat{p}_{ji} = \sum_{j\in I} \pi_i p_{ij} = \pi_i$$

(Y_n) is Markov$(\pi, \hat{P})$:

$$\begin{aligned} P[Y_0 = i_0, \ldots, Y_N = i_N] &= \mathbb{P}[X_0 = i_N, \ldots, X_N = i_0] \\ &= \pi_{i_N} p_{i_N i_{N-1}} \ldots p_{i_1 i_0} \\ &= \pi_{i_{N-1}} p_{i_{N-1} \hat{} i_N} p_{i_{N-1} i_{N-2}} \ldots p_{i_1 i_0} \\ &= \pi_{i_0} \hat{p}_{i_0 i_1} \ldots \hat{p}_{i_{N-1} i_N} \end{aligned}$$

$\hat{P}$ is irreducible since by irreducibility of P, for all $i, j \in I$

$$p_{i_0 i_1} \ldots p_{i_{n-1} i_n} > 0 \text{ for some } i_0, \ldots, i_n \text{ with } i_0 = i, i_n = j$$

$$\implies \hat{p}_{i_1 i_0} \ldots \hat{p}_{i_n i_{n-1}} = \frac{\pi_0}{\pi_1} p_{i_0 i_1} \ldots p_{i_{n-1} i_n} > 0$$

Definition. A stochastic matrix P and a measure λ are in **detailed balance** if

$$\lambda_i p_{ij} = \lambda_j p_{ji} \text{ for all } i, j \in I$$

Lemma. If P and λ are in detailed balance then λ is invariant for P

Proof.

$$(\lambda P)_i = \sum_{j\in I} \lambda_j p_{ji} = \sum_{j\in I} \lambda_i p_{ij} = \lambda_i$$

Definition. Let P be irreducible and (X_n) be Markov(λ, P). Then (X_n) is **reversible** if, for all N, $(X_{N-n})_{0\leq n\leq N}$ is also Markov(λ, P)

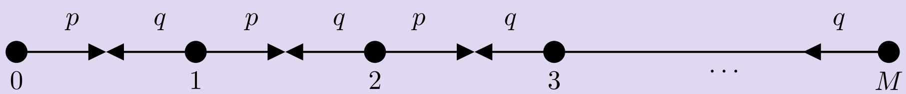
Theorem. Let P be irreducible and let λ be a distribution. Suppose (X_n) is Markov(λ, P). Then the following are equivalent:

(i) (X_n) is reversible

(ii) P and λ are in detailed balance

Proof. Both (i) and (ii) imply that λ is invariant. By the previous theorem, thus both are equivalent to $P = \hat{P}$

Examples.



$$0 < p = 1 - q < 1$$

λ and P are in detailed balance

$$\iff \lambda_i p_{i,i+1} = \lambda_{i+1} p_{i+1,i} \text{ for } i = 0, \dots, M-1$$
$$\iff \lambda_i p = \lambda_{i+1} q$$

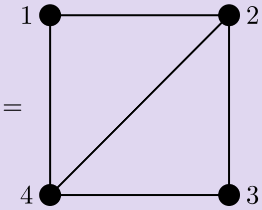
$$\iff \lambda_i = C \left(\frac{p}{q}\right)^i \text{ for some constant } C$$

Thus

$$\pi_i = \frac{\lambda_j}{\sum_j \lambda_j} = \tilde{C} \left(\frac{p}{q}\right)^i$$

for some suitable $\tilde{C}$ is also invariant distribution. Hence the chain started from π is reversible

Example (Random walk on a graph)**.**

$G =$ 

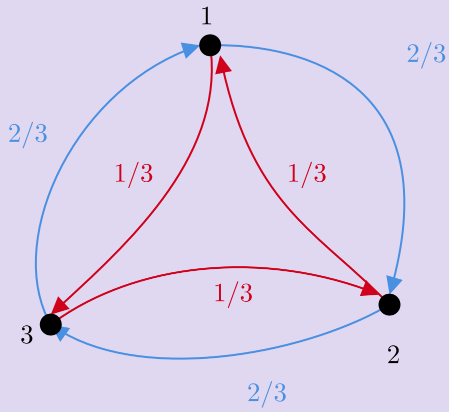
Let v_i be the valency (or degree) of vertex i, i.e., the number of edges incident to i

$$p_{ij} = \begin{cases} 1/v_i & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{otherwise} \end{cases}$$

G connected $\implies$ P irreducible. P is in detailed balance with $v = (v_i)_{i \in I}$:

$$v_i p_{ij} = 1 = v_j p_{ji}$$

Example.



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

$$\pi = (1/3, 1/3, 1/3)$$

$$\hat{P} = P^T$$

10 Ergodic Theorem

Theorem (Strong Law of Large Numbers)**.** Let $(Y_i)_{i=0,\dots}$ be a sequence of i.i.d non-negative random variables with $\mathbb{E}[Y_i] = \mu \in [0, \infty]$. Then

$$\mathbb{P}[\frac{Y_1 + \cdots + Y_{n-1}}{n} \to \mu \text{ as } n \to \infty] = 1$$

Notation. Let $V_i(n) = \sum_{k=1}^{n-1} 1_{X_k} = i =$ number of visits to i before n.

Theorem (Ergodic Theorem)**.** Let P be irreducible andlet λ be any distribution. If (X_n) is Markov(λ, P) then

$$\mathbb{P}[\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty] = 1$$

In particular, if P is positive recurrent (with invariant distribution $\pi_i = 1/m_i$) then

$$\mathbb{P}[\frac{V_i(n)}{n} \to \pi_i \text{ as } n \to \infty] = 1$$

Proof. (i) Case 1: P is transient. In this case, $\mathbb{P}[V_i < \infty] = 1$, $V_i = \sum_{k=0}^{\infty} 1_{X_n} = i$ is the total number of visits

$$\implies \mathbb{P}[\frac{V_i(n)}{n} \leq \frac{V_i}{n} \to 0 = \frac{1}{m_i}] = 1$$

as claimed

(ii) P is recurrent and $\lambda = \delta_i$, i.e.,

$$\mathbb{P}_i[\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty] = 1$$

Let $S_i^{(r)}$ be the rth excursion length between visits to i. We have seen that:

- the $S_i^{(1)}, S_i^{(2)}, \ldots$ are independent
- the $S_i^{(r)}$ are identically distributed with $\mathbb{E}[S_i^{(r)}] = m_i$

$$\implies \mathbb{P}_i[\frac{S_i^{(1)} + \cdots + S_i^{(n)}}{n} \to m_i \text{ as } n \to \infty] = 1$$

To get the claim, note:

$$S_i^{(1)} + \cdots + S_i^{(V_i(n))} \geq n$$

$$S_i^{(1)} + \cdots + S_i^{V_i(n)-1} \leq n - 1$$

$$\implies \frac{S_i^{(1)} + \cdots + S_i^{(V_i(n))}}{V_i(n)} \geq \frac{n}{V_i(n)}$$

$$\implies \frac{S_i^{(1)} + \cdots + S_i^{(V_i(n))}}{V_i(n-1)} \leq \frac{n}{V_i(n)}$$

Since $\mathbb{P}[V_i(n) \to \infty] = 1$ by (*), thus

$$\mathbb{P}[\frac{n}{V_i(n)} \to m_i] = 1$$

(iii) P is recrurrent with a general initial distribution λ. By recurrence, $\mathbb{P}[T_i < \infty] = 1$. By the strong Markov property $(X_{T_I+n})_{n\geq 0}$ is Markov(δ_i, P) and independent of $X_0, \ldots, X_{T_i}$. The general claim now follows since $\lim_n \frac{V_i(n)}{n}$ remains the same if $(X_n)_{n\geq 0}$ is replaces by $(X_{T_i+n})_{n\geq 0}$

Corollary. In the positive recurrent case, for any bounded function $f : I \to \mathbb{R}$,

$$\mathbb{P}[\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \bar{f} \text{ as } n \to \infty] = 1$$

where

$$\bar{f} = \sum_{i \in I} \pi_i f_i$$

Proof. WLOG, $|f| \leq 1$. Then for any $J \subset I$,

$$\begin{aligned} |\frac{1}{n}\sum_{k=1}^{n-1} f(X_k) - \bar{f}| &= |\sum_{i\in I}(\frac{V_i(n)}{n} - \pi_i)f_i| \\ &\leq \sum_{i \in J}|\frac{V_i(n)}{n} - \pi_i| + \sum_{i \notin J}(\frac{V_i(n)}{n} + \pi_i) \\ &\leq 2\sum_{i\in J}|\frac{V_i(n)}{n} - \pi_i| + 2\sum_{i \notin J}\pi_i \end{aligned}$$

Choose $J \subset I$ finite such that $\sum_{i \notin J} \pi_i < \varepsilon$. Choose $N = N(\omega)$ large enough such that

$$\mathbb{P}[\sum_{i \in J}|\frac{V_i(n)}{n} - \pi_i| < \varepsilon \text{ for } n \geq N] = 1$$

Therefore

$$\mathbb{P}[|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \bar{f}| < 4\varepsilon \text{ for } n \geq N] = 1$$

Question: From the observations of a Markov Chain, how can you estimate the transition matrix? Suppose $(X_i)_{i=0,\dots,n}$ is given (observations). For any $\tilde{P} = (\tilde{p}_{ij})$, define

$$\begin{aligned} l(\tilde{p}) &= \log\big(\tilde{p}_{x_0x_1}\tilde{p}_{x_1x_2}\dots\tilde{p}_{x_{n-1}x_n}\big) \\ &= \sum_{i,j\in I} N_{ij}(n)\tilde{p}_{ij} \end{aligned}$$

where

$$N_{ij}(n) = \sum_{m=0}^{m-1} 1_{\{X_m=i, X_{m+1}=j} = \text{number transitions from } i \text{ to } j$$

The maximum likelihood estimator $\hat{P} = \hat{P}(n)$ is the maximiser of $l = l_n$. We can show (using Lagrange multipliers)

$$\hat{p}_{ij}(n) = \frac{N_{ij}(n)}{V_i(n)}$$

where $V_i(n) = \sum_{k=0}^{n-1} 1_{X_k=i}$

Claim. If P is positive recrurrent, then

$$\mathbb{P}[\hat{p}_{ij}(n) \to p_{ij} \text{ as } n \to \infty] = 1$$

Proof. $N_{ij} = \sum_{m=1}^{V_i} Y_m$ where $Y_m = 1$ if the m-th transition is from i is to j and $Y_m = 0$ otherwise. By the strong Markov property, the Y_i are i.i.d with mean p_{ij} and independent from $V_i(n)$. MArkov Chain is positive recurrent so

$$\mathbb{P}[V_i(n) \to \infty \text{ as } n \to \infty] = 1$$

Strong law of large numbers gives

$$\mathbb{P}[\hat{p}_{ij}(n) = \frac{\sum_{k=1}^{V_i(n)} Y_k}{V_i(n)} \to p_{ij} \text{ as } n \to \infty] = 1$$

Outlook: for an aperiodic irreducible finite state Markov Chain, we have seen that

$$\mathbb{P}[X_n = i] \to \pi_i \quad (n \to \infty)$$

Thus, conversely, to sample from a given distribution π (on say N states), one may try to find a Markov Chain as above with π as its invariant distribution, and then run it for a long time (Markov Chain Monte Carlo - MCMC) - Metropolis and Ulam.

There are different ways to find such a Markov Chain. The most famous is the Metropolis algorithm. (Metropolis, Rosenbluth, Teller & Teller (1953))

Question of theoretical and practical relevance: how fast is "$n \to \infty$"? E.g.

$$\min\{n : \sum_i |\mathbb{P}[X_n = i] - \pi_i| < \varepsilon\} =?$$

Depends very much on the particular structure of the Markov Chain. It is a subject of current reearch interest