Part IB — Markov Chains

Based on lectures by Dr P. Sousi and notes by thirdsgames.co.uk

Michaelmas 2022

Contents

§0 History/motivation?

Markov chains are random processes (sequences of rvs) that retain no memory of the past. So the past is independent of the future

Motivation: extend the law of large numbers to the non iid setting

Kolmegorov in 1930 extended them to continuous time Markov processes.

Brownian motion: fundamental object in modern probability theory.

Why study MC? They are one of the simplest mathematical models for various random phenomena that evolve in time. They are simple as they do not rely on the past which makes them amenable to analysis, so we can use tools from probability, analysis, combinatorics.

Applications: population growth (branching processes), mathematical genetics, queuing networks, Monte carlo simulation, ...

Page-Rank algorithm

Model the ... as a directed graph, G: (V, E)

V: set of vertices

§1 Introduction

§1.1 Definitions

Let I be a finite or countable set. All of our random variables will be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1 (Markov Chain)
A stochastic process $(X_n)_{n\geq 0}$ is called a Markov chain if $\forall n \geq 0$ and for $x_1 \dots x_{n+1} \in I$,

$$\mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_1 = x_1\right) = \mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right)$$

We can think of n as a discrete measure of time. If $\mathbb{P}\left(X_{n+1} = y \mid X_n = x\right)$ for all x, y is independent of n, then X is called time-homogeneous. Otherwise, X is called time-inhomogeneous. In this course, we only study time-homogeneous Markov chains. If we consider time-homogeneous chains only, we may as well take $n = 0$ and we can write

$$P(x, y) = \mathbb{P}\left(X_1 = y \mid X_0 = x\right); \quad \forall x, y \in I$$

Definition 1.2 (Stochastic Matrix)

A stochastic matrix is a matrix where the sum of each row is equal to 1.

We call P the transition matrix. It is a stochastic matrix:

$$\sum_{y \in I} P(x,y) = 1$$

Remark 1. The index set does not need to be $\mathbb{N}$; it could alternatively be $\{0, 1, \dots, N\}$ for $N \in \mathbb{N}$.

We say that X is Markov (λ, P) if X_0 has distribution λ, and P is the transition matrix. Hence,

1. $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$
2. $\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0) = P(x_n, x_{n+1}) =: P_{x_n x_{n+1}}$

We usually draw a diagram of the transition matrix using a graph. Directed edges between nodes are labelled with their transition probabilities.

§1.2 Sequence definition

Theorem 1.1

The process X is Markov (λ, P) if and only if $\forall n \geq 0$ and all $x_0, \dots, x_n \in I$, we have

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \lambda_{x_0} P(x_0, x_1) P(x_1, x_2) \dots P(x_{n-1}, x_n)$$

Proof. If X is Markov, then we have

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

as required. Conversely, $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$ satisfies (i). The transition matrix is given by

$$\begin{aligned}
\mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) &= \frac{\lambda_{x_0} P(x_0, x_1) \dots P(x_{n-1}, x_n)}{\lambda_{x_0} P(x_0, x_1) \dots P(x_{n-2}, x_{n-1})} \\
&= P(x_{n-1}, x_n)
\end{aligned}$$

which is exactly the Markov property as required. $\square$

§1.3 Point masses

Definition 1.3

For $i \in I$, the δ_i-mass at i is defined by

$$\delta_{ij} = (i = j)$$

This is a probability measure that has probability 1 at i only.

§1.4 Independence of sequences

Recall that discrete random variables (X_n) are considered independent if for all $x_1, \dots, x_n \in I$, we have

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n)$$

A sequence (X_n) is independent if for all k, $i_1 < i_2 < \dots < i_k$ and for all $x_1, \dots, x_k$, we have

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

Let $X = (X_n), Y = (Y_n)$ be sequences of discrete random variables. They are independent if for all k, m, $i_1 < \dots < i_k$, $j_1 < \dots < j_m$,

$$\mathbb{P}(X_1 = x_1, \dots, X_{i_k} = x_{i_k}, Y_{j_1} = y_{j_1}, \dots, Y_{j_m}) = \mathbb{P}(X_1 = x_1, \dots, X_{i_k} = x_{i_k}) \mathbb{P}(Y_{j_1} = y_{j_1}, \dots, Y_{j_m})$$

§1.5 Simple Markov property

Theorem 1.2 (Simple Markov Property)

Suppose X is Markov(λ, P). Let $m \in \mathbb{N}$ and $i \in I$. Given that $X_m = i$, we have that the process after time m, written $(X_{m+n})_{n \geq 0}$, is Markov(δ_i, P), and it is independent of $X_0, \dots, X_m$.

Informally, the past and the future are independent given the present.

Proof. Let $x_0, x_1, \dots, x_n \in I$. We must show that

$$\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n \mid X_m = i) = \delta_{ix_0} P(x_0, x_1) \dots P(x_{n-1}, x_n)$$

We have

$$\mathbb{P}\left(X_{m+n}=x_{m+n},\dots,X_m=x_m \mid X_m=i\right)=\frac{\mathbb{P}\left(X_{m+n}=x_{m+n},\dots,X_m=x_m\right)}{\mathbb{P}\left(X_m=i\right)}\delta_{ix_m}$$

The numerator is

$$\begin{aligned}
&\mathbb{P}\left(X_{m+n},\dots,X_m=x_m\right)\\
&=\sum_{x_0,\dots,x_{m-1}\in I}\mathbb{P}\left(X_{m+n}=x_{m+n},\dots,X_m=x_m,X_{m-1}=x_{m-1},\dots,X_0=x_0\right)\\
&=\sum_{x_0,\dots,x_{m-1}}\lambda_{x_0}P(x_0,x_1)\dots P(x_{m-1},x_m)P(x_m,x_{m+1})\dots P(x_{m+n-1},x_{m+n})\\
&=P(x_m,x_{m+1})\dots P(x_{m+n-1},x_{m+n})\sum_{x_0,\dots,x_{m-1}}\lambda_{x_0}P(x_0,x_1)\dots P(x_{m-1},x_m)\\
&=P(x_m,x_{m+1}s)\dots P(x_{m+n-1},x_{m+n})\mathbb{P}\left(X_m=x_m\right)
\end{aligned}$$

Thus we have

$$\mathbb{P}\left(X_{m+n}=x_{m+n},\dots,X_m=x_m \mid X_m=i\right)=P(x_m,x_{m+1})\dots P(x_{m+n-1},x_{m+n})\delta_{ix_m}$$

Hence $(X_{m+n})_{n\geq 0}\sim \text{Markov}\left(\delta_i,P\right)$ conditional on $X_m=i$. Now it suffices to show independence between the past and future variables. In particular, we need to show $m\leq i_1<\dots<i_k$ for some $k\in\mathbb{N}$ implies that

$$\begin{aligned}
&\mathbb{P}\left(X_{i_1}=x_{m+1},\dots,X_{i_k}=x_{m+k},X_0=x_0,\dots,X_m=x_m \mid X_m=i\right)\\
&=\mathbb{P}\left(X_{i_1}=x_{m+1},\dots,X_{i_k}=x_{m+k} \mid X_m=i\right)\mathbb{P}\left(X_0=x_0,\dots,X_m=x_m \mid X_m=i\right)
\end{aligned}$$

So let $i=x_m$, and then

$$\begin{aligned}
&=\frac{\mathbb{P}\left(X_{i_1}=x_{m+1},\dots,X_{i_k}=x_{m+k},X_0=x_0,\dots,X_m=x_m\right)}{\mathbb{P}\left(X_m=i\right)}\\
&=\frac{\lambda_{x_0}P(x_0,x_1)\dots P(x_{m-1},x_m)\mathbb{P}\left(X_{i_1}=x_{m+1},\dots,X_{i_k}=x_{m+k} \mid X_m=x_m\right)}{\mathbb{P}\left(x_m=i\right)}\\
&=\frac{\mathbb{P}\left(X_0=x_0,\dots,X_m=x_m\right)}{\mathbb{P}\left(X_m=x_m\right)}\mathbb{P}\left(X_{i_1}=x_{m+1},\dots,X_{i_k}=x_{m+k} \mid X_m=x_m\right)\\
&=\mathbb{P}\left(X_0=x_0,\dots,X_m=x_m \mid X_m=i\right)\mathbb{P}\left(X_{i_1}=x_{m+1},\dots,X_{i_k}=x_{m+k} \mid X_m=i\right)
\end{aligned}$$

which gives the result as required. □

§1.6 Powers of the transition matrix

Suppose $X\sim\text{Markov}\left(\lambda,P\right)$ with values in I. If I is finite, then P is an $|I|\times|I|$ square matrix. In this case, we can label the states as $1,\dots,|I|$. If I is infinite, then we label the states using the natural numbers $\mathbb{N}$. Let $x \in I$ and $n \in \mathbb{N}$. Then,

$$\begin{aligned}\mathbb{P}\left(X_n = x\right) &= \sum_{x_0,\dots,x_{n-1}\in I} \mathbb{P}\left(X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\right) \\ &= \sum_{x_0,\dots,x_{n-1}\in I} \lambda_{x_0} P(x_0, x_1) \dots P(x_{n-1}, x)\end{aligned}$$

We can think of λ as a row vector[1]. So we can write this as

$$= (\lambda P^n)_x$$ [2]

By convention, we take $P^0 = I$, the identity matrix. Now, suppose $m, n \in \mathbb{N}$. By the simple Markov property,

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

We will write $\mathbb{P}_x\left(A\right) := \mathbb{P}\left(A \mid X_0 = x\right)$ as an abbreviation. Further, we write $p_{ij}(n)$ for the (i, j) element of P^n. We have therefore proven the following theorem.

Theorem 1.3

$$\begin{gathered}\mathbb{P}\left(X_n = x\right) = (\lambda P^n)_x \\ \mathbb{P}\left(X_{n+m} = y \mid X_m = x\right) = p_{xy}(n) = \mathbb{P}(X_n = y \mid X_0 = x)\end{gathered}$$

§1.7 Calculating powers

Example 1.1

Consider

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}; \quad \alpha, \beta \in [0, 1]$$

Note that for any stochastic matrix P, P^n is a stochastic matrix. First, we have $P^{n+1} = P^n P$. Let us begin by finding $p_{11}(n+1)$.

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

Note that $p_{11}(n) + p_{12}(n) = 1$ since P^n is stochastic. Therefore,

$$p_{11}(n+1) = p_{11}(n)(1-\alpha-\beta) + \beta, \quad p_{11}(0) = 1.$$

We can solve this recursion relation to find

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

The general procedure for finding P^n is as follows. Suppose that P is a $k \times k$ matrix. Then let $\lambda_1, \dots, \lambda_k$ be its eigenvalues (which may not be all distinct).

- If all λ_i distinct. In this case, P is diagonalisable, and hence we can write $P = UDU^{-1}$ where D is a diagonal matrix, whose diagonal entries are the λ_i. Then, $P^n = UD^nU^{-1}$. Calculating D^n may be done termwise since D is diagonal. In this case, we have terms such as

  $$p_{11}(n) = a_1\lambda_1^n + \cdots + a_k\lambda_k^n; \quad a_i \in \mathbb{R}$$

  First, note $P^0 = I$ hence $p_{11}(0) = 1$. We can substitute small values of n and then solve the system of equations.[3]

  Now, suppose λ_k is complex for some k. In this case, $\overline{\lambda_k}$ is also an eigenvalue. Then, up to reordering,

  $$\lambda_k = re^{i\theta} = r(\cos\theta + i\sin\theta); \lambda_{k-1} = \overline{\lambda_k} = re^{-i\theta} = r(\cos\theta - i\sin\theta)$$

  We can instead write $p_{11}(n)$ as, where b_{k-1}, b_k are set appropriately

  $$p_{11}(n) = a_1\lambda_1^n + \cdots + a_{k-2}\lambda_{k-1}^n + b_{k-1}r^n\cos(n\theta) + b_k r^n \sin(n\theta)$$

  Since $p_{11}(n)$ is real, all the imaginary parts disappear, so we can simply ignore them.

- If not all λ_i distinct. In this case, λ appears with multiplicity 2, then we include also the term $(an+b)\lambda^n$ as well as $b\lambda^n$. This can be shown by considering the Jordan normal form of P.

Example 1.2

Let

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

The eigenvalues are $1, \frac{1}{2}i, -\frac{1}{2}i$. Then, writing $\frac{i}{2} = \frac{1}{2}(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2})$, we can write

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

[3] If we only care about $p_{11}(n)$ then we don't need to work out U we can solve this specific case.

For $n = 0$ we have $p_{11}(0) = 1$, and for $n = 1$ we have $p_{11}(1) = 0$, and for $n = 2$ we can calculate P^2 and find $p_{11}(2) = 0$. Solving this system of equations for α, β, γ, we can find

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

§2 Elementary Properties

§2.1 Communicating classes

Definition 2.1 (Communication)

Let X be a Markov chain with transition matrix P and values in I. For $x, y \in I$, we say that x leads to y, written $x \to y$, if

$$\mathbb{P}_x\left(\exists n \geq 0, X_n = y\right) > 0$$

We say that x communicates with y and write $x \leftrightarrow y$ if $x \to y$ and $y \to x$.

Theorem 2.1

The following are equivalent:

1. $x \to y$
2. There exists a sequence of states $x = x_0, x_1, \ldots, x_k = y$ such that

$$P(x_0, x_1)P(x_1, x_2)\ldots P(x_{k-1}, x_k) > 0$$

3. There exists $n \geq 0$ such that $p_{xy}(n) > 0$.

Proof. First, we show (i) and (iii) are equivalent. If $x \to y$, then $\mathbb{P}_x\left(\exists n \geq 0, X_n = y\right) > 0$. Then if $\mathbb{P}_x\left(\exists n \geq 0, X_n = y\right) > 0$ we must have some $n \geq 0$ such that $\mathbb{P}_x\left(X_n = y\right) = p_{xy}(n) > 0$. Note that we can write (i) as $\mathbb{P}_x\left(\bigcup_{n=0}^{\infty} X_n = y\right) > 0$. If there exists $n \geq 0$ such that $p_{xy}(n) > 0$, then certainly the probability of the union is also positive.

Now we show (ii) and (iii) are equivalent. We can write

$$p_{xy}(n) = \sum_{x_1, \ldots, x_{n-1}} P(x, x_1)\ldots P(x_{n-1}, y)$$

which leads directly to the equivalence of (ii) with (iii). $\square$

Corollary 2.1

Communication. $\leftrightarrow$, is an equivalence relation on I.

Proof. Reflexivity: $x \leftrightarrow x$ since $p_{xx}(0) = 1$.
Transitivity: If $x \to y$ and $y \to z$ then by (ii) above, $x \to z$.

Symmetric by definition. ☐

Definition 2.2 (Communicating Classes)

The equivalence classes induced on I by the communication equivalence relation are called communicating classes.

Definition 2.3 (Closed Communicating Class)

A communicating class C is closed if $x \in C, x \to y \implies y \in C$.

Definition 2.4 (Irreducibility)

A transition matrix P is called irreducible if it has a single communicating class. In other words, $\forall x, y \in I, x \leftrightarrow y$.

Definition 2.5 (Absorption)

A state x is called absorbing if $\{x\}$ is a closed (communicating) class. Equivalently if the markov chain started from x it stays at x forever.

§2.2 Hitting times

Definition 2.6 (Hitting Time)

For $A \subseteq I$, we define the hitting time of A to be a random variable $T_A \colon \Omega \to \{0, 1, 2 \dots\} \cup \{\infty\}$, defined by

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

with the convention that $\inf \varnothing = \infty$.

Definition 2.7 (Hitting Probability)

The hitting probability of A starting at $i \in I$ is $h^A \colon I \to [0, 1]$, defined by

$$h^A(i) = h_i^A = \mathbb{P}_i\left(T_A < \infty\right), \ i \in I$$

Definition 2.8 (Mean Hitting Time)

The mean hitting time of A starting at $i \in I$ is $k^A \colon I \to [0, \infty]$, defined by

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

Example 2.1
Consider

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

Consider $A = \{4\}$.

$$h_1^A = 0$$

$$h_2^A = \mathbb{P}_2\left(T_A < \infty\right) = \frac{1}{2}h_1^A + \frac{1}{2}h_3^A$$

$$h_3^A = \frac{1}{2} \cdot 1 + \frac{1}{2}h_2^A$$

Hence $h_2^A = \frac{1}{3}$.

Now, consider $B = \{1, 4\}$. $k_2^B = \mathbb{E}[T_B]$

$$k_1^B = k_4^B = 0$$

$$k_2^B = 1 + \frac{1}{2}k_1^B + \frac{1}{2}k_3^B$$

$$k_3^B = 1 + \frac{1}{2}k_4^B + \frac{1}{2}k_2^B$$

Hence $k_2^B = 2$.

Theorem 2.2
Let $A \subset I$. Then the vector $(h_i^A)_{i\in I}$ is the minimal non-negative solution to the

system

$$x_i = \begin{cases} 1 & i \in A \\ \sum_j P(i,j) x_j & i \notin A \end{cases}$$

Minimality here means that if $(x_i)_{i \in I}$ is another non-negative solution, then $\forall i, h_i^A \leq x_i$.

Note. The vector $h_i^A = 1$ always satisfies the equation, since P is stochastic, but this solution is typically not minimal.

Proof. First, we will show that $(h_i^A)_{i \in I}$ solves the system of equations.

Certainly if $i \in A$ then $h_i^A = 1$. Suppose $i \notin A$. Consider the event $\{T_A < \infty\}$. We can write this event as a disjoint union of the following events:

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

By countable additivity,

$$\begin{aligned}
\mathbb{P}_i\left(T_A < \infty\right) &= \underbrace{\mathbb{P}_i\left(X_0 \in A\right)}_{=0} + \sum_{n=1}^{\infty} \mathbb{P}_i\left(X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\right) \\
&= \mathbb{P}(X_1 \in A) + \sum_{n=2}^{\infty} \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) \\
\mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) &= \sum_{j \notin A} \mathbb{P}_i(X_1 = j, X_2 \notin A, \dots, X_{n-1} \in A, X_n \in A) \\
&= \sum_{j \notin A} \mathbb{P}_i(X_2 \notin A, \dots, X_{n-1} \notin A, X_n \in A \mid X_0 = i, X_1 = j) P(i,j) \\
&= \sum_{j \notin A} P(i,j) \mathbb{P}_j(X_1 \notin A, \dots, X_{n-2} \notin A, X_{n-1} \in A) \\
h_i^A &= \mathbb{P}_i(X_1 \in A) + \sum_{n=1}^{\infty} \sum_{j \notin A} P(i,j) \mathbb{P}_j(X_1 \notin A, \dots, X_n \in A) \\
&= \sum_{j \in A} h_j^A + \sum_{j \notin A} P(i,j) h_j^A \\
&= \sum_j P(i,j) h_j^A
\end{aligned}$$

So $(h_i^A)_{i \in I}$ satisfies the equation.

Now we must show minimality. If (x_i) is another non-negative solution, we must show that $h_i^A \leq x_i \; \forall \, i$. We have for $i \notin A$

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

Substituting again,

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

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

$$x_i \geq \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \cdots + \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \text{ as } x_k \geq 0.$$

$$x_i \geq \mathbb{P}_i(T_a \leq n) \; \forall \, n \in \mathbb{N}$$

Now, note $\{T_A \leq n\}$ are a set of increasing functions of n, so by continuity of the probability measure, the probability increases to that of the union, $\{T_A < \infty\} = h_i^A$. □

Example 2.2

Consider the Markov chain previously explored:

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

Let $A = \{4\}$. Then $h_1^A = 0$ since there is no route from 1 to 4. From the theorem above, the system of linear equations is

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

$$h_3 = \frac{1}{2}h_4 + \frac{1}{2}h_2$$

$$h_4 = 1$$

Hence,

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

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

So the minimal solution arises at $h_1 = 0$.

Example 2.3 (Simple random walk on $\mathbb{Z}_+$)

Consider $I = \mathbb{Z}_+$, and

$$P(i, i+1) = p \in (0,1); \quad P(i, i-1) = 1 - p = q$$

with $\mathbb{P}(0,1) = 1$. Then $h_i = \mathbb{P}_i (T_0 < \infty)$ hence $h_0 = 1$. The linear equations are

$$h_i = ph_{i+1} + qh_{i-1}$$

$$p(h_{i+1} - h_i) = q(h_i - h_{i-1})$$

Let $u_i = h_i - h_{i-1}$. Then assuming $p \neq q$,

$$\frac{q}{p}u_i = \cdots = \left(\frac{q}{p}\right)^i u_1$$

Hence

$$h_i = \sum_{j=1}^{i}(h_j - h_{j-1}) + 1 = 1 - (1 - h_1)\sum_{j=1}^{i}\left(\frac{q}{p}\right)^j$$

The general solution is therefore

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

with $h_0 = a + b = 1$. If $q > p$, then minimality of h_i implies $b = 0$, $a = 1$. Hence,

$$h_i = 1$$

Otherwise, if $p > q$, minimality of h_i implies $a = 0$, $b = 1$. Hence,

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

If $p = q = \frac{1}{2}$, then

$$h_i = \frac{1}{2}h_{i+1} + \frac{1}{2}h_{i-1}$$

Hence, $h_i = a + bi$. $h_0 = a = 1$ and minimality implies $b = 0$.

$$h_i = 1.$$

§2.3 Birth and death chain

Consider a Markov chain with

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

Now, consider $h_i = \mathbb{P}_i\left(T_0 < \infty\right)$. $h_0 = 1$, and $h_i = p_i h_{i+1} + q_i h_{i-1}$.

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

Let $u_i = h_i - h_{i-1}$ $(u_1 = h_1 - 1)$ to give

$$u_{i+1} = \frac{q_i}{p_i}u_i = \underbrace{\prod_{j=1}^{i} \frac{q_i}{p_i}}_{\gamma_i} u_i$$

Then

$$\begin{aligned} h_i &= \sum_{j=1}^{i}(h_j - h_{j-1}) + 1 \\ &= \sum_{j=1}^{i} u_j \\ &= 1 + u_1 + \sum_{j=1}^{i-1} \gamma_j u_1 \\ h_i &= 1 + (h_1 - 1)(\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}) \\ &= 1 - (1 - h_1)(\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}) \end{aligned}$$

where we let $\gamma_0 = 1$. Since h_i is a non-negative solution,

$$h_i \geq 0 \implies 1 - h_1 \leq \frac{1}{\sum_{j=0}^{i-1} \gamma_j} \leq \frac{1}{\sum_{j=0}^{\infty} \gamma_j}$$

By minimality, we must have exactly this bound. If $\sum_{j=0}^{\infty} \gamma_j = \infty$ then $1 - h_1 = 0 \implies h_i = 1$ for all i. If $\sum_{j=0}^{\infty} \gamma_j < \infty$ then

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

§2.4 Mean hitting times

Recall that

$$k_i^A = \mathbb{E}_i [T_A] = \sum_n n\mathbb{P}_i (T_A = n) + \infty \mathbb{P}_i (T_A = \infty)$$

Theorem 2.3

The vector $(k_i^A)_{i \in I}$ is the minimal non-negative solution to the system of equations

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

Proof. Suppose $i \in A$. Then $k_i = 0$. Now suppose $i \notin A$. Further, we may assume that $\mathbb{P}_i (T_A = \infty) = 0$, since if that probability is positive then the claim is trivial. Indeed, if $\mathbb{P}_i (T_A = \infty) > 0$, then there must exist j such that $P(i,j) > 0$ and $\mathbb{P}_j (T_A = \infty) > 0$ since

$$\mathbb{P}_i (T_A < \infty) = \sum_j P(i,j) h_j^A \implies 1 - \mathbb{P}_i (T_A = \infty) = \sum_j P(i,j)(1 - \mathbb{P}_j (T_A = \infty))$$

Because P is stochastic,

$$\mathbb{P}_i (T_A = \infty) = \sum_j P(i,j) \mathbb{P}_j (T_A = \infty)$$

so since the left hand side is positive, there must exist j with $P(i,j) > 0$ and $\mathbb{P}_j (T_A = \infty > 0)$. For this j, we also have $k_j^A = \infty$. Now we only need to compute $\sum_n n\mathbb{P}_i (T_A = n)$.

$$\mathbb{P}_i (T_A = n) = \mathbb{P}_i (X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A)$$

Then, using the same method as the previous theorem,

$$k_i^A = \sum_n n\mathbb{P}_i\left(T_A = n\right) = 1 + \sum_{j \notin A} P(i,j) k_j^A$$

It now suffices to prove minimality. Suppose (x_i) is another solution to this system of equations. We need to show that $x_i \geq k_i^A$ for all i. Suppose $i \notin A$. Then

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

Expanding inductively,

$$\begin{aligned} x_i = 1 &+ \sum_{j_1 \notin A} P(i,j_1) + \sum_{j_1 \notin A, j_2 \notin A} P(i,j_1)P(j_1,j_2) + \cdots \\ &+ \sum_{j_1 \notin A, \dots, j_n \notin A} P(i,j_1)\dots P(j_{n-1},j_n) + \sum_{j_1 \notin A, \dots, j_{n+1} \notin A} P(i,j)\dots P(j_n,j_{n+1}) x_{j_{n+1}} \end{aligned}$$

Since x is non-negative, we can remove the last term and reach an inequality.

$$x_i \geq 1 + \sum_{j_1 \notin A} P(i,j_1) + \sum_{j_1 \notin A, j_2 \notin A} P(i,j_1)P(j_1,j_2) + \cdots + \sum_{j_1 \notin A, \dots, j_n \notin A} P(i,j_1)\dots P(j_{n-1},j_n)$$

Hence

$$\begin{aligned} x_i &\geq 1 + \mathbb{P}_i\left(T_A > 1\right) + \mathbb{P}_i\left(T_A > 2\right) + \cdots + \mathbb{P}_i\left(T_A > n\right) \\ &= \mathbb{P}_i\left(T_A > 0\right) + \mathbb{P}_i\left(T_A > 1\right) + \mathbb{P}_i\left(T_A > 2\right) + \cdots + \mathbb{P}_i\left(T_A > n\right) \\ &= \sum_{k=0}^{n} \mathbb{P}_i\left(T_A > k\right) \end{aligned}$$

for all n. Hence, the limit of this sum is

$$x_i \geq \sum_{k=0}^{\infty} \mathbb{P}_i\left(T_A > k\right) = \mathbb{E}_i\left[T_A\right]$$

which gives minimality as required. □

§2.5 Strong Markov property

The simple Markov property shows that, if $X_m = i$,

$$X_{m+n} \sim \text{Markov}\left(\delta_i, P\right)$$

and this is independent of $X_0, \ldots, X_m$. The strong Markov property will show that the same property holds when we replace m with a finite random 'time' variable. It is not the case that any random variable will work; indeed, an m very dependent on the Markov chain itself might not satisfy this property.

Definition 2.9 (Stopping Time)

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

Example 2.4

The hitting time $T_A = \inf\{n \geq 0 \colon X_n \in A\}$ is a stopping time. This is because we can write

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

So first hitting times are always stopping times.

Example 2.5

The time $L_A = \sup\{n \geq 0 \colon X_n \in A\}$ is not a stopping time. This is because we need to know information about the future behaviour of X_n in order to guarantee that we are at the supremum of such events.

Theorem 2.4 (Strong Markov Property)

Let $X \sim \operatorname{Markov}(\lambda, P)$ and T be a stopping time. Conditional on $T < \infty$ and $X_T = i$,

$$(X_{n+T})_{n \geq 0} \sim \operatorname{Markov}(\delta_i, P)$$

and this distribution is independent of $X_0, \ldots, X_T$.

Proof. We need to show that, for all $x_0, \ldots, x_n \in I$ and for all vectors w of any length with elements in I,

$$\begin{aligned}&\mathbb{P}\left(X_T = x_0, \ldots, X_{T+n} = x_n, (X_0, \ldots, X_T) = w \mid T < \infty, X_T = i\right)\\ &= \delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n) \mathbb{P}\left((X_0, \ldots, X_T) = w \colon T < \infty, X_T = i\right)\end{aligned}$$

Suppose that w is of the form $w = (w_0, \ldots, w_k)$. Then,

$$\mathbb{P}\left(X_T = X_0, \ldots, X_{T+n} = x_n, (X_0, \ldots, X_T) = w \mid T < \infty, X_T = i\right)$$

$$
\begin{aligned}
&= \mathbb{P}\left(X_k = X_0, \ldots, X_{k+n} = x_n, (X_0, \ldots, X_k) = w \mid T < \infty, X_T = i\right) \\
&= \frac{\mathbb{P}\left(X_k = x_0, \ldots, X_{k+n} = x_n, (X_0, \ldots, X_k) = w, T = k, X_k = i\right)}{\mathbb{P}\left(T < \infty, X_T = i\right)}
\end{aligned}
$$

Now, since $\{T = k\}$ depends only on $X_0, \ldots, X_k$ by the simple Markov property

$$
\begin{aligned}
&\mathbb{P}\left(X_k = x_0, \ldots, X_{k+n} = x_n \mid (X_0, \ldots, X_k) = w, T = k, X_k = i\right) \\
&= \mathbb{P}\left(X_k = x_0, \ldots, X_{k+n} = x_n \mid X_k = i\right) = \delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n)
\end{aligned}
$$

$$
\frac{\mathbb{P}((X_0, \ldots, X_k) = w, T = k, X_k = i)}{\mathbb{P}\left(T < \infty, X_T = i\right)} = \mathbb{P}((X_0, \ldots, X_k) = w \mid T < \infty, X_T = i)
$$

Now,

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

as required. $\square$

Example 2.6

Consider a simple random walk on $I = \mathbb{N} \cup \{0\}$, where $P(x, x \pm 1) = \frac{1}{2}$ for $x \neq 0$, and $P(0, 1) = 1$. Now, let $h_i = \mathbb{P}_i\left(T_0 < \infty\right)$. We want to calculate h_1. We can write

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

but the system of recursion relations this generates is difficult to solve. Instead, we will write

$$
h_2 = \mathbb{P}_2\left(T_0 < \infty\right)
$$

Note that in order to hit 0, we must first hit 1. So conditioning on the first hitting time of 1 being finite, after this time the process starts again from 1. Now,

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

Conditional on $T_1 < \infty$ and X_{T_1}, by the strong Markov property $(X_{T+n})_{n \geq 0}$ is Markov (δ_1, P). So we can express (under the conditioning) $T_0 = T_1 + \widetilde{T}_0$, where $\widetilde{T}_0$

is independent of T_1[a], with the same distribution as T_0 under $\mathbb{P}_1$. Note that

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

But $\mathbb{P}_2\left(T_1 < \infty\right) = \mathbb{P}_1\left(T_0 < \infty\right)$ by translation invariance, so

$$\begin{aligned}h_2 &= \mathbb{P}_2\left(T_1 < \infty\right)\mathbb{P}_1\left(T_0 < \infty\right)\\ &= h_1^2\end{aligned}$$

So $h_1 = \frac{1}{2} + \frac{1}{2}h_2 \implies \frac{1}{2} + \frac{1}{2}h_1^2$ so $h_1 = 1$. In general, therefore, for any $n \in \mathbb{N}$,

$$h_n = h_1^n$$

[a] X_{T_1+n} is independent of $(X_0, \ldots, X_{T_1})$

§3 Transience and recurrence

§3.1 Definitions

Definition 3.1 (Recurrent State)
Let X be a Markov chain, and let $i \in I$. A state i is called recurrent if

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

Definition 3.2 (Transient State)
A state i is called transient if

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

We will prove that any i is either recurrent or transient.

§3.2 Probability of visits

Definition 3.3 (Total number of visits)
The number of visits to i is defined by

$$V_i = \sum_{n=0}^{\infty} 1(X_n = i)^{a}$$

[a] $1(X_0 = i) = 1$ so $V_i \geq 1$.

Definition 3.4 (Return Time)
Let $T_i^{(0)} = 0$ and inductively define

$$T_i^{(r+1)} = \inf\left\{n \geq T_i^{(r)} + 1 \mid X_n = i\right\}^{a}$$

We write $T_i^{(1)} = T_i$, called the first return time (or first passage time) to i. Further, let

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

[a] First time we hit i after $T_i^{(r)}$

Lemma 3.1

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

Proof. For $r = 0$, this is trivially true. Now, suppose that the statement is true for r, and we will show that it is true for $r + 1$.

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

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

$$\begin{aligned}
&= \mathbb{P}_i\left(T_i^{(1)} < \infty\right) f_i^r \\
&= f_i f_i^r \\
&= f_i^{r+1}
\end{aligned}$$

□

§3.3 Duality of transience and recurrence

Theorem 3.1

Let X be a Markov chain with transition matrix P, and let $i \in I$. Then, exactly one of the following is true.

1. If $f_i = 1$, then i is recurrent, and

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

2. If $f_i < 1$, then i is transient, and

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

Proof.

$$\begin{aligned}\mathbb{E}_i[V_i] &= \mathbb{E}_i\left[\sum_{n=0}^{\infty} 1(X_n = i)\right] \\ &= \sum_{n=0}^{\infty} \mathbb{E}_i[1(X_n = i)] \\ &= \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = i) \\ &= \sum_{n=0}^{\infty} p_{ii}(n)\end{aligned}$$

First, suppose $f_i = 1$. Then, for all r, $\mathbb{P}_i(V_i > r) = 1$, so $\mathbb{P}_i(V_i = \infty) = 1$. Hence, i is recurrent. Further, $\mathbb{E}_i[V_i] = \infty$ so $\sum_{n=0}^{\infty} p_{ii}(n) = \infty$.

Now, if $f_i < 1$, by the previous lemma we see that $\mathbb{E}_i[V_i] = \frac{1}{1-f_i} < \infty$ hence $\mathbb{P}_i(V_i < \infty) = 1$. Thus, i is transient. Further, $\mathbb{E}_i[V_i] < \infty$ so $\sum_{n=0}^{\infty} p_{ii}(n) < \infty$. □

Theorem 3.2

Let x, y be states that communicate. Then, either x and y are both recurrent, or they are both transient.

Proof. Suppose x is recurrent. Then, since x and y communicate, $\exists m, \ell \in \mathbb{N}$ such that

$$p_{xy}(m) > 0; \quad p_{yx}(\ell) > 0$$

Note, $\sum_n p_{xx}(n) = \infty$. Then,

$$\begin{aligned}\sum_n p_{yy}(n) \geq \sum_n p_{yy}(n+m+\ell) &\geq \sum_n p_{yx}(\ell) p_{xx}(n) p_{xy}(m) \\ &\geq p_{yx}(\ell) p_{xy}(m) \sum_n p_{xx}(n) = \infty\end{aligned}$$

□

Corollary 3.1

Either all states in a communicating class are recurrent or they are all transient.

§3.4 Recurrent communicating classes

Theorem 3.3

Any recurrent communicating class is closed.

Proof. Suppose a communicating class C is not closed. Then there exists $x \in C$ and $y \notin C$ such that $x \to y$. Let m be such that $p_{xy}(m) > 0$. If, starting from x, we hit y which is outside the communicating class, then we can never return to the communicating class (including x) again. In particular,

$$\mathbb{P}_x\left(V_x < \infty\right) \geq \mathbb{P}_x\left(X_m = y\right) = p_{xy}(m) > 0$$

Hence x is not recurrent, which is a contradiction. □

Theorem 3.4

Any finite closed communicating class is recurrent.

Proof. Let C be a finite closed communicating class. Let $x \in C$. Then, by the pigeonhole principle, there must exist $y \in C$ such that

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

Since x and y communicate, there exists $m \in \mathbb{N}$ such that $p_{yx}(m) > 0$. Now,

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

Thus y is recurrent. Since recurrence is a class property, C is recurrent. □

Theorem 3.5

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

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

Proof. Since y is recurrent,

$$1 = \mathbb{P}_y\left(X_n = y \text{ for infinitely many } n\right)$$

Let m such that $p_{yx}(m) > 0$. Now,

$$\begin{aligned}
1 &= \mathbb{P}_y\left(X_n = y \text{ infinitely often}\right) \\
&= \sum_z \mathbb{P}_y\left(X_m = z, X_n = y \text{ for infinitely many } n \geq m\right) \\
&= \sum_z \mathbb{P}_y\left(X_n = y \text{ for infinitely many } n \geq m \mid X_m = z\right) \mathbb{P}_y\left(X_m = z\right) \\
&= \sum_z \mathbb{P}_z\left(X_n = y \text{ for infinitely many } n\right) \mathbb{P}_y\left(X_m = z\right)
\end{aligned}$$

By the strong Markov property,

$$= \sum_z \mathbb{P}_z\left(T_y < \infty\right) \mathbb{P}_y\left(X_n = y \text{ for infinitely many } n\right) \mathbb{P}_y\left(X_m = z\right)$$

Since y is recurrent,

$$\begin{aligned}
&= \sum_z \mathbb{P}_z\left(T_y < \infty\right) \mathbb{P}_y\left(X_m = z\right) \\
&= \sum_z \mathbb{P}_z\left(T_y < \infty\right) p_{yz}(m)
\end{aligned}$$

Since $p_{yz}(m) > 0$ and $\sum_z p_{yz}(m) = 1$, $\mathbb{P}_x\left(T_y < \infty\right) = 1$. $\square$

§4 Pólya's recurrence theorem

§4.1 Statement of theorem

Definition 4.1

The simple random walk in $\mathbb{Z}^d$ is the Markov chain defined by

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

where e_i is the standard basis.

Theorem 4.1

The simple random walk in $\mathbb{Z}^d$ is recurrent for $d = 1, d = 2$ and transient for $d \geq 3$.

§4.2 One-dimensional proof

Consider $d = 1$. In this case, $P(x, x + 1) = P(x, x - 1) = \frac{1}{2}$. We will show that $\sum_n p_{00}(n) = \infty$, then recurrence will hold. We have $p_{00}(n) = \mathbb{P}_0 (X_n = 0)$. Note that if n is odd, X_n is odd, so $\mathbb{P}_0 (X_{2k+1} = 0) = 0$. So we will consider only even numbers. In order to be back at zero after $2n$ steps, we must make n steps 'up' away from the origin and make n steps 'down'. There are $\binom{2n}{n}$ ways of choosing which steps are 'up' steps. The probability of a specific choice of n 'up' and n 'down' is $\left(\frac{1}{2}\right)^{2n}$. Hence,

$$p_{00}(2n) = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2} \cdot \frac{1}{2^{2n}}$$

Recall Stirling's formula:

$$n! \sim n^n e^{-n} \sqrt{2\pi n}$$

Substituting in,

$$\frac{(2n)!}{(n!)^2} \cdot \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}} = \frac{A}{\sqrt{n}}$$

for $A > 0$; the precise value of A is unnecessary. Hence, for some large n_0, $\forall n \geq n_0$, $p_{00}(2n) \geq \frac{A}{2\sqrt{n}}$. So

$$\sum_n p_{00}(2n) \geq \sum_{n \geq n_0} \frac{A}{2\sqrt{n}} = \infty$$

Now, let us consider the asymmetric random walk for $d = 1$, defined by $P(x, x + 1) = p$ and $P(x, x - 1) = q$. We can compute $p_{00}(2n) = \binom{2n}{n}(pq)^n \sim A\frac{(4pq)^n}{\sqrt{n}}$. If $p \neq q$, then $4pq < 1$ so by the geometric series we have

$$\sum_{n \geq n_0} p_{00}(2n) \leq \sum_{n \geq n_0} 2A(4pq)^n < \infty$$

So the asymmetric random walk is transient.

§4.3 Two-dimensional proof

Now, let us consider the simple random walk on $\mathbb{Z}^2$. For each point $(x, y) \in \mathbb{Z}^2$, we will project this coordinate onto the lines $y = x$ and $y = -x$. In particular, we define

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

If X_n is the simple random walk on $\mathbb{Z}^2$, we consider $f(X_n) = (X_n^+, X_n^-)$.

Lemma 4.1

$(X_n^+), (X_n^-)$ are independent simple random walks on $\frac{1}{\sqrt{2}}\mathbb{Z}$.

Proof. We can write X_n as

$$X_n = \sum_{i=1}^n \xi_i$$

where ξ_i are independent and identically distributed by

$$\mathbb{P}(\xi_1 = (1, 0)) = \mathbb{P}(\xi_1 = (-1, 0)) = \mathbb{P}(\xi_1 = (0, 1)) = \mathbb{P}(\xi_1 = (0, -1)) = \frac{1}{4}$$

and we write $\xi_i = (\xi_i^1, \xi_i^2)$. We can then see that

$$X_n^+ = \sum_{i=1}^n \frac{\xi_i^1 + \xi_i^2}{\sqrt{2}}; \quad X_n^- = \sum_{i=1}^n \frac{\xi_i^1 - \xi_i^2}{\sqrt{2}}$$

We can check that $(X_n^+), (X_n^-)$ are simple random walks on $\frac{1}{\sqrt{2}}\mathbb{Z}$. It now suffices to prove the independence property. Note that it suffices to show that $\xi_i^1 + \xi_1^2$ and $\xi_i^1 - \xi_i^2$ are independent, since the X_n^+, X_n^- are sums of independent and identically distributed copies of these random variables. We can simply enumerate all possible values of ξ_i^1, ξ_i^2 and the result follows. □

We know that $p_{00}(n) = 0$ if n is odd. We want to find $p_{00}(2n) = \mathbb{P}_0(X_{2n} = 0)$. Note, $X_n = 0 \iff X_n^+ = X_n^- = 0$. Using the lemma above,

$$\mathbb{P}_0(X_{2n} = 0) = \mathbb{P}_0\left(X_n^+ = 0, X_n^- = 0\right) = \mathbb{P}_0\left(X_n^+ = 0\right)\mathbb{P}_0\left(X_n^- = 0\right) \sim \frac{A}{\sqrt{n}}\frac{A}{\sqrt{n}} = \frac{A^2}{n}$$

Hence,

$$\sum_{n \geq n_0} \mathbb{P}_0(X_{2n} = 0) \geq \sum_{n \geq n_0} = \frac{A^2}{2n} = \infty$$

which gives recurrence as required.

§4.4 Three-dimensional proof

Consider $d = 3$. Again, $p_{00}(n) = 0$ if n odd. In order to return to zero after $2n$ steps, we must make i steps both up and down, j steps north and south, and k steps east and west, with $i + j + k = n$. There are $\binom{2n}{i,i,j,j,k,k}$ ways of choosing which steps in each direction we take. Each combination has probability $\left(\frac{1}{6}\right)^{2n}$ of happening. Hence,

$$p_{00}(2n) = \sum_{i,j,k \geq 0, i+j+k=n} \binom{2n}{i,i,j,j,k,k} \left(\frac{1}{6}\right)^{2n} = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sum_{i,j,k \geq 0, i+j+k=n} \binom{n}{i,j,k}^2 \left(\frac{1}{3}\right)^{2n}$$

The sum on the right hand side is the total probability of the number of ways of placing n balls in three boxes uniformly at random, so equals one. Suppose $n = 3m$. Then we can show that $\binom{n}{i,j,k} \leq \binom{n}{m,m,m}$.

$$p_{00}(6m) \geq \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \binom{n}{m,m,m} \left(\frac{1}{3}\right)^{n}$$

Applying Stirling's formula again, we have

$$p_{00}(6m) \sim \frac{A}{n^{3/2}}$$

It is sufficient to consider $n = 3m$:

$$p_{00}(6m) \geq \frac{1}{6^2} p_{00}(6m-2); \quad p_{00}(6m) \geq \frac{1}{6^4} p_{00}(6m-4)$$

Hence

$$\sum_n p_{00}(n) < \infty$$

So the Markov chain is transient.