Notes on Discrete Markov Chains

Yangxinyu Xie

Summer 2019

1 Introduction
   1.1 Tail Probability .................................................. 2
   1.2 Generating Functions .......................................... 2
   1.3 Stirling’s formula ............................................... 2
   1.4 Stochastic Process ............................................. 2
   1.5 Markov chain ................................................... 3
   1.6 Chapman-Kolmogorov Equations .............................. 3

2 Classification of States ........................................ 4
   2.1 Classification of States ....................................... 4
   2.2 Classification of Chains ..................................... 7

3 Stationary Distributions and Limiting Probabilities .......... 8
   3.1 Stationary Distribution ....................................... 8
   3.2 Limiting Probabilities .................................... 12

4 Time Reversible Markov Chains ................................ 13

Last updated January 11, 2020
1. Introduction

1.1 Tail Probability

Theorem 1.1. If $X$ is a non-negative integer-valued random variable, then

$$E[X] = \sum_{k=0}^{\infty} P[X = k] \cdot k = \sum_{k=1}^{\infty} P[X \geq k]$$  \hspace{1cm} (1.1.1)

1.2 Generating Functions

Definition 1.2. For a sequence $a_0, a_1, a_2, ... \in \mathbb{R}$, if there exists $s_0 > 0$ such that

$$G(s) := \sum_{i=0}^{\infty} a_i s^i$$  \hspace{1cm} (1.2.1)

converges when $|s| \leq s_0$, then we call $G(s)$ the ordinary generating function of the sequence $\{a_i\}$. If the sum converges if $|s| \leq s_0$ and diverges if $|s| > s_0$, we call $s_0$ the radius of convergence.

Definition 1.3. The convolution of two sequences $\{a_n, n \geq 0\}$ and $\{b_n, n \geq 0\}$ is the new sequence $\{c_n, n \geq 0\}$ whose $n$th element is given by

$$c_n = \sum_{i=0}^{n} a_i b_{n-i}$$  \hspace{1cm} (1.2.2)

we write $\{c_n\} = \{a_n\} \ast \{b_n\}$.

Theorem 1.4. If $\{a_n, n \geq 0\}$ and $\{b_n, n \geq 0\}$ have generating functions $G_a$ and $G_b$, then the generating function of $\{c_n, n \geq 0\} = \{a_n\} \ast \{b_n\}$ is

$$G_c(s) = \sum_{n=0}^{\infty} c_n s^n = \sum_{n=0}^{\infty} \left( \sum_{i=0}^{n} a_i b_{n-i} \right) s^n = \sum_{i=0}^{\infty} a_i s^i \sum_{n=i}^{\infty} b_{n-i} s^{n-i} = G_a(s)G_b(s)$$  \hspace{1cm} (1.2.3)

Theorem 1.5 (Abel’s theorem). If $a_i \geq 0$ for all $i$ and $G_a(s)$ is finite for $|s| < 1$, then

$$\lim_{s \uparrow 1} G_a(s) = \sum_{i=1}^{\infty} a_i$$  \hspace{1cm} (1.2.4)

whether the sum if finite or equals to $\infty$.

1.3 Stirling’s formula

The Stirling’s formula states that

$$n! = \sqrt{2\pi n} \left( \frac{n}{e} \right)^n \left( 1 + O\left( \frac{1}{n} \right) \right)$$  \hspace{1cm} (1.3.1)

Or

$$n! \approx \sqrt{2\pi n}^{n+0.5} e^{-n}$$  \hspace{1cm} (1.3.2)

where $a_n \approx b_n$ if and only if $\lim_{n \to \infty} a_n/b_n = 1$.

1.4 Stochastic Process

Definition 1.6. Let $T$ be a subset of $[0, \infty)$. A stochastic process is a collection of random variables $\{X(t) : t \in T\}$ which take values from some state space $S$. 

2
1.5 Markov chain

**Definition 1.7.** Let \( \{X_n, n = 0, 1, 2, \ldots\} \) be a stochastic process that takes on values from a finite or countable state space \( S \) with positive integers. If \( X_n = i \), then the process is said to be in state \( i \) at time \( n \). Suppose that
\[
P[X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0] = P[X_{n+1} = j|X_n = i] = p_{i,j}
\]
for all states \( i_0, i_1, \ldots, i, j \) and all \( n \geq 0 \). Such a stochastic process is called a **Markov chain**.

Note that the present state \( X_n \) is independent of the past states and depends only on the present state.

**Definition 1.8.** The Markov chain \( X \) is called **temporally homogeneous** if the transition probabilities \( p_{i,j} \) does not depend on time \( n \). That is
\[
P[X_{n+1} = j|X_n = i] = P[X_1 = j|X_0 = i]
\]
The **transition matrix** \( P = (p_{i,j}) \) is a \(|S| \times |S|\) matrix of transition probabilities

\[
P = \begin{bmatrix}
p_{00} & p_{01} & p_{02} & \cdots \\
p_{10} & p_{11} & p_{12} & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
p_{00} & p_{01} & p_{02} & \cdots \\
\vdots & \vdots & \vdots & \ddots 
\end{bmatrix}
\]

where for each row \( i = 0, 1, \ldots \), we have \( p_{ij} \geq 0, \sum_j p_{ij} = 1 \).

**Notation 1.9.** In this note, we restrict our attention to discrete, temporally homogeneous Markov chains unless otherwise specified.

1.6 Chapman-Kolmogorov Equations

We define \( p_{i,j}^n := P[X_n = j|X_0 = i], n \geq 0, i, j \geq 0 \), that is, starting from state \( i \), the process reaches state \( j \) in \( n \) steps. The **Chapman-Kolmogorov Equations** states that
\[
p_{i,j}^{n+m} = P[X_{n+m} = j|X_0 = i]
\]
\[
= \sum_{k=0}^{\infty} P[X_{n+m} = j, X_n = k|X_0 = i]
\]
\[
= \sum_{k=0}^{\infty} P[X_{n+m} = j|X_n = k, X_0 = i]P[X_n = k|X_0 = i]
\]
\[
= \sum_{k=0}^{\infty} p_{k,j}^n p_{i,k}^m = \sum_{k=0}^{\infty} p_{i,k}^m p_{k,j}^n
\]

Generalising this idea, we can let \( P^n \) denote the \( n \)-step transition matrix and have that \( P^{n+m} = P^n \cdot P^m \).

**Example 1.10.** Suppose we have 8 empty urns and we can distribute a ball to an urn in a uniformly random fashion. We are interested in the probability that after distributing 9 balls, there are exactly three non-empty urns. To do this, we first note that with probability 1, there will be one non-empty urn in the first run. Hence, we can construct a transition matrix with states \( 1, 2, 3, 4 \), where state \( i, i \in \{1, 2, 3\} \) denotes the number of non-empty urns and state 4 denotes the state that there are four or more non-empty urns.

\[
P = \begin{bmatrix}
0.125 & 0.875 & 0 & 0 \\
0 & 0.25 & 0.75 & 0 \\
0 & 0 & 0.375 & 0.625 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

We call state 4 an **absorbing state** since once the chain enters this state, there is no way getting out of it. Using Chapman-Kolmogorov Equations, the answer is thus \( P_{1,3}^8 = 0.00757 \).
Theorem 1.11 (Strong Markov Property). Let $X$ be a Markov chain on $S$, and let $T$ be a random variable taking values in $\{0, 1, 2, \ldots\}$ with the property that the indicator function $I\{T = n\}$, of the event that $T = n$, is a function of the variables $X_1, X_2, \ldots, X_n$. Such a random variable $T$ is called a stopping time, and it is decidable whether or not $T = n$ with a knowledge only of the past and present, $X_0, X_1, \ldots, X_n$, and with no further information about the future. We have that

$$P(X_{T+m} = j | X_k = x_k \text{ for } 0 \leq k < T, X_T = i) = P(X_{T+m} = j | X_T = i)$$

(1.6.3)

for $m \geq 0, i, j \in S$ and all sequence $(x_k)$ of states.

2. Classification of States

2.1 Classification of States

Let $f_{i,j}(n) = P\{X_n = j, X_k \neq j \text{ for } 0 < k < n | X_0 = i\}$ and let $f_{i,j}$ be the probability that given $X_0 = i, X_n = j$ for some $n > 0$ and $j \neq i$. That is,

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}(n)$$

(2.1.1)

Assuming $|s| < 1$, we define the generating functions

$$P_{i,j}(s) = \sum_{n=0}^{\infty} s^n p_{i,j}(n)$$

and

$$F_{i,j}(s) = \sum_{n=0}^{\infty} s^n f_{i,j}(n)$$

(2.1.2)

where $p_{i,j}(0) = 1$ if $i = j$ and $p_{i,j}(0) = 0$ otherwise and $p_{i,j}(0) = 0$.

Theorem 2.1. 1. $P_{i,i}(s) = 1 + F_{i,i}(s)P_{i,i}(s)$

2. $P_{i,j}(s) = F_{i,j}(s)P_{j,j}(s)$ if $i \neq j$

Proof. Let $A_n$ be the event that $X_n = i$ and $B_m$ be the event that $X_m = i, X_k \neq i, 0 < k < m$. It is evident that

$$P(A_n) = \sum_{m=0}^{n} P(A_n \mid B_m)P(B_m)$$

(2.1.3)

Due to the strong Markov property,

$$P(A_n) = \sum_{m=0}^{n} P(A_{n-m})P(B_m)$$

(2.1.4)

The two equations in the theorem can then be readily seen by convolution (Theorem 1.4).

Definition 2.2. State $i$ is said to be recurrent if $f_{i,i} = 1$; on the other hand, we say that state $i$ is transient if $f_{i,i} < 1$.

Proposition 2.3. State $i$ is recurrent if $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$ and is transient if $\sum_{n=1}^{\infty} p_{ii}(n) < \infty$.

Proof. To prove the first part of the proposition, we get from Theorem 2.1 that

$$\sum_{n=1}^{\infty} p_{ii}^n = \lim_{s \uparrow 1} P_{i,i}(s) = \lim_{s \uparrow 1} \frac{1}{1 - F_{i,i}(s)}$$

(2.1.5)

which implies

$$1 = \lim_{s \uparrow 1} F_{i,i}(s) = f_{i,i}$$

(2.1.6)
Following this proposition, we can reach the following corollary immediately.

**Corollary 2.4.** If $j$ is transient, then $\sum_{n=1}^{\infty} p_{i,j}(n) < \infty$ for all $i$ and $p_{i,j}(n) \to 0$ as $n \to \infty$ for all $i$.

Let $N_j$ be the number of visits to $j$. That is

$$N_j = \sum_{n=1}^{\infty} I_{\{X_n=j\}} \quad (2.1.7)$$

**Lemma 2.5.**

$$\mathbb{E}[N_j|X_0=i] = \frac{f_{i,j}}{1 - f_{j,j}} \quad (2.1.8)$$

**Proof.** Using tail probability we have

$$\mathbb{E}[N_j|X_0=i] = \sum_{n=1}^{\infty} p_{i,j}(n) = \sum_{n=1}^{\infty} \sum_{m=1}^{n} f_{i,j}(m)p_{j,j}(n-m) \quad (2.1.9)$$

By convolution (Theorem 1.4),

$$\mathbb{E}[N_j|X_0=i] = \lim_{s \uparrow 1} P_{j,j}(s)F_{i,j}(s) = \lim_{s \uparrow 1} \frac{F_{i,j}(s)}{1 - F_{j,j}(s)} = \frac{f_{i,j}}{1 - f_{j,j}} \quad (2.1.10)$$

Using the preceding lemma, we can also obtain the following theorem using the argument similar to Proposition 2.3.

**Theorem 2.6.** State $j$ is recurrent if and only if $\mathbb{E}[N_j|X_0=j] = \infty$.

Let $T_j := \min\{n \geq 1 : X_n = j\}$, i.e., the time passed until the first return to state $j$. It is also easy to see that if $\mathbb{P}(T_j < \infty|X_0=j) < 1$, then state $j$ is transient, while if $\mathbb{P}(T_j < \infty|X_0=j) = 1$, then state $j$ is recurrent. There are a few results that we can readily obtain from this point of view of the classification of states.

**Theorem 2.7.** If $\mathbb{P}(T_j < \infty|X_0=i) > 0$, but $\mathbb{P}(T_i < \infty|X_0=j) < 1$, then state $i$ is transient.

**Proof.** $\mathbb{P}(T_j < \infty|X_0=i) > 0$ implies that $f_{i,j} > 0$. Hence, we let $n := \min\{f_{i,j}(n) > 0\}$. Then,

$$\mathbb{P}(T_j < \infty|X_0=i) \geq [1 - \mathbb{P}(T_i < \infty|X_0=j)]f_{i,j}(n) > 0 \quad (2.1.11)$$

implies $\mathbb{P}(T_j < \infty|X_0=i) < 1$. \hfill \Box

**Corollary 2.8.** If state $i$ is recurrent and $\mathbb{P}(T_j < \infty|X_0=i) > 0$, then $\mathbb{P}(T_i < \infty|X_0=j) = 1$.

**Proof.** If we suppose on the contrary that $\mathbb{P}(T_i < \infty|X_0=j) < 1$, we can reach a contradiction using Theorem 2.7. \hfill \Box

**Definition 2.9.** The **mean recurrence time** $\mu_i$ is defined as, starting from state $i$, the expected value of $T_i$. Moreover,

$$\mu_i = \mathbb{E}(T_i|X_0=j) = \begin{cases} \sum_{n} nf_{i,i}(n) & \text{if state } i \text{ is recurrent} \\ \infty & \text{if state } i \text{ is transient} \end{cases} \quad (2.1.12)$$

**Definition 2.10.** A recurrent state $i$ is called **null recurrent** if $\mu_i = \infty$. A recurrent state $i$ is called **positive recurrent** if $\mu_i < \infty$.

**Theorem 2.11.** A recurrent state $i$ is null recurrent if and only if $p_{i,i}(n) \to 0$ as $n \to \infty$. Moreover, if state $i$ is null recurrent, then $p_{i,i}(n) \to 0$ as $n \to \infty$ for all $j$.

**Definition 2.12.** The **period** of a state $i$ is the largest common divisor of the set $\{n : p_{i,i}(n) > 0, n \geq 1\}$. We write $d(i) = \gcd\{n : p_{i,i}(n) > 0, n \geq 1\}$. We call state $i$ **periodic** if $d(i) > 1$ and **aperiodic** if $d(i) = 1$. 

--

5
Corollary 2.13. If \( p_{i,i}(1) > 0 \), then \( d(i) = 1 \).

Definition 2.14. A state is said to be ergodic if it is positive recurrent and aperiodic.

Example 2.15 (Random Walk). Consider a Markov chain with state space \( \mathbb{Z} \) and transition probability

\[ p_{i,i+1} = 1 - p_{i,i-1} \quad \text{for all } i \in \mathbb{Z} \]  

(2.1.13)

1. We can view this model as a random walk such that on a straight line, we take a step to the right with probability \( p = p_{i,i+1} \) and to the left with probability \( q = 1 - p = p_{i,i-1} \). Suppose we start at state with coordinate 0 and are interested in whether state 0 is recurrent. We do this by using Proposition 2.3. It is easy to see that \( p_{0,0}(m) > 0 \) if and only if \( m \) is even and the total number of steps to took to the left equals the number of steps to the right. That is,

\[
\sum_{m=1}^{\infty} p_{0,0}(m) = \sum_{n=1}^{\infty} p_{0,0}(2n) = \sum_{n=1}^{\infty} \left( \frac{2n}{n} \right) (pq)^n = \sum_{n=1}^{\infty} \frac{2n!}{n!} (pq)^n
\]  

(2.1.14)

Using Sterling’s approximation \( 1.3.2 \) we have that

\[
\sum_{m=1}^{\infty} p_{0,0}(m) \approx \sum_{n=1}^{\infty} \frac{2^{2n}}{\sqrt{n\pi}} (pq)^n = \sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{n\pi}}
\]  

(2.1.15)

Using the fact that if \( a_n \approx b_n \), \( \sum_n a_n < \infty \) if and only if \( \sum_n b_n < \infty \), we obtain that if \( p = q = 1/2 \), then

\[
\sum_{m=1}^{\infty} p_{0,0}(m) \approx \sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}} = \infty
\]  

(2.1.16)

and the state 0 is recurrent. Note also that as \( m \to \infty, p_{0,0}(m) \to 0 \), which implies that state 0 is null recurrent. On the other hand, if \( p \neq 1/2 \), then

\[
\sum_{m=1}^{\infty} p_{0,0}(m) \approx \sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{n\pi}} < \sum_{n=1}^{\infty} (4pq)^n = \lim_{n \to \infty} \frac{1 - (4pq)^n}{1 - (4pq)} < \infty
\]  

(2.1.17)

and the state 0 is transient.

2. Another lively view of this model is usually referred to as the gambler’s ruin, in which we start out with \( i \) units at hand and for each independent gamble, we win a unit with probability \( p = p_{i,i+1} \) and lose one with probability \( q = 1 - p = p_{i,i-1} \). Suppose we are interested in the probability that the gambler reaches at least \( N \) units before going broke (ending up with 0 units). In this case, we can assign two absorbing states \( p_{0,0} = 1, p_{N,N} = 1 \). Letting \( P_i \) denote the probability that starting with \( i \) units, we will eventually reach \( N \) units. We have

\[ P_i = pP_{i+1} + qP_{i-1} \]  

(2.1.18)

That is

\[ pP_i + qP_i = pP_{i+1} + qP_{i-1} \]  

(2.1.19)

then

\[ P_{i+1} - P_i = \frac{q}{p} (P_i - P_{i-1}) \]  

(2.1.20)

Since \( p_{0,0} = 1 \), we have that \( P_0 = 0 \). Thus,

\[ P_2 - P_1 = \frac{q}{p} (P_1 - P_0) = \frac{q}{p} P_1 \]

\[ P_3 - P_2 = \frac{q}{p} (P_2 - P_1) = \left( \frac{q}{p} \right)^2 P_1 \]

\[ \vdots \]

\[ P_{n+1} - P_n = \frac{q}{p} (P_n - P_{n-1}) = \left( \frac{q}{p} \right)^n P_1 \]  

(2.1.21)
which implies if \( p \neq 1/2 \),

\[
P_{n+1} - P_1 = \left[ \frac{q}{p} + \left( \frac{q}{p} \right)^2 + \cdots + \left( \frac{q}{p} \right)^n \right] P_1
\]

\[
P_{n+1} = \left[ 1 + \frac{q}{p} + \left( \frac{q}{p} \right)^2 + \cdots + \left( \frac{q}{p} \right)^n \right] P_1
\]

\[
= \frac{1 - (q/p)^{n+1}}{1 - (q/p)} P_1
\]

Using the fact that \( P_N = 1 \), we have if \( p \neq 1/2 \),

\[
P_1 = \frac{1 - (q/p)}{1 - (q/p)^N}
\]

(2.1.23)

That is, if \( p \neq 1/2 \),

\[
P_i = \frac{1 - (q/p)^{n+1}}{1 - (q/p)^N}
\]

(2.1.24)

On the other hand, if \( p = q = 1/2 \), then \( P_{n+1} = (n + 1)P_1 \). Thus, \( P_1 = 1/N \) and \( P_i = i/N \).

2.2 Classification of Chains

Definition 2.16. We say that state \( j \) is **accessible** from state \( i \) if \( p_{i,j}(n) > 0 \) for some \( n \geq 0 \). We write \( i \rightarrow j \).

Corollary 2.17. If \( x \rightarrow y \) and \( y \rightarrow z \), then \( x \rightarrow z \).

Definition 2.18. Two states \( i \) and \( j \) that are accessible from each other said to **communicate**; we write \( i \leftrightarrow j \).

Lemma 2.19. If \( i \) is recurrent and \( i \leftrightarrow j \), then \( j \) is recurrent.

Proof. Since \( i \leftrightarrow j \), we have there exists \( n = \min \{ n : p_{i,j}(n) > 0 \} \) and \( m = \min \{ m : p_{j,i}(m) > 0 \} \). Therefore,

\[
\sum_{k=1}^{\infty} p_{j,j}(k) \geq p_{i,j}(n)p_{j,i}(m) \sum_{l=1}^{\infty} p_{j,j}(l) = \infty
\]

(2.2.1)

Theorem 2.20. If \( i \leftrightarrow j \), then

1. state \( i \) and state \( j \) have the same period;
2. state \( i \) is transient if and only if state \( j \) is transient;
3. state \( i \) is null recurrent if and only if state \( j \) is null recurrent.

Proof. Since \( i \leftrightarrow j \), then there exists \( m, n, r \geq 1 \) such that \( p_{i,j}(n) > 0, p_{j,i}(m) > 0, p_{i,i}(r) > 0 \).

1. Because \( p_{j,i}(m)p_{i,i}(n) > 0 \) and the steps we take from state \( j \) to itself can be arbitrary, we much have that \( x = y \).
2. Note that \( p_{j,j}(n + r + m) \geq p_{j,i}(m)p_{i,i}(r)p_{i,j}(n) = \alpha p_{i,i}(r) \), then if \( \sum_s p_{j,j}(s) < \infty \), then \( \sum_r p_{i,i}(r) < \infty \).
3. See Proposition [3.7]

Corollary 2.21. If \( i \) is recurrent and \( i \leftrightarrow j \), then \( f_{i,j} = 1 \).
Definition 2.22. A set $C$ of states is called **closed** if $p_{i,j} = 0$ for all $i \in C, j \not\in C$. A set $C$ of states is called **irreducible** if $i \leftrightarrow j$ for all $i, j \in C$. A closed and irreducible set of recurrent states is called a **class**. Two states that communicate are said to be in the same class.

Example 2.23 (Branching Process).

Theorem 2.24 (Decomposition Theorem). The state space $S$ of a Markov chain can be partitioned into disjoint sets as

$$S = T \cup C_1 \cup C_2 \cup ...$$

where $T$ is the set of all transient states and $C_r, r = 1, 2, ...$ are classes.

**Proof.** It suffice to prove that each set $C_r$ is closed. Suppose, on the contrary, there exist states $i \in C_r, j \not\in C_r, i \rightarrow j, j \not\rightarrow i$. Then for all $n \geq 1$,

$$P(X_n \neq i | X_0 = i) \geq p_{i,j}(1) > 0$$

which contradicts with the fact that state $i$ is recurrent.

Lemma 2.25. If the state space $S$ of a Markov chain is finite, then

1. there exists at least one recurrent state;
2. all recurrent states in $S$ are positive.

**Proof.**

1. Suppose, on the contrary, all states are transient, then

$$1 = \lim_{n \to \infty} \sum_{j \in S} p_{i,j}(n) = 0$$

by contradiction, we conclude that there exists at least one recurrent state.

2. Suppose, on the contrary, there is some null state in class $C \subseteq S$. Using Theorem 2.11 and Theorem 2.20 we obtain the contradiction

$$1 = \lim_{n \to \infty} \sum_{j \in C} p_{i,j}(n) = 0$$

3. Stationary Distributions and Limiting Probabilities

3.1 Stationary Distribution

**Definition 3.1.** The vector $\pi$ is called a **stationary distribution** or **stationary measure** of the chain if $\pi$ has entries $(\pi_j, j \in S)$ such that

1. $\pi_j \geq 0$ for all $j \in S, \sum_{j \in S} \pi_j = 1$
2. $\pi = \pi P, \pi_j = \sum_i \pi_i p_{i,j}$ for all $j$

As we start from some fixed state $k$, let $N_i(k)$ denote the number of visits to $i$ before we return to state $k$. That is,

$$N_i(k) = \sum_{n=1}^{\infty} I\{X_n = i, X_0 = k\} \cap \{T_k \geq n\}$$

Further, we let $\rho_i(k)$ denote the expected value of $N_i(k)$. That is,

$$\rho_i(k) = \mathbb{E}[N_i(k)] = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i, T_k \geq n | X_0 = k)$$
Lemma 3.3. For any state $k$ of an irreducible recurrent chain, the vector $\rho(k)$ satisfies $\rho_i(k) < \infty$ for all $i$.

Proof. Let $l_{k,i}(n) = \mathbb{P}(X_n = i, T_k \geq n|X_0 = k)$ and $m = \min\{m : f_{i,k}(m) > 0\}$. It is easy to see that

$$f_{k,k}(m + n) \geq l_{k,i}(n)f_{i,k}(m) \quad (3.1.3)$$

Hence, for all $i$,

$$\rho_i(k) = \sum_{n=1}^{\infty} l_{k,i} \leq \sum_{n=1}^{\infty} f_{k,k}(m + n) \leq \frac{1}{f_{i,k}(m)} < \infty \quad (3.1.4)$$

Lemma 3.3. For any state $k$ of an irreducible recurrent chain, $\rho(k) = \rho(k)\rho$.

Proof. If $n = 1$, we have $l_{k,i}(1) = p_{k,i}$. If $n \geq 2$, using conditional probability, we have

$$l_{k,i}(n) = \sum_{j \in S, j \neq k} \mathbb{P}(X_n = i|X_{n-1} = j)\mathbb{P}(X_{n-1} = j, T_k \geq n - 1|X_0 = k) = \sum_{j \in S, j \neq k} p_{j,i}l_{k,j}(n-1) \quad (3.1.5)$$

Therefore,

$$\rho_i(k) = \sum_{n=1}^{\infty} l_{k,i}(n) = p_{k,i} + \sum_{n=2}^{\infty} \sum_{j \in S, j \neq k} p_{j,i}l_{k,j}(n-1) = p_{k,i} + \sum_{j \in S, j \neq k} p_{j,i} \sum_{n=1}^{\infty} l_{k,j}(n) = p_{k,i} + \sum_{j \in S, j \neq k} p_{j,i}\rho_j(k) \quad (3.1.6)$$

Since $k$ is recurrent, we have $\rho_k(k) = f_{k,k} = 1$, and conclude $\rho(k) = \rho(k)\rho$.

Theorem 3.4. If the chain is irreducible and recurrent, there exists some solution $\mathbf{x} > 0$ of the equation $\mathbf{x} = \mathbf{x}\rho$, unique up to a multiplicative constant. The chain is positive if $\sum x_i < \infty$ and null if $\sum x_i = \infty$.

The preceding theorem (whose proof is beyond the scope of this note) allows us to investigate the following lemma.

Lemma 3.5. If the chain is irreducible and positive recurrent, then it has a stationary distribution $\pi$ which is unique and $\pi_j = \mu_j^{-1}$ for all $j \in S$.

Proof. Note that $\mu_j = \mathbb{E}(T_j|X_0 = j) = \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n|X_0 = j) = \sum_i \rho_i(j)$. Hence,

$$\mu_j\pi_j = \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n|X_0 = j)\mathbb{P}(X_0 = j) = \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n, X_0 = j) \quad (3.1.7)$$

If $n = 1$, we have $\mathbb{P}(T_j \geq 1, X_0 = j) = \mathbb{P}(X_0 = j)$; if $n \geq 2$, then

$$\mathbb{P}(T_j \geq n, X_0 = j) = \mathbb{P}(X_k \neq j, 1 \leq k \leq n - 1, X_0 = j) = \mathbb{P}(X_k \neq j, 1 \leq k \leq n - 1) - \mathbb{P}(X_k \neq j, 0 \leq k \leq n - 1) \quad (3.1.8)$$

By homogeneity, we have that

$$\mathbb{P}(T_j \geq n, X_0 = j) = \mathbb{P}(X_k \neq j, 0 \leq k \leq n - 2) - \mathbb{P}(X_k \neq j, 0 \leq k \leq n - 1) \quad (3.1.9)$$
Therefore,

\[
\mu_j \pi_j = \mathbb{P}(X_0 = j) + \sum_{n=2}^{\infty} \mathbb{P}(T_j \geq n, X_0 = j)
\]

\[
= \mathbb{P}(X_0 = j) + \sum_{n=2}^{\infty} [\mathbb{P}(X_k \neq j, 0 \leq k \leq n - 2) - \mathbb{P}(X_k \neq j, 0 \leq k \leq n - 1)]
\]

\[
= \mathbb{P}(X_0 = j) + \mathbb{P}(X_0 \neq j) - \lim_{n \to \infty} \mathbb{P}(X_k \neq j, 0 \leq k \leq n - 1)
\]

\[
= 1 - \sum_{j} \mathbb{P}(T_j = \infty|X_0 = j)\mathbb{P}(X_0 = j)
\]

(3.1.10)

By Corollary 2.8 we have \( \mu_j \pi_j = 1 \). Furthermore, by Definition 2.9 and 2.10 we have that \( \pi_j \) is unique and \( \pi_j = \mu_j^{-1} \). \( \square \)

Note that if the chain is not positive recurrent, then for all \( j, \mu_j = \infty \) and \( \sum_j \pi_j = \sum_j \mu_j^{-1} = 0 < 1 \). Using the preceding Lemma, we reach the following theorem.

**Theorem 3.6.** An irreducible chain has a stationary distribution \( \pi \) if and only if all states are positive recurrent. In this case, \( \pi \) is unique and \( \pi_i = \mu_i^{-1} \) for all \( i \in S \).

**Proposition 3.7.** If \( i \) is positive recurrent and \( i \leftrightarrow j \), then \( j \) is positive recurrent.

**Proof.** It is easy to see that

\[
\pi_j \geq \pi_i f_{i,j}(n) > 0
\]

(3.1.11)

where \( n = \min\{n : f_{i,j}(n) > 0\} \). Hence, \( \mu_j = \pi_j^{-1} < \infty \). \( \square \)

**Theorem 3.8.** Let \( s \in S \) be any state of an irreducible chain. The chain is transient if and only if there exists a non-zero solution \( \{y_j : j \neq s\} \), satisfying \( |y_i| \leq 1 \) for all \( j \), to the equations

\[
y_i = \sum_{j \neq s} p_{i,j} y_j, \quad i \neq s
\]

(3.1.12)

**Proof.** Suppose \( s \) is transient. We let

\[
\tau_i(n) = \mathbb{P}(X_k \neq s, 1 \leq k \leq n|X_0 = i)\] and \( \tau_i(n + 1) = \sum_{j \neq s} p_{i,j} \tau_j(n) \)

(3.1.13)

and

\[
\tau_i = \lim_{n \to \infty} \tau_i(n) = 1 - f_{i,s} = \lim_{n \to \infty} \sum_{j \neq s} p_{i,j}(n) + p_{i,s}(n) - \sum_{j \neq s} p_{i,j}(n)f_{j,s}(n)
\]

(3.1.14)

Since by Corollary 2.4 \( p_{i,s}(n) \to 0 \) as \( n \to \infty \), \( \tau_i = \sum_{j \neq s} p_{i,j} \tau_j \). We also have \( \tau_i > 0 \). If we suppose on the contrary that \( \tau_i = 0 \), then \( f_{i,s} = 1 \) and we reach a contradiction that

\[
f_{s,s} = p_{s,s} + \sum_{i \neq s} p_{s,i} f_{i,s} = \sum_{p_{s,i}} = 1
\]

(3.1.15)

Conversely, since we have \( |y_i| < 1 \), then

\[
|y_i| \leq \sum_{j \neq s} p_{i,j}|y_j| \leq \tau_i(1)
\]

(3.1.16)

and so on. \( \square \)
Proposition 3.9. Let \( \{X_n, n \geq 1\} \) be an irreducible Markov chain with stationary probabilities \( \pi_j, j \geq 0 \), and let \( r \) be a bounded function on the state space. Then with probability 1,

\[
\lim_{N \to \infty} \frac{\sum_{n=1}^{N} r(X_n)}{N} = \sum_j r(j)\pi_j
\]  

(3.1.17)

Proof. Let \( N_j \) be the number of visits to \( j \). Hence,

\[
\lim_{N \to \infty} \frac{\sum_{n=1}^{N} r(X_n)}{N} = \lim_{N \to \infty} \frac{\sum_j N_j r(j)}{N} = \sum_j r(j)\pi_j
\]  

(3.1.18)

Example 3.10. Suppose that for each day, assuming independence, the number of families check into a hotel has a poisson distribution with mean \( \lambda \), we write \( Y \sim \text{Poisson}(\lambda) \). For each family staying in the hotel, it is likely to check out independently with probability \( p \).

1. We are interested in the transition probability - the number of families staying in the hotel on the very next day, given that \( i \) families stayed. Since each family is independently likely to check out with probability \( p \), the number of families remaining is thus \( r(i) \sim \text{binomial}(i, 1 - p) \). Therefore,

\[
p_{i,j} = \mathbb{P}(r(i) + Y = j)
\]

\[
= \sum_{k=0}^{\min(i,j)} \mathbb{P}(Y = j - r(i)|r(i) = k) \binom{i}{k} p^{i-k}(1-p)^k
\]

(3.1.19)

2. To obtain the expectation \( \mathbb{E}[X_n|X_0 = i] \), we first consider \( \mathbb{E}[X_n|X_{n-1} = i] \).

\[
\mathbb{E}[X_n|X_{n-1} = i] = \mathbb{E}(r(i) + Y) = \mathbb{E}[r(i)] + \mathbb{E}(Y) = iq + \lambda = X_{n-1}q + \lambda
\]

(3.1.20)

where \( q = 1 - p \). Now, we obtain

\[
\mathbb{E}[X_1|X_0 = i] = iq + \lambda
\]

\[
\mathbb{E}[X_2|X_0 = i] = \mathbb{E}[X_2|X_1 = iq + \lambda] = (iq + \lambda)q + \lambda = iq^2 + \lambda(1+q)
\]

\[
\mathbb{E}[X_3|X_0 = i] = iq^3 + \lambda(1+q+q^2)
\]

\[
\vdots
\]

\[
\mathbb{E}[X_n|X_0 = i] = iq^n + \lambda(1+q+q^2+\cdots+q^n) = iq^n + \lambda \frac{1-q^n}{p}
\]

(3.1.21)

3. To obtain the stationary probabilities of number of families staying in the hotel, we can use the fact that \( \mathbf{\pi} = \mathbf{\pi P} \), which lead us to assume that the initial number of people checking into the hotel is Poisson distributed with mean \( \alpha \). Hence, it suffice to use the fact that \( \mathbb{E}[X_1] = \mathbb{E}[X_0] \), which is

\[
\alpha q + \lambda = \alpha
\]

\[
\alpha = \frac{\lambda}{p}
\]

(3.1.22)

Hence, for \( i \geq 0 \)

\[
\pi_i = e^{-\alpha} \frac{\alpha^i}{i!} = e^{-\lambda/p} \frac{(\lambda/p)^i}{i!}
\]

(3.1.23)

Now we can generalise the situation.
### 3.2 Limiting Probabilities

#### Theorem 3.11 (Convergence Theorem)
For an irreducible aperiodic Markov chain, we have that \( p_{i,j}(n) \to 1/\mu_i \) as \( n \to \infty \) for all \( i \) and \( j \).

**Proof.**
1. Suppose the chain is transient, it is evident that by Corollary 2.4, \( p_{i,j}(n) \to 0 \) and \( \mu_j \to \infty \) as \( n \to \infty \).

2. Now consider the case that the chain is recurrent. Let \( X, Y \) be independent, irreducible and aperiodic chains with the same state space \( S \) and transition probability matrix \( P \). We use a "coupled chain" \( Z = (X, Y) \) with transition probabilities

\[
p_{i,j,k,l}(n) = \mathbb{P}(X_n = k, Y_n = l | X_0 = i, Y_0 = j) = \mathbb{P}(X_n = k | X_0 = i) \mathbb{P}(Y_n = l | Y_0 = j) = p_{i,k}(n)p_{j,l}(n)
\]

where \( i, j, k, l \in S \). Since \( X, Y \) are both irreducible, we have that there exists some \( n > 0 \) such that for all \( i, j, k, l \), \( p_{i,k}(n)p_{j,l}(n) > 0 \). Hence, the coupled chain \( Z \) is also irreducible.

(a) Suppose \( X \) and \( Y \) are positive recurrent. Then each state \((i, j)\) is thus positive recurrent. Hence, \( Z \) is also positive recurrent. Since \( Z_0 = (i, j), i, j \in S \), we let \( T = \min\{n \geq 1 : Z_n = (s, s)\} \). If \( T < \infty \), using strong Markov Property, we have that \( X_n \) and \( Y_n \) only depend on \( X_T = Y_T = s \) and the shared probability matrix \( P \); thus \( X_n \) and \( Y_n \) are independent when \( T < n \). Therefore,

\[
p_{i,k}(n) = \mathbb{P}(X_n = k) = \mathbb{P}(X_n = k, T < n) + \mathbb{P}(X_n = k, T \geq n) \leq \mathbb{P}(Y_n = k) + \mathbb{P}(T \geq n) = p_{j,k}(n) + \mathbb{P}(T \geq n)
\]

Similarly, we have \( p_{j,k}(n) \leq p_{i,k}(n) + \mathbb{P}(T \geq n) \). Hence, \( \lim_{n \to \infty} |p_{j,k}(n) - p_{i,k}(n)| \leq \lim_{n \to \infty} \mathbb{P}(T \geq n) = 1 - \mathbb{P}(T < \infty) = 0 \) by Corollary 2.8. Since \( \sum_i \pi_i = 1 \), we then have

\[
\lim_{n \to \infty} \pi_k - p_{j,k}(n) = \lim_{n \to \infty} \sum_i \pi_i(p_{i,k}(n) - p_{j,k}(n)) = 0
\]

(b) Suppose \( X \) and \( Y \) are null recurrent. We need to verify that \( p_{i,j}(n) \to 0 \) as \( n \to \infty \).

i. Suppose \( Z \) is transient. Then we have

\[
0 = \lim_{n \to \infty} \mathbb{P}(X_n = j, Y_n = j | X_0 = i, Y_0 = i) = \lim_{n \to \infty} p_{i,j}(n)p_{i,j}(n)
\]

ii. Suppose \( Z \) is positive recurrent. Let \( \mu_{ii} \) be the expected time elapsed between two consecutive visits to state \((i, i)\). Then we reach the contradiction that

\[
\infty \geq \mu_i \leq \mu_{ii} < \infty
\]

Therefore, \( Z \) cannot be positive recurrent.

iii. Suppose \( Z \) is null recurrent and there exists an \( \alpha \) independent of state \( i \), \( \alpha_j > 0 \) for some \( j \) and a series of steps \( \{n_r, r = 0, 1, 2, \ldots\} \) such that \( p_{i,j}(n_r) \to \alpha_j \) as \( r \to \infty \). For a subset \( F \subseteq S \), we have that

\[
\sum_{j \in F} \alpha_j = \lim_{r \to \infty} \sum_{j \in F} p_{i,j}(n_r) \leq 1
\]

\[
\sum_{k \in F} \alpha_k p_{k,j} = \lim_{r \to \infty} \sum_{k \in F} p_{i,k}(n_r)p_{k,j} \leq p_{i,j}(n_r + 1) = \lim_{r \to \infty} \sum_{k \in S} p_{i,k}(n_r) = \sum_{k \in S} p_{i,k} \alpha_j = \alpha_j
\]
and hence $\sum_k \alpha_k p_{k,j} = \alpha_j$, for if $\sum_k \alpha_k p_{k,j} < \alpha_j$, then as $F \uparrow S$, we have

$$\sum_k \alpha_k = \sum_k \sum_j \alpha_k p_{k,j} < \sum_j \alpha_j \quad (3.2.8)$$

Let $\alpha = \sum_j \alpha_j$, then we have

$$\sum_j \alpha_j / \alpha = 1 \quad \text{and} \quad 1 / \alpha \mathbf{P} = \mathbf{P} \quad (3.2.9)$$

implying that $X$ has a stationary distribution, which contradicts the assumption that $X$ is null recurrent. Therefore, $\alpha = 0$.

The following is a generalised version of the preceding theorem, whose proof is not presented in this note.

**Theorem 3.12 (Ergodic Theorem).** For any aperiodic state $j$ of a Markov chain, $p_{j,j}(n) \to \mu_j^{-1}$ as $n \to \infty$. Furthermore, if $i$ is any other state then $p_{i,j}(n) \to f_{i,j}/\mu_j$ as $n \to \infty$.

**Corollary 3.13.** Let

$$\tau_{i,j}(n) = \frac{1}{n} \sum_{m=1}^n p_{i,j}(m) \quad (3.2.10)$$

be the mean proportion of elapsed time up to the $n$th step during which the chain was in state $j$, starting from $i$. If $j$ is aperiodic, $\tau_{i,j}(n) \to f_{i,j}/\mu_j$ as $n \to \infty$.

## 4. Time Reversible Markov Chains

**Definition 4.1.** Suppose we have an irreducible, positive recurrent Markov chain $\{X_n : 0 \leq n \leq N\}$ with transition probability matrix $\mathbf{P}$ and stationary distribution $\pi$. We further suppose that we run this chain for a long time, that is, $X_n$ has distribution $\pi$ for every $n$. We call a chain $Y$ the **time reversal** of chain $X$ if $Y_n = X_{N-n}$ for $0 \leq n \leq N$.

**Theorem 4.2.** The sequence $Y$ is a Markov chain with

$$\mathbb{P}(Y_{n+1} = j | Y_n = i) = \frac{\pi_j}{\pi_i} p_{j,i} \quad (4.0.1)$$

$$\mathbb{P}(Y_{n+1} = j | Y_n = i) = \mathbb{P}(Y_{n+1} = j, Y_n = i) \frac{\mathbb{P}(Y_n = i)}{\mathbb{P}(X_{N-n} = i)}$$

$$= \mathbb{P}(X_{N-n-1} = j, X_{N-n} = i) \frac{\mathbb{P}(X_{N-n} = i)}{\mathbb{P}(X_{N-n} = i)}$$

$$= \mathbb{P}(X_{N-n-1} = j) \frac{\mathbb{P}(X_{N-n-1} = j)}{\mathbb{P}(X_{N-n} = i)}$$

$$= \frac{p_{j,i} \pi_j}{\pi_i} \quad (4.0.2)$$

**Definition 4.3.** Let $X = \{X_n : 0 \leq n \leq N\}$ be an irreducible Markov chain such that $X_n$ has the stationary distribution $\pi$ for all $n$. The chain is called **reversible** if the transition matrix of $X$ and its time reversal $Y$ are the same. That is,

$$\pi_i p_{i,j} = \pi_j p_{j,i} \text{ for all } i, j \quad (4.0.3)$$
More generally, we say that the transition matrix $P$ and a distribution $\lambda$ are in detailed balance if $\lambda_ip_{i,j} = \lambda_jp_{j,i}$ for all $i, j \in S$. An irreducible chain $X$ with a stationary distribution $\pi$ is called reversible in equilibrium if the transition matrix $P$ is in detailed balance with $\pi$.

**Theorem 4.4.** Let $P$ be the transition probability matrix of an irreducible Markov chain $X$, and suppose that there exists a distribution $\pi$ such that $\pi_ip_{i,j} = \pi_jp_{j,i}$ for all $i, j \in S$. Then $\pi$ is the stationary distribution of the chain. Furthermore, $X$ is reversible in equilibrium.

**Proof.** It is easy to verify that

$$\sum_i \pi_ip_{i,j} = \sum_i \pi_jp_{j,i} = \pi_j \sum_i p_{j,i} = \pi_j$$

(4.0.4)
References

