1 Stationary distributions

Recall the Markov transition matrix $\mathbf{P}$ where $\mathbf{P}_{ij}$ is the probability of going from state i to state j, $i \to j$. We assume there are N states so $\mathbf{P}$ is an $N \times N$ matrix. We know that $\sum_j \mathbf{P}_{ij} = 1$, since this is a Markov transition matrix so each row should sum to 1. Also, note that the largest eigenvalue of $\mathbf{P}$ is 1 with corresponding eigenvector $\mathbf{1} = (1, ...1)$.

We denote a probability vector $\nu = (\nu_1, ..., \nu_N)$ as the initial probability of being in any state$1, ..., N$. Note that the space of probability vectors is the $N-1$ dimensional simplex which we denote

$$\Delta^{N-1} = \left\{ (\nu_1, ..., \nu_N) : \sum_{i=1}^{N} t_i = 1,\, t_i \geq 0, i = 1, ..., N \right\}.$$

So we know

$$\nu_{t+1} = \nu_t \mathbf{P}.$$

We are interested in the following question: given a Markov matrix $\mathbf{P}$ and any initial probability vector $\nu \in \Delta^{N-1}$ does the following converge to a unique probability vector

$$\lim_{t \to \infty} \nu \mathbf{P}^t = \tilde{\nu}.$$

If the above convergence occurs then $\tilde{\nu}$ is called the stationary or invariant distribution. We also will ask can we characterize the above convergence via properties of $\mathbf{P}$.

There are two conditions for a Markov chain to have an invariant distribution. One is the chain is aperiodic the other is that the chain is irreducible. We now consider these two conditions.

2 Irreducible

For a Markov chain with state space S there is a possible path from state i to state j if there is a sequence of states

$$i = i_0 \to i_1 \to \cdots \to i_n = j$$

with $P_{\ell-1,\ell} > 0$, $\ell = 1, ..., n$. States i and j communicate if there is a possible path $i \to j$ and $j \to i$, or $i \leftrightarrow j$. All states communicate with themselves $i \leftrightarrow i$. A communication class is a partition $X \subset S$ where for all $i, j \in X$ $i \leftrightarrow j$ is a communication class. The operation $\leftrightarrow$ partitions the states into communication classes.

A Markov chain is irreducible if there is only one communication class. Irreducibility can also be stated as for all states $i, j \in S$, the probability for getting from state i to j in n steps is positive, $p_{ij}^n > 0$.

We can also ask if one can check for irreducibility based on the Markov transition matrix $\mathbf{P}$. One can think of this as a graph theory problem. Think of each node in the. graph as one of the states. The entry $\mathbf{P}_{ij}$ can be thought of as the weight of the edge between nodes i and j. We can ask the question of whether the weighted graph encoded as $\mathbf{P}$ can be cut, if it cannot then the chain is irreducible.

We now state a function which can help us with measuring if a weighted graph can be cut

$$\lambda_1 = \min_{v \in \mathbb{R}^N, v \perp \mathbf{1}, \|v\|=1} \left[\sum_{i \sim j} (v_i - v_j)^2 \mathbf{P}_{ij} \right]$$

and if $\lambda_1 = 0$ then the weighted graph can be cut. Here $i \sim j$ is the edge between i and j. The vector v assigns a value to each vertex and the idea is to assign as similar as possible values to neighboring vertices or in other words find an assignemnt of vertex values which are positive on one partition of the graph and negative on the other which can perfectly (with 0 edges in between) cut the graph.

The last point is that λ_1 can be computed via an eigenvalue computation. The graph Laplacian is defined as

$$\mathbf{L} = \mathbf{I} - \mathbf{P}.$$

The smallest eigenvalue of $\mathbf{L}$ is $\lambda_0 = 0$ and if the second smallest λ_1 is zero then the Markov chain is reducible if $\lambda_1 > 0$ then the chain is irreducible. Using linear algebra one can show the following relation between the eigenvalue of the graph Laplacian and the cut measure in the previous paragraph

$$\lambda_1 = \min_{v \perp \mathbf{1}} \frac{v \mathbf{L} v^T}{\|v\|^2} = \min_{v \in \mathbb{R}^N, v \perp \mathbf{1}, \|v\|=1} \left[\sum_{i \sim j} (v_i - v_j)^2 \mathbf{P}_{ij} \right].$$

3 Aperiodic

We will ask the question of whether the following limit exists $\lim_{t \to \infty} \mathbf{P}^t$.

Consider the eigen-decomposition of $\mathbf{P}$

$$\mathbf{P} = V \Lambda V^T,$$

where Λ is the diagonal matrix of ordered eigenvalues and $V = (v_1 \cdot v_N)$ where each v_i is the eigenvector corresponding to eigenvalue i.

Given the eigen-decomposition we obtain

$$\mathbf{P}^t = V \Lambda^t V^T = \sum_{i=1}^{t} \lambda_i^t v_i v_i^T.$$

So we can consider

$$\lim_{t \to \infty} \sum_{i=1}^{t} \lambda_i^t v_i v_i^T.$$

We know the largest eigenvalue of the above matrix is 1 and all the eigenvalues are between $[-1, 1]$. So, if there are no eigenvalues that are -1 then the limit exists, $\lim_{t\to\infty} \mathbf{P}^t = \widetilde{\mathbf{P}}$.

If no eigenvalues are -1 then the Markov chain is aperiodic, this raises the question of what aperiodic means.

Definition of an aperiodic state: Let p_{ii}^n denote the probability of returning to state i at step n. State i is periodic with period $t = 2, 3, ...$ iff

$$\begin{aligned} p_{ii}^t &= 0 \text{ for } n \neq t, 2t, 3t... \\ p_{ii}^t &\neq 0 \text{ for } n = t, 2t, 3t... \end{aligned}$$

If there exists no t for which the above holds then the state is aperiodic.

Note that if you have one eigenvalue of -1 and the limit $\lim_{t\to\infty} \mathbf{P}^t$ will oscillate between $v_1 v_1^T + v_N v_N^T$ and $v_1 v_1^T - v_N v_N^T$ based on whether t is odd or even with v_1 corresponding to the $\lambda_i = 1$ and v_N corresponding to $\lambda_N = -1$. This oscillator phenomena is what gives rise to periodicity.

For a irreducible Markov chain there is only one communicating class. It turns out periodicity is a class property, so if one state in a class is periodic or aperiodic the rest of the states will share the periodicity property of the state.

Lastly there is the notion of mixing time or the amount of time it takes for the Markov chain to reach a stationary distribution. For an aperiodic irreducible chain with spectral gap $_1 > 0$ and invariant distribution $\tilde{\nu}$ one can obtain the following rate of convergence

$$\|\nu_0 \mathbf{P}^t - \tilde{\nu}\| \leq C(1 - \lambda_1)^t,$$

where C is some constant.