Random Walks: WEEK 2

1 Recurrence and transience

Consider the event "$\{X_n = i \text{ for some } n > 0\}$" by which we mean "$\{X_1 = i\}or\{X_2 = i, X_1 \neq i\}or\{X_3 = i, X_2 \neq i, X_1 \neq i\}, \cdots$".

Definition 1.1. A state $i \in S$ is recurrent if $\mathbb{P}(X_n = i \text{ for some } n > 0 \mid X_0 = i) = 1$

Definition 1.2. A state $i \in S$ is transient if $\mathbb{P}(X_n = i \text{ for some } n > 0 \mid X_0 = i) < 1$

Note that in the example in Lecture 1, the state "Home" is recurrent (and even absorbing), but all other states are transient.

Definition 1.3. Let $f_{ii}(n)$ be the probability that the first return time to i is n when we start at i:

$$\begin{aligned} f_{ii}(n) &= \mathbb{P}(X_1 \neq i, \ldots, X_{n-1} \neq i, X_n = i | X_0 = i) \\ f_{ii}(0) &= 0, \text{ by convention} \end{aligned}$$

Then we define the probability of eventual return as:

$$f_{ii} = \sum_{n=1}^{+\infty} f_{ii}(n)$$

Theorem 1.4. A rule to determine recurrent and transient states is:

(a) state $i \in S$ is recurrent iff $\sum_{n \geq 1} p_{ii}^{(n)} = +\infty$

(b) state $i \in S$ is transient iff $\sum_{n \geq 1} p_{ii}^{(n)} < +\infty$

Proof: We use generating functions. We define

$$P_{ii}(s) = \sum_{n=0}^{+\infty} s^n p_{ii}(n), \ |s| < 1$$

and

$$F_{ii}(s) = \sum_{n=0}^{+\infty} s^n f_{ii}(n), \ |s| < 1$$

One can show that, for $|s| < 1$:

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

This equation is enough to prove the theorem since, as $s \nearrow 1$ we have

$$P_{ii}(s) \nearrow +\infty \iff f_{ii} = F_{ii}(s = 1) = 1$$

By Abel's theorem,

$$\sum_{n=0}^{+\infty} p_{ii}(n) = \lim_{s \nearrow 1} P_{ii}(s)$$

Thus

$$\sum_{n=0}^{+\infty} p_{ii}(n) = +\infty \iff f_{ii} = +1$$

Theorem 1.5 (Abel's Theorem)**.** If $a_i \geq 0$ for all i and $G(s) = \sum_{n\geq 0} s^n a_n$ is finite for $|s| < 1$, then

$$\lim_{s\nearrow 1} G(s) = \begin{cases} +\infty & \text{if} \sum_{n\geq 0} a_n = +\infty \\ \sum_{n\geq 0} a_n & \text{if} \sum_{n\geq 0} a_n < +\infty \end{cases}$$

Proof of formula $P_{ii}(s) = \frac{1}{1-F_{ii}(s)}$:
Let the events $A_m = \{X_m = i\}$, $B_r = \{X_1 \neq i, \ldots, X_{r-1} \neq i, X_r = i\}$ $\forall 1 \leq r \leq m$. The B_r are disjoint, thus:

$$\begin{aligned} \mathbb{P}(A_m \mid X_0 = i) &= \sum_{r=1}^{m} \mathbb{P}(A_m \cap B_r \mid X_0 = i) \\ &= \sum_{r=1}^{m} \mathbb{P}(A_m \mid B_r, X_0 = i)\mathbb{P}(B_r \mid X_0 = i) \\ &= \sum_{r=1}^{m} \mathbb{P}(A_m \mid X_r = i)\mathbb{P}(B_r \mid X_0 = i) \\ \implies p_{ii}(m) &= \sum_{r=1}^{m} f_{ii}(r) p_{ii}(m-r) \end{aligned}$$

Multiplying the last equation by s^m with $|s| < 1$ and sum over $m \geq 1$ we get:

$$\begin{aligned} \implies \sum_{m\geq 1} s^m p_{ii}(m) &= \sum_{m\geq 1} s^m \sum_{r=1}^{m} f_{ii}(r) p_{ii}(m-r) \\ &= \sum_{m\geq 1} \sum_{k+l=m} f_{ii}(k) s^k p_{ii}(l) s^l \\ &= \sum_{k} f_{ii}(k) s^k \sum_{l} p_{ii}(l) s^l \\ \implies P_{ii}(s) - 1 &= P_{ii}(s) F_{ii}(s) \\ \implies P_{ii}(s) &= \frac{1}{1 - F_{ii}(s)} \end{aligned}$$

2 Positive/null-recurrence

Let $T_i = \inf\{n > 0 : X_n = i\}$ be the first return time of a random walk starting in $X_0 = i$.

Note that

$$\begin{aligned} \mathbb{P}(T_i < +\infty | X_0 = i) &= \sum_{m\geq 1} \mathbb{P}(T_i = m | X_0 = i) = \sum_{n\geq 1} \mathbb{P}(X_n = i, X_{n-1} \neq i, \ldots, X_1 \neq i | X_0 = i) \\ &= \mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) = f_{ii} \end{aligned}$$

Recall also the notation (from Lecture 2) for the probability that *the first return time to* i *is* n

$$f_{ii}(n) = \mathbb{P}(X_n = i, X_{n-1} \neq i, \ldots, X_1 \neq i | X_0 = i) = \mathbb{P}(T_i = n | X_0 = i)$$

For a recurrent state, $1 = f_{ii} = \sum_{n\geq 1} f_{ii}(n) = \mathbb{P}(T_i < +\infty | X_0 = i)$ and for a transient state, $f_{ii} < 1$. So for a transient state, the random variable T_i must have some mass at $T_i = +\infty$: $\mathbb{P}(T_i = +\infty | X_0 = i) > 0$. We define the following:

Definition 2.1. The **mean recurrence time** is given by

$$\mu_i \equiv \mathbb{E}(T_i | X_0 = i) = \sum_{n\geq 1} n\, \mathbb{P}(T_i = n | X_0 = i) = \sum_{n\geq 1} n\, f_{ii}(n).$$

Can we compute this quantity? Let us to that end distinguish two cases:

Transient states: We know that for transient states the event $T_i = +\infty$ occurs with some strictly positive probability $\mathbb{P}(T_i = +\infty | X_0 = i) > 0$. We therefore have that $\mu_i = \mathbb{E}(T_i | X_0 = i) = +\infty$.

Recurrent states: By definition $\sum_{m \geq 1} f_{ii}(m) = 1$ for recurrent states. But $\sum_{n \geq 1} n f_{ii}(n)$ is not necessarily finite. So even though you know for sure the chain will return to state i at some finite time, the average recurrence time may be infinite. Based on this, we say that any recurrent state is either one of the following:

- **Positive-recurrent** if $\mu_i < +\infty$.
- **Null-recurrent** if $\mu_i = +\infty$.

3 Irreducible chains

- We say that $i \to j$ if $\exists n \geq 0$ such that $p_{i \to j}(n) > 0$. In words, we say that i *communicates* with j or that j is *accessible* from i.
- We say that $i \longleftrightarrow j$ if $i \to j$ and $j \to i$. In words, i and j *intercommunicate*. Note, however, that this implies that $\exists n \geq 0$ s.t. $p_{i \to j}(n) > 0$ and $\exists m \geq 0$ s.t. $p_{j \to i}(m) > 0$ (but it is not necessary that $n = m$ holds).
- If $i \longleftrightarrow j$ then these states are reflexive, symmetric and transitive, meaning:
  - Reflexivity: $i \longleftrightarrow i$, because we always have that at least $p_{ii}(0) = 1$ holds.
  - Symmetry: If $i \longleftrightarrow j$ holds then -of course- also $j \longleftrightarrow i$ holds.
  - Transitivity: If $i \longleftrightarrow k$ and $k \longleftrightarrow j$, then $i \longleftrightarrow j$. This statement follows from the Chapman-Kolmogorov equation:

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

  One can always pick an n and m such that the right hand side is strictly positive and hence $\exists n + m$ such that $p_{i \to j}(n+m) > 0$. The same argument shows that there exists n', m' such that $p_{j \to i}(n' + m') > 0$. Thus i and j intercommunicate.
- We can partition the state space $\mathcal{S}$ into *equivalence classes*. For example, the music festival example of Lecture 1 had two such classes: $\{\text{Home}\}$ and $\{\text{Bar}, \text{Dancing}, \text{Concert}\}$. Note that $\{\text{Home}\}$ is here trivially a positive-recurrent state (in fact it is absorbing: once you are at home you return to home at every time step with probability one) and that $\{\text{Bar}, \text{Dancing}, \text{Concert}\}$ are transient states (fortunately or unfortunately ?).

Having stated these notions, we are ready to define irreducible chains as follows:

Definition 3.1. A Markov chain is said to be **irreducible** if it has only one equivalence class, i.e., $\forall i, j \in \mathcal{S}\ \exists n, m$ such that $p_{i \to j}(n)\, p_{j \to i}(m) > 0$.

In other words, in an irreducible Markov chain every state is accessible from every state.

Proposition 3.2. Within an equivalence class of a Markov chain or for an irreducible Markov chain it holds that

1. All states i have the same period.
2. All states i are recurrent or all states are transient.

3. If all states i are recurrent, then either they are all null-recurrent or they are all positive-recurrent.

Proof of point 2. Take two states in the same equivalence class, $i \longleftrightarrow j$. Then, from the Chapman-Kolmogorov equation, we deduce the inequality

$$p_{ii}(n+t+m) \geq \underbrace{p_{i\to j}(n)}_{>0}\, p_{j\to j}(t)\, \underbrace{p_{j\to i}(m)}_{>0} \geq \underbrace{\alpha}_{>0}\, p_{jj}(t).$$

If i is transient, then $\sum_t p_{ii}(t) < +\infty$ (criterion proved in Lecture 2) and thus $\sum_t p_{jj}(n+t+m) < +\infty$ so j is also transient. To complete the proof, we note that the roles of i and j can be interchanged. This way we also get that "if j is transient, then i is transient".

The proof of 1 is similar. The proof of 3 requires more tools that we don't quite have at this point. □

Lemma 3.3. If a Markov chain has a finite state space $\mathcal{S}$ and is irreducible, then all its states are (positive-)recurrent.

Proof. We have the following property:

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

but our state space is finite so we can interchange the order:

$$\sum_{j\in\mathcal{S}} \lim_{n\to+\infty} p_{i\to j}(n) = 1. \tag{1}$$

We continue by contradiction. Assume that all $j \in \mathcal{S}$ are transient, then $p_{jj}(n) \to 0$ as $n \to \infty$. Even stronger, we proved in Homework 2 that $p_{ij}(n) \to 0$ as well. This contradicts (1):

$$\sum_{j\in\mathcal{S}} \underbrace{\lim_{n\to+\infty} p_{i\to j}(n)}_{\to 0} \neq 1.$$

So if there is a j that is recurrent and the chain is irreducible, then all $j \in \mathcal{S}$ must be recurrent.

The proof that all states are in fact positive-recurrent requires more tools that we don't yet have. □

4 Stationary distribution

Definition 4.1. A distribution π^* is called **stationary** if it satisfies the equation $\pi^* = \pi^* P$.

It follows immediately that any stationary distribution also satisfies $\pi^* = \pi^* P^k$ for any $k \geq 0$. In particular, if we initialize a chain in the stationary distribution $\pi^{(0)} = \pi^*$ then at any time n, $\pi^{(n)} = \pi^*$ (and this is why π^* is called stationary).

Discussion: For systems with a *finite state space*, one can show that the finite stochastic matrix P always has a unit eigenvalue and associated left eigenvector with non-negative components $\pi_i \geq 0$[1]. But it may not be unique (as will be clear from the discussion below). Uniqueness requires more conditions on P. For example if P has all its elements strictly positive or if there exists $N \geq 1$ such that P^N has all its elements strictly positive then the standard forms of the Perron-Frobenius theorem imply unicity. However this is not the most general condition. The theory of Markov chains in finite state spaces can be developed through the Perron-Frobenius theorems (at various levels of sophistication) but this is not the route we take in this class because we are also interested in *infinite* (countable) state spaces.

An important theorem is the following.

[1]It is clear that the vector with all ones is a right eigenvector with eigenvalue 1. But the existence of a left eigenvector with non-negative components for the same eigenvalue is not obvious.

Theorem 4.2 (existence and uniqueness of a stationary distribution)**.** Consider an irreducible Markov chain. It has a stationary distribution if and only if the chain is positive-recurrent. Moreover, this distribution is unique and takes on the value

$$\pi_i^* = \frac{1}{\mu_i} = \frac{1}{\mathbb{E}(T_i|X_0 = i)}$$

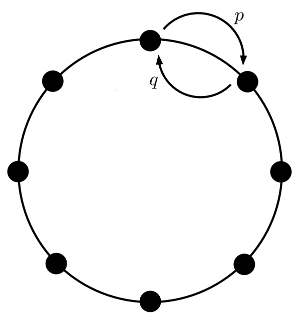
Remark 4.3. Note that μ_i is finite because we assume the chain is positive-recurrent.

Remark 4.4. Take an irreducible Markov chain with a *finite state space*. Then by Lemma 3.3, we know it must be positive-recurrent. Thus an irreducible Markov chain on a finite state space has a unique stationary distribution $\pi_i^* = 1/\mu_i$.

Remark 4.5. This theorem is very powerful. Indeed suppose you have a chain and you know that it is irreducible. With an infinite state space, it might be difficult to prove directly that it is positive-recurrent, but it might be easier to compute the stationary distribution. Then you immediately can conclude that it is necessarily positive-recurrent.

The proof of the theorem is not easy and we do not go through it here. We rather try to motivate the theorem through the following discussion.

Example 4.6. Consider the following random walk on a circle, a finite state space:



The stochastic matrix P of the walk is of size $|\mathcal{S}| \times |\mathcal{S}|$ and looks as follows:

$$\begin{pmatrix} 0 & p & 0 & \cdots & \cdots & q \\ q & 0 & p & & & \vdots \\ 0 & q & 0 & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & 0 & p \\ p & \cdots & & \cdots & q & 0 \end{pmatrix}$$

One can easily verify that $\pi_i^* = \frac{1}{|S|}$ is the stationary distribution. This example suggests that on $\mathbb{Z}^d$ the random walk has no stationary distribution because $|S| \to \infty$ and thus $\pi_i^* = \frac{1}{|S|} \to 0$ would not yield a valid probability distribution. This is true. Indeed the random walk in $\mathbb{Z}^d$ is irreducible and *null recurrent* for $d = 1, 2$ and *transient* for $d \geq 3$, so by the above theorem, it cannot have a stationary distribution.

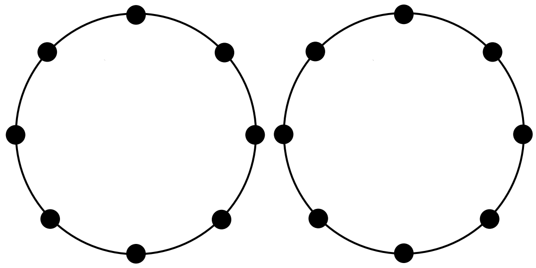
When one verifies that $\pi_i^* = \frac{1}{|S|}$ is the stationary distribution in the example above, one sees that this works because the *columns* sum to one (recall that for a stochastic matrix the *rows* always sum to one). This motivates the following definition.

Definition 4.7. A doubly stochastic matrix is a matrix $P = [p_{ij}]$ with $p_{ij} \geq 0$ of which all the rows and all the columns sum to one.

One can easily see that the mechanism of the example above generalizes to any chain on a finite state space with a doubly stochastic matrix: in this case, $\pi_i^* = \frac{1}{|S|}$ is a stationary distribution because the *columns* sum to one.

Now what about the unicity of the stationary distribution ? The following simple situation suggests that we don't have unicity for reducible chains.

Example 4.8. Consider two separate circles:



The state space is the union of two disconnected finite state spaces, $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. This Markov chain is *not* irreducible! Its transition matrix shows the following block structure:

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \tag{2}$$

Consequently, $\pi^* = \pi^* P$ has many solutions. An example are all distributions that are computed as follows:

$$\pi^* = \left(\frac{\alpha}{|\mathcal{S}_1|}, \cdots, \frac{\alpha}{|\mathcal{S}_1|}, \frac{\beta}{|\mathcal{S}_2|}, \cdots, \frac{\beta}{|\mathcal{S}_2|} \right), \tag{3}$$

where $\alpha + \beta = 1$ and $\alpha, \beta \geq 0$. The first $|\mathcal{S}_1|$ components correspond to the first circle, the last $|\mathcal{S}_2|$ correspond to the second. The uniform stationary distribution corresponds to $\alpha = \frac{|\mathcal{S}_1|}{|\mathcal{S}_1|+|\mathcal{S}_2|}$ and $\beta = \frac{|\mathcal{S}_2|}{|\mathcal{S}_1|+|\mathcal{S}_2|}$. Also note that extreme cases, such as $\{\alpha = 0, \beta = 1\}$ and $\{\alpha = 1, \beta = 0\}$ are also perfectly valid stationary distributions. The general stationary distributions are convex combinations of the two extremal ones.