1 Introduction

A collection of random variables $X = (X_n, n \geq 0)$ forms a Markov chain if for all n and for all possible values of the random variables $(X_n, n \geq 0)$ we have

$$P\left(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, ..., X_0 = i_0\right) = P\left(X_n = i_n | X_{n-1} = i_{n-1}\right),$$

where $i_j \in S$ for $j = 0, 1, ...n$, and S is a countable set.

We will assume that $S = (0, 1, ...)$ and let

$$p_{ij}^{n-1,n} = P\left(X_n = j | X_{n-1} = i\right)\ n \geq 1,\ i, j \in S.$$

Example 1.1 (Coin tossing) Let $X_0 = 0, X_n = \sum_{k=1}^{n} Y_k$, $P\left(Y_k = 1\right) = P\left(Y_k = 0\right) = \frac{1}{2}$, where Y_k's are independent and identically distributed random variables (i.i.d.). Then

$$p_{ij}^{n-1,n} = \begin{cases} \frac{1}{2} & \text{if } \ j = i \\ \frac{1}{2} & \text{if } \ j = i + 1 \\ 0 & \text{otherwise.} \end{cases}$$

In this example $p_{ij}^{n-1,n} = p_{ij}$ for all $n \geq 1, i, j \in S$. If $p_{ij}^{n-1,n} = p_{ij}$ for all $n \geq 1$, we say that a Markov chain has **stationary transition probabilities** or that it is **time homogeneous.** We will restrict attention to Markov chains with stationary transition probabilities. In this case we omit superscripts, that is, we write

$$p_{ij} = P\left(X_n = j | X_{n-1} = i\right),$$

where p_{ij} is called a one-step transition probability. Let $P = [p_{ij}]$ be the matrix of the one-step transition probabilities. P is called a one-step transition probability matrix. P is a stochastic matrix:

$$\begin{aligned} \sum_{j \in S} p_{ij} &= 1, \forall i \in S, \\ p_{ij} &\geq 0, i, j \in S. \end{aligned}$$

Also let

$$\phi_j^{(0)} = P\left(X_0 = j\right), \quad j \in S,$$

where $\sum_{j \in S} \phi_j^{(0)} = 1, \phi_j^{(0)} \geq 0$. The vector $\phi^{(0)} = \left(\phi_j^{(0)},\ j \in S\right)$ is called the initial distribution of X. The one step transition matrix P and the initial distribution $\phi^{(0)}$ determine all finite distributions of X and thus determine the process completely. To see this consider the joint probability function

$$\begin{aligned}
&P\left(X_n = i_n, X_{n-1} = i_{n-1}, ..., X_0 = i_0\right) \\
&= P\left(X_n = i_n | X_{n-1} = i_{n-1}, ..., X_0 = i_0\right) P\left(X_{n-1} = i_{n-1}, ..., X_0 = i_0\right) \\
&= P\left(X_n = i_n | X_{n-1} = i_{n-1}\right) P\left(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}\right) \\
&\times P\left(X_{n-2} = i_{n-2}, ..., X_0 = i_0\right) \\
&= p_{i_{n-1},i_n} p_{i_{n-2},i_{n-1}} ... p_{i_0,i_1} P\left(X_0 = i_0\right) = p_{i_{n-1},i_n} p_{i_{n-2},i_{n-1}} ... p_{i_0,i_1} \phi_{i_0}^{(0)}.
\end{aligned}$$

We now consider some **classic examples of Markov chains**.

Example 1.2 Random walk

Let $X = (X_n,\ n \geq 0)$ be a MC on S=(all integers) with the one-step transition probabilities given by

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

Example 1.3 Birth and death chain

Let $X = (X_n, n \geq 0)$ be a Markov chain on $S = (0, 1, ...)$ with the transition probability matrix P given by

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ . \\ . \\ . \end{array} \begin{pmatrix} r_0 & p_0 & & & & \\ q_1 & r_1 & p_1 & & & \\ & q_2 & r_2 & p_2 & & \\ & & . & . & . & \\ & & & . & . & . \\ & & & & . & . \; . \end{pmatrix}$$

where $p_i > 0, q_i > 0, r_i \geq 0, p_i + q_i + r_i = 1$. Here we interpret X_n as the size of the population at time n. We see that the probability of birth, death or no change depends on the size of the population.

Example 1.4 Gambler's ruin chain on S $=(0, 1, 2, ...)$.

In a birth and death chain, let $p_i = p,\ q_i = q,\ r_i = 0$ for all $i \geq 1,\ r_0 = 1$. Then

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \end{array} \begin{pmatrix} 1 & 0 & & & & & \\ q & 0 & p & & & & \\ & q & 0 & p & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \end{pmatrix}$$

We can interpret X_n as a fortune of the gambler after n games. Here state 0 is an absorbing state.

Example 1.5 Gambler's ruin chain with 2 absorbing barriers on $\mathbf{S} = (\mathbf{0}, \mathbf{1}, \mathbf{2}, ..., \mathbf{N})$. If $N = 4$

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 1 & & & & \\ q & 0 & p & & \\ & q & 0 & p & \\ & & q & 0 & p \\ & & & & 1 \end{pmatrix}$$

Here we suppose that if the capital of the gambler increases to 4 dollars he quits playing.

Example 1.6 A random walk with a reflecting state at 0 and an absorbing state at N where $N = 4$

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 0 & 1 & & & \\ q & 0 & p & & \\ & q & 0 & p & \\ & & q & 0 & p \\ & & & & 1 \end{pmatrix}$$

Here the gambler gets \$1 with the probability of one when his fortune reaches state zero.

2 Chapman-Kolmogorov equations

Definition 2.1 The n-step transition probability $p_{ij}^{(n)}$ is defined as

$$p_{ij}^{(n)} = P\left(X_n = j | X_0 = i\right),$$

for every $n \geq 0$, and all pairs of states (i, j). By time homogeneity of X, for all $m \geq 0$, we have

$$P\left(X_n = j | X_0 = i\right) = P\left(X_{m+n} = j | X_m = i\right).$$

We define the n-step transition probability matrix $P^{(n)}$ as a matrix with the (i,j) entry given by $(p_{ij}^{(n)})$.We write

$$P^{(n)} = (p_{ij}^{(n)})_{i,j \in S}.$$

We also have $p_{ij}^{(1)} = p_{ij}$ and

$$p_{ij}^{(0)} = \begin{matrix} 1 & \text{if } i{=}j \\ 0 & \text{if } i \neq j \end{matrix},$$

and thus $P^{(1)} = P$ and $P^{(0)} = I$ (the identity matrix). To compute the n-step transition probabilities we first derive the **Chapman-Kolmogorov equations.**

Proposition 1 *Transition probabilities of a Markov chain satisfy the following equations*

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)}, n \geq 0, m \geq 0, \; for \; all \; i, j \in S.$$

Proof.

$$\begin{aligned} p_{ij}^{(n+m)} &= P(X_{n+m} = j | X_0 = i) = \sum_{k \in S} P(X_{n+m} = j, X_n = k | X_0 = i) = \\ \sum_{k \in S} P(X_{n+m} &= j | X_n = k, X_0 = i) P(X_n = k | X_0 = i) = \\ \sum_{k \in S} P(X_{n+m} &= j | X_n = k) p_{ik}^{(n)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)}. \end{aligned}$$

■

In matrix notation Chapman-Kolmogorov equations take the form

$$P^{(n+m)} = P^{(n)}P^{(m)} \; n, m \geq 0. \tag{1}$$

By (1)

$$P^{(2)} = P^{(1)}P^{(1)} = PP = P^2$$
$$P^{(3)} = P^{(2)}P^{(1)} = P^2P = P^3$$

so that by induction

$$P^{(n)} = P^n,$$

or in elementwise notation

$$p_{ij}^{(n)} = (P^n)_{ij}.$$

Thus, we see that the n-step transition probability $p_{ij}^{(n)}$ is simply the (i, j) entry in the matrix P^n.

Example 2.1 Consider a Markov chain on $S = \{0, 1\}$ with transition matrix:

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

Then

$$P^6 = \begin{pmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{pmatrix}^6 = \begin{pmatrix} 0.424 & 0.576 \\ 0.384 & 0.616 \end{pmatrix}$$

and thus

$$P(X_6 = 1|X_0 = 0) = 0.576.$$

Corollary 2 *The unconditional probabilities* $\phi_j^{(n)} = P(X_n = j)$ *are computed as follows*

$$\phi_j^{(n)} = P(X_n = j) = \sum_{i \in S} \phi_i^{(0)} p_{ij}^{(n)}, j \in S. \tag{2}$$

Proof.

$$P(X_n = j) = \sum_{k \in S} P(X_n = j,\ X_0 = k) = \sum_{k \in S} P(X_n = j|X_0 = k) P(X_0 = k)$$
$$= \sum_{k \in S} \phi_k^{(0)} p_{kj}^{(n)}. \blacksquare$$

In matrix notation equations (2) take the form

$$\phi^{(n)} = \phi^{(0)} P^n,$$

where $\phi^{(n)} = \left(\phi_0^{(n)}, \phi_1^{(n)}, ...\right)$ is a row vector. ■

Example 2.2 In **Ex 2.1** we assume that $\phi^{(0)} = (0.1, 0.9)$. Then

$$\phi^{(6)} = (0.1, 0.9) P^6 = (0.388, 0.612)$$

or in elementwise notation

$$\begin{aligned} \phi_0^{(6)} &= P(X_6 = 0) = 0.388, \\ \phi_1^{(6)} &= P(X_6 = 1) = 0.612. \end{aligned}$$

For a two-state MC with transition matrix

$$P = \begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, 0 < p, q < 1.$$

one can show, say by induction, that P^n is of the form

$$P^n = (p+q)^{-1} \left\{ \begin{pmatrix} q & p \\ q & p \end{pmatrix} + (1-p-q)^n \begin{pmatrix} p & -p \\ -q & q \end{pmatrix} \right\},$$

which, as $n \to \infty$, converges to

$$P^\infty = (q+p)^{-1} \begin{pmatrix} q & p \\ q & p \end{pmatrix},$$

the matrix with identical rows. Set $\pi = (\pi_0, \pi_1) = (q/(q+p), p/(q+p))$. Note that for any initial distribution $\phi^{(0)}$,

$$\lim_{n \to \infty} \phi^{(0)} P^n = \pi,$$

so π is called a long run (or steady state, or limiting) distribution of a MC. This distribution is **independent** of an initial distribution.

For a MC with

$$P = \begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{pmatrix},$$

$$P^\infty = \begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{pmatrix}$$

and $\pi = (0.4, 0.6)$.

3 Classification of States

To understand the limiting behavior of more general Markov chains we need to classify the states of a MC. As we will see later the limiting behavior of a MC depends on whether or not the chain returns to its starting state with the probability of 1.

Definition 3.1 We say that state i **leads** to state j, written $i \to j$, if there exists $n \geq 0$ such that $p_{ij}^{(n)} > 0$.

Remark 3.1 Since $p_{ii}^{(0)} = 1$ any state leads to itself.

The relation $\to$ is transitive, i.e. if $i \to j$ and $j \to k$ then $i \to k$. To see this note that if $p_{ij}^{(n)} > 0$ and $p_{jk}^{(m)} > 0$ then by Chapman-Kolmogorov's eqs

$$p_{ik}^{(n+m)} = \sum_{l \in S} p_{il}^{(n)} p_{lk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0$$

Definition 3.2 We say that i communicates with j, written $i \leftrightarrow j$, iff $i \to j$ and $j \to i$.

Communication relation satisfies

a. $i \leftrightarrow i$ for every $i \in S$ (reflexivity) [since $p_{ii}^{(0)} = 1$]

b. If $i \leftrightarrow j$ then $j \leftrightarrow i$ (symmetry) [this is immediate from the definition]

c. If $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$ (transitivity) [follows from the transitivity of "leading to" relation shown above]

The relation which satisfies a, b, and c is called an equivalence relation. The equivalence relation partitions the state space into disjoint sets of states called communication classes.

Exercise Find the communication classes for a MC with

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \cdots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Exercise Find the communication classes for a Markov chain on $S = \{0, 1, 2, 3, 4, 5\}$ with

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$$

Remark 3.2 It is possible to leave an equivalence class, but it is not possible to return or else the two classes would together form a single class.

Definition 3.3 A Markov chain is called **irreducible** if there is only one communication class. Otherwise it is called reducible.

3.1 Transient and Recurrent States

Let $X = (X_n,\ n \geq 0)$ be a MC with state space S. Define the first passage time to state $j \in S$ (the first time X enters state j) by

$$T_j \;=\; \min\{n \geq 1 : X_n = j\},$$

with $T_j = \infty$ if $X_n \neq j$ for all $n > 0$ (meaning that X never visits state j). We will use the following notation interchangeably.

$$P_i(\bullet) \equiv P\,(\bullet\,|\,X_0 \;=\; i).$$

Next for $m \geq 1$, define

$$f_{ij}^{(m)} \equiv P_i(T_j \;= m) = P\,(X_m = j, X_{m-1} \;\neq j ... X_2 \neq j,\;\; X_1 \neq j\,|X_0 = i).$$

$f_{ij}^{(m)}$ is the probability that X, starting from state i, makes its **first transition** into state j after m steps.

Remark 3.3 For $m = 1$, $f_{ij}^{(1)} = p_{ij}$. But for $m > 1$, $f_{ij}^{(m)}$ is different from $p_{ij}^{(m)} \;=\; P(X_m \;=\; j|X_0 \;=\; i)$, as $p_{ij}^{(m)}$ includes the possibility that X was in state j at some time before time m.

Note that $\{f_{ij}^{(m)} = P_i(T_j \;= m), m = 1, 2, ...\}$ is the probability distribution of T_j conditional on $X_0 = i$. The probability of ever making a transition to state j, if the process starts in state i is

$$f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)} = P_i(T_j < \infty),$$

so that in particular

$$f_{ii} = P_i(T_i < \infty)$$

is the probability of returning to state i in a finite time.

Definition 3.4 A state i is called recurrent if $f_{ii} = 1$ and transient if $f_{ii} < 1$.

Example 3.1 An absorbing state i is recurrent since

$$f_{ii} = P_i(T_i = 1) = 1.$$

Exercise Consider a Markov chain on $S = \{0, 1, 2\}$ with

$$\begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 0 & 1 & 0 \\ 0.25 & 0 & 0.75 \\ 0 & 0 & 1 \end{pmatrix}$$

Compute $f_{ii}, i = 0, 1, 2$. Which states are recurrent and which are transient?

Exercise Consider a Markov chain on $S = \{0, 1, 2, 3\}$ with transition matrix

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0.3 \end{pmatrix}$$

Compute $f_{ii}, i = 0, 1, 2, 3$. Which states are transient and which are recurrent?

In order to obtain the criterion for recurrence, let

$$N(j) = \sum_{n=1}^{\infty} I(X_n = j),$$

where

$$I(X_n = j) = \begin{cases} 1 \text{ if } X_n = j \\ 0 \text{ if } X_n \neq j \end{cases},$$

so that $N(j) =$ the number of times X_n, $n \geq 1$ visits state j. Note that

$$\{T_j < \infty\} = \{N(j) \geq 1\}$$

and therefore also

$$f_{jj} = P_j(T_j < \infty) = P_j(N(j) \geq 1).$$

From Lemma 18 in In Appendix A,

$$P_j(N(j) \geq m) = f_{jj}^m. \tag{3}$$

This result is important for proving the following Theorem

Theorem 3 *(a) If j is a transient state then*

$$P_j(N(j) < \infty) = 1, \text{ and } E_j N(j) = \frac{f_{jj}}{1 - f_{jj}} < \infty, i \in S.$$

(b) If j is a recurrent state, then

$$P_j(N(j) = \infty) = 1 \text{ and } E_j N(j) = \infty.$$

Proof. (a) If $f_{jj} < 1$, then by (3)

$$P_j(N(j) = m) = (f_{jj})^m (1 - f_{jj}),\ m \geq 0,$$

so that $N(j)|X_0 = j$ has a geometric distribution with the probability of failure equal to f_{jj}. And

$$P_j(N(j) = \infty) = \lim_{m \uparrow \infty} P_j(N(j) = m) = \lim_{m \uparrow \infty} (f_{jj})^m (1 - f_{jj}) = 0$$

and hence $P_j(N(j) < \infty) = 1 - P_j(N(j) = \infty) = 1$. Next compute

$$\begin{aligned} E_j[N(j)] &= \sum_{m=1}^{\infty} m P_j(N(j) = m) = \sum_{m=1}^{\infty} P_j(N(j) \geq m) \\ &= f_{jj} \sum_{m=1}^{\infty} f_{jj}^{m-1} = \frac{f_{jj}}{1 - f_{jj}} < \infty. \end{aligned}$$

(b) If $f_{jj} = 1$ then by (3)

$$P_j(N(j) = \infty) = 1,$$

from which it immediately follows that $E_j(N(j)) = \infty$. ■

Corollary 4 *(Criterion for recurrence) State j is recurrent iff*

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$$

Proof.

$$E_j N(j) = E_j[\sum_{n=1}^{\infty} I\left(X_n = j\right)] = \sum_{n=1}^{\infty} E_j I\left(X_n = j\right) = \sum_{n=1}^{\infty} P_j(X_n = j) = \sum_{n=1}^{\infty} p_{jj}^{(n)}$$

and the corollary follows by Theorem 3(b). ■

Corollary 5 *If $i \leftrightarrow j$ and if i is recurrent then j is recurrent. (Recurrence is a class property.)*

Let m, n be such that $p_{ij}^{(n)} > 0$, and $p_{ji}^{(m)} > 0$. By assumption $\sum\limits_{k=1}^{\infty} p_{ii}^{(k)} = \infty$. We want to show $\sum\limits_{k=1}^{\infty} p_{jj}^{(k)} = \infty$. Now, for every $k \geq 0$

$$p_{jj}^{(m+k+n)} \geq p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)},$$

and hence

$$\sum_{k=1}^{\infty} p_{jj}^{(m+k+n)} \geq \sum_{k=1}^{\infty} p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)} = p_{ji}^{(m)} \left(\sum_{k=1}^{\infty} p_{ii}^{(k)} \right) p_{ij}^{(n)} = \infty.$$

Remark 3.3 In a finite MC not all states can be transient.

Remark 3.4 In a finite irreducible MC all states are recurrent.

Example 3.2 Let $X = (X_n,\ n \geq 0)$ be a random walk on $S = \{\text{all integers}\}$ with transition matrix given by

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

where $q, p > 0$ and $p + q = 1$. We want to evaluate $\sum\limits_{n=1}^{\infty} p_{ii}^{(n)}$, $i \in S$. We have

$$p_{ii}^{(2m)} > 0,\ p_{ii}^{(2m+1)} = 0, m = 1, 2, ...$$

and

$$p_{ii}^{(2m)} = \begin{pmatrix} 2m \\ m \end{pmatrix} p^m q^m.$$

We approximate $m!$ using Stirling's approximation, which works well for large m:

$$m! \cong m^m e^{-m} \sqrt{2\pi\, m},$$

so that

$$\frac{(2m)!}{m!m!} = \frac{(2m)^{2m} e^{-2m} \sqrt{4\pi\, m}}{m^{2m} e^{-2m} 2\pi\, m} = \frac{2^{2m}}{\sqrt{\pi\, m}}.$$

Thus

$$\sum_{m=1}^{\infty} p_{ii}^{(2m)} \approx \sum_{m=1}^{\infty} \frac{4^m}{\sqrt{\pi\, m}} (pq)^m = \sum_{m=1}^{\infty} \frac{(4pq)^m}{\sqrt{\pi\, m}} = \begin{cases} < \infty & if\ pq < \frac{1}{4}, \\ = \infty & if\ pq = \frac{1}{4} \end{cases}$$

where we used the fact: $\sum\limits_{m=1}^{\infty} \frac{1}{m^s}$ converges if $s > 1$ and diverges if $s \leq 1$.

3.2 Periodicity

Definition 3.5 State i is said to have period $d(i)$ if $d(i)$ is the greatest common divisor (gcd) of all integers $n \geq 1$ for which $p_{ii}^{(n)} > 0$. If $p_{ii}^{(n)} = 0$ for all $n \geq 1$ then $d(i) = \infty$. If $d(i) = 1$ then state i is called aperiodic.

Example 3.3 Random walk has period 2.

Proposition 6 *If $i \leftrightarrow j$ then $d(i) = d(j)$, i.e., periodicity is a class property (For a proof see Proposition 4.2.2, page 169 in Ross(B))*

Example 3.4 A chain with the transition probability matrix which has all diagonal entries positive is always aperiodic.

Exercise. Determine periodicity of each state in the chains below.

$$P = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix} \qquad P = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \end{pmatrix}$$

4 Return Times

For a MC on S, let T_j be the *first passage* time to state j. Then

$$\mu_{ij} = E_i(T_j), i, j \in S$$

is the expected time in which a MC that started in state i reaches state j (for the first time). In particular μ_{ii} is the expected return time to state i.

Example 4.1 Consider a MC on $S = \{0, 1\}$ with transition matrix

$$\begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

We now compute μ_{00}

$$\begin{aligned} \mu_{00} &= E_0(T_0) = \sum_{n=1}^{\infty} nP(T_0 = n) = \sum_{n=1}^{\infty} P_0(T_0 \geq n) = 1 + \sum_{n=2}^{\infty} P_0(T_0 \geq n) \\ &= 1 + \sum_{n=2}^{\infty} p(1-q)^{n-2} = (p+q)/q, \end{aligned}$$

and similarly

$$\mu_{11} = E_1(T_1) = (p+q)/p.$$

Definition 4.1 A recurrent state j is called *positive recurrent* if $\mu_{jj} < \infty$ and *null recurrent* if $\mu_{jj} = \infty$.

In **Example 4.1** states 0 and 1 are positive recurrent.

Remark 4.1 By definition of a transient state j, $f_{jj} < 1$ or $P_j(T_j = \infty) > 0$. Hence $\mu_{jj} = E_j(T_j) = \infty$ if j is transient.

Theorem 7 *Let $p_{ij}^{(m)}$ be the m-step transition probability of an irreducible, aperiodic Markov chain. Then, for $i, j \in S$,*

$$\lim_{m\to\infty} p_{ij}^{(m)} = \frac{1}{\mu_{jj}}.$$

If state j is transient or null- recurrent, then

$$\lim_{m\to\infty} p_{ij}^{(m)} = 0. \tag{4}$$

Proof. *We omit the proof of the first statement. If j is a transient state, then $\mu_{jj} = \infty$ follows from Remark 4.1, whereas when j is a null recurrent state then $\mu_{jj} = \infty$ from the definition of null recurrence.* ■

Corollary 8 *Positive and null recurrence are class properties.*

For proof see Appendix B.

Corollary 9 *An aperiodic, irreducible, finite Markov chain is positive recurrent.*

Proof.

$$\lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n)} = 1 = \sum_{j\in S} \lim_{n\to\infty} p_{ij}^{(n)} = \sum_{j\in S} \frac{1}{\mu_{jj}},$$

which implies that $\mu_{jj} < \infty$ for some $j \in S$, i.e., there is at least one positive recurrent state. Hence by **Corollary** 8 (positive recurrence is a class property) all states are positive recurrent. ■

Exercise Consider an infinite MC on $S = (0, 1, 2, ...)$ with transition matrix given by $p_{00} = 0, p_{01} = 1$ and for $i \geq 1, p_{i,i+1} = i/(i+1), p_{i0} = 1/(i+1)$.

a) Show that the chain is recurrent.

b) Is the chain positive or null recurrent?

5 Stationary Distribution

In **Section 2** we showed by explicit computation that for a two-state MC with transition matrix

$$P = \begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = \begin{pmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{pmatrix},$$

we have

$$P^n \to P^\infty = \begin{pmatrix} q/(q+p) & p/(q+p) \\ q/(q+p) & q/(q+p) \end{pmatrix} = \begin{pmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{pmatrix}$$

which implies that independently of an initial distribution $\phi^{(0)}$

$$\lim_{n\to\infty} \phi^{(0)} P^n \to \pi = (\pi_0, \pi_1) = (0.4, 0.6).$$

But note that in this case we can write

$$\pi = \lim_{n\to\infty} \phi^{(0)} P^n = \left[\lim_{n\to\infty} \phi^{(0)} P^{n-1} \right] P = \pi P,$$

obtaining a simple matrix equation for π :

$$\pi = \pi P. \tag{5}$$

Exercise Find the limiting distribution of a MC with transition matrix

$$\begin{array}{c} 0 \\ 1 \end{array} \begin{pmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{pmatrix}$$

using (5).

This discussion motivates the following definition.

Definition 10 *A probability vector $\alpha = \{\alpha_i = P(X_0 = i), i \in S\}$ is called a stationary distribution for P if it satisfies the system of equations*

$$\alpha P = \alpha,$$

or in elementwise notation

$$\alpha_j = \sum_{i \in S} \alpha_i p_{ij}; j \in S.$$

Corollary 11 *If the stationary distribution α is an initial distribution of X, then the distribution of X_n will be α for $n \geq 1$.*

Proof. To show this we can use induction. We clearly have $\phi^{(1)} = \alpha P = \alpha$, and thus the distribution of X_1 is α. Assume now that the distribution of X_m is α, that is, $\phi^{(m)} = \alpha$, then

$$\phi^{(m+1)} = \phi^{(m)} P = \alpha P = \alpha$$

and thus distribution of X_{m+1} is also α for every m. ■

6 Long run behavior of Markov Chains

We now state the main theorem

Theorem 12 *For any irreducible, aperiodic Markov chain, there are two possibilities:*

(a) Markov chain is transient or null-recurrent. Then a stationary distribution does not exist

(b) Markov chain is positive recurrent. Then there exists a unique stationary distribution $(\pi_j, j \in S)$ for transition matrix P of this chain, which satisfies

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_{jj}}, i, j \in S.$$

Proof. Appendix C ■

Definition 6.1 An irreducible, positive recurrent and aperiodic Markov chain is called **ergodic**.

Corollary 13 *An aperiodic, irreducible,* ***finite*** *Markov chain is ergodic.*

Proof. Follows from Corollary 9. ■

Note that Theorem 12 (b) provides an easy way of finding the long run distribution of an ergodic chain with transition matrix P. To find such distribution π one simply solves a system of linear equations: $\pi = \pi P$.

Some interpretations

(i) For an ergodic chain

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

is a long run probability of finding a process in state j, $j \in S$, irrespective of the initial distribution.

(ii) If $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$, then also

$$\lim_{m\uparrow\infty} \frac{1}{m}\sum_{n=1}^{m} p_{ij}^{(n)} = \pi_j.$$

(This statement is true for any sequence of numbers). But recall that

$$\begin{aligned} \frac{1}{m}\sum_{n=1}^{m} p_{ij}^{(n)} &= \frac{1}{m}\sum_{n=1}^{m} P(X_n = j|X_0 = i) = \frac{1}{m}\sum_{n=1}^{m} E_i[I(X_n = j)] \\ &= \frac{1}{m}E_i\sum_{n=1}^{m}[I(X_n = j)] \end{aligned}$$

and thus π_j can also be interpreted as the expected proportion of time that the process spends in state j in the long run.

(iii) If a chain is irreducible, positive recurrent but with period $d > 1$, then there exists a unique stationary distribution

$\pi = (\pi_j,\ j \in S)$, where $\pi_j = 1/\mu_{jj}$, and in this case π_j is interpreted as a long run proportion of time X spends in state j.

Example 6.1 A two state MC with $p_{01} = 1 = p_{10}$ is periodic, but has a unique stationary distribution.

Example 6.2 Consider a simple random walk with reflecting barriers at 0 and $b+1$, i.e. $S = (0, 1, 2, 3...b+1), p_{01} = 1, p_{b+1,b} = 1, p_{i,i+1} = p, p_{i,i-1} = q = 1-p$ for $i = 1, 2, ...b$.

Questions: Does this chain have a stationary distribution? Is it unique? What is the interpretation of this distribution?

Answer also the same questions when $b \to \infty$.

Solution: Since this chain is irreducible and positive recurrent it has a unique stationary distribution. Each state has period equal to 2. To compute the unique stationary distribution we have to solve the following system of equations:

$$\alpha = \alpha P$$

where

$$\alpha = (\alpha_0, \alpha_1, ...\alpha_{b+1}).$$

We have

$$\begin{aligned}
\alpha_0 &= \alpha_1 q \to \alpha_1 = \alpha_0 / q \\
\alpha_1 &= \alpha_0 + \alpha_2 q \to \alpha_2 = (p/q^2)\alpha_0 \\
\alpha_2 &= \alpha_1 p + \alpha_3 q \to \alpha_3 = (p^2/q^3)\alpha_0 \\
\\
\alpha_b &= (p^{b-1}/q^b)\alpha_0 \\
\alpha_{b+1} &= p\alpha_b = (p/q)^b \alpha_0.
\end{aligned}$$

We have expressed the probabilities $\alpha_1, ...\alpha_{b+1}$ in terms of α_0. We can now solve for α_0 using

$$\sum_{k=0}^{b+1} \alpha_k = 1.$$

This gives

$$\alpha_0 \left[1 + \frac{1}{q} \left(1 + \frac{p}{q} + (\frac{p}{q})^2 + ...(\frac{p}{q})^{b-1} \right) + (\frac{p}{q})^b \right] = 1. \tag{6}$$

If $p = q = 1/2$ then (6) becomes

$$\alpha_0 (2 + 2b) = 1,$$

so that in this case

$$\alpha_0 = [2(b+1)]^{-1} = \alpha_{b+1}, \alpha_1 = \alpha_2 = ...\alpha_b = (b+1)^{-1}. \tag{7}$$

If $p \neq q$ then (6) becomes

$$\alpha_0 \left[1 + \frac{1}{q} \left(\frac{1-(p/q)^b}{1-(p/q)} \right) + (\frac{p}{q})^b \right] = 1,$$

Solving for α_0 we obtain

$$\alpha_0 = \frac{q-p}{2q} \times \frac{1}{1-(p/q)^{b+1}}, \tag{8}$$

and

$$\begin{aligned} \alpha_k &= \frac{p^{k-1}}{q^k}\alpha_0 \text{ for } k = 1, 2...b, && (9) \\ \alpha_{b+1} &= (\frac{p}{q})^b \alpha_0. \end{aligned}$$

Since this chain is periodic we interpret α_k as a long run expected proportion of time that the chain spends in state k, and not as a long run probability of finding the chain in state k.

We now consider what happens when $b \to \infty$. If $p = q$, then by (7) we have $\alpha_0 = \alpha_1 = .. = 0$, so that in this case there is no stationary distribution. If $p \neq q$ then we consider two cases: $p > q$ and $p < q$. If $p > q$, then by (8) and (9) we see that again there is no stationary distribution. But if $p < q$ then (8) becomes

$$\alpha_0 = \frac{q-p}{2q}, \tag{10}$$

and in this case there exists a unique stationary distribution given by (10) and

$$\alpha_k = \frac{p^{k-1}}{q^k}\alpha_0 \text{ for } k = 1, 2....$$

Exercise Give an example or quote a relevant result to justify each statement below.

Let k be the number of positive recurrent classes of a Markov chain $(X_n, n \geq 0)$. Then

(a) If $k = 0$, there is no stationary distribution.

(b) If $k = 1$, there is a unique stationary distribution.

(c) If $k \geq 2$, there are infinitely many stationary distributions.

7 Absorption probabilities and expected times to absorption

The following definition is useful for the discussion of the limiting behavior of a reducible Markov chain .

Definition 7.1 A set C of states is *closed* if no state in C leads to any state outside of C, i.e., if $p_{ij} = 0$ when $i \in C, j \notin C$.

Remark 7.1 A recurrent class R of states is closed.

Proof. Suppose that R is not closed, i.e., that $i \in R$ leads to j $\notin$ R. Now i is visited infinitely many times. Hence j must lead to i, but then i and j would be in the same class. Contradiction. ■

Example 7.1 Let

$$P = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

Here closed sets are: $\{1, 2, 3, 4, 5\}$ and {1,3,5}, {2,4} is a transient class.

In general we can decompose the state space in the following way

$$S = R_1 \cup R_2 \cup ... \cup R_k \cup T$$

where $R_1, R_2, ..., R_k$ are closed recurrent classes, and T is a set of transient states.

Example 7.2 Gambler's Ruin Chain on $S = (0, 1, 2, ...N)$.

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

Here $R_1 = \{0\}, R_2 = \{N\}$, and $T = \{1, 2, ..., N - 1\}$.

7.1 Absorption probabilities

Let j be a recurrent state. For $i \in T$ we want to compute

$$f_{ij} = P\left[T_j < \infty \,|\, X_0 = i\right].$$

f_{ij} is the probability of ever entering j given that the process starts in i. In this context, f_{ij} is called an *absorption probability.*

Proposition 14 *If j is recurrent, then the set of probabilities $\{f_{ij},\ i \in T\}$ satisfies*

$$f_{ij} = \sum_{k \in T} p_{ik} f_{kj} + \sum_{k \in R^j} p_{ik}, i \in T$$

where R^j is the set of states which communicate with j (includes j).

Proof.

$$\begin{aligned} f_{ij} &= P(T_j < \infty \,|\, X_0 = i) = \sum_{k \in S} P(T_j < \infty, X_1 = k |\, X_0 = i) \\ &= \sum_{k \in S} P(T_j < \infty \,|\, X_0 = i, X_1 = k) P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} f_{kj} p_{ik} = \sum_{k \in T} p_{ik} f_{kj} + \sum_{k \in R^j} p_{ik} f_{kj} + \sum_{k \notin R^j, k \notin T} p_{ik} f_{kj}. \end{aligned}$$

But $f_{kj} = 0$ if $k \notin R^j$ and if $k \notin T$, and $f_{kj} = 1$ if $k \in R^j$. hence we get the statement of the Proposition. ■

Example 7.3 We consider Gambler's Ruin Chain from **Example 1.5**
Here it is of interest to compute the absorption probabilities f_{iN} or f_{i0} , $i \in (1, 2, ...N-1)$. Obviously f_{iN}+ $f_{i0} = 1$, so we focus on f_{iN}. For brevity, let $f_i = f_{i0}$. The boundary conditions are: $f_0 = 1,\ f_N = 0$.

In this case the equations from **Proposition 14** become

$$f_i = \sum_{k \in T} p_{ik} f_k + p_{i0}, \qquad 1 \leq i \leq N-1$$

More compactly these equations can be written as

$$f_i = qf_{i-1} + pf_{i+1} \qquad\qquad 1 \leq i \leq N-1. \tag{11}$$

This is a system of difference equations which can be solved recursively. First we rewrite (11) as

$$\begin{aligned}
&(p+q)\,f_i = qf_{i-1} + pf_{i+1} \\
&p\,(f_{i+1} - f_i) = q\,(f_i - f_{i-1}) \\
&f_{i+1} - f_i = \tfrac{q}{p}\,(f_i - f_{i-1}) \\
&i.e. \\
&f_2 - f_1 = \tfrac{q}{p}\,(f_1 - f_0) = \tfrac{q}{p} f_1 \\
&f_3 - f_2 = \tfrac{q}{p}\,(f_2 - f_1) = \left(\tfrac{q}{p}\right)^2 f_1 \\
&\vdots \\
&f_i - f_{i-1} = \left(\tfrac{q}{p}\right)^{i-1} f_1 \\
&\vdots \\
&f_N - f_{N-1} = \left(\tfrac{q}{p}\right)^{N-1} f_1
\end{aligned}$$

Adding the first $i-1$ equations gives

$$f_i - f_1 = \left[\frac{q}{p} + \left(\frac{q}{p}\right)^2 + \ldots + \left(\frac{q}{p}\right)^{i-1}\right] f_1, \qquad\qquad 2 \leq i \leq N$$

or

$$f_i = f_1 \left[1 + \frac{q}{p} + \ldots + \left(\frac{q}{p}\right)^{i-1}\right].$$

Hence, if $q = p$

$$f_i = if_1,\ 2 \leq i \leq N,$$

and if $q \neq p$

$$f_i = f_1 \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \frac{q}{p}},\ 2 \leq i \leq N\ . \tag{12}$$

Using $f_N = 1$, we determine f_1,

$$f_1 = \begin{cases} \frac{1}{N} & \text{if } q = p \\ \frac{1-\frac{q}{p}}{1-\left(\frac{q}{p}\right)^N} & \text{if } q \neq p \end{cases} . \tag{13}$$

Substituting (13) into (12) gives

$$f_i = f_{iN} = \begin{cases} \frac{i}{N} & \text{if } q = p \\ \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & \text{if } q \neq p \end{cases} ,$$

and thus the probability of ruin is

$$f_{i0} = \begin{cases} \frac{N-i}{N} & \text{if } q = p = 0.5 \\ \frac{\left(\frac{q}{p}\right)^i-\left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)^N} & \text{if } q \neq p \end{cases}$$

As $N \to \infty$

$$f_{i0} = \begin{cases} \left(\frac{q}{p}\right)^i & \text{if } p > 0.5 \\ 1 & \text{if } p \leq 0.5 \end{cases} .$$

Thus if $p > 0.5$ there is a positive probability of the gambler's fortune becoming infinite.

The fate of a gambler

i	N	p	Prob. of ruin $(1 - f_i)$	Mean gain	Mean duration of game
9	10	.50	0.1	0	9
9	10	.45	0.21	-1.1	11
99	100	.45	0.182	-17.2	171.8
90	100	.50	0.1	0	900
90	100	.45	0.866	-76.6	756.6
90	100	.40	0.983	-88.3	441.3

7.2 Expected times to absorption

As before, let X be a MC with state space $S = R_1 \cup R_2 ... \cup R_k \cup T$, where $R_1, R_2, ..., R_k$ are recurrent classes and T is set of transient states. Let N be the number of transitions until absorption into a recurrent class and $N(j)$ the number of times X visits state j. Then

$$N = 1 + \sum_{j \in T} N(j).$$

and for $i \in T$, let

$$m_i \equiv E_i(N) = E(N|X_0 = i)$$

Proposition 15 *The mean absorption times m_i, $i \in T$ can be obtained as the solution of the system of linear equations:*

$$m_i = 1 + \sum_{k \in T} p_{ik} m_k, i \in T. \tag{14}$$

Proof. By the law of total expectation

$$\begin{aligned} m_i &= E_i(N) = \sum_{k \in S} E_i(N|X_1 = k) p_{ik} \\ &= \sum_{k \in T} E_i(N|X_1 = k) p_{ik} + \sum_{k \in \overline{T}} E_i(N|X_1 = k) p_{ik} \\ &= \sum_{k \in T} [1 + E(N|X_0 = k)] p_{ik} + \sum_{k \in \overline{T}} p_{ik} \\ &= 1 + \sum_{k \in T} p_{ik} m_k. \end{aligned}$$

where $\overline{T}$ denotes the complement of set T. ■

Example 7.4 Consider a Markov Chain $(X_n, n \geq 0)$ on $S = (1, 2, 3, 4, 5)$ with the transition probability matrix

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

a. Classify the states

b. Compute all absorption probabilities $f_{ij}, i \in T, j \in R$.

c. Compute the expected time to absorption when $X_0 = 3$ and when $X_0 = 4$.

Solutions

a. There are two positive recurrent classes of states: $\{1, 2\}$ and $\{5\}$. States 3 and 4 are transient.

b. We first compute f_{31} and f_{41} using equations from Proposition 14:

$$\begin{aligned} f_{31} &= p_{34} f_{41} + p_{31} + p_{32} = \frac{1}{3} f_{41} + \frac{1}{3} + 0, \\ f_{41} &= \frac{2}{3}, \end{aligned}$$

so that $f_{31} = \frac{5}{9}$ and $f_{41} = \frac{2}{3}$. You may easily verify that

$$f_{32} = f_{31} = f_{3,\{1,2\}} = \frac{5}{9},$$

$$f_{42} = f_{41} = f_{4,\{1,2\}} = \frac{2}{3}.$$

Finally we must have

$$\begin{aligned} f_{35} &= 1 - f_{3,\{1,2\}} = \frac{4}{9}, \\ f_{45} &= 1 - f_{4,\{1,2\}} = \frac{1}{3}. \end{aligned}$$

c. We use equations in (14). In our problem $T = \{3, 4\}$, so these equations take the form

$$\begin{aligned} m_3 &= 1 + p_{34} m_4 = 1 + \frac{1}{3} m_4, \\ m_4 &= 1, \end{aligned}$$

and thus the solution is $m_3 = 4/3$ and $m_4 = 1$.

Example 7.5 Gambler's ruin chain on $S = (0, 1, 2, ...N)$. Let m_i be the expected number of plays until the end of the game. Then for $1 \leq i \leq N-1$,

$$m_i = \begin{cases} i(N-i) & p = q \\ \frac{i}{q-p} - \frac{N}{q-p}\frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & p \neq q \end{cases}.$$

(We will derive this result later using martingale methods.)

Similar equations as in (14) also hold in the following situation:

Corollary 16 *Suppose X is an irreducible positive recurrent MC. Let $\mu_{ij} = E_i(T_j)$, $i, j \in S$, be the mean time for the process which starts in state i to reach state j. Then $\{\mu_{ij},\ i,\ j \in S\}$ satisfy the following system of linear equations*

$$\mu_{ij} = 1 + \sum_{k \neq j} p_{ik}\mu_{kj}.$$

$$\begin{aligned} \mu_{ij} &= E_i(T_j) = \sum_{k \in S} E_i(T_j | X_1 = k) p_{ik} \\ &= \sum_{k \in S, k \neq j} p_{ik} E_i(T_j | X_1 = k) + p_{ij} E_i(T_j | X_1 = j) \\ &= \sum_{k \in S, k \neq j} p_{ik} \{1 + E(T_j | X_0 = k)\} + p_{ij} \\ &= 1 + \sum_{k \neq j} p_{ik}\mu_{kj}. \end{aligned}$$

8 Statistical Aspects of Markov Chains

Consider evolution of the voter's preference as a stochastic process, more precisely as a time homogeneous Markov chain. Let X_k = preference of a potential voter at the $k\prime th$ interview.

Assume two possible responses $\{1\}$ = prefers a democrat, $\{2\}$ = prefers a republican and that $X = \{\ X_k,\ k \geq 0\ \}$ is a MC on $S = \{1, 2\}$. We want to estimate a one-step transition matrix P of X based on a random sample of N individuals where for the l'th individual we observe

$$\{X_0^l, X_\Delta^l, ..., X_{n\Delta}^l\}.$$

Here Δ is the time interval between the interviews, which we assume to be the unit time interval for X. Hence $P = P(\Delta)$.

Let $P[(k-1),\ k]$ be the one-step transition matrix between times $(k-1)\Delta$, and $k\Delta$, $1 \leq k \leq n$. Then the natural estimator of $P[(k-1), k]$ is a matrix:

$$\tilde{P}(k-1, k), 1 \leq k \leq n,$$

with the $(i, j)th$ entry given by

$$n_{ij}(k-1, k)/n_i(k-1),$$

where

$n_{ij}(k-1, k)$ = the number of voters in a sample who at time $(k-1)\ \Delta$ were in state i and at time $k\ \Delta$ are in state j,

and

$n_i(k-1) = \sum\limits_{j \in S} n_{ij}((k-1), k)$ = the number of voters in state i at time $(k-1)\Delta$.

In particular

$$\tilde{P}(0, 1) = \begin{pmatrix} \frac{n_{11}(0,1)}{n_1(0)} & \frac{n_{12}(0,1)}{n_1(0)} \\ \frac{n_{21}(0,1)}{n_2(0)} & \frac{n_{22}(0,1)}{n_2(0)} \end{pmatrix}$$

Now, in theory, if X is a time homogeneous MC then

$$P(k-1, k) = P, 1 \leq k \leq n,$$

where P is a common one step transition matrix.

Since we are dealing with sample data the matrices $\tilde{P}(k-1,k), 1 \leq k \leq n$, may differ even if X is a time homogeneous process. Therefore it seems reasonable to take the average of the observed matrices

$$\bar{P} = \frac{1}{n}\sum_{k=1}^{n} \tilde{P}(k-1,k),$$

as an estimate of a true transition probability matrix P. In elementwise notation we estimate the (i,j) entry by

$$\bar{P}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \frac{n_{ij}(k-1,k)}{n_i(k-1)}.$$

It turns out, that under Markov assumption, $\bar{P}$ is not the maximum likelihood estimator (mle) of P.

The mle procedure was described in Anderson and Goodman (1957). The mle $\hat{P}$ of P is the matrix with the (i,j)th entry given by

$$\widehat{P}_{ij} = \frac{\sum_{k=1}^{n} n_{ij}(k-1,k)}{\sum_{k=1}^{n} n_i(k-1)} = \frac{n_{ij}}{n_i},$$

where

$n_{ij} := \sum_{k=1}^{n} n_{ij}(k-1,k) =$ the total # of $i \to j$ transitions in the sample

and $n_i = \sum_{k=1}^{n} n_i(k-1)$ is the total # of visits to state i.

8.1 Modeling issues.

Is a time homogeneous MC model compatible with the data?

1. **Homogeneity test**: are the true one-step transition matrices the same? To answer this question we want to test the hypothesis

$$H_0: \; P(k-1,\; k) = P \qquad\qquad k = 1, 2, ...n$$

This hypothesis may be tested using χ^2 -test for homogeneity (see e.g. Anderson and Goodman (1957).

If H_0 is rejected then X is a time nonhomogeneous chain. (It still could be a Markov chain).

If H_0 is not rejected, we may assume that the process that generates the data is time homogeneous (i.e, $P((k-1),\; k) = P, k \geq 1)$, but is it a Markov chain?

2. **Test of Markov property**

$$H_0 : P(0, k) = P^k, 1 \leq k \leq n.$$

The **informal testing** of H_0, can be based on the comparisons of the observed $k-$step matrices $\tilde{P}(0, k)$ with $\widehat{P}^k$, the matrices predicted by the model, that is we reject H_0 if residual matrices:

$$\tilde{P}(0, k) - \widehat{P}^k$$

have large entries or show a pattern.

In many applications of Markov chains, in particular to social sciences data, the residual matrices show the following pattern

$$\left(\tilde{P}(0, k)\right)_{ii} > \left(\widehat{P}^k\right)_{ii}, 2 \leq k \leq n, \text{ and all } i \in S.$$

(This empirical regularity was first observed by Blumen, Kogan and McCarthy (1955). To account for this regularity one may consider the following mixture of Markov chains, called the Mover-Stayer model.

8.2 The Mover-Stayer Model

This model assumes that population is heterogenous: there are "stayers" and "movers". "Movers" evolve according to a MC with the one-step transition matrix $M = (m_{ij})$. "Stayers" never leave their initial states. In the $i\prime th$ state the proportion of stayers is given by $s_i, i \in S = \{1, 2, ...w\}$. Let $X_n =$ state at time n of an individual chosen at random from this heterogeneous population. We now derive the n-step transition matrix of $X = (X_n, n \geq 0)$.

$$\begin{aligned}
p_{ij}^{(n)} &= P(X_n = j|X_0 = i) = P(X_n = j, \text{mover}|X_0 = i) \\
+P(X_n &= j, \text{stayer}|X_0 = i) \\
&= P(X_n = j|\text{mover, } X_0 = i)P(\text{mover}|X_0 = i) \\
+P(X_n &= j|\text{stayer, } X_0 = i)P(\text{stayer}|X_0 = i) \\
&= \begin{cases} m_{ij}^{(n)}(1 - s_i) \text{ if } i \neq j \\ m_{ii}^{(n)}(1 - s_i) + s_i \text{ if } i = j, \end{cases} \qquad (15)
\end{aligned}$$

where $m_{ij}^{(n)}$ is the $(i, j)th$ entry in M^n. Let $S =$diag$(s_1, s_2, ..., s_w)$ and I is an identity matrix. Then, by (15), the $n-$step transition matrix of a mover-stayer model is

$$P^{(n)} = S + (I - S)M^n, n \geq 1,$$

so that in particular the one-step transition matrix is

$$P = S + (I - S)M.$$

The Mover-Stayer model $(X_n, n \geq 0)$ is not a Markov chain. To see this note that

$$P(X_2 = 3|X_1 = 3, X_0 = 3) = \frac{s_3 + (1 - s_3)m_{33}^2}{s_3 + (1 - s_3)m_{33}},$$

but

$$P(X_2 = 3|X_1 = 3, X_0 = 1) = m_{33}$$

which shows that past history matters. A different way of showing that $(X_n, n \geq 0)$ is not a Markov Chain is to notice, that in general

$$P^{(n)} \neq P^n,$$

that is,

$$S + (I - S)M^n \neq [S + (I - S)M]^n.$$

Recall that the Mover-Stayer model was introduced to account for the empirical regularity:

$$\left(\widetilde{P}(0,k)\right)_{ii} > \left(\widehat{P}^k\right)_{ii}, k = 1, 2, ...n,$$

which may occur for all or some states, observed in many different contexts when Markov Chain model was fitted to the data. To carry out similar comparisons for the M-S model we have first to estimate the parameters of the M-S model, that is, the matrices M and S. The maximum likelihood estimation procedure for the M-S model was described in Frydman (1984). If $\widehat{S}$ and $\widehat{M}$ are the maximum likelihood estimators of S and M respectively, then the estimated $k-$step transition matrix for the M-S model is

$$\widehat{S} + (I - \widehat{S})\widehat{M}^k, k = 1, 2, ...n.$$

In many modeling contexts the comparisons of the observed with the predicted by the M-S model diagonal entries, namely

$$\left(\widetilde{P}(0,k)\right)_{ii} - \left(\widehat{S} + (I - \widehat{S})\widehat{M}^k\right)_{ii} \approx 0, k = 1, 2, ...n, \text{ and every state } i,$$

show that the M-S model accounts for much of the empirical regularity, thus fits the data better than a Markov chain.

9 Applications of Markov chains and the Mover-Stayer model to modelling empirical processes.

9.1 Modeling payment behavior of credit card accounts.

In "**Testing the adequacy of Markov chain and Mover-Stayer models as representations of credit behavior**", Frydman, Kallberg and Kao (1985) applied the methodology described above and more formal tests to modeling the behavior of credit card accounts of a retail department store. The general purpose of this analysis was to be able to forecast cash flows generated by credit accounts. Let

$$X_n = \text{the state of an account at the start of the n'th month.}$$

The following state space $S = (P, C, D)$, was considered where

$$\begin{aligned} P &= \text{(paid up state): the outstanding balance} < \$1, \\ C &= \text{(current state): an outstanding balance} \geq \$1 \\ &\quad \text{ with the last payment at least the minimum required.} \\ D &= \text{(overdue state): the last payment less than the} \\ &\quad \text{ minimum required.} \end{aligned}$$

Data consisted of a random sample of 200 active accounts observed over 16 months period. Three models were fitted to the data: time homogeneous Markov chain, time nonhomogeneous Markov chain, and the M-S model. The estimated models and residual matrices from each model are shown below.

It is seen from the residual matrices, but also from more formal likelihood ratio tests carried out in the paper, that incorporating heterogeneity into the model of payment behavior is more important than incorporating nonstationarity.

The analysis suggests that a mover-stayer model is a better description than a Markov chain of the payment behavior of credit card holders. The further development on modeling payment behavior of credit cards holders using the mover-stayer model can be found in **Behavioral models of credit card usage** by Till&Hand (2003).

9.2 Modeling behavior of corporate credit ratings.

A rating by a credit rating agency represents an overall assessment of a firm's (an obligor's) creditworthiness. The following rating scale from the best to the worst is used in the literature on credit ratings: $\mathcal{AAA}$, $\mathcal{AA}$, $\mathcal{A}$, $\mathcal{BBB}$, $\mathcal{BB}$, $\mathcal{B}$, and $\mathcal{CCC}$ (rating agencies use a more detailed scale). Credit ratings are used, for example to price the corporate bonds with default risk. Until relatively recently, it has been assumed in the extant literature that the ratings process is time homogeneous Markov. This assumption stipulates that all information about credit ratings is contained in the credit ratings transition matrix. In particular rating agencies publish transition matrices. An example of the yearly transition matrix of a Markov chain estimated using S&P US corporate obligors histories over the period 1981-2002 is shown below. Note that in addition to the credit ratings it has two states: $\mathcal{D}$ and NR.

Recently non-Markovian aspects of credit ratings migration have been investigated including the influence of macroeconomic variables on the hazard of default and other credit ratings transitions, see Figlewski, Frydman and Liang (2012), (**Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions.)**

A different non-Markovian model of credit ratings migration is based on an extension of the mover-stayer model is discussed in Frydman and Schuermann (2008), (**Credit Rating Dynamics and Markov Mixture Models**) who proposed a mixture of two Markov chains as an alternative to a single Markov chain for modeling the evolution of credit ratings over time.

9.3 Repayment of car loans

The most recent extension of the mover-stayer model is to an analysis of repayment behavior of customers who took car loans from a bank. (Frydman and Matuszyk, work in progress, 2014). This extension involves modeling the probability of being a stayer using the characteristics of the borrowers.

10 Appendix A

To prove Theorem 3 we need the following lemma

Lemma 17 $P_i(N(j) \geq m+1) = f_{ij}(f_{jj})^m,\ m \geq 0.$

Proof.

$$\begin{aligned} P_i(N(j) &\geq m+1) \\ &= \sum_{n=1}^{\infty} P_i(T_j = n \text{ and } X_s = j \text{ for at least } m \text{ values of } s > n) \\ &= \sum_{n=1}^{\infty} P_i(X_s = j \text{ for at least } m \text{ values of } s > n | T_j = n) P_i(T_j = n) \end{aligned}$$

which by Markov property is equal to

$$\sum_{n=1}^{\infty} P_i(X_s = j \text{ for at least } m \text{ values of } s > n | X_n = j) P_i(T_j = n)$$

and by time homogeneity is further equal to

$$\begin{aligned}\sum_{n=1}^{\infty} P_i(X_s &= j \text{ for at least } m \text{ values of } s > 0 | X_0 = j) P_i(T_j = n) \\ &= P_j N(j) \geq m) \sum_{n=1}^{\infty} P_i(T_j = n) = P_j(N(j) \geq m) f_{ij}.\end{aligned}$$

Hence we have shown

$$P_i(N(j) \geq m+1) = f_{ij} P_j(N(j) \geq m), m \geq 0.$$

Consider this expression for m = 0,1,2,3,...

$$m = 0,\ P_i\,(N(j) \geq 1) = f_{ij}, \text{ in particular } P_j(N(j) \geq 1) = f_{jj}$$

$$m = 1,\ P_i\,(N(j) \geq 2) = f_{ij} P_j(N(j) \geq 1) = f_{ij} f_{jj}$$
$$m = 2,\ P_i\,(N(j) \geq 3) = f_{ij} P_j(N(j) \geq 2) = f_{ij} f_{jj}^2$$

The induction argument now gives the statement of the lemma:

$$P_i(N(j) \geq m+1) = f_{ij}(f_{jj})^m, m \geq 0.$$

■

The following theorem subsumes Theorem 3.

Theorem 18 *(a) If j is a transient state then*

$$P_i(N(j) < \infty) = 1, \text{ and } E_i N(j) = \frac{f_{ij}}{1 - f_{jj}} < \infty, i \in S.$$

(b) If j is a recurrent state, then

$$P_j(N(j) = \infty) = 1 \Rightarrow E_j N(j) = \infty.$$

Also, for any $i \in S$, if $f_{ij} = 0$ then $E_i N(j) = 0$, whereas if $f_{ij} > 0$ then $E_i N(j) = \infty$.

Proof. (a) By definition of a transient state ($f_{jj} < 1$) we obtain

$$P_i(N(j) = \infty) = \lim_{m \uparrow \infty} P_i(N(j) \geq m) = \lim_{m \uparrow \infty} f_{ij} f_{jj}^{m-1} = 0$$

and hence $P_i(N(j) < \infty) = 1 - P_i(N(j) = \infty) = 1$. Next compute

$$\begin{aligned} E_i[N(j)] &= \sum_{m=1}^{\infty} mP(N(j) = m) = \sum_{m=1}^{\infty} P_i(N(j) \geq m) \\ &= f_{ij} \sum_{m=1}^{\infty} f_{jj}^{m-1} = \frac{f_{ij}}{1 - f_{jj}} < \infty. \end{aligned}$$

(b) By definition of a recurrent state ($f_{jj} = 1$)

$$\begin{aligned} P_j(N(j) &= \infty) = \lim_{m \to \infty} f_{jj}^m = 1, \\ P_i(N(j) &= \infty) = \lim_{m \to \infty} f_{ij} f_{jj}^{m-1} = f_{ij}. \end{aligned}$$

Now $f_{ij} > 0$,implies that $E_i(N(j)) = \infty$, and $f_{ij} = 0$, implies that $N(j)|(X_0 = i)$= 0 with probability one, so that indeed $E_i N(j) = 0$. ■

11 Appendix B

Proof of Proposition 8: Suppose i is aperiodic and positive recurrent and $i \leftrightarrow j$. We want to show that j is positive recurrent. Now $i \leftrightarrow j$ implies that

$$p_{ij}^{(n)} > 0, p_{ji}^{(m)} > 0$$

for some $n > 0$ and for some $m > 0$. By applying Chapman-Kolmogorov equations twice, we get

$$p_{jj}^{(n+k+m)} \geq p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)}$$

and hence

$$\lim_{k \to \infty} p_{jj}^{(n+k+m)} \geq p_{ji}^{(m)} (\lim_{k \to \infty} p_{ii}^{(k)}) p_{ij}^{(n)} > 0$$

where ">0" follows from Theorem **??** and the fact that $\mu_{ii} < \infty$.■

12 Appendix C

Proof of Theorem 12

(a) By (4) we know that $\lim p_{ij}^{(n)} = 0$. Suppose that there exists a stationary distribution α, that is, $\alpha_j = \sum_{k\in S} \alpha_k p_{kj}^{(n)}$, for every $j \in S,\ n \geq 1$. Then by the Dominated Convergence Theorem (stated below)

$$\alpha_j = \lim_{n\uparrow\infty} \sum_{k\in S} \alpha_k p_{kj}^{(n)} = \sum_{k\in S} \alpha_k \lim_{n\uparrow\infty} p_{kj}^{(n)} = 0, \qquad \text{for every } j$$

which is a contradiction. Hence there is no stationary distribution.■

(**b**) We give the proof only for a finite chain. For a general proof see (Ross (B)), Theorem 4.3.3, page 175). Consider

$$p_{ij}^{(n+1)} = \sum_{k\in S} p_{ik}^{(n)} p_{kj}^{(1)}$$

and take the limits of both sides as $n \uparrow \infty$, i.e.,

$$\lim_{n\to\infty} p_{ij}^{(n+1)} = \sum_{k\in S} \lim_{n\to\infty} p_{ik}^{(n)} p_{kj}.$$

Then by Theorem 7 we obtain

$$\pi_j = \sum_{k\in S} \pi_k p_{kj}, \quad j \in S$$

or in matrix notation: $\pi = \pi P$, where π is a probability vector, i.e., $\sum_{j\in S} \pi_j = 1$, and $\pi_j > 0$. It remains to show that π is a unique stationary distribution of $(X_n,\ n \geq 0)$. Suppose that α is a different stationary distribution of $(X_n,\ n \geq 0)$ so that $\alpha = \alpha P$. Then also $\alpha = \alpha P^n$, or in elementwise notation

$$\alpha_j = P(X_n = j) = \sum_{i\in s} \alpha_i p_{ij}^{(n)},$$

and letting $n \to \infty$ we obtain

$$\alpha_j = \sum_{i\in S} \alpha_i \lim_{n\uparrow\infty} p_{ij}^{(n)} = \sum_{i\in S} \alpha_i \pi_j = \pi_j \sum_{i\in S} \alpha_i = \pi_j.$$

Hence $\alpha_j = \pi_j$, showing that π is indeed a unique stationary distribution of $(X_n,\ n \geq 0)$ ■.

The following theorem states conditions under which you can interchange infinite summation and limit operations.

Theorem 19 *(Dominated Convergence Theorem) If* $a_{i,n}$ *and* $c_i, i, n = 1, 2,$ *are sequences of numbers for which*

(a) $|a_{i,n}| \le c_i, i, n = 1, 2, \ldots$ *and* $\sum\limits_{i=1}^{\infty} c_i < \infty$

and

(b) $\lim_{n\uparrow\infty} a_{i,n} = a_i$ *for* $i = 1, 2, \ldots$

then

$$\lim_{n\uparrow\infty} \sum_{i=1}^{\infty} a_{i,n} = \sum_{i=1}^{\infty} \lim_{n\uparrow\infty} a_{i,n} = \sum_{i=1}^{\infty} a_i.$$

References

Anderson and Goodman (1957), Statistical Inference about Markov Chains, *Annals of Mathematical Statistics, 28.*

Blumen, Kogan and McCarthy (1955), The industrial mobility of Labor as a Probability Process", Cornell University Press.

Frydman (1984) Maximum likelihood estimation in the Mover-Stayer model, *Journal of the American Statistical Association,* 79, 632 637.

Frydman, Kallberg and Kao (1985) Testing the adequacy of Markov chain and Mover-Stayer models as representations of credit behavior, *Operations Research*, 33, 1203-1214..

Frydman and Schuermann (2008). Credit Rating Dynamics and Markov Mixture Models, *Journal of Banking and Finance*, 32,1062-1075.

Frydman, Matuszyk (2014), Mover-Stayer Model with Covariates; an Application to an Analysis of Car Loans, work in progress.

Till and Hand (2003) Behavioral models of credit card usage, *Journal of Applied Statistics,* vol 30, 1201-1220.