Markov Chains and Algorithmic Applications: WEEK 3

1 Stationary distribution

As a reminder, the distribution of the Markov chain $(X_n,\, n \geq 0)$ at time n is given by:

$$\pi_j^{(n)} = \mathbb{P}(X_n = j), \quad j \in S$$

Let us compute for $j \in S$:

$$\pi_j^{(n+1)} = \mathbb{P}(X_{n+1} = j) = \sum_{i \in S} \mathbb{P}(X_{n+1} = j, X_n = i) = \sum_{i \in S} \mathbb{P}(X_{n+1} = j | X_n = i)\, \mathbb{P}(X_n = i) = \sum_{i \in S} \pi_i^{(n)}\, p_{ij}$$

In vector notation (considering $\pi^{(n)}, \pi^{(n+1)}$ as row vectors), this reads:

$$\pi^{(n+1)} = \pi^{(n)}\, P$$

which further implies that $\pi^{(n)} = \pi^{(0)}\, P^n$. This motivates also the following definition:

Definition 1.1. Let $(X_n,\, n \geq 0)$ be a Markov chain with transition matrix P. A probability distribution $\pi = (\pi_i, i \in S)$ is said to be a **stationary distribution** of the chain X if:

$$\pi_j = \sum_{i \in S} \pi_i\, p_{ij} \quad \forall j \in S \quad \text{i.e.} \quad \pi = \pi\, P$$

Consequences.

- If π is stationary then $\pi P^n = \pi P P^{n-1} = \pi P^{n-1} = ... = \pi$.
- In particular, if the initial distribution $\pi^{(0)} = \pi$, then $\forall n \geq 0$, $\pi^{(n)} = \pi^{(0)} P^n = \pi P^n = \pi$.

Remarks.

- A stationary distribution is a solution of a system of linear equations. Moreover, we can see than π is a (left-)eigenvector of P.
- Thus, π may not always exist (think of the asymmetric random walk on $\mathbb{Z}$, which escapes to infinity for example).
- Moreover, when π exists, it may not be unique.
- And even when π exists and is unique, it is not always the case that $\lim_{n \to \infty} \pi^{(n)} = \pi$.

On the practical side, note that in order to compute π, it is necessary to take into account the condition $\sum_{i \in S} \pi_i = 1$ (as the equation $\pi = \pi\, P$ alone does not impose a normalization on the vector π).

The following theorem is a deep one, especially due to the fact that it provides a *necessary and sufficient* condition guaranteeing the existence of a stationary distribution for an irreducible chain. We shall nevertheless skip its rather lengthy proof here.

Theorem 1.2 (without proof)**.** Let $(X_n,\, n \geq 0)$ be an irreducible Markov chain. Then

$$X \text{ is positive-recurrent if and only if X admits a stationary distribution } \pi$$

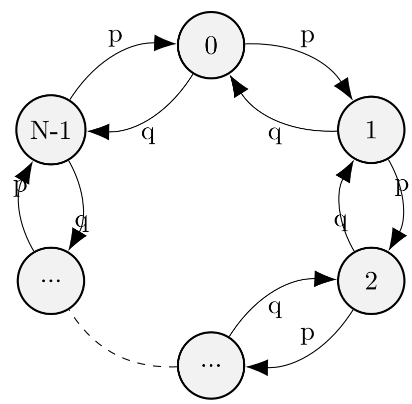
Moreover, if π exists, then it is unique and given by $\pi_i = \dfrac{1}{\mu_i} = \dfrac{1}{\mathbb{E}(T_i | X_0 = i)}$, $i \in S$.

Remark. As X is positive-recurrent, we also have $\mu_i < +\infty$, so $\pi_i > 0$, $\forall i \in S$.

Corollary 1.3. A *finite irreducible* chain always admits a unique stationary distribution π, since a finite irreducible chain is always positive-recurrent.

Example 1.4. Cyclic random walk on $S = \{0, ..., N-1\}$:

$$P = \begin{pmatrix} 0 & p & 0 & ... & 0 & 0 & q \\ q & 0 & p & ... & 0 & 0 & 0 \\ 0 & q & 0 & ... & 0 & 0 & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & 0 & 0 & ... & 0 & p & 0 \\ 0 & 0 & 0 & ... & q & 0 & p \\ p & 0 & 0 & ... & 0 & q & 0 \end{pmatrix}$$



The chain is finite ans irreducible, so it is positive-recurrent. By the theorem, π exists and is unique. In this case, π is given by the uniform distribution: $\pi_i = \frac{1}{N}$, $\forall i \in S$, as we shall see below. Therefore, $\mu_i = \mathbb{E}(T_i | X_0 = i) = N$ for all $i \in S$.

Remarks.

- It is far from evident that μ_i should not depend on the values of p and q.
- It is also far from evident that π should be called a stationary distribution for all values of p and q, as when $p \neq q$, the chain has a tendency to constantly rotate in one direction, so in some sense, it does not seem "fully stationary". We will come back to this later.

Why is the stationary distribution uniform here?

The state space of the chain is *finite* and matrix P is *doubly-stochastic*, i.e., we have both $\sum_{j \in S} p_{ij} = 1$, $\forall i \in S$ and $\sum_{i \in S} p_{ij} = 1$, $\forall j \in S$. In this case, we see that $\pi_i = \frac{1}{N}$ for all $i \in S$ is a solution of $\pi = \pi P$. Indeed, $\forall j \in S$,

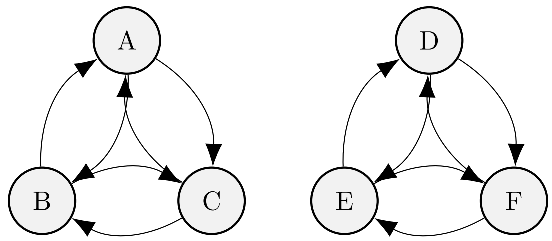
$$(\pi P)_j = \sum_{i \in S} \pi_i \, p_{ij} = \sum_{i \in S} \frac{1}{N} p_{ij} = \frac{1}{N} \underbrace{\sum_{i \in S} p_{ij}}_{=1 \; \forall j} = \frac{1}{N} = \pi_j$$

Example 1.5 (counter-example)**.** The simple symmetric random walk on $\mathbb{Z}$.

It is irreducible, recurrent, but not positive-recurrent, so the above theorem says that it does not admit a stationary distribution π (i.e., there is no distribution π satisfying $\pi = \pi P$). A possible candidate would indeed be the uniform distribution, because the matrix P is doubly-stochastic in this case also, but such a uniform distribution does not exist on the infinite set $\mathbb{Z}$.

Example 1.6. Here are examples where the theorem doesn't apply (in the graphs below, we always assume that back-and-forth arrows between 2 states have the same weight):

Markov chain with two recurrent classes:

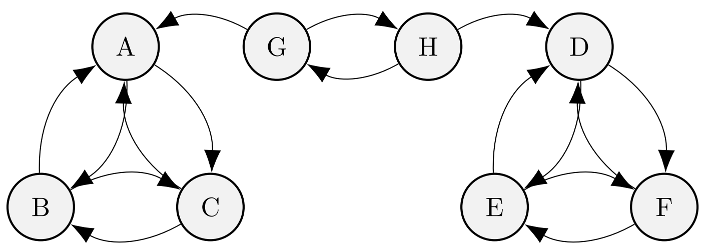


$\pi^{(A,B,C)} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0)$, $\pi^{(D,E,F)} = (0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and any convex combination of these two are possible stationary distributions:

$$\pi = \alpha\, \pi^{(A,B,C)} + (1-\alpha)\, \pi^{(D,E,F)}, \quad \text{where } 0 \leq \alpha \leq 1$$

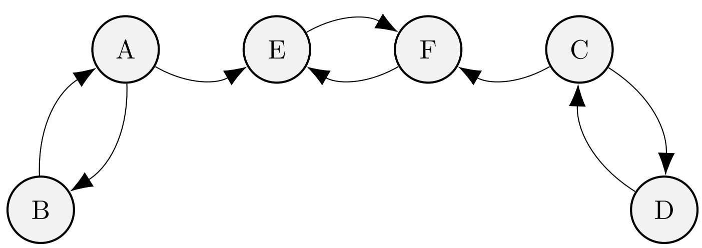
The number of stationary distributions in this case is therefore infinite (and even more: uncountable!).

Another one:



In this case, there is one transient class in the middle (G,H) and two recurrent classes (A,B,C) and (D,E,F). The problem is the same as in the previous example: there are multiple stationary distributions: set $\pi_G = \pi_H = 0$ and consider similar convex combinations of $\pi^{(A,B,C)}$ and $\pi^{(D,E,F)}$ as above.

And yet another one:



In the graph above, the states (A,B) form a transient class, (E,F) a recurrent class and (C,D) another transient class. Taking $\pi_A = \pi_B = \pi_C = \pi_D = 0$ and $\pi_E = \pi_F = \frac{1}{2}$ gives the unique stationary distribution of the chain. So it need not be that the chain is irreducible for a unique stationary distribution to exist (note also that some states have weight $\pi_i = 0$ here).

2 Limiting distribution

Now the question is: for what kind of Markov chains does it hold that, *whatever the initial distribution* $\pi^{(0)}$, $\lim_{n\to\infty} \pi^{(n)} = \pi$?

Definition 2.1. A distribution $\pi = (\pi_i,\, i \in S)$ is said to be a **limiting distribution** of the chain $(X_n,\, n \geq 0)$ if for every initial distribution $\pi^{(0)}$, $\lim_{n\to\infty} \pi^{(n)} = \pi$.

Remarks.

- Such a distribution may not exist.
- If it exists, it is unique!
- If π is a limiting distribution, then it is a also stationary distribution.

Proof. (Finite case; the infinite case requires to handle more technical details).
It always holds that $\pi^{(n+1)} = \pi^(n)\, P$ for every $n \geq 0$. So If π is a limiting distribution, then $\lim_{n\to\infty} \pi^{(n)} = \pi$ and $\lim_{n\to\infty} \pi^{(n+1)} = \pi$ also. Therefore, $\pi = \pi P$, that is, π is a stationary distribution. □

Definition 2.2. A Markov chain is said to be **ergodic** if it is irreducible, aperiodic and positive-recurrent.

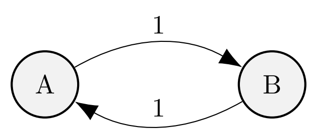
With this definition in hand, we can now state the central theorem of this first part of the course.

Theorem 2.3 (Ergodic theorem)**.** Let X be an ergodic Markov chain. Then it admits a unique limiting and stationary distribution π, i.e., $\forall \pi^{(0)}$, $\lim_{n\to\infty} \pi^{(n)} = \pi$ and $\pi = \pi P$.

Equivalent statements.

- $\forall \pi^{(0)}$ and $\forall j \in S$, $\lim_{n\to\infty} \mathbb{P}(X_n = j) = \pi_j$ and $\pi_j = \sum_{i \in S} \pi_i p_{ij}$
- $\forall i, j \in S$, $\lim_{n\to\infty} \mathbb{P}(X_n = j | X_0 = i) = \pi_j$

Remark. Aperiodicity matters! Indeed, consider the chain with $S = \{A, B\}$ and $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

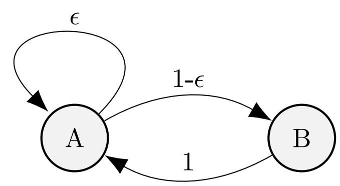


The periodicity is 2. Does a stationary distribution exist ? Yes, $\pi = (\frac{1}{2}, \frac{1}{2})$, as the chain is finite and the matrix P is doubly-stochastic. Moreover, since it is irreducible and positive-recurrent, the previous theorem guarantees uniqueness. However, there is no limiting distribution. Indeed, for $\pi^{(0)} = (1, 0)$, we have for all $n \geq 0$:

$$\pi^{(2n)} = (1, 0) \quad \text{and} \quad \pi^{(2n+1)} = (0, 1)$$

so the sequence $(\pi^{(n)},\, n \geq 0)$ does not converges.

Consider a slightly different case: $S = \{A, B\}$ and $P = \begin{pmatrix} \epsilon & 1-\epsilon \\ 1 & 0 \end{pmatrix}$.



Here the chain is finite, irreducible, positive-recurrent **and** aperiodic, so the ergodic theorem applies and the solution to the equation $\pi = \pi P$ (given here by $\pi = (\frac{1}{2-\epsilon}, \frac{1-\epsilon}{2-\epsilon})$) is also a limiting distribution.

Remark. For a periodic chain, what does it mean to have a stationary distribution π ? Well, it is still true that on average over time, the chain will be in state i a proportion π_i of the time. Said otherwise, $\forall \pi^{(0)}$ and $i \in S$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \underbrace{\mathbb{P}(X_k = i)}_{\pi_i^{(k)}} = \pi_i$$