FINITE MARKOV CHAINS

DARSHAN KOMMANAPALLI, ALLEN WANG

Abstract. This paper examines finite Markov chains and their behavior in terms of both states and times. We first introduce Markov chains through a discussion of stochastic processes and the Markov property. We then consider large-time Markov chains and provide a proof of the Perron-Frobenius theorem, which is necessary to justify the cases when a unique invariant probability distribution exists for Markov chains. We then introduce reducibility and periodicity, two important classifications for Markov chains. We use these classifications to prove properties about return times to states and matrix representations about transient states. Throughout our discussion, we introduce two common applications of Markov chains: random walk and the Gambler's ruin problem.

Contents

1. Introductory Definitions

We begin our discussion of Markov chains by defining some basic properties about stochastic processes.

Definition 1.1. Let X_t be a random variable at time t. A **stochastic process** is a collection of such random variables $\{X_{t_1}, X_{t_2}, \ldots\} = \{X_{t_j}\}_{j\in\mathbb{Z}}$.

Definition 1.2. Supppose $\{X_{t_j}\}_{j\in\mathbb{Z}}$ is a stochastic process. A stochastic process is of **discrete time** if $t_j \in \{0, 1, 2, 3, \ldots\}$ for any X_{t_j}.

Definition 1.3. A **state** is a value that a random variable X_t takes on in a stochastic process. The **state space** S of a stochastic process is the set of all possible values that X_t can take on.

Conventionally, we let the state space $S = \{0, 1, \ldots, N\}$ such that each element in S represents a different possible outcome. Our choice of integers is arbitrary and is typically done by convenience, as future examples will illustrate.

To find the joint probability of a stochastic process, we first allow $(i_0, \ldots, i_n)$ to be a finite sequence of states from S. Consider $\mathbb{P}\{X_0 = i_0, \ldots, X_n = i_n\}$ for every possible sequence $(i_0, \ldots, i_n)$. Furthermore, suppose the initial probability distribution for a stochastic process is given by $\phi_0(i) = \mathbb{P}\{X_0 = i\}$, and a transition probability for a stochastic process is given by $q_n(i_n \mid i_{n-1}, \ldots, i_0) = \mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$. Then:

$$\begin{aligned}
&\mathbb{P}\{X_0 = i_0, \ldots, X_n = i_n\} \\
&= \mathbb{P}\{X_0 = i_0\} \cdot \mathbb{P}\{X_1 = i_1 \mid X_0 = i_0\} \cdot \ldots \cdot \mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} \\
&= \phi_0(i_0) q_1(i_1 \mid i_0) \cdot \ldots \cdot q_n(i_n \mid i_{n-1}, \ldots, i_0) \quad \text{for } i_0 = 0, 1, \ldots, N.
\end{aligned}$$

We are now ready to introduce the Markov property for stochastic processes, which simplifies this expression further.

Definition 1.4. Suppose $\{X_0, X_1, \ldots\}$ is a stochastic process. The **Markov property** states that the probability of a future state is dependent only on the present state. More generally, $\mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}\}$. A discrete-time stochastic process with the Markov property is known as a discrete-time **Markov chain**.

Definition 1.5. Suppose $\{X_0, X_1, \ldots\}$ is a discrete-time Markov chain. $\{X_0, X_1, \ldots\}$ is **time-homogeneous** if $\mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}\} = p(i_{n-1}, i_n)$ for a function $p : [S \times S] \to (0, 1)$. We can generalize this expression to evaluate probabilities for states k time periods ("steps") from the present as $\mathbb{P}\{X_n = i_n \mid X_{n-k} = i_{n-k}\} = p_k(i_{n-k}, i_n)$ for $0 \leq k \leq n$ in discrete time.

For $k = 1$, or a single time period (step) in the future, we typically drop the subscript and write $p_1(i_{n-k}, i_n)$ as $p(i_{n-k}, i_n)$.

There is an important implication of time-homogeneity. Definition 1.5 indicates that only the *difference* in time periods between the present and future, and the states in the present and the future, are relevant when computing a transition probability. The *specific* time periods in which the present and future states occur do not affect the probability. That is:

$$\mathbb{P}\{X_n = i_n \mid X_{n-k} = i_{n-k}\} = \mathbb{P}\{X_{n+c} = i_n \mid X_{n-k+c} = i_{n-k}\}$$

for $c \geq k - n$ in discrete time. We will frequently use X_0 to represent the present, so we often allow $c = k - n$ such that

$$\mathbb{P}\{X_n = i_n \mid X_{n-k} = i_{n-k}\} = \mathbb{P}\{X_k = i_n \mid X_0 = i_{n-k}\}.$$

However, we can index the random variables however we would like depending on their context in larger expressions.

Furthermore, for single step transition probabilities in Markov chains, we have that $q_n(i_n \mid i_{n-1}, \ldots, i_0) = p(i_{n-1}, i_n)$. Thus, the joint probability for a Markov chain becomes

$$\mathbb{P}\{X_0 = i_0, \ldots, X_n = i_n\} = \phi_0(i_0) \cdot p(i_0, i_1) \cdot p(i_1, i_2) \ldots \cdot p(i_{n-1}, i_n).$$

We now consider a more convenient way of expressing all possible transition probabilities for a Markov chain given a number of steps between present and future and a state space S.

Definition 1.6. Suppose $S = \{0,\ 1,\ \ldots,\ \mathrm{N}\}$. The **transition matrix P** for a single step of a Markov chain is given by:

$$\mathbf{P} = \begin{pmatrix} p(0,0) & p(0,1) & \cdots & p(0,N) \\ p(1,0) & p(1,1) & \cdots & p(1,N) \\ \vdots & \vdots & \ddots & \vdots \\ p(N,0) & p(N,1) & \cdots & p(N,N) \end{pmatrix}$$

where the entries P_{ij} are subject to two conditions:

(1) $0 \leq \mathbf{P}_{ij} \leq 1$ for $0 \leq i, j \leq N$
(2) $\sum_{j=0}^{N} \mathbf{P}_{ij} = 1$ for $0 \leq i \leq N$

Note that we are "zero-indexing" such that the first row and first column of $\mathbf{P}$ are denoted by $i = 0$ and $j = 0$ respectively. The first property follows directly from the fact that the probability of a random variable must be a value between 0 and 1. The second property follows from the fact that:

$$\sum_{X_n \in S} P\{X_n \mid X_{n-1} = i\} = 1$$

as the sum of all possible conditional probabilities from the sample space (state space) S for a fixed value of X_{n-1} must equal 1.

A very common application of basic Markov chains is random walk. Consider a collection of points such that each point is connected to at least one other point by an edge. A person is at one of these points, and they randomly take a step onto an adjacent point by traveling along its edge. This process is a Markov chain and a simplified model of random walk. For example, consider a basic graph equally incremented from 0 to N. Suppose with every step, the walker moves to the right with probability p. If the walker is not on either of the extreme points at 0 and N, then $p(i, i+1) = p$ and $p(i, i-1) = 1 - p$ for $0 < i < N$. If the walker is on either extreme point at 0 and N, then we can consider a *reflecting boundary* such that the walker steps in the direction of the middle of the interval with probability 1. That is, $p(0, 1) = p(N, N-1) = 1$. We can create a transition matrix for single steps by the walker:

$$\mathbf{P} = \begin{pmatrix} p(0,0) & p(0,1) & p(0,2) & \cdots & p(0,N) \\ p(1,0) & p(1,1) & p(1,2) & \cdots & p(1,N) \\ p(2,0) & p(2,1) & p(2,2) & \cdots & p(2,N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p(N,0) & p(N,1) & p(N,2) & \cdots & p(N,N) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1-p & 0 & p & \cdots & 0 \\ 0 & 1-p & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

More complicated cases of random walk include a stock price model, where a ticker can move 1 unit up or 1 unit down with given probability and a constant time interval. Additionally, finitely connected random walk models may instead use *absorbing boundaries* at the endpoints such that if the chain reaches the state 0 or N, the process forever stays in that state. In a sense, the walker becomes trapped (or absorbed) at these points.

Because we have also defined transition probabilities for states occurring $k > 1$ steps apart from one another, we need to likewise describe a matrix that consolidates all possible transition probabilities for a state space S given a time interval (or number of steps).

Proposition 1.7. *Suppose* $\mathbf{P}$ *is a transition matrix for a Markov chain with state space* $S = \{0, 1, \ldots, N\}$*. The* k*-step transition matrix is given by:*

$$\mathbf{P}^k = \begin{pmatrix} p(0,0) & p(0,1) & \cdots & p(0,N) \\ p(1,0) & p(1,1) & \cdots & p(1,N) \\ \vdots & \vdots & \ddots & \vdots \\ p(N,0) & p(N,1) & \cdots & p(N,N) \end{pmatrix}^k = \begin{pmatrix} p_k(0,0) & p_k(0,1) & \cdots & p_k(0,N) \\ p_k(1,0) & p_k(1,1) & \cdots & p_k(1,N) \\ \vdots & \vdots & \ddots & \vdots \\ p_k(N,0) & p_k(N,1) & \cdots & p_k(N,N) \end{pmatrix}$$

More generally, every entry in $\mathbf{P}^k$ *is given by* $(\mathbf{P}^k)_{ij} = p_k(i,j)$

Proof. We use induction. Suppose k = 1. Then:

$$(\mathbf{P}^k)_{ij} = \mathbf{P}_{ij} = p_1(i,j) = p(i,j),$$

where we use Definition 1.6 for a single-step transition matrix. Now suppose that $(\mathbf{P}^m)_{ij} = p_m(i,j)$ for $k = m$. Then:

$$\begin{aligned}(\mathbf{P}^{m+1})_{ij} = (\mathbf{P}^m\mathbf{P})_{ij} &= \sum_{c \in S} \big[(\mathbf{P}^m)_{ic} \cdot \mathbf{P}_{cj}\big] = \sum_{c \in S} \big[p_m(i,c) \cdot p(c,j)\big] \\ &= \sum_{c \in S} \big[P\{X_m = c | X_0 = i\} \cdot P\{X_1 = j | X_0 = c\}\big].\end{aligned}$$

By time homogeneity, $P\{X_1 = j \mid X_0 = c\} = P\{X_{m+1} = j \mid X_m = c\}$. Now:

$$\begin{aligned}&\sum_{c \in S} \big[P\{X_m = c \mid X_0 = i\} \cdot P\{X_1 = j \mid X_0 = c\}\big] \\ &\qquad = \sum_{c \in S} \big[P\{X_m = c \mid X_0 = i\} \cdot P\{X_{m+1} = j \mid X_m = c\}\big] \\ &\qquad = P\{X_{m+1} = j \mid X_0 = i\} = p_{m+1}(i,j)\end{aligned}$$

which proves the induction hypothesis. □

We have only mentioned transition probabilities and the initial probability distribution for X_0. We may also be interested in finding the unconditional probability $P\{X_k = i\}$ for a random variable k steps from the present. We can use transition matrices to compute such probabilities.

Proposition 1.8. *Suppose* $\phi_k(i) = P\{X_k = i\}$ *and* $\overline{\phi}_k = (\phi_k(0), \phi_k(1), \ldots, \phi_k(N))$ *for a state space* S*. If* $\mathbf{P}$ *is a transition matrix, then* $\overline{\phi}_k = \overline{\phi}_0\mathbf{P}^k$*.*

Proof. We verify $\overline{\phi}_k = \overline{\phi}_0\mathbf{P}^k$ by direct computation:

$$\begin{aligned}(\overline{\phi}_0\mathbf{P}^k)_{ij} = \sum_{c \in S} \big[(\overline{\phi}_0)_{ic} \cdot (\mathbf{P}^k)_{cj}\big] &= \sum_{c \in S} \big[(P\{X_0 = c\} \cdot p_k(c,j)\big] \\ &= \sum_{c \in S} \big[(P\{X_0 = c\} \cdot P\{X_k = j \mid X_0 = c\}\big] \\ &= P\{X_k = j\}\end{aligned}$$

for $i = 0$ and $j = 0, 1, \ldots, N$. Then $\overline{\phi}_0\mathbf{P}^k = (P\{X_k = 0\}, \ldots, P\{X_k = N\}) = \overline{\phi}_k$.

□

2. Large-Time Behavior and Invariant Probability

The analysis of the long-term behavior of Markov chains is essential for predicting steady states in various systems, from queueing models to genetic populations. This section delves into the underlying mathematical instances of the convergence of Markov chains and the conditions under which certain probability distributions become invariant.

Definition 2.1. Invariant Probability Distribution: A probability vector π is invariant for a transition matrix $\mathbf{P}$ if it satisfies:

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

π represents the average ratio of time that the state stays in each state for sufficiently large n in $\mathbf{P}^n$.

This implies that π is a left eigenvector of $\mathbf{P}$ corresponding to the eigenvalue 1.

Theorem 2.2. ***Perron-Frobenius Theorem:*** *Suppose* $\mathbf{A}$ *is an* $n \times n$ *matrix. If every entry* $\mathbf{A}_{ij} > 0$*, then:*

(1) $\mathbf{A}$ *has a unique maximum eigenvalue* $\lambda_{\max}$*.*
(2) The eigenvector corresponding to $\lambda_{\max}$ *can be written to have all positive entries.*

Proof. Suppose $\vec{v} = (v_1, \ldots, v_n)^T$ is a column vector with $v_j \geq 0$ for $j = 1, \ldots, n$ and at least one entry $v_j > 0$. Since $\mathbf{A}_{ij} > 0$, we have $(\mathbf{A}\vec{v})_{ij} = \sum_{j=1}^{n} \mathbf{A}_{ij} v_j \geq 0$ for $i = 1, \ldots, n$ with at least one entry $(\mathbf{A}\vec{v})_{ij} > 0$. We express this fact as $\mathbf{A}\vec{v} > 0$.

Now suppose λ is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\vec{v}$. Let $g(\vec{v}) := \max\{\lambda \mid \mathbf{A}\vec{v} \geq \lambda\vec{v}\}$. We can see that $\mathbf{A}\vec{v}$ must be bounded below by $\min \sum_{j=1}^{n} \mathbf{A}_{ij} v_j$. Additionally, we can find λ small enough such that $\min \sum_{j=1}^{n} \mathbf{A}_{ij} v_j \geq \lambda v_{\max}$ for $v_{\max} = \max v_j > 0$. Then $g(\vec{v}) > 0$ for all $v \geq 0$. Furthermore, for all $c > 0$, we have that:

$$\begin{aligned} g(c\vec{v}) := \max\{\lambda \mid \mathbf{A}(c\vec{v}) \ \geq \lambda(c\vec{v})\} &= \max\{\lambda \mid c(\mathbf{A}\vec{v}) \geq c(\lambda\vec{v})\} \\ &= \max\{\lambda \mid \mathbf{A}\vec{v} \geq \lambda\vec{v}\} \end{aligned}$$

which implies that $g(c\vec{v}) = g(\vec{v})$.

We now consider the supremum of $g(\vec{v})$:

$$\alpha = \sup_{\vec{v} \geq 0, \vec{v} \neq \vec{0}} g(\vec{v})$$

to find the maximum eigenvalue for matrix A under given conditions.

If $g(\vec{v}) = \alpha$, and $A\vec{v} \geq \alpha\vec{v}$, assume for contradiction that $A\vec{v} \neq \alpha\vec{v}$. Let $\vec{w} = A\vec{v} - \alpha\vec{v}$, then $\vec{w} \geq \vec{0}$ and $\vec{w} \neq \vec{0}$. Thus $A\vec{w} = A^2\vec{v} - \alpha A\vec{v} \geq \alpha\vec{w}$, which implies $\vec{w}$ is an eigenvalue vector corresponding to α, contradicting the uniqueness unless $\vec{w} = \vec{0}$. Therefore, $A\vec{v} = \alpha\vec{v}$.

If $\mathbf{A}\vec{v} = \alpha\vec{v}$, then $\vec{v} = (\frac{1}{\alpha})\mathbf{A}\vec{v}$. Since $\mathbf{A}\vec{v} > 0$ and $\alpha = g(\vec{v}) > 0$, it follows that $\vec{v} > 0$. Now consider $\alpha = \sup_{\|\vec{v}\|=1} g(\vec{v})$. Suppose for contradiction there exist vectors $\vec{v}_1$ and $\vec{v}_2$ such that $\vec{v}_1 \neq c\vec{v}_2$ for $c \in \mathbb{R}$ and

$$g(\vec{v}_1) = g(\vec{v}_2) = \alpha \quad \text{and} \quad \sum_{i=1}^{n} v_1^i = \sum_{i=1}^{n} v_2^i = 1$$

By our previous result, vectors $\vec{v}_1, \vec{v}_2 > 0$. Then for $c \in \mathbb{R}$, the following holds:

$$\mathbf{A}\vec{v}_1 = \alpha\vec{v}_1 \quad \text{and} \quad \mathbf{A}\vec{v}_2 = \alpha\vec{v}_2 \Rightarrow c(\mathbf{A}\vec{v}_2) = c(\alpha\vec{v}_2) \Rightarrow \mathbf{A}(c\vec{v}_2) = \alpha(c\vec{v}_2)$$

We add these equations:

$$\mathbf{A}\vec{v}_1 + \mathbf{A}(c\vec{v}_2) = \alpha\vec{v}_1 + \alpha(c\vec{v}_2) \Rightarrow \mathbf{A}(\vec{v}_1 + c\vec{v}_2) = \alpha(\vec{v}_1 + c\vec{v}_2)$$

If $\vec{v}_1 = (v_1^1, v_1^2, \ldots, v_1^j, \ldots, v_1^n)^T$ and $\vec{v}_2 = (v_2^1, v_2^2, \ldots, v_2^j, \ldots, v_2^n)^T$, then there exists some $c = -v_1^j/v_2^j$ such that $\vec{v}_1 + c\vec{v}_2$ has a 0 entry and $\vec{v}_1 + c\vec{v}_2 \geq 0$, but $\vec{v}_1 + c\vec{v}_2 \neq 0$. However, we also see that

$$\mathbf{A}\vec{v} > 0 \Rightarrow \mathbf{A}(\vec{v}_1 + c\vec{v}_2) = \alpha(\vec{v}_1 + c\vec{v}_2) > 0.$$

Because $\alpha > 0$, the RHS holds only if $\vec{v}_1 + c\vec{v}_2 > 0$. This is a contradiction, since we've shown that we can always find $c \in \mathbb{R}$ such that $\vec{v}_1 + c\vec{v}_2$ has a 0 entry and $\vec{v}_1 + c\vec{v}_2 \not> 0$. So, the eigenvector $\vec{v}$ is unique.

Now suppose that λ is any other eigenvalue of A such that $\lambda \neq \alpha$. If $\vec{u}$ is the eigenvector corresponding to λ, then $\mathbf{A}\vec{u} = \lambda\vec{u}$. If $|A|$ and $|u|$ are given by taking the absolute value of every entry of $\mathbf{A}$ and $\vec{u}$, respectively, then

$$|A||u| \geq |Au| = |\lambda u| = |\lambda||u|.$$

Since $\mathbf{A} > 0$, we have $|A| = \mathbf{A}$. This fact implies that $|A||u| = \mathbf{A}|u| \geq |\lambda||u|$. Furthermore, $|u| \geq 0$. Then λ is an element of the set defined by the function g, which implies that $|\lambda| \leq \sup g = \alpha$. However, since $|\lambda| \neq |\alpha| = \alpha$, equality cannot hold, so $|\lambda| < \alpha$. $\square$

Proposition 2.3. *If* $\mathbf{A}$ *is a stochastic matrix then its maximal eigenvalue is 1*

Proof. We first show that 1 is an eigenvalue of A

Consider the vector $\mathbf{v} = [1, 1, \ldots, 1]^T$, a column vector with all entries equal to 1. Now, multiply A by $\mathbf{v}$:

$$A\mathbf{v} = A \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n A_{1j} \\ \sum_{j=1}^n A_{2j} \\ \vdots \\ \sum_{j=1}^n A_{nj} \end{bmatrix}$$

Since each row of A sums to 1 by the definition of a stochastic matrix, we have:

$$A\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{v}$$

This shows that:

$$A\mathbf{v} = 1 \cdot \mathbf{v}$$

Thus, 1 is an eigenvalue of A.

We then show that 1 is the maximal eigenvalue.

We apply the *Perron-Frobenius theorem*, which states the following for any square matrix with positive entries:

(1) There exists a real eigenvalue $\lambda_{\max}$ such that $|\lambda_{\max}|$ is greater than or equal to the absolute value of any other eigenvalue.

Because eigenvalues correspond to scaling factors of the eigenvectors, and the sum of all rows of A equals to 1, there cannot be an eigenvalue with a magnitude greater than 1. Therefore, because the Perron-Frobenius theorem guarantees a unique maximal eigenvalue, and we know that 1 is indeed an eigenvalue of A, we conclude that:

$$\lambda_{\max} = 1$$

□

Theorem 2.4. *For any Markov chain with a finite state space and positive stochastic matrix, there exists a unique invariant probability distribution.*

Proof. From Proposition 4.3 above, we know for stochastic matrix A

$$\lambda_{\max} = 1$$

and from Theorem 4.2 (2), we know that the left eigenvector corresponding to the maximal eigenvalue can be written to have all positive entries, which gives us our invariant probability distribution π.

□

Consider the stochastic matrix

$$\mathbf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}.$$

We need to find a probability vector $\pi = [\pi_1, \pi_2]$ that satisfies the invariant distribution condition:

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

This translates into the system of equations:

$$\pi_1 = 0.9\pi_1 + 0.5\pi_2, \quad \pi_2 = 0.1\pi_1 + 0.5\pi_2.$$

Additionally, the probability vector must satisfy the normalization condition:

$$\pi_1 + \pi_2 = 1.$$

First, rearrange the first equation to solve for π_1:

$$\pi_1 - 0.9\pi_1 = 0.5\pi_2 \implies 0.1\pi_1 = 0.5\pi_2 \implies \pi_2 = 0.2\pi_1.$$

Substituting π_2 into the normalization condition gives:

$$\pi_1 + 0.2\pi_1 = 1 \implies 1.2\pi_1 = 1 \implies \pi_1 = \frac{5}{6}.$$

Then,

$$\pi_2 = 0.2 \times \frac{5}{6} = \frac{1}{6}.$$

Thus, the invariant probability vector is

$$\pi = \left[\frac{5}{6}, \frac{1}{6}\right].$$

This vector π satisfies both the equation $\pi = \pi\mathbf{P}$ and the normalization condition, confirming that it is the invariant distribution for the given stochastic matrix.

To validate, substituting π back into the equation $\pi = \pi\mathbf{P}$:

$$\left[\frac{5}{6}, \frac{1}{6}\right] \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = \left[0.9 \times \frac{5}{6} + 0.5 \times \frac{1}{6}, 0.1 \times \frac{5}{6} + 0.5 \times \frac{1}{6}\right] = \left[\frac{5}{6}, \frac{1}{6}\right].$$

The results confirm that π is indeed invariant under $\mathbf{P}$.

3. Reducible and Periodic Classifications

Definition 3.1. Communication Classes: Two states i and j of a Markov chain communicate with each other, written $i \leftrightarrow j$, if there exist $m, n \geq 0$ such that $p_m(i,j) > 0$ and $p_n(j,i) > 0$.

This relation is an equivalence relation on the state space, i.e., it is:

- Reflexive: $i \leftrightarrow i$ (since $p_0(i,i) = 1 > 0$).
- Symmetric: $i \leftrightarrow j$ implies $j \leftrightarrow i$.
- Transitive: $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$.

Definition 3.2. Irreducibility: A Markov chain is irreducible if there is only one communication class, i.e., if for all i, j there exists an n such that $p_n(i,j) > 0$.

Consider the transition matrix $\mathbf{P}$ for a Markov chain with states $\{1,2,3,4\}$:

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

In this Markov chain, states $\{1,2\}$ communicate with each other and form one communication class, while states $\{3,4\}$ communicate with each other and form another communication class. However, there is no communication between states in the set $\{1,2\}$ and states in the set $\{3,4\}$. Therefore, the Markov chain is reducible, with two distinct communication classes: $\{1,2\}$ and $\{3,4\}$.

Definition 3.3. Periodicity: The period of a state i, denoted $d(i)$, is defined as the greatest common divisor of the set:

$$J_i := \{n \geq 0 : p_n(i,i) > 0\}.$$

Proposition 3.4. *If $\mathbf{P}$ is irreducible, then all states have the same period d. If $d = 1$, the chain is aperiodic.*

Proof. Let $\mathbf{P}$ be the transition matrix of an irreducible Markov chain. Recall that the period of a state i is defined as:

$$d_i = \gcd\{n \geq 1 : \mathbf{P}^n(i,i) > 0\}$$

where $\mathbf{P}^n(i,i)$ is the probability of returning to state i in exactly n steps. The period d of the Markov chain is the greatest common divisor of all the steps needed to return to any state i.

Since $\mathbf{P}$ is irreducible, for any two states i and j, there exists an integer m such that $\mathbf{P}^m(i,j) > 0$. This means it is possible to get from state i to state j in m steps.

Consider two states i and j. Let d_i be the period of state i and d_j be the period of state j. Since $\mathbf{P}$ is irreducible, there exist positive integers m and n such that:

$$\mathbf{P}^m(i,j) > 0$$

$$\mathbf{P}^n(j,i) > 0$$

This implies that it is possible to go from state i to state j in m steps and from state j to state i in n steps. Hence, it is possible to return to state i from itself in $m+n$ steps by visiting state j in between. Therefore,

$$\mathbf{P}^{m+n}(i,i) > 0$$

The existence of such paths implies that $m+n$ is a multiple of both d_i and d_j. Consequently, d_i and d_j must share a common divisor.

Since this argument holds for any pair of states i and j, all states in the irreducible Markov chain must share the same period. Denote this common period by d.

Now, if $d = 1$, the chain is aperiodic. This means that the greatest common divisor of the lengths of the cycles that return to any state is 1. In other words, it is possible to return to any state at arbitrary times, indicating that the chain does not have a fixed cycle length, which is the definition of aperiodicity.

Therefore, we have shown that if $\mathbf{P}$ is irreducible, all states have the same period d, and if $d = 1$, the chain is aperiodic.

□

Theorem 3.5. *(Irreducible, Aperiodic Chains) If $\mathbf{P}$ is the transition matrix for an irreducible, aperiodic Markov chain, then there exists a unique invariant probability vector π satisfying:*

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

If ϕ is any initial probability vector,

$$\lim_{n\to\infty} \phi\mathbf{P}^n = \pi.$$

Moreover, $\pi(i) > 0$ for each i.

Consider an example transition matrix $\mathbf{P}$

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

we can use computing technologies to calculate that the invariant probability vector π is $\left(\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}\right)$.

This section highlights the crucial role of state classification in understanding the dynamics and long term behavior of Markov chains. The classification not only aids in theoretical analysis but also impacts the practical implementation of these models in various fields such as economics, genetics, and computer science.

4. Return Times

Our discussion of communication classes, reducible Markov chains, and periodicity readily lends itself to a closer examination of chains that begin and "return" to a specified state.

Definition 4.1. Suppose $\{X_0, X_1, \ldots\}$ is an irreducible Markov chain with state space S. The **return time** T_k of a chain starting in state k is the time period when the chain first revisits state k and is given by $T_k = \min\{n \geq 1 \mid X_n = k\}$.

Because we assume the Markov chain is irreducible, there is only one communication class that contains all states within the state space S. We suspect that if $P\{T_k < \infty\} = 1$, or the return time for any state $k \in S$ is finite, then state k is recurrent. To more rigorously develop this idea, we need to introduce the idea of stopping time.

Definition 4.2. Suppose $\{X_0, X_1, \ldots\}$ is Markov chain with state space S. A **stopping time** is a random variable $\tau : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ if the event $\{\tau = n\}$ depends only on the states of the chain $\{X_0, X_1, \ldots, X_n\}$.

Definition 4.2 reveals an important property of return times. The event that the first return time to state k is n steps can be expressed as

$$\{T_k = n\} = \{X_1 \neq k, X_2 \neq k, \ldots, X_{n-1} \neq k, X_n = k\}$$

which satisfies the criterion for a stopping time. Furthermore, we may consider the jth return time as the jth revisit to state k. Given the first return time T_k^1, the jth return time T_k^j to state k is defined recursively as

$$T_k^{j+1} = \min\{n > T_k^j \mid X_n = k\}$$

which must also be a stopping time.

Stopping times also provide a further extension of the Markov property.

Proposition 4.3. ***Strong Markov Property:*** *Suppose τ is a stopping time of the Markov chain $\{X_0, X_1, \ldots\}$ with state space S. Given $\tau < \infty$ and $X_\tau = i$, the state of $X_{\tau+n}$ for $n \geq 0$ is independent of the states of $X_0, \ldots, X_{\tau-1}$.*

Proof. Suppose B is an event determined by $X_0, \ldots, X_{\tau-1}$. Then $B \cap \{\tau = m\}$ is determined by $X_0, \ldots, X_m$, and

$$P\{(X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n), B, (\tau = m), (X_\tau = i)\} \tag{4.4}$$

$$= P\{(X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n), B \;\mid \tau = m, X_\tau = i\} \cdot P\{\tau = m, X_\tau = i\} \tag{4.5}$$

However, we can also express (4.4) slightly differently:

$$\begin{aligned}
&P\{(X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n), B, (\tau = m), (X_\tau = i)\} \\
&= P\{X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n \mid B, \tau = m, X_\tau = i\} \\
&\quad \cdot P\{B, \tau = m, X_\tau = i\} \\
&= P\{X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n \mid B, \tau = m, X_0 = i\} \\
&\quad \cdot P\{B, \tau = m, X_\tau = i\} \\
&= P\{X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n \mid B, \tau = m, X_0 = i\} \qquad (4.6) \\
&\quad \cdot P\{B \mid \tau = m, X_\tau = i\} \cdot P\{\tau = m, X_\tau = i\}
\end{aligned}$$

where we used the Markov property to "shift" the chain from beginning at time τ to time 0. We now have two equivalent expressions (4.5) and (4.6). Then:

$$\begin{aligned}
&P\{(X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n), B \ \mid \tau = m, X_\tau = i\} \\
&\qquad \cdot P\{\tau = m, X_\tau = i\} \\
&= P\{X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n \mid B, \tau = m, X_0 = i\} \\
&\qquad \cdot P\{B \mid \tau = m, X_\tau = i\} \cdot P\{\tau = m, X_\tau = i\} \\
&\Rightarrow P\{(X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n), B \ \mid \tau = m, X_\tau = i\} \\
&\qquad = P\{X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n \mid B, \tau = m, X_0 = i\} \\
&\qquad\qquad \cdot P\{B \mid \tau = m, X_\tau = i\}
\end{aligned}$$

□

We can likewise relate transition probabilities to the strong Markov property. Proposition 4.3 implies that

$$P\{X_{\tau+n} - X_\tau = j \mid X_\tau = i\} = P\{X_n = i + j \mid X_0 = i\} \tag{4.7}$$

if $i + j \in S$. In other words, the Markov chain starting at time τ has the same transition probabilities as the original Markov chain starting at time 0.

We return to our discussion of the jth return time for $j > 1$. We are often interested in the time interval between two consecutive return times, which describes the number of steps to revisit state k after the $(j-1)$th return. Suppose we define this value as

$$S_k^j = \begin{cases} T_k^j - T_k^{j-1} & \text{if } \ T_k^{j-1} < \infty \\ 0 & \text{otherwise} \end{cases}$$

Lemma 4.8. *If the (j-1)th return time $T_k^{j-1} < \infty$ for all $j = 2, 3, \ldots$, then:*

(1) S_k^j is independent of the event $\{X_m = t \mid m \leq T_k^{j-1}\}$
(2) $P\{S_k^j = n \mid T_k^{j-1} < \infty\} = P\{T_k = n\}$

Proof. We will prove (1) first. Since the the *(j-1)*th return to state k given by T_k^{j-1} is a stopping time, suppose $\tau = T_k^{j-1}$. By the definition of a return time, $X_\tau = k$. Then by the Strong Markov Property, the event $\{X_{\tau+n} = k \mid \tau < \infty, X_\tau = k\}$ for $n \geq 0$ is independent of $X_0, X_1, \ldots, X_{\tau-1}$, and likewise S_k^j is independent of $X_0, X_1, \ldots, X_{\tau-1}$.

To prove (2), we realize that $S_k^j = \min\{n \geq 1 \mid X_{\tau+n} = k\}$. By (1) and (6.7), the event that $\{T_k^j - T_k^{j-1} = n \mid T_k^{j-1} < \infty\}$ has the same distribution as the first return time to state k, or the event $\{T_k = n \mid X_0 = k\}$. Then $P\{T_k^j - T_k^{j-1} = n \mid T_k^{j-1} < \infty\} = P\{S_k^j = n \mid T_k^{j-1} < \infty\} = P\{T_k = n\}$. □

We are almost ready to prove that if the return time for a state is finite, then the state is recurrent. Suppose we define a variable V_k as the number of visits to state k in a Markov chain:

$$V_k = \sum_{n=1}^{\infty} I\{X_n = k\}$$

where I is a random "indicator" variable such that $I = 1$ if $X_n = k$ and $I = 0$ if $X_n \neq k$ for $k \in S$. Then:

$$\begin{aligned}\mathbb{E}[V_k] = \mathbb{E}\left[\sum_{n=1}^{\infty} I\{X_n = k\}\right] &= \sum_{n=1}^{\infty} \mathbb{E}[I\{X_n = k\}] \\ &= \sum_{n=1}^{\infty} p_n(k,k) = \sum_{n=1}^{\infty} P\{X_n = k \mid X_0 = k\}\end{aligned}$$

Lemma 4.9. *Suppose for a Markov chain, $f_k = P\{T_k < \infty\}$. Then $(f_k)^r = P\{V_k > r\}$ for $r = 0, 1, 2, \ldots$.*

Proof. We proceed by induction. For the base case $r = 0$, it follows by definition that $P\{V_k > 0\} = 1 = (f_k)^0$. Now suppose that$(f_k)^m = P\{V_k > m\}$ for $r = m$. Also, note that if $X_0 = k$, then the event $\{V_k > m\}$ occurs if and only if the mth return time is finite, or $\{T_k^m < \infty\}$. By the inductive step:

$$\begin{aligned}P\{V_k > m+1\} = P\{T_k^{m+1} < \infty\} &= P\{T_k^{m+1} < \infty, S_k^{m+1} < \infty\} \\ &= P\{S_k^{m+1} < \infty \mid T_k^m < \infty\} \cdot P\{T_k^m < \infty\}\end{aligned}$$

By Lemma 4.8, $P\{S_k^{m+1} < \infty \mid T_k^m < \infty\} = P\{T_k < \infty\}$. Then:

$$\begin{aligned}P\{S_k^{m+1} < \infty \mid T_k^m < \infty\} \cdot P\{T_k^m < \infty\} &= P\{T_k < \infty\} \cdot P\{T_k^m < \infty\} \\ &= f_k \cdot (f_k)^m = (f_k)^{m+1}\end{aligned}$$

□

We are finally ready to classify whether a state is recurrent or transient based on return times.

Theorem 4.10. *Suppose $\{X_0, X_1, \ldots\}$ is a Markov chain with state space S and return time T_k for $k \in S$.*

(1) If $f_k = P\{T_k < \infty\} < 1$, then state k is transient and $\sum_{n=0}^{\infty} p_n(k,k) < \infty$
(2) If $f_k = P\{T_k < \infty\} = 1$, then state k is recurrent and $\sum_{n=0}^{\infty} p_n(k,k) = \infty$

Proof. For (1), we have that $\sum_{n=0}^{\infty} p_n(k,k) = \mathbb{E}[V_k] = \sum_{r=0}^{\infty} P\{V_k > r\}$. By Lemma 4.9,

$$\sum_{r=0}^{\infty} P\{V_k > r\} = \sum_{r=0}^{\infty} (f_k)^r = \frac{1}{1 - f_k} < \infty$$

since $0 \leq f_k < 1$. Then because the chain visits state k only finitely often—and eventually leaves the transition class containing k—state k must be transient.

For (2), we again apply Lemma 4.9. If $P\{T_k < \infty\} = f_k = 1$, then $(f_k)^r = 1 = P\{V_k > r\}$ for $r = 0, 1, 2, \ldots$. Then:

$$\lim_{r \to \infty} P\{V_k > r\} = P\{V_k = \infty\} = 1$$

In other words, the probability that there is an infinite number of visits in the Markov chain to state k is 1, and therefore state k must be recurrent. When a chain enters a recurrent class, it never leaves the recurrent class. Then we also have that $\sum_{n=0}^{\infty} p_n(k,k) = \infty$. □

Recall the invariant probability vector $\overline{\pi} = (\pi(0), \dots, \pi(N))$ obtained from a row of $\Pi = \lim_{n\to\infty} \mathbf{P}^n$ for a stochastic matrix satisfying the relevant properties. Any entry $\pi(k)$ for $0 \le k \le N$ represents the probability that the chain *eventually* ends in state k. Likewise, $\pi(k)$ also represents the fraction of time that the chain stays at state k for sufficiently large n.

Intuitively, each $\pi(k)$ must be related to the 1st, 2nd, ..., nth return of the chain to state k as $n \to \infty$. Suppose $S_k^1, S_k^2, \dots, S_k^n$ represents the time elapsed between the consecutive *(n-1)*th and nth returns to state k. The total time elapsed during the chain for n returns to state k is given by $R = \sum_{j=1}^n S_k^j$. Then:

$$\frac{R}{\pi(k)\cdot R} = \frac{S_k^1 + \ldots + S_k^n}{n} \approx \mathbb{E}[T_k] \Rightarrow \mathbb{E}[T_k] \approx \frac{1}{\pi(k)}$$

where we approximate the expectation by the Law of Large Numbers as $n \to \infty$.

We now rigorously prove this fact.

Theorem 4.11. *Suppose $\{X_0, X_1, \ldots\}$ is an irreducible Markov chain with state space $S = \{0, \ldots, N\}$, transition matrix $\mathbf{P}$, and invariant probability vector $\overline{\pi} = (\pi(0), \dots, \pi(N))$. Let the return time of a chain starting in state k be given by $T_k = \min\{n \ge 1 \mid X_n = k\}$. Then $\mathbb{E}[T_k] = \frac{1}{\pi(k)}$.*

Proof. Recall the indicator variable I. For fixed state k and each state $j \in S$, consider the time a chain spends in state j before returning to state k:

$$\gamma_j^k = \mathbb{E}\left[\sum_{n=0}^{T_k - 1} I\{X_n = j\}\right]$$

Now suppose we have the row vector $\vec{\gamma_k} = (\gamma_0^k, \dots, \gamma_m^k)$. We want to show that $\vec{\gamma_k}$ is an invariant probability vector such that $\vec{\gamma_k}\mathbf{P} = \vec{\gamma_k}$. Note that $\mathbb{P}\{X_0 = k\} = \mathbb{P}\{X_{T_k} = k\} = 1$ by the definition of return time. Then we can change the indices of summation and use some properties of conditional probabilities:

$$\begin{aligned}
\gamma_j^k = \mathbb{E}\left[\sum_{n=1}^{T_k} I\{X_n = j\}\right] &= \mathbb{E}\left[\sum_{n=1}^{\infty} I\{X_n = j, T_k \ge n\}\right] \\
&= \sum_{n=1}^{\infty} \mathbb{P}\{X_n = j, T_k \ge n \mid X_0 = k\} \\
&= \sum_{i\in S}\sum_{n=1}^{\infty} \mathbb{P}\{X_n = j, X_{n-1} = i, T_k \ge n \mid X_0 = k\}
\end{aligned}$$

The probability that the chain returns to state k by at least time n is equivalent to the probability that the chain does not return to state k up to time $n-1$. The probability of this event depends on the states of $X_0, X_1, \dots, X_{n-1}$, which enables us to use the Markov property. Then:

$$\begin{aligned}
&\mathbb{P}\{X_n = j, X_{n-1} = i, T_k \ge n \mid X_0 = k\} \\
&\qquad = \mathbb{P}\{X_{n-1} = i, T_k \ge n \mid X_0 = k\} \cdot \mathbb{P}\{X_n = j \mid X_{n-1} = i\}.
\end{aligned}$$

Applying this result, we have the following:

$$\begin{aligned}
\gamma_j^k &= \sum_{i\in S}\sum_{n=1}^{\infty} \mathbb{P}\{X_n = j, X_{n-1} = i, T_k \geq n \mid X_0 = k\} \\
&= \sum_{i\in S} \mathbb{P}\{X_n = j \mid X_{n-1} = i\} \cdot \mathbb{E}\left[\sum_{p=1}^{\infty} I\{X_p = i, T_k - 1 \geq p\}\right] \\
&= \sum_{i\in S} p(i,j) \cdot \mathbb{E}\left[\sum_{p=0}^{T_k-1} I\{X_p = i\}\right] = \sum_{i\in S} \mathbf{P}_{ij} \cdot \gamma_i^k = \sum_{i\in S} \gamma_i^k \cdot \mathbf{P}_{ij}
\end{aligned}$$

which written in matrix notation is $\vec{\gamma_k} = \vec{\gamma_k}\mathbf{P}$.

Lastly, we need to show that γ_j^k is positive and finite for all $j \in S$. When $j = k$, we can see that

$$\gamma_k^k = \mathbb{E}\left[\sum_{n=0}^{T_k-1} I\{X_n = k\}\right] = \mathbb{E}[I\{X_0 = k\}] = \mathbb{E}[1] = 1$$

because the chain begins in state k itself and will not return to state k until time T_k, which is not included in the terms we sum over. Since we assume the Markov chain is irreducible, there exist $s, t \geq 0$ such that for all states $j \in S$, $p_s(j,k) > 0$ and $p_t(k,j) > 0$. Since both probabilities have an upper bound of 1 by definition, then $\gamma_i^k \geq \gamma_k^k \cdot p_t(k,j) = p_t(k,j) > 0$. Recall that the irreducibility of the Markov chain also implies that the return time is finite, or $P\{T_k < \infty\} = 1$. Then $\mathbb{E}[T_k]$ must be finite, so γ_j^k must be finite. Then $0 < \gamma_j^k < \infty$ for all $j \in S$.

We are finally ready to relate γ_j^k to $\overline{\pi}$. By Theorem 2.4, the invariant probability vector $\overline{\pi}$ must be unique. Then $\vec{\gamma_k} = c\overline{\pi}$ for $c > 0$, and $\gamma_j^k = c\pi(j)$. When $j = k$:

$$\gamma_k^k = 1 = c\pi(k) \Rightarrow c = \frac{1}{\pi(k)}$$

so $\gamma_j^k = c\pi(j) = \frac{\pi(j)}{\pi(k)}$, and $\sum_{j\in S}\gamma_j^k = \frac{1}{\pi(k)}\sum_{j\in S}\pi(j)$.

Notice that

$$\begin{aligned}
\sum_{j\in S}\gamma_j^k = \sum_{j\in S}\mathbb{E}\left[\sum_{n=0}^{T_k-1} I\{X_n = j\}\right] &= \mathbb{E}\left[\sum_{j\in S}\sum_{n=0}^{T_k-1} I\{X_n = j\}\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{T_k-1}\sum_{j\in S} I\{X_n = j\}\right]
\end{aligned}$$

At each time n, the chain can only be at one state. Then $\sum_{j\in S} I\{X_n = j\} = 1$ for each n. Applying this result:

$$\sum_{j\in S}\gamma_j^k = \mathbb{E}\left[\sum_{n=0}^{T_k-1}\sum_{j\in S} I\{X_n = j\}\right] = \mathbb{E}\left[\sum_{n=0}^{T_k-1} 1\right] = \mathbb{E}[T_k]$$

Furthermore, $\sum_{j\in S}\pi(j) = 1$ because we are taking the sum over the sample (state) space. Then:

$$\sum_{j\in S}\gamma_j^k = \frac{1}{\pi(k)}\sum_{j\in S}\pi(j) \Rightarrow \mathbb{E}[T_k] = \frac{1}{\pi(k)}.$$

□

We now illustrate an example of computing the return times for states by using this relationship.

Example 4.12. Return Times for a 3-State Markov Chain. Consider a Markov chain with state space $S = \{0, 1, 2\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/8 & 2/3 & 5/24 \\ 0 & 1/6 & 5/6 \end{pmatrix}$$

We compute the large-time stochastic matrix Π by

$$\lim_{n\to\infty} \mathbf{P}^n \approx \begin{pmatrix} 0.182 & 0.364 & 0.455 \\ 0.182 & 0.364 & 0.455 \\ 0.182 & 0.364 & 0.455 \end{pmatrix}$$

The invariant probability vector is given by any of the identical rows of Π. That is, $\overline{\pi} = (\pi(0), \pi(1), \pi(2)) = (0.182, 0.364, 0.455)$. We can compute the expected return times for each state in S as follows:

$$\mathbb{E}[T_0] = \frac{1}{\pi(0)} \approx 5.495 \qquad \mathbb{E}[T_1] = \frac{1}{\pi(1)} \approx 2.747 \qquad \mathbb{E}[T_2] = \frac{1}{\pi(2)} \approx 2.198$$

5. Transient States

Definition 5.1. A state $i \in S$ is called *transient* if, starting from state i, there is a non-zero probability that the process will never return to i. Formally, state i is transient if

$$\mathbb{P}(X_n = i \text{ for some } n > 0 \mid X_0 = i) < 1.$$

Definition 5.2. State i is *recurrent* if

$$\mathbb{P}(X_n = i \text{ for some } n > 0 \mid X_0 = i) = 1.$$

Let $\mathbf{P}$ be the transition matrix of a finite Markov chain. If the chain has transient states, we can partition $\mathbf{P}$ into submatrices as follows:

$$\mathbf{P} = \begin{bmatrix} \tilde{\mathbf{P}} & \mathbf{0} \\ \mathbf{S} & \mathbf{Q} \end{bmatrix},$$

where Q is the submatrix corresponding to the transient states. The n-step transition matrix can be written as:

$$P^n = \begin{bmatrix} \tilde{\mathbf{P}}^n & \mathbf{0} \\ \mathbf{S}^n & \mathbf{Q}^n \end{bmatrix}.$$

Definition 5.3. Substochastic and Fundamental Matrix: The matrix $\mathbf{Q}$ is a substochastic matrix, meaning that it has non-negative entries and the row sums are less than or equal to 1. Since the states represented by $\mathbf{Q}$ are transient, we have $\mathbf{Q} \to 0$ as $n \to \infty$.

The fundamental matrix $\mathbf{M}$ is defined as:

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}.$$

This matrix plays an important role in determining the expected number of visits to transient states.

Proposition 5.4. *Let i be a transient state. The expected number of visits to state i starting from a transient state j is given by the (j,i) entry of the matrix $\mathbf{M}$. Formally,*

$$\mathbb{E}(Y_i \mid X_0 = j) = \mathbf{M}_{ji},$$

where Y_i is the total number of visits to state i.

Proof. Consider the total number of visits to state i:

$$Y_i = \sum_{n=0}^{\infty} I_{\{X_n = i\}}.$$

Taking the expectation, we get:

$$\mathbb{E}(Y_i \mid X_0 = j) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = i \mid X_0 = j) = \sum_{n=0}^{\infty} p^n(j,i),$$

which corresponds to the (j,i) entry of the matrix series $I + Q + Q^2 + \cdots = (I - Q)^{-1}$. □

The expected number of steps until absorption, starting from state j, can be computed using the fundamental matrix M. For a finite Markov chain with absorbing states, the transition matrix P can be written as:

$$P = \begin{bmatrix} I & 0 \\ S & Q \end{bmatrix}.$$

The expected number of steps to absorption is then given by:

$$\mathbf{t} = M\mathbf{v}_1,$$

where $\mathbf{v}_1$ is a column vector of ones.

The fundamental matrix M is a crucial tool in analyzing not only the behavior of transient states, but also behavior regarding recurrent states. The expected number of steps until the chain enters a recurrent class can also be determined using the fundamental matrix M. For an irreducible Markov chain with transient states and recurrent states, the transition matrix can be written as:

$$P = \begin{bmatrix} I & 0 \\ S & Q \end{bmatrix}.$$

Let T_i be the number of steps needed to reach state i. The expected number of steps is given by:

$$\mathbb{E}(T_i \mid X_0 = j) = \sum_{k \neq i} M_{jk}.$$

To determine the probability that the Markov chain eventually reaches a particular recurrent class, let A be the matrix with entries $\alpha(t_i, r_j)$, where $\alpha(t_i, r_j)$ is the probability that the chain starting at transient state t_i ends up in recurrent state r_j. The matrix A can be computed as:

$$A = S + QA = (I - Q)^{-1}S = MS.$$

Example 5.5. Consider the gambler's ruin problem where a gambler starts with i dollars and plays a game where at each step they win or lose one dollar with equal probability until they either go broke (reach 0 dollars) or reach a target of N dollars. The states $\{1, 2, \ldots, N-1\}$ are transient, while states 0 and N are absorbing.

The transition matrix $\mathbf{P}$ for this Markov chain can be written as

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The submatrix $\mathbf{Q}$, representing the transitions among the transient states, is given by

$$\mathbf{Q} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & \cdots & 0 \\ 0 & \frac{1}{2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{2} \end{bmatrix}.$$

The matrix $\mathbf{M}$ is then

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}.$$

We are interested in the expected number of steps until absorption, or the process where the chain enters a state from which it cannot leave, starting from state i. Let $\mathbf{t}$ be the column vector whose i-th entry t_i is the expected number of steps to absorption starting from state i. Then $\mathbf{t}$ can be computed using

$$\mathbf{t} = \mathbf{M}\mathbf{1},$$

where $\mathbf{1}$ is a column vector of ones.

For the gambler's ruin problem with $N = 5$, the submatrix $\mathbf{Q}$ and its inverse $\mathbf{M}$ are:

$$\mathbf{Q} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix},$$

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 4 & 4 & 4 \\ 2 & 4 & 6 & 6 \\ 2 & 4 & 6 & 8 \end{bmatrix}.$$

The expected number of steps to absorption starting from each transient state can be calculated as

$$\mathbf{t} = \mathbf{M}\mathbf{1} = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 4 & 4 & 4 \\ 2 & 4 & 6 & 6 \\ 2 & 4 & 6 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 14 \\ 18 \\ 20 \end{bmatrix}.$$

The Gambler's Ruin problem illustrates the significance of the application of transient state analysis in Markov chains. By constructing the transition matrix and identifying transient and absorbing states, we can simplify the computation of important quantities such as the expected number of steps until absorption and the probabilities of reaching specific absorbing states. This example encapsulates the broader applicability of Markov chain theory in modeling and solving real-world stochastic processes.

Acknowledgments

We would like to thank our mentor, Baichen Tan, for introducing us to Markov chains in an accessible and compelling way. His patience and attention to detail while explaining proofs and examples was instrumental in our approach to writing this paper. We hope this paper reflects our appreciation for both Baichen's guidance and the material. Additionally, we would like to thank Professor Peter May for organizing the 2024 REU. We are grateful for the opportunity to learn and refine the elements of writing a mathematical paper, and we look forward to further developing as mathematicians and academic writers after completing this experience.

References

[1] Gregory Lawler. Stochastic Processes, Second Edition. Taylor and Francis Group. 2006.