Markov Chains

Shing Tak Lam

May 12, 2022

Contents

1 Definitions

In this course, let I be a finite or countable set. The random variables will be defined in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1 (Markov Chain)
A stochastic process $(X_n)_{n\geq 0}$ is a Markov chain if for all $n \geq 0$, and $x_0, \dots, x_{n+1} \in I$, we have that

$$\mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0\right) = \mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right)$$

Definition 1.2 (Time homogeneous Markov chains)
If $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ is independent of n, then X is called time homogeneous. Otherwise it is called time inhomogeous.

Remark 1.3. In this course, we will only study time homogeneous Markov chains.

Definition 1.4 (Transition matrix)
The matrix P given by

$$P_{ij} = P(i, j) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

is known as the transition matrix of the Markov chain X.

Remark 1.5. P is a stochastic matrix, that is, the sum of each row is

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

In particular, this means that 1 is always an eigenvalue of P, with eigenvector $(1, \ldots, 1)^{\mathsf{T}}$.

Definition 1.6 (Markov Distribution)
We say that $X \sim \text{Markov}(\lambda, P)$ if $\lambda(x) = \mathbb{P}(X_0 = x)$ is a probability distribution on I, and P is the transition matrix of X.

Theorem 1.7. A stochastic process X is $\text{Markov}(\lambda, P)$ if and only if for all $n \geq 0$, $x_0, \ldots, x_n \in I$, we have that

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

Proof. For ($\implies$), condition and use Markov property. For ($\impliedby$), induct on n. $\square$

2 Simple Markov Property

Definition 2.1 (δ-mass)
For $i \in I$, define the probability mass function δ_i on I by

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

Theorem 2.2 (Simple Markov Property). Suppose $X \sim \text{Markov}(\lambda, P)$. Fix $m \in \mathbb{N}$ and $i \in I$. Then conditional on $X_m = i$, the process $(X_{m+n})_{n \geq 0}$ is $\text{Markov}(\delta_i, P)$, and independent of $X_0, \ldots, X_m$.

Proof. First, we need to show that it is a Markov chain. We use theorem 1.7. By the law of total probability, we have that

$$\begin{aligned}
\mathbb{P}(X_m = x_m, \ldots, X_{m+n} = x_{m+n}) &= \sum_{x_0, \ldots, x_{m-1} \in I} \mathbb{P}(X_0 = x_0, \ldots, X_{m-1} = x_{m-1}, X_m = x_m, \ldots, X_{m+n} = x_{m+n}) \\
&= \sum_{x_0, \ldots, x_{m-1} \in I} \lambda(x_0)P(x_0, x_1) \cdots P(x_{m-1}, x_m)P(x_m, x_{m+1}) \cdots P(x_{m+n-1}, x_{m+n}) \\
&= \mathbb{P}(X_m = x_m)P(x_m, x_{m+1}) \cdots P(x_{m+n-1}, x_{m+n})
\end{aligned}$$

Thus, the conditional probability is

$$
\begin{aligned}
\mathbb{P}(X_m = x_m, \ldots, X_{m+n} = x_{m+n} \mid X_m = i) &= \frac{\mathbb{P}(X_m = x_m, \ldots, X_{m+n} = x_{m+n}, X_m = i)}{\mathbb{P}(X_m = i)} \\
&= \frac{\delta_{ix_m}\mathbb{P}(X_m = x_m, \ldots, X_{m+n} = x_{m+n})}{\mathbb{P}(X_m = i)} \\
&= \frac{\delta_{ix_m}\mathbb{P}(X_m = x_m)P(x_m, x_{m+1}) \cdots P(x_{m+n-1}, x_{m+n})}{\mathbb{P}(X_m = i)} \\
&= \delta_{ix_m} P(x_m, x_{m+1}) \cdots P(x_{m+n-1}, x_{m+n})
\end{aligned}
$$

and this shows that $(X_{m+n})_n$ is Markov(δ_i, P). To show that this is independent to $X_0, \ldots, X_m$, let $m \leq i_1 \leq \cdots \leq i_k$, $x_0, \ldots, x_m, x_{i_1}, \ldots, x_{i_k} \in I$. Then

$$
\begin{aligned}
&\mathbb{P}(X_0 = x_0, \ldots, X_m = x_m, X_{i_i} = x_{i_1}, \ldots, X_{i_k} = x_{i_k} \mid X_m = i) \\
&\qquad = \frac{\mathbb{P}(X_0 = x_0, \ldots, X_m = x_m, X_{i_i} = x_{i_1}, \ldots, X_{i_k} = x_{i_k})}{\mathbb{P}(X_m = i)} \quad \text{where we assume that } x_m = i \\
&\qquad = \frac{\lambda_{x_0} P(x_0, x_1) \cdots P(x_{m-1}, x_m)\mathbb{P}(X_{i_1} = x_{i_1}, \ldots, X_{i_k} = x_{i_k} \mid X_m = i)}{\mathbb{P}(X_m = i)} \\
&\qquad = \mathbb{P}(X_0 = x_0, \ldots, X_m = x_m \mid X_m = i)\mathbb{P}(X_{m+1} = x_{m+1}, \ldots, X_{m+n} = x_{m+n} \mid X_m = i)
\end{aligned}
$$

□

3 Eigenvalues of the transition matrix

Notation 3.1. Define the notations

$$
\begin{aligned}
\mathbb{P}_i(\cdot) &= \mathbb{P}(\cdot \mid X_0 = i) \\
p_{ij}(n) &= (P^n)_{ij}
\end{aligned}
$$

Theorem 3.2. Suppose $X \sim$ Markov(λ, P). Considering X as a row vector, we have that

$$
\begin{aligned}
\mathbb{P}(X_n = x) &= (\lambda P^n)_x \\
\mathbb{P}(X_n = y \mid X_0 = x) &= (\delta_x P^n)_y
\end{aligned}
$$

Proof. Expand. □

Suppose P is a $k \times k$ matrix, and let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues. There are three cases.

- If all are real and distinct, then we can diagonalise $P = UDU^{-1}$. Then $P^n = UD^nU^{-1}$, and (WLOG) $p_{11}(n) = a_1\lambda_1^n + \cdots + a_k\lambda_k^n$ for some $a_1, \ldots, a_k$.
- If $\lambda_i, \overline{\lambda_i}$ are complex conjugate eigenvalues, say $\lambda_i = re^{i\theta}$, then $p_{11}(n) =$ (terms from other eigenvalues) $+$ $a\cos(n\theta) + b\sin(n\theta)$.
- If they are not distict, then we can consider the Jordan normal form of P, and we have terms of the form $q(n)\lambda^n$ where $q(n)$ is a polynomial, with degree one less than the algebraic multiplicity of λ.

4 Hitting times

Definition 4.1 (Hitting time, hitting probability, mean hitting time)
Let $A \subseteq I$, then the hitting time T_A is a random variable, known as the hitting time of A, defined by

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

where we take $\inf \varnothing = \infty$. The hitting probability h^A is

$$h^A = \mathbb{P}(T_A < \infty)$$

and the mean hitting time is

$$k^A = \mathbb{E}(T_A)$$

Remark 4.2. The hitting probability and mean hitting time starting at a given state, h_i^A and k_i^A are defined similarly, by replacing $\mathbb{P}$ with $\mathbb{P}_i$ and $\mathbb{E}$ with $\mathbb{E}_i$.

Theorem 4.3. Let $A \subseteq I$. Then the vector $(h_i^A)_{i \in I}$ is the minimal nonnegative solution to

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

Proof. First we show that (h_i^A) is a solution. Suppose $i \in A$, then $h_i^A = 1$. Now suppose $i \notin A$. Then we can write

$$\begin{aligned}
\mathbb{P}_i(T_A < \infty) &= \sum_{n=0}^{\infty} \mathbb{P}_i(T_A = n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_i(X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{P}(X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A, X_1 = j \mid X_0 = i) \qquad \text{(i)} \\
&= \sum_{j \in I} \mathbb{P}(X_1 \in A, X_1 = j \mid X_0 = i) + \sum_{n=2}^{\infty} \sum_{j \in I} \mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A, X_1 = j \mid X_0 = i) \\
&= \sum_j P(i,j) \mathbb{P}(X_1 \in A \mid X_1 = j, X_0 = i) + \\
&\quad \sum_j P(i,j) \sum_{n=2}^{\infty} \mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A \mid X_0 = i, X_1 = j) \qquad \text{(ii)} \\
&= \sum_j P(i,j) \mathbb{P}(X_0 \in A \mid X_0 = j) + \sum_j P(i,j) \sum_{n=2}^{\infty} \mathbb{P}(X_0 \notin A, \ldots, X_{n-2} \notin A, X_{n-1} \in A \mid X_0 = j) \\
&\qquad \text{(iii)} \\
&= \sum_j P(i,j) \sum_{n=0}^{\infty} \mathbb{P}_j(T_A = n) \\
&= \sum_j P(i,j) \mathbb{P}(T_A < \infty) \\
&= \sum_j P(i,j) h_i^A
\end{aligned}$$

Where for (i) we used the law of total probability, for (ii) we used properties of conditional probability, and for (iii) we used the Markov property.

Now we will show that h_i^A minimal. For $i \in A$ this clearly holds. Now fix $i \notin A$, and suppose x_i is another solution to the equations. Then

$$\begin{aligned}
x_i &= \sum_j P(i,j)x_j \\
&= \sum_{j_1 \in A} P(i,j_1)x_{j_1} + \sum_{j_1 \notin A} P(i,j_1)x_{j_1} \\
&= \sum_{j_1 \in A} P(i,j_1) + \sum_{j_1 \notin A, j_2 \in A} P(i,j_1)P(j_1,j_2) + \cdots \\
&\geq \sum_{n=1}^{N} \mathbb{P}_i(T_A = n)
\end{aligned}$$

for all N. Taking the limit $N \to \infty$ we get the required result. □

Theorem 4.4. The vector (k_i^A) is the minimal nonnegative solution to

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

Proof. The proof that k_i^A is a solution follows as in the case of the hitting probability. The proof for minimality follows in a similar fashion, where we use

$$\mathbb{E}(T) = \sum_{n=0}^{\infty} \mathbb{P}(T > n)$$

□

5 Strong Markov Property

Definition 5.1 (Stopping time)

A stopping time T is a random variable in $\mathbb{N} \cup \{\infty\}$, where for all $n \in \mathbb{N}$, the event $\{T = n\}$ depends only on $X_0, \ldots, X_n$.

Theorem 5.2 (Strong Markov Property). Let $X \sim \text{Markov}(\lambda, P)$, and T be a stopping time. Conditional on $T < \infty$ and $X_T = i$, we have that the process

$$(X_{T+m})_{m \geq 0} \sim \text{Markov}(\lambda, P)$$

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

Proof. We want to show that

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

For all $x_0, \ldots, x_n \in I$, and for all w. Suppose $T = k$, and $w = (w_0, \ldots, w_k) \in I^{k+1}$. Then the LHS of (*) becomes

$$\frac{\mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n, (X_0, \ldots, X_n) = w, T = k, X_k = i)}{\mathbb{P}(T < \infty, X_T = i)} \qquad (\dagger)$$

Applying the simple Markov property, we find that

$$\begin{aligned}\mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n \mid (X_0, \ldots, X_k) = w, T = k, X_k = i) \\ = \mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n, X_k = i) \quad \text{as } T \text{ is a stopping time} \\ = \delta_{ix_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n)\end{aligned}$$

Substituting this into the RHS of (*) and using † we get the required result. □

6 Communicating classes

Definition 6.1 (Leads to, communicates with)
Let $X \sim \text{Markov}(\lambda, P)$. Then for $x, y \in I$, we say that x leads to y, $x \to y$ if

$$\mathbb{P}_x(\exists n \text{ s.t. } X_n = y) > 0$$

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

Theorem 6.2. The following are equivalent.

(i) $x \to y$

(ii) There exists a sequence $x = x_0, \ldots, x_k = y$ s.t. $P(x_0, x_1) \cdots P(x_{k-1}, x_k) > 0$.

(iii) There exists n s.t. $p_{xy}(n) > 0$.

Proof. First we show that $(i) \iff (iii)$. Consider the events

$$\{\exists n \text{ s.t. } X_n = y\} = \bigcup_n X_n = y$$

If the LHS has positive probability, so must the RHS for some n. Conversely, if the RHS has a positive probability for some n, so must the LHS. To show that $(ii) \iff (iii)$, we note that

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

□

Corollary 6.3. $\leftrightarrow$ defines an equivalence relation.

Proof. Reflexivity follows by setting $n = 0$ in (iii). Symmetry is true by definition. Transitivity follows by (ii). □

Definition 6.4 (Communicating classes)
The equivalence classes of I under $\leftrightarrow$ are known as communicating classes.

Definition 6.5 (Closed)
A communicating class C is closed if when $x \in C$, $x \to y$, we have that $y \in C$. Equivalently, if $x \in C$ and $x \to y$, then $y \to x$.

Definition 6.6 (Irreducible)
A transition matrix P is irreducible if it has a single communicating class.

Definition 6.7 (Absorbing state)
A state $x \in I$ is called absorbing if $\{x\}$ is a closed class. Equivalently, there is no $y \neq x$ s.t. $x \to y$.

7 Transience and recurrence

Definition 7.1 (Transient)
A state $x \in I$ is transient if

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

Definition 7.2 (Recurrent)
A state $x \in I$ is recurrent if

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

Definition 7.3 (Return time)
Define the r-th return time inductively by

$$\begin{aligned} T_i^{(0)} &= 0 \\ T_i^{(r+1)} &= \inf\{n \geq T_i^{(r)} + 1 : X_n = i\} \end{aligned}$$

and we write $T_i = T_i^{(1)}$ for the first return time.

Definition 7.4 (Return probability)
The probability that the Markov chain will return to the state i is

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

Definition 7.5 (Visits)
The number of visits by the Markov chain to the state i is

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

Lemma 7.6. $\mathbb{P}_i(V_i > r) = f_i^r$ for all $r \in \mathbb{N}$.

Proof. By induction on r. The case $r = 0$ is true by definition as $V_i \geq 1$. Now suppose it holds for r. Then

$$
\begin{aligned}
\mathbb{P}_i(V_i \geq r+1) &= \mathbb{P}_i(T_i^{(r+1)} < \infty) \\
&= \mathbb{P}_i(T_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty)\mathbb{P}_i(T_i^{(r)} < \infty) \\
&= \mathbb{P}_i(T_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty) f_i^r \\
&= \mathbb{P}_i(T_i < \infty) f_i^r \quad \text{by Strong Markov on the stopping time } T_i^{(r)} \\
&= f_i^{r+1}
\end{aligned}
$$

□

Theorem 7.7. Let $X \sim \text{Markov}(\lambda, P)$, $i \in I$. Then we have that

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. First, we note that

$$\mathbb{E}_i(V_i) = \mathbb{E}_i\left(\sum_{n=1}^{\infty} \mathbb{1}(X_n = i)\right) = \sum_{n=1}^{\infty} \mathbb{E}_i(\mathbb{1}(X_n = i)) = \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = i) = \sum_{n=1}^{\infty} p_{ii}(n)$$

For (i), if $f_i = 1$, then for all r, $\mathbb{P}_i(V_i > r) = 1$. So $\mathbb{P}_i(V_i = \infty) = 1$. Hence i is recurrent, and $\mathbb{E}_i(V_i) = \infty$. For (ii), if $f_i < 1$, then $V_i \sim \text{Geo}(f_i)$. This has finite expectation, and $\mathbb{P}_i(V_i = \infty) = 0$. So i is transient. □

Theorem 7.8. Let $x, y \in I$, $x \leftrightarrow y$. Then either x, y are both recurrent, or x, y are both transient.

Proof. Since $x \leftrightarrow y$, we have m, l s.t. $p_{xy}(m), p_{yx}(l) > 0$. Suppose x recurrent. Then we have that

$$\sum_n p_{yy}(n) \geq \sum_n p_{yy}(n+m+l) \geq \sum_n p_{yx}(l)p_{xx}(n)p_{xy}(m) = p_{yx}(l)p_{xy}(m)\sum_n p_{xx}(n) = \infty$$

So y is recurrent as well. □

Corollary 7.9. Transience and recurrence is a communicating class property.

Theorem 7.10. Let C be a recurrent communicating class. Then C is closed.

Proof. Suppose not. Then we have $x \in C, y \notin C, x \to y, y \not\to x$. Let m be s.t. $p_{xy}(m) > 0$. Since if we ever reach y we cannot reach x again, we must have that

$$\mathbb{P}_i(V_i < \infty) \geq \mathbb{P}_i(X_m = y) > 0$$

So x is not recurrent. Contradiction. □

Theorem 7.11. A finite closed class is recurrent.

Proof. Let C be a finite closed class, $x \in C$. Then by (infinite) Pigeonhole, we must have $y \in C$ such that

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

Since $x \leftrightarrow y$, there must exist m such that $p_{yx}(m) > 0$. Then

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

Hence y is recurrent, so C is recurrent. □

Theorem 7.12. Let P be irreducible and recurrent (ie. the only communicating class is recurrent). Then for all x, y, $\mathbb{P}_x(T_y < \infty) = 1$.

Proof. Let m be such that $p_{xy}(m) > 0$. Then we have that

$$\begin{aligned}
\mathbb{P}_y(X_n = y \text{ for infinitely many } n) &= \sum_z \mathbb{P}_y(X_m = z, X_n = y \text{ for infinitely many } n) \\
&= \sum_z \mathbb{P}_z(X_n = y \text{ for infinitely many } n)p_{yz}(m) \quad \text{by Markov property} \qquad (*)
\end{aligned}$$

Now we note that

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

by Strong Markov on the stopping time T_y, under the probability measure $\mathbb{P}_z$. Substituting into (*), we find that

$$1 = \sum_z \mathbb{P}_z(T_y < \infty)p_{yz}(m)$$

Since $\sum_z p_{yz}(m) = 1$, $p_{yx}(m) > 0$ and $\mathbb{P}_z(T_y < \infty) \leq 1$ for all z, we must have that

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

□

7.1 Random walks on $\mathbb{Z}^d$

Definition 7.13 (Simple Random Walk on Z^d)
Define the SRW on Z^d to have the transition matrix

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

where the e_i are the basis vectors for $\mathbb{Z}^d$.

Theorem 7.14 (Polya). The simple random walk on $\mathbb{Z}^d$ is recurrent for $d = 1, 2$, and transient for $d \geq 3$.

Here, we will not prove the full theorem. We will prove it for $d = 1, 2, 3$. For $d > 3$, we can embed the SRW into a SRW on $\mathbb{Z}^3$, and the result follows.

Proof (d=1). In this case, $P(x, x + 1) = P(x, x - 1) = \frac{1}{2}$, and we will show that $\sum_n p_{00}(n)$ diverges. In this case,

$$p_{00}(n) = \mathbb{P}_0(X_n = 0) = \begin{cases} 0 & \text{if } n \text{ odd} \\ \binom{2m}{m}2^{-2m} & \text{if } n = 2m \end{cases}$$

Recall Stirling's Formula, $n! \sim n^n e^{-n}\sqrt{2\pi n}$ as $n \to \infty$. Substituting, we find that

$$\begin{aligned} p_{00}(2m) &= \frac{(2m)!}{m!m!}\left(\frac{1}{2}\right)^{2m} \\ &\sim \left(\frac{(2m)^{2m}e^{-2m}\sqrt{4\pi m}}{(m^m e^{-m}\sqrt{2\pi m})^2}\right)\frac{1}{2^{2m}} \\ &= \frac{1}{\sqrt{\pi m}} \end{aligned}$$

Let n_0 be s.t. $\forall n \geq n_0, p_{00}(2n) \geq \dfrac{1}{2\sqrt{\pi n}}$. Then

$$\sum_n p_{00}(n) \geq \sum_{n \geq n_0} p_{00}(2n) \geq \sum_{n\geq n_0} \frac{1}{2\sqrt{\pi n}}$$

diverges. □

Remark 7.15. In the asymmetric case,

$$p_{00}(2n) = \binom{2n}{n}(pq)^n \sim A\frac{(4pq)^n}{\sqrt{n}}$$

So if $p \neq q$, then $4pq < 1$, and we can upper bound this by a geometric series, so the random walk is transient.

Proof (d = 2). In $\mathbb{R}^2$, we can project and get two independent simple random walks. In particular, let

$$f(x,y) = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right) \quad \text{and} \quad f(X_n) = (X_n^+, X_n^-)$$

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

Proof of claim. We can write $X_n = \sum_{i=1}^n \xi_i$, where

$$\xi_i \overset{\text{iid}}{\sim} \text{Unif}\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\-1\end{pmatrix}\right\}$$

Then we have that

$$X_n^+ = \sum_{i=1}^n \frac{\xi_i^{(1)} + \xi_i^{(2)}}{\sqrt{2}} \quad \text{and} \quad X_n^- = \sum_{i=1}^n \frac{\xi_i^{(1)} - \xi_i^{(2)}}{\sqrt{2}}$$

From this we can show it is a simple random walk, and that the two are independent by considering the cases of ξ_i. □

Using the claim,

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

□

Proof (d = 3). In order to return 0 after $2n$ steps, we must have travelled i steps along e_1 and $-e_1$, j along e_2 and $-e_2$, and k along e_3 and $-e_3$, where $i+j+k = n$. Then given i, j, k, we have

$$\binom{2n}{i, i, j, j, k, k} = \frac{(2n)!}{i!i!j!j!k!k!}$$

different paths, each with probability $(\frac{1}{6})^{2n}$. So,

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

Claim.
$$\sum_{i+j+k=n} \binom{n}{i,j,k} \frac{1}{3^n} = 1$$

Proof of claim. This is the total probability for placing n balls into 3 bins uniformly at random. □

Furthermore, if $n = 3m$, then we have that

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

So we get that

$$p_{00}(6m) \le \binom{2n}{n} \frac{1}{2^{2n}} \binom{n}{m,m,m} \frac{1}{3^n} \sim \frac{A}{n^{3/2}} \quad \text{by Stirling}$$

Note that we can bound $p_{00}(6m-2)$ and $p_{00}(6m-4)$ by $p_{00}(6m)$, so $\sum_n p_{00}(2n)$ is finite as $\sum_n n^{-3/2}$ converges. Hence 0 is transient. □

8 Invariant distribution

Definition 8.1 (Invariant, (or Equilibrium, Stationary) distribution)
Let P be a transition matrix. Then a probability distribution π is invariant if $\pi = \pi P$.

Theorem 8.2. If $X \sim \text{Markov}(\pi, P)$, then $X_n \sim \pi$ for all n.

Proof. $\pi P^n = \pi$ for all n. □

Theorem 8.3. Suppose I finite, and there exists $i \in I$ such that $p_{ij}(n) \to \pi_j$ as $n \to \infty$. Then π is an invariant distribution.

Proof.
$$\sum_j \pi_j = \sum_j \lim_{n\to\infty} p_{ij}(n) = \lim_{n\to\infty} \sum_j p_{ij}(n) = \lim_{n\to\infty} 1 = 1$$

So π is a distribution. Furthermore,

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

□

Remark 8.4. This does not have to be true for I infinite.

Remark 8.5. By Perron-Frobenius, an invariant distribution always exists.

Definition 8.6 (Expected visits)
Fix $k \in I$. Then for $i \in I$, define

$$\nu_k(i) = \mathbb{E}_k\left(\sum_{n=0}^{T_k-1} \mathbb{1}(X_n = i)\right)$$

That is, the expected number of visits to i during an excursion from k.

Theorem 8.7. If P is irreducible and recurrent, then ν_k is an invariant measure. That is, $\nu_k = \nu_k P$. Furthermore, $\nu_k(k) = 1$, and $0 < \nu_k(i) < \infty$ for all i.

Proof. Since P is recurrent, we have that T_k finite with probability 1, and $X_{T_k} = X_0 = k$ by definition. So we have that

$$\begin{aligned}
\nu_k(i) &= \mathbb{E}_k\left(\sum_{n=1}^{T_k} \mathbb{1}(X_n = i)\right) \\
&= \mathbb{E}_k\left(\sum_{n=1}^{\infty} \mathbb{1}(X_n = i, T_k \geq n)\right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_k(X_n = i, T_k \geq n) \\
&= \sum_{n=1}^{\infty} \sum_j \mathbb{P}_k(X_n = i, X_{n-1} = j, T_k \geq n) \\
&= \sum_{n=1}^{\infty} \sum_j \mathbb{P}_k(X_n = i \mid X_{n-1} = j, T_k \geq n)\mathbb{P}_k(X_{n-1} = j, T_k \geq n)
\end{aligned}$$

Since T_k is a stopping time, $\{T_k \geq n\} = \{T_k \leq n-1\}^C$ only depends on $X_0, \dots, X_{n-1}$, we have by the Markov property that

$$\mathbb{P}_k(X_n = i \mid X_{n-1} = j, T_k \geq n) = \mathbb{P}_k(X_n = i \mid X_{n-1} = j) = P(j, i)$$

So

$$\begin{aligned}
\nu_k(i) &= \sum_{n=1}^{\infty} \sum_j P(j, i)\mathbb{P}_k(X_{n-1} = j, T_k \geq n) \\
&= \sum_j P(j, i) \sum_{n=1}^{\infty} \mathbb{P}_k(X_{n-1} = j, T_k \geq n) \\
&= \sum_j P(j, i)\mathbb{E}_k\left(\sum_{n=1}^{\infty} \mathbb{1}(X_n = j, T_k \geq n)\right) \\
&= \sum_j P(j, i)\nu_k(j)
\end{aligned}$$

Thus $\nu_k = \nu_k P$, and ν_k is an invariant measure. It is clear by definition that $\nu_k(k) = 1$. Since P irreducible, we must have n such that $p_{ki}(n) > 0$, and m such that $p_{ik}(m) > 0$. then

$$v_k(i) = \sum_j v_k(j) P^n(j,i) \geq v_k(k) P_{ki}(n) > 0$$

and

$$1 = v_k(k) = \sum_j v_k(j) P^m(j,k) \geq v_k(i) P^m(i,k)$$

So $v_k(i) \leq \dfrac{1}{P^m(i,k)} < \infty$ $\square$

Theorem 8.8. Let P be irreducible, λ be an invariant measure with $\lambda_k = 1$. Then $\lambda \geq v_k$. Furthermore, if P is recurrent, then $\lambda = v_k$.

Proof. Fix $i \in I$. Then $\lambda_i = \sum_j \lambda_j P(j,i)$ as $\lambda = \lambda P$. Thus,

$$\begin{aligned}
\lambda_i &= P(k,i) + \sum_{j_1 \neq k} \lambda_{j_1} P(j_1, i) \\
&= P(k,i) + \sum_{j_1 \neq k} P(j_1,i) + \cdots + \sum_{j_1,\ldots,j_{n-1} \neq k} P(j_{n-1}, j_{n-2}) \cdots P(j_2, j_1) P(j_1, i) \\
&\quad + \sum_{j_1,\ldots,j_n \neq k} \lambda_{j_n} P(j_n, j_{n-1}) \cdots P(j_2, j_1) P(j_1, i)
\end{aligned}$$

Since λ is a measure, $\lambda_x \geq 0$ for all x. So we must have that

$$\begin{aligned}
\lambda_i &\geq P(k,i) + \sum_{j_1 \neq k} P(j_1,i) + \cdots + \sum_{j_1,\ldots,j_{n-1} \neq k} P(j_{n-1}, j_{n-2}) \cdots P(j_2, j_1) P(j_1, i) \\
&= \mathbb{P}_k(X_1 = i, T_k \geq 1) + \cdots + \mathbb{P}_k(X_n = i, T_k \geq n) \\
&= \sum_{\ell=1}^{n} \mathbb{P}_k(X_\ell = i, T_k \geq \ell) \\
&\to \sum_{\ell=1}^{\infty} \mathbb{P}_k(X_\ell = i, T_k \geq \ell) \\
&= v_k(i)
\end{aligned}$$

So $\lambda_i \geq v_k(i)$ for all i. Now suppose if P is recurrent. Then v_k is an invariant measure, and so is $\lambda - v_k$. By irreducibility, we have m such that $p_{ik}(m) > 0$. Then

$$0 = \lambda_i - v_k(i) = \sum_j (\lambda_j - v_k(j)) p_{jk}(m) \geq (\lambda_i - v_k(i)) p_{ik}(m)$$

So we must have that $\lambda_i = v_k(i)$ for all i. $\square$

Remark 8.9. Thus, in the case where P is irreducible and recurrent, all invariant measures are unique up to multiplication by a constant.

Remark 8.10. Fix k. If $\sum_i v_k(i) < \infty$, then

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

is an invariant distribution. Furthermore,

$$\sum_j \nu_k(j) = \sum_j \mathbb{E}_k\left(\sum_{n=0}^{T_k-1} \mathbb{1}(X_n = j)\right) = \mathbb{E}_k\left(\sum_{n=0}^{T_k-1} \underbrace{\sum_i \mathbb{1}(X_n = i)}_{=1}\right) = \mathbb{E}_k(T_k)$$

Thus, if $\mathbb{E}_k(T_k)$ is finite, then we can normalise and get an invariant distribution.

Definition 8.11

Let $k \in I$ be recurrent. Then

- k is positive recurrent if $\mathbb{E}_k(T_k) < \infty$, and
- k is null recurrent if $\mathbb{E}_k(T_k) = \infty$.

Theorem 8.12. Suppose P is irreducible. Then the following are equivalent.

(i) Every state is positive recurrent.

(ii) Some state is positive recurrent.

(iii) P has an invariant distribution π.

If any of the above hold, then we also have that

$$\pi_i = \frac{1}{\mathbb{E}_i(T_i)}$$

Proof. (i) immediately implies (ii). To show that (ii) implies (iii), let k be positive recurrent. Then ν_k is an invariant measure, and

$$\pi_i = \frac{\nu_k(i)}{\mathbb{E}_k(T_k)}$$

is an invariant distribution. To show that (iii) implies (i), let k be any state. In addition, we must have i such that $\pi_i > 0$. By irreducibility, we must have n such that $p_{ik}(n) > 0$. Then

$$\pi_k = \sum_j \pi_j P^n(j,k) \geq \pi_i P^n(j,k) > 0$$

So $\pi_k > 0$ for all k. Define a new measure $\lambda_i = \dfrac{\pi_i}{\pi_k}$. Then this is an invariant measure, with $\lambda_k = 1$. So we get that $\lambda \geq \nu_k$, and

$$\mathbb{E}_k(T_k) = \sum_i \nu_k(i) \leq \sum_i \lambda_i = \sum_i \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty \tag{*}$$

So k is positive recurrent. Furthermore, suppose (i), (ii) and (iii) hold. Then P is recurrent, and $\lambda = \nu_k$, so we get equalities in (*), and

$$\pi_k = \frac{1}{\mathbb{E}_k(T_k)}$$

□

Corollary 8.13. If P is irreducible with invariant distribution π, then for all x, y,

$$v_x(y) = \frac{\pi_y}{\pi_x}$$

9 Reversability

Theorem 9.1. Suppose $X \sim \text{Markov}(\pi, P)$, where P is irreducible, and has invariant distribution π. Fix $N \in \mathbb{N}$. Then $(Y_n)_{0 \le n \le N}$ defined by $Y_n = X_{N-n}$ is a Markov chain, with transition matrix $\hat{P}$ given by

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

Moreover, $\hat{P}$ is irreducible, and has invariant distribution π.

Proof. First we check that $\hat{P}$ is a stochastic matrix.

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

Now we show that Y is a Markov chain. Let $y_0, \dots, y_N \in I$. Then

$$\begin{aligned}
\mathbb{P}(Y_0 = y_0, \dots, Y_N = y_N) &= \mathbb{P}(X_0 = y_N, \dots, X_N = y_0) \\
&= \pi(y_N) P(y_N, y_{N_1}) \cdots P(y_1, y_0) \\
&= \pi(y_0) \hat{P}(y_0, y_1) \cdots \hat{P}(y_{n-1}, y_N)
\end{aligned}$$

which shows that $Y \sim \text{Markov}(\pi, \hat{P})$. To show that π is invariant for $\hat{P}$, we have that

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

and finally to show that $\hat{P}$ is irreducible, we note that for $x, y \in I$, as P is irreducible we have $x = x_0, \dots, x_k = y$ such that $P(x_0, x_1), \dots, P(x_{k-1}, x_k) > 0$. Then

$$\hat{P}(x_k, x_{k-1}) \cdots \hat{P}(x_1, x_0) = \frac{\pi(x_0) P(x_0, x_1) \dots P(x_{k-1}, x_k)}{\pi(x_k)} > 0$$

So $\hat{P}$ is irreducible. □

Definition 9.2 (Reversible invariant distribution)
Let $X \sim \text{Markov}(\pi, P)$. Then the invariant distribution π is called reversible if $\hat{P} = P$. Equivalently, π satisfies the detailed balance equations

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

for all $x, y \in I$.

Remark 9.3. Equivalently, X is reversible if for all N,

$$(X_0, \dots, X_N) \overset{\text{distr.}}{=} (X_N, \dots, X_0)$$

Lemma 9.4. Let P be a transition matrix, μ a distribution satisfying the detailed balance equations. Then μ is an invariant distribution.

Proof.

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

□

10 Convergence to equilibrium

Definition 10.1 (Period)
Let P be a transition matrix, for $i \in I$, let

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

This is known as the period of i. If $d_i = 0$, then i is called aperiodic.

Lemma 10.2. $d_i = 1$ if and only if $P^n(i,i) > 0$ for all n sufficiently large.

Proof. Suppose $d_i = 1$. Let $D(i) = \{n \geq 1 : P^n(i,i) > 0\}$. First, we note that $D(i)$ is closed under addition and scalar multiplication. Let $r = \min_{m \neq n, m, n \in D(i)} |m - n|$. If $r = 1$ we are done. Suppose if $r \geq 2$. Let $m, n \in D(i)$ with $n = m + r$. Furthermore, as $d_i = 1$ we must also have k such that $k = \ell r + s$ with $0 < s < r$. Let $a = (\ell + 1)n$ and $b = (\ell + 1)m + k$, we have that a and b are in $D_(i)$, with

$$0 < a - b = r - s < r$$

Contradiction as r minimal. So $D(i)$ contains consecutive integers $n_1, n_1 + 1$. Then for $n \geq n_1^2$, $n \in D(i)$. Converse is clear. □

Lemma 10.3. Let P be irreducible, and i is aperiodic. Then j is aperiodic.

Proof. We have n, m such that $P^n(i,j), P^m(j,i) > 0$. Then

$$P^{n+m+r}(j,j) \geq P^m(j,i)P^r(i,i)P^n(i,j) > 0 \quad \text{for } r \text{ large}$$

□

Remark 10.4. Thus aperiodicity is a property of communicating classes.

Theorem 10.5. Let P be irreducible, aperiodic with invariant distribution π, and let $X \sim \text{Markov}(\lambda, P)$. Then for all $y \in I$,

$$\mathbb{P}(X_n = y) \to \pi(y) \quad \text{as} \quad n \to \infty$$

In particular, taking $\lambda = \delta_x$, we have that

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

Proof. Let $Y \sim \text{Markov}(\pi, P)$ be independent to X. Then (X_n, Y_n) defines a Markov chain on $I \times I$, with initial distribution $\lambda \times \pi$, and transition matrix

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

and invariant distribution $\tilde{\pi} = \pi \times \pi$. Now fix $a \in I$, and define

$$T = \inf\{n \geq 1 : (X_n, Y_n) = (a,a)\}$$

Claim.

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

Proof of claim. Suffices to show that $\tilde{P}$ is irreducible, as this means that $\tilde{P}$ is irreducible with an invariant distribution, and so is positive recurrent. Then T is a first return time, so $bbP(T < \infty) = 1$.

Let $x, x', y, y' \in I$. Then as P is irreducible, we have ℓ, m such that $P^\ell(x, x') > 0$ and $P^m(y, y') > 0$. Then

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

For n sufficiently large, by aperiodicity of P. Hence $\tilde{P}$ is irreducible. $\square$

Now define

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

Claim. $Z \sim \text{Markov}(\lambda, P)$

Proof of claim. $\mathbb{P}(Z_0 = x) = \mathbb{P}(X_0 = x) = \lambda(x)$, so suffices to show that Z is a Markov chain with the given transition matrix. Let $A = \{Z_{n-1} = z_{n-1}, \dots, Z_0 = z_0$. We want to show that $\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) = P(x, y)$.

$$\begin{aligned}\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) &= \mathbb{P}(Z_{n+1} = y, T > n \mid Z_n = x, A) \\ &\quad \mathbb{P}(Z_{n+1} = y, T \leq n \mid Z_n = x, A) \\ &= \mathbb{P}(X_{n+1} = y \mid T > n, Z_n = x, A)\mathbb{P}(T > n \mid Z_n = x, A) \\ &\quad \mathbb{P}(Y_{n+1} = y \mid T \leq n, Z_n = x, A)\mathbb{P}(T \leq n \mid Z_n = x, A)\end{aligned}$$

Now note the event $\{T > n\}$ depends only on $\{(X_0, Y_0), \dots, (X_n, Y_n)\}$, so we have that

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

So $\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) = P(x, y)$, and $Z \sim \text{Markov}(\lambda, P)$ as required. $\square$

Hence, we have that

$$\begin{aligned}|\mathbb{P}(X_n = y) - \pi(y)| &= |\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)| \\ &= |\mathbb{P}(X_n = y, n < T) + \mathbb{P}(Y_n = y, n > T) - \mathbb{P}(Y_n = y, n \leq T) - \mathbb{P}(Y_n = y, n > T)| \\ &\geq \mathbb{P}(n < T) \to 0 \quad \text{as} \quad n \to \infty \text{ since } \mathbb{P}(T < \infty) = 1\end{aligned}$$

$\square$

Theorem 10.6. Let P be irreducible, aperiodic, null-recurrent. Then for all x, y, $P^n(x, y) \to 0$ as $n \to \infty$.

Proof. Define $\tilde{P}$ as in the previous theorem. Then we have already shown that $\tilde{P}$ is irreducible, given P is irreducible.

Suppose if $\tilde{P}$ transient. Then $\sum_n \tilde{P}^n((x, y), (x, y)) < \infty$, but this implies that

$$\tilde{P}^n((x, y), (x, y)) = (P^n(x, y))^2 \to 0$$

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

Now suppose if $\tilde{P}$ recurrent. Fix $y \in I$, and define the probability measure $v_y(x)$ as before. We have shown that this is an invariant measure, and as P is null recurrent, we have that

$$v_y(I) = \mathbb{E}_y(T_y) = \infty$$

Then for all $M > 0$, there must exists $A \subseteq I$ finite such that $\nu_y(A) > M$. Define a probability measure

$$\mu(z) = \mu_{y,A}(z) = \frac{\nu_y(z)}{\nu_y(A)} \mathbb{1}(z \in A)$$

Then for $z \in A$, we have that

$$\begin{aligned}
\mu P^n(z) &= \sum_x \mu(z) P^n(x,z) \\
&= \sum_{x \in A} \frac{\nu_y(x)}{\nu_y(A)} P^n(x,z) \\
&\leq \frac{1}{\nu_y(A)} \sum_x \nu_y(x) P^n(x,z) \\
&= \frac{\nu_y(z)}{\nu_y(A)}
\end{aligned}$$

and note that the case where $z \notin A$ holds trivially. Let $(X, Y) \sim \text{Markov}(\mu \times \delta_x, \tilde{P})$, and define the stopping time

$$T = \inf n \geq 0 : (X_n, Y_n) = (x, x)$$

which is finite with probability 1. Let

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

We have shown that $Z \sim \text{Markov}(\mu, P)$. We have that $\mathbb{P}(Z_n = y) = \mu P^n(y) \leq \dfrac{\nu_y(y)}{\nu_y(A)} = \dfrac{1}{\nu_y(A)} < \dfrac{1}{M}$. Substituting, we find that

$$\begin{aligned}
P^n(x,y) &= \mathbb{P}_x(Y_n = y) \\
&= \mathbb{P}_x(Y_n = y, n \geq T) + \mathbb{P}_x(Y_n = y, T > n) \\
&\leq \mathbb{P}(Z_n = y) + \mathbb{P}(T > n)
\end{aligned}$$

This then means that

$$\limsup_{n \to \infty} \mathbb{P}_x(Y_n = y) \leq \frac{1}{M}$$

as $\mathbb{P}(T < \infty) = 1$. Since this holds for all $M > 0$, we get the required result. □