# Markov Chains Review Sheet

Michael Li

April 25, 2016

**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

# 1 Markov Chains

## 1.1 What is a Markov Chain?

Let's review the definition of a homogeneous Markov Chain (the ones learned in class).

**Definition** (Markov chain)**.** Let $X = (X_0, X_1, \cdots)$ be a sequence of random variables taking values in some (assumed to be countable) set $S$, the *state space*.

We say $X$ has the *Markov property* if for all $n \geq 0, i_0, \cdots, i_{n+1} \in S$, we have

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \cdots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n).$$

If $X$ has the Markov property, we call it a *Markov chain*. If the conditional probabilities above does not depend on $n$, we call it a *homogeneous* chain.

Now this property is quite limited. We can easily see how this can be extended to the following *Extended Markov Property*:

**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

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

For test purposes, these two properties can be used interchangeably.

## 1.2 Transition Probabilities

We have the following definition of the transition probability:

**Definition** ($n$-step transition probability)**.** The $n$-step transition probability from $i$ to $j$ is the $ij$th element of the $n$th Transition Matrix $P(n)$, defined as:

$$p_{i,j}(n) = \mathbb{P}(X_n = j \mid X_0 = i).$$

Now, we have our famous Chapman-Kolmogorov Equations:

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

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

Now it may look messy, but it is really written common sense: To get to $j$ in $m+n$ steps, you have to get somewhere ($k$) in $m$ steps, and take it from there. So we add up all the possible $k$ and this is it. If you can't remember the form on the left, remember the one on the right, which is much more intuitive.

But how do we compute $n$-step probabilities? Do we really have to multiply $n$ matrices together? No, of course not. We are lazy:

### 1.2.1 Computing $n$-step Probabilities

We list 3 common ways for the test:

Diagonal Let $S = \{1, 2\}$, with

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

We assume $0 < \alpha, \beta < 1$. We want to find the $n$-step transition probability.

We can achieve this via diagonalization. We can write $P$ as

$$P = U^{-1} \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} U,$$

where the $\kappa_i$ are eigenvalues of $P$, and $U$ is composed of the eigenvectors.

To find the eigenvalues (VM refresher), we calculate

$$\det(P - \lambda I) = (1 - \alpha - \lambda)(1 - \beta - \lambda) - \alpha\beta = 0.$$

To get the answer. Simple.

Reduce then Diagonal But what if the matrix is large and we don't want to do a lot of calculations? You may have help. From Markov Chain Ex1 Q7, we have:

A flea hops about at random on the vertices of a triangle (i.e., each hop is from the currently occupied vertex to one of the other two vertices each with probability $\frac{1}{2}$ 2 ). Find the probability that after $n$ hops the flea is back where it started.

Now we have a $3\times3$ matrix. But since it is *symmetric*, and we are only interested in its original state, we can reduce the state space to $\{1\}$ and $\{2, 3\}$. So the matrix becomes:

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

Difference Equation Now using the example in the first case, you can use conditional probabilities to easily show that:

$$p_{1,1}(n + 1) = p_{1,1}(n)(1 - \alpha) + p_{1,2}(n)\beta.$$

Use the fact that each row of the matrix add up to 1 to get:

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

# 2 Classification of Chains and States

## 2.1 Communicating Classes

**Definition** (Leading to and communicate)**.** Suppose we have two states $i, j \in S$. We write $i \to j$ ($i$ *leads to* $j$) if there is some $n \geq 0$ such that $p_{i,j}(n) > 0$, ie. it is possible for us to get from $i$ to $j$ (in multiple steps). Note that we allow $n = 0$. So we always have $i \to i$.

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

It is clear that $\leftrightarrow$ is an equivalence relation, so we form equivalence classes called *communicating classes*. You could see them as separate "bundles" of lines connecting a set of possible states. Formalizing this, each bundle is a closed set of states:

**Definition** (Closed)**.** A subset $C \subseteq S$ is *closed* if $p_{i,j} = 0$ for all $i \in C, j \notin C$.

But in this class, we really only cared about:

**Definition** (Irreducible chain)**.** A Markov chain is *irreducible* if there is a unique communication class.

## 2.2 Recurrence and Communicating Classes

Now we first introduce some notation:

**Notation.** For convenience, we will introduce some notations. We write

$$\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i),$$

and

$$\mathbb{E}_i(Z) = \mathbb{E}(Z \mid X_0 = i).$$

Moreover, we have:

$$F_{i,j}(S) = \sum_{n=0}^{\infty} f_{i,j}(n)s^n \qquad P_{i,i}(s) = \frac{1}{1 - F_{i,i}(s)}$$

The *first passage time* of $j \in S$ starting from $i$ is

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

Now we refresh ourselves of the definition of a recurrent state:

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

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

This basically means we will eventually get back to the state. Otherwise, we call the state *transient*.

Now here is an important theorem with an important proof, but we start from Abel's lemma and a useful result:

**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

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

The proof for Abel's lemma is in Analysis 2. Here is the useful result and its proof:

**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

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

□

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

*Proof.* Now we use $j = i$ in the useful result above, for $0 < s < 1$, we have

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

Now we check we are not dividing by 0:

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

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

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

So for $|s| < 1$, $F_{i,i}(s) < 1$. So we are cool. 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 $\sum_n p_{i,i}(n) = \infty$. □

Now here is a theorem that would let us check recurrent states much easier:

**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.

□

#### 2.2.1 Finite State Spaces

We have some nice results if our state space is finite:

**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) as the result from communicating classes show that if a chain is irreducible, then all states are either recurrent or transient. 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 $p_{i,j}(n) \to 0$. If every state is transient, then since the sum is finite, we know $\sum p_{i,j}(n) \to 0$ as $n \to \infty$. But we also know that:

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

And this is true for all $n$. But our sum is always decreasing (to 0). So we have a contradiction. □

### 2.2.2 Polya's Theorem

This is so important it deserves a special section. From Polya, we have this beautiful result:

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

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

Note that $j$ and $i$ are points in $d$-dimensional spaces. This is an irreducible chain, so the points are either all recurrent or all transient.

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

*Proof.*

$d = 1$ We want to show that $\sum p_{0,0}(n) = \infty$. Then we know the origin is recurrent. Since it is impossible to come back in odd number of steps, we consider $\sum p_{0,0}(2n)$, which is just combinatorics now, as we need $n$ steps forward and back:

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

We use the Stirling's approximation ($n! \simeq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$) to get:

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

So the sum of this diverges.

$d = 2$ This is similar to $d = 1$. We need $m$ steps right and left, and $n - m$ steps up and down, So we have:

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

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.

$d \geq 3$ We do this again:

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

And then we employ some bounding magic:

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

We are going to use the identity $\displaystyle\sum_{i+j+k=n}\frac{n!}{3^n i!j!k!} = 1$. This comes from the following: 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 gives 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 is 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. Then similarly we can use this to prove for $d > 3$.

□

## 2.3 Hitting Probabilities

We first introduce some definitions:

**Definition** (Hitting time)**.** The *hitting time* of $A \subseteq S$ is the random variable $H^A = \min\{n \geq 0 : X_n \in A\}$. In particular, if we start in $A$, then $H^A = 0$. We also have

$$h_i^A = \mathbb{P}_i(H^A < \infty) = \mathbb{P}_i(\text{ever reach } A).$$

And we define $k_i^A = E_i(H^A)$.

Now the following is the core theorem on hitting times:

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

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

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 $i$.

The formula is intuitive (second part follows from conditional probabilities), but proving minimality is slightly harder:

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

$$x_i^A = \begin{cases} 1 & i \in A \\ \sum_{j \in S} p_{i,j} x_j^A & i \notin A \end{cases},$$

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

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

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. □

In the exact same way (just copy and paste the proof above), we have:

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

$$k_i^A = \begin{cases} 0 & i \in A \\ 1 + \sum_j p_{i,k} k_j^A & i \not\in A. \end{cases}$$

This confirms the fact that in Probability 1A, we said the solution to expected times/probability was the minimal positive solution to the equation we had. Now we know why. We now give a Birth-death chain example, which is important:

**Example** (Birth-death chain)**.** Let $(p_i : i \geq 1)$ be an arbitrary sequence such that $p_i \in (0, 1)$. We let $q_i = 1 - p_i$. We let $\mathbb{N}$ be our state space and define the transition probabilities to be

$$p_{i,i+1} = p_i, \quad p_{i,i-1} = q_i.$$

This is really just the random walk, but every time you walk, the probability changes. We have the *absorbing state*

$$p_{0,0} = 1.$$

Let $h_i = h_i^{\{0\}}$. We know that

$$h_0 = 1, \quad p_i h_{i+1} - h_i + q_i h_{i-1} = 0, \quad i \geq 1.$$

This is no longer a difference equation, since the coefficients depends on the index $i$. We now employ magic:

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

We let $u_i = h_{i-1} - h_i$ and this becomes:

$$u_{i+1} = \frac{q_i}{p_i} u_i.$$

We can iterate this to become

$$u_{i+1} = \left(\frac{q_i}{p_i}\right)\left(\frac{q_{i-1}}{p_{i-1}}\right) \cdots \left(\frac{q_1}{p_1}\right) u_1.$$

Let

$$\gamma_i = \frac{q_1 q_2 \cdots q_i}{p_1 p_2 \cdots p_i}.$$

Then we get $u_{i+1} = \gamma_i u_1$. For convenience, we let $\gamma_0 = 1$. Now we want to retrieve our $h_i$. We can do this by summing the equation $u_i = h_{i-1} - h_i$:

$$h_0 - h_i = u_1 + u_2 + \cdots + u_i.$$

Using the fact that $h_0 = 1$, we get

$$h_i = 1 - u_1(\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}).$$

Here we have a parameter $u_1$, and we need to find out what this is. Our theorem tells us the value of $u_1$ minimizes $h_i$. This all depends on the value of

$$S = \sum_{i=0}^{\infty} \gamma_i.$$

If it diverges, then $u_1 = 0$. If it converges, then since $u_1$ minimizes $h_i$ (but it can't let it go negative), and $S_n = \sum_{i=0}^n \gamma_i$ is always increasing, $h_i$ is decreasing, so $u_1$ can take a maximum of $\frac{1}{S}$ to make $h_\infty$ 0. So we have

$$h_i = \frac{\sum_{k=i}^\infty \gamma_k}{\sum_{k=0}^\infty \gamma_k}.$$

## 2.4 Strong Markov Property and Applications

We state the Strong Markov Property. In probability, this usually means we restate the "Weak" one in random variables:

**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 $X_0, \cdots, X_T$.

But what is a stopping time?

**Definition** (Stopping time). Let $X$ be a Markov chain. A random variable $T$ (which is a function $\Omega \to \mathbb{N} \cup \{\infty\}$) is a *stopping time* for the chain $X = (X_n)$ if for $n \geq 0$, the event $\{T = n\}$ is given in terms of $X_0, \cdots, X_n$.

This can be seen as a *stopping* time because we can treat $T$ as the time for which a condition on the current Markov Chain $X_i$ is achieved. Then we leave when $T = n$, which can be expressed in $X_0, \cdots, X_n$. Now we list an important example, which is gambling:

**Example** (Gambler's ruin). Again, this is the Markov chain taking values on the non-negative integers, moving to the right with probability $p$ and left with probability $q = 1 - p$. 0 is an absorbing state, since we have no money left to bet if we are broke. We want to find the time it takes to get to 0, ie.

$$H = \inf\{n \geq 0 : X_n = 0\}.$$

For this type of questions, we usually use the generating function:

$$G_i(s) = \mathbb{E}_i(s^H) = \sum_{n=0}^\infty s^n \mathbb{P}_i(H = n), \quad |s| < 1.$$

We have

$$G_1(s) = \mathbb{E}_1(s^H) = p\mathbb{E}_1(s^H \mid X_1 = 2) + q\mathbb{E}_1(s^H \mid X_1 = 0).$$

The second term is easy, since if $X_1 = 0$, then we must have $H = 1$. So $\mathbb{E}_1(s^H \mid X_1 = 0) = s$.

For the first one, we can realize that the expected time to reach 2 from 0 is exactly the sum of the expected time to reach 1 from 2 and the time to reach 0 from 1. Now they are independent by Strong Markov Property, so in terms of generating functions (which makes sums into products):

$$G_1 = psG_1^2 + qs. \tag{$*$}$$

Solving this, we get two solutions

$$G_1(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps}.$$

This is for each value of $s$. We know that if we suddenly change the sign, then $G_1(s)$ will be discontinuous at that point, but $G_1$, being a power series, has to be continuous. So the solution must be either $+$ for all $s$, or $-$ for all $s$.

We look at $s = 0$. We see that the numerator becomes $1 \pm 1$, while the denominator is 0. As $G$ converges at $s = 0$, we must pick $-$, so

$$G_1(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}.$$

We can find $\mathbb{P}_1(H = k)$ by expanding the Taylor series.

What is the probability of ever hitting 0? This is

$$\mathbb{P}_1(H < \infty) = \lim_{s \to 1} G_1(s) = \frac{1 - \sqrt{1 - 4pq}}{2p}. = \frac{1 - |p - q|}{2p} = \begin{cases} 1 & p \leq q \\ \frac{q}{p} & p > q \end{cases}.$$

We can also find $\mu = \mathbb{E}_1(H)$. Technically, we can use the derivative of $G$ and take $s = 1$, but that is too hard. A easier way is to see that for $p \geq \frac{1}{2}$, it is clearly $\infty$, as we can escape. But for $p < \frac{1}{2}$, using the Strong Markov property:

$$\mu = 1 + p(2\mu)$$

This is because with probability $q$ you hit 0, but with probability $p$ you hit 2, which now has *twice* the expected time as 1 to get to 0. So this gives:

$$\mu = \frac{1}{q - p}$$

## 2.5 Periodicity, Ergodicity, and Positivity

Here are some more classification that makes our lives easier:

**Definition** (Mean recurrence time)**.** Let $T_i$ be the returning time to a state $i$. Then the *mean recurrence time* of $i$ is

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

**Definition** (Null and positive state)**.** If $i$ is recurrent, we call $i$ a *null state* if $\mu_i = \infty$. Otherwise $i$ is *non-null* or *positive*.

**Definition** (Period)**.** The *period* of a state $i$ is $d_i = \gcd\{n \geq 1 : p_{i,i}(n) > 0\}$.

A state is *aperiodic* if $d_i = 1$.

**Definition** (Ergodic state)**.** A state $i$ is *ergodic* if it is aperiodic and positive recurrent.

Now aperiodic, ergodic, positive recurrent states are the states we love. Why? Read the invariant distribution section. We can first prove a result for communicating states:

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

(i) $d_i = d_j$.

(ii) $i$ is recurrent iff $j$ is recurrent.

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

□

# 3 Long-run Behavior

## 3.1 Invariant Distribution

An invariant distribution is one that just won't change under a Markov chain and is a one we would *like* a Markov chain to tend to:

**Definition** (Invariant distriubtion)**.** Let $X_j$ be a Markov chain with transition probabilities $P$. The distribution $\pi = (\pi_k : k \in S)$ is an *invariant distribution* if

(i) $\pi_k \geq 0$, $\sum_k \pi_k = 1$.

(ii) $\pi = \pi P$.

The first condition just ensures that this is a genuine distribution.

Now we prove something big! Remember when we said the classification would make our lives nice?

**Note.** This theorem , although not non-examinable, has not appeared in tests yet. It is not supposed to, as it is VERY long. But it includes some helpful methods for thinking:

- Contradict Transience by finding appropriate constant sum and showing that if all were transient, this sum goes to 0.

- Don't know how to do things under a countably infinite state space $S$? Take a finite one $F$ and approximate $S$.

Before we introduce the main theorem, let's prove this:

**Notation.** Let $W_i$ denote the number of visits to $i$ before the next visit to $k$. Formally, we have

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

where $T_k$ is the recurrence time of $m$ and $1$ is the indicator function. In particular, $W_i = 1$ for $i = k$ (if $T_k$ is finite). We can also write this as

$$W_i = \sum_{m=1}^{T_k} 1(X_m = i).$$

This is a random variable. So we can look at its expectation. We define

$$\rho_i = \mathbb{E}_k(W_i).$$

**Proposition.** For an irreducible recurrent chain and $k \in S$, $\rho = (\rho_i : i \in S)$ defined as above by $\rho_i = \mathbb{E}_k(W_i)$, 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 $\rho_i$, 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

$$\sum_i \rho_i = \sum_i \mathbb{E}_k(W_i) = \mathbb{E}_k\left(\sum_i W_i\right) = \mathbb{E}_k(T_k) = \mu_k.$$

We swapped sums and expectation, but that's an Analysis problem. Not ours.

(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 $k$. So it is useless:

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

Let $r = m - 1$, and get

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

We look at the 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. And all other terms are 0 by definition. So the sum is 1, which is $\rho_k$.

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 $m, n$,

$$\begin{aligned}
\rho_i &\geq \rho_k p_{k,i}(n) \\
\rho_k &\geq \rho_i p_{i,k}(m)
\end{aligned}$$

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.

□

Now we can prove our big theorem.

**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 $\pi$, then it is unique, every state is positive recurrent, and

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

for $i \in S$, where $\mu_i$ is the mean recurrence time of $i$.

*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 $\pi$ 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 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 $\pi_i$ 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 $\mu_i$ 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 $\pi_i$ be the probability of starting at $i$. So suppose $X_0$ has distribution $\pi$. 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? $\pi_i$ 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

Now when does convergence actually occur? It occurs precisely under the following conditions:

**Theorem** (Proof Non-examinable)**.** Consider a Markov chain that is irreducible, positive recurrent and aperiodic. Then

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

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

We can see the invariant distribution as the amount of time the Markov chain spends in each state in the long run. We can't really state the question formally (what is an average of a random variable anyways?) so we would not prove it here.

## 4 Time Reversal

Are we going to time travel?! No, don't get your hopes high. This is just a section to show that (some) Markov chains still behave nicely if we go back in time. Just some.

**Theorem.** Let $X$ be positive recurrent, irreducible with invariant distribution $\pi$. Suppose that $X_0$ has distribution $\pi$. 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 $\pi$ is invariant for $\hat{P}$.

Now it should be clear that $\pi$ is invariant for $\hat{P}$ from the definition, and the transition matrix is essentially the transpose. But why should $Y$ be Markov?

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

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

Therefore:

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

As we continue to switch $\pi$ around. So $Y$ is a Markov chain. □

But this does not mean that the chain is reversible. A reversible chain needs the same dynamics going there and back, so we need:

**Definition** (Reversible chain)**.** An irreducible Markov chain $X = (X_0, \cdots, X_N)$ in its invariant distribution $\pi$ is *reversible* if its reversal has the same transition probabilities as does $X$, aka we need the *detailed balance equation*:

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

for all $i, j \in S$. If a distribution $\lambda$ other than $\pi$ satisfy this equation above, we say $(P, \lambda)$ is in detailed balance.

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

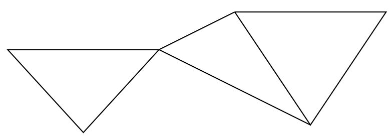
*Proof.* We really just need to show this is invariant. This is easy:

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

So $\lambda$ is invariant, so it is *unique* and the chain is by definition reversible. □

Now we end this review sheet with an interesting example: A random walk on a finite graph.

**Example** (Random walk on a finite graph)**.** A graph is a collection of points with edges between them. For example, the following is a graph:



More precisely, a graph is a pair $G = (V, E)$, where $E$ contains distinct unordered pairs of distinct vertices $(u, v)$, drawn as edges from $u$ to $v$.

Note that the restriction of distinct pairs and distinct vertices are there to prevent loops and parallel edges, and the fact that the pairs are unordered means our edges don't have orientations.

A graph $G$ is connected if for all $u, v \in V$, there exists a path along the edges from $u$ to $v$.

Let $G = (V, E)$ be a connected graph with $|V| \leq \infty$. Let $X = (X_n)$ be a random walk on $G$. Here we live on the vertices, and on each step, we move to one an adjacent vertex. More precisely, if $X_n = x$, then $X_{n+1}$ is chosen uniformly at random from the set of neighbours of $x$, ie. the set $\{y \in V : (x, y) \in E\}$, independently of the past. This is a Markov chain. Our transition probabilities are

$$p_{i,j} = \begin{cases} 0 & j \text{ is a neighbour of } i \\ \frac{1}{d_i} & j \text{ is a neighbour of } i \end{cases},$$

where $d_i$ is the number of neighbours of $i$, commonly known as the *degree* of $i$.

By connectivity, the Markov chain is irreducible. Since it is finite, it is recurrent, and in fact positive recurrent. Now we would like to guess its invariant distribution by solving:

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

This is balanced if $i$ is not a neighbour of $j$. Otherwise:

$$\lambda_i \frac{1}{d_i} = \lambda_j \frac{1}{d_j}.$$

Hmmm, this is easy, take $\lambda_i = d_i$! Now we normalize it. $\sum d_i$ is really just 2 times the sum of number of edges (every edge counted twice), so we have:

$$\lambda_i = \frac{d_i}{2|E|}.$$