Part IB — Markov Chains

Theorems with proof

Based on lectures by G. R. Grimmett

Notes taken by Dexter Chua

Michaelmas 2015

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Discrete-time chains

Definition and basic properties, the transition matrix. Calculation of n-step transition probabilities. Communicating classes, closed classes, absorption, irreducibility. Calculation of hitting probabilities and mean hitting times; survival probability for birth and death chains. Stopping times and statement of the strong Markov property. [5]

Recurrence and transience; equivalence of transience and summability of n-step transition probabilities; equivalence of recurrence and certainty of return. Recurrence as a class property, relation with closed classes. Simple random walks in dimensions one, two and three. [3]

Invariant distributions, statement of existence and uniqueness up to constant multiples. Mean return time, positive recurrence; equivalence of positive recurrence and the existence of an invariant distribution. Convergence to equilibrium for irreducible, positive recurrent, aperiodic chains *and proof by coupling*. Long-run proportion of time spent in a given state. [3]

Time reversal, detailed balance, reversibility, random walk on a graph. [1]

Contents

0 Introduction

1 Markov chains

1.1 The Markov property

Proposition.

(i) λ is a *distribution*, i.e. $\lambda_i \geq 0, \sum_i \lambda_i = 1$.

(ii) P is a *stochastic matrix*, i.e. $p_{i,j} \geq 0$ and $\sum_j p_{i,j} = 1$ for all i.

Proof.

(i) Obvious since λ is a probability distribution.

(ii) $p_{i,j} \geq 0$ since p_{ij} is a probability. We also have

$$\sum_j p_{i,j} = \sum_j \mathbb{P}(X_1 = j \mid X_0 = i) = 1$$

since $\mathbb{P}(X_1 = \cdot \mid X_0 = i)$ is a probability distribution function. □

Theorem. Let λ be a distribution (on S) and P a stochastic matrix. The sequence $X = (X_0, X_1, \cdots)$ is a Markov chain with initial distribution λ and transition matrix P iff

$$\mathbb{P}(X_0 = i, X_1 = i_1, \cdots, X_n = i_n) = \lambda_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n} \qquad (*)$$

for all $n, i_0, \cdots, i_n$

Proof. Let A_k be the event $X_k = i_k$. Then we can write $(*)$ as

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

We first assume that X is a Markov chain. We prove $(*)$ by induction on n.

When $n = 0$, $(*)$ says $\mathbb{P}(A_0) = \lambda_{i_0}$. This is true by definition of λ.

Assume that it is true for all $n < N$. Then

$$\begin{aligned}
\mathbb{P}(A_0 \cap A_1 \cap \cdots \cap A_N) &= \mathbb{P}(A_0, \cdots, A_{N-1}) \mathbb{P}(A_0, \cdots, A_n \mid A_0, \cdots, A_{N-1})\\
&= \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{N-2},i_{N-1}} \mathbb{P}(A_N \mid A_0, \cdots, A_{N-1})\\
&= \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{N-2},i_{N-1}} \mathbb{P}(A_N \mid A_{N-1})\\
&= \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{N-2},i_{N-1}} p_{i_{N-1},i_N}.
\end{aligned}$$

So it is true for N as well. Hence we are done by induction.

Conversely, suppose that $(*)$ holds. Then for $n = 0$, we have $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$. Otherwise,

$$\begin{aligned}
\mathbb{P}(X_n = i_n \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}) &= \mathbb{P}(A_n \mid A_0 \cap \cdots \cap A_{n-1})\\
&= \frac{\mathbb{P}(A_0 \cap \cdots \cap A_n)}{\mathbb{P}(A_0 \cap \cdots \cap A_{n-1})}\\
&= p_{i_{n-1},i_n},
\end{aligned}$$

which is independent of $i_0, \cdots, i_{n-2}$. So this is Markov. □

Theorem (Extended Markov property)**.** Let X be a Markov chain. For $n \geq 0$, any H given in terms of the past $\{X_i : i < n\}$, and any F given in terms of the future $\{X_i : i > n\}$, we have

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

1.2 Transition probability

Theorem (Chapman-Kolmogorov equation)**.**

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

2 Classification of chains and states

2.1 Communicating classes

Proposition. $\leftrightarrow$ is an equivalence relation.

Proof.

(i) Reflexive: we have $i \leftrightarrow i$ since $p_{i,i}(0) = 1$.

(ii) Symmetric: trivial by definition.

(iii) Transitive: suppose $i \to j$ and $j \to k$. Since $i \to j$, there is some $m > 0$ such that $p_{i,j}(m) > 0$. Since $j \to k$, there is some n such that $p_{j,k}(n) > 0$. Then $p_{i,k}(m+n) = \sum_r p_{i,r}(m)p_{rk}(n) \geq p_{i,j}(m)p_{j,k}(n) > 0$. So $i \to k$.

Similarly, if $j \to i$ and $k \to j$, then $k \to i$. So $i \leftrightarrow j$ and $j \leftrightarrow k$ implies that $i \leftrightarrow k$. □

Proposition. A subset C is closed iff "$i \in C, i \to j$ implies $j \in C$".

Proof. Assume C is closed. Let $i \in C, i \to j$. Since $i \to j$, there is some m such that $p_{i,j}(m) > 0$. Expanding the Chapman-Kolmogorov equation, we have

$$p_{i,j}(m) = \sum_{i_1, \cdots, i_{m-1}} p_{i,i_1}p_{i_1,i_2}, \cdots, p_{i_{m-1},j} > 0.$$

So there is some route $i, i_1, \cdots, i_{m-1}, j$ such that $p_{i,i_1}, p_{i_1,i_2}, \cdots, p_{i_{m-1},j} > 0$. Since $p_{i,i_1} > 0$, we have $i_1 \in C$. Since $p_{i_1,i_2} > 0$, we have $i_2 \in C$. By induction, we get that $j \in C$.

To prove the other direction, assume that "$i \in C, i \to j$ implies $j \in C$". So for any $i \in C, j \notin C$, then $i \not\to j$. So in particular $p_{i,j} = 0$. □

2.2 Recurrence or transience

Theorem. i is recurrent iff $\sum_n p_{i,i}(n) = \infty$.

Theorem.

$$P_{i,j}(s) = \delta_{i,j} + F_{i,j}(s)P_{j,j}(s),$$

for $-1 < s \leq 1$.

Proof. Using the law of total probability

$$p_{i,j}(n) = \sum_{m=1}^n \mathbb{P}_i(X_n = j \mid T_j = m)\mathbb{P}_i(T_j = m)$$

Using the Markov property, we can write this as

$$\begin{aligned} &= \sum_{m=1}^n \mathbb{P}(X_n = j \mid X_m = j)\mathbb{P}_i(T_j = m)\\ &= \sum_{m=1}^n p_{j,j}(n-m)f_{i,j}(m). \end{aligned}$$

We can multiply through by s^n and sum over all n to obtain

$$\sum_{n=1}^{\infty} p_{i,j}(n)s^n = \sum_{n=1}^{\infty}\sum_{m=1}^{n} p_{j,j}(n-m)s^{n-m} f_{i,j}(m)s^m.$$

The left hand side is *almost* the generating function $P_{i,j}(s)$, except that we are missing an $n = 0$ term, which is $p_{i,j}(0) = \delta_{i,j}$. The right hand side is the "convolution" of the power series $P_{j,j}(s)$ and $F_{i,j}(s)$, which we can write as the product $P_{j,j}(s)F_{i,j}(s)$. So

$$P_{i,j}(s) - \delta_{i,j} = P_{i,j}(s)F_{i,j}(s).$$

□

Lemma (Abel's lemma)**.** Let $u_1, u_2, \cdots$ be real numbers such that $U(s) = \sum_n u_n s^n$ converges for $0 < s < 1$. Then

$$\lim_{s\to 1^-} U(s) = \sum_n u_n.$$

Theorem. i is recurrent iff $\sum_n p_{ii}(n) = \infty$.

Proof. Using $j = i$ in the above formula, for $0 < s < 1$, we have

$$P_{i,i}(s) = \frac{1}{1 - F_{i,i}(s)}.$$

Here we need to be careful that we are not dividing by 0. This would be a problem if $F_{ii}(s) = 1$. By definition, we have

$$F_{i,i}(s) = \sum_{n=1}^{\infty} f_{i,i}(n)s^n.$$

Also, by definition of f_{ii}, we have

$$F_{i,i}(1) = \sum_n f_{i,i}(n) = \mathbb{P}(\text{ever returning to } i) \leq 1.$$

So for $|s| < 1$, $F_{i,i}(s) < 1$. So we are not dividing by zero. Now we use our original equation

$$P_{i,i}(s) = \frac{1}{1 - F_{i,i}(s)},$$

and take the limit as $s \to 1$. By Abel's lemma, we know that the left hand side is

$$\lim_{s\to 1} P_{i,i}(s) = P_{i,i}(1) = \sum_n p_{i,i}(n).$$

The other side is

$$\lim_{s\to 1} \frac{1}{1 - F_{i,i}(s)} = \frac{1}{1 - \sum f_{i,i}(n)}.$$

Hence we have

$$\sum_n p_{i,i}(n) = \frac{1}{1 - \sum f_{i,i}(n)}.$$

Since $\sum f_{i,i}(n)$ is the probability of ever returning, the probability of ever returning is 1 if and only if $\sum_n p_{i,i}(n) = \infty$. □

Theorem. Let C be a communicating class. Then

(i) Either every state in C is recurrent, or every state is transient.

(ii) If C contains a recurrent state, then C is closed.

Proof.

(i) Let $i \leftrightarrow j$ and $i \neq j$. Then by definition of communicating, there is some m such that $p_{i,j}(m) = \alpha > 0$, and some n such that $p_{j,i}(n) = \beta > 0$. So for each k, we have

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

So if $\sum_k p_{j,j}(k) = \infty$, then $\sum_r p_{i,i}(r) = \infty$. So j recurrent implies i recurrent. Similarly, i recurrent implies j recurrent.

(ii) If C is not closed, then there is a non-zero probability that we leave the class and never get back. So the states are not recurrent. □

Theorem. In a finite state space,

(i) There exists at least one recurrent state.

(ii) If the chain is irreducible, every state is recurrent.

Proof. (ii) follows immediately from (i) since if a chain is irreducible, either all states are transient or all states are recurrent. So we just have to prove (i).

We first fix an arbitrary i. Recall that

$$P_{i,j}(s) = \delta_{i,j} + P_{j,j}(s)F_{i,j}(s).$$

If j is transient, then $\sum_n p_{j,j}(n) = P_{j,j}(1) < \infty$. Also, $F_{i,j}(1)$ is the probability of ever reaching j from i, and is hence finite as well. So we have $P_{i,j}(1) < \infty$. By Abel's lemma, $P_{i,j}(1)$ is given by

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

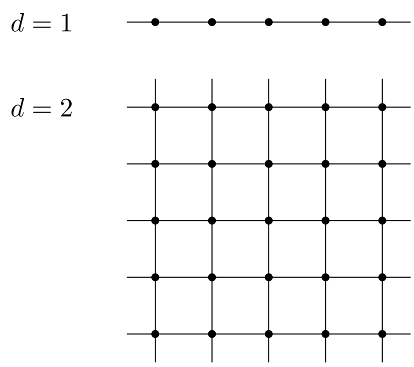
Since this is finite, we must have $p_{i,j}(n) \to 0$.

Since we know that

$$\sum_{j\in S} p_{i,j}(n) = 1,$$

If every state is transient, then since the sum is finite, we know $\sum p_{i,j}(n) \to 0$ as $n \to \infty$. This is a contradiction. So we must have a recurrent state. □

Theorem (Pólya's theorem)**.** Consider $\mathbb{Z}^d = \{(x_1, x_2, \cdots, x_d) : x_i \in \mathbb{Z}\}$. This generates a graph with x adjacent to y if $|x - y| = 1$, where $| \cdot |$ is the Euclidean norm.



Consider a random walk in $\mathbb{Z}^d$. At each step, it moves to a neighbour, each chosen with equal probability, i.e.

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \frac{1}{2d} & |j - i| = 1 \\ 0 & \text{otherwise} \end{cases}$$

This is an irreducible chain, since it is possible to get from one point to any other point. So the points are either all recurrent or all transient.

The theorem says this is recurrent iff $d = 1$ or 2.

Proof. We will start with the case $d = 1$. We want to show that $\sum p_{0,0}(n) = \infty$. Then we know the origin is recurrent. However, we can simplify this a bit. It is impossible to get back to the origin in an odd number of steps. So we can instead consider $\sum p_{0,0}(2n)$. However, we can write down this expression immediately. To return to the origin after $2n$ steps, we need to have made n steps to the left, and n steps to the right, in any order. So we have

$$p_{0,0}(2n) = \mathbb{P}(n \text{ steps left}, n \text{ steps right}) = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}.$$

To show if this converges, it is not helpful to work with these binomial coefficients and factorials. So we use Stirling's formula $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$. If we plug this in, we get

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

This tends to 0, but really slowly, and even more slowly than the harmonic series. So we have $\sum p_{0,0}(2n) = \infty$.

In the $d = 2$ case, suppose after $2n$ steps, I have taken r steps right, ℓ steps left, u steps up and d steps down. We must have $r + \ell + u + d = 2n$, and we need

$r = \ell, u = d$ to return the origin. So we let $r = \ell = m, u = d = n - m$. So we get

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

We now use a well-known identity (proved in IA Numbers and Sets) to obtain

$$\begin{aligned}
&= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}\binom{2n}{n}\\
&= \left[\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right]^2\\
&\sim \frac{1}{\pi n}.
\end{aligned}$$

So the sum diverges. So this is recurrent. Note that the two-dimensional probability turns out to be the square of the one-dimensional probability. This is not a coincidence, and we will explain this after the proof. However, this does not extend to higher dimensions.

In the $d = 3$ case, we have

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

This time, there is no neat combinatorial formula. Since we want to show this is summable, we can try to bound this from above. We have

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

Why do we write it like this? We are going to use the identity $\displaystyle\sum_{i+j+k=n} \frac{n!}{3^n i!j!k!} = 1$. Where does this come from? Suppose we have three urns, and throw n balls into it. Then the probability of getting i balls in the first, j in the second and k in the third is exactly $\frac{n!}{3^n i!j!k!}$. Summing over all possible combinations of i, j and k gives the total probability of getting in any configuration, which is 1. So we can bound this by

$$\begin{aligned}
&\leq \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \max\left(\frac{n!}{3^n i!j!k!}\right) \sum \frac{n!}{3^n i!j!k!}\\
&= \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \max\left(\frac{n!}{3^n i!j!k!}\right)
\end{aligned}$$

To find the maximum, we can replace the factorial by the gamma function and use Lagrange multipliers. However, we would just argue that the maximum can be achieved when i, j and k are as close to each other as possible. So we get

$$\begin{aligned} &\leq \left(\frac{1}{2}\right)^{2n}\binom{2n}{n}\frac{n!}{3^n}\left(\frac{1}{\lfloor n/3\rfloor !}\right)^3 \\ &\leq Cn^{-3/2} \end{aligned}$$

for some constant C using Stirling's formula. So $\sum p_{0,0}(2n) < \infty$ and the chain is transient. We can prove similarly for higher dimensions. □

2.3 Hitting probabilities

Theorem. The vector $(h_i^A : i \in S)$ satisfies

$$h_i^A = \begin{cases} 1 & i \in A \\ \sum_{j\in S} p_{i,j} h_j^A & i \not\in A \end{cases},$$

and is *minimal* in that for any non-negative solution $(x_i : i \in S)$ to these equations, we have $h_i^A \leq x_i$ for all i.

Proof. By definition, $h_i^A = 1$ if $i \in A$. Otherwise, we have

$$h_i^A = \mathbb{P}_i(H^A < \infty) = \sum_{j\in S}\mathbb{P}_i(H^A < \infty \mid X_1 = j)p_{i,j} = \sum_{j \in S} h_j^A p_{i,j}.$$

So h_i^A is indeed a solution to the equations.

To show that h_i^A is the minimal solution, suppose $x = (x_i : i \in S)$ is a non-negative solution, i.e.

$$x_i^A = \begin{cases} 1 & i \in A \\ \sum_{j\in S} p_{i,j} x_j^A & i \not\in A \end{cases},$$

If $i \in A$, we have $h_i^A = x_i = 1$. Otherwise, we can write

$$\begin{aligned} x_i &= \sum_j p_{i,j}x_j \\ &= \sum_{j\in A} p_{i,j}x_j + \sum_{j\not\in A} p_{i,j}x_j \\ &= \sum_{j\in A} p_{i,j} + \sum_{j\not\in A} p_{i,j}x_j \\ &\geq \sum_{j\in A} p_{i,j} \\ &= \mathbb{P}_i(H^A = 1). \end{aligned}$$

By iterating this process, we can write

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

By induction, we obtain

$$x_i \geq \mathbb{P}_i(H^A \leq n)$$

for all n. Taking the limit as $n \to \infty$, we get

$$x_i \geq \mathbb{P}_i(H^A \leq \infty) = h_i^A.$$

So h_i^A is minimal. □

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

$$k_i^A = \begin{cases} 0 & i \in A\\ 1 + \sum_j p_{i,k}k_j^A & i \notin A.\end{cases}$$

Proof. The proof that (k_i^A) satisfies the equations is the same as before.

Now let $(y_i : i \in S)$ be a non-negative solution. We show that $y_i \geq k_i^A$.

If $i \in A$, we get $y_i = k_i^A = 0$. Otherwise, suppose $i \notin A$. Then we have

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

By induction, we know that

$$y_i \geq \mathbb{P}_i(H^A \geq 1) + \cdots + \mathbb{P}_i(H^A \geq n)$$

for all n. Let $n \to \infty$. Then we get

$$y_i \geq \sum_{m\geq 1} \mathbb{P}_i(H^A \geq m) = \sum_{m\geq 1} m\mathbb{P}_i(H^a = m) = k_i^A.$$ □

2.4 The strong Markov property and applications

Theorem (Strong Markov property)**.** Let X be a Markov chain with transition matrix P, and let T be a stopping time for X. Given $T < \infty$ and $X_T = i$, the chain $(X_{T+k} : k \geq 0)$ is a Markov chain with transition matrix P with initial distribution $X_{T+0} = i$, and this Markov chain is independent of $X_0, \cdots, X_T$.

2.5 Further classification of states

Theorem. Suppose $X_0 = i$. Let $V_i = |\{n \geq 1 : X_n = i\}|$. Let $f_{i,i} = \mathbb{P}_i(T_i < \infty)$. Then

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

since we have to return r times, each with probability $f_{i,i}$, and then never return.

Hence, if $f_{i,i} = 1$, then $\mathbb{P}_i(V_i = r) = 0$ for all r. So $\mathbb{P}_i(V_i = \infty) = 1$. Otherwise, $\mathbb{P}_i(V_i = r)$ is a genuine geometric distribution, and we get $\mathbb{P}_i(V_i < \infty) = 1$.

Proof. Exercise, using the strong Markov property. □

Theorem. If $i \leftrightarrow j$ are communicating, then

(i) $d_i = d_j$.

(ii) i is recurrent iff j is recurrent.

(iii) i is positive recurrent iff j is positive recurrent.

(iv) i is ergodic iff j is ergodic.

Proof.

(i) Assume $i \leftrightarrow j$. Then there are $m, n \geq 1$ with $p_{i,j}(m), p_{j,i}(n) > 0$. By the Chapman-Kolmogorov equation, we know that

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

where $\alpha = p_{i,j}(m)p_{j,i}(n) > 0$. Now let $D_j = \{r \geq 1 : p_{j,j}(r) > 0\}$. Then by definition, $d_j = \gcd D_j$.

We have just shown that if $r \in D_j$, then we have $m + r + n \in D_i$. We also know that $n + m \in D_i$, since $p_{i,i}(n + m) \geq p_{i,j}(n)p_{ji}(m) > 0$. Hence for any $r \in D_j$, we know that $d_i \mid m + r + n$, and also $d_i \mid m + n$. So $d_i \mid r$. Hence $\gcd D_i \mid \gcd D_j$. By symmetry, $\gcd D_j \mid \gcd D_i$ as well. So $\gcd D_i = \gcd D_j$.

(ii) Proved before.

(iii) This is deferred to a later time.

(iv) Follows directly from (i), (ii) and (iii) by definition. □

Proposition. If the chain is irreducible and $j \in S$ is recurrent, then

$$\mathbb{P}(X_n = j \text{ for some } n \geq 1) = 1,$$

regardless of the distribution of X_0.

Proof. Let

$$f_{i,j} = \mathbb{P}_i(X_n = j \text{ for some } n \geq 1).$$

Since $j \to i$, there exists a *least* integer $m \geq 1$ with $p_{j,i}(m) > 0$. Since m is least, we know that

$$p_{j,i}(m) = \mathbb{P}_j(X_m = i, X_r \neq j \text{ for } r < m).$$

This is since we cannot visit j in the path, or else we can truncate our path and get a shorter path from j to i. Then

$$p_{j,i}(m)(1 - f_{i,j}) \leq 1 - f_{j,j}.$$

This is since the left hand side is the probability that we first go from j to i in m steps, and then never go from i to j again; while the right is just the probability of never returning to j starting from j; and we know that it is easier to just not get back to j than to go to i in exactly m steps and never returning to j. Hence if $f_{j,j} = 1$, then $f_{i,j} = 1$.

Now let $\lambda_k = \mathbb{P}(X_j = k)$ be our initial distribution. Then

$$\mathbb{P}(X_n = j \text{ for some } n \geq 1) = \sum_i \lambda_i \mathbb{P}_i(X_n = j \text{ for some } n \geq 1) = 1. \qquad \square$$

3 Long-run behaviour

3.1 Invariant distributions

Theorem. Consider an irreducible Markov chain. Then

(i) There exists an invariant distribution if some state is positive recurrent.

(ii) If there is 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 of i. In particular, π is unique.

Proposition. For an irreducible recurrent chain and $k \in S$, $\rho = (\rho_i : i \in S)$ defined as above by
$$\rho_i = \mathbb{E}_k(W_i), \quad W_i = \sum_{m=1}^{\infty} 1(X_m = i, T_k \geq m),$$
we have

(i) $\rho_k = 1$

(ii) $\sum_i \rho_i = \mu_k$

(iii) $\rho = \rho P$

(iv) $0 < \rho_i < \infty$ for all $i \in S$.

Proof.

(i) This follows from definition of ρ_i, since for $m < T_k$, $X_m \neq k$.

(ii) Note that $\sum_i W_i = T_k$, since in each step we hit exactly one thing. We have
$$\begin{aligned}\sum_i \rho_i &= \sum_i \mathbb{E}_k(W_i)\\ &= \mathbb{E}_k\left(\sum_i W_i\right)\\ &= \mathbb{E}_k(T_k)\\ &= \mu_k.\end{aligned}$$

Note that we secretly swapped the sum and expectation, which is in general bad because both are potentially infinite sums. However, there is a theorem (bounded convergence) that tells us this is okay whenever the summands are non-negative, which is left as an Analysis exercise.

(iii) We have

$$\begin{aligned}
\rho_j &= \mathbb{E}_k(W_j)\\
&= \mathbb{E}_k\left(\sum_{m\geq 1} 1(X_m = j, T_k \geq m)\right)\\
&= \sum_{m \geq 1}\mathbb{P}_k(X_m = j, T_k \geq m)\\
&= \sum_{m\geq 1}\sum_{i \in S}\mathbb{P}_k(X_m = j \mid X_{m-1} = i, T_k \geq m)\mathbb{P}_k(X_{m-1} = i, T_k \geq m)
\end{aligned}$$

We now use the Markov property. Note that $T_k \geq m$ means $X_1, \cdots, X_{m-1}$ are all not k. The Markov property thus tells us the condition $T_k \geq m$ is useless. So we are left with

$$\begin{aligned}
&= \sum_{m\geq 1}\sum_{i \in S}\mathbb{P}_k(X_m = j \mid X_{m-1} = i)\mathbb{P}_k(X_{m-1} = i, T_k \geq m)\\
&= \sum_{m\geq 1}\sum_{i \in S}p_{i,j}\mathbb{P}_k(X_{m-1} = i, T_k \geq m)\\
&= \sum_{i \in S}p_{i,j}\sum_{m\geq 1}\mathbb{P}_k(X_{m-1} = i, T_k \geq m)
\end{aligned}$$

The last term looks really ρ_i, but the indices are slightly off. We shall have faith in ourselves, and show that this is indeed equal to ρ_i.

First we let $r = m - 1$, and get

$$\sum_{m\geq 1}\mathbb{P}_k(X_{m-1} = i, T_k \geq m) = \sum_{r = 0}^\infty \mathbb{P}_k(X_r = i, T_k \geq r + 1).$$

Of course this does not fix the problem. We will look at the different possible cases. First, if $i = k$, then the $r = 0$ term is 1 since $T_k \geq 1$ is always true by definition and $X_0 = k$, also by construction. On the other hand, the other terms are all zero since it is impossible for the return time to be greater or equal to $r + 1$ if we are at k at time r. So the sum is 1, which is ρ_k.

In the case where $i \neq k$, first note that when $r = 0$ we know that $X_0 = k \neq i$. So the term is zero. For $r \geq 1$, we know that if $X_r = i$ and $T_k \geq r$, then we must also have $T_k \geq r + 1$, since it is impossible for the return time to k to be exactly r if we are not at k at time r. So $\mathbb{P}_k(X_r = i, T_k \geq r + 1) = \mathbb{P}_k(X_r = i, T_k \geq r)$. So indeed we have

$$\sum_{m\geq 0}\mathbb{P}_k(X_{m-1} = i, T_k \geq m) = \rho_i.$$

Hence we get

$$\rho_j = \sum_{i \in S}p_{ij}\rho_i.$$

So done.

(iv) To show that $0 < \rho_i < \infty$, first fix our i, and note that $\rho_k = 1$. We know that $\rho = \rho P = \rho P^n$ for $n \geq 1$. So by expanding the matrix sum, we know that for any m, n,

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

By irreducibility, we now choose m, n such that $p_{i,k}(m), p_{k,i}(n) > 0$. So we have

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

Since $\rho_k = 1$, the result follows. □

Theorem. Consider an irreducible Markov chain. Then

(i) There exists an invariant distribution if and only if some state is positive recurrent.

(ii) If there is 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 of i. In particular, π is unique.

Proof.

(i) Let k be a positive recurrent state. Then

$$\pi_i = \frac{\rho_i}{\mu_k}$$

satisfies $\pi_i \geq 0$ with $\sum_i \pi_i = 1$, and is an invariant distribution.

(ii) Let π be an invariant distribution. We first show that all entries are non-zero. For all n, we have

$$\pi = \pi P^n.$$

Hence for all $i, j \in S$, $n \in \mathbb{N}$, we have

$$\pi_i \geq \pi_j p_{j,i}(n). \tag{$*$}$$

Since $\sum \pi_1 = 1$, there is some k such that $\pi_k > 0$.

By $(*)$ with $j = k$, we know that

$$\pi_i \geq \pi_k p_{k,i}(n) > 0$$

for some n, by irreducibility. So $\pi_i > 0$ for all i.

Now we show that all states are positive recurrent. So we need to rule out the cases of transience and null recurrence.

So assume all states are transient. So $p_{j,i}(n) \to 0$ for all $i, j \in S$, $n \in \mathbb{N}$. However, we know that

$$\pi_i = \sum_j \pi_j p_{j,i}(n).$$

If our state space is finite, then since $p_{j,i}(n) \to 0$, the sum tends to 0, and we reach a contradiction, since π_i is non-zero. If we have a countably infinite set, we have to be more careful. We have a huge state space S, and we don't know how to work with it. So we approximate it by a finite F, and split S into F and $S \setminus F$. So we get

$$\begin{aligned} 0 &\leq \sum_j \pi_j p_{j,i}(n) \\ &= \sum_{j \in F} \pi_j p_{j,i}(n) + \sum_{j \notin F} \pi_j p_{j,i}(n) \\ &\leq \sum_{j \in F} p_{j,i}(n) + \sum_{j \notin F} \pi_j \\ &\to \sum_{j \notin F} \pi_j \end{aligned}$$

as we take the limit $n \to \infty$. We now want to take the limit as $F \to S$. We know that $\sum_{j \in S} \pi_i = 1$. So as we put more and more things into F, $\sum_{j \notin F} \pi_i \to 0$. So $\sum \pi_j p_{j,i}(n) \to 0$. So we get the desired contradiction. Hence we know that all states are recurrent.

To rule out the case of null recurrence, recall that in the previous discussion, we said that we "should" have $\pi_i \mu_i = 1$. So we attempt to prove this. Then this would imply that μ_i is finite since $\pi_i > 0$.

By definition $\mu_i = \mathbb{E}_i(T_i)$, and we have the general formula

$$\mathbb{E}(N) = \sum_n \mathbb{P}(N \geq n).$$

So we get

$$\pi_i \mu_i = \sum_{n=1}^{\infty} \pi_i \mathbb{P}_i(T_i \geq n).$$

Note that $\mathbb{P}_i$ is a probability conditional on starting at i. So to work with the expression $\pi_i \mathbb{P}_i(T_i \geq n)$, it is helpful to let π_i be the probability of starting at i. So suppose X_0 has distribution π. Then

$$\pi_i \mu_i = \sum_{n=1}^{\infty} \mathbb{P}(T_i \geq n, X_0 = i).$$

Let's work out what the terms are. What is the first term? It is

$$\mathbb{P}(T_i \geq 1, X_0 = i) = \mathbb{P}(X_0 = i) = \pi_i,$$

since we know that we always have $T_i \geq 1$ by definition.

For other $n \geq 2$, we want to compute $\mathbb{P}(T_i \geq n, X_0 = i)$. This is the probability of starting at i, and then not return to i in the next $n-1$ steps. So we have

$$\mathbb{P}(T_i \geq n, X_0 = i) = \mathbb{P}(X_0 = i, X_m \neq i \text{ for } 1 \leq m \leq n-1)$$

Note that all the expressions on the right look rather similar, except that the first term is $= i$ while the others are $\neq i$. We can make them look more similar by writing

$$\begin{aligned}\mathbb{P}(T_i \geq n, X_0 = i) = \mathbb{P}(X_m \neq i \text{ for } 1 \leq m \leq n-1) \\ - \mathbb{P}(X_m \neq i \text{ for } 0 \leq m \leq n-1)\end{aligned}$$

What can we do now? The trick here is to use invariance. Since we started with an invariant distribution, we always live in an invariant distribution. Looking at the time interval $1 \leq m \leq n-1$ is the same as looking at $0 \leq m \leq n-2$). In other words, the sequence $(X_0, \cdots, X_{n-2})$ has the same distribution as $(X_1, \cdots, X_{n-1})$. So we can write the expression as

$$\mathbb{P}(T_i \geq n, X_0 = i) = a_{n-2} - a_{n-1},$$

where

$$a_r = \mathbb{P}(X_m \neq i \text{ for } 0 \leq i \leq r).$$

Now we are summing differences, and when we sum differences everything cancels term by term. Then we have

$$\pi_i \mu_i = \pi_i + (a_0 - a_1) + (a_1 - a_2) + \cdots$$

Note that we cannot do the cancellation directly, since this is an infinite sum, and infinity behaves weirdly. We have to look at a finite truncation, do the cancellation, and take the limit. So we have

$$\begin{aligned}\pi_i \mu_i &= \lim_{N \to \infty} [\pi_i + (a_0 - a_1) + (a_1 - a_2) + \cdots + (a_{N-2} - a_{N-1})] \\ &= \lim_{N \to \infty} [\pi_i + a_0 - a_{N-1}] \\ &= \pi_i + a_0 + \lim_{N \to \infty} a_N.\end{aligned}$$

What is each term? π_i is the probability that $X_0 = i$, and a_0 is the probability that $X_0 \neq i$. So we know that $\pi_i + a_0 = 1$. What about $\lim a_N$? We know that

$$\lim_{N \to \infty} a_N = \mathbb{P}(X_m \neq i \text{ for all } m).$$

Since the state is recurrent, the probability of never visiting i is 0. So we get

$$\pi_i \mu_i = 1.$$

Since $\pi_i > 0$, we get $\mu_i = \frac{1}{\pi_i} < \infty$ for all i. Hence we have positive recurrence. We have also proved the formula we wanted. □

3.2 Convergence to equilibrium

Theorem. Consider a Markov chain that is irreducible, positive recurrent and aperiodic. Then

$$p_{i,k}(n) \to \pi_k$$

as $n \to \infty$, where π is the unique invariant distribution.

Proof. (non-examinable) The idea of the proof is to show that for any $i, j, k \in S$, we have $p_{i,k}(n) \to p_{j,k}(n)$ as $n \to \infty$. Then we can argue that no matter where we start, we will tend to the same distribution, and hence any distribution tends to the same distribution as π, since π doesn't change.

As mentioned, instead of working with probability distributions, we will work with the chains themselves. In particular, we have *two* Markov chains, and we imagine one starts at i and the other starts at j. To do so, we define the pair $Z = (X, Y)$ of *two* independent chains, with $X = (X_n)$ and $Y = (Y_n)$ each having the state space S and transition matrix P.

We can let $Z = (Z_n)$, where $Z_n = (X_n, Y_n)$ is a Markov chain on state space S^2. This has transition probabilities

$$p_{ij,k\ell} = p_{i,k}p_{j,\ell}$$

by independence of the chains. We would like to apply theorems to Z, so we need to make sure it has nice properties. First, we want to check that Z is irreducible. We have

$$p_{ij,k\ell}(n) = p_{i,k}(n)p_{j,\ell}(n).$$

We want this to be strictly positive for some n. We know that there is m such that $p_{i,k}(m) > 0$, and some r such that $p_{j,\ell}(r) > 0$. However, what we need is an n that makes them *simultaneously* positive. We can indeed find such an n, based on the assumption that we have aperiodic chains and waffling something about number theory.

Now we want to show positive recurrence. We know that X, and hence Y is positive recurrent. By our previous theorem, there is a unique invariant distribution π for P. It is then easy to check that Z has invariant distribution

$$\nu = (\nu_{ij} : ij \in S^2)$$

given by

$$\nu_{i,j} = \pi_i \pi_j.$$

This works because X and Y are independent. So Z is also positive recurrent.

So Z is nice.

The next step is to couple the two chains together. The idea is to fix some state $s \in S$, and let T be the earliest time at which $X_n = Y_n = s$. Because of recurrence, we can always find such at T. After this time T, X and Y behave under the exact same distribution.

We define

$$T = \inf\{n : Z_n = (X_n, Y_n) = (s, s)\}.$$

We have

$$\begin{aligned}
p_{i,k}(n) &= \mathbb{P}_i(X_n = k)\\
&= \mathbb{P}_{ij}(X_n = k)\\
&= \mathbb{P}_{ij}(X_n = k, T \leq n) + \mathbb{P}_{ij}(X_n = k, T > n)
\end{aligned}$$

Note that if $T \leq n$, then at time T, $X_T = Y_T$. Thus the evolution of X and Y after time T is equal. So this is equal to

$$\begin{aligned}
&= \mathbb{P}_{ij}(Y_n = k, T \leq n) + \mathbb{P}_{ij}(X_n = k, T > n)\\
&\leq \mathbb{P}_{ij}(Y_n = k) + \mathbb{P}_{ij}(T > n)\\
&= p_{j,k}(n) + \mathbb{P}_{ij}(T > n).
\end{aligned}$$

Hence we know that

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

As $n \to \infty$, we know that $\mathbb{P}_{ij}(T > n) \to 0$ since Z is recurrent. So

$$|p_{i,k}(n) - p_{j,k}(n)| \to 0$$

With this result, we can prove what we want. First, by the invariance of π, we have

$$\pi = \pi P^n$$

for all n. So we can write

$$\pi_k = \sum_j \pi_j p_{j,k}(n).$$

Hence we have

$$|\pi_k - p_{i,k}(n)| = \left|\sum_j \pi_j (p_{j,k}(n) - p_{i,k}(n))\right| \leq \sum_j \pi_j |p_{j,k}(n) - p_{i,k}|.$$

We know that each individual $|p_{j,k}(n) - p_{i,k}(n)|$ tends to zero. So by bounded convergence, we know

$$\pi_k - p_{i,k}(n) \to 0.$$

So done. □

4 Time reversal

Theorem. Let X be positive recurrent, irreducible with invariant distribution π. Suppose that X_0 has distribution π. Then Y defined by

$$Y_k = X_{N-k}$$

is a Markov chain with transition matrix $\hat{P} = (\hat{p}_{i,j} : i, j \in S)$, where

$$\hat{p}_{i,j} = \left(\frac{\pi_j}{\pi_i}\right) p_{j,i}.$$

Also π is invariant for $\hat{P}$.

Proof. First we show that $\hat{p}$ is a stochastic matrix. We clearly have $\hat{p}_{i,j} \geq 0$. We also have that for each i, we have

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

using the fact that $\pi = \pi P$.

We now show π is invariant for $\hat{P}$: We have

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

since P is a stochastic matrix and $\sum_i p_{ji} = 1$.

Note that our formula for $\hat{p}_{i,j}$ gives

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

Now we have to show that Y is a Markov chain. We have

$$\begin{aligned}
\mathbb{P}(Y_0 = i_0, \cdots, Y_k = i_k) &= \mathbb{P}(X_{N-k} = i_k, X_{N-k+1} = i_{k-1}, \cdots, X_N = i_0) \\
&= \pi_{i_k} p_{i_k, i_{k-1}} p_{i_{k-1}, p_{k-2}} \cdots p_{i_1, i_0} \\
&= (\pi_{i_k} p_{i_k, i_{k-1}}) p_{i_{k-1}, p_{k-2}} \cdots p_{i_1, i_0} \\
&= \hat{p}_{i_{k-1} i_k} (\pi_{i_{k-1}} p_{i_{k-1}, p_{k-2}}) \cdots p_{i_1, i_0} \\
&= \cdots \\
&= \pi_{i_0} \hat{p}_{i_0, i} \hat{p}_{i_1, i_2} \cdots \hat{p}_{i_{k-1}, i_k}.
\end{aligned}$$

So Y is a Markov chain. □

Proposition. Let P be the transition matrix of an irreducible Markov chain X. Suppose (P, λ) is in detailed balance. Then λ is the *unique* invariant distribution and the chain is reversible (when X_0 has distribution λ).

Proof. It suffices to show that λ is invariant. Then it is automatically unique and the chain is by definition reversible. This is easy to check. We have

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

So λ is invariant. □