

IFT 6760A - Lecture 7

Markov Chains

Scribe(s): Samy Coulombe, Shishir Sharma

Instructor: Guillaume Rabusseau

1 Summary

In the previous lecture, we looked at Dimensionality Reduction using PCA, covering the different perspectives of PCA i.e. that it seeks to minimize the reconstruction error as well as maximize the variance. We also briefly looked at Canonical Correlation Analysis (CCA) as an approach to find low dimensional projections with maximum correlation for two random variables of different dimensions.

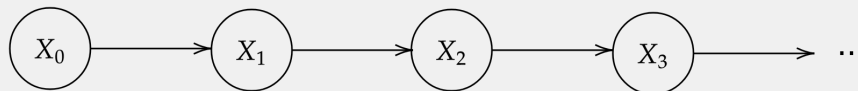
In this lecture we examine Markov Chains (MC) and their various attributes. We also discuss the different types of Markov Chains such as Reducible, Periodic and Ergodic Markov Chains and their properties.

2 Markov Chains

Definition 1 (Markov Chains). *A sequence of random variables X_0, X_1, \dots taking their values in a set of states S and following the Markov property:*

$$\mathbb{P}(X_t = i_t | X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}) = \mathbb{P}(X_t = i_t | X_{t-1} = i_{t-1})$$

i.e. the probability distribution of X_t , given the previous state X_{t-1} , is independent of the value of X_0, X_1, \dots, X_{t-2} .



Definition 2 (Time Homogeneous Markov Chains). *A Markov chain is time homogeneous if*

$$\mathbb{P}(X_t = j | X_{t-1} = i) = \mathbb{P}(X_{t'} = j | X_{t'-1} = i)$$

for all t, t' i.e. the transition probability for going from state i to j is independent of t .

Remark 3. *We shall only consider time homogeneous Markov chains with finite state spaces $S = \{1, 2, \dots, n\}$.*

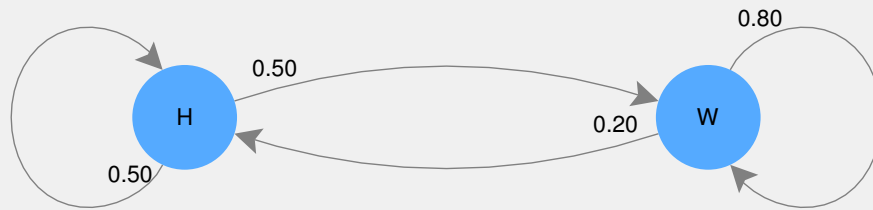


Figure 1: Markov Chain representing the process of going from work to home and vice versa.

Example 4. We consider representing the routine of a person i.e. whether he is at home all day or goes to work using Markov Chains.

For this purpose, we start with defining the two states H and W, representing whether the person remains at home or goes to work respectively. The person follows the routine of working on weekdays and staying at home on the weekend. If the person is currently at home, then depending upon whether it's a Saturday or a Sunday, he will either remain at home or leave for office respectively. Thus, we have

$$\mathbb{P}(X_{t+1} = H|X_t = H) = 1/2 = 0.5$$

$$\mathbb{P}(X_{t+1} = W|X_t = H) = 1/2 = 0.5$$

Similarly, if it's a week day, the person will stay at home tomorrow only if it's a Friday.

$$\mathbb{P}(X_{t+1} = H|X_t = W) = 1/5 = 0.2$$

$$\mathbb{P}(X_{t+1} = W|X_t = W) = 4/5 = 0.8$$

Definition 5 (Transition Matrix). The transition matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ of a Markov chain, is the matrix containing transition probabilities of the Markov Chain such that:

$$\mathbf{P}_{i,j} = \mathbb{P}(X_t = j|X_{t-1} = i)$$

- The rows of the matrix sum to 1 as the entries correspond to transition probabilities from one state to every other state.

For the above Markov Chain example, the Transition Matrix \mathbf{P} can be written as:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix} \begin{matrix} H \\ W \end{matrix}$$

where the 1st and 2nd row correspond to the transitions from state H and W respectively.

Definition 6 (Initial Distribution). The initial distribution is the probability distribution corresponding to X_0 , denoted by the row vector $\mathbf{p}_0 \in \mathbb{R}^n$ such that

$$(\mathbf{p}_0)_i = \mathbb{P}(X_0 = i)$$

i.e. the i^{th} entry of the vector corresponds to the probability of X_0 taking value i .

Similar to the Initial distribution, the distribution over states at time t is denoted by the row vector $\mathbf{p}_t \in \mathbb{R}^n$:

$$(\mathbf{p}_t)_i = \mathbb{P}(X_t = i)$$

Writing \mathbf{p}_t as a function of \mathbf{p}_{t-1} , we have:

$$\begin{aligned}
(\mathbf{p}_t)_j &= \mathbb{P}(X_t = j) \\
&= \sum_{i=1}^n \mathbb{P}(X_t = j | X_{t-1} = i) \mathbb{P}(X_{t-1} = i) \\
&= \sum_{i=1}^n \mathbf{P}_{i,j} (\mathbf{p}_{t-1})_i \\
&= (\mathbf{p}_{t-1} \mathbf{P})_j,
\end{aligned} \tag{1}$$

Therefore,

$$\mathbf{p}_t = \mathbf{p}_{t-1} \mathbf{P} = \mathbf{p}_{t-2} \mathbf{P} \mathbf{P} = \mathbf{p}_{t-2} \mathbf{P}^2 = \mathbf{p}_{t-3} \mathbf{P}^3 = \dots = \mathbf{p}_0 \mathbf{P}^t$$

Assume that $\boldsymbol{\pi} = \lim_{t \rightarrow \infty} \mathbf{p}_t$ exists. Then,

$$\boldsymbol{\pi} = \lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^t = \lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^{t+1} = (\lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^t) \mathbf{P} = \boldsymbol{\pi} \mathbf{P}$$

meaning that $\boldsymbol{\pi}$ is a left eigenvector of \mathbf{P} (i.e. an eigenvector of \mathbf{P}^T) with eigenvalue 1!

Remark 7. We represent distributions $\mathbf{p}_0, \mathbf{p}_t, \boldsymbol{\pi}$ as row vectors throughout. As these row vectors represent actual probability distributions, they are bound by the condition that the sum of the row must be equal to 1.

Definition 8 (Stationary distribution). Any distribution $\boldsymbol{\pi} \in [0, 1]^n$ over a Markov Chain's n states is said to be a stationary distribution of that Markov Chain if-and-only-if

$$\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}$$

(where $\mathbf{P} \in [0, 1]^{n \times n}$ is the Markov Chain's transition matrix).

Remarks:

- any distribution which satisfies the aforementioned condition is necessarily a left eigenvector of the Markov Chain's transition matrix; specifically, an eigenvector associated with the eigenvalue 1.
- every finite homogeneous Markov Chain has at least 1 stationary distribution.

Definition 9 (Limiting distribution). Any distribution $\boldsymbol{\pi} \in [0, 1]^n$ over a Markov Chain's n states is said to be a limiting distribution of that Markov Chain if

$$\lim_{k \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^k = \boldsymbol{\pi}$$

for some initial distribution \mathbf{p}_0 (where $\mathbf{P} \in [0, 1]^{n \times n}$ is the Markov Chain's transition matrix).

It is important to note the difference between the stationary and the limiting distribution i.e. the limiting distribution depends on the initial distribution while the stationary distribution does not. We later cover the conditions under which the two distributions end up being the same.

2.1 Reducible Markov Chains

Definition 10 (Reducible Markov Chain). A Markov Chain is said to be reducible if it possesses at least 1 state(s) from which it is impossible to reach any other state in a finite number of steps. (Figure 2, 3)

Stationary Distributions of Reducible Markov Chains:

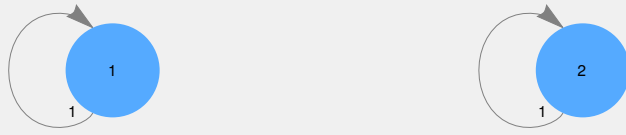


Figure 2: Example of a (trivially) reducible Markov Chain

Example 11.

Consider the simplest case: the reducible Markov Chain in Figure 2, whose transition matrix is:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For any distribution π , $\pi = \pi\mathbf{P}$ i.e. any distribution π can be a stationary distribution for this Markov Chain, thus highlighting that π need not be unique for a Markov Chain.

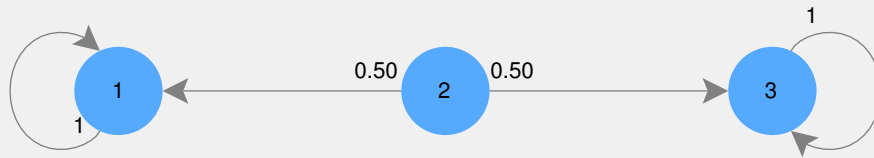


Figure 3: Example of a reducible Markov Chain

Example 12.

A similar, though less trivial case, is exemplified by the Markov Chain in Figure 3, whose transition matrix is:

$$\mathbf{P} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

For this Markov Chain, following stationary distributions can be trivially identified :

$$\pi_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \pi_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

These are derived from the fact that states 1 and 3 are absorbing states i.e once we are in either of the states, we tend to remain in that state perpetually. There is clearly no unique stationary distribution for this Markov Chain and we could try solving the linear system $\pi\mathbf{P} = \pi$ (while being mindful of the restrictions on π) to get all the possible stationary distributions, but there exists a more efficient approach leveraging the first remark raised above relating stationary distributions and left eigenvectors; we could simply find the left eigenvector of \mathbf{P} associated with the eigenvalue 1. Doing so would yield the general solution of left eigenvectors described as $[\alpha, 0, \beta]$. Due to the restrictions on π , this result is further refined to

$$\pi = \begin{pmatrix} \alpha \\ 0 \\ 1 - \alpha \end{pmatrix} \quad (2)$$

The above Markov Chain examples highlight the fact that whenever there are several non-communicating components in a Markov Chain, there does not exist a unique stationary distribution. We define non-communicating components as:

Definition 13 (Non-communicating components). For any two states i and j ,

$$\mathbb{P}(X_t = j | X_0 = i) = 0, \forall t$$

i.e. there does not exist a path from i to j after any value of time t , then those two states are part of two non-communicating components.

Remark 14. These two examples are indicative of general properties of reducible Markov Chains:

1. Reducible Markov Chains have an infinite number of stationary distributions.
2. This was not covered in the lecture, but readers inquisitive about how to derive the limiting distribution of a reducible Markov Chain are encouraged to read procedure outlined in [3].

2.2 Periodic Markov Chains

Definition 15 (Periodic Markov Chain). Periodic Markov Chains are irreducible Markov Chains which possess > 1 state(s) that can only be reached every $p > 1$ time step (where p is called the state's periodicity). See Figure 4.

Stationary Distributions of Periodic Markov Chains:

Like reducible Markov Chains, the stationary distribution of a periodic Markov Chains holds no implications regarding the corresponding limiting distribution. Furthermore, these lectures will place more emphasis on ergodic Markov Chains.

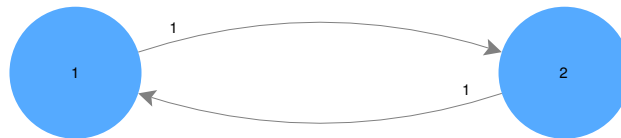


Figure 4: Example of a periodic Markov Chain; Should state 1 be visited at timepoint $t = 0$, then the Markov Chain will necessarily revisit state 1 at each completion of its periodicity.

For the above Markov Chain example, the Transition Matrix \mathbf{P} can be written as:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This Markov Chain has the unique stationary distribution $\pi_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$.

Looking at distribution over states starting with the initial distribution $\mathbf{p}_0 = \begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}$, we have:

$$\mathbf{p}_1 = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}, \mathbf{p}_2 = \begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}, \mathbf{p}_3 = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix} \dots$$

It can be easily observed that the distribution doesn't converge in this case. For this Markov Chain, we observe that the limit $\lim_{k \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^k$ exists only for initial distribution $\mathbf{p}_0 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$, because for this specific \mathbf{p}_0

$$\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \mathbf{P}^k = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \mathbf{P}^t, \text{ for non-negative integers } k \text{ and } t$$

The corresponding limiting distribution π is :

$$\pi = \lim_{k \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^k = \lim_{k \rightarrow \infty} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^k = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Remark 16. A finite homogeneous Markov chain that is irreducible and aperiodic, has a unique stationary distribution π , and the chain will converge towards this distribution from any initial distribution \mathbf{p}_0 i.e. the Markov Chain will have the same stationary and limiting distribution.

2.3 Ergodic Markov Chains

Definition 17 (Ergodic Markov Chain). Markov Chains which are neither reducible nor periodic are said to be Ergodic Markov Chains (or Regular Markov Chains). See Figure 1 for an example.

The remainder of these notes will focus on ergodic Markov Chains. Before considering the properties of the stationary distributions of ergodic Markov Chains, we must first examine some properties of the transition matrix of ergodic Markov Chains.

Property 18. Let matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ such that $\mathbf{P} \geq 0$ (i.e. all the elements of \mathbf{P} are non-negative as they correspond to probabilities) be the transition probability matrix of an ergodic Markov Chain, and recall that \mathbf{P} is row-stochastic (meaning its rows sum to 1). The following properties can be said of any such matrix \mathbf{P} :

1. \mathbf{P} possesses a right eigenvector whose corresponding eigenvalue is 1.
2. The eigenvalue of \mathbf{P} of largest magnitude has a value of 1.
3. \mathbf{P} possesses a left eigenvector π associated with the eigenvalue 1 such that π is non-negative.

Proof.

1. Recall that \mathbf{P} is row-stochastic; its rows all sum to 1. Now consider the vector $\mathbf{1} \in \mathbb{R}^n$. Clearly $\mathbf{P}\mathbf{1} = \mathbf{1}$, proving that $\lambda = 1, \mathbf{x} = \mathbf{1} \in \mathbb{R}^n$ is an eigenpair of \mathbf{P} : $\mathbf{P}\mathbf{x} = \lambda\mathbf{x}$.
2. Consider $\mathbf{P}\mathbf{x} = \lambda_1\mathbf{x}$, and let \mathbf{x}_i be the component of \mathbf{x} of largest magnitude (e.g. $|\mathbf{x}_i| \geq |\mathbf{x}_j| \forall j \in \{1..n\}$). Then by definition,

$$\begin{aligned}\mathbf{P}\mathbf{x} &= \lambda_1\mathbf{x} \\ \lambda_1\mathbf{x}_i &= \sum_{j=1}^n \mathbf{P}_{i,j}\mathbf{x}_j \\ |\lambda_1\mathbf{x}_i| &= \left| \sum_{j=1}^n \mathbf{P}_{i,j}\mathbf{x}_j \right|\end{aligned}$$

Since by definition $\mathbf{P} \geq 0$, the following inequality is raised:

$$|\lambda_1\mathbf{x}_i| = \left| \sum_{j=1}^n \mathbf{P}_{i,j}\mathbf{x}_j \right| \leq \sum_{j=1}^n \mathbf{P}_{i,j}|\mathbf{x}_j|$$

Furthermore, since $|\mathbf{x}_i| \geq |\mathbf{x}_j| \forall j \in \{1..n\}$,

$$|\lambda_1\mathbf{x}_i| \leq \sum_{j=1}^n \mathbf{P}_{i,j}|\mathbf{x}_j| \leq \sum_{j=1}^n \mathbf{P}_{i,j}|\mathbf{x}_i|$$

Since \mathbf{P} is row-stochastic, the term $\sum_{j=1}^n \mathbf{P}_{i,j}$ will always sum to 1. Therefore,

$$\begin{aligned}|\lambda_1\mathbf{x}_i| &\leq 1|\mathbf{x}_i| \\ |\lambda_1||\mathbf{x}_i| &\leq 1|\mathbf{x}_i| \\ |\lambda_1| &\leq 1\end{aligned}$$

3. From property 18.1, we know that there exists a right eigenvector paired with the eigenvalue of 1. This implies the existence of a left eigenvector paired with the same eigenvalue. Let π be that left eigenvector such that $\pi\mathbf{P} = \pi$; from properties 18.1 and 18.2, we know that \mathbf{P} possesses such an eigenpair. For sake of simplicity, assume that $\pi_i \neq 0 \forall i$ (the proof can be adapted if it is not the case). Then π can be decomposed in a unique way into a sum $\pi = \pi_+ - \pi_-$, where both π_+ and π_- are vectors with non-negative entries. This leads to:

$$\begin{aligned}\pi\mathbf{P} &= \pi \\ (\pi_+ - \pi_-)\mathbf{P} &= \pi_+ - \pi_- \\ \pi_+\mathbf{P} - \pi_-\mathbf{P} &= \pi_+ - \pi_-\end{aligned}$$

By positivity of \mathbf{P} , both $\pi_+\mathbf{P}$ and $\pi_-\mathbf{P}$ are vectors with non-negative entries, and since we assumed that such a decomposition of π is unique, we have

$$\pi_+\mathbf{P} = \pi_+, \pi_-\mathbf{P} = \pi_-$$

Hence, π_+ is a non-negative left eigenvector of \mathbf{P} associated with the eigenvalue 1. □

Having established the properties of the transition matrix \mathbf{P} of any ergodic Markov Chain, we now examine the following properties of any such Markov Chain :

- The stationary distribution exists and is unique.
- The limiting distribution converges to the stationary distribution of the Markov Chain as the number of time steps t approaches infinity from any initial distribution.

Following sections will provide a deeper discussion of the aforementioned properties, but we first introduce the Perron-Frobenius Theorem, as it will be useful in for subsequent proofs.

2.4 Perron-Frobenius Theorem

Theorem 19 (Perron-Frobenius Theorem). *For any strictly positive matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ (such that \mathbf{P} 's smallest entry > 0), there exists an eigenpair λ_1, \mathbf{x}_1 such that:*

1. $\lambda_1 > 0$, and \mathbf{x}_1 is strictly positive (its smallest entry > 0).
2. for any other eigenvalue λ of \mathbf{P} (such that $\lambda \neq \lambda_1$), $|\lambda| < \lambda_1$.
3. λ_1 's algebraic and geometric multiplicities are both 1.

Note: The proof of the Perron-Frobenius Theorem was omitted from the lecture material. Inquisitive readers are referred to [1] for its formal demonstration.

Corollary 20. *The Perron-Frobenius Theorem also holds when \mathbf{P} is non-negative (meaning \mathbf{P} may contain 0 entries), but with the additional requirement that there must exist some power t of \mathbf{P} for which \mathbf{P}^t is strictly positive.*

The following lemma relates the transition matrix of any ergodic Markov Chain with the Perron-Frobenius theorem (along with its corollary):

Lemma 21. *Let $\mathbf{P} \in \mathbb{R}^{n \times n}$ such that $\mathbf{P} \geq 0$ (i.e. \mathbf{P} only has non-negative elements) be the transition matrix of an ergodic Markov Chain with n states. Then, for some time step $t \geq 1$, we can see that $\mathbf{P}^{t'} > 0 \forall t' \geq t$.*

A formal proof for this lemma can be found in [1] (page 10, Lemma 1.6), but the intuitive explanation is straightforward: given that \mathbf{P} defines a regular, fully-communicating Markov Chain, there must exist a number of time steps $t' \geq 1$ after which it is possible to reach any of the chain's n states; proving this to be false would contradict the claim that the Markov Chain is ergodic, as it would imply the existence of ≥ 1 non-communicating component in the chain.

Together with the Perron-Frobenius theorem, this lemma can be used to prove that ergodic all ergodic Markov Chains have a unique stationary distribution $\pi \in \mathbb{R}^n$ such that $\pi \geq 0$.

2.5 Unicity of Stationary Distribution of Ergodic Markov Chains

Property 22. Let $\mathbf{P} \in \mathbb{R}^{n \times n}$ such that $\mathbf{P} \geq 0$ define the transition matrix of an ergodic Markov Chain. Then, there exists a unique probability distribution vector $\boldsymbol{\pi}$ over the states of the chain such that $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$.

Proof. (adapted from [2])

Lemma 21 relates the corollary to the Perron-Frobenius theorem to the transition matrix of ergodic Markov Chains. From the Perron-Frobenius theorem, we can therefore claim that \mathbf{P} 's largest eigenvalue λ_1 has a geometric and algebraic multiplicity of 1. From Property 19.2, we know that λ_1 has a value of 1. Therefore, there is a *unique* left eigenvector $\boldsymbol{\pi}$ which satisfies $\boldsymbol{\pi}\mathbf{P} = \lambda_1\boldsymbol{\pi} = 1\boldsymbol{\pi}$. And by definition, this *unique* left eigenpair ($\lambda_1 = 1, \boldsymbol{\pi}$) satisfies the definition of a stationary distribution. \square

2.6 Convergence to the Stationary Distribution of Ergodic Markov Chains

Having established that every ergodic Markov Chain possesses a unique stationary distribution, we can show a second key property of ergodic Markov Chains: that their limiting distribution converges to the chain's stationary distribution as the number of time steps t approaches infinity.

Property 23. Let $\mathbf{P} \in \mathbb{R}^{n \times n}$ such that $\mathbf{P} \geq 0$ define the transition matrix of an ergodic Markov Chain, and $\boldsymbol{\pi} \in \mathbb{R}^n$ such that $\boldsymbol{\pi} \geq 0$ define its stationary distribution. Then the chain's stationary distribution converges to its limiting distribution, which is obtained as $\lim_{t \rightarrow \infty} \mathbf{p}_0\mathbf{P}^t$ for any initial probability distribution \mathbf{p}_0 .

Proof. (adapted from [1, 2])

We first consider the eigendecomposition of the transition matrix \mathbf{P} :

$$\mathbf{P} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{U}$$

Where entries of the diagonal matrix $\boldsymbol{\Lambda}$ (eigenvalues of \mathbf{P}) are in non-increasing order with respect to their absolute value ($1 > |\lambda_2| > \dots > |\lambda_n|$)¹. As expected, the columns of \mathbf{V} and rows of \mathbf{U} respectively represent the right and left eigenvectors associated with the eigenvalues in $\boldsymbol{\Lambda}$ (for simplicity, we assume that these eigenvectors are all real and unique - up to normalisation - but the following also hold in the general case). Of particular importance are the first right and left eigenvectors: $\mathbf{V}[:, 0]$, and $\mathbf{U}[0, :]$ (where row and column indices start at 0). By definition, the first right eigenvector is a vector of 1s, and the first left eigenvector corresponds to the stationary distribution of the Markov Chain.

To describe the probability distribution over the states of the Markov Chain as the number of time steps approaches infinity we consider the limit:

$$\lim_{t \rightarrow \infty} \mathbf{P}^t = \mathbf{V}\boldsymbol{\Lambda}^t\mathbf{U}$$

Recall the property that the dominant eigenvalue of \mathbf{P} is 1, and all other eigenvalues are of lesser magnitude (ref); this implies that:

$$\lim_{t \rightarrow \infty} \boldsymbol{\Lambda}^t = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 \end{bmatrix} \quad (3)$$

Taking this into account,

$$\lim_{t \rightarrow \infty} \mathbf{V}\boldsymbol{\Lambda}^t\mathbf{U} = \mathbf{V} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 \end{bmatrix} \mathbf{U} = \mathbf{V} \begin{bmatrix} \mathbf{U}_{0,0} & \mathbf{U}_{0,1} & \dots & \mathbf{U}_{0,n} \\ 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} -\mathbf{U}[0, :] - \\ -\mathbf{U}[0, :] - \\ \vdots \\ -\mathbf{U}[0, :] - \end{bmatrix} \quad (4)$$

Where the first row of \mathbf{U} ($\mathbf{U}[0, :]$) is the stationary distribution $\boldsymbol{\pi}$ of the Markov Chain defined by \mathbf{P} :

$$\lim_{t \rightarrow \infty} \mathbf{P}^t = \mathbf{V}\boldsymbol{\Lambda}^t\mathbf{U} = \begin{bmatrix} -\mathbf{U}[0, :] - \\ -\mathbf{U}[0, :] - \\ \vdots \\ -\mathbf{U}[0, :] - \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\pi} - \\ -\boldsymbol{\pi} - \\ \vdots \\ -\boldsymbol{\pi} - \end{bmatrix} \quad (5)$$

¹This strict inequality is a simplifying assumption. Interested readers are encouraged to consult section 1.3 in [2] for the general case

As \mathbf{p}_0 is by definition a row stochastic vector, it can be seen that

$$\lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^t = \mathbf{p}_0 \begin{bmatrix} -\pi- \\ -\pi- \\ \vdots \\ -\pi- \end{bmatrix} = \pi \quad (6)$$

Hence, the stationary distribution of an ergodic Markov Chain converges to the chain's limiting distribution as the number of time steps (t) taken over the Markov Chain approaches infinity. \square

Remark 24.

1. The limit above holds regardless of the initial probability distribution \mathbf{p}_0 ; $\lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^t = \pi$ (see note above).
2. For any initial distribution \mathbf{p}_0 , the limiting distribution $\mathbf{p}_t = \lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^t$ is proportional to $\frac{1}{c} \langle \mathbf{1}, \mathbf{p}_0 \rangle \pi$, where c is a normalizing factor; $\mathbf{1}$ is a vector of n 1s, and π is the ergodic Markov Chain's stationary distribution. By definition, \mathbf{p}_0 and π 's components sum to 1, which implies that $\mathbf{p}_t = \lim_{t \rightarrow \infty} \mathbf{p}_0 \mathbf{P}^t = \frac{\pi}{c} = \pi$.
3. An upper-bound on the rate at which any ergodic Markov Chain converges to its limiting distribution (and by definition to its stationary distribution) starting a distribution $\mathbf{p}_t = \mathbf{p}_0 \mathbf{P}^t$ can be derived from the two largest eigenvalues of the chain's transition matrix \mathbf{P} (e.g. $|\lambda_1| = 1 > |\lambda_2| > |\lambda_3| > \dots$):

Let \mathbf{p}_t denote the probability distribution over the chain's states at time t , \mathbf{p}_0 denote the initial probability distribution, π be the Markov Chain's stationary distribution, and \mathbf{P} be its transition matrix (such that $\mathbf{p}_t = \mathbf{p}_0 \mathbf{P}^t$). Then, the difference between \mathbf{p}_t and π is bounded by:

$$\|\mathbf{p}_t - \pi\|_2 = \mathcal{O}\left(\left(\frac{|\lambda_2|}{|\lambda_1|}\right)^t\right)$$

Since $\lambda_1 = 1$, this simplifies to $\mathcal{O}(|\lambda_2|^t)$. A Markov Chain with a smaller $|\lambda_2|$ will therefore require fewer time steps (t) for its empirical state distribution to converge to its stationary and limiting distributions compared to Markov Chains with larger values for $|\lambda_2|$.

References

- [1] P. Orponen. Lecture notes from t-79.250 combinatorial models and stochastic algorithms (2005): Markov chains and stochastic sampling (part i). http://www.tcs.hut.fi/Studies/T-79.250/tekstit/lecnotes_01.pdf, . Accessed: 2020-02.
- [2] P. Orponen. Lecture notes from t-79.250 combinatorial models and stochastic algorithms (2005): Markov chains and stochastic sampling (part ii). http://www.tcs.hut.fi/Studies/T-79.250/tekstit/lecnotes_02.pdf, . Accessed: 2020-02.
- [3] J. Resing. Lecture notes from stochastic operations research (2dd27): Limiting behaviour of markov chains. https://www.win.tue.nl/~resing/SOR/eng/college3_09_eng.pdf. Accessed: 2020-02.