University of
Cambridge

Mathematics Tripos

Part IB

Markov Chains

Michaelmas, 2017

Lectures by
G. R. Grimmett

Notes by
Qiangru Kuang

Contents

1 Definition

Let $X : \Omega \to S$ be a random variable. Then $(X_0, X_1, ...)$, a sequence of random variables, is called a *stochastic/random process*. The problem we are interested in is whether there is any dependence between the random variables. Another example of a stochastic process is $(X_t, t \in \mathbb{R})$, representing, for example, the evolution of a system with respect to time.

Definition. Let $X = (X_n : n = 0, 1, 2, ...)$ be a sequence taking values in some *state space* S, which is either finite or countably infinite. X is a *Markov chain* if it satisfies the *Markov condition*:

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) \quad \forall n \geq 0, i_0, \dots, i_{n+1} \in S$$

X is called *homogeneous* if $\mathbb{P}(X_{n+1} = j | X_n = i)$ does not depend on the value of n.

Example.

1. Random walk is a Markov chain: let $Z_1, Z_2, ...$ be independent, $\mathbb{P}(Z_i = 1) = p, \mathbb{P}(Z_i = -1) = 1 - p$, $X_n = Z_1 + \dots + Z_n$.
2. Branching process: let X_n be the size of the nth generation, then (X_n) is a Markov chain.

Convention. Henceforth, unless contradicted, all chains are assumed to be homogeneous.

Two quantities associated with a chain are:

1. initial distribution $\lambda = (\lambda_i : i \in S)$ where $\lambda_i = \mathbb{P}(X_0 = i)$, the probability mass function at 0,
2. transition matrix $P = (p_{i,j} : i, j \in S)$ given by $p_{i,j} = \mathbb{P}(X_1 = j | X_0 = i)$.

Proposition 1.1.

1. *λ is a distribution in that $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$.*
2. *P is a* stochastic matrix *in that $p_{i,j} \geq 0, \sum_j p_{i,j} = 1$.*

Proof.

1. $\lambda_i = \mathbb{P}(X_0 = i) \geq 0$, $\sum_i \lambda_i = \sum_i \mathbb{P}(X_0 = i) = 1$, i.e. $\{X_0 = i\}_{i \in X}$ partitions Ω.
2. $p_{i,j} = \mathbb{P}(X_1 = j | X_0 = i) \geq 0$, $\sum_j p_{i,j} = \sum_j \mathbb{P}(X_1 = j | X_0 = i) = 1$.

$\square$

Theorem 1.2. *Let λ be a distribution on S and P be a stochastic matrix. The sequence $X = (X_n : n \geq 0)$ is a Markov chain with initial distribution λ and transition matrix P if and only if*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n} \qquad \forall n \geq 0, i_0, \cdots, i_n \in S \quad (*)$$

Proof. Let $A_k = \{X_k = i_k\}$. Equation (**??**) is

$$\mathbb{P}(A_0 \cap A_1 \cap \cdots \cap A_n) = \lambda_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n} \tag{$\star$}$$

Suppose X is a (λ, P) Markov chain. Proof of equation (**??**) by induction on n. When $n = 0$, $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$. Suppose equation (**??**) holds for $n < N$.

$$\begin{aligned}
\mathbb{P}(A_0 \cap \cdots \cap A_N) &= \mathbb{P}(A_0 \cap \cdots \cap A_N | A_0 \cap \cdots \cap A_{N-1}) \mathbb{P}(A_0 \cap \cdots \cap A_{N-1}) \\
&= \mathbb{P}(A_N | A_0 \cap \cdots \cap A_{N-1}) \mathbb{P}(A_0 \cap \cdots \cap A_{N-1}) \\
&\overset{\text{MP}}{=} \mathbb{P}(A_N | A_{N-1}) \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{N-2},i_{N-1}}
\end{aligned}$$

Conversely, suppose equation (**??**) holds. By the equation, when $n = 0$,

$$\mathbb{P}(X_0 = i_0) = \lambda_{i_0},$$

so X_0 has p.m.f. λ. Then

$$\mathbb{P}(A_{n+1} | A_0 \cap \cdots \cap A_n) = \frac{\mathbb{P}(A_0 \cap \cdots \cap A_{n+1})}{\mathbb{P}(A_0 \cap \cdots \cap A_n)} = p_{i_n,i_{n+1}}.$$

Therefore X is a Markov chain with transition matrix P. □

Theorem 1.3 (Extended Markov Property)**.** *Let X be a Markov chain and $n \geq 1$. Let H be a historic event, i.e. H is given in terms of $\{X_0, X_1, \dots, X_{n-1}\}$, and let F be a future event, i.e. F is given in terms of $\{X_{n+1}, X_{n+2}, \dots\}$. Then*

$$\mathbb{P}(F | H, X_n = i) = \mathbb{P}(F | X_n = i).$$

Proof. For F that depends only on finitely many of the future variables,

$$\begin{aligned}
\mathbb{P}(F | H, X_n = i) &= \frac{\sum_{>n} \sum_{<n} \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_n,i} p_{i,i_{n+1}} \cdots}{\sum_{<n} \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i}} \\
&= \sum_{>n} p_{i,i_{n+1}} \cdots \\
&= \mathbb{P}(F | X_n = i)
\end{aligned}$$

The case for infinite variables can be deduced using continuity of probability measure. □

Notation. $\sum_{<n}$ denotes the summation over all $i_0, \dots, i_{n-1}$ contibutions to H/F.

2 Transition Probabilities

The one-step transition probability is $p_{i,j} = \mathbb{P}(X_1 = j|X_0 = i)$. The n-step transition probability is $p_{i,j}(n) = \mathbb{P}(X_n = j|X_0 = i)$.

Question. How to compute the n-step probabilities from the one-step probabilities?

The answer is: matrix. The one-step transition matrix is $P = (p_{i,j})_{i,j\in S}$. Similarly $P(n) = (p_{i,j}(n))_{i,j\in S}$.

Theorem 2.1.

$$P(n) = P^n.$$

Proposition 2.2 (Chapman-Kolmogorov equations)**.**

$$p_{i,j}(m+n) = \sum_{k\in S} p_{i,k}(m)p_{k,j}(n).$$

Proof.

$$\mathbb{P}(X_{m+n} = j|X_0 = i) = \sum_{k\in S} \mathbb{P}(X_{m+n} = j, X_m = k|X_0 = i)$$

Use the equality $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|B \cap C)\mathbb{P}(B|C)$,

$$= \sum_{k\in S} \mathbb{P}(X_{m+n} = j|X_m = k, X_0 = i)\mathbb{P}(X_m = k|X_0 = i)$$

By Markov property, the first term can be simplified

$$= \sum_{k\in S} p_{i,j}(m)p_{k,j}(n)$$

□

Proof of Theorem. By Chapman-Kolmogorov equation, $P(m+n) = P(m)P(n)$. Thus $P(n) = P(1)P(n-1) = \cdots = P(1)^n = P^n$. □

Example. Let $S = \{1, 2\}$, $P = \left(\begin{smallmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{smallmatrix}\right)$ where $0 < \alpha, \beta < 1$.

1. Diagonalisation method: $\det(P - \kappa I) = 0$, has roots $\kappa_1 = 1, \kappa_2 = 1 - \alpha - \beta$. Then
$$P = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix} U,$$
for some invertible U. Then
$$P^n = U^{-1} \begin{pmatrix} 1^n & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} U$$
Thus we may write
$$P_{1,1}(n) = A + B(1-\alpha-\beta)^n,$$

with $P_{1,1}(0) = 1, P_{1,1}(1) = 1 - \alpha$. Solve to get

$$\begin{aligned} A &= \frac{\beta}{\alpha + \beta} \\ B &= \frac{\alpha}{\alpha + \beta} \end{aligned}$$

For the other entries, note $P_{1,2}(n) = 1 - P_{1,1}(n)$ and $P_{2,1}(n)$ and $P_{2,2}(n)$ can be obtained by exchanging α and β due to symmetry.

2. Difference equation method:

$$\begin{aligned} p_{1,1}(n+1) &= \sum_{k=1,2} p_{1,k}(n) p_{k,1}(1) \\ &= p_{1,1}(n)(1-\alpha) + p_{1,2}(n)\beta \\ &= p_{1,1}(n)(1-\alpha) + (1 - p_{1,1}(n))\beta \end{aligned}$$

Thus $\pi_n = p_{1,1}(n)$ satisfies

$$\pi_{n+1} = (1 - \alpha - \beta)\pi_n + \beta$$

which can be solved.

It is convenient to think of λ as a row vector and P as a matrix. Then $\mathbb{P}(X_1 = j) = \sum \lambda_i p_{i,j}$ becomes $(\mathbb{P}(X_i = j) : j \in S) = \lambda P$, i.e. pre-multiplying P by λ.

3 Class Structure

Definition. We say a state i *leads* to a state j if $p_{i,j}(n) > 0$ for some $n \geq 0$, write $i \to j$.

If $i \to j$ and $j \to i$, we say i and j *communicate* and write $i \leftrightarrow j$.

Proposition 3.1. *$\leftrightarrow$ is an equivalence relation.*

Proof.

- reflexive: $p_{i,i} = 1 > 0$,
- symmetric: if $i \leftrightarrow j$ then $j \leftrightarrow i$ by definition,
- transitive: if $i \leftrightarrow j, j \leftrightarrow k$ then exist m, n such that $p_{i,j}(m) > 0, p_{j,k}(n) > 0$. Then

  $$p_{i,k}(m+n) \overset{\text{CK}}{=} \sum_r p_{i,r}(m) p_{r,k}(n) \geq p_{i,j}(m) p_{j,k}(n) > 0$$

  so $i \to k$. Similarly $k \to i$.

□

Definition. The equivalence classes of $\leftrightarrow$, i.e. subsets of S of the form $C_i = \{j \in S : i \leftrightarrow j\}$, are called *communicating classes*.

Definition. If there is a unique equivalence class S, call S (or the chain) *irreducible*.

Definition. A subset $C \subseteq S$ is called *closed* if

$$i \in C, i \to j \Rightarrow j \in C.$$

If $i \in S$ is such that $\{i\}$ is closed, i is called *absorbing*.

A closed set is different from communicating class as it is "one-way".

Proposition 3.2. *$C \subseteq S$ is closed if and only if*

$$p_{i,j} = 0 \text{ for } i \in C, j \notin C. \qquad (*)$$

Proof. Let $C \subseteq S$. If (**??**) fails, then there exists $i \in C, j \notin C$ such that $p_{i,j} > 0$ so $i \to j$ and so C is not closed.

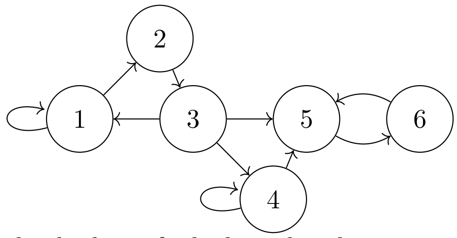
Suppose (**??**) holds and exists $m > 0$ such that $p_{i,j}(m) > 0$. Then

$$0 < p_{i,j}(m) = \sum_{x_1, \dots, x_{m-1} \in S} p_{i,x_1} p_{x_1,x_2} \cdots p_{x_{m-1},j}.$$

Thus exists $x_1, \dots, x_{m-1} \in S$ such that a summand on RHS is larger than 0. By (**??**) $x_1, \dots, x_{m-1}, j \in C$ so C is closed. □

Example.

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



has communicating classes $\{1, 2, 3\}, \{4\}, \{5, 6\}$, only the last of which is closed.

4 Recurrence & Transience

4.1 Definition

Introduce notation $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot|X_0 = i), \mathbb{E}_i(\cdot) = \mathbb{E}(\cdot|X_0 = i)$.

Definition. The *first-passage time* of $j \in S$ is

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

The *first-passage probability* are

$$f_{i,j}(n) = \mathbb{P}_i(T_j = n).$$

Definition. $i \in S$ is *recurrent (or persistent)* if

$$\mathbb{P}_i(T_i < \infty) = 1,$$

and *transient* otherwise.

4.2 Recurrent Condition

Theorem 4.1. *i is recurrent if and only if*

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

Before we prove the theorem, introduce two generating functions:

$$\begin{aligned} P_{i,j}(s) &= \sum_{n \geq 0} p_{i,j}(n) s^n \\ F_{i,j}(s) &= \sum_{n \geq 0} f_{i,j}(n) s^n \end{aligned}$$

with the convention that

$$\begin{aligned} p_{i,j}(0) &= \delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \\ f_{i,j}(0) &= 0 \, \forall i, j \end{aligned}$$

Note that

$$\begin{aligned} p_{i,j}(n) &= \sum_{m=1}^{n} \mathbb{P}_i(X_n = j | T_j = m) \mathbb{P}_i(T_j = m) \\ &\overset{\text{MP}}{=} \sum_{m=1}^{n} p_{j,j}(n-m) f_{i,j}(m) \end{aligned}$$

which is the convolution of p and f.

It follows that for $n \geq 1$,

$$\begin{aligned}
\sum_{n\geq 1} p_{i,j}(n)s^n &= \sum_{n\geq 1}\sum_{m=1}^{n} \Big(f_{i,j}(m)s^m\Big)\Big(p_{j,j}(n-m)s^{n-m}\Big) \\
P_{i,j}(s) - \delta_{i,j} &= \sum_{m=1}^{\infty}\sum_{n=m}^{\infty} \Big(f_{i,j}(m)s^m\Big)\Big(p_{j,j}(n-m)s^{n-m}\Big) \\
&= \sum_{m\geq 1} f_{i,j}(m)s^m \sum_{r\geq 0} p_{j,j}(r)s^r \\
&= F_{i,j}(s)P_{j,j}(s)
\end{aligned}$$

be careful when dealing with double summations.

Theorem 4.2.

$$P_{i,j}(s) = \delta_{i,j} + F_{i,j}(s)P_{j,j}(s) \textit{ for } |s| < 1.$$

The $|s| < 1$ condition: since $|F_{i,j}(s)| < \infty$ if $|s| < 2$,

$$|P_{i,j}(s)| \leq \sum_n |s|^n < \infty$$

A lemma from analysis before we finally prove Theorem **??**:

Lemma 4.3 (Abel's Lemma)**.** *Let $(u_i)_{i\geq 0}$ be a non-negative sequence such that*

$$\mathcal{U}(s) = \sum_{i=0}^{\infty} u_i s^i$$

converges when $|s| < 1$, then

$$\sum_{i=0}^{\infty} u_i = \lim_{s\to 1^-} \mathcal{U}(s),$$

whether or not this limit is finite.

Proof. Exercise. □

Proof of Theorem **??**. For $|s| < 1$, then

$$\begin{aligned}
p_{i,i}(s) &= 1 + F_{i,i}P_{i,i} \\
P_{i,i} &= \frac{1}{1 - F_{i,j}}
\end{aligned}$$

Let $s \to 1^-$, then by Abel's Lemma

$$P_{i,i}(1) = \infty \Leftrightarrow F_{i,i}(1) = 1 \tag{1}$$

i.e. i is recurrent. □

4.3 Properties of Recurrence

Theorem 4.4 (Recurrence as a class property)**.** *Let C be a communicating class, then*

1. *either every state in C is recurrent or every state is transient,*
2. *if C contains some recurrent state then C is closed.*

Proof.

1. Let $i, j \in C$, $i \neq j$. There exist $m, n \geq 1$ such that

$$\alpha = p_{i,j}(m)p_{j,i}(n) > 0.$$

This is simply a restatement of the communicating property. Then

$$p_{i,j}(m+k+n) \geq p_{i,j}(m)p_{j,j}(k)p_{j,i}(n) = \alpha p_{j,j}(k)$$

Intuitively, the middle term is the probability of going from i to i by passing j at step m and $m+k$. Thus

$$\sum_k p_{i,i}(m+k+n) \geq \alpha \sum_k p_{j,j}(k).$$

Thus j recurrent implies that i is recurrent. Vice versa.

2. Suppose $i \in C$ is recurrent but C is not closed. Then there exists $j \in C, k \notin C$ with $p_{j,k} > 0$. By the previous part j is recurrent so

$$1 = \mathbb{P}_j(T_j < \infty) = 1 - \mathbb{P}_j(\text{no return to } j) \leq 1 - p_{j,k} < 1$$

Absurd.

□

Theorem 4.5. *Assume $|S| < \infty$, then*

1. *S contains some recurrent state,*
2. *if the chain is irreducible, all states are recurrent.*

Similar as before, we need a proposition before the proof:

Proposition 4.6. *If j is a transient state then*

$$\forall i,\ p_{i,j}(n) \to 0 \text{ as } n \to \infty$$

Proof. Asuume j is transient. By (**??**),

$$P_{j,j}(1) < \infty.$$

By Theorem **??**,

$$\sum_n p_{i,j}(n) = P_{i,j}(1) < \infty$$

so nth term $p_{i,j}(n)$ tends to 0 as $n \to \infty$. □

Proof.

1. $\sum_{j \in S} p_{i,j}(n) = 1$ since P is a stochastic matrix. If j is transient then each summand tends to 0 as $n \to \infty$, which is absurd since $|S| < \infty$.

2. Obvious.

□

4.4 Random Walks and Pólya's Theorem

In this section, we discuss simple symmetric random walk on d-dimensional lattices, i.e. $\mathbb{Z}^d$, in particular answering the question when the chain is recurrent.[1] It turns out that there is a surprisingly beautiful result.

We call two points x and y *neighbours* if

$$\sum_i |x_i - y_i| = 1,$$

i.e. they differ by 1 in only one coordinate. Let X be a symmetric random walk on $\mathbb{Z}^d$ where $d \geq 1$, i.e. $X = (X_1, X_2, \dots)$ is a Markov chains with state space $S = \mathbb{Z}^d$ and transition probability

$$\mathbb{P}(X_{n+1} = y | X_n = x) = \begin{cases} 0 & \text{if } x \text{ and } y \text{ are not neighbours} \\ \frac{1}{2d} & \text{if } x \text{ and } y \text{ are neighbours} \end{cases}$$

Theorem 4.7 (Pólya's)**.** *X is recurrent if $d \leq 2$ and transient if $d \geq 3$.*

Proof. First set $d = 1$. Recall that 0 is recurrent if and only if $\sum_n p_{0,0}(n) = 0$. However, it is not possible to return to the same place after an odd number of steps so the expression simplifies to

$$\sum_n p_{0,0}(2n) \tag{$*$}$$

In the even case, the random walk returns to 0 if and only if there are equal number of movement to either direction so by applying the binomial distribution,

$$p_{0,0}(2n) = \left(\frac{1}{2}\right)^{2n} \binom{2n}{n}$$

To simplify this, recall Sterling's formula

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \text{ as } n \to \infty.$$

So

$$p_{0,0}(2n) \sim \frac{1}{\sqrt{\pi n}}$$

and the sum (**??**) tend to infinity.

[1] Note that this chain is irreducible so by Theorem **??** either we can talk about recurrence as a chain property.

Now let $d = 2$. By the same reasoning and generalising binomial to multinomial coefficients, we get

$$\begin{aligned} p_{0,0}(2n) &= \left(\frac{1}{4}\right)^{2n} \sum_{m=0}^{n} \binom{2n}{m, m, n-m, n-m} \\ &= \left(\frac{1}{4}\right)^{2n} \sum_{m=0}^{n} \frac{(2n)!}{(m!)^2((n-m)!)^2} \\ &= \left(\frac{1}{4}\right)^{2n} \frac{(2n)!}{(n!)^2} \sum_{m=0}^{n} \binom{n}{m} \binom{n}{n-m} \end{aligned}$$

Now pause and think: the summation represents the number of ways to take m balls from a bag of n balls and take $n - m$ balls from another bag of n balls, for $0 \le m \le n$, but this is precisely the number of ways to take n balls from $2n$ balls!

$$\begin{aligned} &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \binom{2n}{n} \\ &= \left(p_{0,0}^{d=1}(2n)\right)^2 \end{aligned}$$

thus the sum (**??**) also tends to infinity and 0 is recurrent.

Let $d = 3$ (similar for $d \ge 4$) and we have

$$\begin{aligned} p_{0,0}(2n) &= \left(\frac{1}{6}\right)^{2n} \sum_{i+j+k=n} \binom{2n}{i, i, j, j, k, k} \\ &= \left(\frac{1}{6}\right)^{2n} \sum_{i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} \\ &= \left(\frac{1}{6}\right)^{2n} \binom{2n}{n} \sum_{i+j+k=n} \left(\frac{n!}{i!j!k!}\right)^2 \\ &= \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \sum_{i+j+k=n} \left(\frac{n!}{3^n i!j!k!}\right)^2 \\ &\le \left(\frac{1}{2}\right)^{2n} M_n \sum_{i+j+k=n} \frac{1}{3^n i!j!k!} \end{aligned}$$

where $M_n = \max\{\frac{n!}{3^n i!j!k!}, i + j + k = n\}$.

The reason we introduce 3^n becomes apparent in this step: $\frac{n!}{3^n i!j!k!}$ is the probability of, upon throwing n balls into 3 urns, finding i, j, k in each respectively. Thus they sum up to 1.

$$\le \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \frac{n!}{3^n \left(\lfloor n/3 \rfloor !\right)^3}$$

The upper bound of M_n is left as an exercise. By Sterling's formula,

$$p_{0,0}(2n) \le \frac{C}{n^{3/2}}$$

so the sum is finite and 0 is recurrent. □

We have seen that the probability $p_{0,0}(2n)$ when $d = 2$ is the square of the probability when $d = 1$, but when $d = 3$ it doesn't become cubed. It should inspire us to suspect that the $d = 2$ case is simple enough such that the random walks in two directions are "independent", but when $n \geq 3$ there is some hidden structure that destroys such independence. What is so special about dimension two?

There is an alternative way to tackle this problem that might be more lucid and shed some light on the magical property of $d = 2$. Instead of cartesian coordinates $X_n = (A_n, B_n)$, rotate the axes by 45° clockwise. The new coordinates, scaled by a constant factor for convenience, are

$$Y_n = \begin{pmatrix} U_n \\ V_n \end{pmatrix} = \sqrt{2} \begin{pmatrix} \cos 45^\circ & -\sin 45^\circ \\ \sin 45^\circ & \cos 45^\circ \end{pmatrix} X_n = \begin{pmatrix} A_n - B_n \\ A_n + B_n \end{pmatrix}$$

Claim $U = (U_n)$ and $V = (V_n)$ are *independent* random walks on $\mathbb{Z}$:

Proof.

$$\begin{aligned} \mathbb{P}(Y_{n+1} = Y_n + (1,1)) = \mathbb{P}(X_{n+1} = X_n + (1,0)) = \frac{1}{4} \\ = \mathbb{P}(U_{n+1} - U_n = 1, V_{n+1} - V_n = 1) \end{aligned}$$

since during one step X_n can only change by 1 in one coordinate. Similar for the other three cases.

So

$$\mathbb{P}(U_{n+1} - U_n = \alpha, V_{n+1} - V_n = \beta) = \left(\frac{1}{2}\right)^2 = \mathbb{P}(U_{n+1} - U_n = \alpha)\mathbb{P}(V_{n+1} - V_n = \beta)$$

for $\alpha, \beta = \pm 1$. U_n and V_n are independent and each generates a random walk on $\mathbb{Z}$. □

By independence

$$\mathbb{P}_0(X_n = (0,0)) = \mathbb{P}_0(Y_n = (0,0)) = \mathbb{P}_0(U_n = 0)\mathbb{P}_0(V_n = 0)$$

so

$$p_{0,0}^{d=2}(2n) = \left(p_{0,0}^{d=1}\right)^2.$$

The moral of this calculation is, two dimensional random walk is indeed the product of two one dimensional cases, but in a not-entirely-straightforward way.

5 Hitting Time and Probability

Definition (Hitting Time)**.** Given a subset $A \subseteq S$, the *hitting time* of A is

$$H^A = \inf\{n \geq 0 : X_n \in A\}$$

Note that $\inf \emptyset = \infty$ so $H^A : \Omega \to \{0, 1, \dots\} \cup \{\infty\}$.

Definition (Hitting probability)**.** The *hitting probability* is $h_i^A = \mathbb{P}_i(H^A < \infty)$.

By Markov property, hitting probability satisfies the equation

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

Theorem 5.1. *The vector $h^A = (h_i^A : i \in S)$ is the minimal non-negative solution to (**??**) in that for any $x = (x_i : i \in S)$ satisfy*

$$x_i = \begin{cases} 1 & i \in A \\ \sum_j p_{ij} x_j & i \notin A \end{cases}$$

and $x_i \geq 0$ for $i \in S$, $h_i^A \leq x_i$ for $i \in S$.

Proof. By MP, h^A satisfies (**??**). Suppose $x = (x_i : i \in S)$ satisfies the hypothesis in the theorem. If $i \in A$, $x_i = 1 = h_i^A$ so $h_i^A \leq x_i$. Let $i \notin A$, then

$$\begin{aligned} x_i &= \sum_j p_{ij} x_j \\ &= \sum_{j \in A} p_{ij} \cdot 1 + \sum_{j \notin A} p_{ij} x_j \\ &\geq \sum_{j \in A} p_{ij} \\ &= \mathbb{P}_i(H^A = 1) \end{aligned}$$

Similarly

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

so by induction

$$x_i \geq \sum_{m=1}^{n} \mathbb{P}_i(H^A = m) = \mathbb{P}_i(H^A \leq n) \to \mathbb{P}_i(H^A < \infty) = h_i^A$$

as $n \to \infty$. $\square$

Definition (Mean hitting time)**.** The *mean hitting time* is

$$k_i^A = \mathbb{E}_i(H^A).$$

Note. $k_i^A = \infty$ if $h_i^A < 1$.

Theorem 5.2. *The vector $k^A = (k_i^A : i \in S)$ is the minimal non-negative solution to the equation*

$$y_i = \begin{cases} 0 & i \in A \\ 1 + \sum_j p_{ij} y_j & i \notin A \end{cases} \tag{$\dagger$}$$

Proof. By MP, k^A satisfies (**??**). Let $y = (y_i : i \in S)$ be a non-negative solution to (**??**). Let $i \in A$ then $y_i = 0 = k_1$. Let $i \notin A$.

$$\begin{aligned} y_i &= 1 + \sum_{j \in S} p_{ij} y_j \\ &= 1 + \sum_{j \notin A} p_{ij} y_j \\ &= 1 + \sum_{j \notin A} p_{ij} \left(1 + \sum_{k \notin A} p_{jk} y_k\right) \\ &\geq \mathbb{P}_i(H^A \geq 1) + \mathbb{P}_i(H^A \geq 2) \end{aligned}$$

By induction

$$y_i \geq \sum_{m=1}^{n} \mathbb{P}_i(H^A \geq m) \to \sum_{m=1}^{\infty} \mathbb{P}_i(H^A \geq m) = \mathbb{E}_i(H^A)$$

as $n \to \infty$. □

Example (Gambler's ruin)**.** Let $S = \{0, 1, 2, \dots\}$ and $0 < p < 1$. Take a random walk on S which moves one step right with probability p and left with probability $q = 1 - p$. 0 is an *absorbing barrier* as as soon as the random walk hits 0 it ends.

Question. What is the probability of absorption at 0 starting at i?

Answer. Let $h_i = h_i^{\{0\}}$, then

$$\begin{aligned} h_0 &= 1 \\ h_i &= p h_{i+1} + q h_{i-1} \end{aligned}$$

which can be regarded as a second order difference equation with initial conditions.

First suppose $p \neq q$. The general solution is

$$h_i = A + B \left(\frac{q}{p}\right)^i \text{ for } i \geq 0.$$

Suppose $p < q$. Since $h^i \leq 1$, $B = 0$ or otherwise for large i, h^i will blow up. Thus $h_i = h_0 = 1$ for $i \geq 0$. Now suppose $p > q$. Eliminate B to get

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

The minimality of h^i requires that $A = 0$ so

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

Finally, consider the case $p = q = \frac{1}{2}$. As above $B = 0, A = 1$ so $h_i = 1$ for $i \geq 0$.

Example (Birth-death chain)**.** This is similar to gambler's ruin but with inhomogeneuous transition probabilites: $p_i + q_i = 1, p_i \in (0, 1)$. The governing equations are

$$\begin{aligned} h_0 &= 1 \\ h_i &= p_i h_{i+1} + q_i h_{i-1} \end{aligned}$$

At first glance, it seems nothing like a second order differential equation. However, rearrange to get

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

Define $u_i = h_{i-1} - h_i$ then

$$p_i u_{i+1} = q_i u_i$$

so

$$\begin{aligned} u_i &= \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1} u_1 \\ &= \gamma_i u_1 \end{aligned}$$

where

$$\gamma_i = \frac{\prod_{j=1}^i q_j}{\prod_{j=1}^i p_j}.$$

As

$$\begin{aligned} h_i &= 1 - (u_1 + u_2 + \cdots + u_i) \\ &= 1 - u_1(\gamma_0 + \cdots + \gamma_{i-1}) \text{ for } i \geq 1 \end{aligned}$$

where $\gamma_0 = 1$. Let $S = \sum_{i=0}^\infty \gamma_i$. If $S = \infty$ then $u_1 = 0$ and $h_i = 1$. If $S < \infty$, $1 - u_1 S = 0$ so

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

6 Stopping Times & Strong Markov Property

Definition (Stopping time)**.** Let X be a Markov chain. A *stopping time (or Markov time)* is a random variable $T : \Omega \to \{0, 1, \dots\} \cup \{\infty\}$ such that for $n \geq 0$, the event $\{T = n\}$ is given in terms of $X_0, \dots, X_n$.

Note. The definition can be equivalently formulated using languages of meausure theory. Recall that in a probability space $\Omega, \mathcal{F}, \mathbb{P}$, a random variable $X : \Omega \to \mathbb{R}$ is measurable if

$$X^{-1}((-\infty, n]) \in \mathcal{F} \; \forall n \in \mathbb{R}.$$

Define the σ-field generated by X to be

$$\sigma(X) = \sigma(\{X^{-1}((-\infty, n]) : n \in \mathbb{R}\}) \subseteq \mathcal{F}$$

and the definition basically says $\{T = n\} \in \sigma(\{X_0, \dots X_n\})$.

Theorem 6.1 (Strong Markov property)**.** *Let X be a Markov chain with transition matrix P and let T be a stopping time. Given $T < \infty$ and $X_T = i$,*

$$Y := (X_T, X_{T+1}, X_{T+2}, \dots),$$

the future process of X, is a Markov chain with transition matrix P and Y is independent of $X_0, \dots, X_{T-1}$.

Example.

1. Hitting time H^A is a stopping time:

$$\{H^A = n\} = \{X_n \in A\} \cap \left(\bigcap_{m=0}^{n-1} \{X_m \notin A\} \right)$$

2. $H^A + 1$ is a stopping time.
3. $H^A - 1$ is *not* a stopping time: $\{H^A - 1 = n\}$ obviously depends on X_{n+1}.

Example (Gambler's ruin)**.** Let $H = H^{\{0\}}$. We have proved that

$$\mathbb{P}_i(H < \infty) = \begin{cases} = 1 & q \geq p \\ < 1 & q < p \end{cases}$$

if $i \geq 1$. However, we want to find the probability mass function of H given $X_0 = 1$. Use probability generating function:

$$G(s) = \mathbb{E}_1(s^H)$$

where $|s| < 1$ so that s^∞ can be interpreted as 0

$$= \sum_{n=1}^{\infty} s^n \mathbb{P}_1(H = n)$$

where the limit $s \to 1^-$ is studied via Abel's Lemma.

$$\begin{aligned} G(s) &= \mathbb{E}_1(s^H | X_1 = 0)q + \mathbb{E}_1(s^H | X_1 = 2)p \\ &= qs + p\mathbb{E}_1(s^{1+H'+H''}) \end{aligned}$$

where H' is the hitting time of 1 starting at 2 and H'' is the subsequent time needed to reach 0. By strong Markov property, H' and H'' are independent and distributed as H so

$$= qs + psG(s)^2$$

Hence

$$G(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps}$$

Note that G is continuous on $(-1,1)$ since it is the a power series. Since $\sqrt{1-4pqs^2} \neq 0$ for $|s| < 1$, we must choose a sign and stick with it on $(-1,1)$. Since G convergens $(-1,1)$, $+$ is impossible as otherwise G does not converge at $s = 0$. Thus

$$G(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps} = \sum_{n=1}^{\infty} s^n \mathbb{P}_1(H = n)$$

hence $\mathbb{P}_1(H = n)$ can be found by expanding G.

In addition

$$\begin{aligned} \lim_{s \to 1^-} G(s) &= \sum_{n=1}^{\infty} \mathbb{P}_1(H = n) \\ &= \mathbb{P}_1(H < \infty) \\ &= \frac{1 - \sqrt{1 - 4pq}}{2p} \\ &= \frac{1 - |p - 1|}{2p} \\ &= \begin{cases} 1 & q \geq p \\ \frac{q}{p} & q < p \end{cases} \end{aligned}$$

To find $\mathbb{E}_1(H)$ when $q \geq p$, differentiate to get

$$G' = q + pG^2 + 2psGG'$$

so

$$G'(s) = \frac{q + pG^2}{1 - 2psG}$$

By **??** Abel's theorem

$$E_1(H) = \lim_{s \to 1^-} G'(s) = \frac{q + p}{1 - 2p} = \frac{1}{q - p}.$$

7 Classification of States

Depending on ht mean time to return to the recurrent state: finite or infinite

Theorem 7.1. *Let $X_0 = i$ and $V_i = |\{n \geq 1 : X_n = i\}|$. Then V_i has a geometric distribution*

$$\mathbb{P}_i(V_i = r) = f^r(1-f), r \geq 1$$

where $f = f_{i,i} \in [0,1]$.

Proof.

$$\begin{aligned}\mathbb{P}_i(V_i \geq r) &= \mathbb{P}_i(T^r < \infty)\\ &= \mathbb{P}_i(T^4 < \infty | T^{r-1} < \infty)\mathbb{P}_i(T^{r-1} < \infty)\\ &= f \cdot \mathbb{P}_i(V_i \geq r-1) SMP\\ &= f^r\end{aligned}$$

where

$$T^r = \begin{cases} timeofrthreturn & \\ \infty & \text{if } V_i < r\end{cases}$$

Then

$$\mathbb{P}_i(V_i = r) = \mathbb{P}_i(V_i \geq r) - \mathbb{P}_i(V_i \geq r+1) = f^r(1-r)$$

□

Note. If $f < 1$, $\mathbb{P}_i(V_i < \infty) = 1$ and if $f = 1$, $\mathbb{P}_i(V_i = \infty) = 1$.

Definition (Mean recurrence time)**.** The *mean recurrence time* of $i \in S$ is

$$\mu_i = \mathbb{E}_i(T_i) = \begin{cases}\infty & \text{if } i \text{ is transient}\\ \sum_{n=1}^{\infty} n f_{i,i}(n) & \end{cases}$$

Let i be recurrnet. Then i is *null* if $\mu_i = \infty$ and *positive* if $\mu_i < \infty$.
The *period* of $i \in S$ is

$$d_i = \gcd\{n : p_{i,i}(n) > 0\}.$$

i is *aperiodic* if $d_i = 1$.
i is *ergodic* if it is recurrent, positive and aperiodic.

Theorem 7.2. *Let $i \leftrightarrow j$. Then*

1. *$d_i = d_j$, i.e. period is a class property.*
2. *i is recurrent if and only if j is recurrent.*
3. *i is positive recurrent if and only if j is positive recurrent.*
4. *i is ergodic if and only if j is ergodic.*

Proof. 2 has already been proved and 4 follows from 1, 2 and 3. 3 will be proved later. To prove 1:

Let $i \leftrightarrow j$ and $i \neq j$.

$$\begin{aligned} D_k &= \{n \geq 1 : p_{k,k}(n) > 0\} \\ d-k &= \gcd\{D_k\} \end{aligned}$$

Since $i \to j$, there exists $m, n \geq 1$ such that $\alpha = p_{i,j}(m)p_{i,j}(n) > 0$. By CK,

$$p_{i,i}(m+r+n) \geq \alpha p_{j,j}(r).$$

Thus if $r \in D_j \cup \{0\}$, $p_{i,i}(m+r+n) > 0$ and hence $d_i \mid m+r+n$ and hence $d_i \mid r$ since if $r = 0$, $d_i \mid m+n$.

Thus $d_i \mid d_j$. Similarly $d_j \mid d_i$ and hence $d_i = d_j$. □

Proposition 7.3. *If a chain is irreducible and let $j \in S$ be recurrent. Then*

$$\mathbb{P}(T_j < \infty) = \mathbb{P}(X_n = j \text{ for some } n \geq 1) = 1.$$

Compare to the definition of recurrence, $\mathbb{P}_j$

Proof. $f_{i,j} = \mathbb{P}_i(T_j < \infty)$. Let $i \neq j$. Claim that $p_{j,i}(m)(1-f_{i,j}) \leq 1-f_{j,j}$ where $m = \inf\{r : p_{j,i}(r) > 0\}$. Then $p_{j,i}(m) = \mathbb{P}_j(X_m = i, X_r \neq j \text{ for } 1 \leq r < m)$. We have $f_{j,j} = 1$ and hence $f_{i,j} = 1$.

Thus $\mathbb{P}_i(T_j < \infty) = 1$. Let $\lambda_i = \mathbb{P}(X_0 = i)$,

$$\mathbb{P}(T_j < \infty) = \sum_{i \in S} \mathbb{P}_i(T_j < \infty)\lambda_i = 1.$$

□

8 Invariant Distributions

What happens to X_n as $n \to \infty$? Random variables are functions so we are talking about convergence of a sequence of functions. There are lots of modes of convergence on function space. When studying Markov chains, it turns out there is a unique convergence that we need: does $\mathbb{P}(X_n = i)$ converge as $n \to \infty$.

$$\mathbb{P}(X_{n+1} = j) = \sum_i \mathbb{P}(X_{n+1} = j | X_n = 0)\mathbb{P}(X_n = i)$$

$$\pi_j = \sum_i p_{i,j}\pi_i$$

$\pi = \pi P$, eigenvalue problem.

Definition (Invariant distribution)**.** X is aMarkov chain with transition matrix P. The vector $\pi = (\pi_i : i \in S)$ is an *invariant distribution* if

1. $\pi_i \geq 0, \sum_i \pi_i = 1$,
2. $\pi = \pi P$

If X_0 has distribution π, X_n has distribution

$$\pi P^n = (\pi P)P^{n-1} = \pi P^{n-1} = \cdots \pi$$

Theorem 8.1. *Let X be an irreducible Markov chain. Then*

1. *There exists an invariant distribution if and only if some state of the chain is positive recurrent.*
2. *If there exists an invariant distribution π then every state is positive recurrent and*
$$\pi_i = \frac{1}{\mu_i}$$
for $i \in S$ where μ_i is the mean recurrence time. In particular π is unique.

Fix $k \in S$, start at k. Let W_i be the number of visits to i up to the first return time to k, i.e.

$$W_i = \sum_{m=1}^{\infty} \mathbf{1}(X_m = i, T_k \geq m) = \sum_{m=1}^{T_k} \mathbf{1}(X_m = i)$$

where $\mathbf{1}(\cdot)$ is the indicator function.

Let $\rho(i) = \mathbb{E}_k(W_i)$.

Proposition 8.2. *Suppose the chain is irreducible and recurrent, $k \in S$. $\rho = (\rho_i : i \in S)$ satisfies*

1. *$\rho_k = 1$.*
2. *$\sum_i \rho(i) = \mu_k$ whether or not $\mu_k < \infty$.*
3. *$\rho = \rho P$.*

4. $0 < \rho_i < \infty$ *for* $i \in S$.

Proof.

1. Immediate from the definition.

2. $$\mathbb{E}_k \sum_{i \in S} W_i = E_k T_k$$

   Assuming we can interchange summations (since $\mathbb{E}$ is the limit of a series), since all summands are non-negative,

   $$\sum_{i \in S} E_k(W_i) = \mu_k$$

   but the summand on LHS is precisely ρ_i.

3. $$\begin{aligned}
\rho_j &= \mathbb{E}_k(W_j) \\
&= \sum_{m \geq 1} \mathbb{P}_k(X_m = j, T_k \geq m) \text{ again we interchange the summations} \\
&= \sum_{m \geq 1} \sum_{i \in S} \mathbb{P}_k(X_m = j, X_{m-1} = i, T_k \geq m) \\
&= \sum_{m \geq 1} \sum_{i \in S} \mathbb{P}_k(X_m = j | X_{m-1} = i, T_k \geq m) \mathbb{P}_k(X_{m-1} = i, T_k \geq m) \\
&= \sum_{m \geq 1} \sum_{i \in S} \mathbb{P}_k(X_m = j | X_{m-1} = i, T_k \geq m) \mathbb{P}_k(X_{m-1} = i) \text{ Markov property} \\
&= \sum_{m \geq 1} \sum_{i \in S} p_{i,j} \mathbb{P}_k(X_{m-1} = i, T_k \geq m) \\
&= \sum_{i \in S} p_{i,j} \sum_{m \geq 1} \mathbb{P}_k(X_{m-1} = i, T_k \geq m) \\
&= \sum_{i \in S} p_{i,j} \sum_{r \geq 1} \mathbb{P}_k(X_r = i, T_k \geq r + 1)
\end{aligned}$$

   consider two cases: if $i \neq k$, the term when $r = 0$ is 0 and $T_k \geq r + 1$ if and only if $T_k \geq r$. If $i = k$, the term when $r = 0$ is 1 and all the other terms are zero. Thus

   $$= \sum_{i \in S} p_{i,j} \rho_i$$

4. Pivot off the fact that $\rho_k = 1$. Since $\rho = \rho P$, we have $\rho = \rho P^r$ for $r \geq 1$. Thus

   $$\rho_i \geq \rho_k p_{k,i}(m), \rho_k \geq \rho_i p_{i,k}(n)$$

   By irreducibility, there exists $m, n \geq 1$ with $p_{k,i}(m), p_{i,k}(m) > 0$ so

   $$0 < p_{k,i}(m) \leq \rho_i \leq \frac{1}{p_{i,k}(n)} < \infty$$

□

Proof.

1. Let k be positive recurrent, hence $\mu_k < \infty$. Then $\pi_i := \rho_i/\mu_i k$ is an invariant distribution.

2. Let π be an invariant distribution. Claim $\pi_i > 0$ for all $i \in S$:

   Proof. Since $\pi = \pi p$, we have $\pi = \pi p^n$ for $n \geq 0$ and hence
   $$\pi_i = \sum_j \pi_j p_{j,i}(n) \geq \pi_k p_{k,i}(n)$$
   for $k \in S$. Since $\sum_i \pi_i = 1$ we may pick $k \in S$ with $\pi_k > 0$. By irreducibility there exists $n \geq 0$ such tat $p_{k,i}(n) > 0$. Hence $\pi_i > 0$.
   Suppose every state is transient, since $\pi = \pi p^n$,
   $$\pi_j = \sum_i \pi_i p_{i,j}(n)$$
   taking limit as $n \to \infty$,
   $$\sum_i \pi_i 0 = 0$$
   absurd.

   Proof of the limiting process.
   $$\begin{aligned} 0 \leq \sum_i \pi_i p_{i,j}(n) &= \sum_{i \in F} \cdots + \sum_i \notin F \cdots \text{ where } F \subseteq S, |F| < \infty \\ &\leq \sum_{i \in F} p_{i,j}(n) + \sum_{i \notin F} \pi_i \text{ by boundedness} \\ &\to 0 + \sum_{i \notin F} \pi_i \text{ as } n \to \infty \\ &\to 0 \text{ as } F \to S^- \end{aligned}$$
   □

   Thus every state is recurrent. □

$$\begin{aligned} \pi_i \mu_i &= \sum_{n=1}^{\infty} \mathbb{P}_i(T_i \geq n) \mathbb{P}(X_0 = i) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X_0 = i, T_i \geq n) \end{aligned}$$

Using stationarity,

$$\begin{aligned} \pi_i \mu_i &= \pi_i + \sum_{n=2}^{\infty} (a_{n-2} - a_{n-1}) \text{ where } a_r = \mathbb{P}(X_0 \neq i, \dots, X_r \neq i) \\ &= \pi_i + a_0 - \lim_{m \to \infty} a_m \\ &= \pi_i + (1 - \pi_i) - \underbrace{\mathbb{P}(T_i = \infty)}_{=0 \text{ by recurrence}} \\ &= 1 \end{aligned}$$

Since $\pi_i \mu_i = 1$,

$$\mu_i = \frac{1}{\pi_i} < \infty$$

since $\pi_i > 0$. Thus i is positive recurrent.

$\square$

9 Convergence to Equilibrium

Theorem 9.1. *Consider an irreducible aperiodic, positive recurrent Markov chain. For $i, j \in S$, $p_{i,j}(n) \to \pi_j$ as $n \to \infty$ where π is the unique invariant distribution.*

Ergodic theorem

Proof. "coupling" is the main idea.

Let $X = (X_n), Y = (Y_n)$ be independent Markov chains with the appropriate common invariant distribution. Let $Z = (Z_n = (X_n, Y_n))_{n \geq 0}$. Then Z is a Markov chain with state space $S \times S$ and transition matrix

$$p_{ij,kl} = p_{i,k} p_{j,l}$$

Fix $s \in S$, let

$$T = \inf\{n \geq 1 : Z_n = (s, s)\}.$$

Since X and Y have invariant distribution π, Z has invariant distribution $v_{ij} = \pi_i \pi_j$ since

$$\sum_{i,j} v_{ij} p_{ij,kl} = \sum_i \pi_i p_{i,k} \sum_j \pi_j p_{j,l} = \pi_k \pi_l = v_{kl}$$

Hence Z is positive recurrent and $\mathbb{P}(T < \infty) = 1$.

We are still lacking one thing: Z is irreducible. This has something to do with aperiodicity.

A digression about number theory: if D is a finite subset of non-negative integers with $\gcd(D) = 1$, there exists N such that for $n > N$ and expression

$$n = \sum_{d \in D} \alpha_d d$$

with $\alpha_d \in \{0, 1, 2, \dots\}$.

Since X is aperiodic, we deduce that $p_{i,i}(n) > 0$ for all large n. Thus

$$p_{ij,ij}(n) = p_{i,i}(n) p_{j,j}(n) > 0$$

for all large n. Therefore Z is aperiodic.

Similarly Z is irreducible (oops, check the book!) □

$$\begin{aligned}
p_{i,k}(n) &= \mathbb{P}_i(X_n = k) \\
&= \mathbb{P}_{ij}(X_n = k) \\
&= \sum_{t=1}^{\infty} \mathbb{P}_{ij}(X_n = k | T = t) \mathbb{P}_{ij}(T = t) \\
&= \sum_{t \leq n} \mathbb{P}_{ij}(X_n = k | T = t) \mathbb{P}_{ij}(T = k) + \sum_{t > n} \mathbb{P}_{ij}(T = t) \\
&= \sum_{t \leq n} \mathbb{P}_{ij}(Y_n = k, T = k) + \mathbb{P}_{ij}(T > n) \\
&\leq p_{j,k}(n) + \mathbb{P}_{ij}(T > n)
\end{aligned}$$

so

$$|p_{i,k}(n) - p_{j,k}(n)| \le \mathbb{P}_{ij}(T > n) \to 0$$

as $n \to \infty$. This says that if it converges they converge the the same value. Now we are just one line from the final result:

$$\pi_k - p_{ij}(n) = \sum_i \pi(p_{i,k}(n) - p_{j,k}(n)) \to 0$$

by bounded convergence theorem.

This is an extremely elegant proof and it took a long time before this proof was found.

We are left with one final bit

Theorem 9.2. *Let $V_i(n) = \sum_{k=1}^n \mathbf{1}(x_k = i)$ be the total number of visits to i up to time n. If the chain is irreducible and positive recurrent then*

$$\frac{V_i(n)}{n} \Rightarrow \frac{1}{\mu_i}$$

as $n \to \infty$, where $\Rightarrow$ means weak convergence, i.e.

$$\mathbb{P}\left(\frac{V_i(n)}{n} \le a\frac{1}{\mu_i}\right) \to \begin{cases} 0 & a < 1 \\ 1 & a > 1 \end{cases}$$

We are not going prove this. Renewal theorem.

Remark. Let u_i be a typical interval length between successive visits to i. $V_i(n) \ge x$ if and only if $\sum_{m=1}^x u_i(m) \le n$ where the $u_i(m)$ are iid copies of u_i.

10 Time Reversal

some ruminations bout physics: in real life time reversal is possible but extremely unlikely. The typical explanation is entropy.

Let $X = (X_n : n = 0, 1, \cdots, N)$ be an irreducible and positive recurrent Markov chain with transition matrix P and invariant distribution π. Let $Y_n = X_{N-n}$, so $Y = (Y_0, \cdots, Y_N) = (X_n, \cdots, X_0)$, the reverse of X.

We have to make some assumptions for our reversed chain to make sense: assume that X_0 has distribution π.

Theorem 10.1. *Y is an irreducible Markov chain with transition matrix*

$$\hat{p}_{i,j} = \frac{\pi_j}{\pi_i} p_{j,i}$$

and invariant distribution π.

Proof. First check that $\hat{P} = (\hat{p}_{i,j})$ is a stochastic matrix: the entries are non-negative and

$$\sum_j \hat{p}_{i,j} = \sum_j \frac{\pi_j}{\pi_i} p_{j,i} = \frac{1}{\pi_i} \sum_j \pi_j p_{j,i} = 1.$$

Claim $\pi = \pi \hat{P}$:

$$\sum_i \pi_i \hat{p}_{i,j} = \sum_i \pi_j p_{j,i} = \pi_j$$

Now to prove it is a Markov chain,

$$\begin{aligned}
\mathbb{P}(Y_0 = i_0, \dots, Y_n = i_n) &= \mathbb{P}(X_{N-n} = i_n, \dots, X_N = i_0) \\
&= \pi_{i_n} p_{i_n, i_{n-1}} \cdots p_{i_1, i_0} \\
&= \pi_{i_{n-1}} \hat{p}_{i_{n+1}, i_n} p_{i_{n-1}, i_{n-1}} \cdots, p_{i_1, i_n} \\
&= \pi_{i_0} \hat{p}_{i_0, i_1} \cdots \hat{p}_{i_{n-1}, i_n}
\end{aligned}$$

Hence Y has the stated properties. □

We call mY the *time-reversal* of X and we say X is *reversible* if Y and X have the same transition probabilities. By (*) (the equation is the statement of the theorem), X is recursive if and only if

$$\pi_i p_{i,j} = \pi_j p_{j,i}$$

for all $i, j \in S$. This is the detailed balance equation.

More generally, we say a transition matrix P and a distribution λ are in *detailed balance* if

$$\lambda_i p_{i,j} = \lambda_j p_{j,i}$$

for all $i, j \in S$. An irreducible chain X with invariant distribution π is called *reversible in equilibrium* if its P is in detailed balance with π.

Equation such as $\pi = \pi P$ can be difficult and may depend on some special structure on P. On the other hand the detailed balance equation is almost trivial.

Proposition 10.2. *If π is a distribution satisfying*

$$\pi_i p_{i,j} = \pi_j p_{j,i}$$

for all $i, j \in S$ and S is irreducible, then π is the only invariant distribution of the chain and the chain is reversible in equilibrium

Proof. Let π be a distribution satisfying the hypothesis. Then

$$\sum_i \pi_j p_{j,i} = \sum_j \pi_i p_{i,j} = \pi_i$$

as $(\pi P)_i = \pi_i$. Therefore $\pi = \pi P$. $\square$

Example (Birth-death with retaining barrier)**.** Try the detailed balance equation

$$\pi_{i-1} p_{i-1} = \pi_i q_i$$

so

$$\pi_i = \frac{p_{i-1}}{q_i} \frac{p_{i-2}}{q_{i-1}} \cdots \frac{p_0}{q_1} \pi_0 = \rho_i \pi_0 \tag{2}$$

$$\sum_i \pi_i = \pi_0 \sum_i \rho_i$$

If $S = \sum_i \rho_i$ satisfies $S < \infty$ then $\pi_i = \rho_i / S$ is an invariant distribution and if $S = \infty$ there is no invariant distribution.

11 Random Walk on a Graph

A finite graph consists of vertices and edges and is denoted $G = (V, E)$. We discuss graphs are simple (in which there are no parallel edges and loops) and connected. If $(u, v) \in E$ then v is called an *neighbour* of u. The *degree* of u, $d(u)$, is the number of its neighbours.

A random walk on G is a Markov chain with state space V and transition probability

$$p_{u,n} = \begin{cases} 0 & \text{if } v \text{ is not a neighbour of } u \\ \frac{1}{d(u)} & \text{if } v \text{ is a neighbour of } u \end{cases}$$

This is irreducible if and only if G is connected, which we assume henceforth.

As always, the natural question to ask is if there is an invariant distribution. Try to solve

$$\cdot p_{u,v} = \cdot p_{v,u}$$

for $(u, v) \in E$. We try to find "things" to multiply for the above relation to hold. The obvious choice is $\pi_u = d(u)$. But we have to normalise it since

$$\sum_u d(u) = 2|E|$$

Then $\pi_u = \frac{d(u)}{2|E|}$ satsifies the above detailed balance equation, and hence is the unique invariant distribution.

Example (Erratic Knight)**.** A knight performs independent legal knight moves about a 8×8 chessboard. This is a Markov chain on the state space S, the smallest square on hte board. (Exercise: show this is irreducible). The question is: what is its invariant distribution?

The answer is simple:

$$\pi_i = \frac{\text{No. of legal moves from square}}{336}$$

Exercise (Erratic Bishop)**.** There are two types of bishops, depending on the colour of the intial checkerboard. Consequently there are two commutative classes.