Markov Chains and Algorithmic Applications: WEEK 5

1 Preliminary: reversible chains

The ergodic theorem provides us with a nice convergence result, that is $\lim_{n\to\infty} p_{ij}(n) = \pi_j$ for any $i, j \in S$. But for the purpose of any practical application, we would like to know more about the rate at which this convergence occurs. We will start by talking about reversible chains and detailed balance.

Definition 1.1. An ergodic Markov chain $(X_n, n \geq 0)$ is said to be reversible if its stationary distribution π^* satisfies the following detailed balance equation:

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j \in S$$

Remarks.

- We can still talk about reversibility if the chain is only irreducible and positive-recurrent.
- If one assumes that the chain is in stationary distribution from the start, then the backwards chain $X_n, X_{n-1}, \ldots$ has the same transition probabilities as the original chain, hence the name "reversible".
- If π satisfies the detailed balance equation, then $\pi = \pi P$.
- The reciprocal statement is wrong, as we will see in some counter-examples.
- Note that in general, the detailed balance equation is easier to solve than the equation $\pi = \pi P$, but there are unfortunately no simple conditions that ensure that the detailed balance equation is satisfied.

Example 1.2 (Ehrenfest urns)**.** Consider 2 urns with N numbered balls. At each step, we pick uniformly at random a number between 1 and N, take the ball with this number and put it in the other urn. The state is the number of balls in the right urn. The transition probabilities are the following:

$$\begin{aligned} p_{i,i+1} &= \frac{N-i}{N} \\ p_{i,i-1} &= \frac{i}{N} \end{aligned}$$

Solving the detailed balance equation, we get:

$$\begin{aligned} & \pi_{i+1} = \frac{p_{i,i+1}}{p_{i+1,i}} \pi_i = \frac{N-i}{i+1} \pi_i \\ \Rightarrow \quad & \pi_{i+1} = \frac{(N-i)(N-i+1)\ldots N}{(i+1)i(i-1)\ldots 2} \pi_0 = \frac{N!}{(N-i-1)!(i+1)!} \pi_0 \\ \Rightarrow \quad & \pi_{i+1} = \binom{N}{i+1} \pi_0 \end{aligned}$$

which leads to the conclusion that $\pi_0 = \frac{1}{2^N}$. This chain is therefore reversible.

Example 1.3. All irreducible birth-death chains satisfy the detailed balance equation.

Example 1.4. If for a given $i, j \in S$, we have $p_{ij} > 0$ and $p_{ji} = 0$, then the chain is *not* reversible.

Example 1.5 (Random walk on the cycle)**.** We know that the stationary distribution for the cyclic random walk with transition probabilities p and q $(p + q = 1)$ is simply the uniform distribution, $\pi_i = \frac{1}{N}$. To be verified, the detailed balance equation requires $\pi_i p = \pi_{i+1} q$ i.e. $p = q = \frac{1}{2}$. In all other cases, the detailed balance equation is not satisfied. In a general manner, as soon as there is a cycle with such an asymmetry in a chain, the chain is not reversible.

2 Rate of convergence, spectral gap and mixing times

2.1 Setup, motivation and assumptions

Let $(X_n, n \geq 0)$ be a time-homogeneous, ergodic Markov chain on a state space $\mathcal{S}$ and let P be its transition matrix. Therefore, there exists a limiting and stationary distribution which we call π. In other words, $p_{ij}(n) \xrightarrow[n\to\infty]{} \pi_j, \forall i, j \in \mathcal{S}$ (limiting distribution) and $\pi = \pi P$ (stationary distribution).

The question is now: *For what values of n is $p_{ij}(n)$ "really close" to π_j? In other words, "how fast" does $p_{i,\cdot}(n)$ converge to its limiting distribution π?* The answer to this question is useful for practical applications (see for example Section 6.14 of Grimmett & Stirzaker) where it is not enough to only know what happens when $n \to \infty$; in some cases, we also need to have a notion of how soon the behavior of a Markov chain becomes similar to the behavior at infinity.

We make the following simplifying assumptions:

1. $\mathcal{S}$ is finite ($|\mathcal{S}| = N$).
2. The detailed balance equation is satisfied:

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j \in \mathcal{S}. \tag{1}$$

2.2 Total variation norm and convergence of distribution

Here, we consider *convergence of distribution*. We want to see when $p_{ij}(n) = \mathbb{P}(X_n = j|X_0 = i)$ and $\pi_j = \mathbb{P}(X_* = j)$ "get close to each other". To clarify what this means, we use the *total variation distance*, already introduced in Lecture 4.

Definition 2.1. Total variation distance. Let $\mu = (\mu_i, i \in \mathcal{S}),\ \nu = (\nu_i, i \in \mathcal{S})$ such that $\mu_i \geq 0,\ \nu_i \geq 0,\ \sum_{i\in\mathcal{S}} \mu_i = 1,\ \sum_{i\in\mathcal{S}} \nu_i = 1$. We define the total variation distance between μ and ν as $\frac{1}{2}\sum_{i\in\mathcal{S}} |\mu_i - \nu_i|$ and we denote it by $\|\mu - \nu\|_{\mathrm{TV}}$.

Note: It is easy to check that $\|\mu - \nu\|_{\mathrm{TV}} \in [0, 1]$.

In what follows, we find an upper-bound on the total variation distance $\|P_i^n - \pi\|_{\mathrm{TV}} = \frac{1}{2}\sum_{j\in\mathcal{S}} |p_{ij}(n) - \pi_j|$. By studying how fast this upper-bound goes to 0, we will find out how fast $\|P_i^n - \pi\|_{\mathrm{TV}}$ goes to 0.

2.3 Eigenvalues and eigenvectors of P

Define a new matrix Q as follows:

$$q_{ij} = \sqrt{\pi_i}\, p_{ij}\, \frac{1}{\sqrt{\pi_j}}, \qquad \forall i, j \in \mathcal{S}$$

Two observations: 1. $q_{ii} = p_{ii},, \forall i \in \mathcal{S}$, 2. $q_{ij} \geq 0$, but $\sum_{j\in\mathcal{S}} q_{ij} \neq 1$ in general.

Proposition 2.2. *Q is symmetric.*

Proof.

$$q_{ji} = \sqrt{\pi_j} p_{ji} \frac{1}{\sqrt{\pi_i}} = \frac{1}{\sqrt{\pi_i \pi_j}} \pi_j p_{ji} \overset{(*)}{=} \frac{1}{\sqrt{\pi_i \pi_j}} \pi_i p_{ij} = q_{ij}.$$

where $(*)$ follows from the detailed balance equation. $\square$

Since Q is symmetric, we can use the spectral theorem to conclude the following:

Proposition 2.3. There exist real numbers $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1}$ (the eigenvalues of Q) and vectors $u^{(0)}, \ldots, u^{(N-1)} \in \mathbb{R}^N$ (the eigenvectors of Q) such that $Qu^{(k)} = \lambda_k u^{(k)}, \forall k = 0, ..., N-1$. Moreover, $u^{(0)}, \ldots, u^{(N-1)}$ forms an orthonormal basis of $\mathbb{R}^N$ (equipped with the standard scalar product).

Proposition 2.4. Define the vector $\phi^{(k)} = \left(\dfrac{u_j^{(k)}}{\sqrt{\pi_j}}, j \in \mathcal{S}\right)$. Then,

$$P\phi^{(k)} = \lambda_k \phi^{(k)}.$$

Proof. By Proposition 2.3, for every $k = 0, ..., N-1$

$$\begin{aligned} &Qu^{(k)} = \lambda_k u^{(k)} && \Leftrightarrow \sum_{j \in \mathcal{S}} q_{ij} u_j^{(k)} = \lambda_k u_i^{(k)} \\ \Leftrightarrow &\sum_{j \in \mathcal{S}} \sqrt{\pi_i} p_{ij} \frac{1}{\sqrt{\pi_j}} u_j^{(k)} = \lambda_k u_i^{(k)} && \Leftrightarrow \sum_{j \in \mathcal{S}} p_{ij} \underbrace{\left(\frac{u_j^{(k)}}{\sqrt{\pi_j}}\right)}_{=\phi_j^{(k)}} = \lambda_k \underbrace{\left(\frac{u_i^{(k)}}{\sqrt{\pi_i}}\right)}_{=\phi_i^{(k)}} && \Leftrightarrow P\phi^{(k)} = \lambda_k \phi^{(k)} \end{aligned}$$

□

Proposition 2.4 says that the eigenvalues of P are $\lambda_0, \lambda_1, ..., \lambda_{N-1}$ (the same as those of Q) and the eigenvectors of P are $\phi^{(0)}, ..., \phi^{(N-1)}$. Note that $\phi^{(0)}, ..., \phi^{(N-1)}$ is not in general an orthonormal basis of $\mathbb{R}^N$ (equipped with the standard scalar product).

2.4 Main results

Facts about the λ's and ϕ's (proof: next time):

1. $\lambda_0 = 1, \quad \phi^{(0)} = [1, ..., 1]^T$.
2. $|\lambda_k| \leq 1, \quad \forall k = 1, ..., N-1$.
3. $\lambda_1 < 1$ and $\lambda_{N-1} > -1$.

Definition 2.5. $\lambda_* = \max\limits_{1 \leq k \leq N-1} |\lambda_k| = \max\{\lambda_1, -\lambda_{N-1}\} (< 1)$

Theorem 2.6. Rate of convergence. Under all the assumptions made, it holds that:

$$\|P_i^n - \pi\|_{\mathrm{TV}} \leq \frac{\lambda_*^n}{2\sqrt{\pi_i}} \quad \forall i \in \mathcal{S}, \forall n \geq 1 \tag{2}$$

Theorem 2.6 says that $\|P_i^n - \pi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{j \in \mathcal{S}} |p_{ij}(n) - \pi_j|$ decays exponentially fast to 0 as $n \to \infty$.

Definition 2.7. Spectral gap. $\gamma = 1 - \lambda_*$.

Note: $\lambda_*^n = (1-\gamma)^n \leq e^{-\gamma n}$ (since $1 - x \leq e^{-x}, \quad \forall x \geq 0$). This shows that if the spectral gap is large, convergence is fast; if it is small, convergence is slow.

Definition 2.8. Mixing time. For a given $\epsilon > 0$, define $T_\epsilon = \inf\{n \geq 1 : \max\limits_{i \in \mathcal{S}} \|P_i^n - \pi\|_{\mathrm{TV}} \leq \epsilon\}$.

Remark. One can show that the sequence $\max_{i \in \mathcal{S}} \|P_i^n - \pi\|_{\mathrm{TV}}$ is decreasing in n, so the above definition makes sense.

2.5 Examples

Example 2.9. Cyclic random walk.

Remember from last week that the stationary distribution is $\pi_j = \frac{1}{N}$, $\forall j \in \mathcal{S}$, and that $p = q = \frac{1}{2}$ implies detailed balance.

$$P = \begin{pmatrix} 0 & 1/2 & 0 & \cdots & 0 & 1/2 \\ 1/2 & 0 & 1/2 & \cdots & 0 & 0 \\ 0 & 1/2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1/2 \\ 1/2 & 0 & 0 & \cdots & 1/2 & 0 \end{pmatrix}$$

The eigenvalues of P are $\lambda_k = \cos\left(\frac{2k\pi}{N}\right)$, $\forall k = 0, ..., N-1$ (note that the eigenvalues are not ordered here).

If N is even, the chain is periodic of period 2, therefore not ergodic, therefore a limiting distribution does not exist (and the spectral gap is equal to zero).

If N is odd,

$$\lambda_* = \left|\cos\left(\frac{2\pi(N-1)/2}{N}\right)\right| = \left|\cos\left(\pi\left(1 - \frac{1}{N}\right)\right)\right| = \cos\left(\frac{\pi}{N}\right)$$

The spectral gap is:

$$\gamma = 1 - \cos\frac{\pi}{N} \simeq 1 - \left(1 - \frac{\pi^2}{2N^2}\right) = \frac{\pi^2}{2N^2}$$

because $\cos(x) \simeq 1 - x^2/2$ close to $x = 0$. The spectral gap is therefore $\mathcal{O}\left(\frac{1}{N^2}\right)$. To compute the mixing time T_ϵ, we use Theorem 2.6:

$$\max_{i \in \mathcal{S}} \|P_i^n - \pi\|_{\text{TV}} \leq \frac{\lambda_*^n}{2\sqrt{\pi_i}} \leq \frac{e^{-\gamma n}}{2\sqrt{1/N}} \simeq \frac{\sqrt{N}}{2}\exp\left(-\frac{\pi^2 n}{2N^2}\right)$$

This goes fast to 0 if $n \gg N^2$, for example if $n = cN^2 \log N$, with $c > 0$ a sufficiently large constant. This confirms our intuition that the larger the circle is, the longer we have to wait for the chain to reach equilibrium.

Example 2.10. Complete graph of N vertices.

$$p_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{N-1} & \text{otherwise} \end{cases}$$

$$P = \frac{1}{N-1}\begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}$$

The stationary distribution is uniform:

$$\pi_i = \frac{1}{N}, \ \forall i \in \mathcal{S}$$

The eigenvalues of P are $\lambda_0 = 1$, $\lambda_k = -\frac{1}{N-1}$, $\forall 1 \leq k \leq N-1 \Rightarrow \lambda_* = \frac{1}{N-1}$. The spectral gap is therefore $\gamma = 1 - \frac{1}{N-1} = \frac{N-2}{N-1}$.

To compute the mixing time T_ϵ, we use Theorem 2.6:

$$\|P_i^n - \pi\|_{\text{TV}} \leq \frac{\lambda_*^n}{2\sqrt{1/N}} \leq \frac{\sqrt{N}}{2}\exp\left(-n\frac{N-2}{N-1}\right)$$

which shows that the mixing time is roughly $\mathcal{O}(1)$.