Applied Random Processes

Notes for MATH 466, Aniket Deshpande
As instructed by Soobin Cho and Partha Dey

"The theory of probability can be established on a completely objective basis by means of the concept of equally likely cases." – Andrey Markov

---

Contents

Preface

These notes were compiled during the Spring 2025 semester for MATH 466: Applied Random Processes at the University of Illinois at Urbana-Champaign. The course covers the fundamental theory and applications of stochastic processes, with particular emphasis on Markov chains and their applications in modeling real-world phenomena.

The material presented here builds upon basic probability theory to explore discrete-time and continuous-time Markov chains, their long-run behavior, and computational methods for analysis. Special attention is given to the Metropolis-Hastings algorithm and its role in modern computational statistics.

These notes, which closely follow *Markov Chains* by J.R. Norris, are intended as a supplement to, not a replacement for, the assigned textbook and lecture materials. Any errors or omissions are my own.

Aniket Deshpande

August 2025

1 Set Theory and Linear Algebra

Before we begin studying discrete and continuous-time Markov chains, we must develop our fundamentals in set theory and matrix theory.

1.1 Sets and Sequences

We begin with a set $S \subseteq \mathbb{R}$ and define two very important properties of sets: the *supremum* and *infimum*.

Definition 1.1.1 (Infimum and Supremum). Let $S \subseteq \mathbb{R}$ be a non-empty set. The *infimum* of S is the greatest lower bound of S, denoted by $\inf S$:

$$\inf S = \sup\{x \in \mathbb{R} : x \leq s \text{ for all } s \in S\} \tag{1}$$

The *supremum* of S is the least upper bound of S, denoted by $\sup S$:

$$\sup S = \inf\{x \in \mathbb{R} : x \geq s \text{ for all } s \in S\} \tag{2}$$

The *well-ordering principle* states that the infimum and supremum always exist in $\mathbb{R}$. We also extend $\mathbb{R}$ to the *extended reals* to include $\pm\infty$.

Definition 1.1.2 (Extended Real Numbers). We extend the real numbers as

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} \tag{3}$$

This allows us to say, for some $a, b \in \overline{\mathbb{R}}$ with $a < b$

$$\begin{aligned} &\inf[a,b] = \inf(a,b] = \inf[a,b) = \inf(a,b) = a \\ &\sup[a,b] = \sup(a,b] = \sup[a,b) = \sup(a,b) = b \\ &\sup(a,\infty) = \infty \quad \inf(-\infty,b) = -\infty \end{aligned} \tag{4}$$

Example 1.1.1. Let $\varnothing$ denote the empty set. Then $\inf\varnothing = \infty$ and $\sup\varnothing = -\infty$.

We now expand these definitions to sequences and their limits.

Definition 1.1.3 (Limit). Given a sequence $(a_n)_{n\in\mathbb{N}}$ and a constant $L \in \overline{\mathbb{R}}$

$$\lim_{n\to\infty} a_n = L \tag{5}$$

if and only if for every $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for all $n \geq N$, $|a_n - L| < \varepsilon$.

We say that $a_n \to \infty$ if and only if for every $M \in \mathbb{R}$, there exists $N(M)$ such that for all $n \geq N$, $a_n > M$.

Similarly, $a_n \to -\infty$ if and only if for every $M \in \mathbb{R}$, there exists $N(M)$ such that for all $n \geq N$, $a_n < M$.

Note that a sequence $(a_n)_{n\in\mathbb{N}}$ can diverge without approaching an infinity, for example $a_n = (-1)^n$, an *oscillating sequence*.

Theorem 1.1.1 (Monotone Convergence Theorem on $\mathbb{R}$). Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. If $(a_n)_{n\in\mathbb{N}}$ is bounded and monotonically increasing, then

$$\lim_{n\to\infty} a_n = \sup\{a_n : n \in \mathbb{N}\} \tag{6}$$

If $(a_n)_{n\in\mathbb{N}}$ is bounded and monotonically decreasing, then

$$\lim_{n\to\infty} a_n = \inf\{a_n : n \in \mathbb{N}\} \tag{7}$$

Note that the limit of a_n exists if the limit of the supremum and infimum are equivalent. Additionally, we can extend this definition of limits to the sets and sequences of them.

Given a sequence of sets $A_n \subseteq \mathbb{R}$, $n \in \mathbb{N}$, we define the limit of the unions and intersections as

$$\begin{aligned} \bigcup_{n=1}^{\infty} A_n &= \{x : x \in A_n \exists\, n \in \mathbb{N}\} \\ \bigcap_{n=1}^{\infty} A_n &= \{x : x \in A_n \forall\, n \in \mathbb{N}\} \end{aligned} \tag{8}$$

We can think of unions as finding a "least upper bound" and intersections as finding a "greatest lower bound", or at least develop some sort of analogy between the two. We now define limits to supremums and infimums of sets.

Definition 1.1.4. lim sup and lim inf. Given a sequence of sets $A_n \subseteq \mathbb{R}$, $n \in \mathbb{N}$, we define the *limit superior* and *limit inferior* as

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k \qquad \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k \tag{9}$$

This can be a struggle to gain some intuition on, so we will provide some examples.

Example 1.1.2. Consider the sequence of sets $(A_n)_{n\in\mathbb{N}}$ where $A_n = [\frac{1}{n}, 1]$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty}[0,1] = (0,1] \\ \liminf_{n\to\infty} A_n &= \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty}\left[\frac{1}{n},1\right] = \{1\} \end{aligned} \tag{10}$$

Example 1.1.3. Consider the sequence of sets $(A_n)_{n\in\mathbb{N}}$ where $A_n = [0, 1-(-1)^n]$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty}[0,2] = [0,2] \\ \liminf_{n\to\infty} A_n &= \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty}[0,1] = \{0\} \end{aligned} \tag{11}$$

Definition 1.1.5 (Power Set). Given a set E, the *power set* of E, denoted by 2^E or $\mathscr{P}(E)$, is the set of all subsets of E.

$$2^E := \{A : A \subseteq E\} \tag{12}$$

Note that the notation 2^E is used to denote the cardinality of the power set of E. When $|E| < \infty$, there is a surjective correspondence between 2^E and $\{0,1\}^{|E|}$.

Thus, the cardinality of the power set of E is $2^{|E|}$, which is a very important result.

1.2 Fubini's Theorem

Theorem 1.2.1 (Fubini's Theorem). Given countable sets I and J and terms $a_{ij} \geq 0$, then

$$\sum_{i \in I} \sum_{j \in J} a_{ij} = \sum_{j \in J} \sum_{i \in I} a_{ij} \tag{13}$$

In other words, we can rearrange the summation of non-negative terms.

Proof. We can prove this by considering the double sum as a sum over the set $I \times J$. Then, we can rearrange the terms in the sum to obtain the desired result. □

1.3 Linear Algebra

We begin with a formal defintion of a *matrix*.

Definition 1.3.1 (Matrix). A *matrix* is an array of numbers indexed by two countable sets. With the index sets I, J, we denote a matrix A as

$$A = (a_{ij})_{i \in I, j \in J} \tag{14}$$

Matrix multiplication can be understood with the following summation

$$(AB)_{ij} = \sum_{k \in K} a_{ik} b_{kj} \tag{15}$$

The kth power of a matrix is defined inductively with $A^k = A^{k-1}A$ and $A^0 = I$, the identity matrix. This leads us to the defintions of eigenvalues and eigenvectors.

Definition 1.3.2 (Eigenvalues and Eigenvectors). Given a matrix A and a vector v, we say that v is an *eigenvector* of A with *eigenvalue* $\lambda \in \mathbb{C}$ if

$$Av = \lambda v \tag{16}$$

The set of all eigenvalues of A is called the *spectrum* of A, denoted $\mathrm{spec}(A)$.

$$\mathrm{spec}(A) = \{\lambda \in \mathbb{C} : \exists\, v \neq 0 \text{ such that } Av = \lambda v\} \tag{17}$$

Note that λ is an eigenvalue if and only if $A - \lambda I$ is not invertible, this is equivalent to the determinant of $A - \lambda I$ being zero. This leads us to the characteristic polynomial of a matrix.

Definition 1.3.3 (Characteristic Polynomial). Given a matrix A, the *characteristic polynomial* of A is defined as

$$\chi_A(\lambda) = \det(A - \lambda I) \tag{18}$$

We now define the diagonalizabilty of a matrix, a process that uses eigenvalues.

Theorem 1.3.1. If a matrix A has n linearly independent eigenvectors, then A is *diagonalizable*. Then, we define S as the matrix with the eigenvectors as columns and Λ as the diagonal matrix with the eigenvalues on the diagonal. Then

$$A = S\Lambda S^{-1} \tag{19}$$

Note that powers of A can be computed with the diagonalization of A.

$$A^k = S\Lambda S^{-1}S\Lambda S^{-1}\cdots S\Lambda S^{-1} = S\Lambda^k S^{-1}$$

Thus, a power of a diagonalizable matrix is a power on its eigenvalues. We now define a theorem that allows for us to find approximations to eigenvalues when the characteristic polynomial is difficult to solve.

Theorem 1.3.2 (Gershgorin Circle Theorem). Given a matrix A with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and n discs D_i with center a_{ii} and radius $R_i = \sum_{j\neq i} |a_{ij}|$, then

$$\mathrm{spec}(A) \subseteq \bigcup_{i=1}^{n} D_i \tag{20}$$

Proof. Let $\lambda \in \mathrm{spec}(A)$ and v be the corresponding eigenvector. We wish to prove that there exists an i such that $\lambda \in D_i$, the ith disk.

We pick the largest magnitude component of v, say v_k. Thus, $|v_k| > 0$, we can assume that $v_k > 0$. Then the ith equation of $Av = \lambda v$ is

$$\lambda v_i = \sum_{j=1}^{n} a_{ij} v_j = a_{ii} v_i + \sum_{j\neq i} a_{ij} v_j \tag{21}$$

Rearranging, we have

$$|\lambda - a_{ii}| \cdot v_i = \left| \sum_{j\neq i} a_{ij} v_j \right| \leq \sum_{j\neq i} |a_{ij}||v_j| \leq \sum_{j\neq i} |a_{ij}| v_k = R_i v_k$$

Since $v_k > 0$, we have $\lambda \in B_i$. □

1.4 Some Measure Theory

Measure theory is an extension of real analysis that allows us to define the concept of a *measure* on a set. We begin with the definition of a σ-algebra.

Definition 1.4.1 (σ-algebras). Given a set E, a collection of subsets $\mathscr{E}$ is a σ-*algebra* if

1. $\varnothing \in \mathscr{E}$
2. $A \in \mathscr{E} \implies A^c \in \mathscr{E}$
3. $A_1, A_2, \ldots \in \mathscr{E} \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{E}$

Note that using conditions 2 and 3, we can show that the intersection of countable sets is also in $\mathscr{E}$.

The tuple $(E, \mathscr{E})$ is called a *measurable space*.

Given two σ-algebras $\mathscr{E}_1$ and $\mathscr{E}_2$, if $\mathscr{E}_1 \subseteq \mathscr{E}_2$, then we say that $\mathscr{E}_2$ is *finer* than $\mathscr{E}_1$ and $\mathscr{E}_1$ is *coarser* than $\mathscr{E}_2$.

Proposition 1.4.1. The power set of a set E, 2^E, is a σ-algebra.

Proof. We can show that the power set of E satisfies the three conditions of a σ-algebra. Firstly, $\varnothing \in 2^E$ since $\varnothing \subseteq E$. Secondly, if $A \in 2^E$, then $A^c \in 2^E$ since $A^c \subseteq E$. Lastly, if $A_1, A_2, \ldots \in 2^E$, then $\bigcup_{n=1}^{\infty} A_n \in 2^E$ since $\bigcup_{n=1}^{\infty} A_n \subseteq E$. □

In fact, the power set of E is the *largest* σ-algebra on E. This is known as the *trivial* σ-algebra.

Definition 1.4.2 (Smallest σ-algebra). If $\mathscr{E}$ is a collection of subsets of a set E, then the *smallest* σ-algebra containing $\mathscr{E}$ is the intersection of all σ-algebras containing $\mathscr{E}$. This is denoted by $\sigma(\mathscr{E})$.

While these objects can seem rather abstract, we can motivate this with some examples.

Example 1.4.1. Let $E = \{1, 2, 3\}$ and $\mathscr{E} = \{\{1\}\}$. Then we find the smallest σ-algebra containing $\mathscr{E}$ as

$$\sigma(\mathscr{E}) = \{\varnothing, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$$

Consider another collection of subsets, $\mathscr{B} = \{\{1\}, \{2\}\}$. Then the smallest σ-algebra containing $\mathscr{B}$ is

$$\sigma(\mathscr{B}) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\} = 2^E$$

Definition 1.4.3 (Partition). A *partition* Π of a set E is a collection of non-empty, disjoint subsets of E such that

$$\bigcup_{A \in \Pi} A = E \tag{22}$$

In other words, Π is both *exhaustive* (covers all of E) and *mutually exclusive* (disjoint).

We now use some examples to demonstrate the relationship between partitions and generated σ-algebras.

Example 1.4.2. For a set $E = \{1, 2, 3\}$ and partition $\Pi = \{\{1\}\{2, 3\}\}$, we can generate the σ-algebra $\sigma(\Pi)$ as

$$\sigma(\Pi) = \{\varnothing, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$$

Proposition 1.4.2. Given a partition $\Pi = \{E_i\}_{i\in I}$ of a set E, the σ-algebra generated by Π is

$$\sigma(\Pi) = \left\{ \bigcup_{i\in I'} E_i : I' \subseteq I \right\} \tag{23}$$

Proof. We can show that the collection of sets on the right-hand side of the equation is a σ-algebra. Firstly, $\varnothing = \bigcup_{i\in\varnothing} E_i \in \sigma(\Pi)$. Secondly, if $A = \bigcup_{i\in I'} E_i \in \sigma(\Pi)$, then $A^c = \bigcup_{i\in I\setminus I'} E_i \in \sigma(\Pi)$. Lastly, if $A_1, A_2, \ldots \in \sigma(\Pi)$, then $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \bigcup_{i\in I_n} E_i = \bigcup_{i\in\bigcup_{n=1}^{\infty} I_n} E_i \in \sigma(\Pi)$. □

2 Basics of Probability Theory

2.1 Probability and Measure

With our knowledge of measure theory, we are equipped to develop rigorous definitions in probability theory.

Definition 2.1.1 (Probability Measures). A *measure* on a measurable space $(E,\mathscr{E})$ is a function $\mu : \mathscr{E} \to [0,\infty]$ that satisfies the following properties:

1. $\mu(\emptyset) = 0$.
2. For any countable collection of pairwise disjoint sets $\{A_i\}_{i=1}^{\infty}$, we have

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{24}$$

The triple $(E,\mathscr{E},\mu)$ is called a *measure space*. If $\mu(E) = 1$, then μ is called a *probability measure*.

We now introduce some standard notation in measure-theoretic probability theory.

Definition 2.1.2 (Probability Space). A *probability space* is a triple $(\Omega,\mathscr{F},\mathbb{P})$.

1. Ω is a set of outcomes.
2. $\mathscr{F}$ is a σ-algebra of events.
3. $\mathbb{P}$ is a probability measure on $(\Omega,\mathscr{F})$.

Example 2.1.1. Let $\Omega = \{1,2,3,4,5,6\}$ be the set of outcomes of a fair six-sided die. Let $\mathscr{F} = 2^{\Omega}$ be the power set of Ω. Define $\mathbb{P} : \mathscr{F} \to [0,1]$ by

$$\mathbb{P}(A) = \frac{|A|}{6}, \tag{25}$$

where $|A|$ denotes the cardinality of A. Then $(\Omega,\mathscr{F},\mathbb{P})$ is a probability space.

Generally, given a set $\Omega = \{1,2,3,\ldots,n\}$, we can define the probability measure $\mathbb{P}$ by

$$\mathbb{P}(A) = \frac{|A|}{n}. \tag{26}$$

A probability space $(\Omega,\mathscr{F},\mathbb{P})$ has outcomes $\omega \in \Omega$ and events $A \in \mathscr{F}$.

Theorem 2.1.1 (Properties of Probability). Given a probability space $(\Omega,\mathscr{F},\mathbb{P})$

1. $\mathbb{P}(\emptyset) = 0$.
2. For some $A \in \mathscr{F}$, we have $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.
3. For any $A,B \in \mathscr{F}$, we have $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

2.2 Random Variables

Definition 2.2.1 (Random Variables). Given a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and a measurable space $(E,\mathscr{E})$, an E-valued *random variable* is a measurable function $X : \Omega \to E$.

$$X^{-1}(A) \in \mathscr{F} \text{ for all } A \in \mathscr{E}. \tag{27}$$

A special case of this definition is if E is countable, then $X : \Omega \to E$ is a $(2^E, \mathscr{F})$-measurable RV if and only if $X^{-1}(A) \in \mathscr{F}$ for all $A \subseteq E$.

Definition 2.2.2 (Probability Mass Function). The *probability mass function* (pmf) or distribution of a random variable X is a function $p_X : E \to [0,1]$ defined by

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}). \tag{28}$$

This is best described with an example.

Example 2.2.1 (Two Fair Coins.). Consider two fair coins with the set of outcomes $\Omega = \{HH, HT, TH, TT\}$. Let X be the random variable representing the number of heads in two tosses, with $E = \{0,1,2\}$. Then, X is a $(2^E, 2^\Omega)$-measurable RV. The pmf of X is

$$p_X(x) = \mathbb{P}(X = x) = \begin{cases} \frac{1}{4} & x \in \{0,2\} \\ \frac{1}{2} & x = 1 \end{cases}. \tag{29}$$

2.3 Conditional Probability

The notion of a conditional probability is the probability of an event A *given* that another event B has occurred. This is denoted by $\mathbb{P}(A|B)$.

Definition 2.3.1 (Conditional Probability). Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and events $A, B \in \mathscr{F}$ with $\mathbb{P}(B) > 0$, the *conditional probability* of A given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{30}$$

Theorem 2.3.1 (Law of Total Probability). Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and events $B_1, B_2, \dots, B_n \in \mathscr{F}$ that form a partition of Ω, we have

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \tag{31}$$

Definition 2.3.2 (Conditional Random Variables). Let Ω be a countable set and $\Pi = \{\Omega_i, i \in I\}$ is a partition of Ω. Denote $\mathscr{F} = \sigma(\Pi)$, or the σ-algebra generated by Π. Let $\mathbb{P}$ be a probability measure on the space $(\Omega, \mathscr{F})$.

Given a σ-algebra $\mathscr{B} \subseteq \mathscr{E}$ and an event $A \in \mathscr{E}$, the conditional probabilty of A given $\mathscr{B}$ is a random variable that is constant on each E_i.

$$\mathbb{P}(A|\mathscr{B})(\omega) := \mathbb{P}(A|E_i) \text{ for } \omega \in E_i. \tag{32}$$

Example 2.3.1 (Two Fair Coins). Let $\Omega = \{HH, HT, TH, TT\}$ be the set of outcomes of two fair coins with $\mathscr{F} = 2^\Omega$ and $\mathbb{P}(A) = \frac{|A|}{4}$ for any $A \in \mathscr{F}$. Let $\mathscr{B} = \{\varnothing, \{HH, HT\}, \{TT, TH\}, \Omega\}$. The partition $E_1 = \{HH, HT\}$

and $E_2 = \{TT, TH\}$ generate $\mathscr{B}$.

$$\mathbb{P}(A|\mathscr{B})(\omega) = \begin{cases} \mathbb{P}(A|\{HH,HT\}) & \omega \in \{HH,HT\} \\ \mathbb{P}(A|\{TT,TH\}) & \omega \in \{TT,TH\} \end{cases} \tag{33}$$

If $A = \{HH\}$, then

$$\mathbb{P}(A|\mathscr{B})(\omega) = \begin{cases} \frac{1}{2} & \omega \in \{HH,HT\} \\ 0 & \omega \in \{TT,TH\} \end{cases}. \tag{34}$$

2.4 Expectation

We now define *expected values* or *expectations* of random variables.

Definition 2.4.1 (Expectation). Given a random variable $X : \Omega \to I$ with $I \subset \mathbb{R}$, the *expectation* of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) = \sum_{x \in I} x\mathbb{P}(X = x). \tag{35}$$

Proposition 2.4.1. The expectation of a random variable is linear.

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]. \tag{36}$$

Proof. Let $X = \sum_{i=1}^{n} a_i X_i$. Then

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) = \sum_{\omega \in \Omega} \left(\sum_{i=1}^{n} a_i X_i(\omega)\right)\mathbb{P}(\omega) = \sum_{i=1}^{n} a_i \sum_{\omega \in \Omega} X_i(\omega)\mathbb{P}(\omega) = \sum_{i=1}^{n} a_i \mathbb{E}[X_i].$$

□

Example 2.4.1. Consider $X_1, \ldots, X_n$ as random variables taking values in the set $\{0, 1\}$. The probabilities are uniform.

$$\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = \frac{1}{2}.$$

Our set of outcomes is $\Omega = \{0, 1\}^n$, and for some event $A \in \mathscr{F}$, we have

$$\mathbb{P}(A) = \frac{|A|}{2^n}.$$

Let $Y = \sum_{i=1}^{n} X_i$. Then to *predict* Y, we have

$$\mathbb{E}[Y] = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2}.$$

Definition 2.4.2 (Independence). Two random variables are independent if and only if

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \forall\, x, y. \tag{37}$$

Note that this implies that for independent random variables X and Y, we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]. \tag{38}$$

The converse however, is *not* true. That is, if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, then X and Y are not necessarily independent.

Definition 2.4.3 (Conditional Expectation). Given a random variable $X : \Omega \to I$ and an event $B \in \mathscr{F}$ with $\mathbb{P}(B) > 0$, the *conditional expectation* of X given B is

$$\mathbb{E}(X|B) = \sum_{i \in I} i\ \mathbb{P}(X = i|B) = \frac{\sum_{i \in I} i\ \mathbb{P}(\{X = i\} \cap B)}{\mathbb{P}(B)} \tag{39}$$

Definition 2.4.4 (Conditional Expectation on σ-algebras). Given a random variable $X : \Omega \to I$ and a σ-algebra $\mathscr{B} \subseteq \mathscr{F}$ generated by a countable partition $\Pi = \{E_i, i \geq 1\}$ of Ω. The conditional expectation of X given $\mathscr{B}$ is a piecewise constant random variable.

$$\mathbb{E}(X|\mathscr{B})(\omega) = \mathbb{E}(X|E_i) \text{ for } \omega \in E_i. \tag{40}$$

Proposition 2.4.2. If $\mathscr{B} = \sigma(\varnothing)$, then $\mathbb{E}(X|\mathscr{B}) = \mathbb{E}(X)$.

Proof. Note that $\sigma(\varnothing) = \{\varnothing, \Omega\}$. Then the conditional expectation takes the form.

$$\mathbb{E}(X|\mathscr{B})(\omega) = \begin{cases} \mathbb{E}(X|\varnothing) & \omega \in \varnothing \\ \mathbb{E}(X|\Omega) & \omega \in \Omega \end{cases}. \tag{41}$$

Since $\omega \in \varnothing$ is a contradiction, we have $\mathbb{E}(X|\mathscr{B}) = \mathbb{E}(X|\Omega) = \mathbb{E}(X)$. □

Example 2.4.2. We will use the same example of $\Omega = \{0,1\}^n$ with $X_1, \ldots, X_n$ as random variables taking values in $\{0,1\}$. The probabilities are uniform.

$$\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = \frac{1}{2}.$$

Then, the partition $\Pi = \{E_1 = \{0\}, E_2 = \{1\}\}$ generates $\mathscr{B} = \sigma(\Pi)$. The conditional expectation of X_1 given $\mathscr{B}$ is

$$\mathbb{E}(X_1|\mathscr{B})(\omega) = \begin{cases} \mathbb{E}(X_1|E_1) & \omega \in E_1 \\ \mathbb{E}(X_1|E_2) & \omega \in E_2 \end{cases}.$$

If $X_1 = 1$, then

$$\mathbb{E}(X_1|\mathscr{B})(\omega) = \begin{cases} \frac{1}{2} & \omega \in E_1 \\ \frac{1}{2} & \omega \in E_2 \end{cases}.$$

Qualitatively, this is an expectation of X given whatever information $\mathscr{B}$ provides. Note that $\mathbb{P}(A|B)$ and $\mathbb{E}(A|B)$ are both scalars. But when the conditioning object is a σ-algebra, the conditional expectation is a random variable.

Proposition 2.4.3. If X is measurable with respect to $\mathscr{B}$, then $\mathbb{E}(X|\mathscr{B}) = X$.

Proof. Let X be measurable with respect to $\mathscr{B}$. Then X is constant on each E_i. Thus, $\mathbb{E}(X|E_i) = X$ for all i. Therefore, $\mathbb{E}(X|\mathscr{B}) = X$. □

Example 2.4.3. The trivial σ-algebra $\mathscr{B}_0 = \{\varnothing, \Omega\}$ has partition $\Pi = \{\Omega\}$.

$$\mathbb{E}(X|\mathscr{B}_0)(\omega) = \mathbb{E}(X|\Omega) = \mathbb{E}(X). \tag{42}$$

as $\omega \in \Omega$ for all ω and $\mathscr{B}$ provides no information. Conversely, consider the finest σ-algebra $\mathscr{B}_f = 2^\Omega$ and the partition it generates $\Pi = \{E_i = \{\omega_i\}, \omega \in \Omega\}$.

$$\mathbb{E}(X|\mathscr{B}_f) = \mathbb{E}(X|\{\omega_i\}) = \sum_i i\ \mathbb{P}(X = i|\{\omega_i\}) = X(\omega). \tag{43}$$

Therefore, $\mathbb{E}(X|\mathscr{B}_f) = X$, random variables given the most information.

Theorem 2.4.1 (Law of Total Expectation). On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, given a σ-algebra $\mathscr{B} \subseteq \mathscr{F}$ and a random variable $X : \Omega \to I$, we have

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\mathscr{B})). \tag{44}$$

Proof. Construct a partition $\Pi = \{E_j, j \geq 1\}$ of Ω that generates $\mathscr{B}$.

$$\sum_j \mathbb{E}(X|E_j)\mathbb{P}(E_j) = \sum_j \sum_i i\ \mathbb{P}(X = i|E_j)\mathbb{P}(E_j) = \sum_i i \sum_j \mathbb{P}(X = i|E_j)\mathbb{P}(E_j) = \sum_i i\ \mathbb{P}(X = i) = \mathbb{E}(X).$$

□

2.5 Filtrations

Filtrations are sequences of σ-algebras that represent the information available at each time step.

Definition 2.5.1 (Filtration). Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, a *filtration* is a sequence of increasing σ-algebras on Ω giving more/finer information.

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots \subseteq \mathscr{F}_n \subseteq \ldots \subseteq \mathscr{F}. \tag{45}$$

Example 2.5.1. Consider n random variables $X_1, \ldots, X_n$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The filtration $\mathscr{F}_n$

takes the form:

$$\mathscr{F}_n = \sigma(X_1, \ldots, X_n).$$

Making $\mathscr{F}_0 = \sigma(\varnothing)$, $\mathscr{F}_1 = \sigma(X_1)$, $\mathscr{F}_2 = \sigma(X_1, X_2)$, and so on.

2.6 Moments and the Generating Function

Before we define moments, we must define a vital measure of the spread of a random variable.

Definition 2.6.1 (Variation). The *variation* of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{46}$$

This is, qualitatively, a measure of the spread of X.

Definition 2.6.2 (Moments). For a natural number $k \in \mathbb{N}$, the k-th moment of a random variable X on a probability $(\Omega, \mathscr{F}, \mathbb{P})$ is $\mathbb{E}(X^k)$.

Definition 2.6.3 (Moment Generating Function). The *moment generating function* of a random variable X is

$$M_X(t) = \mathbb{E}[e^{tX}]. \tag{47}$$

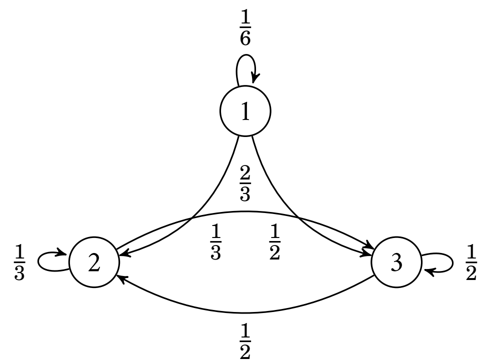
The reason this function *generates* moments is that the k-th derivative of $M_X(t)$ at $t = 0$ is the k-th moment of X.

$$M_X^{(k)}(0) = \mathbb{E}[X^k]. \tag{48}$$

3 Discrete-Time Markov Chains

3.1 The Definition

The figure below describes a *Markov Chain* with three states. The numbers in the circles represent the states, and the arrows represent the probabilities of transitioning from one state to another. For example, the probability of transitioning from state 1 to state 2 is $\frac{1}{3}$.



Note that these processes are *memoryless*, meaning that the probability of transitioning to a state depends only on the current state and not on the history of the process. This is known as the *Markov Property*.

List most things in stochastic theory, these processes can be represented using matrices.

Definition 3.1.1 (Stochastic Matrix). Let $P = (P_{ij})_{i,j=1}^{n}$ be a square matrix with non-negative entries.

1. P is *stochastic* if
$$\sum_{j=1}^{n} P_{ij} = 1 \quad \text{for all } i = 1, 2, \ldots, n \tag{49}$$
2. P is *sub-stochastic* if
$$\sum_{j=1}^{n} P_{ij} \leq 1 \quad \text{for all } i = 1, 2, \ldots, n \tag{50}$$
3. P is *doubly stochastic* if P and P^T are both stochastic.

Example 3.1.1. Let us construct the stochastic matrix for the figure shown above. There are three states, so the matrix will be 3×3. The matrix is given by

$$P = \begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Here, the ij-th entry of the matrix is the probability of transitioning from state i to state j. The rows of the matrix sum to 1, so the matrix is stochastic.

Example 3.1.2. DO THIS EXAMPLE.

Theorem 3.1.1. Given a Markov chain, the probability of transitioning from state i to state j in k steps happens with probability $q_{ij}^{(k)}$, where $q_{ij}^{(k)}$ is the ij-th entry of the matrix P^k.

Lemma 3.1.1. If P and Q are stochastic matrices, then PQ is also a stochastic matrix.

Proof. Let P and Q be stochastic matrices. Then, for all $i = 1, 2, \dots, n$,

$$\sum_{j=1}^{n} (PQ)_{ij} = \sum_{j=1}^{n} \sum_{k=1}^{n} P_{ik} Q_{kj} = \sum_{k=1}^{n} P_{ik} \sum_{j=1}^{n} Q_{kj} = 1$$

Therefore, PQ is a stochastic matrix. □

If it wasn't obvious before, we use weighted graphs to represent Markov chains. We define some basic terms from graph theory to aid us in our representations.

Definition 3.1.2 (Paths and Cycles). Given a directed graph $G = (V, E)$, a *path* is a sequence of edges $\{e_1, e_2, \dots, e_n\}$ such that the end vertex of e_i is the start vertex of e_{i+1}. A *cycle* is a path where the start and end vertices are the same.

Proposition 3.1.1. If P is a stochastic matrix, then P^k is also a stochastic matrix for all $k \in \mathbb{N}$.

Proof. Using the previous lemma, we can show that P^2 is a stochastic matrix. By induction, we can show that P^k is a stochastic matrix for all $k \in \mathbb{N}$. □

Theorem 3.1.2. Let P be a stochastic matrix. Then we have the following properties.

1. $\text{spec}(P) \subset B_1$, where B_1 is the unit ball in $\mathbb{C}$ centered at 0. In other words, the eigenvalues of P are all within the unit circle.
2. If $P_{ii} > 0$ for all $i = 1, 2, \dots, n$, then $\text{spec}(P) \subset B_1 \cup 1$. In other words, the eigenvalues of P are all within the unit circle and include 1.

Definition 3.1.3 (Diagonalization of Stochastic Matrices). Suppose we have a stochastic matrix P with spectrum $\text{spec}(P) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$. Then, P can be diagonalized as

$$UPU^{-1} = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{51}$$

for some invertible matrix U. Note that for an arbitrary power of P, we have

$$P^k = UD^kU^{-1} = U \begin{pmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{pmatrix} U^{-1} \tag{52}$$

So, the spectrum of a stochastic matrix will determine the long-time behavior of the Markov chain. With this, we continue to some formal definitions.

Definition 3.1.4 (Stochastic Process). Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, an index set I and a countable set $\mathscr{T}$, a *discrete-time stochastic process* is a collection of random variables $\{X_t : t \in \mathscr{T}\}$, where each X_t is defined by

$$X_t : (\Omega, \mathscr{F}, \mathbb{P}) \to I \tag{53}$$

for each $t \in \mathscr{T}$. I is defined to be the *state space* and $\mathscr{T}$ is the *time domain*.

Definition 3.1.5 (Markov Process). Consider a discrete-time stochastic process $(X_t)_{t\geq 0}$ defined on $(\Omega, \mathscr{F}, \mathbb{P})$. The process is said to be a *Markov process* with initial distribution λ and transition matrix P if

1. $X_0 \sim \lambda$.
2. $\mathbb{P}(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{t+1} = j | X_t = i)$, implying that the probability of transitioning to a state at time $t+1$ depends only on the state at time t.

We will use the notation $(X_t)_{t\geq 0} \sim \text{Markov}(\lambda, P)$ to denote a Markov process with initial distribution λ and transition matrix P.

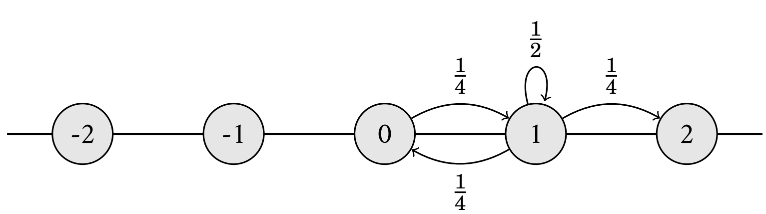
If $p_{ij}'s$ are independent of time (making P time-independent), then the Markov process is said to be *homogeneous*.

Note that a stochastic process is a function of two variables and could be notated as $X(t, \omega)$. However, we will use the standard notation of $X_t(\omega)$, as we think of the two variables differently. The first variable is the time index, and the second variable is the sample point.

Example 3.1.3 (Lazy Random Walk). Consider $\mathbb{Z}$, the set of integers and a Markov chain $(X_t)_{t\geq 0}$. The random variables take values in $\mathbb{Z}$ ($I = \mathbb{Z}$). Consider an initial distribution of $X_0 = 1$. The value of the random variable in the next step increases by 1 with probability 1/4, decreases by 1 with probability 1/4, and remains the same with probability 1/2. The elements of the transition matrix are given below.

$$p_{ij} = \begin{cases} 1/4 & \text{if } j = i+1 \\ 1/2 & \text{if } j = i \\ 1/4 & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

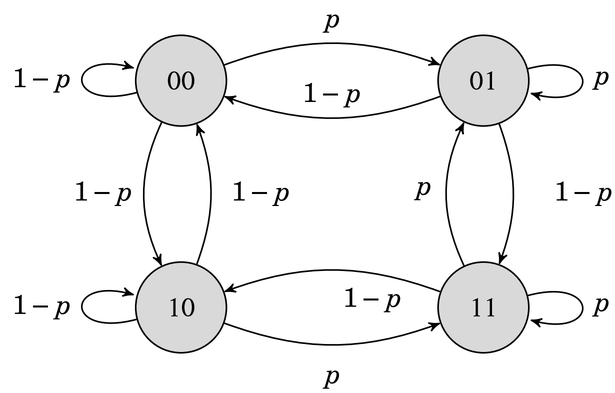
This process is called a *lazy random walk* on $\mathbb{Z}$. The term "lazy" arises from the fact that the process remains in the same state with probability 1/2.



Example 3.1.4 (Pattern Recognition). Consider a sequence $(Y_t)_{t\geq 0} \overset{\text{i.i.d.}}{\sim}$ Bernoulli(p) where $Y_0 = 1$ with probability p and $Y_0 = 0$ with probability $1-p$. Now, we define random variables X_t as follows.

$$X_0 = (Y_0, Y_1), \quad X_1 = (Y_1, Y_2), \quad X_t = (Y_t, Y_{t+1}) \text{ for } t \geq 1$$

The random variable X_t then takes values in $I = \{0,1\}^2$.



We can now construct the transition matrix for this process. The matrix is given by

$$P = \begin{pmatrix} 1-p & p & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 1-p & p \end{pmatrix}$$

3.2 Some Properties

Theorem 3.2.1. Given some initial distribution λ and transition matrix P on a countable set I, $(X_t)_{t\geq 0} \sim$ Markov(λ, P) if and only if,

$$P(X_0 = x_0, \ldots, X_n = x_n) = \lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x_n} \tag{54}$$

where $x_0, x_1, \ldots, x_n \in I$.

Proof. ($\Longrightarrow$) We begin with assuming that $(X_t)_{t\geq 0} \sim$ Markov(λ, P). We can write the joint distribution of the random variables as

$$\begin{aligned} \mathbb{P}(X_n = x_n \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) &= \frac{\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n)}{\mathbb{P}(X_0 = x_0, \ldots, X_{n-1} = x_{n-1})} \\ &= \frac{\lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x_n}}{\lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-2} x_{n-1}}} \end{aligned}$$

The final line simplifies to $P_{x_{n-1} x_n}$, which is the probability of transitioning from state x_{n-1} to state x_n. This implies that the process is a Markov process.

($\Longleftarrow$) We now assume that the joint distribution of the random variables is given by the expression in the theorem. We can write the conditional probability $\mathbb{P}(X_0 = x_0, \dots X_n = x_n)$ as

$$\begin{aligned}\mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \cdot \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ = P_{x_{n-1}x_n} \cdot \lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-2}x_{n-1}} \\ = \lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1}x_n}\end{aligned}$$

This recursive property implies that the process is a Markov process. □

Definition 3.2.1 (Conditional Probabilities). Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a Markov chain $(X_t)_{t\geq 0} \sim \text{Markov}(\lambda, P)$, the *conditional probability* of the process is given by

$$\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i) \tag{55}$$

where $A \in \mathscr{F}$. The conditional probability is a probability measure on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Definition 3.2.2 (Dirac Measure). We can define the *Dirac measure* δ_x as a probability measure on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

Proposition 3.2.1. For a Markov chain $(X_t)_{t\geq 0} \sim \text{Markov}(\delta_x, P)$,

$$\mathbb{P} = \mathbb{P}_x$$

Proof. We can write the conditional probability with some $A \in \mathscr{F}$ as

$$\mathbb{P}_x(A) = \mathbb{P}(A \mid X_0 = x)$$

Here, $X_0 \sim \delta_x$. By definition of the Dirac measure, $\mathbb{P}(X_0 = x) = 1$. Therefore, we have

$$\mathbb{P}_x(A) = \mathbb{P}(A \mid X_0 = x) = \frac{\mathbb{P}(A, X_0 = x)}{\mathbb{P}(X_0 = x)} = \mathbb{P}(A)$$

This implies that $\mathbb{P} = \mathbb{P}_x$ for all $A \in \mathscr{F}$. □

Proposition 3.2.2. Given a Markov chain $(X_t)_{t\geq 0} \sim \text{Markov}(\lambda, P)$, we have the two properties below.

(I) $\mathbb{P}(X_n = i) = \lambda P^n_i$

(II) $\mathbb{P}(X_n = j \mid X_0 = i) = P^n_{ij}$

Proof. We can write the joint distribution of the random variables as

$$\begin{aligned}\mathbb{P}(X_n = i) &= \sum_{x_{n-1}} \sum_{x_{n-2}} \cdots \sum_{x_0} \lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x_n} \\ &= \sum_{x_{n-2}} \cdots \sum_{x_0} \lambda P_{x_0 x_1} \cdots \left(\sum_{x_{n-1}} P_{x_{n-1} x_n} \right) \\ &= \lambda P^n_i\end{aligned}$$

This proves the first property. The proof of the second property is as follows.

$$\mathbb{P}(X_n = j \mid X_0 = i) = \frac{\mathbb{P}(X_n = j, X_0 = i)}{\mathbb{P}(X_0 = i)} = \frac{\sum_{x_{n-1}} \cdots \sum_{x_0} \lambda_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x_n}}{\lambda_i} = P^n_{ij}$$

□

Lemma 3.2.1. Let $(X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)$ be a Markov chain. Then, $(X_{\alpha+\beta t})_{t \geq 0} \sim \text{Markov}(\lambda P^\alpha, P^\beta)$ is also a Markov chain for all $\alpha, \beta \in \mathbb{N}$.

3.3 Hitting Time

The *hitting time* of a Markov chain is the time it takes for the process to reach a certain state. We define the hitting time formally below.

Definition 3.3.1 (Hitting Time). Let $(X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)$ take values in I and denote $A \subseteq I$. We define the *hitting time* of A, denoted by H^A, as follows.

$$H^A := \inf\{n \geq 0 \mid X_n \in A\} \tag{57}$$

We denote the hitting time of A given an initial state $i \in I$ as H^A_i.

Note that the hitting time of a set A is a random variable that takes values in $\{1, 2, 3, \ldots\} \cup \{\infty\}$. We can denote the probability that the process hits A with h^A.

$$h^A = \mathbb{P}(H^A < \infty)$$

We also consider the probability of finite hitting times given an initial state $i \in I$.

$$h^A_i = \mathbb{P}(H^A < \infty \mid X_0 = i) = \mathbb{P}_i(H^A < \infty)$$

With the law of total probabilities, we can say that $h^A = \sum_{i \in I} \lambda_i h^A_i$.

Theorem 3.3.1. The finite hitting time probabilities $(h^A_i)_{i \in I}$ are the unique solution to the system of equations

$$\begin{cases} h^A_i = 1 & \text{if } i \in A \\ h^A_i = \sum_{j \in I} P_{ij} h^A_j & \text{if } i \notin A \end{cases} \tag{58}$$

Additionally, $(h^A_i)_{i \in I}$ is the *smallest* non-negative solution of the system of equations.

Note that we can develop a trivial solution to this system if we have another non-negative solution $(x_i)_{i\in I}$ such that $x_i := 1$ for all i. This is not a minimal solution, but it guarantees its existence.

Proof. We first prove that the finite hitting time probabilities satisfy the system. If $i \in A$, then since $H_i^A = 0$, we get $h_i^A = 1$.

Let $i \notin A$. Then, we need to perform at least one step before reaching A. This means we must compute the marginal probability on the value of X_1 to find

$$\begin{aligned} h_i^A &= \mathbb{P}(H^A < \infty \mid X_0 = i) \\ &= \sum_j \mathbb{P}(H^A < \infty \mid X_1 = j, X_0 = i)\mathbb{P}(X_1 = j \mid X_0 = i) \\ &= \sum_j \mathbb{P}(H^A < \infty \mid X_1 = j) P_{ij} \\ &= \sum_j h_j^A P_{ij} \quad \textit{by the Markov property.} \end{aligned}$$

Suppose now that $(x_i)_{i\in I}$ is another non-negative solution. Then $x_i = h_i^A = 1$ for all $i \in A$. We pick some $i \notin A$.

$$x_i = \sum_{j\in I} p_{ij} x_j = \sum_{j\in A} p_{ij} + \sum_{j\notin A} p_{ij} x_j.$$

Substituting for x_j, we arrive at

$$\begin{aligned} x_i &= \sum_{j_1\in A} p_{ij_1} + \sum_{j_1\notin A} p_{ij_1}\left(\sum_{j_2\in A} p_{j_1 j_2} + \sum_{j_2\notin A} p_{j_1 j_2} x_{j_2}\right) \\ &= \sum_{j_1\in A} p_{ij_1} + \sum_{j_1\notin A, j_2\in A} p_{ij_1} p_{j_1 j_2} + \sum_{j_1\notin A, j_2\notin A} p_{ij_1} p_{j_1 j_2} x_{j_2}. \end{aligned}$$

□

The Gambler's Ruin Problem

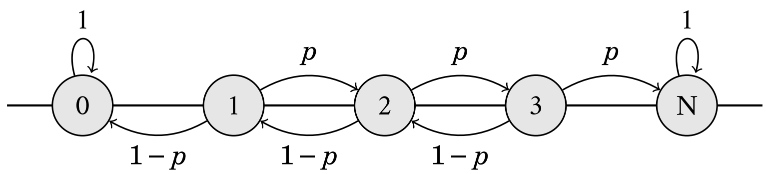
A gambler starts with k dollars and plays a game. Every round, he either loses or wins 1 dollar. He will play until he has N dollars or loses all his money. We are interested in the probability of the latter event.

Let X_n be the gambler's fortune after n rounds, and let $p \in (0,1)$ is the fixed probability of winning a round. Then, $(X_n)_{n\geq 0}$ is a Markov chain with state space $I = \{0, 1, \ldots, N\}$ and transition matrix P given by

$$P_{ij} = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

where $i \in \{0,1,\ldots,N-1\}$. Both 0 and N are called *absorbing states*, $P_{00} = P_{NN} = 1$. The event $A = \{0\}$ is the event that the gambler loses all his money. We are interested in the hitting time of A.

$$h_k = \mathbb{P}(H^{\{0\}} < \infty \mid X_0 = k)$$



Above, is the transition diagram for the gambler's ruin problem ($N = 3$). The gambler starts at state k and plays until he reaches state 0 or state N. The gambler's ruin problem is a special case of the lazy random walk problem. We can solve the gambler's ruin problem by solving the lazy random walk problem.

Since $0 \in A$, $h_0 = 1$. But, since X could never reach 0 if it starts at N, $h_N = 0$. We can now solve for h_i for $i \in \{1,2,\ldots,N-1\}$.

$$h_i = ph_{i+1} + (1-p)h_{i-1}$$

This implies the following.

$$p(h_{i+1} - h_i) = (1-p)(h_i - h_{i-1}) \iff h_{i+1} - h_i = \frac{1-p}{p}(h_i - h_{i-1})$$

Define the parameter $\eta := \frac{1-p}{p}$. We can now write the recursive equation as

$$h_{i+1} - h_i = \eta(h_i - h_{i-1}) = \ldots = \eta^i(h_1 - h_0) = \eta^i(h_1 - 1).$$

Now, sum over i to $k-1$.

$$h_{k-1} = (\eta^{k-1} + \eta^{k-2} + \ldots + \eta + 1)(h_1 - 1) = \begin{cases} \frac{1-\eta^k}{1-\eta}(h_1 - 1) & \text{if } \eta \neq 1 \\ k(h_1 - 1) & \text{if } \eta = 1 \end{cases}$$

Since $h_N = 0$, we obtain the following result.

$$h_1 - 1 = \begin{cases} -\frac{1-\eta}{1-\eta^N} & \text{if } \eta \neq 1 \\ -\frac{1}{N} & \text{if } \eta = 1 \end{cases} \implies h_k = \begin{cases} 1 - \frac{1-\eta^k}{1-\eta^N} & \text{if } \eta \neq 1 \\ 1 - \frac{k}{N} & \text{if } \eta = 1 \end{cases}$$

We can generalize the finite hitting time probabilities to the expected value of any non-negative function φ on H^A.

Theorem 3.3.2 (Mean Hitting Time). If $i \in A$, then $H^A = 0$. If $i \notin A$, for any function $\varphi : \{0,1,\ldots\} \to [0,\infty)$, we have

$$\mathbb{E}_i[\varphi(H^A)] = \sum_{j \in I} p_{ij} \mathbb{E}_j[\varphi(H^A + 1)] \tag{59}$$

Proof. Denote $H_i^A = H^A(i, X_1, X_2, \ldots)$. If $i \in A$, then $H_i^A = 0$. If $i \notin A$, note that at least one step has to be taken.

$$H^A(i, X_1, \ldots) = H^A(X_1, X_2, \ldots) + 1$$

Applying φ on both sides and taking the expected value, we get

$$\begin{aligned}
\mathbb{E}_i[\varphi(H^A)] = \mathbb{E}_i[\varphi(H^A(i, X_1, X_2, \ldots))] &= \mathbb{E}_i[\varphi(H^A(X_1, X_2, \ldots) + 1)] \\
&= \sum_{j \in I} \mathbb{E}_i[\varphi(H^A(X_1, X_2, \ldots) + 1)]\mathbb{P}_i(X_1 = j) \\
&= \sum_{j \in I} P_{ij}\mathbb{E}_j[\varphi(H^A(X_1, X_2, \ldots) + 1)].
\end{aligned}$$

This completes the proof. □

3.4 Stopping and Return Times

Recall the definition of a filtration; a sequence of σ-algebras $(\mathscr{F}_t)_{t \geq 0}$ is a filtration if $\mathscr{F}_s \subseteq \mathscr{F}_t$ for all $s \leq t$. We can define a stopping time with respect to a filtration.

Definition 3.4.1 (Stopping Time). Let T be a $\{0, 1, 2, \ldots, \infty\}$-valued random variable. We say that T is a *stopping time* with respect to a filtration $(\mathscr{F}_t)_{t \geq 0}$ if

$$\{T = n\} \in \mathscr{F}_n \quad \text{for all } n \in \{0, 1, 2, \ldots, \infty\}$$

We say T is a stopping time with respect to a process $(X_t)_{t \geq 0}$ if T is a stopping time with respect to the natural filtration $(\mathscr{F}_t)_{t \geq 0} = \sigma(X_0, X_1, \ldots, X_t)$.

If for any time k, by observing $X_0, X_1, \ldots, X_k$, we can determine whether or not T has occured before time k, then T is a stopping time.

Example 3.4.1. The hitting time $H^A = \inf\{n \geq 0 \mid X_n \in A\}$ is a stopping time with respect to the natural filtration $(\mathscr{F}_t)_{t \geq 0}$.

$$\{H^A = k\} = \{X_0 \notin A, X_1 \notin A, \ldots, X_{k-1} \notin A, X_k \in A\} \in \mathscr{F}_k.$$

Example 3.4.2. The return time $R^A = \inf\{n \geq 1 \mid X_n \in A\}$ is a stopping time with respect to the natural filtration $(\mathscr{F}_t)_{t \geq 0}$.

$$\{R^A = k\} = \{X_1 \notin A, X_2 \notin A, \ldots, X_{k-1} \notin A, X_k \in A\} \in \mathscr{F}_k.$$

Example 3.4.3. The last return time $L^A = \sup\{n \geq 0 \mid X_n \in A\}$ is *not* a stopping time with respect to the natural filtration $(\mathscr{F}_t)_{t \geq 0}$.

$$\{L^A = k\} = \{X_k \in A, X_{k+1} \notin A, X_{k+2} \notin A, \ldots\} \notin \mathscr{F}_k.$$

Note that the event $\{L^A = k\}$ depends on the *future* values of the process.

Proposition 3.4.1. Recall that the hitting time of a set A given a Markov chain $(X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)$ is defined as

$$H^A = \inf\{n \geq 0 \mid X_n \in A\}$$

We can claim that H^A is a stopping time with respect to the natural filtration $(\mathscr{F}_t)_{t\geq 0}$.

Proof. Note that $\{H^A \leq n\}$ can be written as the following union.

$$\{H^A \leq n\} = \bigcup_{k=0}^{n} \{X_k \in A\} \in \mathscr{F}_n.$$

This implies that H^A is a stopping time with respect to the natural filtration. □

The first *return time* of a Markov chain is the time it takes for the process to return to a certain state.

Definition 3.4.2 (Return Time). Let $(X_t)_{t\geq 0} \sim \text{Markov}(\lambda, P)$ be a Markov chain. The *return time* of state i is defined as

$$R^A = \inf\{n \geq 1 \mid X_n \in A\} \tag{60}$$

Note the difference between the hitting time and the return time. The hitting time is the time it takes to reach a state, while the return time is the time it takes to return to a state.

Proposition 3.4.2. The last return time of a Markov chain is expressed below.

$$L^A = \sup\{n \geq 0 \mid X_n \in A\}$$

This is *not* a stopping time.

Proof. The event $\{L^A \leq n\}$ can be written as

$$\{L^A \leq n\} = \{X_n \in A\} \cap \bigcup_{k=n+1}^{\infty} \{X_k \notin A\} \notin \mathscr{F}_n.$$

This implies that L^A is not a stopping time. □

The Markov property states that for each time t, conditional on $X_t = i$, the process after t begins anew from state i. Now, instead of conditioning on $X_t = i$, let us wait for th process to hit a set $A \subset I$, at some random time H^A. What can we say about the process after time H^A?

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. For a given filtration $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \ldots \subseteq \mathscr{F}$, we can define the stopped σ-algebra at time T.

Definition 3.4.3 (Stopped σ-Algebra). Let T be a stopping time with respect to a filtration $\mathscr{F}_n = \sigma(X_0, X_1, \ldots, X_n)$. The *stopped σ-algebra* at time T is a σ-algebra $F_T \subseteq \mathscr{F}$.

$$A \in F_T \iff A \cap \{T = n\} \in \mathscr{F}_n \quad \text{for all } n \geq 0. \tag{61}$$

Proposition 3.4.3. F_T is a σ-algebra.

Proof. We must show the three properties of a σ-algebra.

1. We have $\varnothing \cap \{T = n\} = \varnothing \in \mathscr{F}_n$ for all $n \geq 0$. Therefore, $\varnothing \in F_T$.
2. Suppose $A \in F_T$. For all $n \geq 0$, $A^c \cap \{T = n|\} = (A \cap \{T = n\})^c \in \mathscr{F}_n$. Therefore, $A^c \in F_T$.
3. Given a sequence $A_i \in \mathscr{F}_T$ for all $i \in \mathbb{N}$, we have that $A_i \cap \{T = n\} \in \mathscr{F}_n$ for all $n \geq 0$. Therefore, $\bigcup_{i=1}^{\infty} A_i \cap \{T = n\} \in \mathscr{F}_n$ for all $n \geq 0$. This implies that $\bigcup_{i=1}^{\infty} A_i \in F_T$.

□

Heuristically, $\mathscr{F}_T$ is the information available in $X_0, X_1, \ldots$ up to a *random* time T.

Theorem 3.4.1 (The Strong Markov Property). Let $(X_n)_{n\geq 0} \sim \text{Markov}(\lambda, P)$ be a Markov chain and T be a stopping time.

1. Conditioned on $\{T < \infty, X_T = i\}$, the distribution of $(X_{T+n})_{n\geq 0}$ is independent of $\mathscr{F}_T$.
2. Conditioned on $\{T < \infty, X_T = i\}$, $(X_{T+n})_{n\geq 0} \sim \text{Markov}(\delta_i, P)$.

The proof is omitted for brevity.

Consider the process $(X_n)_{n\geq 0} \sim \text{Markov}(\lambda, P)$ and some subset $A \subset I$. With A, we have *almost sure* finite hitting time $\mathbb{P}_i(H^A < \infty) = 1$. We can define the following sequence of random variables.

$$
\begin{aligned}
T_0 &= H^A && \text{the first hitting time of } A \\
T_1 &= \inf\{n \geq 0 \mid X_n \in A\} && \text{the first return time of } A \\
&\vdots \\
T_k &= \inf\{n \geq T_{k-1} + 1 \mid X_n \in A\} && \text{the } k\text{-th return time of } A
\end{aligned}
$$

Theorem 3.4.2. The random variables T_m are stopping times for all $m \geq 0$.

Proof. We have a base case for $T_0 = H^A$, which we proved in Ex. 17. The proof is by induction. Suppose T_{m-1} is a stopping time. We can write T_m as

$$T_m = \inf\{n \geq T_{m-1} + 1 \mid X_n \in A\} = \inf\{n \geq 0 \mid X_{n+T_{m-1}+1} \in A\}.$$

This implies that T_m is a stopping time. □

Recall the definition of *return times* or *first passage times* to A.

$$R^A = \inf\{n \geq 1 \mid X_n \in A\}$$

This is a stopping time.

Theorem 3.4.3. The discrete-time stochastic process $(X_{T_j})_{j\geq 0}$ is a Markov chain with transition matrix P.

$$\widehat{P}_{ij} := \mathbb{P}_i(X_{R^A} = j), \quad i,j \in A.$$

The process $(X_{T_j})_{j\geq 0}$ is called the *trace process* of $(X_n)_{n\geq 0}$ on A.

Example 3.4.4. Let $(X_n)_{n\geq 0}$ be a Markov chain with state space $I = \{0,1,\dots,N\}$ with transition probabilities.

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1-p, \quad i \in \{1,2,\dots,N-1\}$$
$$p_{0,1} = 1, \quad p_{N,N-1} = 1, \quad \text{absorbing states}$$

Here, $p \in (0,1)$. With $\eta := \frac{1-p}{p}$, we can write the hitting time as

$$q_k := \mathbb{P}_k(\text{hitting 0 before } N) = \begin{cases} 1 - \frac{1-\eta^k}{1-\eta^N} & \text{if } p \neq 1/2 \\ 1 - \frac{k}{N} & \text{if } p = 1/2 \end{cases}$$

Let $A = \{0,N\}$. The trace process of $(X_n)_{n\geq 0}$ on A is a Markov chain with transition probabilities

$$p_{0,0} = q_1, \quad p_{0,N} = 1-q_1, \quad p_{N,0} = q_{N-1}, \quad p_{N,N} = 1-q_{N-1}.$$

Suppose $p < 1/2$ ($\eta > 1$) and consider the limiting case $N \to \infty$.

$$q_1 = 1 - \frac{\eta - 1}{\eta^N - 1} \sim 1 - \frac{1}{\eta^N - 1} \text{ and } q_{N-1} = 1 - \frac{\eta^{N-1}-1}{\eta^N - 1} \simeq 1 - \frac{1}{\eta}.$$

The Probability Generating Function

The probability generating function (PGF) is a power series representation of the probability mass function of a discrete random variable.

Definition 3.4.4 (Probability Generating Function). Let Y be a non-negative integer-valued random variable. The *probability generating function* of Y is defined as

$$\phi_Y(t) = \mathbb{E}[t^Y] = \sum_{k=0}^{\infty} t^k \mathbb{P}(Y = k), \quad t \in [0,1]. \tag{62}$$

It is clear that the mapping $t \mapsto \phi_Y(t)$ is non-decreasing, $\phi_Y(0) = 0$, and $\phi_Y(1) = 1$. So, the function is bounded by $[0,1]$.

FINISH 05.

3.5 Class Structure

It is possible to break Markov chains into smaller chains that are easier to analyze. We do this by investigating the communicating classes of a Markov chain.

Definition 3.5.1 (Leading). Let $(X_n)_{n\geq 0}$ be a Markov chain with state space I. We say that state i *leads to* state j if the probability of reaching state j from state i in a finite number of steps is positive.

$$i \to j \implies \mathbb{P}_i(H^{\{j\}} < \infty) > 0. \tag{63}$$

We say i does not lead to j if $i \nrightarrow j$.

$$i \nrightarrow j \implies \mathbb{P}_i(H^{\{j\}} = \infty) = 1.$$

Lemma 3.5.1. Given a Markov chain $(X_n)_{n\geq 0} \sim \text{Markov}(\lambda, P)$ with state space I. Then, for $i, j \in I$, $i \neq j$, the following are equivalent.

(i) $i \to j$.

(ii) There exists a path from i to j in the corresponding weighted directed graph for P.

(iii) There exists an $m \geq 0$ such that $(P^m)_{ij} > 0$.

Proof. First, we prove that (i) implies (iii). Suppose $i \to j$. Then, by definition,

$$0 < \mathbb{P}_i(H^{\{j\}} < \infty) = \sum_{m=0}^{\infty} \mathbb{P}_i(H^{\{j\}} = m).$$

So, there exists an $m \geq 0$ such that $\mathbb{P}_i(H^{\{j\}} = m) > 0$. This implies the following.

$$(P^m)_{ij} = \mathbb{P}_i(X_m = j) \geq \mathbb{P}_i(H^{\{j\}} = m) > 0.$$

Now, we prove that (iii) implies (i). Suppose there exists an $m \geq 0$ such that $(P^m)_{ij} > 0$.

$$0 < (P^m)_{ij} = \mathbb{P}_i(X_m = j) \leq \mathbb{P}_i(H^{\{j\}} < \infty).$$

□

Definition 3.5.2 (Communicating). We say that states i and j *communicate* if $i \to j$ and $j \to i$. In other words, if i leads to j and j leads to i, then $i \leftrightarrow j$.

Proposition 3.5.1. The relation $\leftrightarrow$ is an equivalence relation.

Proof. We must show that $\leftrightarrow$ is reflexive, symmetric, and transitive.

(i) Reflexive: $i \to i$ since $\mathbb{P}_i(H^{\{i\}} = 0) = 1$.

(ii) Symmetric: If $i \to j$, then there exists an $m \geq 0$ such that $(P^m)_{ij} > 0$. This implies that $(P^m)_{ji} > 0$, so $j \to i$.

(iii) Transitive: If $i \to j$ and $j \to k$, then there exists an $m \geq 0$ such that $(P^m)_{ij} > 0$ and an $n \geq 0$ such that $(P^n)_{jk} > 0$. This implies that $(P^{m+n})_{ik} = \sum_{l \in I} (P^m)_{il} (P^n)_{lk} > 0$, so $i \to k$.

□

By applying the equivalence relation $\leftrightarrow$, we can partition the state space I into communicating classes.

Definition 3.5.3 (Communicating Class). The equivalence classes of the relation $\leftrightarrow$ are called *communicating classes*.

Definition 3.5.4 (Closed and Absorbing Classes). A class (set of states) C of state space I is *closed* if for all $i \in C$, the following is true.

$$i \to j \implies j \in C.$$

We say that a state $i \in I$ is an *absorbing state* if $\{i\}$ is a closed class.

$$i \nrightarrow j \quad \text{for all } j \neq i.$$

If i is an absorbing state, then once a Markov chain reaches state i, it will never leave. The state i is an *absorbing state* because it absorbs the chain. Also, note the trivial case of the entire state space I being a closed class. This is because $i \to j$ for all $i, j \in I$.

Definition 3.5.5 (Irreducibility). A Markov chain is *irreducible* if ther exists a unique closed class C, which is the entire state space I.

Theorem 3.5.1. A Markov chain with state space I is irreducible if and only if

$$i \leftrightarrow j \quad \text{for all } i, j \in I. \tag{64}$$

Proof. ($\Longleftarrow$) Assume $i \leftrightarrow j$ for all $i, j \in I$. Let C be a closed class. We wish to prove that $C = I$. We will do this by contradiction; assume $C \neq I$. Then choose $j \in I \setminus C$. Pick any $i \in C$, we have $i \to j$, so $j \in C$. This is a contradiction, so $C = I$, making the Markov chain irreducible.

($\Longrightarrow$) Assume the Markov chain is irreducible. We wish to show that $i \leftrightarrow j$ for all $i, j \in I$. We will do this by contradiction; assume $i \nleftrightarrow j$ for some $i, j \in I$. Then, $i \to j$ and $j \to i$ are false. This implies that i and j are in different closed classes, which contradicts the Markov chain being irreducible. Therefore, $i \leftrightarrow j$ for all $i, j \in I$. This completes the proof. □

Definition 3.5.6 (Recurrence). A state $i \in I$ is *recurrent* if

$$\mathbb{P}_i(X_t = i \text{ for infinitely many } t) = 1. \tag{65}$$

A state $i \in I$ is *transient* if

$$\mathbb{P}_i(X_t = i \text{ for infinitely many } t) = 0. \tag{66}$$

This leads us to the *dichotomy* theorem.

Theorem 3.5.2 (Dichotomy). All states are *either* recurrent or transient.

We will not prove this theoreom yet, we will first introduce some more definitions.

Definition 3.5.7 (mth Return Time). Given a state space I, the mth return time of state i is defined as

$$R_0^{\{i\}} = 0 \text{ and } R_m^{\{i\}} = \inf\{t > R_{m-1}^{\{i\}} \mid X_t = i\}. \tag{67}$$

Intuitively, the mth return time is the time it takes for the Markov chain to return to state i for the mth time.

Note that the "hit", $t = 0$, is *not* counted as a return time.

Theorem 3.5.3. For all $m \geq 0$ and $i \in I$, conditioned on $\{R_m^{\{i\}} < \infty\}$, the random variable

$$G_{m+1}^{\{i\}} := R_{m+1}^{\{i\}} - R_m^{\{i\}}$$

is *independent* of $\mathscr{F}_{R_m^{\{i\}}}$ and has the same distribution as $R_1^{\{i\}}$ under $\mathbb{P}_i$.

This theorem is a consequence of the strong Markov property. Once a Markov chain returns to a state i, the future "waiting time" until it returns is independent of what has happened before. Also, the distribution of the waiting time is the same as the distribution of the first return time.

Proof. Given $R_m^{\{i\}} < \infty$, we use the strong Markov property.

$$\left(X_{R_m^{\{i\}}+k}\right)_{k\geq 0} \text{ is independent of } \mathscr{F}_{R_m^{\{i\}}}.$$
$$\left(X_{R_m^{\{i\}}+k}\right)_{k\geq 0} \sim \text{Markov}(\delta_i, P).$$

The second line states that the process after $R_m^{\{i\}}$ is a Markov chain with the same transition matrix as the original chain. Using the defintition, we can rewrite the m-th gap, G, as follows.

$$\begin{aligned} G_m^{\{i\}} = R_{m+1}^{\{i\}} - R_m^{\{i\}} &= \inf\{t > R_m^{\{i\}} \mid X_t = i\} - R_m^{\{i\}} \\ &= \inf\{t = R_m^{\{i\}} + k \mid X_t = i\} - R_m^{\{i\}} \\ &= \inf\{k > 0 \mid X_{R_m^{\{i\}}+k} = i\} \end{aligned}$$

The last line is a restatement of the definition of the gap. This implies that $G_m^{\{i\}}$ is the first return time of the process after $R_m^{\{i\}}$ to state i. This completes the proof. □

Definition 3.5.8 (Number of Returns). We define the total *number of returns* for a state i as V_i.

$$V_i = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = i\}}. \tag{68}$$

Here, $\mathbb{1}_{\{X_n=i\}}$ is the indicator function

$$\mathbb{1}_{\{X_n=i\}} = \begin{cases} 1 & \text{if } X_n = i \\ 0 & \text{if } X_n \neq i \end{cases}$$

V_i represents the number of times $X_n = i$ in the entire lifetime of the Markov chain.

Using this definition, we can redefine recurrence and transience. For any $i \in I$, we have the following.

$$i \text{ is recurrent} \iff \mathbb{P}_i(V_i = \infty) = 1$$
$$i \text{ is transient} \iff \mathbb{P}_i(V_i = \infty) = 0$$

We can find the expectation of V_i using Fubini's theorem.

$$\mathbb{E}(V_i) = \sum_{n=0}^{\infty} \mathbb{E}\left(\mathbb{1}_{\{X_n=i\}}\right) = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = i) = \sum_{n=0}^{\infty} (P^n)_{ii}.$$

We can apply Theorem 18 to find the following.

$$\mathbb{P}_i(V_i > m) = \mathbb{P}_i\left(G_j^{\{i\}} < \infty, j \in [0, m)\right) = \prod_{j=0}^{m-1} \mathbb{P}_i(G_j^{\{i\}} < \infty) = \mathbb{P}_i\left(R_1^{\{i\}} < \infty\right)^m.$$

Define the following.

$$a_i = \mathbb{P}\left(R^{\{i\}} < \infty\right), \quad i \in I. \implies \mathbb{P}_i(V_i = \infty) = \lim_{m \to \infty} \mathbb{P}_i(V_i > m) = \begin{cases} 1 & \text{if } a_i = 1 \\ 0 & \text{if } a_i < 1 \end{cases}$$

Moreover, we can calculate the expectation using a_i.

$$\mathbb{E}_i(V_i) = \sum_{m=0}^{\infty} \mathbb{P}_i(V_i > m) = \sum_{m=0}^{\infty} a_i^m = \begin{cases} \infty & \text{if } a_i = 1 \\ \frac{1}{1-a_i} & \text{if } a_i < 1 \end{cases}.$$

This leads us to the following result regarding recurrence and transience.

Theorem 3.5.4. For any $i \in I$, the following are equivalent.

$$\mathbb{P}_i\left(R^{\{i\}} < \infty\right) = 1 \iff \mathbb{P}_i\left(V_i = \infty\right) \iff \mathbb{E}_i(V_i) = \infty,$$
$$\mathbb{P}_i\left(R^{\{i\}} = \infty\right) < 1 \iff \mathbb{P}_i\left(V_i < \infty\right) \iff \mathbb{E}_i(V_i) < \infty.$$

Using these definitions, we can now prove Theorem 17.

Proof. Let $i \in I$. If $\mathbb{P}_i(R^{\{i\}} < \infty) = 1$, then i is recurrent. If $\mathbb{P}_i(R^{\{i\}} < \infty) < 1$, then i is transient. This completes the proof. □

Recurrence and transience are *class properties.*

Theorem 3.5.5. For any communication class C, all states are recurrent or all states are transient.

Proof. Let C be a communication class. Take two states $i, j \in C$. It is enough to prove that if i is recurrent, then so is j. By the previous theorems, with a state $a \in I$, we have the following.

$$a \text{ is recurrent} \iff \mathbb{E}_a(V_a = \infty) = \infty \iff \sum_{n=0}^{\infty} (P^n)_{aa} = \infty.$$

Since $i \leftrightarrow j$, there exists $s, t, > 0$ such that $(P^s)_{ij} > 0$ and $(P^t)_{ji} > 0$. So, we obtain the following.

$$\sum_{n=0}^{\infty} (P^n)_{jj} \geq \sum_{n=0}^{\infty} (P^{n+s+t})_{jj} \geq \sum_{n=0}^{\infty} (P^t)_{ji}(P^n)_{ii}(P^s)_{ij} = (P^t)_{ji}(P^s)_{ij} \sum_{n=0}^{\infty} (P^n)_{ii} = \infty.$$

This implies that j is recurrent. This completes the proof. □

Theorem 3.5.6. Any recurrent communication class is closed.

Proof. Let C be a recurrent communication class and suppose for contradiction that C is not closed. Then there exists states $i \in C$ and $j \notin C$ such that $P_{ij} > 0$. Since i is recurrent, we have $\mathbb{P}_i(R^{\{i\}} < \infty) = 1$.

However, once the chain moves from i to j, since $j \notin C$ and C is a communication class, there is no path back from j to i (otherwise j would be in C). This means $\mathbb{P}_i(R^{\{i\}} < \infty) < 1$, which contradicts the recurrence of i. Therefore, C must be closed. □

Theorem 3.5.7. Any irreducible Markov chain on a *finite* state space I is recurrent.

Proof. Suppose for contradiction that there exists a transient state $i \in I$. Since the state space is finite and the chain is irreducible, we can partition the states into recurrent and transient states. Let R be the set of recurrent states and T be the set of transient states, with $i \in T$.

From any transient state, the chain will eventually leave the set T and never return (by definition of transience). Since I is finite, the chain must eventually reach and remain in the recurrent states R. But this contradicts irreducibility: if the chain starts at a transient state and eventually gets trapped in R, then states in T cannot communicate with states in R in both directions.

Therefore, there can be no transient states, and all states must be recurrent. □

Theorem 3.5.8. Suppose P is irreducible and recurrent. Then, for any initial condition, the following holds.

$$\mathbb{P}(R^{\{i\}} < \infty) = 1 \quad \text{for all } i \in I. \tag{69}$$

3.6 Invariant Distributions

Now, we will focus on the *long-term* behavior of irreducible and recurrent Markov chains.

Definition 3.6.1 (Invariant Distribution). A probability distribution μ is called *invariant* or stationary for a transition matrix P if

$$\mu P = \mu \tag{70}$$

Based on the above definition, two facts immediately follow.

1. By induction, we get that $X_0 \sim \mu \implies X_t \sim \mu$ for all $t \geq 0$.
2. μ is a left eigenvector of P with eigenvalue 1.

Theorem 3.6.1. Let P be a stochastic matrix on a *finite* state space I. Suppose that for some $i \in I$, $(P^n)_{ij}$ has a limit as $n \to \infty$ for all $j \in I$. Let

$$\mu_j := \lim_{n\to\infty} (P^n)_{ij}. \tag{71}$$

Then, $\mu = \left(\mu_j\right)_{j \in I}$ is an invariant distribution for P.

Proof. We first assure ourselves that μ is a probability distribution. Since I is finite, we can interchange limits and summations over I. Thus, since P^n is a stochastic matrix, we have the following.

$$\sum_{j\in I} \mu_j = \sum_{j\in I} \lim_{n\to\infty} (P^n)_{ij} = \lim_{n\to\infty} \sum_{j\in I} (P^n)_{ij} = \lim_{n\to\infty} 1 = 1.$$

Now, we prove that μ_j is invariant. Observe the following.

$$(P^{n+1})_{ij} = \sum_{k\in I} (P^n)_{ik} P_{kj}.$$

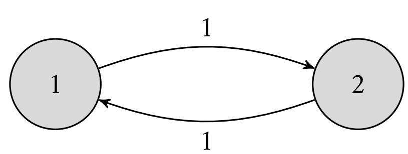
Taking the limit as $n \to \infty$, we have the following.

$$\mu_j = \lim_{n\to\infty} (P^{n+1})_{ij} = \lim_{n\to\infty} \sum_{k\in I} (P^n)_{ik} P_{kj} = \sum_{k\in I} \lim_{n\to\infty} (P^n)_{ik} P_{kj} = \sum_{k\in I} \mu_k P_{kj} = (\mu P)_j.$$

This completes the proof. □

Note that the above theorem assumes that the limit $\lim_{n\to\infty}(P^n)_{ij}$ exists for all $j \in I$. This is not always the case. However, if the limit does exist, then the invariant distribution is unique.

Example 3.6.1. Consider the following trivial Markov chain.



The transition matrix is

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

The limit $\lim_{n\to\infty}(P^n)_{ij}$ does not exist for all $j \in I$. This is because the powers of P oscillate between the identity matrix and P. However, the invariant distribution is $\mu = (1/2, 1/2)$.

Invariant Distributions using the First Return Time

Recall that the first return time of a Markov chain is defined as

$$R^{\{i\}} := \inf\{n \geq 1 \mid X_n = i\}. \tag{72}$$

For $j \in I$, define the following.

$$\gamma_j^{(i)} := \mathbb{E}_i \sum_{n=0}^{R^{\{i\}}-1} \mathbb{1}_{\{X_n = j\}}.$$

$\gamma_j^{(i)}$ is the expected number of visits to state j until the first return to state i. With the strong Markov chain, we can establish the following theorem.

Theorem 3.6.2. Let the Markov chain be irreducible and recurrent. Fix some state $i \in I$. Then, the following hold.

(i) $\gamma_i^{(i)} = 1$.
(ii) $\gamma_i^{(i)} = \left(\gamma_k^{(i)}\right)_{k \in I}$ is an invariant distribution for P.
(iii) $\gamma_k^{(i)} \in (0, \infty)$ for all $k \in I$.
(iv) $\sum_{k \in I} \gamma_k^{(i)} = \mathbb{E}_i R^{\{i\}}$.

Proof. Fix $i \in I$. For ease of notation, let $R_i = R^{\{i\}}$ and $\gamma_j = \gamma_j^{(i)}$. We will prove each part of the theorem.

(i) By definition, we have that $X_n \neq i$ for $n \in [0, R_i - 1]$. Thus, we get the following.

$$\gamma_i = \mathbb{E}_i \sum_{n=0}^{R_i - 1} \mathbb{1}_{\{X_n = i\}} = \mathbb{E}_i \mathbb{1}_{X_0 = i} = 1.$$

(ii) We can rewrite the expectation as follows.

$$\begin{aligned}
\gamma_k = \mathbb{E}_i \sum_{n=0}^{R_i - 1} \mathbb{1}_{X_n = k} &= \mathbb{E}_i \sum_{n=1}^{R_i} \mathbb{1}_{X_n = k} \\
&= \mathbb{E}_i \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{1}_{X_{n-1} = j, X_n = k, R_i > n-1} \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{P}_i(X_n = k \mid X_{n-1} = k, R_i > n-1) \mathbb{P}_i(X_{n-1} = j, R_i > n-1) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} P_{jk} \mathbb{P}_i(X_{n-1} = j, R_i > n-1) \\
&= \sum_{j \in I} P_{jk} \mathbb{E}_i \sum_{n=1}^{\infty} \mathbb{1}_{X_{n-1} = j, R_i > n-1} \\
&= \sum_{j \in I} P_{jk} \mathbb{E}_i \sum_{n=0}^{R_i - 1} \mathbb{1}_{X_n = j} = \sum_{j \in I} P_{jk} \gamma_j.
\end{aligned}$$

Since the above result is true for any k, we have $\gamma P = \gamma$.

(iii) Let $k \in I$. Since the Markov chain is irreducible and recurrent, we have $i \leftrightarrow k$, meaning that $(P^n)_{ik} > 0$ and $(P^m)_{ki} > 0$ for some $n, m \geq 1$. By (ii), we have that $\gamma = \gamma P = \gamma P^2 = \ldots = \gamma P^N$. Thus, we have the following.

$$\gamma_k = \sum_{j \in I} \gamma_j (P^n)_{jk} \geq \gamma_i (P^n)_{ik} > 0.$$

and

$$1 = \gamma_i = \sum_{j \in I} \gamma_j (P^m)_{ji} \geq \gamma_k (P^m)_{ki} \implies \gamma_k \leq 1/(P^m)_{ki} < \infty.$$

Based on the above inequalities, we conclude that $\gamma_k \in (0, \infty)$.

(iv) By linearity, we have the following.

$$\mathbb{E}_i R_i = \mathbb{E}_i \sum_{n=0}^{R_i - 1} 1 = \mathbb{E} \sum_{n=0}^{R_i - 1} \mathbb{1}_{X_n \in I} = \sum_{k \in I} \mathbb{E}_i \sum_{n=0}^{R_i - 1} \mathbb{1}_{X_n = k} = \sum_{k \in I} \gamma_k.$$

□

Next, we establish the *minimality* of $\gamma^{(i)}$.

Theorem 3.6.3. Let P be irreducible and let λ be an invariant measure for P with $\lambda_i = 1$. Then, $\lambda \geq \gamma^{(i)}$. If in addition P is recurrent, then $\lambda = \gamma^{(i)}$.

Proof. For any $j \in I$, we have the following.

$$\begin{aligned}
\lambda_j = \sum_{k_1 \in I} \lambda_{k_1} P_{k_1 j} &= P_{ij} + \sum_{k_1 \neq i} \lambda_{k_1} P_{k_1 j} \\
&= P_{ij} + \sum_{k_1 \neq i} P_{i k_1} P_{k_1 j} + \sum_{k_1, k_2 \leq i} \lambda_{k_2} P_{k_2 k_1} P_{k_1 j} \\
&\vdots \\
&= P_{ij} + \sum_{k_1 \neq i} P_{i k_1} P_{k_1 j} + \ldots + \sum_{k_1, \ldots, k_{n-1} \neq i} P_{i k_{n-1}} \cdots P_{k_1 j} \\
&+ \sum_{k_1, \ldots, k_n \neq i} \lambda_{k_n} P_{k_n k_{n-1}} \cdots P_{k_1 j} \\
&= \mathbb{P}_i (X_1 = j, R^{\{i\}} \geq 1) + \mathbb{P}_i (X_2 = j, R^{\{i\}} \geq 2) \\
&+ \ldots + \mathbb{P}_i (X_n = j, R^{\{i\}} \geq n) + \sum_{k_1, \ldots, k_n \neq i} \lambda_{k_n} P_{k_n k_{n-1}} \cdots P_{k_1 j}
\end{aligned}$$

We drop the nonnegative term and the limit as $n \to \infty$ to get the following.

$$\lambda_j \geq \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = j, R^{\{i\}} \geq n) = \gamma_j.$$

Now suppose P is recurrent. Then, γ is an invariant measure for P.

$$\mu_j := \lambda_j - \gamma_j^{(i)}$$

Consider the above definition. By linearity, it is easy to see that μ is an invariant measure for P. Additionally, we have that $\mu_i = \lambda_i - 1 = 0$ and $\mu_j \geq 0$ for all $j \in I$. Pick any $j \neq i$. Since P is irreducible, we have $(P^n)_{ij} > 0$ for some $n \geq 1$. Thus, we have the following.

$$0 = \mu_i = \sum_{j \in I} \mu_j (P^n)_{ij} \geq \mu_j (P^n)_{ij} \implies \mu_j = 0.$$

We conclude that $\mu = 0$, so $\lambda = \gamma^{(i)}$. □

Recall the following implications

$$i \in I \text{ is recurrent} \iff \mathbb{P}_i(X_n = i \text{ infinitely often}) = 1 \iff \mathbb{P}_i(R^{\{i\}} < \infty) = 1.$$

Definition 3.6.2 (Positive and Null Recurrence). A state i is *positive recurrent* if $\mathbb{E}_i R^{\{i\}} < \infty$. If i is recurrent but not positive recurrent, then i is *null recurrent*.

Note that if i is positive recurrent, then $\mathbb{P}_i(R^{\{i\}} < \infty) = 1$, and thus, must be recurrent.

Theorem 3.6.4. Let P be irreducible and recurrent. Suppose that i is a positive recurrent state.

$$\mu_j := \frac{\gamma_j^{(i)}}{\mathbb{E}_i R^{\{i\}}}, \quad j \in I. \tag{73}$$

The above is an invariant distribution for P.

Proof. By property (iv) of of Theorem 25, $\mu = (\mu_j)_{j \in I}$ is a probability distribution. Thus, the result follows from (iii). □

Theorem 3.6.5. Let a transition matrix P be irreducible. Then the following are equivalent.

(i) Every state is positive recurrent.
(ii) Some state is positive recurrent.
(iii) P has an invariant distribution π.

Additionally, when (iii) holds,

$$\mathbb{E}_i R^{\{i\}} = 1/\pi_i \text{ for all } i \in I. \tag{74}$$

Example 3.6.2. Let there be a Markov chain on a state space $I = \{1, 2, 3, 4\}$. The transition matrix is given as follows.

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

Let μ be an invariant probability distribution.

$$\mu_1 = \frac{1}{2}\mu_2 + \frac{1}{2}\mu_4 \quad \mu_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_3$$
$$\mu_3 = \frac{1}{2}\mu_2 + \frac{1}{2}\mu_4 \quad \mu_4 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_3$$

This enforces the following probability distribution: $\mu = (1/4, 1/4, 1/4, 1/4)$.

3.7 Aperiodicity & Convergence

Consider

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

We have $P^{2n} = I$ and $P^{2n+1} = P$ for all integer $n \geq 0$. Clearly, $p_{ij}^{(n)}$ does not converge for all i, j. In this case, the states are *not* aperiodic.

Definition 3.7.1 (Aperiodicity). A state i is *aperiodic* if $p_{ii}^{(n)} > 0$ for all sufficiently large n.

Lemma 3.7.1. Suppose P is irreducible with an aperiodic state i. Then for all states j and k, we have that $p_{jk}^{(n)} > 0$ for all sufficiently large n. Consequently, all states are aperiodic

Proof. Let $j, k \in I$. Since P is irreducible, for some $r, s \geq 0$, we have $p_{ji}^{(r)} > 0$ and $p_{ik}^{(s)} > 0$. Then,

$$p_{ij}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ik}^{(s)} > 0$$

for all sufficiently large n. ☐

Recall that if P is irreducible and positive recurrent, then there exists a unique invariant distribution $\pi = (\pi_i)_{i \in I}$ given by

$$\pi_i := \frac{1}{\mathbb{E}_i R_i}, \quad R_i = R^{\{i\}} := \inf\{t \geq 1 \mid X_t = i\}.$$

Theorem 3.7.1. Let $(X_n)_{n\geq 0} \sim \text{Markov}(\lambda, P)$. Suppose P is irreducible and positive recurrent with a unique invariant distribution $\pi = (\pi_i)_{i \in I}$. If P is aperiodic, then

$$\mathbb{P}(X_n = j) \to \pi_j \text{ as } n \to \infty \text{ for all } j$$

In particular, for any fixed $i \in I$, $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$.

We write

$$F_n(i) = \frac{1}{n}\sum_{t=0}^{n-1} \mathbb{1}_{\{X_t=i\}}, \quad i \in I$$

as the fraction of time the Markov chain stays at the state i in the time interval 0 to $n-1$. Then, we have

$$\mathbb{E}F_n(i) = \frac{1}{n}\sum_{t=0}^{n-1} \mathbb{P}(X_t = i) \to \pi_i$$

as $n \to \infty$ if P is irreducible and aperiodic with invariant distribution π. Thus, in this case, the asymptotic fraction of times the Markov chain spends at i is π_i. We will prove Theorem 29 using the *coupling technique*.

Definition 3.7.2. Let Λ be an index set such as $\{0,1,2,\ldots\}$ or $[0,\infty)$. A coupling between two stochastic processes $(X_i)_{i\in\Lambda}$ and $(Y_i)_{i\in\Lambda}$ is a joint distribution $\left((\widehat{X}_i,\widehat{Y}_i)\right)_{i\in\Lambda}$ such that

$$\left(\widehat{X}_i\right)_{i\in\Lambda} \sim (X_i)_{i\in\Lambda} \text{ and } \left(\widehat{Y}_i\right)_{i\in\Lambda} \sim (Y_i)_{i\in\Lambda}.$$

The goal of such a coupling is to define the processes in the same space. There is no unique coupling.

Proof. Consider two independent discrete-time Markov chains

$$(X_n)_{n\ge 0} \sim \mathrm{Markov}(\lambda,P), \quad (Y_n)_{n\ge 0} \sim \mathrm{Markov}(\pi,P).$$

Observe that for any $n \ge 0$, the distribution of Y_n is π, that is,

$$\mathbb{P}(Y_n = j) = \pi_j, \quad j \in I.$$

since π is an invariant distribution. Let

$$\widetilde{\lambda}_{ij} := \lambda_i \pi_j, \quad (i,j) \in I \times I$$

and

$$\widetilde{p}_{ij,kl} := p_{ik}p_{jl}, \quad (i,j),(k,l) \in I \times I.$$

Define a stochastic process W on $I \times I$ as $W_n = (X_n, Y_n)$. Since X and Y are independent, we obtain $(W_n)_{n\ge 0} \sim \mathrm{Markov}(\widetilde{\lambda},\widetilde{P})$. We claim that W is irreducible, aperiodic, and positive recurrent. Since P is irreducible, for any states $(i,j),(k,l) \in I \times I$, there exist $n_1, n_2 \ge 0$ such that $p_{ik}^{(n_1)} > 0$ and $p_{jl}^{(n_2)} > 0$. Taking $n = \max(n_1,n_2)$, we have $\widetilde{p}_{ij,kl}^{(n)} \ge p_{ik}^{(n)} p_{jl}^{(n)} > 0$, so W is irreducible. Since P is aperiodic, there exists N such that $p_{ii}^{(n)} > 0$ for all $n \ge N$ and all $i \in I$. Then $\widetilde{p}_{ij,ij}^{(n)} = p_{ii}^{(n)} p_{jj}^{(n)} > 0$ for all $n \ge N$, so W is aperiodic. Since I is finite (or P is positive recurrent), $I \times I$ is also finite, so W is positive recurrent. The invariant distribution of W is $\widetilde{\pi}_{ij} = \pi_i\pi_j$. By the convergence theorem for irreducible, aperiodic, positive recurrent chains:

$$\mathbb{P}(W_n = (i,j)) \to \pi_i\pi_j \text{ as } n \to \infty.$$

Now, observe that $\mathbb{P}(X_n = i, Y_n = j) = \mathbb{P}(X_n = i)\mathbb{P}(Y_n = j) = \mathbb{P}(X_n = i)\pi_j$ by independence. Therefore:

$$\mathbb{P}(X_n = i)\pi_j \to \pi_i \pi_j \text{ as } n \to \infty.$$

Since $\pi_j > 0$ for all j (by irreducibility and positive recurrence), we can divide by π_j to get:

$$\mathbb{P}(X_n = i) \to \pi_i \text{ as } n \to \infty.$$

This completes the proof. □

3.8 Time Reversal & Detailed Balance

Theorem 3.8.1. Let P be irreducible with an invariant distribution π, and let $(X_n)_{n\geq 0} \sim \text{Markov}(\pi, P)$. Take $N \geq 1$ and set $Y_n = X_{N-n}$ for $n = 0, 1, \ldots, N$. Then, we have

$$(Y_n)_{n=0}^N \sim \text{Markov}\left(\pi, \widehat{P}\right)$$

where $\widehat{p}_{ij} = \pi_j p_{ij} / \pi_i$ for $i, j \in I$. Moreover, π is an invariant distribution for $\widehat{P}$.

The chain $(Y_n)_{0\leq n\leq N}$ is called the *time-reversal* of $(X_n)_{0\leq n\leq N}$.

Proof. First, we check that $\widehat{P}$ is a transition matrix. We have $\widehat{p}_{ij} \geq 0$ for all i, j and

$$\sum_{j\in I} \widehat{p}_{ij} = \frac{1}{\pi_i} \sum_{j\in I} \pi_j p_{ji} = \frac{1}{\pi_i} \pi_i = 1.$$

since π is invariant for P. Next, we check that $\sum_{i\in I} \pi_i \widehat{p}_{ij} = \pi_j$ for all i. Indeed, we have that

$$\sum_{i\in I} \pi_i \widehat{p}_{ij} = \sum_{i\in I} \pi_i \cdot \frac{\pi_j p_{ji}}{\pi_i} = \sum_{i\in I} \pi_j p_{ji} = \pi_j \sum_{i\in I} p_{ji} = \pi_j.$$

Now we verify that $(Y_n)_{n=0}^N$ is indeed a Markov chain with transition matrix $\widehat{P}$. For any $0 \leq n \leq N-1$ and states $i_0, i_1, \ldots, i_n, j \in I$, we need to show:

$$\mathbb{P}(Y_{n+1} = j \mid Y_0 = i_0, Y_1 = i_1, \ldots, Y_n = i_n) = \widehat{p}_{i_n j}.$$

Since $Y_k = X_{N-k}$, this is equivalent to:

$$\mathbb{P}(X_{N-n-1} = j \mid X_N = i_0, X_{N-1} = i_1, \ldots, X_{N-n} = i_n) = \widehat{p}_{i_n j}.$$

Using Bayes' theorem and the Markov property of (X_n):

$$\begin{aligned}
&\mathbb{P}(X_{N-n-1} = j \mid X_N = i_0, \ldots, X_{N-n} = i_n) \\
&= \frac{\mathbb{P}(X_{N-n-1} = j, X_{N-n} = i_n, \ldots, X_N = i_0)}{\mathbb{P}(X_{N-n} = i_n, \ldots, X_N = i_0)} = \frac{\mathbb{P}(X_{N-n-1} = j) p_{j,i_n} p_{i_n,i_{n-1}} \cdots p_{i_1,i_0}}{\mathbb{P}(X_{N-n} = i_n) p_{i_n,i_{n-1}} \cdots p_{i_1,i_0}} \\
&= \frac{\pi_j p_{j,i_n}}{\pi_{i_n}} = \widehat{p}_{i_n j}.
\end{aligned}$$

Finally, since $(X_n) \sim \mathbf{Markov}(\pi, P)$, we have $\mathbb{P}(Y_0 = i) = \mathbb{P}(X_N = i) = \pi_i$, so (Y_n) starts with distribution π. □

Definition 3.8.1. A Markov chain with transition matrix P and positive invariant distribution π is called *reversible* if $\widehat{P} = P$.

Definition 3.8.2. A distribution λ and a transition matrix P are said to be in *detailed balance* if

$$\lambda_i p_{ij} = \lambda_j p_{ji} \quad \text{for all } i, j \in I.$$

Theorem 3.8.2. Suppose that (λ, P) is in detailed balance. Then λ is an invariant distribution for P. Moreover, if $(X_n)_{n \geq 0} \sim \mathbf{Markov}(\lambda, P)$, then $(X_n)_{n \geq 0}$ is reversible.

Let us consider random walks on locally finite graphs. Let $G = (V, E)$ be an undirected connected graph. For $v \in V$, the degree of v is defined as

$$\deg(v) = |\{u \in V : (u, v) \in E\}|.$$

We say that G is *locally finite* if $\deg(v) < \infty$ for all $v \in V$. A locally finite graph G is *finite* if V is finite. Let us assume that G is locally finite. For $u, v \in V$, we write $u \sim v$ if $(u, v) \in E$, and $u \nsim v$ if $(u, v) \notin E$.

$$p_{uv} = \begin{cases} \frac{1}{\deg(u)} & \text{if } u \sim v \\ 0 & \text{if } u \nsim v \end{cases}$$

It is easy to see that $P = (p_{uv})$ is a stochastic matrix on $V \times V$. Additionally, since G is connected, P is irreducible. A Markov chain X on G corresponding to P is called a *simple random walk* on G. For finite G,

$$\lambda_u = \frac{\deg(u)}{2|E|}, \quad u \in V$$

Theorem 3.8.3. If G is finite, then any random walk on G is positive recurrent with the unique invariant distribution λ.

It suffices to show that (λ, P) is in detailed balance.

Proof. We wish to show that $\lambda_u p_{uv} = \lambda_v p_{vu}$ for all u, v. If $(u, v) \notin E$, then both are 0. If $(u, v) \in E$, then

$$\lambda_u p_{uv} = \frac{\deg(u)}{2|E|} \cdot \frac{1}{\deg(u)} = \frac{1}{2|E|} = \lambda_v p_{vu}$$

□

If G is infinite, then any random walk on G is either null recurrent or transient. Let us consider a random walk on the integer lattice $\mathbb{Z}^d$. If $d \leq 2$, then the walk is null recurrent. If $d \geq 3$, then it is transient.

3.9 Ergodic Theory & Sampling Algorithms

Ergodic theorems concern the limiting behavior of averages over time. The following theorem is vital in this context.

Theorem 3.9.1 (The Strong Law of Large Numbers). Let $Y_0, Y_1, Y_2, \dots$ be a sequence of i.i.d non-negative random variables with mean $\mathbb{E} Y_1 = \mu \in [0, \infty)$. Then

$$\frac{1}{n} \sum_{k=0}^{n-1} Y_k \to \mu \text{ as } n \to \infty \text{ almost surely.}$$

That is,

$$\mathbb{P}\left(\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} Y_k = \mu \right) = 1.$$

For any state $i \in I$ and $n \geq 1$, we define

$$V_i(n) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = i\}}.$$

Then $V_i(n)/n$ is the proportion of times before n spent in state i.

Theorem 3.9.2 (The Ergodic Theorem). Assume that P is irreducible. Then

$$\frac{V_i(n)}{n} \to \frac{1}{\mathbb{E}_i R_i} \text{ as } n \to \infty \text{ almost surely.}$$

Moreover, if P is positive recurrent, then for any bounded function $f : I \to \mathbb{R}$, we have

$$\frac{1}{n} \sum_{t=0}^{n-1} f(X_t) \to \sum_{x \in I} f(x) \pi_x \text{ as } n \to \infty \text{ almost surely.}$$

where π is the invariant distribution.

Suppose that I is finite and π exists on I. We want to generate a sample from π and, for a given function f, approximate

$$\pi f := \sum_{x \in I} \pi_x f(x)$$

Suppose we know a transition matrix P on $I \times I$, which is easy to sample from. Moreover, suppose P is irreducible, aperiodic, and positive recurrent. Under P, the chain moves $x \to y$ with probability p_{xy}. We now describe the Metropolis-Hastings algorithm to sample from π.

Algorithm 1 Metropolis-Hastings Algorithm

Require: Target distribution π, proposal matrix P, number of steps N

Ensure: Sample sequence $(X_0, X_1, \ldots, X_N)$

Initialize $X_0 \in I$ arbitrarily

for $t = 0, 1, \ldots, N-1$ **do**

  Propose $Y \sim P(X_t, \cdot)$ ▷ i.e., $\mathbb{P}(Y = y) = p_{X_t, y}$

  Compute acceptance probability:

  $\alpha(X_t, Y) \leftarrow \min\left\{1, \frac{\pi_Y p_{Y, X_t}}{\pi_{X_t} p_{X_t, Y}}\right\}$

  Generate $U \sim \text{Uniform}(0, 1)$

  if $U \leq \alpha(X_t, Y)$ **then**

    $X_{t+1} \leftarrow Y$ ▷ accept the proposal

  else

    $X_{t+1} \leftarrow X_t$ ▷ reject the proposal

  end if

end for

return $(X_0, X_1, \ldots, X_N)$

The key insight is that this algorithm constructs a Markov chain with transition probabilities:

$$\tilde{p}_{xy} = \begin{cases} p_{xy}\alpha(x, y) & \text{if } x \neq y \\ p_{xx} + \sum_{z \neq x} p_{xz}(1 - \alpha(x, z)) & \text{if } x = y \end{cases}$$

where the acceptance probability $\alpha(x, y)$ ensures that the detailed balance condition $\pi_x \tilde{p}_{xy} = \pi_y \tilde{p}_{yx}$ is satisfied, making π the invariant distribution of the constructed chain.

Example 3.9.1 (Uniform Sampling on Graphs). Let $G = (V, E)$ be a connected finite graph with large $|V|$. Our goal is to generate a sample uniformly at random from V, i.e., $\pi_x = 1/|V|$ for all $x \in V$. We can use simple random walk as our proposal distribution with transition matrix P given by $p_{xy} = 1/d_x$ for $x \sim y$ (where d_x is the degree of vertex x), and $p_{xy} = 0$ otherwise.

For adjacent vertices $x \neq y$ (i.e., $x \sim y$), the acceptance probability becomes:

$$\alpha(x, y) = \min\left\{1, \frac{\pi_y p_{y,x}}{\pi_x p_{x,y}}\right\} = \min\left\{1, \frac{(1/|V|) \cdot (1/d_y)}{(1/|V|) \cdot (1/d_x)}\right\} = \min\left\{1, \frac{d_x}{d_y}\right\}$$

This gives us the Metropolis-Hastings transition probabilities:

$$\tilde{p}_{xy} = \begin{cases} \frac{1}{d_x} \min\left\{1, \frac{d_x}{d_y}\right\} = \frac{1}{\max\{d_x, d_y\}} & \text{if } x \sim y \\ 0 & \text{if } x \not\sim y, x \neq y \\ 1 - \sum_{z \sim x} \tilde{p}_{xz} & \text{if } x = y \end{cases}$$

This chain has uniform distribution π as its invariant distribution and tends to be more efficient than simple random walk on graphs with varying vertex degrees, as it reduces the tendency to get trapped at high-degree vertices.

4 Continuous-Time Markov Chains

4.1 Equivalent Definitions

Let I be a countable set. A *continuous-time Markov chain* (CTMC) on the state space I is a *right-continuous* stochastic process $(X_t)_{t\geq 0}$ taking values in I, indexed by continuous time $[0,\infty)$ and defined using

(i) an initial probability distribution λ on I

(ii) a rate matrix $Q = (q_{ij})_{i,j\in I}$ such that

  (a) $q_{ij} \geq 0$ for all $i \neq j$ and

  (b) $\mathbf{Q1} = \mathbf{0}$. That is, $q_{ii} = -\sum_{j\neq i} q_{ij}$ for all $i \in I$.

The right-continuity is needed to ensure that that finite-dimensional distributions characterize the joint distribution of the whole process.

$$\mathbb{P}\left(X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n\right) = \mathbb{P}\left(Y_{t_1} = x_1, Y_{t_2} = x_2, \dots, Y_{t_n} = x_n\right)$$

for all $n \geq 1$ and $0 \leq t_1 < t_2 < \cdots < t_n$ implies that $(X_t)_{t\geq 0} \overset{d}{=} (Y_t)_{t\geq 0}$. There are three equivalent ways of defining a CTMC $(X_t)_{t\geq 0} \sim \text{Markov}(\lambda, Q)$.

For given positive functions f, g, the notation $f(h) = o(g(h))$ means that $\lim_{h\to 0} f(h)/g(h) = 0$.

Definition 4.1.1 (The Infinitesimal Generator). We say that, a right-continuous stochastic process $(X_t)_{t\geq 0}$ taking values in I is a CTMC with initial distribution λ and rate matrix Q if

(i) $\mathbb{P}(X_0 = x) = \lambda_x$ for all $x \in I$,

(ii) X_t jumps from state i to j with rate $q_{ij} \geq 0$, i.e.

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = \begin{cases} hq_{ij} + o(h) & \text{if } i \neq j \\ 1 + hq_{ii} + o(h) & \text{if } i = j \end{cases}$$

as $h \to 0$ for all $t \geq 0$.

If we write the transition matrix as

$$P(t) = (\mathbb{P}(X_t = j \mid X_0 = i))_{i,j\in I},$$

then the above definition says that

$$P(h) = I + hQ + o(h) \text{ as } h \to 0.$$

Note that, this shows that we definitely need $\mathbf{Q1} = \mathbf{0}$ since $P(h)\mathbf{1} = \mathbf{1}$ for all $h \geq 0$.

In the discrete example, the transition matrix at time n, given the state at 0, is given by P^n. For continuous time, we want to define P^t for real $t \geq 0$. Note that, for a positive real number $p > 0$, we can define p^t with $e^{t\log p}$. Here, Q is analogous to $\log p$. First, let us define some linear algebra machinery.

Definition 4.1.2 (Spectral Norm). The *spectral norm* of a matrix A is defined as

$$||A|| := \sup_{||v||=1} ||A\mathbf{v}||, \quad \text{where } ||\mathbf{v}|| = \left(\sum_{i\in I} v_i^2\right)^{1/2}.$$

Definition 4.1.3 (The Matrix Exponential). For a matrix A with $||A|| \le \infty$, the exponential of A is defined as

$$e^A := I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}A^n.$$

Note that for commuting matrices A, B, we have $e^{A+B} = e^A e^B$. This holds with the polynomial expansion of the exponential function. Additionally, $\left(e^A\right)^{-1} = e^{-A}$.

Theorem 4.1.1. Let I be a countable set. An $I \times I$ matrix Q with $||Q|| < \infty$ is a rate matrix if and only if e^{tQ} is a stochastic matrix for all $t \ge 0$.

Proof. ($\Longrightarrow$) All entries in e^{tQ} are non-negative. Further, using $Q\mathbf{1} = \mathbf{0}$, we have

$$e^{tQ}\mathbf{1} = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k \mathbf{1} = Q^0 \mathbf{1} = \mathbf{1}$$

Hence, e^{tQ} is a stochastic matrix for all $t \ge 0$.
($\Longleftarrow$) Suppose that e^{tQ} is a stochastic matrix for all $t \ge 0$. Since $Q = \lim_{t\to 0} \frac{e^{tQ}-I}{t}$, we have $q_{ij} \ge 0$ for all $i \ne j$. Further, since $e^{tQ}\mathbf{1} = \mathbf{1}$ for all $t \ge 0$, we have

$$Q\mathbf{1} = \lim_{t\to 0} \frac{e^{tQ} - I}{t}\mathbf{1} = \lim_{t\to 0} \frac{e^{tQ}\mathbf{1} - \mathbf{1}}{t} = \lim_{t\to 0} \frac{\mathbf{1} - \mathbf{1}}{t} = \mathbf{0} \quad \text{so, } e^{tQ} \text{ is stochastic.}$$

□

Theorem 4.1.2. Let Q be a rate matrix on I with $||Q|| < \infty$. We define $P(t) := e^{tQ}$. Then, $(P(t))_{t\ge 0}$ satisfies the following properties:

(i) $P(t+s) = P(t)P(s)$ for all $t, s \ge 0$. This is the semi-group property.

(ii) $(P(t))_{t\ge 0}$ is the unique solution to the forward equation

$$\frac{d}{dt}P(t) = P(t)Q \text{ for } t > 0, \quad P(0) = I. \quad \textit{the forward equation.}$$

(iii) $(P(t))_{t\ge 0}$ is the unique solution to the backward equation

$$\frac{d}{dt}P(t) = QP(t) \text{ for } t > 0, \quad P(0) = I. \quad \textit{the backward equation.}$$

Let us consider the *Poisson process*. Let $I = \{0,1,\dots\}$ and $\lambda > 0$ be constant. Consider the rate matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Observe that $||Q|| \le 2\lambda < \infty$. Indeed, for any $\mathbf{v} = (v_i)_{i\in I}$ with $||\mathbf{v}|| = 1$, we have

$$||Q\mathbf{v}||^2 = \sum_{i\ge 0} (-\lambda v_i + \lambda v_{i+1})^2 = 2\lambda^2 \sum_{i\ge 0} \left(v_i^2 + v_{i+1}^2\right) \le 4\lambda^2 \sum_{i\ge 0} v_i^2 = 4\lambda^2.$$

So, $P(t) := e^{tQ}$ is well-defined for all $t \ge 0$. By the forward equation, we have $P'(t) = P(t)Q$. It follows that for all $i, j \in I$,

$$P'_{i,j}(t) = \sum_{k\in I} P_{i,k}(t) Q_{k,j} = -\lambda P_{i,j}(t) + \lambda P_{i,j-1}(t), \quad P_{i,j}(0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

With induction, one can show that

$$P_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} \cdot \mathbb{1}_{i\le j} = \mathbb{P}(\text{Poisson}(\lambda t) = j - i)$$

Definition 4.1.4 (The Transition Matirx). We say that a right-continuous stochastic process $(X_t)_{t\ge 0}$ taking values in I is a CTMC with initial distribution λ and rate matrix Q if

$$\mathbb{P}\left(X_{t_0} = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n\right) = \lambda_{x_0} \prod_{i=0}^{n} P_{x_i x_{i+1}}(t_{i+1} - t_i)$$

4.2 Explosion Times & Minimal Chains

Let $(X_t)_{t\ge 0}$ be a CTMC taking values in I. The following three properties hold.

(i) $t \mapsto X_t(\omega)$ is right-continuous in $[0,\infty)$ almost surely.

(ii) Conditional on $\{X_t = i\}$, $(X_s)_{s\ge t}$ is independent of $X(_r)_{r\ge t}$. This is the *Markov property*.

(iii) $\mathbb{P}(X_{t+h} = j \mid X_t = i) = \delta_{ij} + q_{ij} h + o(h)$ as $h \to 0$.

Definition 4.2.1 (Explosion Time). Let $X = (X_t)_{t\ge 0}$ be a CTMC constructed from its jump chain and holding times. The explosion time ζ is defined as

$$\zeta = \lim_{n\to\infty} J_n$$

where J_n is the time of the n-th jump.

Theorem 4.2.1. If any of the following hold, we have $\mathbb{P}(\zeta = \infty) = 1$.

(i) $\sup_{i\in I} q_i < \infty$.

(ii) I is finite.

(iii) $X_0 = i$ and i is recurrent for the jump chain.

Proof. Let $M := \sup_{i\in I} q_i < \infty$. We have $T_1, T_2, \cdots \overset{iid}{\sim} \mathrm{Exp}(1)$. By a previously proven theorem, $\sum_{k=1}^{\infty} T_k = \infty$ almost surely, and thus

$$\zeta = \lim_{n\to\infty} J_n = \sum_{k=1}^{\infty} T_k/q_{Y_{k-1}} \geq \sum_{k=1}^{\infty} T_k/M = \infty \quad \text{almost surely.}$$

If I is finite, then $\sup_{i\in I} q_i < \infty$ trivially. We know that $(Y_n)_{n\geq 0}$ visits i infinitely often, at times $R_1^i, R_2^i, \ldots$, say. If $q_i = 0$, then $\zeta \geq J_1 = \infty$. If $q_i > 0$, then $\zeta \geq q_i^{-1} \sum_{k=1}^{\infty} T_{R_k^i} = \infty$. □

Let us fix a parameter $\theta > 0$ and define $z_i = \mathbb{E}_i\, e^{-\theta\zeta}$ for $i \in I$. Then, $\mathbf{z} = (z_i)_{i\in I}$ satisfies

(i) $|z_i| \leq 1$ and

(ii) $Q\mathbf{z} = \theta\mathbf{z}$.

Moreover, if $\hat{\mathbf{z}}$ satisfies both of these, then $|\hat{z}_i| \leq z_i$ for all i.

Proof. Since $\theta > 0$, it holds that $|z_i| \leq 1$. Recall that $S_n = T_n/q_{Y_{n-1}}$ with $T_n \overset{iid}{\sim} \mathrm{Exp}(1)$. We have for all i,

$$z_i = \mathbb{E}_i\left(e^{-\theta\sum_{n=1}^{\infty} S_n}\right) = \sum_{k\neq i} \mathbb{E}_i\left(e^{-\theta S_1} \mathbb{1}_{Y_1=k} \cdot \mathbb{E}_k\left(e^{-\theta\zeta}\right)\right) = \sum_{k\neq i} \mathbb{E}_i\left(e^{-\theta T_1/q_i}\right) \cdot \Pi_{ik} z_k.$$

Since $\mathbb{E}\left(e^{-cT}\right) = \int_0^{\infty} e^{-cx} e^{-x}\, dx = (1+c)^{-1}$ for all $c \geq 0$, it follows that

$$z_i = \frac{q_i}{\theta + q_i} \sum_{k\neq i} \frac{q_{ik}}{q_i} z_k = \frac{1}{\theta + q_i} \sum_{k\neq i} q_{ik} z_k = \frac{1}{\theta + q_i}\left((Q\mathbf{z})_i - q_{ii} z_i\right) = \frac{1}{\theta + q_i}\left((Q\mathbf{z})_i + q_i z_i\right).$$

This proves $Q\mathbf{z} = \theta\mathbf{z}$. For the second assertion, it is enough to show that for all $n \geq 0$,

$$|\hat{z}_i| \leq \mathbb{E}_i\, e^{-\theta J_n} \text{ for all } i \in I.$$

If the above equation holds, then by taking the limit as $n \to \infty$, since $\lim_{n\to\infty} \mathbb{E}_i\, e^{-\theta J_n} = z_i$ by the monotone convergence theorem, we are done. The above equation clearly holds if $n = 0$. We induct on n and obtain:

$$\begin{aligned} |\hat{z}_i| &= \frac{1}{\theta + q_i}\left(|(Q\hat{\mathbf{z}})_i| + q_i|\hat{z}_i|\right) \leq \frac{1}{\theta + q_i} \sum_{k\neq i} q_{ik}|\hat{z}_k| \leq \frac{1}{\theta + q_i} \sum_{k\neq i} q_{ik} \mathbb{E}_k\, e^{-\theta J_N} \\ &= \frac{q_i}{\theta + q_i} \sum_{k\neq i} \Pi_{ik} \mathbb{E}_k\, e^{-\theta J_N} = \sum_{k\neq i} \mathbb{E}_i\left(e^{-\theta S_1} \mathbb{1}_{Y_1=k} \cdot \mathbb{E}_k\left(e^{-\theta J_N}\right)\right) = \mathbb{E}_i\, e^{-\theta J_{N+1}}. \end{aligned}$$

The proof is complete by induction. □

Theorem 4.2.2. The following statements are equivalent.

(a) $\mathbb{P}(\zeta = \infty) = 1$ for all $i \in I$.

(b) If $Q\mathbf{z} = \mathbf{z}$ and $\sup_{i\in I} |z_i| \leq 1$, then $\mathbf{z} = \mathbf{0}$.

(c) If $Q\mathbf{z} = \theta\mathbf{z}$ for some $\theta > 0$, and $\sup_{i\in I} |z_i| < 1$, then $\mathbf{z} = \mathbf{0}$.

We say a rate matrix Q is *non-explosive* if $Q\mathbf{z} = \theta\mathbf{z}$ for some $\theta > 0$ and $\sup_{i \in I} |z_i| \le 1$ imply $\mathbf{z} = \mathbf{0}$. Otherwise, Q is *explosive.* A CTMC X is non-explosive or *conservative* if

$$\mathbb{P}_i(\zeta = \infty) = 1 \text{ for all } i \in I.$$

Note that for a CTMC X with rate matrix Q, X is explosive if and only if Q is. A process X is a *minimal chain* if it is either non-explosive or satisfies

$$X_t = \partial \quad \text{for all } t \ge \zeta$$

where ∂ is an added state called the *cemetery state*.

notes on jump chains, stopping times, and birth (poisson) processes are missing, will fill in soon.

4.3 Class Structure, Recurrence, and Transience

Consider a CTMC $(X_t)_{t \ge 0} \sim \text{Markov}(\lambda, Q)$ on I. Let $P(t) = \big(p_{ij}(t)\big)_{i,j \in I}$ be the transition matrix of X.

$$p_{ij}(t) := \mathbb{P}(X(t) = j \mid X(0) = i).$$

Definition 4.3.1. We say that i *leads to* j, $i \to j$ if

$$\mathbb{P}_i(X_t = j \text{ for some } t \ge 0) > 0.$$

We say that i *communicates* with j, $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Communication is an equivalence relation and I can be decomposed into disjoint communicating classes. The notions of closed classes, absorbing states, and irreducibility are all inherited from the jump chains.

Theorem 4.3.1. For distinct $i, j \in I$, the following are equivalent.

(i) $i \to j$.

(ii) $i \to j$ for the jump chain.

(iii) $q_{ik_1} q_{ik_1,k_2} \cdots q_{k_n j} > 0$ for some states $k_1, \ldots, k_n \in I$.

(iv) $p_{ij}(t) > 0$ for some $t \ge 0$.

(v) $p_{ij}(t) > 0$ for all $t \ge 0$.

The CTMC $X(t)_{t \ge 0}$ is irreducible if and only if $p_{ij}(t) > 0$ for all $i, j \in I$ and $t \ge 0$. We now discuss hitting times and return times for a minimal Markov chain X with rate matrix Q.

Definition 4.3.2 (Hitting Time). For a subset $A \subseteq I$, the *hitting time* of A is defined as

$$D^A := \inf\{t \ge 0 \mid X_t \in A\}.$$

We will use H^A to denote the hitting time for the jump chain $(Y_n)_{n \ge 0}$.

Definition 4.3.3 (Return Time). For a subset $A \subseteq I$, the *return time* of A is defined as

$$R^A := \inf\{t \ge J_1 \mid X_t \in A\}.$$

Note that

$$D^A = \begin{cases} J_{H^A} & \text{if } H^A < \infty \\ \infty & \text{otherwise} \end{cases}$$

Proof. We use the fact that $X_t = Y_n$ if $J_n \le t \le J_{n+1}$. If H^A is the hitting time of A for the jump chain, then $\{H^A < \infty\} = \{D^A < \infty\}$ and, on this set, we have $D^A = J_{H^A}$. □

Let us define the hitting probabilities and the expected hitting time, respectively, as

$$h_i^A = \mathbb{P}_i\left(D^A < \infty\right) = \mathbb{P}_i\left(H^A < \infty\right), \quad k_i^A = \mathbb{E}_i D^A.$$

We write $h^A := \left(h_i^A\right)_{i\in I}$ and $k^A := \left(k_i^A\right)_{i\in I}$. Note that h^A is a vector of probabilities and k^A is a vector of expected hitting times.

Theorem 4.3.2. The following are true.

(i) h^A is the minimal non-negative solution to the following equation

$$\begin{cases} h_i^A = 1 & \text{if } i \in A \\ \sum_{j\in I} q_{ij} h_j^A = 0 & \text{if } i \notin A \end{cases}$$

(ii) Assume $q_i > 0$ and $k_i^A < \infty$ for all $i \notin A$. Then k^A is minimal non-negative solutions to the following equation.

$$\begin{cases} k_i^A = 0 & \text{if } i \in A \\ -\sum_{j\in I} q_{ij} k_j^A = 1 & \text{if } i \notin A \end{cases}$$

Definition 4.3.4 (Recurrence & Transience). We stay that a state is *recurrent* if

$$\mathbb{P}_i\left(\{t \ge 0 \mid X_t = i\} \text{ is unbounded}\right) = 1.$$

We say that a state is *transient* if

$$\mathbb{P}_i\left(\{t \ge 0 \mid X_t = i\} \text{ is unbounded}\right) = 0.$$

Dichotomy is defined as the following holding.

(a) If $q_i = 0$ or i is recurrent for the jump chain, then i is recurrent for the CTMC and

$$\int_0^\infty p_{ii}(t)\,dt = \infty.$$

(b) If $q_i > 0$ and i is transient for the jump chain, then i is transient for the CTMC and

$$\int_0^\infty p_{ii}(t)\,dt < \infty.$$

Consequently, i is recurrent for the CTMC if and only if $\int_0^\infty p_{ii}(t)\,dt = \infty$.

4.4 Invariant Measures

Let $(X_t)_{t\geq 0} \sim \mathbf{Markov}(\lambda, Q)$ be an irreducible minimal CTMC on I. Let $\Pi = (\pi_{ij})$ be the transition matrix of its jump chain.

Definition 4.4.1. We say λ is invariant for X if $\lambda Q = 0$.

Theorem 4.4.1. A measure λ is invariant if and only if $(\lambda_i q_i)_{i\in I}$ is invariant for Π.

This connection with invariant measures for the jump matrix allows us to apply existence and uniqueness results for discrete-time Markov chains to establish the following result.

Theorem 4.4.2. Suppose that Q is irreducible and recurrent. Then, Q has an invariant measure λ which is unique up to a multiplicative constant.

Let us introduce some notation. The first return time to state i for the corresponding jump chain $(Y_n)_{n\geq 0}$ is defined as

$$N_i := \inf\{n \geq 1 \mid Y_n = i\}.$$

An invariant measure for Π normalized so that $\gamma_i^i = 1$

$$\gamma_j^i = \mathbb{E}_i\left[\sum_{n=0}^{N_i-1} \mathbb{1}_{\{Y_n=j\}}\right], \quad j \in I.$$

The first return time to state i for the CTMC $(X_t)_{t\geq 0}$ is defined as

$$R_i := \inf\{t \geq J_1 \mid X_t = i\}.$$

An invariant measure for Q normalized that so that $\mu_i^i = \mathbb{E}_i[J_1] = 1/q_i$ is given by

$$\mu_j^i = \mathbb{E}_i\left[\int_0^{R_i} \mathbb{1}_{\{X_t=j\}}\,dt\right], \quad j \in I.$$

wip. still missing some theorems on time reversal, convergence, and branching processes.