# PStat 213A: Probability Theory and Stochastic Processes

Simon Rubinstein–Salzedo

Fall 2005

## 0.1 Introduction

These notes are based on a graduate course on probability theory and stochastic processes I took from Professor Raya Feldman in the Fall of 2005. The primary textbook was *Adventures in Stochastic Processes* by Sidney Resnick. Other recommended books were *Probability and Random Processes* by Geoffrey Grimmett and David Stirzaker, *Probability: Theory and Examples* by Richard Durrett, *A First Course in Stochastic Processes* by Samuel Karlin and Howard Taylor, and *Probability* by Leo Breiman.

# Chapter 1

# Generating Functions and Branching Processes

## 1.1 Generating Functions

Generating functions are the main tool of dealing with sums of discrete nonnegative integer valued random variables. Let $X$ be a nonnegative integer valued random variable $\{0, 1, \ldots\}$ with probabilities $p_k = \Pr\{X = k\}$, $k = 0, 1, 2, \ldots$ The generating function of $X$ is

$$P_X(s) \triangleq E(s^X) = \sum_{k=0}^{\infty} s^k p_k, \qquad s \geq 0.$$

We have the following properties:

1. $P_X(s)$ is an increasing function of $s$ on $s \geq 0$, $P_X(0) = p_0 = \Pr(X = 0)$, and $P_X(1) = 1$.

2. $P_X(s)$ is convex on $s \geq 0$ assuming $P(X \geq 2) > 0$. We have $P'_X(s) = \sum_{k=1}^{\infty} k s^{k-1} p_k \geq 0$ and $P''_X(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} p_k \geq 0$. The last inequality is strict if $\Pr(X \geq 2) > 0$.

3. $p_k = \Pr(X = k) = \frac{P_X^{(k)}(0)}{k!}$.

4. $E[X(X-1)\cdots(X-n+1)] = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1)p_k = p_X^{(n)}(1)$.

5. If $X_1, \ldots, X_n$ are independent random variables with values in $\{0, 1, 2 \ldots\}$ and $S_n = X_1 + \cdots + X_n$, then

$$\begin{aligned} P_{S_n}(s) &= E(s^{X_1+\cdots+X_n}) \\ &= E(s^{X_1} s^{X_2} \cdots s^{X_n}) \\ &= Es^{X_1} Es^{X_2} \cdots Es^{X_n} \\ &= P_{X_1}(s) \cdots P_{X_n}(s). \end{aligned}$$

6. Let $\{X_n\}$ be independent identically distributed random variables with values in $\{0, 1, 2, \ldots\}$, each with generating function $P_X(s)$. Let $N \geq 0$ be integer-valued and independent of the $\{X_n\}$. Let $\Pr(N = j) = \alpha_j$ for $j = 0, 1, 2 \ldots$. Then $P_N(s) = \sum_{j=0}^{\infty} s^j \alpha_j$. Let $S_N = X_1 + \cdots + X_N$. Then $P_{S_N}(t) = P_N(P_X(t))$.

   **Proof.** We have $P_{S_N}(t) = E(t^{S_N})$. By the tower property, we have $E(X) = E(E(X \mid Y))$. Hence

$$\begin{aligned} E(t^{S_N}) &= E(E(t^{S_N} \mid N)) \\ &= \sum_{j=0}^{\infty} E(t^{S_j} \mid N = j)\alpha_j \\ &= \sum_{j=0}^{\infty} \alpha_j E(t^{X_1+\cdots+X_j} \mid N = j) \\ &= \sum_{j=0}^{\infty} \alpha_j E(t^{X_1+\cdots+X_j}) \\ &= \sum_{j=0}^{\infty} \alpha_j (E(t^{X_1}))^j \\ &= \sum_{j=0}^{\infty} \alpha_j (P_X(t))^j \\ &= P_N(P_X(t)), \end{aligned}$$

   as desired. ■

## 1.2 Branching Processes (Galton-Watson-Bienaymé)

The Galton-Watson-Bienaymé branching process is the simplest stochastic model for population growth. A family starts with one progenitor (one founding member) who forms generation 0.

1. $Z_0 = 1$.
2. The family sizes of the individuals $\{Z_{n,j}\}$ are independent and identically distributed. $Z_{n,j}$ is the size of the $j^{\text{th}}$ family in the $n^{\text{th}}$ generation.

$Z_n$ is the size of the $n^{\text{th}}$ generation. We therefore have $Z_n = Z_{n,1}+Z_{n,2}+\cdots+Z_{n,Z_{n-1}}$. Of course, $Z_1 \triangleq Z_{n,j}$. We set $\Pr\{Z_{n,j} = k\} = p_k$ for $k = 0, 1, 2, \ldots$ for all $n$ and $j$. We let $P(s) = \sum_{k=0}^{\infty} s^k p_k$ be the generating function for each $Z_{n,j}$. The mean is $m = EZ_{n,j}$, and $\sigma^2 = \text{Var}(Z_{n,j})$.

**Theorem 1.1** Let $P_n(s) = E(s^{Z_n})$ be the generating function on $Z_n$. Then $P_{n+m}(s) = P_m(P_n(s)) = P(P(P(\cdots(P(s))\cdots)))$.

**Proof.** Let $S_N = X_1 + \cdots + X_N$ so that $P_{S_N}(s) = P_N(P_X(s))$. We have $Z_n = Z_{n,1} + \cdots + Z_{n,Z_{n-1}}$. Then $\{Z_{n,j}\}$ are independent identically distributed random variables with generating function $P$, and $Z_{n-1}$ is independent of $\{Z_{n,j}\}$ and has generating function $P_{n-1}$. Thus $P_n(s) = P_{n-1}(P(s)) = P_{n-2}(P(P(s))) = \cdots$. ■

## 1.3 Moments of $Z_n$

Let $m_n := E(Z_n)$. Thus $m_1 = m = EZ_{n,j}$. Then $m_n = m^n$ and

$$\text{Var}(Z_n) = \begin{cases} n\sigma^2 & \text{if } m = 1, \\ \sigma^2(m^n - 1)m^{n-1}(m-1)^{-1} & \text{if } m \neq 1. \end{cases}$$

**Proof.** We have

$$\begin{aligned} E(Z_n) &= P_n'(1) \\ &= [P_{n-1}(P(s))]'_{s=1} \\ &= P_{n-1}'(P(1))P'(1) \\ &= P_{n-1}'(1)m \\ &= m_{n-1}m. \end{aligned}$$

We also have

$$\begin{aligned} \mathrm{Var}(Z_n) &= EZ_n^2 - (EZ_n)^2 \\ &= E[Z_n(Z_n-1)] + EZ_n - (EZ_n)^2 \\ &= P_n''(1) + m^n - m^{2n}, \end{aligned}$$

and the result follows. ■

## 1.4 Probability of Extinction

**Definition.** By **extinction**, we mean an event that the random sequence $\{Z_n\}$ consists of zeros for all but finitely many values of $n$.

Mathematically, $\{\text{extinction}\} = \{Z_n = 0 \text{ for some } n\} = \{Z_1 = 0 \text{ or } Z_2 = 0 \text{ or } \cdots\} = \bigcup_{n=1}^{\infty}\{Z_n = 0\} = \bigcup_{n=1}^{\infty} A_n$, where $A_n = \{Z_n = 0\}$. The sequence $\{A_n\}$ is monotonic:

$$\underbrace{\{Z_n = 0\}}_{A_n} \subseteq \underbrace{\{Z_{n+1} = 0\}}_{A_{n+1}} \subseteq \underbrace{\{Z_{n+2} = 0\}}_{A_{n+2}} \subseteq \cdots .$$

Hence $\lim_{n\to\infty} \Pr(A_n) = \Pr\left(\bigcup_{n=1}^{\infty} A_n\right)$ exists. Let

$$\pi = \Pr(\{\text{extinction}\}) = \Pr\left(\bigcup_{n=1}^{\infty}\{Z_n = 0\}\right) = \lim_{n\to\infty} \Pr(Z_n = 0) = \lim_{n\to\infty} P_n(0)$$

and $\pi_n = \Pr(Z_n = 0) = \Pr(A_n)$. Clearly $\pi = \lim_{n\to\infty} \pi_n$.

**Theorem 1.2** $\pi = \Pr(\text{extinction}) = \lim_{n\to\infty} \Pr(Z_n = 0)$ is the smallest nonnegative root of the equation $s = P(s)$. If $m = E(Z_1) < 1$, then $\pi = 1$ (this is the subcritical case). If $m = 1$ and $\sigma^2 > 0$, then $\pi = 1$ (this is the critical case). If $m > 1$, then $\pi < 1$ (this is the supercritical case). (We assume that $0 < p_0 = \Pr(Z_1 = 0) < 1$. If $p_0 = 1$, then all die, and $\pi = 1$. If $p_0 = 0$, then there are no deaths, so $\pi = 0$.)

If $m < 1$, then $E(Z_n) = m^n \downarrow 0$. If $m > 1$, then $E(Z_n) = m^n \nearrow \infty$. Then $Z_0 \to 0$ with probability $\pi$, and $Z_n \to \infty$ with probability $1 - \pi$.

**Proof.** We show that $\pi = P(\pi)$. We have $P_n(s) = P(P_{n-1}(s))$. If $s = 0$, then $\pi_n = P_n(0) = P(P_{n-1}(0)) = P(\pi_{n-1})$. We have

$$\pi_n = \Pr(Z_n = 0) \leq \Pr(Z_{n+1} = 0) = \pi_{n+1}. \qquad (*)$$

So $\{\pi_n\}$ is increasing: $0 \leq \pi_n \leq 1$, $\pi_n \uparrow \pi$. By $(*)$, $P(s) \overset{s \to s_0}{\longrightarrow} P(s_0)$.

We now show that $\pi$ is the smallest nonnegative root of $s = P(s)$. Let $q \geq 0$ be another root, so $q = P(q)$. We show that $\pi \leq q$. On $s \geq 0$, $P(s)$ is nondecreasing: If $q \geq 0$, then $q = P(q) \geq P(0) = \pi_!$. Since $q \geq \pi_1$, we have $q = P(q) \geq P(\pi_1) = \pi_2$, so $q \geq \pi_2$, and so forth. Hence $q \geq \pi_n$ for each $n$. Therefore $q \geq \pi$.

Now let $y(s) = P(s) - s$. We have $y(1) = 1 - 1 = 0$, and $y'(1) = P'(1) - 1 = m - 1$. If $m > 1$, then $y'(1) > 0$, so $y$ is increasing, so $y(s) < 0$ for $s < 1$. Therefore $P(s) < s$ for $s < 1$ near $s = 1$. In this case, $\pi < 1$. If $m < 1$, then $y'(1) < 0$, and $y$ is decreasing. If $s < 1$, then $y(s) > 0$, so $P(s) > s$, and $\pi = 1$. If $m = 1$, then $y'(1) = 0$. We have

$$P(s) = P(1) + (s-1)P'(1) + \frac{(s-1)^2}{2!} P''(1 + \theta(s-1)),$$

the second-order Taylor expansion in a neighborhood of 1. Thus $P(s) = 1 + (s-1) + \frac{(s-1)^2}{2!} P''(1 + \theta(s-1)) = 1 + (s-1) +$(something nonnegative)$\geq s$, so $\pi = 1$. ■

## 1.5 Instability of $Z_n$

The result of Theorem 1.2 is that $\Pr(\text{extinction}) = \pi < 1$ when $m > 1$.

**Question.** What happens when there is no extinction?

**Theorem 1.3** No matter what the finite value of $m = EZ_{n,j}$ is, we have $\Pr(Z_n = k) \overset{n\to\infty}{\to} 0$ for all $k = 1, 2, \ldots$ Moreover, $Z_n \to \infty$ with probability $1 - \pi$, and $Z_n \to 0$ with probability $\pi$.

**Question.** What is the rate of growth when $Z_n \to \infty$?

**Theorem 1.4** If $m > 1$ and $EZ_{n,j}^2 < \infty$, then the random variables $W_n := \frac{Z_n}{m^n} \to W$ almost surely, where $W$ has finite $EW = 1$, $\mathrm{Var}(W) > 0$.

If $n$ is large, then $Z_n \sim m^n W$; thus $\log Z_n \sim n \log m + \log W$ (which is linear in $n$).

# Chapter 2

# Markov Chains: Definitions and Examples

## 2.1 Definitions

A **stochastic process** is a random process evolving in time. Mathematically, it is a collection of random variables indexed by time $X = \{X_t, t \in T\}$. Here all $X_t$'s are random variables on the same probability space $(\Omega, \mathcal{F}, P)$, and they all take values in the same state space $S$. In this section, $S$ is a finite or countable set, so the $X_i$'s are discrete random variables. We also assume that $T = \{0, 1, 2, \ldots\}$ is discrete.

To describe the probabilities of such a process, we can consider $\Pr(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)$ for all $n$, $i_0, \ldots, i_n \in S$. Equivalently, we have

- an initial distribution $\{a_k\} = \{\Pr(X_0 = i_k)\}$, $i_k \in S$.
- transition probabilities $q_n(i_n \mid i_0, \ldots, i_{n-1}) = \Pr(X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1})$.

So then $\Pr(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = a_{i_0} q_1(i_1 \mid i_0) q_2(i_2 \mid i_0, i_1) \cdots q_n(i_n \mid i_0, \ldots, i_{n-1})$.

A **Markov chain** is a stochastic process that satisfies the **Markov property**:

$$\Pr(X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \Pr(X_n = i_n \mid X_{n-1} = i_{n-1}).$$

Here $X_{n-1}$ is the last observed value, or the present case, $X_0, \ldots, X_{n-2}$ are the past, and $X_n$ is the value of tomorrow.

We generally make the additional assumption of time-homogeneity. A Markov chain is **time-homogeneous** (or has stationary transition probabilities) if

$$\begin{aligned}\Pr(X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}) &= \Pr(X_n = i_n \mid X_{n-1} = i_{n-1}) \\ &= \Pr(X_1 = i_n \mid X_0 = i_{n-1}) \\ &= p_{i_{n-1}, i_n}.\end{aligned}$$

Then

$$\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = a_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}.$$

The $p_{ij}$'s are the one-step transition probabilities.

We define the **transition matrix** to be $P = (p_{ij})_{i,j \in S}$. $P$ and the initial distribution $a = (a_k)_{k \in S}$ determine the distribution of the chain.

**Note.** The matrix $P$ is **stochastic**:

1. $0 \leq p_{ij} \leq 1$ for all $i, j \in S$.
2. $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$. (The rows sum to 1.)

## 2.2 Examples.

1. Independent trials. Let $X_n$ be independent identically distributed random variables with probability mass function $\{a_k\}$. Then

   $$\Pr(X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}) = \Pr(X_n = i_n) = a_n.$$

   We have $p_{i_{n-1},i_n} = a_n$. The transition matrix is

   $$P = \begin{matrix} a_0 & a_1 & a_2 & \cdots & a_k & \cdots \\ a_0 & a_1 & a_2 & \cdots & a_k & \cdots. \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{matrix}$$

2. Galton-Watson Branching Process. Let $Z_n = Z_{n,1} + Z_{n,2} + \cdots + Z_{n,Z_{n-1}}$, where $Z_{n,j}$ are independent identically distributed with probability mass function $p_k$, $Z_0 = 1$, generating function $P(s)$, and state space $S = \{0, 1, 2, \ldots\}$. Then

$$\begin{aligned}
&\Pr(Z_n = i_n \mid Z_0 = 1, Z_i = i_1, \ldots, Z_{n-1} = i_{n-1}) \\
&= \Pr(Z_{n,1} + \cdots + Z_{n,i_{n-1}} = i_n \mid Z_0 = 1, \ldots, Z_{n-1} = i_{n-1}) \\
&= \Pr(Z_{n,1} + Z_{n,2} + \cdots + Z_{n,i_{n-1}} = i_n) \\
&= p_{i_{n-1},i_n}.
\end{aligned}$$

If we let $Z_{n,1} = Y_1, \ldots, Z_{n,i_{n-1}} = Y_{i_{n-1}}$, then $Y_1, \ldots, Y_{i_{n-1}}$ are independent identically distributed random variables. Then $P_{\sum Z_{n,j}}(s) = (P(s))^{i_{n-1}}$.

3. Random walk. Let $\{\xi_i\}$ be independent identically distributed random variables, and let $\Pr(\xi_i = k) = a_k$ for $k \in \mathbb{Z}$. We let our state space be $S = \mathbb{Z}$. Start with $X_0 = 0$ and let $X_n = X_{n-1} + \xi_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Then

$$\begin{aligned}
&Pr(X_n \equiv X_{n-1} + \xi_n = j \mid X_0 = 0, \ldots, X_{n-1} = i_{n-1}) \\
&= \Pr(X_n = i_{n-1} + \xi_n = j \mid X_0 = 0, \ldots, X_{n-1} = i_{n-1}) \\
&= \Pr(\xi_n = j - i_{n-1}) \\
&= a_{j-i_{n-1}}.
\end{aligned}$$

We call a random walk simple if $\xi_i = +1$ with probability $p$ and $\xi_i = -1$ with probability $q = 1 - p$. We say a simple random walk is symmetric if $p = q = \frac{1}{2}$.

4. Random walk with reflecting boundary. Consider a "random walker" moving along $\{0, \ldots, N\}$. If $0 < i < N$, let $p_{i,i+1} = p$ and $p_{i,i-1} = q = 1 - p$. Let $p_{0,1} = p_{N,N-1} = 1$.

5. Random walk with absorbing boundary (gambler's ruin chain). Let $S = \{0, \ldots, N\}$. We have the same probabilities as above except that $p_{0,0} = p_{N,N} = 1$.

6. Random walk on a graph. Let $G$ be a finite unoriented simple graph with vertex set $S$. We let $S$ be the state space, and we define transition probabilities by

$$p_{v_i,v_j} = \begin{cases} 0 & v_i \text{ is not adjacent to } v_j, \\ \frac{1}{d(v_i)} & v_i \text{ is adjacent to } v_j. \end{cases}$$

# Chapter 3

# Long-Term Evolution of Markov Chains

**Definition.** Define $n$-step transition probabilities by

$$p_{ij}^{(n)} = \Pr(X_{m+n} = j \mid X_m = i).$$

Define the $n$-step transition matrix by $P^{(n)} = (p_{i,j}^{(n)})$. We have $P^{(1)} = P$.

We have

$$\begin{aligned} a_j^{(n)} &:= \Pr(X_n = j) \\ &= \sum_{i \in S} P_i(X_0 = i) \Pr(X_n = j \mid X_0 = i) \\ &= \sum_{i \in S} a_i p_{ij}^{(n)}, \end{aligned}$$

or $a^n = ap^{(n)}$.

**Chapman-Kolmogorov Equation.** We have

$$\begin{aligned} p_{ij}^{(n+1)} &= \Pr(X_{n+1} = j \mid X_0 = i) \\ &= \sum_{k \in S} \Pr(X_n = k \mid X_0 = i) \Pr(X_{n+1} = j \mid X_n = k) \\ &= \sum_{k \in S} p_{ik}^{(n)} p_{kj}. \end{aligned}$$

If $n = 1$, then we have $p_{ij}^{(2)} = \sum_k p_{ik} p_{kj}$. Hence $P^{(2)} = P \cdot P = P^2$. In general, $P^{(n)} = P^n$. Therefore $P^{(n+m)} = P^n \cdot P^m = P^{(n)} P^{(m)}$. We therefore have the Chapman-Kolmogorov Equation

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}.$$

# Chapter 4

# Decomposition of the State Space

**Questions.**

- Which states can be reached from a starting point $j$?
- If one state is reachable from another, is there a return path?

**Definition.** Let $\tau_j = \min\{n \geq 0 : X_n = j\}$ be the first hitting time of state $j$ (the first passage time). Let $\tau_B = \min\{n \geq 0 : X_n \in B\}$ be the first hitting time of a set $B$. Then $\tau_j$ and $\tau_B$ are discrete random variables taking the values $0, 1, 2, \ldots$ We say $\tau_j = \infty$ if $X_n$ never visits $j$.

**Definition.**

1. State $j$ is **accessible** from $i$ or $i$ **communicates** with $j$ (and we write $i \to j$) if $\Pr(\tau_j < \infty \mid X_0 = i) > 0$.
2. States $i$ and $j$ **(inter-)communicate** (and we write $i \leftrightarrow j$) if $i \to j$ and $j \to i$.
3. A set of states $C$ is **irreducible** if for all $i, j \in C$, $i \leftrightarrow j$. A Markov chain is irreducible if $i \leftrightarrow j$ for all $i, j \in S$.
4. A set of states $C$ is **closed** if for all $i \in C$, $j \notin C$, $p_{ij} = 0$ (the chain never leaves $C$) if and only if for all $i \in C$, $P_i(\tau_{C^c} = \infty) = 1$.
5. State $i$ is absorbing if $p_{ii} = 1$ (i.e. if $\{i\}$ is a closed set).

**Examples.**

1. Random walk with absorbing boundaries (gambler's ruin chain). Let $N = 3$ and $S = \{0, 1, 2, 3\}$. We have

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad a = (0, 1, 0, 0).$$

Then $\{0\}$ and $\{3\}$ are absorbing. We have $1 \leftrightarrow 2$ and $2 \to 3$, but $3 \not\to 2$. There are three classes of communicating states: $\{0\}$, $\{1, 2\}$, and $\{3\}$.

2. Random walk with reflecting boundaries. Let $S = \{0, 1, 2, 3\}$. Then

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

is irreducible.

**Lemma 4.1** We have $i \to j$ if and only if there exists $n \geq 0$ such that $p_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i) > 0$.

**Proof.** We first show the reverse direction. We know that there exists an $n > 0$ such that $p_{ij}^{(n)} > 0$. We show that $P_i(\tau_j < \infty) > 0$. We have $\{X_n = j\} \subseteq \{\tau_j \leq n\} \subseteq \{\tau_j < \infty\}$. So

$$P_i(X_n = j) \equiv p_{ij}^{(n)} \leq P_i(\tau_j < \infty).$$

Since $p_{ij}^{(n)} > 0$, we also have $P_i(\tau_j < \infty) > 0$.

We now show the forward direction. We know that $i \to j$. We show that there exists an $n > 0$ such that $p_{ij}^{(n)} > 0$. Assume that $p_{ij}^{(n)} = 0$ for all $n = 0, 1, 2, \ldots$ We show that $i \not\to j$, i.e. $P_i(\tau_j < \infty) = 0$. Then

$$\begin{aligned} P_i(\tau_j < \infty) &= \lim_{n\to\infty} P_i(\tau_j \le n) \\ &= \lim_{n\to\infty} P_i\left(\bigcup_{k=0}^{n} \{X_k = j\}\right) \\ &\le \lim_{n\to\infty} \sum_{k=0}^{n} P_i(X_k = j) \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} p_{ij}^{(k)} \\ &= 0. \end{aligned}$$

We conclude that $P_i(\tau_j < \infty) = 0$, so $i \not\to j$, which is a contradiction. ■

## 4.1 Equivalence Classes

1. Recall that $i \leftrightarrow j$ means that there exist $n, m > 0$ such that $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$.

2. The relation $\leftrightarrow$ is an equivalence relation; that is, it is reflexive ($i \leftrightarrow i$), symmetric ($(i \leftrightarrow j) \Rightarrow (j \leftrightarrow i)$), and transitive (if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$; this follows from Chapman-Kolmogorov).

3. This equivalence relation partitions the space $S$ into disjoint equivalence classes. Pick a state, say $i_0$. Put all states $i$ such that $i \leftrightarrow i_0$ into $C_0$. Pick a state $i_1 \in S \setminus C_0$. Put all states $i \leftrightarrow i_1$ into $C_1$. Continue the process. Then $S = \bigcup_i C_i$, and $C_i \cap C_j = \varnothing$ if $i \neq j$.

# Chapter 5

# Classification of States

## 5.1 Reducibility

**Definition.** A state $i$ is called **recurrent** or **persistent** if $P_i(X_n = i$ for some $n \geq 1) = 1$. A state is **transient** if $P_i(X_n = i$ for some $n \geq 1) < 1$.

We can also define these terms by using halting time $\tau_j(1) := \min\{n \geq 1 : X_n = j\}$. Then $i$ is recurrent if $P_i(\tau_i(1) < \infty) = 1$ and transient if $P_i(\tau_i(1) = \infty) > 0$.

**Question.** For recurrent states, how long does it take to return?

**Definition.** Let

$$f_{ij}^{(n)} := P_i(\tau_j(i) = n) \equiv P_i(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j).$$

Set $f_{ij}^{(0)} = 0$. Let

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \equiv \sum_{n=1}^{\infty} P_i(\tau_j(1) = n) \equiv P_i(\tau_j(1) < \infty).$$

Note that $i$ is recurrent if and only if $f_{ii} = 1$.

**Definition.** The mean recurrent time is

$$m_i := E_i(\tau_i(1)) = \begin{cases} \sum_{n=1}^{\infty} nP_i(\tau_i(1) = n) = \sum_{n=1}^{\infty} nf_{ii}^{(n)} & i \text{ recurrent,} \\ \infty & i \text{ transient.} \end{cases}$$

**Definition.** A recurrent state is **non-null** or **positive recurrent** if $m_i < \infty$ and **null** if $m_i = \infty$.

## 5.2 Calculations

We now present a recursive algorithm for calculating the $f_{ij}^{(n)}$'s. We have

$$f_{ij}^{(n)} = \begin{cases} p_{ij} & n = 1, \\ \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)} & n > 1. \end{cases}$$

**Proof.** Suppose $n = 1$. Then $f_{ij}^{(1)} = P_i(\tau_j(1) = 1) = P_i(X_1 = j) = p_{ij}$. Now suppose $n > 1$. Then

$$\begin{aligned} f_{ij}^{(n)} &= P_i(\tau_j(1) = n) \\ &= P_i(X_1 \neq j, \ldots, X_{n-1} \neq j, X_n = j) \\ &= \sum_{k \neq j} P_i(X_1 = k, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j) \\ &= \sum_{k \neq j} P_i(X_2 \neq j, \ldots, X_{n-1} \neq j, X_n - j \mid X_1 = k) P_i(X_1 = k) \\ &= \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}, \end{aligned}$$

as desired. ∎

**Proposition 5.1**

- For all $i, j \in S$, $p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}$ whenever $n \geq 1$.

- For all $0 < s < 1$ and $i \neq j$, $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$, while $P_{ii}(s) = 1 + F_{ii}(s)P_{ii}(s)$.

**Proof.**

- $$\begin{aligned} p_{ij}^{(n)} &= P_i(X_n = j) \\ &= \sum_{k=1}^{n} P_i(X_n = j, \tau_j(1) = k) \\ &= \sum_{k=1}^{n} P_i(X_n = j \mid \tau_j(1) = k) P_i(\tau_j(1) = k) \\ &= \sum_{k=1}^{n} \Pr(X_n = j \mid X_0 = i, X_1 \neq j, \ldots, X_{k-1} \neq j, X_k = j) f_{jj}^{(k)} \\ &= \sum_{k=1}^{n} f_{ij}^{(k)} f_{jj}^{(n-k)}, \end{aligned}$$

  as desired.

- $$\begin{aligned} \sum_{n=1}^{\infty} s^n p_{ij}^{(n)} &= \sum_{n=1}^{\infty} s^n \left( \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \right) \\ &= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} s^{n-k} p_{jj}^{(n-k)} s^k f_{jj}^{(k)} \\ &= \sum_{k=1}^{\infty} s^k f_{ij}^{(k)} \sum_{m=0}^{\infty} s^m p_{jj}^{(m)} \\ &= F_{ij}^{(s)} P_{jj}(s), \end{aligned}$$

  as desired. ■

**Proposition 5.2**

1. $j$ is recurrent iff $\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$ iff $f_{jj} = 1$. If this holds, then $\sum_n p_{ij}^{(n)} = \infty$ for all $i$ such that $f_{ij} > 0$.

2. $j$ is transient iff $\sum_n p_{jj}^{(n)} < \infty$ iff $f_{jj} < 1$. If this holds, then $\sum_n p_{ij}^{(n)} < \infty$ for all $i$.

**Proof.** We show only (1).

1. We have
$$\lim_{s \nearrow 1} P_{jj}(s) = \lim_{s \nearrow 1} \sum_n s^n P_{jj}^{(n)} = \sum_n p_{jj}^{(n)}.$$
Also
$$\lim_{s \nearrow 1} P_{jj}(s) = \lim_{s \nearrow 1} \frac{1}{1 - F_{jj}(s)}.$$
Also
$$\lim_{s \nearrow 1} (1 - F_{jj}(s)) = 0$$
iff $F_{jj}(1) = \sum_n f_{jj}^{(n)} = f_{jj} = 1$. If $i \neq j$, then
$$\sum_n p_{ij}^{(n)} \equiv P_{ij}(1) = F_{ij}(1) P_{jj}(1) = f_{ij} \sum_n p_{jj}^{(n)} = \infty$$
iff $f_{ij} > 0$ and $\sum_n p_{jj}^{(n)} = \infty$, as desired. ■

## 5.3 Markov and Strong Markov Property

We called $(X_n)$ a homogeneous Markov chain if

$$\Pr(X_{n+1} = j \mid X_0 = i_0, \ldots, X_n = i_n) = \Pr(X_{n+1} = j \mid X_{n=1} = i_n) = \P_{i_n}(X_n = j) = p_{i_n,j}.$$

Then
$$\Pr(X_0 = i_0, \ldots, X_n = i_n) = a_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}.$$

Write $\boldsymbol{\delta}_i = (\delta_{ij}, j \in S)$.

**Theorem 5.3** (Markov Property.) Let $(X_n)$ be a Markov chain with parameters $(\mathbf{a}, P)$. Then, conditional on $X_m = i$, $(\widetilde{X}_n := X_{m+n})_{n\geq 0}$ is Markov with parameter $(\boldsymbol{\delta}_i, P)$ and is independent of the random variables $X_0, \ldots, X_m$.

**Proof.** Take $A = \{X_0 = i_0, \ldots, X_m = i_m\}$. We show that

$$\begin{aligned}
&\Pr(\{\widetilde{X}_0 \equiv X_m = j_0, \ldots, \widetilde{X}_m \equiv X_{m+n} = j_n\} \cap A \mid X_m = i) \\
&= \Pr(\{X_0 = j_0, \ldots, X_n = j_n\} \mid X_0 = i) \Pr(A \mid X_m = i) \\
&\equiv \delta_{ij_0} p_{j_0 j_1} \cdots p_{j_{n-1} j_n} \Pr(A \mid X_m = i).
\end{aligned}$$

We have

$$\begin{aligned}
LHS &= \frac{\Pr(X_0 = i_0, \ldots, X_m = i_m = i, \ldots, X_{m+n} = j_n \text{ and } i_m = j_0 = i)}{\Pr(X_m = i)} \\
&= \Pr(X_m = j_0, X_{m+1} = j_1, \ldots, X_{m+n} = j_n \mid X_0 = i_0, \ldots, X_{m-1} = i_{m-1}, X_m = i_m) \\
&\quad \times \frac{\Pr(X_0 = i_0, \ldots, X_{m-1} = i_{m-1}, X_m = i_m \text{ and } i_m = i)}{\Pr(X_m = i)} \delta_{ij_0} \\
&= \delta_{ij_0} p_{j_0 j_1} \cdots p_{j_{n-1} j_n} \Pr(A \mid X_m = i).
\end{aligned}$$

### 5.3.1 Stopping Time

**Definition.** A random variable $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a **stopping time** if the event $\{T_n\}$ depends only on $X_0, \ldots, X_n$ (for $n = 0, 1, 2, \ldots$).

**Example.**

- $T = \tau_j(1) = \min\{n \geq 1 : X_n = j\}$ is a stopping time since $\{T = n\} \equiv \{\tau_j(1) = n\} = \{X_1 \neq j, \ldots, X_{n-1} \neq j, X_n = j\}$ depends on $X_1, \ldots, X_n$.
- $T = \tau_j(1) - 1$ is not a stopping time since $\{T = \tau_j(1) - 1 = n\} = \{\tau_j(1) = n+1\}$ depends on $X_{n+1}$.

### 5.3.2 Strong Markov Property

**Theorem 5.4** Let $X = (X_n)_{n\geq 0}$ be Markov with parameters $(\mathbf{a}, P)$, and let $T$ be a stopping time of $X$. Then, conditional on $\{T < \infty\}$ and $X_T = i$, $(\widetilde{X}_n = X_{T+n})_{n\geq 0}$ is Markov with parameters $(\boldsymbol{\delta}_i, P)$ and is independent of $X_0, \ldots, X_{T-1}$.

**Proof.** Let $A$ be an event determined by $X_0, \ldots, X_T$. We want to show that

$$\begin{aligned}&\Pr(\{\widetilde{X}_0 = X_T = j_0, \ldots, \widetilde{X}_n = X_{p_n} = j_n\} \cap A \mid T < \infty, X_T = i)\\ &= \Pr(\{X_0 = j_0, \ldots, X_n = j_n \mid X_0 = i) \Pr(A \mid T < \infty, X_T = i).\end{aligned}$$

We have

$$\begin{aligned} LHS &= \sum_{m=0}^{\infty} \frac{\Pr(\{X_T = j_0, \ldots, X_{T+n} = j_n\} \cap A \cap \{T = m\} \cap \{X_T = i\})}{\Pr(T < \infty, X_T = i)} \\ &= \sum_{m=0}^{\infty} \frac{\Pr(\{X_m = j_0, \ldots, X_{m+n} = j_n\} \cap A \cap \{T = m\})}{\Pr(T < \infty, X_T = i)} \\ &= \sum_{m=0}^{\infty} \frac{\Pr(X_0 = j_0, \ldots, X_n = j_n \mid X_0 = i) \Pr(A \cap \{T = m\} \cap \{X_T = i\})}{\Pr(T < \infty, X_T = 1)} \\ &= RHS, \end{aligned}$$

as desired. ■

## 5.4 Excursions

Assume $X_0 = i$. Then $\tau_i(1) = \min\{n \geq 1 : X_n = i\}$. On $\{\tau_i(1) < \infty\}$, define $\tau_i(2) = \min\{n > \tau_i(1) : X_n = 1\}$. On $\{\tau_i(1) < \infty, \ldots, \tau_i(n) < \infty\}$, define $\tau_i(n+1) = \min\{m > \tau_i(n) : X_m = i\}$. The block $(X_{\tau_i(n-1)+1}, \ldots, X_{\tau_i(n)})$ is the $n^{\text{th}}$ excursion. The length is $\alpha_n = \tau_i(n) - \tau_i(n-1)$, and $\alpha_0 = 0$, $\alpha_i = \tau_i(1)$.

**Proposition 5.5** Conditional on $\tau_i(1) < \infty$ and $\tau_i(2) < \infty$, random vectors $(\alpha_1, X_1, \ldots, X_{\tau_i(1)})$ and $(\alpha_2, X_{\tau_i(1)+1}, \ldots, X_{\tau_i(2)})$ are independent and identically distributed.

## 5.5 Recurrence and the Number of Visits to a State

Let $N_j = \sum_{n=1}^{\infty} I_{\{X_n=j\}}$ be the number of visits to state $j$ after time zero. Define $\mathbb{1}_A = \chi_A = I_A = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A. \end{cases}$ Calculate

$$E_i N_j = \sum_{n=1}^{\infty} E_i I_{\{X_n=j\}} = \sum_{n=1}^{\infty} P_i(X_n = j) = \sum_{n=1}^{\infty} p_{ij}^{(n)}.$$

Then the result Proposition 5.2 says that when starting from $j$ ($X_0 = j$), $j$ is recurrent iff

$$E_j N_j \equiv \sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty.$$

**Proposition 5.6** For all $i, j \in S$ and $k \geq 0$,

$$P_i(N_j = k) = \begin{cases} 1 - f_{ij} & k = 0, \\ f_{ij}(f_{jj})^{k-1}(1 - f_{jj}) & k = 1. \end{cases}$$

**Conclusions.** If $j$ is transient ($f_{jj} < 1$), i.e. $N_j \sim \text{Geom}(f_{jj})$,

- $P_j(N_j = k) = f_{jj}^k(1 - f_{jj})$ for $k \geq 0$.
- $E_i N_j \equiv \sum_{n=1}^{\infty} p_{ij}^{(n)} = \frac{f_{ij}}{1-f_{jj}} < \infty$.
- For all $i$, $P_i(N_j < \infty) = 1$ or $P_i(N_j = \infty) = 0$. (The chain returns to state $j$ finitely often.)

If $j$ is recurrent, then

- $P_j(N_j = k) = 0$ for all $k$.
- $P_j(N_j < \infty) = 1$, i.e. $P_j(\{X_n = j\}$ infinitely often$) = 1$.

**Proof.** $P_i(N_j = 0) = 1 - f_{ij}$, the probability that the chain never hits $j$ from $i$. Then

$$\begin{aligned} P_i(N_j \geq n+1 \mid N_j \geq n) &\equiv \Pr(j \text{ was hit at least } n \text{ times — known; } j \text{ was hit at least once more}) \\ &= \Pr(\text{returning to } j \text{ from } j \text{ at least once}) \\ &= f_{jj}. \end{aligned}$$

The rest follows easily. ■

## 5.6 Periodicity

**Example.** Simple random walk with steps $\pm 1$. The chain can only return to state 0 on even steps.

**Definition.**

- The period of a state $i$ is defined as $d(i) = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$.
- If $d(i) = 1$, then $i$ is **aperiodic**.
- If $d(i) > 1$, then $i$ is **periodic**.
- A state is **ergodic** if it is positive recurrent and aperiodic.

# Chapter 6

# Canonical Decomposition

**Proposition 6.1** If $i \leftrightarrow j$, then

1. $i$ and $j$ have the same period.
2. $i$ is transient if and only if $j$ is transient.
3. $i$ is null recurrent if and only if $j$ is null recurrent.

**Proof.** See pages 92–93 of Resnick. ■

**Proposition 6.2** The state space $S$ may be decomposed as $S = T \cup C_1 \cup C_2 \cup \cdots$ (see Section 4.1, part (iii)), where $T$ is the set of transient states (not necessarily one class of intercommunicating states), and $C_1, C_2, \ldots$ are closed, disjoint, irreducible classes of recurring states. If $j \in C_\alpha$, then $f_{jk} = \begin{cases} 1 & k \in C_\alpha, \\ 0 & \text{otherwise.} \end{cases}$ For each class $C_\alpha$, we obtain a stochastic matrix $P_\alpha$ by considering only rows and columns of $P$ for states in $C_\alpha$. Then after reordering states, $P$ can be written as

$$P = \begin{pmatrix} P_1 & 0 & \cdots \\ 0 & P_2 & \cdots \\ \vdots & \vdots & \cdots \\ Q_1 & Q_2 & \cdots \end{pmatrix}.$$

Transitions from states in $T$ are governed by the matrices $Q_i$.

**Example.** Gambler's ruin chain, $S = \{0, 1, 2, 3\}$. In this case, $C_1 = \{0\}$, $C_2 = \{3\}$, and $T = \{1, 2\}$. Then we have

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

The rows and columns, in order, correspond to states 0, 3, 1, 2.

**Note.** If $X_0 \in C_\alpha$, since the chain never leaves $C_\alpha$, it can be taken as the whole space.

## 6.1 Finite Markov Chains

**Lemma.**

1. If $S$ is finite, then at least one state is recurrent.
2. All recurrent states are positive.

For all $i$, $\sum_{j \in S} p_{ij}^{(n)} = 1$. We can then let $n \to \infty$ to obtain $\sum_{j \in S} \left( \lim_{n \to \infty} p_{ij}^{(n)} \right) = 1$. If $j$ is transient, then $p_{ij}^{(n)} \xrightarrow{n \to \infty} 0$.

If $S$ is finite, first look for all recurrent states. At least one exists by the Lemma. Look for all states which intercommunicate with this one.

## 6.2 Random Walk Example

Consider a simple random walk. Let $\xi_i$ be independent and identically distributed, with $\Pr(\xi_i = 1) = p$ and $\Pr(\xi_i = -1) = 1 - p = q$. We set $S_0 = 0$ and $S_n = \xi_1 + \cdots + \xi_n$.

The strong law of large numbers tells us that if $xi_i$ are independent and identically distributed and $E\xi_i^2 < \infty$, then $\Pr\left(\left\{\omega : \frac{1}{n}\sum_{i=1}^n \xi_i(\omega) \stackrel{n\to\infty}{\longrightarrow} E\xi_i = \mu\right\}\right) = 1$. If $p = q = 1/2$, then $\Pr(S_n/n \to 0) = 1$. If $p \neq q$, then $\Pr(S_n/n \to p - q \neq 0) = 1$. In particular, if $p > q$, then $S_n(\omega) \to +\infty$ for almost all $\omega$, and if $p < q$, then $S_n(\omega) \to -\infty$ for almost all $\omega$.

We can conclude that for asymmetric random walks ($p \neq q$), 0 is transient. Therefore every state is transient. However, if $p = q$, then $\frac{S_n(\omega)}{n} \to 0$.

**Lemma.** If $p = q = 1/2$, then 0 is recurrent.

**Proof.** By Proposition 5.2, 0 is recurrent iff $\sum_n p_{00}^{(n)} = \infty$. We have $p_{00}^{(2n+1)} = 0$. We also have

$$p_{00}^{(2n)} = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}.$$

Stirling's formula tells us that $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$. Hence $p_{00}^{(n)} \approx \frac{1}{\sqrt{\pi n}}$. This implies that $\sum_n p_{00}^{(n)} = \infty$.

**Definition.** The $d$-dimensional symmetric random walk is given by $\mathbf{S}_0 = 0$, $\mathbf{S}_n = \boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_n$, where $\boldsymbol{\xi}_i = (\xi_i^{(1)}, \ldots, \xi_i^{(d)}) \in \{-1, 1\}^d$. The components are independent and identically distributed. Hence $\Pr(\boldsymbol{\xi}_i = (i_1, \ldots, i_d)) = \left(\frac{1}{2}\right)^d$, where each $i_d = \pm 1$.

**Lemma.** For symmetric random walks in $\mathbb{Z}^d$, 0 is recurrent if $d = 1$ or 2; 0 is transient if $d \geq 3$.

**Proof.** We have

$$p_{00}^{(2n)} = P_0(\mathbf{S}_{2n} = 0) = (P_0(S_{2n}^{(1)} = 0)^d) \simeq (\pi n)^{-d/2}.$$

Hence we have

$$\sum_n p_{00}^{(n)} \simeq c \sum \frac{1}{n^{d/2}} = \begin{cases} \infty & d = 1, 2; \\ \text{finite} & d \geq 3. \end{cases}$$

## 6.3 Branching Process Example

Let $(Z_n)$ be a Galton-Watson branching process. We have $Z_0 = 1$ and $Z_n = Z_{n,1} + \cdots + Z_{n,Z_{n-1}}$.

**Theorem 6.3** Let $p_1 = \Pr(Z_{n,j} = 1) < 1$. Then $\Pr(Z_n = j) \overset{n\to\infty}{\longrightarrow} 0$ for $j = 1, 2, \ldots$

**Proof.** This is a question of classification of states. 0 is absorbing and hence recurrent. We now show that $j = 1, 2, \ldots$ are transient. We have $p_{ij}^{(n)} \overset{n\to\infty}{\longrightarrow} 0$. We have transience if and only if

$$f_{jj} = \Pr(Z_{m+n} = j \text{ for some } n \geq 1 \mid Z_m = j) < 1.$$

Suppose that $p_0 = \Pr(Z_{n,j} = 0) = 0$ (i.e. there are no deaths without children). Then $f_{jj}^{(1)} = \Pr(Z_{m+1} = j \mid Z_m = j) = (p_1)^j$. We have $f_{jj}^{(2)} = \Pr(Z_{m+1} \neq j, Z_{m+2} = j \mid Z_m = j) = 0$, so $f_{jj}^{(n)} = 0$ for $n \geq 2$. Hence $f_{ij} = f_{ij}^{(1)} = (p_1)^j < 1$.

Now suppose $p_0 = \Pr(Z_{n,j} = 0) < 1$. Then $\{Z_{m+n} = j \text{ for some } n \geq 1\} \subseteq \{Z_{m+1} > 0\}$. Then

$$f_{jj} \leq \Pr(Z_{m+1} > 0 \mid Z_m = j) = 1 - \Pr(Z_{m+1} = 0 \mid Z_m = j) = 1 - (p_0)^j < 1,$$

as desired. ∎

# Chapter 7

# Steady State

The goal of this chapter is to work out the long-term behavior of the chain. What happens when $n$ is large?

## 7.1 Stationarity of a Markov Chain: Stationary Distributions

**Definition.** $\{Y_n, n \geq 0\}$ is **(strictly) stationary** if for all integers $m \geq 0$, $k > 0$, we have $(Y_0, \ldots, Y_m) \stackrel{\Delta}{=} (Y_k, \ldots, Y_{k+m})$, i.e. distribution doesn't change under time-shift translation.

**Definition.** A vector $\pi = (\pi_j, j \in S)$ is a stationary distribution of the Markov chain $(\cdot, P)$ if

1. $\pi_j \geq 0$, $\sum_{j \in S} \pi_j = 1$ (distribution),
2. $\pi = \pi P$, i.e. $\pi_j = \sum_{i \in S} \pi_i p_{ij}$ for all $j \in S$.

**Note.** $\pi P^2 = (\pi P)P = \pi P = P$, and in general $\pi P^n = \pi$.

**Proposition 7.1** Denote by $P_\pi$ the distribution of the Markov chain $(\pi, P)$, i.e. $\Pr(X_0 = j) = \pi_j$ or $P_\pi(\cdot) = \sum_{j \in S} \pi_j \Pr(\cdot \mid X_0 = j)$. With respect to $P_\pi$, $(X_n)$ is a strictly stationary process:

$$P_\pi(X_n = i_0, \ldots, X_{n+m} = i_m) = \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} \equiv P_\pi(X_0 = i_0, \ldots, X_m = i_m).$$

**Proof.**

$$\begin{aligned} LHS &= \sum_{i \in S} \pi_i \Pr(X_n = i_0, \ldots, X_{n+m} = i_m \mid X_0 = i) \\ &= \sum_{i \in S} \pi_i p_{i i_0}^{(n)} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} \\ &= \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}, \end{aligned}$$

as desired. ■

**Definition.** A vector $\gamma = (\gamma_j, j \in S)$ is an **invariant measure** if

1. $\gamma_j \geq 0$,
2. $\gamma = \gamma P$.

(Of course, if $\sum \gamma_i < \infty$, then $\pi_i = \frac{\gamma_i}{\sum \gamma_j}$ is a stationary distribution.)

## 7.2 A Fundamental Result on the Existence and Uniqueness of Invariant Measures and Stationary Distributions

**Theorem 7.2** An *irreducible* Markov chain has a stationary distribution $\pi$ iff all states are positive recurrent. In this case, $\pi$ is the *unique* stationary distribution and is given by $\pi_j = \frac{1}{m_j}$, $j \in S$, where $m_j = E_i \tau_j(1)$. Also in this case, the equation $\pi = \pi P$ has a positive root which is unique up to a multiplicative constant, and for which $\sum x_i < \infty$. If the chain is irreducible but *null recurrent*, the previous statement holds, but $\sum x_i = \infty$.

**Proof.**

1. We first show that for a transient irreducible Markov chain, a stationary distribution doesn't exist. Suppose $\pi P^n = \pi$, so $\pi_j = \sum_{i\in S} \pi_i p_{ij}^{(n)}$. Letting $n \to \infty$, we have $p_{ij}^{(n)} \to 0$ (since the chain is transient). Hence $\pi_j = \sum_{i\in S} \pi_i \lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all $j$, which contradicts the assumption that $\sum \pi_j = 1$.

2. For an irreducible recurrent Markov chain, we construct an invariant measure. Consider one excursion $i \to i$. Define $N_j^{(i)} = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=j\}\cap\{\tau_j(1)\geq n\}}$, and let $\gamma_j^{(i)} = \sum_{n=1}^{\infty} P_i(X_n = j, \tau_j(1) \geq n) = E_j N_j^{(i)}$ be the expected number of visits to $j$ between two successive visits to $i$. Note that $N_i^{(i)} = 1$, $\gamma_i^{(i)} = 1$, $\tau_i(1) = \sum_{j\in S} N_j(1)$, and

$$m_i(1) = E_i\tau_i(1) = \sum_{j\in S} E_i N_j^{(i)} = \sum_{j\in S} \gamma_j^{(i)}. \qquad (*)$$

We prove that $\gamma^{(i)} = (\gamma_j^{(i)}, j \in S)$ is invariant: $\gamma^{(i)} = \gamma^{(i)}P$. Set $\ell_{ij}^{(n)} := P_i(X_n = j, \tau_i(1) \geq n)$ so that $\gamma_j^{(i)} = \sum_{n=1}^{\infty} \ell_{ij}^{(n)}$. If $n = 1$, we have $\ell_{ij}^{(1)} = P_i(X_1 = j, \tau_i(1) \geq 1) = p_{ij}$. If $n \geq 2$, we have $\ell_{ij}^{(n)} = \sum_{k\neq i} ell l_{ik}^{(n-1)} p_{kj}$. Then we have

$$\begin{aligned}
\gamma_j^{(i)} &= \sum_{n=1}^{\infty} \ell_{ij}^{(n)} \\
&= p_{ij} + \sum_{n=2}^{\infty}\sum_{k\neq j} \ell_{ik}^{(n-1)} p_{kj} \\
&= p_{ij} + \sum_{k\neq i}\left(\sum_{n=2}^{\infty} \ell_{ij}^{(n-1)}\right) p_{kj} \\
&= \gamma_i^{(i)} p_{ij} + \sum_{k\neq i} \gamma_k^{(i)} p_{kj}.
\end{aligned}$$

We have showed that $\gamma_j^{(i)} = \sum_k \gamma_k^{(i)} p_{kj}$.

3. If $i$ is positive recurrent, i.e. $m_i = E_i\tau_i(1) < \infty$, then by $(*)$ we have $\sum_j \gamma_j^{(i)} = m_i < \infty$, and so $\{\gamma_j^{(i)}/m_i, j \in S\}$ is a stationary distribution.

4. For all recurrent states $i$, we show that $0 < \gamma_j^{(i)} < \infty$. Since the Markov chain is irreducible, for each $j \in S$ there exists an $m > 0$ such that $p_{ji}^{(m)} > 0$. Since $\gamma^{(i)} = \gamma^{(i)} P^m$, we have

$$1 = \gamma_i^{(i)} = \sum_{k\in S} \gamma_k^{(i)} p_{ki}^{(m)} \geq \gamma_j^{(i)} p_{jj}^{(m)} > 0,$$

so $\gamma_j^{(i)} \leq \frac{1}{p_{ji}^{(m)}} < \infty$ for all $j$. Since $i \to j$, there exists an $m$ so that $p_{ij}^{(m)} > 0$. So

$$\gamma_j^{(i)} = \sum_k \gamma_k^{(i)} p_{kj}^{(m)} \geq \gamma_i^{(j)} p_{ij}^m = p_{ij}^{(m)} > 0.$$

5. If a Markov chain is irreducible and recurrent, then an invariant measure $\gamma$ is unique up to a multiplicative constant; if $\gamma = \gamma P$ and $\mu = \mu P$ $(0 < \mu_j < \infty)$, then $\mu =$const$\cdot T$. (The proof is in Resnick.)

6. From (5), it follows that $\gamma^{(i)} =$const$\cdot\gamma^{(k)}$ for all $i, k \in S$. Since $\gamma_i^{(i)} = 1$ in the positive recurrent case, $\pi_i = \frac{\gamma_i^{(i)}}{m_i} = \frac{1}{m_i}$. ■

**Remarks.**

1. If we have an irreducible Markov chain which is recurrent, then there exists an invariant measure $\gamma$ so that $\gamma = \gamma P$ which is unique up to a multiplicative constant $C\gamma = (c\gamma)P$. If the Markov chain is positive recurrent, then $\sum_j \gamma_j < \infty$, and there exists a unique stationary distribution $\pi_j = \frac{\gamma_j}{\sum \gamma_j}$. If the Markov chain is null recurrent, then $\sum \gamma_j = \infty$, and there is no stationary distribution. (If the Markov chain is transient, then there is no stationary distribution.)

2. Because the Markov chain can be decomposed into $S = T \cup C_1 \cup C_2 \cup \cdots$, it is enough to consider irreducible Markov chains.

3. We have a new method to calculate mean recurrence times: $m_j = \frac{1}{\pi_j}$, where $\pi = \pi P$.

4. We have a new method to determine whether an irreducible Markov chain is positive recurrent: solve $\pi = \pi P$. See whether there exists a unique solution so that $\sum_j \pi_j = 1$.

## 7.3 Examples

There is no uniqueness without irreducibility. Consider the gambler's ruin chain $S = \{0, 1, 2, 3\}$ with

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Take $\alpha \in [0, 1]$, and set $\pi_\alpha = (\alpha, 0, 0, 1 - \alpha)$. Then $\pi_\alpha P = \pi_\alpha$. Hence there are infinitely many stationary distributions.

Consider the simple symmetric random walk on $\mathbb{Z}$. We showed that it is recurrent, but we can now show that it is null recurrent. First note that it is irreducible with $p_{ij} = \begin{cases} 1/2 & |i-j| = 1, \\ 0 & \text{otherwise.} \end{cases}$ If it is positive recurrent, then there must be a unique stationary distribution. We must then have $\pi_j = \frac{1}{2}\pi_{j-1} + \frac{1}{2}\pi_{j+1}$. Thus $\pi = (1, 1, \ldots)$ is an invariant measure. Since $\sum \pi_j = \infty$, there is no stationary distribution; hence the chain is null recurrent.

Consider the asymmetric random walk on $\mathbb{Z}$ with $p_{ij} = \begin{cases} p & j = i+1, \\ q = 1-p & j = i-1, \\ 0 & \text{otherwise.} \end{cases}$ Assume that $p > q$. We showed that the chain is transient, and it is also irreducible. Suppose $\pi = \pi P$. Then $\pi_j = q\pi_{j-1} + p\pi_{j+1}$. We can check (see the problems) that $\pi + j = A + B\left(\frac{p}{q}\right)^j$ is a solution for all $A, B > 0$. There are infinitely many invariant measures. See Karlin and Taylor for more details.

# Chapter 8

# Limit Theorems

**Questions.**

- Does $\lim_{n\to\infty} p_{ij}^{(n)}$ exist?
- Do we visit $j$ "at $n = \infty$"?

We know that if $j$ is transient, then $\lim_{n\to\infty} p_{ij}^{(n)} = 0$, and if the limit distribution exists, it must be stationary.

**Lemma 8.1** If for all $j$, $p_{ij}^{(n)} \to \pi)j$, where $(\pi_j)$ is a distribution, then $\pi = (\pi_j, j \in S)$ is a stationary distribution.

**Proof.**

- Suppose $S$ is finite. Then if $p_{ij}^{(n)} \to \pi_j$, then $(\pi_j)$ is a distribution because $1 \geq \pi_j \geq 0$, and we have

  $$\sum_j \pi_j = \sum_j \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_j p_{ij}^{(n)} = 1,$$

  where we can interchange the sum and the limit because the sum is finite. We then have

$$\begin{aligned}\pi_j &= \lim_{n\to\infty} p_{ij}^{(n)} \\ &= \lim_{n\to\infty} \sum_k p_{ik}^{(n-1)} p_{kj} \\ &= \sum_k \left( \lim_{n\to\infty} p_{ik}^{(n-1)} \right) p_{kj} \\ &= \sum_k \pi_k p_{kj}.\end{aligned}$$

Hence the distribution is stationary.

- For the proof when $S$ is countable, see pages 126–127 of Resnick. ■

For periodic states, the limit might not exist.

**Example.** Let $S = \{0, 1\}$ and

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

Then the period is $d = 2$, so $p_{11}^{(n)} = (0, 1, 0, 1, \ldots)$. Hence $\lim_{n\to\infty} p_{ii}^{(n)}$ does not exist. Note that there exists a unique stationary distribution $\pi = (1/2, 1/2)$ and that $p_{ii}^{(2n)} = \lim 1 = 1 = 1/2 \cdot 2 = d\pi_i$.

**Theorem 8.2** (Ergodic Theorem.) Let $(X_n)$ be an irreducible positive recurrent Markov chain with stationary distribution $\pi$. If the chain is aperiodic, then there exists $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j = \frac{1}{m_j}$. If the chain is periodic with period $d$, then for all $i, j \in S$, there exists an integer $r$ with $0 \le r \le d$ such that $p_{ij}^{(n)} = 0$ unless $n = md + r$ for some $m \ge 0$, and there exists $\lim_{n\to\infty} p_{ij}^{(md+r)} = d\pi_j = \frac{d}{m_j}$. For null recurrent chains, these results hold with $m_j = \infty$, namely there exists $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ in the aperiodic case, and there exists $\lim_{n\to\infty} p_{ij}^{(md+r)} = 0$ in the periodic case.

**Note.** If $(X_n)$ is irreducible and aperiodic, the stationary distributions exist iff the chain is positive recurrent iff the limit distributions exist, so that $p_{ij}^{(n)} \to \pi_j$.

**Note.** If $(X_n)$ is irreducible, aperiodic, and finite, then it is positive recurrent, its stationary distributions exist, and $p_{ij}^{(n)} \to \pi_j$.

**Note.** In the ergodic case, $p_{ij}^{(n)} \to \pi_j = \frac{1}{m_j}$, and chain forgets its origin:

$$\Pr(X_n = j) = \sum_i \Pr(X_0 = i) p_{ij}^{(n)} \stackrel{n\to\infty}{\longrightarrow} \sum_j \Pr(X_0 = i) \lim_{n\to\infty} p_{ij}^{(n)}.$$

**Proof of Theorem 8.2** We use the coupling method.

1. "Couple" $X$ with $Y$, creating a Markov chain $Y$ with stationary distribution $\pi$, and let $\xi_n = (X_n, Y_n)$. Let $X_0 = i$ for some $i$. $Y$ starts according to $\pi$: $\Pr(Y_0 = j) = \pi_j$. $X$ and $Y$ are independent. By the problems, $(\xi_n)$ is a Markov chain on $S \times S$ with transition probabilities $(\widetilde{p}_{(i,j)(k,\ell)} = p_{ik}p_{j\ell})$ and stationary distributions $\widetilde{\pi}_{(i,k)} = \pi_i \pi_k$ (and $(\xi_n)$ is positive recurrent).

2. Fix $i_0 \in S$, and set $\tau := \min\{n \geq 0 : X_n = Y_n = i_0\}$. Then $\tau$ is the first hitting time of $(\xi_n)$ to $(i_0, i_0)$. Because $(\xi_n)$ is positive recurrent, $\Pr(\tau < \infty) = 1$. By the strong Markov property for $\xi_n$, $\xi_{\tau+n} = (X_{\tau+n}, Y_{\tau+n})$ is Markov with parameters $(\delta_{(i_0,i_0)}, \widetilde{P})$ and is independent of $(X_0, Y_0), \ldots, (X_\tau, Y_\tau)$. The main point is that $X_\tau = Y_\tau = i_0$, and both have transition matrix $P$, so $(X_{\tau+n})_{n\geq 0} \stackrel{\Delta}{=} (Y_{\tau+n})_{n\geq 0}$.

3. We show that

   $$|p_{ij}^{(n)} - \pi_j| = |\Pr(X_n = j) - \Pr(Y_n = j)| \stackrel{n\to\infty}{\longrightarrow} 0.$$

   Note that $\Pr(Y_n = j) = \pi_j$ because $(Y_n)$ is Markov with parameters $(\pi, P)$. We have

$$
\begin{aligned}
|\Pr(X_n = j) - \Pr(Y_n = j)| &\le |\Pr(X_n = j, \tau \le n) - \Pr(Y_n = j, \tau \le n)| \\
&\qquad + |\Pr(X_n = j, \tau > n) - \Pr(Y_n = j, \tau > n)| \\
&= 0 + |\Pr(X_n = j, \tau > n) - \Pr(Y_n = j, \tau > n)| \\
&= E(\mathbb{1}_{\{X_n=j\}}\mathbb{1}_{\{\tau>n\}} - \mathbb{1}_{\{Y_n=j\}}\mathbb{1}_{\{\tau>n\}}) \\
&\le E\left(|\mathbb{1}_{\{X_n=j\}} - \mathbb{1}_{\{Y_n=j\}}|\mathbb{1}_{\{\tau>n\}}\right) \\
&\le E\mathbb{1}_{\{\tau>n\}} \\
&= \Pr(\tau > n) \\
&\xrightarrow{n\to\infty} 0
\end{aligned}
$$

because $\Pr(\tau < \infty) = 1$. ■

What goes wrong in the periodic case?

**Example.** Suppose we have $S = \{0, 1\}$ and $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $\pi = (1/2, 1/2)$. If $P(X_0 = 0) = 1$ and $Y_0 = 1$ (this happens with probability 1/2), then they never meet.

# Chapter 9

# Time Averages

In class, we stated most of the formulae to look something like

$$\frac{1}{N}\sum_{n=0}^{N} f(X_n).$$

I know that this is equivalent to what I'm writing here, but it just seems like such a silly way of thinking, so I changed it to $\frac{1}{N+1}$. Do statisticians really insist on dividing by $N$ when dealing with $N+1$ terms?

Recall the following results:

1. The strong law of large numbers for independent identically distributed random variables. If $\{Y_n\}$ is a sequence of independent identically distributed random variables with $E|Y_i| < \infty$ and $EY_i = \mu$, then

$$\Pr\left(\omega : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i(\omega) = \mu\right) = 1,$$

   or we write

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$

   with probability 1 or almost surely (a.s.).

2. Let $(X_n)$ be a Markov chain and $f : S \to \mathbb{R}$ a bounded or positive function.

   - We interpret $f(i)$ as the reward or cost of having the process at state $j$.

- We consider

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(X_n)$$

to be the average long-term reward (cost) rate.

- If $f(k) = I_j(k) = \delta_{kj}$, then $f(X_n) = \mathbb{1}_{\{X_n=j\}} \equiv I_j(X_n)$. Then

$$\frac{1}{N+1} \sum_{n=0}^{N} I_j(X_n) = \frac{\text{number of visits to } j \text{ on steps } 0, \ldots, N}{N+1}$$

is the relative frequency or proportion of time $(X_0, \ldots, X_N)$ spent in $j$.

**Proposition 9.1** Let $X$ be irreducible and positive recurrent, and let $\pi$ be the unique stationary distribution of $X$. Let $f : S \to \mathbb{R}$ be bounded or positive. Then for all initial distributions of $X$, we have

$$\Pr\left(\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} f(X_n) = \pi(f) := \sum_{k\in S} \pi_k f(k)\right) = 1.$$

The left side is

$$\frac{f(X_0) + \cdots + f(X_N)}{N+1},$$

the sample average.

**Corollary 9.2** For bounded functions $f$, for all $i \in S$

$$\frac{1}{N+1} \sum_{n=0}^{N} E_i(f(X_n)) \to \pi(f).$$

**Proof.** By Proposition 9.1,

$$Y_N := \frac{1}{N+1} \sum_{n=0}^{N} f(X_n) \to \pi(f)$$

with probability 1. We have

$$|Y_N| \leq \sup_{x \in S} |f(x)| = \mu < \infty.$$

By the bounded convergence theorem,

$$E_i Y_N \to E\pi(f) = \pi(f). \qquad \blacksquare$$

**Corollary 9.3**

$$\frac{1}{N} \sum_{n=1}^{N} p_{ij}^{(n)} \to \pi_j$$

as $N \to \infty$.

**Proof.** Take $f(k) = I_j(k)$. Then

$$E_i I_j(X_n) = P_i(X_n = j) = p_{ij}^{(n)},$$

and $\pi(I_j) = \pi_j$. Then use Corollary 9.2. $\blacksquare$

**Example.** Let $f(k) = \mathbb{1}_{\{i\}}(k)$. Then

$$\pi(f) = \sum_k \mathbb{1}_{\{j\}}(k) \pi_k = x_j.$$

Then $\frac{1}{N+1} \sum_{n=0}^{N} \mathbb{1}_{\{j\}}(X_n)$ is the proportion of time that $(X_0, \ldots, X_N)$ spends at $j$.

**Proof of Proposition 9.1** Suppose $f \geq 0$.

- We consider excursions. The $k^{\text{th}}$ excursion is $(X_{\tau_i(k-1)+1}, \ldots, X_{\tau_i(k)})$. The excursions are independent identically distributed random variables. Set

$$\eta_0 := \sum_{n=0}^{\tau_i(1)} f(X_n),$$

$$\eta_k := \sum_{n=\tau_i(k)+1}^{\tau_i(k+1)} f(X_n), \qquad k \geq 1.$$

Since these sums correspond to different excursions, $\eta_0, \eta_1, \eta_2, \ldots$ are independent and identically distributed.

- By the strong law of large numbers for independent identically distributed random variables $\eta_k$,

$$\frac{1}{m}\sum_{k=1}^{m} \eta_k \to E_i \eta_1$$

almost surely. Note that

$$\sum_{k=1}^{m} \eta_k = \sum_{n=\tau_i(1)+1}^{\tau_i(B(N))} f(X_n),$$

where $B(N) = \max\{k \geq 0 : \tau_i(k) \leq N\}$ is the number of completed excursions. Then

$$\frac{1}{N+1}\sum_{n=0}^{\tau_i(B(N))} f(X_n) \leq \frac{1}{N+1}\sum_{n=0}^{N} f(X_n) \leq \frac{1}{N+1}\sum_{n=0}^{\tau_i(B(N)+1)} f(X_n). \qquad (*)$$

The left side of $(*)$ is $\frac{1}{N+1}\eta_0 + \frac{1}{N+1}\sum_{k=1}^{B(N-1)} \eta_k$. The right side of $(*)$ is $\frac{1}{N+1}\eta_0 + \frac{1}{N+1}\sum_{k=1}^{B(N)} \eta_k$. We have

$$\frac{\eta_0}{N+1} \stackrel{N\to\infty}{\longrightarrow} 0 \qquad \text{almost surely.} \qquad (\dagger)$$

We show that

$$\frac{B(N)}{N+1} \stackrel{N\to\infty}{\longrightarrow} \frac{1}{m_j} = \pi_j. \qquad (**)$$

To prove $(**)$, let $\alpha_0$, $\alpha_1 = \tau_i(1)$, and $\alpha_n = \tau_i(n) - \tau_i(n-1)$. These are independent identically distributed random variables representing random variables with mean $m_i$. Thus

$$\frac{\tau_i(k)}{k} \stackrel{k\to\infty}{\longrightarrow} m_i \qquad \text{almost surely} \qquad (\blacktriangle)$$

and

$$\tau_i(B(N)) \leq N < \tau_i(B(N)+1). \qquad (\blacktriangle\blacktriangle)$$

We have

$$\frac{\tau_i(B(N))}{B(N)} \leq \frac{N}{B(N)} < \frac{\tau_i(B(N)+1)}{B(N)} \equiv \frac{\tau_i(B(N)+1)}{B(N)+1} \cdot \frac{B(N)+1}{B(N)}.$$

By (▲), the left side goes to $m_i$, and the right side also goes to $m_i$. Thus

$$\frac{B(N)}{N} \to \frac{1}{m_i}$$

as $N \to \infty$ almost surely. We therefore have

$$\frac{1}{N}\sum_{k=1}^{B(N)} \eta_k = \frac{B(N)}{N} \cdot \frac{1}{B(N)} \sum_{k=1}^{B(N)} \eta_k \overset{N\to\infty}{\longrightarrow} \frac{1}{m_i} E_i \eta_1 \qquad \text{almost surely.} \qquad (***)$$

The left and right sides of $(*)$ tend to $\frac{1}{m_i} E_i \eta_1$, and so also

$$\frac{1}{N+1} f(X_n) \overset{N\to\infty}{\longrightarrow} \frac{1}{m_i} E_i \eta_1$$

almost surely. We now show

$$\sum_j f(j)\pi_j = \frac{1}{m_i} E_i \eta_1. \qquad (****)$$

We have

$$\sum_j f(j)\pi_j = \sum_j f(j) \frac{\gamma_j^{(i)}}{m_i},$$

where by construction

$$\gamma_j^{(i)} = E_i \left( \sum_{m=1}^{\infty} \mathbb{1}_{\{X_n = j\} \cap \{\tau_i(1) \geq n\}} \right)$$

is the average number of visits to $j$ in one excursion determined by $i$. We therefore have

$$\begin{aligned}
\sum_j f(j)\pi_j &= E_i \left( \sum_{j\in S} \sum_{n=1}^{\tau_i(1)} f(j) \mathbb{1}_{\{X_n = j\}} \right) \frac{1}{m_i} \\
&= E_i \sum_{n=1}^{\tau_i(1)} \left( \sum_{j \in S} f(j) \mathbb{1}_{\{X_n=j\}} \right) \frac{1}{m_i} \\
&= E_i \left( \sum_{n=1}^{\tau_i(1)} f(X_n) \right) \frac{1}{m_i} \\
&= (E_i \eta_1) \frac{1}{m_i},
\end{aligned}$$

as desired. ■

# Chapter 10

# Time Reversal

Let $(X_n)$ be an irreducible positive recurrent Markov chain with matrix $P$ and stationary distribution $\pi$. Take $X_0 \sim \pi$ so that the system is in steady state. We consider the **reversed chain** $\{Y_n := X_{-n}, n = 0, \pm 1, \pm 2, \ldots\}$. The transition probabilities for $Y_n$ are

$$\begin{aligned} q_{ij} &= \Pr(Y_{n+1} = X_{-n-1} = j \mid Y_n = X_{-n} = i) \\ &= \Pr(X_m = j \mid X_{m+1} = j) \\ &= \frac{\Pr(X_{m+1} = i \mid X_m = j) \Pr(X_m = j)}{\Pr(X_{m+1} = i)} \\ &= p_{ji} \frac{\pi_j}{\pi_i}. \end{aligned}$$

**Definition.** $X_n$ is reversible if $q_{ij} = p_{ij}$.

**Theorem 10.1** $X$ is time reversible iff $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$. These are called the balance equations.

**Theorem 10.2** For an irreducible Markov chain, if there exists $\pi$ such that $0 \leq \pi_i \leq 1$, $\sum_i \pi_i = 1$, $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$, then the chain is time reversible (in equilibrium) and positive recurrent with stationary distribution $\pi$.

**Proof.** We have $\pi_i p_{ij} = \pi_j p_{ji}$, so $\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j$. We showed $\pi$ satisfies $\pi = \pi P$ iff $X$ is positive recurrent. ■

At this point, we turned to Markov chain Monte Carlo methods, but these don't seem to fit at this point; therefore I am relegating them to Appendix A.

# Chapter 11

# Continuous Time Markov Chains

**Definition.** Let $\{X_t, t \geq 0\}$ be a family of random variables taking values in some *countable* state space $S$. Then $\{X_t\}$ is a Markov chain if it satisfies the **Markov property**: For all $n \geq 1$, for all $t_1 < t_2 < \cdots < t_n$, for all $j, i_1, \ldots, i_{n-1} \in S$,

$$\Pr(X_{t_n} = j \mid X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}) = \Pr(X_{t_n} = j \mid X_{t_{n-1}} = i_{n-1}).$$

Transition probabilities are $P_{ij}(s,t) := \Pr(X_t = j \mid X_s = i)$, $s < t$, for all $i, j$.

We make the assumption of time-homogeneity here: $P_{ij}(s,t) = P_{ij}(s-t, 0) =: P_{ij}(s-t)$.

We set $P_t := (P_ij(t))_{i,j \in S}$, a $|S| \times |S|$ matrix, for all $t \geq 0$.

**Theorem 11.1** The family $\{P_t, t \geq 0\}$ is a **stochastic semigroup**, i.e.

1. $P_0 = I$, the identity matrix (i.e. $P_{ij}(0) = \delta_{ij}$).
2. For each $t \geq 0$, $P_t$ is **stochastic** (i.e. $P_{ij}(t) \geq 0$; the rows sum to 1).
3. $P_{t+s} = P_t P_s$ for all $s, t \geq 0$ (Chapman-Kolmogorov).

**Proof.**

2.
$$\sum_{j \in S} P_{ij}(t) = \sum_{j \in S} \Pr(X_t = j \mid X_0 = i) = \Pr(X_t \in S \mid X_0 = i).$$

3.

$$\begin{aligned} P_{ij}(t+s) &= \Pr(X_{t+s} = j \mid X_0 = i) \\ &= \sum_{k \in S} \Pr(X_{t+s} = j \mid X_0 = i, X_s = k) \\ &= \sum_{k \in S} P_{kj}(t) P_{ij}(s), \end{aligned}$$

as desired. ■

**Definition.** The semigroup $(P_t)_{t \geq 0}$ is called **standard** if $P_t \xrightarrow{t \searrow 0} I \equiv P_0$. This assures continuity by Chapman-Kolmogorov.

We consider only standard $P_t$.

# Chapter 12

# Construction of Continuous Time Markov Chains

## 12.1 Ingredients

1. Take $(X_n)$, a discrete time Markov chain with parameters $(\{a_k\}, Q)$, state space $S$, and $Q = (Q_{ij})$. Assume $Q_{ii} = 0$ for all $i \in S$.

2. $\{E_n, n \geq 0\}$ independent identically distributed random variables distributed by $\exp(i)$, independent of $(X_n)$.

3. $\{\lambda_k, k \in S\}$, which is a function on $S$, with $\lambda(k) > 0$ for all $k \in S$.

It follows that continuous time Markov chains are "usually constant" except for discrete jumps. We call $T_n$ jump times; these are times when transitions $i \to j$ occur. $T_1 - T_0 \sim \exp(\lambda_i) = \exp(\lambda(X_0))$, $T_2 - T_1 \sim \exp(\lambda_j) = \exp(\lambda(X_1))$, and so forth.

## 12.2 Construction

- $T_0 = 0$. $W(0) = E_0/\lambda(X_0)$. Then

$$\Pr(W(0) > x \mid X_0 = i) = \Pr(E_0 > x\lambda(i) \mid X_0 = i) = e^{-\lambda(i)x}$$

  so that $W(0) \sim \exp(\lambda(X_0)) \mid_{X_0=i}$.

- $T_1 = T_0 + W(0)$; $X(t) = X_0$ for $T_0 \le t < T_1$.
- Let $W(T_1) := E_1(\lambda(X_1)) \sim \exp(\lambda(X_1))$. Then $T_2 = T_1 + W(T_1)$, and $X_t = X_1$ for $T_1 \le t < T_2$.

Now suppose that $\{W(T_m), m \le n-1\}$ and $\{T_m, m \le n\}$ and $\{X(t), 0 \le t \le T_n\}$ are already defined. Set $T_\infty := \lim_{n\to\infty} T_n$. On $[0, T_\infty)$, $X_t := \sum_{n=0}^{\infty} X_n \mathbb{1}_{[T_n, T_{n+1})}(t)$, $t < T_\infty$.

## 12.3 Properties

**Lemma 12.1** $\{T_m - T_{m-1} \equiv W(T_{m-1}), m \ge 1\}$ are independent and exponential, given $(X_n)$, i.e.

$$\Pr(T_m - T_{m-1} > u_m, 1 \le m \le n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \prod_{m=1}^{n} e^{-\lambda(i_{m-1})u_m}.$$

**Lemma 12.2**

$$\begin{aligned}\Pr(X_{n+1} = j, T_{n+1} - T_n > u \mid X_0, \ldots, X_n, T_1, \ldots, T_n) &= \Pr(X_{n+1} = j, T_{n+1} - T_n > u \mid X_n) \\ &= Q_{X_n,j} e^{-\lambda(X_n)u}.\end{aligned}$$

**Proof.**

$$\begin{aligned}&\Pr(X_{n+1} = j, T_{n+1} - T_n > u \mid X_0 = i_0, \ldots, X_n = i_n, T_1, \ldots, T_n) \\ &= \Pr(X_{n+1} = j, E_n > u\lambda(i_n) \mid X_0 = i_0, \ldots, X_n = i_n, \\ &\qquad T_m - T_{m-1} = W(T_{m-1}) = \frac{E_{m-1}}{\lambda(i_{m-1})}, m = 1, \ldots, n, E_1, \ldots, E_{i-1}) \\ &= \Pr(X_{n+1} = j, E_n > u\lambda(i_n) \mid X_n = i_n) \\ &= Q_{ij} e^{-u\lambda(i_n)},\end{aligned}$$

as desired. $\blacksquare$

## 12.4 Summary

A Markov chain $(X_t, t \geq 0)$ is determined by

1. a **jump chain** $(X_n)$, which is a Markov chain with matrix $Q$, where $Q_{ii} = 0$.
2. a function $\{\lambda(k), k \in S\}$, the rates of holding times $T_m = T_{m-1} = W(T_m)$. Then
   - $\{T_0, T_1, \ldots\}$ are the jump times, with $T_m - T_{m-1} \sim \exp(\lambda(X_{m-1}))$.
   - $X_n = X(T_n^+)$ is the **embedded** discrete time Markov chain with probabilities $Q_{ij}$; this governs the jumps.

# Chapter 13

# Birth-Death Processes

## 13.1 Description

Let $X_t = i$ (i.e. at time $t$, there are $i$ individuals alive). Let $B(i) \sim \exp(\lambda_i)$ be the time until the next birth and $D(i) \sim \exp(\mu_i)$ be the time until the next death ($\mu_0 = 0$). Assume that only one events happens at a time. Then the population increases by 1, i.e. $X_t$ makes the transition $i \to i+1$ if $B(t) < D(t)$, with probability $\Pr(B(i) < D(i)) = \frac{\lambda_i}{\lambda_i+\mu_i}$. $X_t$ makes the transition $i \to i-1$ if $D(i) < B(i)$ with probability $\Pr(D(i) < B(i)) = \frac{\mu_i}{\lambda_i+\mu_i}$. The holding time is $\min(B(i), D(i)) \sim \exp(\lambda_i + \mu_i)$.

Let $\{X_t, t \geq 0\}$ be a Markov chain.

- Jumps. We have an embedded discrete time chain $(X_n)$, where $n$ corresponds to the "$n^{\text{th}}$ jump." We have $Q = (Q_{ij})$, where $Q_{ij}$ is the probability of jumping from $i$ to $j$ at the time of a jump; $Q_{ii} = 0$.
- Rates $\{\lambda(k)\}$ describe how long, on average, the chain stays at $k$. We have jump times $T_0 = 0, T_1, T_2, \ldots$. Then $T_m - T_{m-1} = \frac{E_m}{\lambda(X_{m-1})} \sim \exp(\lambda(X_n))$. $X_t$ is defined on $[0, T_\infty)$, where $T_\infty = \lim_{n\to\infty} T_n$.

We call $\lambda_i$ the birth rate and $\mu_i$ the death rate. Note that if $p_i = \frac{\lambda_i}{\lambda_i+\mu_i}$ and $q_i = \frac{\mu_i}{\lambda_i+\mu_i}$, then $\lambda_i = p_i\lambda(i)$ and $\mu_i = q_i\lambda(i)$.

## 13.2 Special Cases

### 13.2.1 Linear Birth

Suppose $\lambda_i = i\lambda$ and $\mu_i = i\mu$ for $i \geq 0$. We calculate $\lambda(i) = \lambda_i + \mu_i = i(\lambda + \mu)$. Then $p_i = \frac{\lambda_i}{\lambda(i)} = \frac{\lambda}{\lambda+\mu}$ and $q_i = \frac{\mu}{\lambda+\mu}$ are constants.

Let $X_t = i$, i.e. $i$ individuals are alive. Each individual, independently, lives a lifetime $L_k \sim \exp(\lambda + \mu)$ for $k = 1, 2, \ldots, i$. At its deathbed, the individual

- dies childless with probability $\frac{\mu}{\lambda+\mu} = q$, or
- has two children with probability $\frac{\lambda}{\lambda+\mu} = p$.

From this description, we find that the holding time is $\min(L_1, \ldots, L_i) \sim \exp(i(\lambda+\mu))$. Then $\lambda_i = p_i\lambda(i) = \frac{\lambda}{\lambda+\mu} i(\lambda + \mu) = i\lambda$. Similarly, $\mu_i = i\mu$.

### 13.2.2 The Pure-Birth Process

Let $\mu_i = 0$ so that $Q_{i,i+1} = 1 - p_i$ for all $i$. Then $(X_n)$ determines when the chain increases by 1. Then $\lambda(i) = \lambda_i > 0$. An example is the Yule process, or linear pure birth process, in which $\lambda_i = i\lambda$, where $\lambda > 0$.

Let $X_t = i$ so that $i$ individuals are alive. Each independently lives $L_k \sim \exp(\lambda)$, for $k = 1, 2, \ldots, i$. At the end of his or her life, the individual is replaced by 2 (or, equivalently, gives birth to 1 and continues living). The holding time is $B(i) = \min(L_1, \ldots, L_i) \sim \exp(i\lambda)$.

### 13.2.3 The Poisson Process

This is a pure birth process, with $\lambda(i) = \lambda i = \lambda$. Let $N_0 = 0$ and $T_0, T_1, T_2, \ldots$ be arrival times, with $T_n - T_{n-1} \sim \exp(\lambda)$. Then $N_t$ is the number of arrivals or occurrences by time $t$. This is often called the counting process or import process.

Later we shall show

- $N_t \sim P_i(\lambda t)$, $t \geq 0$. $\lambda$ is the intensity or rate.
- The increments $N_t - N_s$ and $N_s - N_0$ are independent and stationary $(0 < s < t)$.

### 13.2.4 Thinned Poisson Process

Customers arrive according to a Poisson process given by $N_t$ and intensity $\lambda$. Each customer stays with probability $p$ or gives up (balks). The resulting process $N'_t$ is Poisson with intensity $\lambda' = \lambda p$. (The birth rate for $N' = \lambda'_i = p'\lambda(i) = p\lambda$.)

# Chapter 14

# Stability and Explosions

**Questions.**

- When is $T_\infty = \infty$, i.e., when is $(X_t)$ defined for all $t \geq 0$?
- What is the importance of the condition $0 < \lambda(k) < \infty$?

**Definition.** A state $i$ for which $0 < \lambda(i) < \infty$ is called **stable**. A state $i$ for which $\lambda(i) = \infty$ is called **instantaneous**.

**Note.**

- If $\lambda(i) = \infty$, then the mean holding time at state $i$ is $\frac{1}{\lambda(i)} = 0$.
- $\lambda(i) = 0$ means that the mean holding time at state $i$ is $\frac{1}{0} = \infty$ so that $i$ is absorbing.

In this class we make the assumption that all states are stable.

**Definition.** If for all $i \in S$, $P_i(T_\infty = \infty) = 1$, then $(X_t)$ is **regular**.

**Note.** If $T_\infty < \infty$, then on a finite interval $[0, T_\infty]$, $(X_t)$ has infinitely many jumps at time $T_1, T_2, \ldots$. We say that an explosion occurs.

**Proposition 14.1** For all $i \in S$,

$$P_i(T_\infty < \infty) = P_i\left(\sum_n \frac{1}{\lambda(X_n)} < \infty\right).$$

Thus $(X_t)$ is regular iff $P_i\left(\sum_n \frac{1}{\lambda(X_n)} = \infty\right) = 1$ for all $i \in S$.

For example, the Poisson process is regular because $\sum \frac{1}{\lambda} = \infty$. The Yule process is regular because $\sum \frac{1}{i\lambda} = \infty$. The pure birth process is regular iff $\sum \frac{1}{\lambda_i} = \infty$. For example, if $\lambda_i = i^2\lambda$, then we have an explosion.

**Proof.** We have

$$\begin{aligned} T_\infty &= \lim_{N\to\infty} T_n \\ &= \lim_{N\to\infty} \sum_{n=0}^{N-1}(T_{n+1} - T_n) \\ &= \sum_{n=0}^{\infty}(T_{n+1} - T_n) \\ &= \sum_{n=0}^{\infty} \frac{E_n}{\lambda(X_n)}. \end{aligned}$$

Hence by the problems we have

$$P_i(T_\infty < \infty \mid (X_n)) = P_i\left(\sum_{n=0}^{\infty} \frac{1}{\lambda(X_n)} < \infty \mid (X_n)\right),$$

as desired. ∎

**Corollary 14.2** The following are sufficient conditions for regularity:

1. If $\max_i \lambda(i) < \infty$, then $(X_t)$ is regular. This is because $\sum \frac{1}{\lambda(X_n)} \geq \sum \frac{1}{\max_i \lambda(i)} = \infty$.

2. If $S$ is finite, then the chain is regular. This follows from (1) since $\max_{i \in S} \lambda(i) < \infty$.

3. If $X_0 = i$ and $i$ is recurrent for the jump chain $(X_n)$, then $(X_t)$ is regular. This is because if $i$ is recurrent, then $X_n = i$ infinitely often, e.g. on steps $N_1, N_2, \ldots$ Then $\sum \frac{1}{\lambda(X_n)} \geq \sum_j \frac{1}{\lambda(X_{n_j})} = \infty$.

# Chapter 15

# The Backward and Forward Equations and the Generator Matrix

We make the following assumptions:

- $(X_t)$ is regular (i.e. $T_\infty = \infty$).
- All states are stable ($0 < \lambda(k) < \infty$ for all $k$).

The goal is to find a relationship between $(Q, \lambda)$ and $(P_{ij}(t))_{i,j \in S}$.

**Proposition 15.1** For all $t > 0$ and for all $i, j \in S$, we have

$$P_{ij}(t) = \delta_{ij} e^{-\lambda_i(t)} + \int_0^t \lambda(i) e^{-\lambda(i)s} \sum_{k \neq i} Q_{ik} P_{kj}(t-s) \, ds.$$

**Proof.** The technique is to decompose according to the value of the firm jump $T_1$ (conditioning on $T_1$). We have

$$\begin{aligned} P_{ij}(t) &= P_i(X_t = j) \\ &= P_i(X_t = j, t < T_1) + P_i(X_t = j, t \geq T_1). \end{aligned}$$

We have $P_i(T_1 > t) = e^{-\lambda(i)t}$. The first term is $\delta_{ij} e^{-\lambda_i(t)}$. The second term is

$$\sum_{k \neq i} P_i(T_1 \leq t, X_{T_1} = k, X_t = j) = \sum_{k \neq i} \int_0^t \lambda(i) e^{-\lambda(i)s} Q_{ik} P_{kj}(t-s)\, ds,$$

and the result follows. ■

**Corollary 15.2** $P_t \searrow I$ as $t \searrow 0$ (which means that $P_i$ is standard).

**Corollary 15.3** For all $i, j \in S$, $P_{ij}(t)$ is differentiable with continuous derivative. At $t = 0$,

$$\frac{d}{dt} P_{ij}(0) \equiv A_{ij} = \begin{cases} -\lambda(i) & i = j, \\ \lambda(i) Q_{ij} & i \neq j. \end{cases}$$

We call the $|S| \times |S|$ matrix $P'(0) = A = (A_{ij})$ the generator matrix. We then have the backward difference equation $P'(t) = AP(t)$, i.e.

$$P'_{ij}(t) = \sum_k A_{ik} P_{kj}(t).$$

**Proof.** Differentiate the result of Proposition 15.1, or see page 387 of Resnick. ■

We can interpret the generator matrix as flow rates of probability.

1. $$A_{ii} = -\lambda(i) = P'_{ii}(0) = \lim_{h \searrow 0} \frac{P_{ii}(h) - P_{ii}(0)}{h} - \lim_{h \searrow 0} \frac{P_{ii}(h) - 1}{h},$$

   or

   $$P_{ii}(h) = 1 - \lambda(i)h + o(h) = 1 + A_{ii}h + o(h).$$

2. If $i \neq j$, then

   $$A_{ij} = \lambda(i) Q_{ij} = \lim_{h \searrow 0} \frac{P_{ij}(h) - P_{ij}(0)}{h} = \lim_{h \searrow 0} \frac{P_{ij}(h)}{h},$$

   or

   $$P_{ij}(h) = \lambda(i) Q_{ij} h + o(h) = A_{ij} h + o(h).$$

3. The probability of two of more transitions in the interval of length $h$ is $o(h)$:

$$\begin{aligned}
\Pr(T_2 < h \mid X_0 = i, X_{T_1} = j) &= \Pr\left(\frac{E_1}{\lambda(i)} + \frac{E_2}{\lambda(j)} < h\right) \\
&\leq \Pr\left(\frac{E_1}{\lambda(i)} < h \text{ and } \frac{E_2}{\lambda(j)} < h\right) \\
&= (1 - e^{-\lambda(i)h})(1 - e^{-\lambda(j)h}) \\
&= (\lambda(i)h + o(h))(\lambda(j)h + o(h)) \\
&= o(h).
\end{aligned}$$

## 15.1 Dynamics of the Chain

Fix $X_t = i$; consider a small time interval $(t, t+h)$. By (3), the probability of more than one transition in this interval is $o(h)$. Thus in $(t, t+h)$, the probability that nothing happens (the chain remains in state $i$) is

$$P_{ii}(h) + o(h) = 1 - \lambda(i)h + o(h) = 1 + A_{ii}(h) + o(h)$$

(the diagonal of the generator) by (1). In this case, the holding time $\frac{E_1}{\lambda(i)} > h$ with probability $e^{-\lambda(i)h} = 1 - \lambda(i)h + o(h)$. So $-A_{ii}h = \lambda(i)h \approx 1 - P_{ii}(h)$ is the probability that the chain leaves state $i$, so $-A_{ii} = \lambda(i)$ is the flow rate for the probability to leave state $i$. The probability for the chain to move $i \to j$ in $(t, t+h)$ is

$$P_{ij}(t) + o(h) = \lambda(i)Q_{ij}h + o(h) = A_{ij}h + o(h)$$

by (2). Note that the move $i \to j$ happens in $(t, t+h)$ if

- the holding time $\frac{E_1}{\lambda(i)} < h$ with probability $1 - e^{-\lambda(i)h} = \lambda(i)h + o(h)$,
- the chain jumps $i \to j$ with probability $Q_{ij}$.

Suppose $X \sim \exp(\lambda)$. Then $E(X) = \frac{1}{\lambda}$. In this case $f_X(x) = \lambda e^{-\lambda x}$ (for $x > 0$) is the probability density function. We have $P(X \leq x) = 1 - e^{-\lambda x}$ and $P(X > x) = e^{-\lambda x}$ for $x > 0$. The Markov chain jumps from $i$ to $j$ with probability $P_{ij}(h) + o(h)$ (where the

$o(h)$ accounts for the possibility of $i \to j \to k \to j$) and remains in $i$ with probability $P_{ii}(h) + o(h)$.

We have the relationship $P_{ij}(h) = \lambda(i)Q_{ij}(h) + o(h)$ and $P_{ii}(h) = 1 - \lambda(i)h + o(h)$.

**Example.** (Thinned Poisson.) Customers arrive according to a Poisson process with intensity $\lambda$ ($N_t$). Each customer stays with probability $p$ or balks. The resulting process is $N'_t$. Show that $N'_t \sim P$ and $\lambda' = \lambda p$.

**Solution.** Suppose $N'_t = i$. Consider the interval $(t, t+h)$. Then

$$\begin{aligned}
P_{i,i+1}(h) &= \Pr(N'_{t+h} = i+1 \mid N'_t = i) \\
&= \Pr(\text{one new arrival according to input process } N_t \text{ and stay}) \\
&= \Pr(\text{interarrival time for } N_t < h) \cdot p + o(h) \\
&= (1 - e^{-\lambda h})p + o(h) \\
&= \lambda h p + o(h).
\end{aligned}$$

Then $P_{ii}(h) = \Pr(\text{no arrivals in } N_t \text{ or arrival in } N_t \text{ leaves})$. Also, $P_{ii}(h) = 1 - P_{i,i+1}(h) + o(h) = 1 - \lambda p h + o(h)$. Then $\lambda(i)Q_{i,i+1} = \lambda p$ and $\lambda(i) = \lambda p$, so $Q_{ij} = \delta_{i+1,j}$. These are parameters of a Poisson process with intensity $\lambda' = \lambda p$.

There is a second method based on waiting time $W_i = \min(L_1, \dots, L_i, E) \sim \exp(iq + \lambda)$.

Let $A = (A_{ij})$.

**Note.** $\sum_j A_{ij} = 0$. (The rows of the generator matrix sum to zero.)

**Proof.** $\sum_{j \neq i} Q_{ij} = 1$ iff $\sum_{j \neq i} \frac{A_{ij}}{\lambda(i)} = 1$ iff $\sum_{j \neq i} A_{ij} = \lambda(i) = -A_{ii}$. ■

We now work out the generator for the birth-death process: $A_{ii} = -\lambda(i) = -(\lambda_i + \mu_i)$, and $A_{ij} = \lambda(i)Q_{ij}$, which is $\lambda_i$ for $j = i+1$ and $\mu_i$ for $j = i-1$ and 0 otherwise.

We now have the forward equation: $P'(t) = P(t)A$.

**Proof.** We use Chapman-Kolmogorov: $P(t+s) = P(t)P(s)$. Since $P'(t+s) = P(t)P'(s)$, we have $P'(t) = P(t)A$. ■

**Remark.** $P(t) = e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$.

The forward equation means that $P'_{ij}(t) = \sum_k P_{ik}(t)A_{kj} = \sum_{k \neq j} P_{ik}(t)\lambda(k)Q_{kj} - \lambda(j)P_{ij}(t)$.

**Example.** The generator matrix for the birth-death process is

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

**Example.** Consider a Poisson process with $\mu_i = 0$ for all $i$, $\lambda_i = \lambda$, and $N_0 = 0$. The goal is to use the forward equations to find $P_{0j}(t) = \Pr(N_t = j \mid N_0 = 0)$. We have $P'_{00}(t) = -\lambda P_{00}$, and $P'_{0j}(t) = \lambda P_{0,j-1} - \lambda P_{0j}(t)$ and $P_{0j}(t) = \delta_{0j}$ for $j \geq 1$. (These are called the differential difference equations.) We use the generating function $P(t,s) = \sum_{j=0}^{\infty} s^j P_{0j}(t) = E(s^{N_t} \mid N_0 = 0)$. Multiply the $j^{\text{th}}$ equation by $s^j$ and sum:

$$\frac{\partial P(t,s)}{\partial t} = \lambda s P - \lambda P = \lambda(s-1)P$$
$$P(s,0) = s^0 = 1.$$

Thus $P(s,t) = e^{\lambda(s-1)t}$. This is the generating function of a Poisson $(\lambda t)$ random variable. Thus $N_t$ is Poisson $(\lambda t)$ and $P_{0j}(t) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}$.

# Chapter 16

# Stationary and Limiting Distributions

## 16.1 Classification of States

For all $i, j \in S$, either $P_{ij}(t) > 0$ for all $t > 0$ or $P_{ij}(t) = 0$ for all $t > 0$.

**Definition.** The chain is irreducible if for all $i, j \in S$, $P_{ij}(t) > 0$ for some $t$ (and thus for all $t > 0$). An irreducible continuous time Markov chain is transient or recurrent iff the the embedded discrete time Markov chain is transient or recurrent.

## 16.2 Stationary Distributions

**Definition.**

- A measure $\eta = \{\eta_j : j \in S\}$ on $S$ is invariant if for all $t > 0$, $\eta P(t) = \eta$.
- If $\eta$ is an invariant probability measure, then it is called a **stationary distribution** $\pi$.

If an initial distribution of the chain is $\pi$, then $X_t$ is strictly stationary.

**Theorem 16.1** (Ergodic Theorem) Let $(X_t)$ be irreducible.

1. If there exists a stationary distribution $\pi$, then it is unique, and $P_{ij}(t) \to \pi_j$ as $t \to \infty$ for all $i, j$.

2. If a stationary distribution does not exist, then $P_{ij}(t) \to 0$ as $t \to \infty$ for all $i, j$.

**Sketch of Proof.** Fix step $h > 0$, and let $Y_n := X(nh)$ be an irreducible discrete time aperiodic Markov chain (since $P_{ij}(t) > 0$ for all $t$). This is a skeleton of $X$.

- From Theorem 8.2, we know that if $(Y_n)$ is null recurrent or transient, then $P_{ij}(nh) \to 0$ as $n \to \infty$; if $(Y_n)$ is positive recurrent, then there exists a unique stationary distribution $(\pi^h)$, and $P_{ij}(nh) \to \pi_j^h$ as $n \to \infty$ for all $i$.
- Notice that for all *rational* $h_1$ and $h_2$, $\pi^{h_1} = \pi^{h_2}$ because $\{nh_1, n \geq 0\}$ and $\{nh_2, n \geq 0\}$ have infinitely many common points. Thus the limit of $\{P_{ij}(t)\}$ exists along all sequences $\{nh, n \geq 0\}$.
- (See Resnick.) Use continuity of $P_{ij}(t)$ to fill the gaps. ■

**Claim 16.2** $\pi = \pi P_t$ iff $\pi A = 0$. (These are called the balance equations.) That is, $\sum_i \pi_i A_{ij} = \sum_{i \neq j} \pi_i A_{ij} + \pi_j A_{jj} = 0$.

**Sketch of Proof.** $\pi A = 0$ iff $\pi A^n = 0$ for all $n \geq 1$ iff $\sum_{n=1}^{\infty} \frac{t^n}{n!} \pi A^n = 0$ for all $t > 0$ iff $\pi \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n = \pi$ for all $t > 0$ iff $\pi e^{tA} = \pi$ iff $\pi P_t = \pi$. ■

**Example.** Birth-death process. The balance equations are $\pi A = 0$ or $\pi_j A_{jj} + \sum_{i \neq j} \pi_i A_{ij} = 0$. When $j = 0$, we have $\pi_0(-\lambda_0) + \pi_1 \mu_1 = 0$, so $\pi_1 = \pi_0 \frac{\lambda_0}{\mu_0}$. When $j = 1$, we have $\pi_1(-(\lambda_1 + \mu_1)) + \pi_0 \lambda_0 + \pi_2 \mu_2 = 0$, so $\pi_2 = \pi_0 \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2}$, and so forth. Since $\pi$ is a stationary distribution iff $\sum_{n=0}^{\infty} \pi_n = 1$ and $\pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \pi_0$, the chain has a stationary distribution if $\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} < \infty$.

# Chapter 17

# Problems

1. Let $X$ have the Poisson distribution with parameter $Y$, where $Y$ has the Poisson distribution with parameter $\mu$. Show that $P_{X+Y}(x) = \exp\{\mu(xe^{x-1} - 1)\}$.

2. A hen lays $N$ eggs where $N$ is Poisson with parameter $\lambda$. The weight of the $n^{\text{th}}$ egg is $W_n$, where $W_1, W_2, \ldots$ are independent identically distributed random variables with probability generating function $P(s)$, independent of $N$. Show that the generating function $P_W$ of the total weight $W = \sum_{i=1}^{N} W_i$ is given by $P_W(s) = \exp\{-\lambda + \lambda P(s)\}$. $W$ is said to have a *compound Poisson distribution*.

3. Let $\{X_n, n \geq 1\}$ be independent identically distributed nonnegative integer valued random variables independent of the nonnegative integer valued random variable $N$ and suppose that $X$'s and $N$ have finite second moments. Let $S_n = \sum_{k=0}^{n} X_i$. Use generating functions to check $\operatorname{Var}(S_N) = E(N)\operatorname{Var}(X_1) + (EX_1)^2 \operatorname{Var}(N)$.

4. Let $X$ have probability mass function $\{p_k\}$ satisfying $\sum_{k=0}^{\infty} p_k = 1$. Let $P(s)$ be the generating function of $X$ and define $Q(s)$ to be the generating function of the sequence $\{q_k\} = \{P(X > k), k = 0, 1, 2, \ldots\}$ (i.e. $Q(s) = \sum_{k=0}^{\infty} s^k q_k$). Show that $Q(s) = (1 - P(s))/(1 - s)$ for $0 \leq s < 1$. Use this formula to show that $EX \equiv \sum_{k=0}^{\infty} P(X > k) = P'(1)$.

5. Consider a branching process (Galton-Watson). Suppose that each family size has geometric distribution, i.e. $p_k = qp^k$, $k \geq 0$, where $p + q = 1$, $p, q > 0$.

   (a) Show (by induction) that

      i. if $q \neq p$, then $P_n(s) = q\frac{p^n - q^n - (p^{n-1} - q^{n-1})ps}{p^{n+1} - q^{n+1} - (p^n - q^n)ps}$;

      ii. if $q = p$, then $P_n(s) = \frac{n - (n-1)s}{n+1-ns}$.

(b) Use (a) to show that $P(Z_n = 0) \to 1$ if $p \leq q$, $\quad P(Z_n = 0) \to q/p$, if $p > q$.

6. Consider a branching process (Galton-Watson). Suppose that each family size has mean $m = EZ_{n,j} = EZ_1$ and variance $\sigma^2 = \mathrm{Var}\, Z_{n,j} = \mathrm{Var}\, Z_1$. Show that

   (a) $\mathrm{Var}\, Z_n = P_n''(1) + m^n - m^{2n}$;

   (b) $P_n''(1) = (\sigma^2 + m^2 - m)m^{n-1} + m^2 P_{n-1}''(1)$;

   (c)
   $$\mathrm{Var}\, Z_n = \begin{cases} n\sigma^2, & \text{if } m = 1; \\ \sigma^2(m^n - 1)m^{n-1}(m-1)^{-1} & \text{if } m \neq 1. \end{cases}$$

7. Suppose that every man in a certain society has exactly three children, which independently have probability one-half of being a boy and one-half of being a girl. Suppose also that the number of *males* in the $n^{th}$ generation forms a branching process.

   (a) Find the probability that the male line of a given man eventually becomes extinct.

   (b) If a given man has two boys and one girl, what is the probability that his male line will continue forever?

8. Let $Y \geq 0$ and let $\phi \geq 0$ be a function that is increasing on $[0, \infty)$. Show that

   (a) $\phi(a)P(Y \geq a) \leq E\phi(Y)$;

   (b) $P(|X| \geq \varepsilon) \leq EX^2/\varepsilon^2$.

9. Show that any sequence of independent random variables taking values in a countable set $S$ is a Markov chain. Under what conditions is this chain homogeneous?

10. Let $X$ and $Y$ be two independent homogeneous Markov chains, each with the same discrete state space $S$. Show that the sequence $Z_n = (X_n, Y_n)$ is a Markov chain with state space $S \times S$ and give the transition probability matrix.

11. Let $X$ be a Markov chain with state space $S$ and suppose that $h : S \to T$ is one-one.

    (a) Show that $Y_n = h(X_n)$ is a Markov chain on $T$.

(b) Give an example of a function $f$ such that $Z_n = f(X_n)$ is not a Markov chain.

12. Show that

    (a)

$$\begin{aligned}&\Pr(X_{n+1} = i_{n+1}, \ldots, X_{n+m} = i_{n+m} \mid X_0 = i_0, \ldots, X_n = i_n)\\&= p_{i_n, i_{n+1}} \cdots p_{i_{n+m-1}, i_{n+m}}.\end{aligned}$$

    (b)

$$\begin{aligned}&\Pr(X_{n+m} = i_{n+m} \mid X_0 = i_0, \ldots, X_n = i_n)\\&= p^{(m)}_{i_n, i_{n+m}} \equiv \Pr(X_{n+m} = i_{n+m} \mid X_n = i_n).\end{aligned}$$

13. Let $\{Z_n\}_n$ be independent identically distributed random variables and $X_0 = f(Z_0)$, $X_{n+1} = g(X_n, Z_{n+1})$ for some functions $f$ and $g$. Then $X_n$ is a Markov chain (this is a way to generate Markov chains!).

14. Assume that once a day (e.g. at noon), the weather is observed as being one of the following: state 1: rain (or snow); state 2: cloudy; state 3: sunny. We postulate that the weather on day $t$ is characterized by a single one of the three states above, and that all of the transition probabilities $p_{ij}$, $i, j = 1, 2, 3$ are positive (and known).

    (a) Given that the weather today is sunny, what is the probability that the weather for the next 7 days will be "sun-sun-rain-rain-sun-cloudy-sun"?

    (b) Given that the weather today is sunny, what is the probability that it stays sunny for exactly $d$ days?

15. (a) Let $\{X_n, n \geq 0\}$ be a two-state Markov chain having state space $S = \{0, 1\}$ with transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

    Show that

$$p^{(n)}_{0,0} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n, & \text{for } \alpha+\beta > 0, \\ 1, & \text{for } \alpha+\beta = 0. \end{cases}$$

(b) Suppose a virus can exist in $N$ different strains and in each generating either stays the same, or with probability $\alpha$ mutates to another strain, which is chosen at random. What is the probability that the strain in the $n^{\text{th}}$ generation is the same as that in the $0^{\text{th}}$?

16. Give an example to show that for a Markov chain to be irreducible, it is sufficient but not necessary that for some $n \geq 1$, $p_{ij}^{(n)} > 0$ **for all** $i, j \in S$.

17. Let $\{X_n, n \geq 0\}$ be a Markov chain having state space $S = \{0, 1\}$. Let $P(X_{n+1} = 1 \mid X_n = 0) = p$, $P(X_{n+1} = 0 \mid X_n = 1) = q$, $P(X_0 = 0) = \mu_0$. Here $0 < p, q, \mu_0 < 1$. Find $P(X_1 = 0 \mid X_0 = 0, X_2 = 0)$ and $P(X_1 \neq X_2)$.

18. Suppose $\{Z_n, n \geq 1\}$ are independent and identically distributed representing outcomes of successive throws of a die. Define $X_n = \max\{Z_1, \ldots, Z_n\}$. Show that $\{X_n, n \geq 1\}$ is a Markov chain and give its transition matrix $P$. Determine which states are transient and which are recurrent.

19. Classify the states of the discrete Markov chain with state space $S = \{1, 2, 3, 4\}$ and transition matrix

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

Calculate $f_{34}^{(n)}$ and deduce that the probability of ultimate absorption in state 4, starting from 3, equals 2/3.

20. Harry's restaurant business fluctuates in successive years between three states: 0 (bankruptcy), 1 (verge of bankruptcy), and 2 (solvency). The transition matrix giving the probabilities of evolving from state to state is

$$P = \begin{pmatrix} 1 & 0 & 0 \\ .5 & .25 & .25 \\ .5 & .25 & .25 \end{pmatrix}.$$

What is the expected number of years until Harry's restaurant goes bankrupt assuming that he starts from the state of solvency?

21. The Media Police have identified six states associated with television watching: 0 (never watch TV), 1 (watch only PBS), 2 (watch TV fairly frequently), 3 (addict), 4 (undergoing behavior modification), 5 (brain dead). Transitions from

state to state can be modelled as a Markov chain with the following transition matrix:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 & 0 & 0 \\ .1 & 0 & .5 & 0 & 0 & .4 \\ 0 & 0 & 0 & .7 & .1 & .2 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

(a) Which states are transient and which are recurrent?

(b) Starting from state 1, what is the probability that state 5 is entered before state 0; i.e. what is the probability that a PBS viewer will wind up brain dead?

22. Consider a Markov chain on $S = \{0, 1, 2\}$ with transition matrix

$$P = \begin{pmatrix} 0 & .5 & .5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

(a) Find $n$-step probabilities $p_{ij}^{(n)}$, $n = 1, 2, \ldots$

(b) Show that all states are recurrent.

(c) Find all probabilities $f_{ii}^{(n)}$, $n \geq 1$, $i = 0, 1, 2$.

(d) Find the mean recurrence times $m_i$, $i = 0, 1, 2$.

(e) Does the chain have ergodic (i.e. positive recurrent and aperiodic) states?

23. Let $\{X_n, n \geq 0\}$ be a two-state Markov chain having state space $S = \{0, 1\}$ with transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Let $X_0 = 0$ and $T = \tau_0(1)$ the first time that the chain returns to state 0. Determine the distribution of $T$ and calculate $E(T)$ (average length of an excursion).

24. Consider a Markov chain having the transition matrix

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

Determine which states are transient and which are non-null (positive) or null recurrent. Put the matrix $P$ in canonical form.

25. Show that every recurrent class is closed.

26. In a finite Markov chain, $j$ is transient iff there exists some state $k$ such that $j \to k$, but $j$ is not accessible from $k$. Explain why. Give an example to show that this is false if the Markov chain has an infinite number of states.

27. Let $\{Z_n, -\infty < n < \infty\}$ be the sequence of independent identically distributed random variables with $P(Z_1 = 0) = P(Z_1 = 1) = 1/2$. Define the stochastic process $\{X_n\}$ with state space $\{0, \dots, 6\}$ by

$$X_n = Z_{n-1} + 2Z_n + 3Z_{n+1}, \qquad -\infty < n < \infty.$$

    (a) Determine $P(X_0 = 1, X_1 = 3, X_2 = 2)$ and $P(X_1 = 3, X_2 = 2)$.

    (b) Is $\{X_n\}$ Markov? Why or why not?

28. Let $X_1, X_2, \dots$ be discrete independent identically distributed random variables with mean $\mu < \infty$ and finite variance. Let $S_0 = 0$, $S_n = X_1 + \cdots + X_n$. Show that if $\mu \neq 0$, then the state $\{0\}$ is transient for $S_n$.

29. Show that for an irreducible positive recurrent Markov chain with stationary distribution $\pi$, the expected number of visits to state $j$ between two successive visits to state $i$ is $\pi_j/\pi_i$.

30. (a) Let function $f$ satisfy the set of linear difference equations

$$f(n) = qf(n-1) + pf(n+1), \qquad n = 1, 2, \dots, \qquad p+q = 1, p \in (0,1), p \neq q. \quad (*)$$

    i. Seek a solution of $(*)$ in the form $f(n) = \alpha^n$. Show that such an $f$ solves $(*)$ for $\alpha = 1$ and $\alpha = q/p$.

    ii. Show that $f(n) = A \cdot 1^n + B \cdot (q/p)^n$ with $A$ and $B$ arbitrary constants solve $(*)$ (i.e. this is the general solution of $(*)$).

    (b) Let $X_n$ be an asymmetric random walk on $\mathbb{Z}$. Find invariant measures.

31. Let $\{X_n\}$ be an aperiodic positive recurrent irreducible Markov chain with stationary distribution $\pi$. Let $\{Y_n\}$ be an independent copy of $X$.

    (a) Show that $\{\xi_n = (X_n, Y_n), n \geq 0\}$ is Markov and find its transition probabilities.

(b) Give the stationary distribution of $\{\xi_n\}$.

32. Consider a Markov chain having transition matrix

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

Find the stationary distribution concentrated on each of the irreducible closed sets.

33. A Markov chain has state space $\{0, 1, 2, \ldots\}$ and transition probabilities $p_{i,i+1} = \lambda/(i + \nu + 1)$, $p_{i,0} = 1 - p_{i,i+1}$, where $\lambda > 0$ and $\nu \geq 0$ are constants. State any other necessary restrictions on the values of $\lambda$ and $\nu$. Show that the chain is irreducible, aperiodic, and positive recurrent. Find explicit forms for the stationary distribution in the cases $\nu = 0$ and $\nu = 1$.

34. Give an example of an irreducible Markov chain such that it has a stationary distribution $\pi$, but $n$-step transition probabilities $p_{ij}^{(n)}$ do not converge to $\pi_j$.

35. Let $\{X_n\}$ be an aperiodic positive recurrent irreducible Markov chain with stationary distribution $\pi$.

(a) Show that $\{\xi_n = (X_n, X_{n+1}), n \geq 0\}$ is Markov and find its transition probabilities.

(b) Give the stationary distribution of $\{\xi_n\}$.

(c) Suppose each time there is a transition from state $i$ to state $j$ there is a reward of $g(i, j)$ which is received. Assume that $g$ is bounded. What is the long-term reward rate $\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n} E_i g(X_m, X_{m+1})$? Why does this limit exist? Does it depend on the initial state $i$?

36. Suppose that the probability of a dry day (state 0) following a rainy day (state 1) is 1/3 and that the probability of a rainy day following a dry day is 1/2.

(a) Given that December $1^{\text{st}}$ is a dry day, find the probability that December $3^{\text{rd}}$ is a dry day.

(b) Find the long run proportion of rainy days.

(c) Find the expected number of days between two rainy days.

37. Give an example of a Markov chain with

    (a) at least two stationary distributions;

    (b) a unique stationary distribution;

    (c) no stationary distribution.

38. Four boys (denoted by 1, 2, 3, and 4) arranged in a circle play a game of throwing a ball to one another. At each stage the child having the ball is likely to throw it to a boy standing next to him clockwise with probability $p$, $0 < p < 1$, and to his neighbor counterclockwise with probability $q = 1 - p$. Suppose that $X_0$ denotes the child who had the ball initially and $X_n$ ($n \geq 1$) denotes the boy who had the ball after $n$ throws.

    (a) Write the state space and the transition probability matrix for the Markov chain $X_n$.

    (b) Find all the classes and determine which ones are recurrent and which ones are transient.

    (c) Does the chain have a stationary distribution? Find the expected number of steps it takes the boy who originally had the ball to get the ball back.

39. A problem of interest to sociologists is to determine the proportion of society that has an upper- or lower-class occupation. One possible mathematical model would be to assume that the transitions between social classes of the successive generations in a family can be regarded as transitions of a Markov chain. That is, we assume that the occupation of a child depends only on his or her parent's occupation. Let us suppose that such a model is appropriate and that the transition probability matrix is given by

$$P = \begin{pmatrix} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.70 & 0.25 \\ 0.01 & 0.50 & 0.49 \end{pmatrix}.$$

That is, for instance, we suppose that the child of a middle-class worker will attain an upper-, middle-, or lower-class occupation with respective probabilities 0.05, 0.70, 0.25.

    (a) Is this chain

        i. irreducible?

        ii. non-null persistent?

iii. reversible?

(b) In the long run, what percent of people in such a society have upper-, middle-, or lower-class jobs?

(c) A worker has a middle-class occupation. On average, how many generations pass until his/her descendent will have a middle-class job as well?

40. Let $X_n$ be a discrete-time Markov chain with state space $S = \{1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

(a) Classify the states of the chain.

(b) Suppose that $\alpha\beta > 0$ and $\alpha\beta \neq 1$.

i. Find the $n$-step transition probabilities and show that they converge to the unique stationary distribution as $n \to \infty$.

ii. For what values of $\alpha$ and $\beta$ is the chain time-reversible in equilibrium?

41. Each of the processes described below is a birth and death process. Find the birth and death rates, holding rates, and specify transition probabilities of the embedded discrete time jump chain (i.e., write $\lambda_i$, $\mu_i$, $\lambda(i)$, $Q_{i,j}$'s).

(a) $M/M/1$ Queue. Suppose customers arrive according to a Poisson process with parameter $\lambda > 0$. They are served by a single server and leave. Suppose the service times are exponentially distributed with parameter $\mu > 0$ and that whenever there is more than one customer waiting for service the excess customers form a queue and wait until their turn. $X(t)$ denotes the number of customers in the system at time $t$.

(b) $M/M/N$ Queue. Read part (a) first! Customers arrive according to a Poisson process with parameter $\lambda > 0$, but now they are served by $N$ servers, each of whom serves at rate $\mu$. $X(t)$ denotes the number of customers in the system at time $t$.

(c) $M/M/\infty$ Queue. Read part (a) first! Arrivals are Poisson with rate $\lambda$, service times are exponential with parameter $\mu$, and there are an infinite number of servers. $X(t)$ denotes the number of customers in the system at time $t$.

(d) $M/M/1$ Queue with balking. Read part (a) first! Potential customers arrive at a single-server station in accordance with a Poisson process with rate $\lambda$. However, if the arrival finds $n$ customers already in the station, then he will enter the system with probability $\alpha_n$. Assume an exponential service rate $\mu$. $X(t)$ denotes the number of customers in the system at time $t$.

(e) Branching process. Consider a collection of particles which act independently in giving rise to succeeding generations of particles. Suppose that each particle, from the time it appears, waits a random length of time having an exponential distribution with parameter $q$ and then splits into two identical particles with probability $p$ or vanishes with probability $1-p$. $\{X(t),\ 0 \le t < \infty\}$ denotes the number of particles present at time $t$.

(f) Branching process with immigration. Consider the branching process introduced in part (e). Suppose that new particles immigrate into the system at random times that form a Poisson process with parameter $\lambda$ and then give rise to succeeding generations as described in part (e). $X(t)$ denotes the number of particles present at time $t$.

42. Superposition of independent Poisson processes is a Poisson process. Flies and wasps land on your dinner plate in the manner of independent Poisson process with respective intensities $\lambda$ and $\mu$. Show that the arrivals of flying objects form a Poisson process with intensity $\lambda + \mu$.

43. Thinning. Insects land in the soup in the manner of a Poisson process with intensity $\lambda$. Each such insect is green with probability $p$, independently of the colors of all other insects. Show that the arrivals of the green insects form a Poisson process with intensity $\lambda p$.

44. Let $\{X_t, t \ge 0\}$ be a Markov chain on $\{1, 2\}$ with generator

$$A = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}$$

where $\lambda\mu > 0$.

(a) Describe the dynamics of the chain.

(b) Write down the forward equations and solve them for the transition probabilities $P_{ij}(t)$, $i, j = 1, 2$.

(c) Calculate $A^n$ and hence find

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} A^n.$$

Compare your answer with that of part (b).

(d) Find $P(X(t) = 2 \mid X(0) = 1, X(3t) = 1)$.

(e) Solve the equation $\pi A = 0$ in order to find the stationary distribution. Verify that $P_{ij}(t) \to \pi_j$ as $t \to \infty$.

45. Consider the following continuous-time branching process with binary splitting: if a given particle is alive at a certain time, its additional life length is an exponential random variable with parameter $a$. Upon death, the particle either splits in two with probability $p$ or vanishes with probability $q$. The particles act independently of other particles and of the history of the process. Let $X(t)$ be the number of particles in the population at time $t$.

  (a) Find the generator matrix and write out the forward equations.

  (b) Find the stationary distribution of $X(t)$.

46. $M/M/\infty$ Queue. Arrivals are Poisson with rate $a$, service times are exponential with parameter $b$, and there are an infinite number of servers. Let $X(t)$ be the number in the system at time $t$.

  (a) Give the generator matrix.

  (b) Show by solving balance equations that the stationary and limiting distribution is Poisson.

47. Potential customers arrive at a single-server station in accordance with a Poisson process with rate $\lambda$. However, if the arrival finds $n$ customers already in the station, then he will enter the system with probability $\alpha_n$. Assuming an exponential service rate $\mu$, set this up as a birth and death process.

  (a) Give the generator matrix $A$.

  (b) Determine the stationary distributions when $\alpha_0 = 1$, $\alpha_n = a$ $(0 < a < 1)$, $n \geq 1$, $\rho = \lambda/\mu < 1$.

  (a) Give an example of a Markov chain on a finite state space $S$ such that three of the states each have a different period. (Total number of states is $|S| = N > 3$.)

  (b) Give an example of a Markov chain on a finite state space that has two or more stationary distributions.

  (c) Let $(X_n)$ be a symmetric random walk on the nonnegative integers with the following transition probabilities: $p_{i,i+1} = 1/2$ for $i \geq 0$; $p_{i,i-1} = 1/2$ for $i \geq 1$, and $p_{0,0} = 1/2$. Show that $(X_n)$ is null recurrent.

(d) Give an example of a Markov chain for which some states are positive recurrent, some states are null recurrent, and some states are transient.

48. Consider two boxes $A$ and $B$. Suppose there are $d$ balls divided between the boxes. Initially, some of the balls are in box $A$ and some of the balls are in box $B$. Suppose at each step, we choose one ball uniformly at random from among the $d$ balls and switch it to the opposite box. Let $X_n$ be the number of balls in box $A$ at time $n$.

   (a) Describe the state space and transition probabilities for $X_n$. Classify its states.

   (b) Determine the stationary distribution. Is the chain time-reversible? How many balls should one expect to find in box $A$ at equilibrium?

   (c) Does the limit of the sequence $\{P_0(X_n = 0)\}$ exist when $n \to \infty$? If the limit exists, find the limit. If not, explain whether this contradicts the ergodic theorem.

49. A mature individual produces offspring according to the probability generating function $f(s)$. Each immature individual grows to maturity with probability $p$ and then reproduces independently of the other individuals.

   (a) Suppose that we have a population of $k$ immature individuals. Find the probability generating function of the number of (immature) individuals at the next generation.

   (b) Suppose that there are $k$ mature individuals in the parent population. Find the probability generating function of the number of mature individuals at the next generation.

50. Consider a continuous time random walk on the half line. After a mean 1 exponential time, the walker jumps one unit to the left with probability $q$ or one unit to the right with probability $p$. Describe this process in terms of parameters $(\lambda, Q)$, where $\lambda(i)$'s are the holding rates of a continuous time Markov chain and $Q_{i,j}$'s are the transition probabilities of the embedded jump processes. Does this process have a stationary distribution? If yes, find it. If necessary, introduce conditions on $p$ and $q$.

51. Assume that each member in a population has a probability $\beta h + o(h)$ of giving birth to a new member in an interval of time $h$ ($\beta > 0$). Assume that the population is started with one individual and let $X_t$ be the number of births in this population by time $t$.

(a) Write forward equations for $P_{1,n}(t) = P(X_t = n \mid X_0 = 1)$ and show that $X_t$, the number of births in the population, has a geometric distribution with mean $e^{\beta t}$.

(b) Does the process $X_t$ have any stationary distributions?

# Appendix A

# Markov Chain Monte Carlo Methods

Consider the crude Monte Carlo method: $I = \int_0^1 f(x)\ dx = Ef(U)$, where $U = U(0,1)$ is a random variable with uniform distribution. By the strong law of large numbers, we can simulate by letting $U_1, \ldots, U_N$ be independent and identically random variables with uniform distribution on $(0,1)$. Then

$$I = \int_0^1 f(x)\ dx \equiv Ef(U) \approx \frac{1}{N}\sum_{i=1}^{N} f(U_i).$$

We can take a sample $U_1, \ldots, U_N$ of independent identically distributed random variables. Then

$$\frac{f(U_1) + \cdots + f(U_N)}{N} \to Ef(U) = \mu$$

almost surely.

**Chebyshev's Inequality** (See problems.) $ES_N = I \equiv Ef(U)$. Then

$$\begin{aligned} \Pr(|S_N - I| > \varepsilon) &\le \frac{\operatorname{Var} S_N}{\varepsilon} \\ &= \frac{\operatorname{Var} f(U)}{N\varepsilon^2}; \\ \operatorname{Var} S_N &= \frac{1}{N^2} N \operatorname{Var} f(U) \\ &= \frac{\operatorname{Var} f(U)}{N}. \end{aligned}$$

## A.1 Importance Sampling

We have

$$\begin{aligned} I &= \int_X f(x)\,dx \\ &= \int_X \frac{f(x)}{p(x)} p(x)\,dx \\ &\equiv E\left[\frac{f(x)}{p(x)}\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)}{p(X_i)}, \end{aligned}$$

where $p(x)$ is a density function, $p(x) > 0$, and $X_1, \ldots, X_N$ are independent identically distributed random variables with probability density function $p(x)$.

**Summary.** Rewrite $I = \int f(x)\pi(x)\,dx = Ef(x)$, where $X \sim \pi(x)$ for some $f$ and some distribution $\pi$. Approximate

$$I = Ef(X) \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i).$$

Then $X_i$'s are sampled from a distribution $\pi$ independently.

If $X_1, \ldots, X_N$ come from a Markov chain with stationary distribution $\pi$, then the strong law of large numbers holds, and

$$\frac{1}{N} \sum_{i=m} m + N f(X_i) \to \pi(f) = Ef(X), \qquad X \sim \pi.$$

**Observation.** Let $X_1, \ldots, X_n$ have probability density functio $p(x)$. Define a vector $(X, Y)$ such that

$$\begin{cases} X \mid Y \sim p(x|y) \\ Y \mid X \sim p(y \mid x). \end{cases}$$

Then

$$\begin{aligned} p(x) &= \int p(x \mid y) p(y)\, dy \\ &= \int p(x \mid y) \left[ \int p(y \mid x') p(x')\, dx' \right] dy \\ &= \int h(x, x') p(x')\, dx', \end{aligned}$$

where

$$h(x, x') = \int p(x \mid y) p(y \mid x')\, dy.$$

**Summary.** We have

$$p(x) = \int h(x, x') p(x')\, dx'.$$

Hence $p(x)$ is a solution of this integral equation. Under some conditions, the solution exists and is unique.

**Conclusion.** Conditional densities $p(x \mid y)$ and $p(y \mid x)$ uniquely determine $p(x)$.

Markov chain Monte Carlo methods are techniques for generating random variables from a density without having the density explicitly, but given a variety of facts about it.

## A.2 Motivation

We discuss Bayesian interference. Let us say we would like to estimate a proportion $\theta$. Take an independent identically distributed sample $X_1, \ldots, X_n$, where

$$X_i = \begin{cases} 1 & \text{if “no” on Proposition 80,} \\ 0 & \text{otherwise.} \end{cases}$$

I think this lecture was on the same day as the California election. Then $X_i \sim \text{Bin}(1, \theta)$ with observations $x_1, \ldots, x_n$.

Let $p(\theta)$ be the prior distribution of $\theta$ (which comes from intuition, prior knowledge, or something else). Let $p(x \mid \theta)$ be the likelihood function, the joint density of observed values $x = (x_1, \ldots, x_n)$ with parameter $\theta$. Then

$$\begin{aligned} p(x \mid \theta) &= \Pr(X_1 = x_1, \ldots, X_n = x_n \mid \theta) \\ &= \theta^{x_1}(1-\theta)^{1-x_1} \cdots \theta^{x_n}(1-\theta)^{1-x_n} \\ &= \theta^{x_1+\cdots+x_n}(1-\theta)^{n-x_1-\cdots-x_n}. \end{aligned}$$

If $k$ individuals in the sample vote “no” on 80, then $p(x \mid \theta) = \theta^k(1-\theta)^{n-k}$. We want $p(\theta \mid x)$, the probability density function of $\theta$ given a sample $x = (x_1, \ldots, x_n)$.

By Bayes's formula we have

$$p(\theta \mid x) = \frac{p(x \mid \theta)p(\theta)}{p(x)}.$$

Often we are interested in

$$E(g(\theta) \mid x) = \int g(\theta)p(\theta \mid x)\, d\theta.$$

Then

$$p(x) = \int p(x \mid \theta)p(\theta)\, d\theta$$

is a normalizing factor.

There are two approaches to solving integration problems:

- Laplace Approximations (Leonard and Hsu).
- Markov chain Monte Carlo methods or Gibbs samples (Bernardo and Smith).

## A.3 Physics Example: The Ising Model

Consider a finite graph $\mathcal{G} = (V, E)$. Each vertex may be in either of two states $+1$ or $-1$. A configuration is $\theta = (\theta_v : v \in V)$. We have $\theta \in \Theta = \{-1, 1\}^{|V|}$, where $|V|$ is the number of vertices. Each configuration has probability

$$p(\theta) = \frac{1}{Z} \exp\left\{ \sum_{\substack{v \neq w \\ v \sim w}} \theta_v \theta_w \right\},$$

where $v \sim w$ means that these two vertices are adjacent and $Z$ is a partition function or normalizing constant given by

$$Z = \sum_{\theta \in \Theta} \exp\left\{ \sum_{\substack{v \neq w \\ v \sim w}} \theta_v \theta_w \right\}.$$

## A.4 General Metropolis-Hastings Algorithm

This is a method to produce a random sequence from a given density.

- We generate a vector $X$ such that $\Pr(X = x_j) = \frac{b_j}{B}$, where $B = \sum_{j=1}^{\infty} b_j < \infty$ cannot be computed. (In the continuous case, $\frac{b_j}{B}$ is the density function; for example, $b_j \to b(\theta) = p(x \mid \theta)p(\theta)$ and $B \to p(x) = \int p(x \mid \theta)p(\theta)\, d\theta$.)
- The Metropolis-Hastings algorithm generates a Markov chain which is *time-reversible* and has stationary distributions $\pi_j = \frac{b_j}{B}$. If $\{X_j\}$ is such a Markov chain, then
$$\frac{1}{N} \sum_{j=1}^{N} f(X_j) \to Ef(X) = \frac{1}{B} \sum_j f(X_j) b_j.$$

1. Let $Q = \{q_{ij}\}$ be *any* stochastic matrix, called the **proposal matrix**.

2. Let $A = \{\alpha_{ij}\}$ be a matrix with values $0 \leq \alpha_{ij} \leq 1$ (probabilities), called acceptance probabilities.

3. Generate a Markov chain as follows: if $X_n = i$, generate a candidate for $X_{n+1}$: $X^c_{n+1} \sim q(i, \cdot)$, i.e. $\Pr(X^c_{n+1} = j \mid X_n = i) = q_{ij}$ for $j = 1, 2$. Then we choose $X_{n+1}$ as follows:

   - $X_{n+1} = X^c_{n+1} = j$ with probability $\alpha_{ij}$ (accept the proposed candidate), or
   - $X_{n+1} = X_n = i$ with probability $1 - \alpha_{ij}$ (reject the proposal and repeat $X_n$).

4. The resulting $(X_n)$ has the following transition probabilities:

$$
\begin{aligned}
i \neq j \qquad & p_{ij} = \alpha_{ij} q_{ij}, \\
i = j \qquad & p_{ii} = q_{ii} + \sum_{j \neq i} q_{ij}(1 - \alpha_{ij}).
\end{aligned}
$$

5. We would like to generate a time-reversible Markov chain $(X_n)$ with stationary distribution $\pi_j = \frac{b_j}{B}$. How do we choose $Q$ and $A$?

   - Choose $Q$ so that it is easy and cheap to generate variables.
   - Choose $A$ to assume the time-reversibility of $(X_n)$ and correct stationary distributions. Use the balance equations $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j$. Then from (4), if $i \neq j$, $\pi_i \alpha_{ij} q_{ij} = \pi_j \alpha_{ji} q_{ji}$. Choose $\alpha_{ij} = \min\left\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\}$. If $\alpha_{ij} = 1$, then $\alpha_{ji} = \frac{\pi_i q_{ij}}{\pi_j q_{ji}}$.
   - To generate $\{\pi_j = b_j/B\}$, then $\alpha_{ij} = \min\left\{1, \frac{b_j q_{ji}}{b_i q_{ij}}\right\}$, which is independent of $B$. We can then choose $X_n = i$ and $X^c_{n+1} \sim q(i, \cdot)$ so that

$$
X_{n+1} = \begin{cases} X^c_{n+1} = j & \text{with probability } \alpha_{ij}, \\ X_n = i & \text{with probability } 1 - \alpha_{ij}. \end{cases}
$$

   If $i \neq j$, then $p_{ij} = \alpha_{ij} q_{ij}$. If $i = j$, then $p_{ii} = q_{ii} + \sum_{j \neq i} q_{ij}(1 - \alpha_{ij})$.

**Note.** There are possible improvements to increase the acceptance rate. For example, choose $T = \{t_{ij}\}$ symmetric with $t_{ij} \geq 0$ such that $\alpha'_{ij} = \alpha_{ij} t_{ij} < 1$ for all $i, j$. Then the $\alpha$'s also satisfy balance equations $\pi_i \alpha_{ij} q_{ij} = \pi_j \alpha_{ji} q_{ji}$ and

$$\pi_i [\alpha_{ij} t_{ij}] q_{ij} = \pi_j [\alpha_{ji} t_{ji}] q_{ji} \qquad \text{(Hastings)}.$$

Then the acceptance probability is $\pi(X_{n+1}^c)/\pi(X_n)$.

A special case is the Gibbs sampler, used in Bayesian interference and introduced by Geman and Geman in 1984.

The Gibbs sampler breaks down the problem of drawing samples from multivariate density into one of drawing successive samples from densities of smaller dimensions (e.g. univariate).

**Recall.**

- We observe a sample $(x_1, \ldots, x_n)$ from a distribution with parameter $\theta$ unknown.
- We know
  - $p(\theta)$, the prior distribution,
  - $p(x \mid \theta)$, the likelihood function.
- The ultimate goal is to calculate $Eg(\theta) = \int g(\theta) p(\theta \mid \pi)\, d\theta$ or to produce a sample of values of $\theta$ from a posterior density $p(\theta \mid x) = \frac{p(x|\theta)p(\theta)}{p(x)}$ and to use this sample to estimate the density $p(\theta)$ and other statistics.

We have $p(\theta \mid x) = \frac{b(\theta)}{B}$, with $b(\theta) = p(x \mid \theta) p(\theta)$ and $B = p(x) = \int p(x \mid \theta) p(\theta)\, d\theta$. The acceptance probabilities are $\alpha_{\theta_n, \theta_{n+1}^c} = \min\{1, r\}$, where

$$r = \frac{p(\theta_{n+1}^c) q(\theta_{n+1}^c, \theta_n)}{b(\theta_n) q(\theta_n, \theta_{n+1}^c)} = \frac{p(\theta_{n+1}^c)/q(\theta_n, \theta_{n+1}^c)}{p(\theta_n \mid x)/q(\theta_{n+1}^c, \theta_n)}.$$

Assume that $\theta = (\theta_1, \theta_2)$ and that $\theta^c_{1,n+1}$ is sampled from $p(\theta^c_{1,n+1} \mid \theta_{2,n}, x)$, $\theta^c_{2,n+1} = \theta_{2,n}$ and that $\theta^c_{1,n+2} = \theta_{1,n+1}, \theta^c_{2,n+2}$ is sampled from $p(\theta^c_{2,n+2} \mid \theta_{1,n+1}, x)$. Then

$$q(\theta_n, \theta^c_{n+1}) = \begin{cases} p(\theta^c_{1,n+1} \mid \theta_{2,n}, x) & \text{if } \theta^c_{2,n+1} = \theta_{2,n}, \\ p(\theta^c_{2,n+1} \mid \theta_{2,n}, x) & \text{if } \theta^c_{1,n+1} = \theta_{1,n}, \\ 0 & \text{otherwise.} \end{cases}$$

One can calculate $r = 1$.