Markov Chains

Consider the sequence

$$X_0, X_1, X_2, ...$$

of random variables taking values on a finite or countable infinite set of integer values:

$$D \quad = \quad \{0, 1, 2, ..., k\} \quad \text{(finite case)}$$

or

$$D \quad = \quad \{0, 1, 2, ...\} \quad \text{(infinite, countable case)}$$

Definition (Markov Chain): The sequence $X_0, X_1, X_2, ...$ is a *stationary Markov chain* if

$$P\left(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i\right) \quad =$$

$$= \quad P\left(X_{n+1} = j \mid X_n = i\right) = P\left(X_1 = j \mid X_0 = i\right)$$

$$= \quad P_{ij},$$

for all $i_0, i_1, \cdots, i_{n-1}, i$ in D and for all $n \geq 0$.

Definition (Transition Matrix): The transition matrix P is a squared stochastic matrix with generic element $P_{ij} = P(X_1 = j \mid X_0 = i)$. That is

$$P = (P_{ij})$$

where the P_{ij} satisfy the condition

$$\sum_{j \in D} P_{ij} = 1, \quad \text{for all } i$$

Definition (Initial Probability). The initial probability is a vector (or sequence in the infinite case) $\mathbf{p}$ which entries are the initial probabilities for the chain, that is,

$$p_i = P(X_0 = i), \quad \text{for all } i \in D.$$

Example 1 (**A simple weather Markov Chain**). Suppose that the weather condition tomorrow (Sunny=0, Cloudy=1 and Rainy/Snowy=2) only depends on the weather condition today, with transition probabilities

$$P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.2 & 0.4 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

Example 2 (**A random walk with absorbing barrier**). Suppose that

$$
\begin{aligned}
P_{ij} &= \begin{cases} \alpha & j = i \\ \\ 1-\alpha & j = i+1 \end{cases}, \qquad i = 0,1,2,3 \\
P_{4j} &= \begin{cases} 1 & j = 4 \\ \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

where $k = 4$ is an absorbing state. For instance, taking $\alpha = 0.4$

$$
P = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
$$

Example 3 (A simple gene model) The simplest type of inheritance occurs when a trait is governed by a gene with two *alleles*, say G and g.

Genotype: An individual may have a GG combination or Gg (which is genetically the same as gG) or gg.

Dominance: Very often the GG and Gg types are indistinguishable in appearance, and then we say that the G gene dominates the g gene. An individual is called *dominant* if it is GG, recessive if it is gg, and *hybrid* if it is Gg.

Mating: In the mating of two individuals, the offspring inherits one allele of the pair from each parent.

Basic assumption of genetics: the alleles are selected at random, independently of each other. This assumption determines the probability of occurrence of each type of offspring.

Offsprings: The offsprings of two purely dominant parents must be dominant, of two recessive parents must be recessive, and of one dominant and one recessive parent must be hybrid.

In the mating of a dominant and a hybrid, each offspring must get a G allele from the former and has an equal chance of getting either G or g from the latter. Hence there is an equal probability, 1/2, for getting a dominant or a hybrid offspring.

In the mating of a recessive and a hybrid, there is an even chance, 1/2, for getting either a recessive or a hybrid.

In the mating of two hybrids, the offspring has an equal chance of getting G or g from each parent. Hence the probabilities are

$$\begin{array}{ll} 1/4 & \text{for } GG \\ 1/2 & \text{for } Gg \\ 1/4 & \text{for } gg \end{array}$$

Continuous Mating: We start with an individual of known genetic character and mate it with a hybrid.

We assume that there is at least one offspring...

The oldest offspring is mated with a hybrid and this process is repeated through a number of generations.

The genetic type of the chosen offspring in successive generations can be represented by a Markov chain with states GG, Gg, and gg. The corresponding transition matrix is:

	GG	Gg	gg
GG	0.5	0.5	0
Gg	0.25	0.5	0.25
gg	0	0.5	0.5

As an exercise, you can derive the transition matrix for a process where we mate the oldest offspring with a dominant.

Example 4 (Another gene model) As another exercise (a bit more involved this time) derive the transition matrix for the following mating process: we start with two specimens of opposite sex, mate them, select two of their offspring of opposite sex, and mate those, and so forth. Here a state is determined by a pair of specimens. Hence, the states of this process are:

$$\begin{aligned} s1 &= (GG; GG), \\ s2 &= (GG; Gg), \\ s3 &= (GG; gg), \\ s4 &= (Gg; Gg), \\ s5 &= (Gg; gg), \\ s6 &= (gg; gg). \end{aligned}$$

We illustrate the calculation of the corresponding transition probabilities in terms of the state $s2$:

	$GG\,(1/2)$	$Gg\,(1/2)$
$GG\,(1/2)$	$GG - GG\,(1/4)$	$GG - Gg\,(1/4)$
$Gg\,(1/2)$	$Gg - GG\,(1/4)$	$Gg - Gg\,(1/4)$

$$\rightarrow \quad s1 = (GG, GG)\,(1/4) \quad s2 = (GG, Gg)\ (1/2) \quad s4 = (Gg, Gg)\ (1/4)$$

If we start with $s3 = (GG; gg)$ then the next state is $s4 = (Gg; Gg)$ with probability one. As an exercise verify that the transition probabilities are as displayed in the table below:

	$s1$	$s2$	$s3$	$s4$	$s5$	$s6$
$s1$	1	0	0	0	0	0
$s2$	0.25	0.5	0	0.25	0	0
$s3$	0	0	0	1	0	0
$s4$	0.062	0.25	0.125	0.25	0.25	0.062
$s5$	0	0	0	0.25	0.5	0.25
$s6$	0	0	0	0	0	1

We will see that in this case the limiting state will be either $s1$ or $s6$. So this method could be used to obtain "pure breed" individuals.

Result 1: Let

$$A = \{(X_0, X_1, X_2, \cdots, X_{n-1}) \in B\}$$

where $B \subset D^n$. Then

$$P(X_{n+1} = j \mid A, X_n = i) = P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i).$$

Proof:

$$P\left(X_{n+1}=j|A,X_n=i\right)=$$

$$= \frac{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_{n+1}=j,X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}$$

$$= \frac{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_{n+1}=j \mid X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right) P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}$$

$$= \frac{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_{n+1}=j \mid X_n=i\right) P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)} \qquad \text{by the Markov Property}$$

$$= P\left(X_{n+1}=j \mid X_n=i\right) \frac{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}{\sum_{\mathbf{x}\in\mathbf{B}} P\left(X_n=i,\mathbf{X}_{n-1}=\mathbf{x}\right)}$$

$$= P\left(X_{n+1}=j \mid X_n=i\right) = P\left(X_1=j \mid X_0=i\right).$$

Result 2:

$$P\left(X_{n+1}=j_1,\cdots,X_{n+m}=j_m \mid X_n=i\right) \quad = \quad P_{i\,j_1}P_{j_1j_2}\cdots P_{j_{m-1}j_m}$$

Proof:

$$P\left(X_{n+1}=j_1,\cdots,X_{n+m}=j_m \mid X_n=i\right)=$$

$$= \quad P\left(X_{n+1}=j_1 \mid X_n=i\right)\; P\left(X_{n+2}=j_2 \mid X_{n+1}=j_1,X_n=i\right)\times$$

$$P\left(X_{n+3}=j_m \mid X_{n+2}=j_2,X_{n+1}=j_1,X_n=i\right)\times\cdots\times$$

$$P\left(X_{n+m}=j_m \mid X_{n+m-1}=j_{m-1},\cdots,X_{n+1}=j_1,X_n=i\right)$$

$$= \quad P_{i\,j_1}P_{j_1j_2}\cdots P_{j_{m-1}j_m} \qquad \text{by Markov Property}$$

Result 3: Let

$$A = \{(X_0, X_1, X_2, \cdots, X_{n-1}) \in B\} \quad \text{(the past)}$$

and

$$D_{n,m} = \{(X_{n+1}, X_{n+2}, ..., X_{n+m}) \in C\}, \quad C \subset D^m \quad \text{(the near future)}$$

Then

$$P(D_{n,m} \mid A, X_n = i) = P(D_{n,m} \mid X_n = i) = P(D_{0,m} \mid X_0 = i).$$

Proof:

$$P(D_{n,m} \mid A, X_n = i) =$$

$$= \sum_{(x_1,\cdots,x_m)\in C} P\left(X_{n+1}=x_1,\cdots,X_{n+m}=x_m \mid A, X_n=i\right)$$

$$= \sum_{(x_1,\cdots,x_m)\in C} P\left(X_{n+1}=x_1 \mid A, X_n=i\right)\times\cdots\times P\left(X_{n+m}=x_m \mid A, X_n=i,\cdots,X_{n+m-1}=x_{m-1}\right)$$

$$= \sum_{(x_1,\cdots,x_m)\in C} P\left(X_{n+1}=x_1 \mid X_n=i\right)\times\cdots\times P\left(X_{n+m}=x_m \mid X_{n+m-1}=x_{m-1}\right) \quad \text{by Result 1}$$

$$= \sum_{(x_1,\cdots,x_m)\in C} P\left(X_{n+1}=x_1, X_{n+2}=x_2,\cdots,X_{n+m}=x_m \mid X_n=i\right) \quad \text{by Result 2}$$

$$= P\left(D_{n,m} \mid X_n=i\right)$$

$$= \sum_{(x_1,\cdots,x_m)\in C} P\left(X_1=x_1, X_2=x_2,\cdots,X_m=x_m \mid X_0=i\right) \quad \text{by stationarity}$$

$$= P\left(D_{0,m} \mid X_0=i\right)$$

Result 4: More generally, let

$$A = \{(X_0, X_1, X_2, \cdots, X_{n-1}) \in B\}$$

and

$$D_n = \{(X_{n+1}, X_{n+m}, ...) \in C\}, \quad C \subset D^{\infty}.$$

Then

$$P\left(D_n \mid A, X_n=i\right) = P\left(D_n \mid X_n=i\right) = P\left(D_0 \mid X_0=i\right).$$

Proof: to prove this result one can use a monotone convergence argument, with sets of the form

$$\tilde{D}_{n,m} = \{(x_{n+1}, x_{n+2}, \cdots, x_{n+m}) \times D^{\infty} : (x_{n+1}, x_{n+2}, \cdots, x_{n+m}, \cdots) \in C\}, \qquad C \subset D^{\infty},$$

which clearly decrease toward C. That is,

$$\lim_{m\to\infty} P\left(\tilde{D}_{n,m} \mid A, X_n = i\right) = P\left(\lim_{m\to\infty} \tilde{D}_{n,m} \mid A, X_n = i\right) = P\left(D_n \mid A, X_n = i\right)$$

On the other hand, by Result 3,

$$\begin{aligned}
\lim_{m\to\infty} P\left(\tilde{D}_{n,m} \mid A, X_n = i\right) &= \lim_{m\to\infty} P\left(\tilde{D}_{n,m} \mid X_n = i\right) \\
&= P\left(\lim_{m\to\infty} \tilde{D}_{n,m} \mid X_n = i\right) = P\left(D_n \mid X_n = i\right)
\end{aligned}$$

Result 5: Giving the present, the past and future are independent. More precisely

$$P\left(A \cap D_n \mid X_n = i\right) = P\left(A \mid X_n = i\right) P\left(D_n \mid X_n = i\right),$$

where A and D_n are as in **Result 1** and **Result 4**.

Proof

$$\begin{aligned}
P\left(A \cap D_n \mid X_n = i\right) &= P\left(A \mid X_n = i\right) P\left(D_n \mid A, X_n = i\right) \\
&= P\left(A \mid X_n = i\right) P\left(D_n \mid X_n = i\right), \quad \text{by Result 4.}
\end{aligned}$$

Transition Probabilities (and Transition Matrix)

$$P_{ij}^{(m)} \quad = \quad P\left(X_{n+m}=j \mid X_n=i\right)=P\left(X_m=j \mid X_0=i\right)$$

Obviously, $P_{ij}^{(1)} = P_{ij}$. These probabilities can be arranged in the (possibly infinite) matrix

$$P^{(m)} = \left(P_{ij}^{(m)}\right)$$

Notice that i^{th} row of $P^{(m)}$ gives the probabilities of being in state j at time m, given that the chain was at state i at time zero. Since the chain must be in some state at time m, it is clear that

$$\sum_j P_{ij}^{(m)} = 1, \quad \text{for all } i.$$

Result 6 (Chapman-Kolmogorov Equations): For all $m \geq 1$,

$$P^{(m)} = P^m \qquad \text{(matrix multiplication in the RHS)}$$

Proof By induction. Clearly the property holds for $m = 1$. Suppose, now that $P^{(m)} = P^m$ for some $m > 1$. We will show that $P^{(m+1)} = P^{m+1}$. In fact,

$$\begin{aligned}
P_{ij}^{(m+1)} &= P\left(X_{m+1}=j \mid X_0=i\right) \\
&= \sum_k P\left(X_{m+1}=j, X_m=k \mid X_0=i\right) \\
&= \sum_k P\left(X_m=k \mid X_0=i\right) P\left(X_{m+1}=j \mid X_0=i, X_m=k\right) \\
&= \sum_k P\left(X_m=k \mid X_0=i\right) P\left(X_{m+1}=j \mid X_m=k\right) \quad \text{by Markov Property} \\
&= \sum_k P_{ik}^{(m)} P_{kj}
\end{aligned}$$

Therefore

$$P^{(m+1)} = P^m P = P^{m+1}.$$

A simple consequence of this result is that

$$P^{(m+n)} = P^m P^n, \qquad \text{for all } m, n$$

That is,

$$P_{ij}^{(m+n)} \quad = \quad \sum_k P_{ik}^{(m)} P_{kj}^{(n)}, \qquad \text{for all } i,\ j,\ m,\ n.$$

Transient and Recurrent States

Let

$$T_k = \min\{i \geq 1 : X_i = k\} \qquad \text{(return time for state } k\text{)}$$

Obviously,

$$X_{T_k} = k \quad \text{and} \quad X_1 \neq k, \cdots, X_{T_k - 1} \neq k$$

Also define

$$\begin{aligned} N_k &= \sum_{i \geq 0} I(X_i = k) \qquad \text{(number of visits to state } k\text{)} \\ &= f(X_0, X_1, X_2, \cdots) \end{aligned}$$

Since we will assume that the initial state of the chain is k - that is $X_0 = k$ - we have that $N_k \geq 1$.

Let

$$\begin{aligned} f_{kk} &= P(T_k < \infty \mid X_0 = k) = P(N_k > 1 \mid X_0 = k) = 1 - P(N_k = 1 \mid X_0 = k) \\ &= \text{“Probability that the chain will eventually return to state k”} \end{aligned}$$

The state k is called **recurrent** if $f_{kk} = 1$. On the other hand, if $f_{kk} < 1$ the the state k is called **transient**.

Recurrent	Transient
$f_{kk} = 1$	$f_{kk} < 1$

Clearly,

$$P(N_k = 1 \mid X_0 = k) = P(T_k = \infty \mid X_0 = k) = 1 - f_{kk}$$

More generally, we have the following result:

Result 7:

(a) For all $r > 1$, we have

$$P(N_k = r \mid X_0 = k) = f_{kk} P(N_k = r - 1 \mid X_0 = k)$$

(b) For all $r \geq 1$,

$$P(N_k = r \mid X_0 = k) = f_{kk}^{\,r-1}(1 - f_{kk})$$

That is, N_r is a Geometric random variable with

$$\begin{aligned} p &= 1 - f_{kk} \\ &= \text{“Probability that the chain will never return to state k”} \end{aligned}$$

and

$$E(N_k) = \frac{1}{1 - f_{kk}}$$

Proof
(a) Since we are assuming that $X_0 = k$, we have that

$$N_k = I(X_0 = k) + \sum_{i \geq 1} I(X_i = k) = 1 + \sum_{i \geq 1} I(X_i = k)$$

Now, if $T_k = \min\{i \geq 1 \ : \ X_i = k\} = \infty$, then $N_k = 1$. On the other hand, if $T_k < \infty$, then

$$N_k = 1 + \sum_{i \geq T_k} I(X_i = k) = 1 + f(X_{T_k}, X_{T_k+1}, X_{T_k+2}, \cdots)$$

Let $r > 1$. Then,

$$P(N_k = r \mid X_0 = k) = P(N_k = r \ , T_k < \infty \mid X_0 = k)$$

because $T_k = \infty \Rightarrow N_k = 1 < r$. Moreover,

$$\begin{aligned} P(N_k = r \mid X_0 = k) &= P(f(X_0, X_1, X_2, \cdots) = r \ , T_k < \infty \mid X_0 = k) \\ &= \sum_{j \geq 1} P(f(X_j, X_{j+1}, X_{j+2}, \cdots) = r - 1 \ , T_k = j \mid X_0 = k) \end{aligned}$$

$$= \sum_{j \geq 1} P(f(X_j, X_{j+1}, X_{j+2}, \cdots) = r - 1 \mid T_k = j, \ X_0 = k) P(T_k = j \mid X_0 = k)$$

We now notice that

$$P(f(X_j, X_{j+1}, X_{j+2}, \cdots) = r - 1 \mid T_k = 1, \ X_0 = k) =$$

$$= P\left(f\left(X_1, X_2, X_3, \cdots\right) = r-1 \mid X_0 = k, X_1 = k\right)$$

$$= P\left(f\left(X_1, X_2, X_3, \cdots\right) = r-1 \mid X_1 = k\right)$$

$$= P\left(f\left(X_0, X_1, X_2, \cdots\right) = r-1 \mid X_0 = k\right)$$

and for any $j > 1$,

$$P\left(f\left(X_j, X_{j+1}, X_{j+2}, \cdots\right) = r-1 \mid T_k = j,\ X_0 = k\right) =$$

$$= P\left(f\left(X_j, X_{j+1}, X_{j+2}, \cdots\right) = r-1 \mid X_0 = k, X_1 \neq k, \cdots, X_{j-1} \neq k, X_j = k\right)$$

$$= P\left(f\left(X_j, X_{j+1}, X_{j+2}, \cdots\right) = r-1 \mid X_j = k\right)$$

$$= P\left(f\left(X_0, X_1, X_2, \cdots\right) = r-1 \mid X_0 = k\right)$$

Therefore

$$\begin{gathered} P\left(N_k = r \mid X_0 = k\right) = \\ P\left(f\left(X_0, X_1, X_2, \cdots\right) = r-1 \mid X_0 = k\right) \sum_{j \geq 1} P\left(T_k = j \mid X_0 = k\right) = \\ = P\left(f\left(X_0, X_1, X_2, \cdots\right) = r-1 \mid X_0 = k\right) P\left(T_k < \infty \mid X_0 = k\right) = \\ = P\left(N_k = r-1 \mid X_0 = k\right) f_{kk}. \end{gathered}$$

This proves Part (a).

(b) Suppose that $f_{kk} < 1$. By Part (a),

$$P\left(N_k = r \mid X_0 = k\right) = P\left(N_k = r-1 \mid X_0 = k\right) f_{kk}$$

Therefore,

$$\begin{aligned} P\left(N_k = r \mid X_0 = k\right) &= P\left(N_k = r-1 \mid X_0 = k\right) f_{kk} \\ &= P\left(N_k = r-2 \mid X_0 = k\right) f_{kk}^2 \\ &= P\left(N_k = r-3 \mid X_0 = k\right) f_{kk}^3 \\ &\ \ \vdots \\ &= P\left(N_k = 1 \mid X_0 = k\right) f_{kk}^{r-1} \\ &= P\left(T_k = \infty \mid X_0 = k\right) f_{kk}^{r-1} \\ &= \left(1 - f_{kk}\right) f_{kk}^{r-1} \end{aligned}$$

NOTE 1: If the state is **recurrent**, that is, if $f_{kk} = 1$, then by **Result 7 (a)** we have

$$P(N_k = r \mid X_0 = k) = P(N_k = r - 1 \mid X_0 = k) \quad \text{for all } r > 1,$$

and this implies that

$$P(N_k = r \mid X_0 = k) = 0 \text{ for all } r > 1.$$

That is, the chain visits a recurrent state k an infinite number of times.

NOTE 2: If the state is transient ($f_{kk} < 1$) then the number of visits to state k is a Geometric distribution with "probability of success" $p = 1 - f_{kk}$. Therefore, the expected number of visits to a transient estate k is

$$E(N_k | X_0 = k) = \frac{1}{1 - f_{kk}} < \infty \Leftrightarrow f_{kk} < 1.$$

NOTE 3: Notice that

$$\begin{aligned}
N_k &= \sum_{i \geq 0} I(X_i = k) \\
E(N_k \mid X_0 = k) &= \sum_{i \geq 0} E\{I(X_i = k) \mid X_0 = k\} \\
&= \sum_{i \geq 0} P(X_i = k | X_0 = k) \\
&= \sum_{i \geq 0} P_{kk}^{(i)} \\
&= 1 + \sum_{i > 0} P_{kk}^{(i)}
\end{aligned}$$

From Notes 2 and 3,

$$\sum_{i \geq 0} P_{kk}^{(i)} < \infty \Leftrightarrow f_{kk} < 1 \Leftrightarrow k \text{ is transient.}$$

Remark: If the state k is transient, then it will not occur infinitely often. Let

$$A_i = \{X_i = k\}$$

and let Q be the conditional probability given $X_0 = k$. Then

$$Q(A_i) = P_{kk}^{(i)}$$

By the Borel-Cantelli Lemma:

$$\sum^{\infty} Q(A_i) \quad < \quad \infty \Rightarrow Q(A_i \ \text{ i.o.}) = 0$$

COMMUNICATION OF STATES

Notation: We will write

$$P_{ij}^{(n)} = P\left(X_n = j | X_0 = i\right), \quad \text{for all } i, j \in D \ \text{ and } n \geq 1$$

Moreover, set

$$P_{ij}^{(0)} = P\left(X_0 = j | X_0 = i\right).$$

Then,

$$P_{ij}^{(0)} = \begin{cases} 1 & i = j \\ \\ 0 & \text{otherwise} \end{cases}$$

Definition (Communicating states): States i and j communicate (and write $i \leftrightarrow j$) if there exist integers $r \geq 0, s \geq 0$ such that $P_{ij}^{(r)} > 0$ and $P_{ji}^{(s)} > 0$.

Notice that "**communicate**" is an equivalence relation , that is,
(i) **reflexive** ($i \leftrightarrow i$) because $P_{ii}^{(0)} = 1 > 0$,
(ii) **symmetric** ($i \leftrightarrow j$ iff $j \leftrightarrow i$, by definition) and
(iii) **transitive** ($i \leftrightarrow j$ and $j \leftrightarrow k$ implies that $i \leftrightarrow k$) because $P_{ij}^{(r)} > 0$, $P_{jk}^{(s)} > 0 \Rightarrow P_{ik}^{(r+s)} = \sum_{l \in D} P_{il}^{(r)} P_{lk}^{(s)} \geq P_{ij}^{(r)} P_{jk}^{(s)} > 0$.

Result 8: If two states communicate then they are both persistent or both transient.

Proof. Suppose that states i and j communicate. Then there exist $r \geq 0, s \geq 0$ such that $P_{ij}^{(r)} > 0$ and $P_{ji}^{(s)} > 0$. Suppose that state i is transient. By **Note 3** above i is transient $\Leftrightarrow \sum_{n \geq 0} P_{ii}^{(n)} < \infty$. For all $n \geq 0$

$$P_{ii}^{(r+n+s)} = \sum_{l_1 \in D} \sum_{l_2 \in D} P_{il_1}^{(r)} P_{l_1 l_2}^{(n)} P_{l_2 i}^{(s)} \geq P_{ij}^{(r)} P_{jj}^{(n)} P_{ji}^{(s)}$$

and so

$$\infty > \sum_{n \geq 0} P_{ii}^{(n+r+s)} \geq P_{ij}^{(r)} \left(\sum_{n \geq 0} P_{jj}^{(n)} \right) P_{ji}^{(s)} \Rightarrow \infty > \sum_{n \geq 0} P_{jj}^{(n)},$$

since $P_{ij}^{(r)} > 0$ and $P_{ji}^{(s)} > 0$. Therefore state j is also transient.

Definition (Irreducible Chain): If all the state of the chain communicate (that is if $P_{ij}^{(n_{ij})} > 0$ for some $n_{ij} \geq 0$ for all $i, j \in D$) then the chain is called **irreducible**.

NOTE: From Result 8 we conclude that all the states of an irreducible chain are either transient or persistent. **In other words, persistence is a class property.**

PERIODICITY

Definition (Period): the period of the state i is the largest common divisor of the set

$$\left\{n \geq 1 : P_{ii}^{(n)} > 0\right\}$$

FACT: It can be shown that all communicating states have the same period (that is, **periodicity is a class property**).

Definition (Aperiodic Chain): A chain is **aperiodic** if all its states have period equal to one.

STATIONARY DISTRIBUTION

Definition. A **stationary distribution** for a Markov process with transition matrix P is an initial distribution

$$\mathbf{p} = \begin{pmatrix} p_1 \\ \\ p_m \\ \vdots \end{pmatrix}$$

satisfying:

$$\mathbf{p}' = \mathbf{p}' P \quad \text{or} \quad P'\mathbf{p} = \mathbf{p}$$

Equivalently

$$p_j = P(X_1 = j) = \sum_{i \in D} p_i P_{ij}, \quad \text{for all } i \in D.$$

In this case we obviously have

$$\mathbf{p}' = \mathbf{p}' P^n = \mathbf{p}' P^{(n)}, \quad \text{for all } n \geq 1.$$

Therefore,

$$p_j = p_j^{(n)} = P(X_n = j) = \sum_{i \in D} p_i P_{ij}^{(n)}, \quad \text{for all } i \in D, \quad \text{for all } n.$$

We have the following important result:

Result 9: Suppose that the Markov Process is **aperiodic, irreducible** and **has a stationary distribution p**. Then the stationary distribution is unique, satisfies $p_j > 0$ for all $j \in D$, and

$$\lim_{n \to \infty} P_{ij}^{(n)} = p_j \quad \text{for all } i, j \in D.$$

Moreover, the process is persistent and the average return time of state j, is

$$\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)} = 1/p_j \quad \text{for all } j \in D$$

where in general

$$f_{ij}^{(n)} = P\left(X_1 \neq j, \cdots, X_{n-1} \neq j, X_n = j \mid X_0 = i\right), \quad \text{for all } i, j \in D.$$

Strong Law of Large Numbers for Markov Chains

Result 10: If the Markov chain has a unique stationary distribution $\mathbf{p}$ and the function $g(x)$ is such that

$$E_{\mathbf{p}}\left(\left|g\left(X\right)\right|\right) \quad < \quad \infty$$

then

$$\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}g\left(X_i\right) \quad = \quad E_{\mathbf{p}}\left(g\left(X\right)\right), \qquad \text{a.s.}$$

Finding the Invariant (Stationary) Distribution for a Finite State Chain

Note: to find the stationary distribution of a finite state Markov Chain we must solve the equation

$$\mathbf{p}' = \mathbf{p}'P, \quad \mathbf{p} = P'\mathbf{p}$$

That is,

$$(P' - I)\,\mathbf{p} = 0$$

From a practical point of view we can compute the eigen values of the matrix P' and verify that one of them is equal to 1.

For example, in the case of the transition matrix of Example 3,

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/9 & 4/9 & 4/9 & 0 \\ 0 & 4/9 & 4/9 & 1/9 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

we have:

```
===========================
> eigen(t(P))
$values
[1] 1.0000000 0.3333333 -0.3333333 -0.1111111
$vectors
[,1] [,2] [,3] [,4]
[1,] 0.07808688 -0.2236068 0.2236068 0.5
[2,] 0.70278193 -0.6708204 -0.6708204 -0.5
[3,] 0.70278193 0.6708204 0.6708204 -0.5
[4,] 0.07808688 0.2236068 -0.2236068 0.5
>
> a=eigen(t(P))$vectors[,1]
> a/sum(a)
[1] 0.05 0.45 0.45 0.05
```

>

==============================

Therefore the stationary distribution is

$$\mathbf{p} = \begin{pmatrix} 0.05 \\ 0.45 \\ 0.45 \\ 0.05 \end{pmatrix}$$

and the mean return times are

$$\boldsymbol{\mu} = \begin{pmatrix} 20 \\ 2.222 \\ 2.222 \\ 20 \end{pmatrix}.$$

On the other hand, in the case of the transition matrix of example 2

$$P = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

we have:

```
===========================
> eigen(t(P))
$values
[1] 1.0 0.4 0.4 0.4 0.4
$vectors
[,1] [,2] [,3] [,4] [,5]
[1,] 0 0.0000000 0.000000e+00 0.000000e+00 2.293675e-48
[2,] 0 0.0000000 0.000000e+00 1.549469e-32 -1.549469e-32
[3,] 0 0.0000000 1.046728e-16 -1.046728e-16 1.046728e-16
[4,] 0 0.7071068 -7.071068e-01 7.071068e-01 -7.071068e-01
[5,] 1 -0.7071068 7.071068e-01 -7.071068e-01 7.071068e-01

===========================
```

Clearly, this transition matrix doesn't have a proper initial distribution (with $p_i > 0$, $i = 1, 2, 3, 4, 5$).

This was to be expected because this transition matrix is not irreducible: it has two communicating classes: $\{1, 2, 3, 4\}$ which are transient states and $\{5\}$ which is a persistent state (absorbing state).

Finally considering the transition matrix

$$P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.2 & 0.4 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

of Example 1,

```
===========================
> eigen(t(P))
$values
[1] 1.0 0.4 -0.2
$vectors
[,1] [,2] [,3]
[1,] -0.5773503 7.071068e-01 0.4082483
[2,] -0.5773503 8.527521e-16 -0.8164966
[3,] -0.5773503 -7.071068e-01 0.4082483
```

```
> a=eigen(t(P))$vectors[,1] / sum(eigen(t(P))$vectors[,1])
> a
[1] 0.3333333 0.3333333 0.3333333
>
=============================
```

the stationary distribution is

$$\mathbf{p} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

and the mean return times are

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}.$$