Markov Chains

John C. Chao

Econ 721 Lecture Notes

November 1, 2022

Stochastic Process

- **Definition:** A stochastic process $\{X_t : t \in \mathbb{T}\}$ is a collection of random variables, where the variables X_t take values in some set $\mathcal{S}$ called the state space and where the set $\mathbb{T}$ is called the index set.
- **Remarks:**

(i) For our purposes, we can think of t as time.

(ii) Note that $\mathbb{T}$ can be discrete, e.g., $\mathbb{T} = \{0, 1, 2, ...\}$, or $\mathbb{T}$ can be continuous, e.g., $\mathbb{T} = [0, \infty)$; and the same goes for the state space $\mathcal{S}$. Our discussion here will focus only on stochastic processes for which the state space is discrete, either $\mathcal{S} = \{s_1, s_2, ..., s_N\}$ or $\mathcal{S} = \{s_1, s_2, ...\}$ and for which the index set is $\mathbb{T} = \{0, 1, 2, ...\}$.

Markov Chain

- Roughly speaking, a Markov chain is a stochastic process for which the distribution of X_t depends only on X_{t-1}. More formally, we have the following definition
- **Definition:** The process $\{X_n : n \in \mathbb{T}\}$ is a Markov chain if

$$\Pr\left(X_n = x | X_0, .., X_{n-1}\right) = \Pr\left(X_n = x | X_{n-1}\right)$$

for all n and for all $x \in \mathcal{S}$.

- **Remark:** Note that, for a Markov chain, the joint probability mass function (pmf) factors as follows:

$$f\left(x_1, ..., x_n\right) = f\left(x_1\right) f\left(x_2 | x_1\right) f\left(x_3 | x_2\right) \times \cdots \times f\left(x_n | x_{n-1}\right).$$

Markov Chain

- The key quantities of a Markov chain are the probabilities of jumping from one state to another. Hence, it is important for us to define the so-called transition probabilities.
- **Transition Probabilities:** A Markov chain is said to be homogeneous if $\Pr(X_{n+1} = s_j | X_n = s_i)$ does not change with time. Thus, for a homogenous Markov chain, we have

$$\Pr(X_{n+1} = s_j | X_n = s_i) = \Pr(X_1 = s_j | X_0 = s_i).$$

  We will focus our discussion here only on homogeneous Markov chains.
- **Definition:** The conditional probabilities

$$p_{ij} = \Pr(X_{n+1} = s_j | X_n = s_i) \text{ for } s_i, s_j \in \mathcal{S}$$

  are called the transition probabilities of the chain. For Markov chain with finite state space, the matrix $\mathbf{P}$ whose $(i,j)^{th}$ element is p_{ij} is called the transition matrix.

Markov Chain

- **Some Properties of P**:

(i) $p_{ij} \geq 0$ for all i, j.
(ii) $\sum_{j=1}^{N} p_{ij} = 1$ for all i (Hence, each row of $\mathbf{P}$ can be viewed as a pmf.)

- **Remark:** It follows from property (ii) that if we let $\mathbf{e} = (1, 1, .., 1)'$ be an $N \times 1$ vector of ones; then,

$$\mathbf{Pe} = \mathbf{e}$$

so that $\mathbf{e}$ can be viewed as a (positive) eigenvector of $\mathbf{P}$ associated with a unit eigenvalue.

- Furthermore, let

$$\begin{aligned} p_{ij}(n) &= \Pr\left(X_{m+n} = s_j \middle| X_m = s_i\right) \\ &= \Pr\left(X_n = s_j \middle| X_0 = s_i\right) \end{aligned}$$

be the probability of going from state s_i to state s_j in n steps. Let $\mathbf{P}_n$ be the matrix whose $(i, j)^{th}$ element is $p_{ij}(n)$.

- The elements of $\mathbf{P}_n$ are called the *n*-step transition probabilities.
- **Theorem (The Chapman-Kolmogorov equations):** The *n*-step transition probabilities satisfy

$$\begin{aligned} p_{ij}(m+n) &= \sum_{k=1}^{N} p_{ik}(m)\, p_{kj}(n) \\ &= \sum_{k=1}^{N} \Pr\left(X_{m+n} = s_j \middle| X_m = s_k\right) \Pr\left(X_m = s_k \middle| X_0 = s_i\right) \end{aligned}$$

for every $i, j \in \{1, ..., N\}$

Markov Chain

- **Proof:**

$$
\begin{aligned}
& p_{ij}(m+n) \\
= \; & \Pr\left(X_{m+n} = s_j \middle| X_0 = s_i\right) \\
= \; & \sum_{k=1}^{N} \Pr\left(X_{m+n} = s_j, X_m = s_k \middle| X_0 = s_i\right) \\
= \; & \sum_{k=1}^{N} \Pr\left(X_{m+n} = s_j \middle| X_m = s_k, X_0 = s_i\right) \Pr\left(X_m = s_k \middle| X_0 = s_i\right) \\
= \; & \sum_{k=1}^{N} \Pr\left(X_{m+n} = s_j \middle| X_m = s_k\right) \Pr\left(X_m = s_k \middle| X_0 = s_i\right) \\
& \text{(by Markov property)} \\
= \; & \sum_{k=1}^{N} p_{ik}(m)\, p_{kj}(n)
\end{aligned}
$$

Markov Chain

- Since the Chapman-Kolmogorov equation holds for every element of $\mathbf{P}_{m+n}$, so what we have shown is that

$$\mathbf{P}_{m+n} = \mathbf{P}_m \mathbf{P}_n$$

- By definition, $\mathbf{P}_1 = \mathbf{P}$. The above result can be used to show that

$$\mathbf{P}_2 = \mathbf{P}_{1+1} = \mathbf{P}_1 \mathbf{P}_1 = \mathbf{P}^2.$$

  By induction, we have

$$\mathbf{P}_n = \mathbf{P}^n = \underbrace{\mathbf{P} \times \mathbf{P} \times \cdots \times \mathbf{P}}_{\text{product of } n \text{ transition matrices}}.$$

Markov Chain

- Next, let

$$\mu_n = (\mu_n(1), \ldots, \mu_n(N))$$

be a row vector where

$$\mu_n(i) = \Pr(X_n = s_i)$$

so that $\mu_n(i)$ is the marginal probability that the chain is in state s_i at time n.

- In particular, μ_0 is called the initial distribution.
- To simulate a Markov chain, it suffices to know only μ_0 and $\mathbf{P}$ using the following algorithm.

Step 1: Draw $X_0 \sim \mu_0$, so that $\Pr(X_0 = s_i) = \mu_0(i)$.

Step 2: Suppose that $X_0 = s_i$. Draw X_1 using the conditional distribution

X_1	$\Pr(\cdot \mid X_0 = s_i)$
s_1	$\Pr(X_1 = s_1 \mid X_0 = s_i)$
s_2	$\Pr(X_1 = s_2 \mid X_0 = s_i)$
$\vdots$	$\vdots$
s_N	$\Pr(X_1 = s_N \mid X_0 = s_i)$

Step 3: Suppose that $X_1 = s_j$. Draw X_2 using the conditional distribution

X_2	$\Pr(\cdot \mid X_1 = s_j)$
s_1	$\Pr(X_2 = s_1 \mid X_1 = s_j)$
s_2	$\Pr(X_2 = s_2 \mid X_1 = s_j)$
$\vdots$	$\vdots$
s_N	$\Pr(X_2 = s_N \mid X_1 = s_j)$

and so on.

Markov Chain

- The following proposition gives an interpretation for the marginal distribution μ_n at time n.
- **Proposition:** The marginal probabilities are given by

$$\mu_n = \mu_0 \mathbf{P}^n$$

Markov Chain

- **Proof:** Let e_j denote an $N \times 1$ elementary vector whose j^{th} component is 1 and all other components are 0. Note that for $j = 1, ..., N$

$$
\begin{aligned}
\mu_n e_j &= \mu_n(j) \\
&= \Pr(X_n = s_j) \\
&= \sum_{i=1}^{N} \Pr(X_n = s_j, X_0 = s_i) \\
&= \sum_{i=1}^{N} \Pr(X_n = s_j | X_0 = s_i) \Pr(X_0 = s_i) \\
&= \sum_{i=1}^{N} \mu_0(i) p_{ij}(n) \\
&= \mu_0 \mathbf{P}_n e_j \\
&= \mu_0 \mathbf{P}^n e_j. \ \square
\end{aligned}
$$

- **Definition: (Hitting Times)** Let A be a subset of $\mathcal{S}$. The hitting time $\mathbb{T}_A$ of A is defined by

$$\mathbb{T}_A = \min\{n > 0 : X_n \in A\}$$

if $X_n \in A$ for some $n > 0$ and by

$$\mathbb{T}_A = \infty$$

if $X_n \notin A$ for all $n > 0$.

- Set

$$\rho^{ij} = \Pr\left(\mathbb{T}_j < \infty \middle| X_0 = s_i\right)$$

where we abbreviate $\mathbb{T}_j = \mathbb{T}_{s_j}$. It follows that ρ^{ij} denotes the probability that a Markov chain starting in state s_i will reach state s_j in some (positive) finite time.

- **Definition:** We say that s_i reaches s_j (or s_j is accessible from s_i) if

$$\rho^{ij} = \Pr\left(\mathbb{T}_j < \infty \middle| X_0 = s_i\right) > 0$$

and we write

$$s_i \longrightarrow s_j$$

- If $s_i \longrightarrow s_j$ and $s_j \longrightarrow s_i$; then, we say that the states s_i and s_j communicate and write $s_i \leftrightarrow s_j$.
- **Theorem:** The communication relation satisfies the following properties.

1. If $s_i \leftrightarrow s_j$, then $s_j \leftrightarrow s_i$. (symmetric property)
2. If $s_i \leftrightarrow s_j$ and $s_j \leftrightarrow s_k$; then, $s_i \leftrightarrow s_k$. (transitive property)
3. The set of states $\mathcal{S}$ can be written as a disjoint union of classes

$$\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots$$

where two states s_i and s_j communicate with each other if and only if they are in the same class.

- If all states communicate with each other, then the chain is called irreducible.
- A set of states is closed if once you enter this set of states you will never leave.
- A closed set consisting of a single state is called an absorbing state.
- **Example:** Suppose that $\mathcal{S} = \{1, 2, 3, 4\}$ and

$$\mathbf{P} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In this case, the classes are $\{1, 2\}$, $\{3\}$, and $\{4\}$ and state 4 is an absorbing state.

Markov Chain

- If all states communicate with each other, then the chain is called irreducible.
- A set of states is closed if once you enter this set of states you will never leave.
- A closed set consisting of a single state is called an absorbing state.
- **Example:** Suppose that $\mathcal{S} = \{1, 2, 3, 4\}$ and

$$\mathbf{P} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In this case, the classes are $\{1, 2\}$, $\{3\}$, and $\{4\}$ and state 4 is an absorbing state.

- **Definition:** State s_i is recurrent or persistent if

$$\rho^{ii} = \Pr\left(\mathbb{T}_i < \infty \middle| X_0 = s_i\right) = 1.$$

Otherwise, state s_i is said to be transient.

- Define

$$\mathbb{I}_n\left(i\right) = \begin{cases} 1 & \text{if } X_n = s_i \\ 0 & \text{if } X_n \neq s_i \end{cases}$$

and note that the number of times that the chain is in state s_i is given by

$$N\left(i\right) = \sum_{n=1}^{\infty} \mathbb{I}_n\left(i\right)$$

Moreover, note that the event $\left\{N\left(i\right) \geq 1\right\}$ (i.e., that the chain is in state s_i at least once) is the same as the event $\left\{\mathbb{T}_i < \infty\right\}$ (i.e., that the hitting time for state s_i is finite).

Markov Chain

- **Theorem:**

(a) Suppose that state s_i is recurrent. Then,

$$\Pr\left(N\left(i\right)=\infty\middle|\ X_0=s_i\right)=1 \text{ and } \sum_{n=1}^{\infty} p_{ii}\left(n\right)=\infty$$

(b) Suppose that state s_i is transient. Then,

$$\Pr\left(N\left(i\right)<\infty\middle|\ X_0=s_i\right)=1 \text{ and } \sum_{n=1}^{\infty} p_{ii}\left(n\right)<\infty$$

Here,

$$p_{ii}\left(n\right)=\Pr\left(X_{m+n}=s_i\middle|X_m=s_i\right)=\Pr\left(X_n=s_i\middle|X_0=s_i\right).$$

- **Definition:** A Markov chain is called a transient chain if all its states are transient and a recurrent chain if all its states are recurrent.
- **Remark:** A Markov chain having a finite state space must have at least one recurrent state, so it cannot possibly be a transient chain.

Positive and Non-negative Matrices

- **Definition:** The set of distinct eigenvalues of an $N \times N$ matrix A, denoted by $\sigma(A)$, is called the spectrum of A.
- **Definition:** For $N \times N$ matrix A, the number

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$$

is called the spectral radius of A.

- **Review of Multiplicities:** For A $(N \times N)$ and $\lambda \in \sigma(A) = \{\lambda_1, \lambda_2, ..., \lambda_s\}$, we adopt the following definitions

(a) The algebraic multiplicity of λ is the number of times it is repeated as a root of the characteristic polynomial. In other words,

$$alg\ mult_A(\lambda_i) = a_i \text{ for } i = 1, ..., s$$

if and only if

$$(x - \lambda_1)^{a_1} \times \cdots \times (x - \lambda_s)^{a_s} = 0$$

is the characteristic polynomial of A.

(a) Note that here

$$\sum_{i=1}^{s} a_i = N.$$

(b) When $alg\ mult_A(\lambda) = 1$, it is called a simple eigenvalue.

(c) The geometric multiplicity of λ is $\dim N(A - \lambda I)$ (i.e., the dimension of the null space, or kernel, of $A - \lambda I$). In other words, $\text{geo mult}_A(\lambda)$ is the maximal number of linearly independent eigenvectors associated with λ.

(d) Eigenvalues such that

$$alg\ mult_A(\lambda) = geo\ mult_A(\lambda)$$

are called semisimple eigenvalues of A. A simple eigenvalue is always semisimple, but not conversely.

- Multiplicity Inequality: For each $A \in \mathbb{R}^{N \times N}$ and for each $\lambda \in \sigma(A)$,

$$geo\ mult_A(\lambda) \leq alg\ mult_A(\lambda)$$

Positive Matrices

- **Perron's Theorem (PT):** If A is an $N \times N$ matrix such that $A >_e 0$ and let $r = \rho(A)$, then the following statements are true.

(i) $r > 0$.

(ii) $r \in \sigma(A)$ (i.e., r is an eigenvalue or root of A, it is commonly called the Perron root).

(iii) $alg\ mult_A(r) = 1$

(iv) There exists an eigenvector $x >_e 0$ such that $Ax = rx$ (this part does not assert uniqueness)

(v) The Perron vector is the unique vector defined by

$$A\mathbf{p} = r\mathbf{p},\ \mathbf{p} >_e 0 \text{ and } \|\mathbf{p}\|_1 = 1$$

and except for positive multiples of $\mathbf{p}$, there are no other non-negative eigenvectors for A, regardless of the eigenvalue.

(vi) r is the only eigenvalue on the spectral circle of A.

- Now, suppose we allow zeros to creep into the matrix, and we wish to investigate to what extend Perron's theorem can be generalized to non-negative matrices with at least one zero entry. Suppose we sacrifice the existence of a positive eigenvector (among other things), then we get the following result.
- **Theorem (Non-negative Eigenpair):** Suppose that A is an $N \times N$ matrix such that $A \geq_e 0$ and let $r = \rho(A)$; then, the following statements are true.

(i) $r \in \sigma(A)$ (i.e., r is still an eigenvalue or root of A but $r = 0$ is now possible)

(ii) $Az = rz$ for some $z \in \mathcal{N} = \{x | x \geq_e 0 \text{ with } x \neq 0\}$.

- **Remarks:** Note that this is as far as Perron's theorem will generalize to non-negative matrices without additional condition. Comparing with Perron's theorem (which holds for positive matrices), we see that we have lost the following properties

$$\begin{aligned} r &> 0 \quad \text{(PT part (i))}; \\ alg\ mult_A\,(r) &= 1 \quad \text{(PT part (iii))}; \end{aligned}$$

positivity of eigenvector (PT part (iv))

r being the only eigenvalue on the spectral circle (PT part (vi)).

- That we will not have these other properties in general for non-negative matrices can be shown by the following counterexamples

(a) **Counterexample 1:**

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Here, the eigenvalues are $\lambda = 0, 0$ (i.e., the root 0 occurs with an algebraic multiplicity of 2). Moreover, there is only one linearly independent eigenvector given by

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

so that the geometric multiplicity in this case is 1, and we have in this case

$$1 = geo\ mult_A(\lambda) < alg\ mult_A(\lambda) = 2.$$

(a) **Counterexample 1 (con't):** Note that, in this example, $r = \rho(A) = 0$ (not positive, so property PT(i) does not hold), the algebraic multiplicity of r is 2 (so property PT (iii) does not hold), and eigenvectors are only non-negative and not positive (so property PT (iv) does not hold).

(b) **Counterexample 2:**

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Here, the eigen-pairs are

$$\lambda_1 = 1,\ x_1 = \begin{pmatrix} x_{1,1} \\ x_{1,2} \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and

$$\lambda_2 = -1,\ x_1 = \begin{pmatrix} x_{2,1} \\ x_{2,2} \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

(b) **Counterexample 2 (con't):** Here, we do have $r = \rho(A) = 1 > 0$ (so property PT (i) does hold); the algebraic multiplicity of r is 1 (so property PT (iii) does hold), and there does exist a positive eigenvector associated with $r = 1$, i.e.,

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(so property PT (iv) does hold). However, there are more than one eigenvalues on the spectral circle, i.e., $\lambda = 1$ and $\lambda = -1$ (so that property PT (vi) does not hold).

Non-negative Matrices

- An interesting extension of Perron's theorem was given by Frobenius, who introduced the additional hypothesis of irreducibility. Frobenius recognized that what prevents the generalization of the original Perron's theorem for positive matrices to the case of non-negative matrices is not so much the existence of zero entries but rather the positions of the zero entries. For example, property PT (iii) (i.e., $alg\ mult_A(r) = 1$) and property PT (iv) (i.e., the existence of a positive eigenvector) does not hold for

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

but does hold for

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Non-negative Matrices

- Note that the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

has eigenvalues $\lambda = 1, 1$ so $r = 1$ has an algebraic multiplicity of 2. In addition, there is only one linearly independent eigenvector given by

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

which is a non-negative but not a positive eigenvector.

- On the other hand, the matrix

$$\widetilde{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

has eigenvalues

$$\lambda_1 = \frac{1+\sqrt{5}}{2} \approx 1.62 \text{ and } \lambda_1 = \frac{1-\sqrt{5}}{2} \approx -0.62$$

so that $r = \left(1+\sqrt{5}\right)/2 > 0$ has an algebraic multiplicity of 1. Moreover, the eigenvector associated with the eigenvalue $r = \left(1+\sqrt{5}\right)/2 > 0$ can be shown to be

$$x_1 = \begin{pmatrix} x_{1,1} \\ x_{1,2} \end{pmatrix} = \begin{pmatrix} \left(1+\sqrt{5}\right)/2 \\ 1 \end{pmatrix} >_e 0$$

so it is a positive eigenvector.

Perron-Frobenius Theorem (PFT)

- **Theorem:** Suppose that the $N \times N$ matrix A is non-negative (i.e., $A \geq_e 0$). Suppose also that A is irreducible, and let $r = \rho(A)$. Then, the following statements are true.

(i) $r \in \sigma(A)$ (i.e., r is an eigenvalue) and $r > 0$.

(ii) $alg\ mult_A(r) = 1$.

(iii) There exists an eigenvector $x >_e 0$ such that $Ax = rx$ (this part does not assert uniqueness).

(iv) The unique vector defined by

$$A\mathbf{p} = r\mathbf{p}, \mathbf{p} >_e 0, \text{ and } \|\mathbf{p}\|_1 = 1$$

is called the Perron vector. There are no other non-negative eigenvectors for A, except for positive multiples of $\mathbf{p}$, regardless of the eigenvalue.

Perron-Frobenius Theorem (PFT)

- **Remark:** Comparing PFT and PT, note that by imposing irreducibility, the only property which we cannot recover from PT is (vi) which asserts that there is only one eigenvalue on the the spectral circle. Indeed, the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is non-negative and irreducible, but the eigenvalues ± 1 are both on the unit circle. The property of not having a unique eigenvalue on the spectral circle divides the set of non-negative irreducible matrices into two important classes.

Primitive Matrices

- **Definition:** A non-negative, irreducible matrix A having only one eigenvalue
$$r = \rho(A),$$
on its spectral circle is said to be a primitive matrix.
- **Definition:** A non-negative, irreducible matrix A having $h > 1$ eigenvalues on its spectral circle is called imprimitive and h is referred to as the index of imprimitivity.

Periodic and Non-periodic Markov Chain

- The positive integer d is said to be a divisor of the positive integer n if n/d is an integer. If I is a nonempty set of positive integers, the greatest common divisor of I, denoted by $g.c.d.\ I$, is defined to be the largest integer d such that d is a divisor of every integer in I. It follows that

$$1 \leq g.c.d.I \leq \min \{n : n \in I\}.$$

- Note, in particular, that if $1 \in I$, then $g.c.d.\ I = 1$, and the greatest common divisor of the set of even positive integers is 2.
- Let s_i be the state of a Markov chain such that

$$\begin{aligned} p_{ii}(n) &= \Pr(X_{m+n} = s_i | X_m = s_i) \\ &= \Pr(X_n = s_i | X_0 = s_i) \\ &> 0 \end{aligned}$$

for some finite $n \geq 1$, i.e., such that

$$\rho^{ii} = \Pr(\mathbb{T}_i < \infty | X_0 = s_i) > 0.$$

Periodic and Non-periodic Markov Chain

- In this case, we can define the period $d(i)$ to be

$$d(i) = g.c.d. \{n \geq 1 : p_{ii}(n) > 0\}.$$

- Note that

$$1 \leq d(i) \leq \min\{n \geq 1 : p_{ii}(n) > 0\}$$

  so that if

$$\begin{aligned} p_{ii} &= \Pr(X_n = s_i | X_{n-1} = s_i) \\ &= \Pr(X_1 = s_i | X_0 = s_i) \\ &> 0 \end{aligned}$$

  then $d(i) = 1$.

- **Definition:** State s_i is periodic if $d(i) > 1$ and aperiodic if $d(i) = 1$.

Periodic and Non-periodic Markov Chain

- **Claim:** Given a homogeneous, irreducible Markov chain, let s_i and s_j be any two states in its state space S; then,

$$d(i) = d(j).$$

- **Example 1 (Periodic Chain):** Consider the transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}.$$

It can be readily shown that $\mathbf{P}$ is irreducible, $\sigma(\mathbf{P}) = \{-1, 0, 1\}$, and the left-hand Perron vector is

$$\pi = \begin{pmatrix} 0.25 & 0.5 & 0.25 \end{pmatrix}.$$

Periodic and Non-periodic Markov Chain

- **Example 1 (con't):** Note, however, that

$$\mathbf{P}^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

$$\mathbf{P}^3 = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{P}^4 = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

and so on. As n increases, $\mathbf{P}^n$ oscillates between these two cases, and so this chain does not converge.

Periodic and Non-periodic Markov Chain

- **Remark:** An irreducible Markov chain is said to be a periodic chain when its transition matrix $\mathbf{P}$ is imprimitive (with the period of the chain being given by the index of imprimitivity for $\mathbf{P}$). On the other hand, an irreducible Markov chain for which $\mathbf{P}$ is primitive is called an aperiodic chain.
- **Example 2: aperiodic Markov chain**

$$\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Here, $\sigma(\mathbf{P}) = \{1, -0.5 + 0.5i, -0.5 - 0.5i\}$.

Periodic and Non-periodic Markov Chain

- **Example 3: Periodic Markov chain with a period of 2**

$$\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

In this case, $\sigma(\mathbf{P}) = \{-1, 0, 1\}$, and note that

$$\mathbf{P}^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}, \ \mathbf{P}^3 = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\mathbf{P}^4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}, ..$$

Periodic and Non-periodic Markov Chain

- **Theorem 8.5.3 of Horn and Johnson (1985):** Let A be an $N \times N$ matrix such that A is non-negative and irreducible, and let $[P_i]$ denote the set of nodes of the directed graph $\Gamma(A)$. Denote by

$$L_i = \left\{ k_1^{(i)}, k_2^{(i)}, \ldots \right\}$$

the set of lengths of all directed paths in $\Gamma(A)$ that both starts and ends at the node P_i, $i = 1, \ldots, N$. Denote by g_i the greatest common divisor of all the lengths in L_i. Then, A is primitive if and only if

$$g_i = 1 \text{ for every } i \in \{1, \ldots, N\}.$$

Convergence of Markov Chain

- Let A be an $N \times N$ (not necessarily non-negative) matrix. We want to first consider in what situations $\lim_{n\to\infty} A^n$ exists. It turns out that if some of the eigenvalues of A are such that $|\lambda| > 1$; then, this limit does not exist. On the other hand, we have the following result for the case where all eigenvalues are such that $|\lambda| < 1$.
- Proposition: For $A \in \mathbb{C}^{N\times N}$

$$\lim_{n\to\infty} A^n = 0$$

if and only if

$$\rho(A) < 1.$$

Convergence of Markov Chain

- Now, consider the case where some of the eigenvalues are on the unit circle, i.e., $|\lambda| = 1$, while the remaining eigenvalues are inside, i.e., $|\lambda| < 1$. In this case, it turns out that $\lim_{n\to\infty} A^n$ exists if and only if there exists a nonsingular $N \times N$ matrix B such that the Jordan form has the structure

$$J = B^{-1}AB = \begin{pmatrix} I_q & 0 \\ 0 & K \end{pmatrix},$$

where

$$q = \ alg\ mult_A (1) \ \text{ and } \rho(K) < 1.$$

- Define

$$C = B^{-1},$$

and partition

$$B = \left[\begin{array}{cc} \underset{N\times q}{B_1} & \underset{N\times(N-q)}{B_2} \end{array} \right], \ C = \left[\begin{array}{c} C_1' \ \ (q \times N) \\ C_2' \ \ ((N-q) \times N) \end{array} \right].$$

- Note that

$$
\begin{aligned}
A^n &= \underbrace{\left(BJB^{-1}\right) \times \left(BJB^{-1}\right) \times \cdots \times \left(BJB^{-1}\right)}_{\text{product of } n \text{ matrices of the form } BJB^{-1}} \\
&= BJ^n B^{-1}
\end{aligned}
$$

so that

$$
\begin{aligned}
\lim_{n \to \infty} A^n &= \lim_{n \to \infty} B \begin{pmatrix} I_q & 0 \\ 0 & K^n \end{pmatrix} B^{-1} \\
&= B \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix} B^{-1} \\
&= \begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} C_1' \\ C_2' \end{bmatrix} \\
&= B_1 C_1'.
\end{aligned}
$$

Convergence of Markov Chain

- In the special case where $q = 1$ (i.e., *alg mult*$_A(1) = 1$, so that $\lambda = 1$ is a simple eigenvalue), we have

$$\lim_{n \to \infty} A^n = b_1 c_1'.$$

- **Claim:** $A' c_1 = c_1$, so that c_1 is an eigenvector of A' associated with the eigevalue of 1.
- **Proof of Claim:** Note that, since $\sigma(A') = \sigma(A)$, 1 is also an eigenvalue of A'. In fact, it is of course the largest eigenvalue, in modulus, of A'.) Moreover, form the relationship

$$J = B^{-1} A B = CAB$$

we have that

$$J' = B' A' C' \Longrightarrow B'^{-1} J' = A' C' \Longrightarrow C' J' = A' C'.$$

- **Proof of Claim (con't):** Writing this out, we have

$$\begin{bmatrix} c_1 & C_2 \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 0 & K' \end{pmatrix} = A' \begin{bmatrix} c_1 & C_2 \end{bmatrix}$$

which implies that

$$\begin{bmatrix} c_1 & C_2 K' \end{bmatrix} = \begin{bmatrix} A' c_1 & A' C_2 \end{bmatrix}$$

from which we deduce that

$$A' c_1 = c_1. \ \square$$

- **Remark:** Transposing the equation above, we have

$$c_1' A = c_1'.$$

It follows that we can call c_1' the left-hand eigenvector of A associated with the eigenvalue 1.

Convergence of Markov Chain

- Returning to the convergence of Markov chains, we give the following definition. Let

$$\pi = (\pi_1, \pi_2, ..., \pi_N)$$

such that

$$\begin{aligned} \pi_i &\geq 0 \text{ for } i = 1, ..., N; \\ \sum_{i=1}^{N} \pi_i &= 1 \end{aligned}$$

so that π can be thought of as a probability mass function.

- **Definition:** π is said to be a stationary (or invariant) distribution if

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

- **Remark:** The intuition behind the above definition is that if we draw X_0 from $\mu_0 = \pi$, where π is the stationary distribution. Then, next period if we were to draw X_1 from the next period marginal distribution μ_1, we would have

$$X_1 \sim \mu_1 = \mu_0 \mathbf{P} = \boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}.$$

Similarly,

$$X_2 \sim \mu_2 = \mu_0 \mathbf{P}^2 = \boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}.$$

Continuing in this way

$$X_n \sim \mu_n = \mu_0 \mathbf{P}^n = \boldsymbol{\pi} \mathbf{P}^n = \boldsymbol{\pi}.$$

Hence, if at any time, the Markov chain has distribution π, it will then have this distribution in all subsequent periods.

Convergence of Markov Chain

- **Theorem (Fundamental Theorem of Markov Chain):** If a Markov chain $\mathbf{P}$ is irreducible and aperiodic, then it has a unique stationary distribution π. this is the unique (normalized such that the entries sum to one) left-hand eigenvector of $\mathbf{P}$ associated with the eigenvalue $\lambda = 1$. Moreover,

$$\mathbf{P}^n \longrightarrow \iota_N \pi \quad \text{as } n \longrightarrow \infty,$$

where

$$\underset{N \times 1}{\iota_N} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

General State Space

- For a general Markov chain with a state space that possibly contains an uncountable number of elements, we can define the transition kernel as follows.
- **Definition:** A transition kernel is a function K defined on $\mathcal{S} \times \mathcal{B}(\mathcal{S})$ such that

(i) $\forall x \in \mathcal{S}$, $K(x, \cdot)$ is a probability measure;

(ii) $\forall A \in \mathcal{B}(\mathcal{S})$, $K(\cdot, A)$ is measurable.

- **Remark:** For the case where X_n is a continuous random variable, the terminology kernel is also used sometimes to denote the conditional density $K(x, x')$, i.e.,

$$\Pr(X_n = A | X_{n-1} = x) = \int_A K(x, x') \, dx'.$$

General State Space

- **Definition:** A probability distribution π is the invariant (or stationary) distribution for the Markov chain with transition probabilities $P(x, \cdot)$ if

$$
\begin{aligned}
\pi(A) &= \int_{\mathcal{S}} P(y, A)\, \pi(dy) \\
&= \int_{\mathcal{S}} P(y, A)\, \pi(y)\, dy
\end{aligned}
$$

  for all $A \in \mathcal{B}(S)$ where $\mathcal{B}(S)$ denotes the Borel sigma field of S.

- To study the question of whether a Markov chain will converge to a stationary distribution if it starts from some arbitrary initial distribution μ_0, we need a measure of the sensitivity of the Markov chain to initial condition. As we have seen in the finite state space case, a key concept here is that of irreducibility.

General State Space

- However, note that in the discrete case, the chain is irreducible if all states communicate, i.e.,

$$\Pr\left(\mathbb{T}_j < \infty | X_0 = s_i\right) > 0 \ \forall s_i, s_j \in S$$

with $\mathbb{T}_j$ being the first time s_j is visited. In the more general cases (such as the case of an uncountable state space), $\Pr\left(\mathbb{T}_y < \infty | X_0 = x\right)$ may be uniformly equal to zero. Hence, it is necessary to introduce a slightly modified notion of irreducibility.
- **Definition:** A Markov chain $\{X_n\}$ with transition probability $P(x, \cdot)$ is said to be φ-irreducible if there exists a measure φ such that for every $A \in \mathcal{B}(S)$ with $\varphi(A) > 0$, the condition

$$\Pr\left(\mathbb{T}_A < \infty | X_0 = x\right) > 0$$

is satisfied for all $x \in S$.
- **Remark:** Note that in words φ-irreducibility means that any non-negligible set (i.e., one with positive φ measure) can be reached with positive probability from any starting point x.

General State Space

- Convergence of Markov chain in general state space is usually stated in terms of total variation distance. Note that, given two probability measures μ and v, their total variation distance is given by

$$\|\mu - v\|_{TV} = \sup_{A \in \mathcal{B}(S)} |\mu(A) - v(A)|$$

- **General Definition of Aperiodicity:** A Markov chain is aperiodic if there does not exist a partition

$$S = S_1 \cup \cdots \cup S_d,\ d \geq 2,\ S_i \cap S_j = \varnothing \text{ for all } i \neq j$$

such that

$$P(x, S_{i+1}) = 1 \text{ for all } x \in S_i \text{ for } i = 1, ..., d-1$$

and

$$P(x, S_1) = 1 \text{ for all } x \in S_d.$$

- **Example:**

$$\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Here, we take

$$S_1 = \{s_1\} \text{ and } S_2 = \{s_2, s_3\} \text{ so } d = 2.$$

- **Lemma:** If a Markov chain is φ-irreducible and has a stationary distribution π, then $\varphi \ll \pi$ (φ is absolutely continuous with respect to π, i.e., $\varphi(A) > 0$ then $\pi(A) > 0$ (alternatlvely, $\pi(A) = 0$ implies that $\varphi(A) = 0$).

- **Theorem:** Let $P(x, dy)$ be the transition probabilities for a φ-irreducible, aperiodic Markov chain on a general state space S, having a stationary distribution π. Then, for π-*a.e.* $x \in S$, we have

$$\lim_{n\to\infty} \|P^n(x,\cdot) - \pi(\cdot)\|_{TV} = 0,$$

where $P^n(x,\cdot)$ denotes the n-step transition probability given by

$$\begin{aligned} P^n(x,\cdot) &= \int_{\mathcal{S}} P(y,A)\, P^{n-1}(x,dy) \\ &= \int_{\mathcal{S}} P(y,A)\, p^{n-1}(x,y)\, dy. \end{aligned}$$

- **Remark:** Note that another way to write

$$\lim_{n\to\infty} \|P^n(x,\cdot) - \pi(\cdot)\|_{TV} = 0,$$

for π-*a.e.* $x \in S$, is

$$\pi\left(x \in S : \lim_{n\to\infty} \|P^n(x,\cdot) - \pi(\cdot)\|_{TV} = 0\right) = 1$$

General State Space

- **Definition:** A Markov chain is reversible (or said to satisfy detailed balance) with respect to $\pi(\cdot)$ if

$$\pi(dx) P(x, dy) = \pi(dy) P(y, dx) \text{ for all } x, y \in S.$$

- **Claim:** If a Markov chain is reversible with respect to $\pi(\cdot)$, then $\pi(\cdot)$ is stationary or invariant.
- **Proof of Claim:** If $\pi(\cdot)$ is reversible, then

$$\begin{aligned}
\int_{x \in \mathcal{S}} \pi(dx) P(x, dy) &= \int_{x \in \mathcal{S}} P(x, dy) \pi(x) dx \\
&= \int_{x \in \mathcal{S}} \pi(dy) P(y, dx) \\
&= \pi(dy) \int_{x \in \mathcal{S}} p(y, x) dx \\
&= \pi(dy). \quad \square
\end{aligned}$$