Notes for Math 450
Matlab listings for Markov chains

Renato Feres

1 Classification of States

Consider a Markov chain $X_0, X_1, X_2 \ldots$, with transition probability matrix P and set of states S. A state j is said to be *accessible* from i if for some $n \geq 0$ the probability of going from i to j in n steps is positive, that is, $p_{ij}^{(n)} \geq 0$. We write $i \to j$ to represent this. If $i \to j$ and $j \to i$, we say that i and j *communicate* and denote it by $i \leftrightarrow j$.

The definition of communicating states introduces an *equivalence relation* on the set of states. This means, by definition, that $\leftrightarrow$ satisfies the following properties:

1. The relation is *reflexive*: $i \leftrightarrow i$, for all $i \in I$;

2. it is *symmetric*: $i \leftrightarrow j$ if and only if $j \leftrightarrow i$;

3. it is *transitive*: if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.

An equivalence relation on a set S decomposes the set into *equivalence classes*. If S is countable, this means that S can be partitioned into subsets $C_1, C_2, C_3, \ldots$ of S, such that two elements i, j of S satisfy $i \leftrightarrow j$ if and only if they belong to the same subset of the partition. If a state i belongs to C_u for some u, we say that i is a *representative* of the equivalence class C_u. For the specific equivalence relation we are considering here, we call each set C_u a *communicating class* of P. Note, in particular, that any two communicating classes are either equal or disjoint, and their union is the whole set of states.

1.1 Closed classes and irreducible chains

A communicating class is said to be *closed* if no states outside of the class can be reached from any state inside it. Therefore, once the Markov chain reaches a closed communicating class, it can no longer escape it. If the single point set $\{i\}$ comprises a closed communicating class, we say that i is an *absorbing state*. The Markov chain, or the stochastic matrix, are called *irreducible* if S consists of a single communicating class.

As a simple example, consider the stochastic matrix

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

The set of states is $\{1, 2\}$. The communicating class containing 1 is the single point set $\{1\}$, and the communicating class containing 2 is $\{2\}$. The class $\{2\}$ is closed since 1 cannot be reached from 2, but $\{1\}$ is not closed since there is a positive probability of leaving it. Therefore, 2 is an absorbing state and P (or any chain defined by it) is not irreducible.

We wish now to obtain an algorithm for finding the communicating classes of a stochastic matrix P, and for determining whether not they are closed. It is convenient to use the function notation $C(i)$ to denote the communication class containing i. It follows from the definition of $C(i)$ that it is the intersection of two sets:

1. $T(i)$: the set of all states in S that are accessible from i, or the *to-set*;
2. $F(i)$: the set of all states in S from which i can be reached, or the *from-set*.

In other words, j belongs to $T(i)$ if and only if $i \to j$; and j belongs to $F(i)$ if and only if $j \to i$. Notice that the communicating class of i is the intersection of the two:

$$C(i) = T(i) \cap F(i).$$

Moreover, the class $C(i)$ is closed exactly when $C(i) = T(i)$, i.e. when any state that can be arrived at from i already belongs to $C(i)$.

1.2 Algorithm for finding $C(i)$

The following algorithm partitions a finite set of states S into communicating classes. Let m denote the number of elements in S.

1. For each i in S, let $T(i) = \{i\}$;
2. For each i in S, do the following: for each k in $T(i)$, add to $T(i)$ all states j such that $p_{kj} > 0$. Repeat this step until the number of elements in $T(i)$ stops growing. When there are no further elements to add, we have obtained to-sets $T(i)$ for all the states in S. A convenient way to express the set $T(i)$ is as a row vector of length m of 0s and 1s, where the jth entry is 1 if j belongs to $T(i)$ and 0 otherwise. Viewed this way, we have just constructed an m-by-m matrix T of 0s and 1s such that $T(i, j) = 1$ if $i \to j$, and 0 otherwise.
3. To obtain $F(i)$ for all i, first define the m-by-m matrix F equal to the transpose of T. In other words, $F(i, j) = T(j, i)$. Thus, the ith row of F is a vector of 0s and 1s and an entry 1 at position j indicates that state i can be reached from state j.

4. Now defined C as the m-by-m matrix such that

$$C(i,j) = T(i,j)F(i,j).$$

   Notice that $C(i,j)$ is 1 if j is both in the to-set and in the from-set of i, and it is 0 otherwise.

5. The class $C(i)$ is now the set of indices j for which $C(i,j) = 1$. The class is closed exactly when $C(i) = T(i)$.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [C,v]=commclasses(P)
%Input  - P is a stochastic matrix
%Output - C is a matrix of 0s and 1s.
%       - C(i,j) is 1 if and only if j is in the
%       - communicating class of i.
%       - v is a row vector of 0s and 1s. v(i)=1 if
%       - the class C(i) is closed, and 0 otherwise.
[m m]=size(P);
T=zeros(m,m);
i=1;
while i<=m
    a=[i];
    b=zeros(1,m);
    b(1,i)=1;
    old=1;
    new=0;
    while old ~= new
        old=sum(find(b>0));
        [ignore,n]=size(a);
        c=sum(P(a,:),1);
        d=find(c>0);
        [ignore,n]=size(d);
        b(1,d)=ones(1,n);
        new=sum(find(b>0));
        a=d;
    end
    T(i,:)=b;
    i=i+1;
end
F=T';
C=T&F;
v=(sum(C'==T')==m);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Once the matrix C has been obtained using the above program, one can use the command

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
find(C(i,:)==1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

to obtain the set of states in the communicating class of i.

2 Canonical form of P

Suppose that we have found the communicating classes of P and know which ones are closed. We can now use this information to rewrite P by re-indexing the set of states in a way that makes the general structure of the matrix more apparent. First, let $E_1, \dots, E_n$ be the closed communicating classes. All the other classes are lumped together into a set T (for *transient*). Now re-index S so that the elements of E_1 come first, followed by the elements of E_2, etc. The elements of T are listed last. In particular, 1 now represents a state in E_1 and m (the size of S) represents a state in T. We still denote the resulting stochastic matrix by P. Notice that $p_{ij} = 0$ if i and j belong to different closed classes; it is also zero if i is in a closed class and j is in the transient set T. Thus the matrix P takes the block form

$$P = \begin{pmatrix} P_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & P_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & P_n & 0 \\ R_1 & R_2 & R_3 & \cdots & R_n & V \end{pmatrix}$$

The square block P_i defines a stochastic matrix on the set E_i.

The following program gives the canonical form of P. It uses the program `commclasses(P)`.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [Q p]=canform(P)
%Obtain the canonical form Q of a stochastic matrix P.
%The permutation of indices is p.
%Uses the function commclasses(P)
[m m]=size(P);
[C,v]=commclasses(P);
u=find(v==1); %indices in u comprise union of closed classes
w=find(v==0);
R=[];
while length(u)>0
    R=[R u(1)];
    v=v.*(C(u(1),:)==0);
    u=find(v==1);
end
```

```
%R is now the set of representatives of closed classes
%Each closed class has a unique representative in R.
p=[];
for i=1:length(R)
    a=find(C(R(i),:));
    p=[p a];
end
p=[p w];
%We have now a permutation p of indices, p, that
%gives the new stochastic matrix Q.
Q=P(p,p);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Example 2.1 Consider the stochastic matrix

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

We wish to find the communication classes, determine which ones are closed, and put P in canonical form. First, let us write P in Matlab:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
P=zeros(10,10);
P(1,[1 3])=1/2;
P(2,2)=1/3; P(2,7)=2/3;
P(3,1)=1;
P(4,5)=1;
P(5,[4 5 9])=1/3;
P(6,6)=1;
P(7,7)=1/4; P(7,9)=3/4;
P(8,[3 4 8 10])=1/4;
P(9,2)=1;
P(10,[2 5 10])=1/3;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The command `[C,v]=commclasses(P)` gives:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C =
```

```
     1     0     1     0     0     0     0     0     0     0
     0     1     0     0     0     0     1     0     1     0
     1     0     1     0     0     0     0     0     0     0
     0     0     0     1     1     0     0     0     0     0
     0     0     0     1     1     0     0     0     0     0
     0     0     0     0     0     1     0     0     0     0
     0     1     0     0     0     0     1     0     1     0
     0     0     0     0     0     0     0     1     0     0
     0     1     0     0     0     0     1     0     1     0
     0     0     0     0     0     0     0     0     0     1


v =

     1     1     1     0     0     1     1     0     1     0
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Thus we obtain the communication classes

$$\begin{aligned} C(1) &= \{1, 3\} \\ C(2) &= \{2, 7, 9\} \\ C(4) &= \{4, 5\} \\ C(6) &= \{6\} \\ C(8) &= \{8\} \\ C(10) &= \{10\}. \end{aligned}$$

The classes $C(1)$, $C(2)$ and $C(6)$ are closed, while $C(4)$, $C(8)$, and $C(10)$ are not. The permutation of indices that puts P in canonical form, as well as the canonical form itself, are obtained using `[Q p]=canform(P)`. The permutation p is given by [1 3 2 7 9 6 4 5 8 10]. The matrix Q is

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

Therefore, if we ignore the transient states (which the chain will leave, eventually, and never return to), the chain reduces to a simpler one having stochastic matrix P_1, P_2, or P_3, where

$$P_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

involves only the states 1 and 3,

$$P_1 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \\ 1 & 0 & 0 \end{pmatrix}$$

involves the states $2, 7, 9$ and $P_3 = (1)$ describes the constant process at state 6. The following diagram shows more clearly the classes.

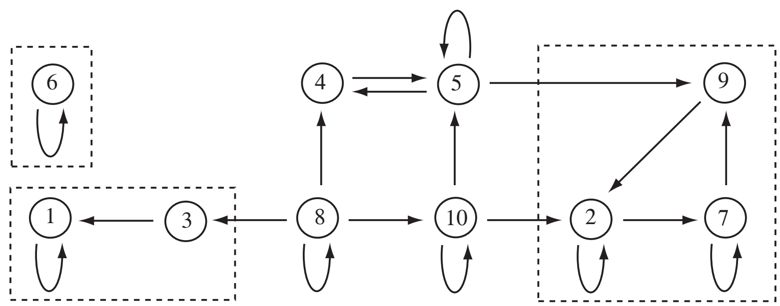

Figure 1: Digraph representing the communication properties of the stochastic matrix P of the example. The closed classes are boxed.

3 Period of an irreducible Markov chain

Consider the graph on the left-hand side of figure 2, representing an irreducible Markov chain. Bunching together the states as in the graph on the right-hand side we note that the set of states decomposes into three subsets that are visited in cyclic order. This type of cyclic structure (possibly consisting of a single subset) is a general feature, as indicated in the next theorem.

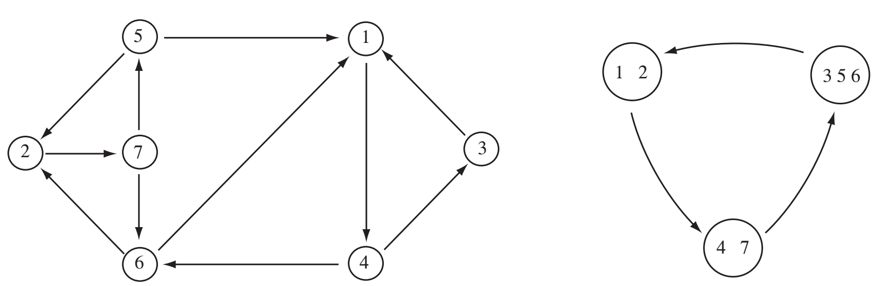

Figure 2: Transition diagram of an irreducible Markov chain of period 3 with cyclic classes $S_1 = \{1, 2\}$, $S_2 = \{4, 7\}$, and $S_3 = \{3, 5, 6\}$.

3.1 G.C.D. and period

We first need a few definitions. Recall that the positive integer d is said to be a *divisor* of the positive integer n if n/d is an integer. If I is a nonempty set of positive integers, the *greatest common divisor*, or g.c.d. of I, is defined to be the largest integer d such that d is a divisor of every integer in I. It follows immediately that the g.c.d. of I is an integer between 1 and the least among $n \in I$. In particular, if $1 \in I$, the g.c.d. of I is 1.

The following simple program can be used to obtain the g.c.d. of a set of numbers.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y=gcd(a,b)
%Obtain the greatest common divisor
%of a and b by the Euclidean algorithm.
n=min(abs(a),abs(b));
N=max(abs(a),abs(b));
if n==0
    y=N;
    return
end
u=1;
while u~=0
    u=rem(N,n);
    if u==0
        y=n;
        return
    end
    N=n;
    n=u;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Let $i \in S$ be a state of a Markov chain such that $p_{ii}^{(n)} > 0$ for some $n \geq 1$. We define the *period* d_i of i by

$$d_i = \text{g.c.d}\{n \geq 1 : p_{ii}^{(n)} > 0\}.$$

Note that if $p_{ii} > 0$, the period of i is 1.

Proposition 3.1 *If i, j are two states in the same communication class of a possibly non-irreducible Markov chain, then $d_i = d_j$.*

Proof. Let n_1 and n_2 be positive integers such that $p_{ij}^{(n_1)} > 0$ and $p_{ji}^{(n_2)} > 0$. Then

$$p_{ii}^{(n_1+n_2)} \geq p_{ij}^{(n_1)} p_{ji}^{(n_2)} > 0,$$

so d_i divides $n_1 + n_2$. If $p_{jj}^{(n)} > 0$, then

$$p_{ii}^{(n_1+n+n_2)} \geq p_{ij}^{(n_1)} p_{jj}^{(n)} p_{ji}^{(n_2)} > 0,$$

so d_i also divides $n_1 + n + n_2$. Therefore, d_i divides n. This means that d_i divides the period of j, so $d_i \leq d_j$. By symmetry the inequality holds in the other direction and we have $d_i = d_j$. □

The proposition shows that the states of an irreducible Markov chain all have the same period, d, which is called the *period* of the Markov chain. The chain is said to be *aperiodic* if its period is $d = 1$. For an irreducible Markov chain to be aperiodic, it is sufficient (but not necessary) that $p_{ii} > 0$ for some i. For example, the transition graph of the figure 3 defines an irreducible and aperiodic Markov chain.

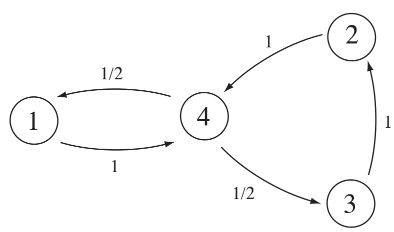

Figure 3: Transition diagram for an irreducible and aperiodic chain with $p_{ii} = 0$ for all i.

3.2 Cyclic decomposition

The sets $S_1, S_2, \ldots, S_d$ in the theorem below are called the *cyclic* classes of the irreducible Markov chain with period d. The theorem says that the Markov chain moves from one cyclic class to the next at each transition in the cyclic order of the classes.

Theorem 3.1 (Cyclic decomposition) *For any irreducible Markov chain, the set of states S can be partitioned in a unique way into k subsets $S_1, S_2, \ldots, S_d$ such that for each S_r and each $i \in S_r$,*

$$\sum_{j \in S_{r+1}} p_{ij} = 1,$$

where by convention $S_d = S_0$, and where d is maximal. (That is, it is not possible to find any other partition with a greater number of elements having the same property.) Furthermore, $Q = P^d$ is a stochastic matrix such that $q_{ij} \neq 0$ only if i, j are in the same set S_k, for some k. Therefore, Q defines a Markov chain on each S_k, which is irreducible and aperiodic.

Proof. Starting with state 1, consider the set of all states that can be reached from 1 in nd steps, where d is the period of the Markov chain and n is a positive integer. Note that S_1 contains 1. Then define S_i the set of states that can be reached from any state in S_1 in $i-1$ steps, for $i = 1, 2, \ldots, d$. It is left as an exercise (until I get around to writing the details here) that this decomposition has the properties claimed. □

The following program calculates the period, d, of a Markov chain and the cyclic classes, indexed by $\{0, 1, \ldots, d-1\}$, to which each state belongs.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [d v]=period(P)
%Obtain the period of an irreducible transition
%probability matrix P of size n-by-n.
%The cyclic classes are numbered 0, 1, ..., d-1
%and v=[a_1 ... a_n] is a vector with entries in
%{0, 1, ..., d-1} such that a_i is the cyclic class
%of state i. (Algorithm by Eric V. Denardo.)
%Uses the program gcd.
n=size(P,2);
v=zeros(1,n);
v(1,1)=1;
w=[];
d=0;
T=[1];
m=size(T,2);
while (m>0 & d~=1)
    i=T(1,1);
    T(:,1)=[];
    w=[w i];
    j=1;
    while j<=n
        if P(i,j)>0
            r=[w T];
            k=sum(r==j);
            if k>0
                b=v(1,i)+1-v(1,j);
                d=gcd(d,b);
            else
                T=[T j];
                v(1,j)=v(1,i)+1;
            end
        end
        j=j+1;
    end
    m=size(T,2);
```

```
end
v=rem(v,d);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

4 Passage and hitting times

Let $X_0, X_1, \ldots$ be a Markov chain with state space S, initial probability distribution π, and transition probabilities matrix P. Define the *first passage time* from state i to state j as the number T_{ij} of steps taken by the chain until it arrives for the first time at state j given that $X_0 = i$. This is a random variable with values in the set of non-negative integers. Its probability distribution function is given by

$$h_{ij}^{(n)} = P(T_{ij} = n) = P(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j | X_0 = i).$$

The first passage times can be found recursively as follows: $h_{ij}^{(1)} = p_{ij}$ and, for $n \geq 2$,

$$h_{ij}^{(n)} = \sum_{k \in S - \{j\}} p_{ik} h_{kj}^{(n-1)}.$$

Let $H^{(n)}$ denote the matrix with entries $h_{ij}^{(n)}$ and $H_0^{(n)}$ the same matrix except that the diagonal entries are set equal to 0. Then $H^{(1)} = P$ and an easy calculation gives

$$H^{(n)} = P H_0^{(n-1)}.$$

Let h_{ij} (without upper-script) be the *reaching probability* from state i to j, i.e., the probability that state j is ever reached from state i. Then

$$h_{ij} = P(T_{ij} < \infty) = \sum_{n=1}^{\infty} P(T_{ij} = n) = \sum_{n=1}^{\infty} h_{ij}^{(n)}.$$

The following program gives the first passage time matrix $H^{(n)}$. The (i,j)-entry is the probability of arriving at j for the first time at time n given the initial state i.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function H=firstpassage(P,i,n)
%For a transition probability matrix P
%obtain first passage probabilities from state
%i to all states in 1:n steps. The output is
%the matrix H with (k,j)-entry is hij(k), where
%k=1:n. In other words, the columns are indexed
%by the destination and the rows are indexed by
%the number of time steps till first passage.
G=P;
H=[P(i,:)];
```

```
E=1-eye(size(P));
for m=2:n
    G=P*(G.*E);
    H=[H;G(i,:)];
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

More generally, we define the *hitting time*, T_A, of a subset $A \subseteq S$ as the first time (possibly infinite) that $X_n \in A$. The probability starting from i that $\{X_n\}$ ever hits A is then

$$h_{iA} = P(T_A < \infty | X_0 = i) = P(T_{iA}).$$

If $A = \{j\}$ consists of a single state, we are back to the previous definitions. If A is a closed communicating class, then h_{iA} is called the *absorption probability* of A starting from i.

State i is called *recurrent* if $h_{ii} = 1$, so that starting at state i the chain with probability 1 eventually returns to i. If $h_{ii} < 1$ state i is called transient, so there is in this case a positive probability that starting at i, the chain never again returns to i. For any i, define the *recurrence time* of state i as the random variable T_{ii}. Then if state i is recurrent we have $P(T_{ii} < \infty) = 1$. Denote the expected recurrence time to i by

$$\mu_{ii} = E[T_{ii}].$$

The expected time for $\{X_n\}$ to reach a set of states A from i is

$$\mu_{iA} = E[T_{iA}] = \sum_{n=0}^{\infty} nP(T_{iA} = n)$$

if with probability 1 A is eventually reached, and ∞ otherwise. Thus we have the following quantities of interest:

$$h_{iA} = P(\text{hit } A \text{ from } i), \quad \mu_{iA} = E[\text{time to hitting } A \text{ from } i].$$

4.1 Number of visits

Given $X_0 = i$, we are now interested in counting the number of visits to state j over a period of time. Define the function $I_{ij}(n)$ to be 1 if $X_n = j$ given that $X_0 = i$, and 0 otherwise. The number of visits to state j, starting at state i, by time n is defined as

$$N_{ij}(n) = \sum_{k=1}^{n} I_{ij}(k).$$

The initial passage time from i to j is distributed according to $h_{ij}^{(n)}$ and all the subsequent return times to j follow the distribution $h_{jj}^{(n)}$. If the chain is presently in a given state, the first time it will visit state j is a stopping time.

By the strong Markov property, we conclude that these interarrival times are conditionally independent. Using these facts, the *mean state-occupancy time*, defined as

$$M_{ij}(n) = E[N_{ij}(n)],$$

can be obtained as follows:

$$\begin{aligned} M_{ij}(n) &= E\left[\sum_{k=1}^{n} I_{ij}(k)\right] \\ &= \sum_{k=1}^{n} E[I_{ij}(k)] \\ &= \sum_{k=1}^{n} p_{ij}^{(k)}. \end{aligned}$$

If $M(n)$ denotes the matrix with entries $M_{ij}(n)$, then

$$M(n) = \sum_{k=1}^{n} P^{(k)}.$$

Recall that if state j is recurrent, then $h_{jj} = 1$. This means that state j will be visited infinitely often, that is, $(N_{jj}(\infty) = \infty) = 1$ or $M_{jj}(\infty) = \infty$. On the other hand, if state j is transient, then $h_{jj} < 1$, and $N_{jj}(\infty)$ is a geometric random variable with probability distribution function

$$P(N_{jj}(\infty) = k) = (h_{jj})^k (1 - h_{jj})$$

for $k = 0, 1, 2, \dots$, with mean

$$M_{jj}(\infty) = E[N_{jj}(\infty)] = \frac{1}{1 - h_{jj}} < \infty.$$

Therefore, state j is recurrent if and only if

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

This gives another way to characterize a recurrent state.

5 Stationary distributions

Consider an irreducible Markov chain with state space $S = \{0, 1, 2, \dots\}$ consisting of a single closed communicating class. Let $N_{ij}(n)$ denote the number of visits to state j in n transition steps given that $X_0 = i$. Let T_{ij} denote the first passage time from state i to state j. Then the following holds:

$$\lim_{n\to\infty} \frac{N_{ij}(n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \frac{1}{\mu_{jj}}$$

where $\mu_{jj} = E[T_{jj}]$ is the expected recurrence time to state j.

If state j is aperiodic, then we have the stronger result:

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

independent of the initial state i. If the state has a period d, then

$$\lim_{n\to\infty} p_{jj}^{(nd)} = \frac{d}{\mu_{jj}},$$

where it is assumed here that $X_0 = j$. Denote the limiting state probability by

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

For an aperiodic chain, we have

$$\pi_j = \frac{1}{\mu_{jj}}.$$

Thus the limiting probability distribution is the reciprocal of the mean recurrence time. Recall that state j is said to be positive recurrent if $\mu_{jj} < \infty$ and null recurrent if $\mu_{jj} = \infty$. Hence for the former case we have $\pi_j > 0$ and for the latter $\pi_j = 0$.

A probability distribution π_i, $i \geq 0$, is a *stationary distribution* of a Markov chain with transition matrix P if $\pi = \pi P$, that is,

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

for all j.

We say that a Markov chain is *ergodic* if it is irreducible, aperiodic, and positive recurrent. The limiting distribution of an ergodic chain is the unique nonnegative solution of the equation $\pi = \pi P$ such that $\sum_j \pi_j = 1$.

The ratio π_j/π_i for a stationary distribution π has the following useful interpretation. Consider a discrete process in which each random time step corresponds to the return time to a state i. The interarrival time in this process is the recurrence time T_{ii}. Let V_j denote the number of visits to state j between two successive visits to i. Then

$$\pi_j = \lim_{n\to\infty} P(X_j = j) = \frac{E[V_j]}{E[T_{ii}]} = E[V_j]\pi_i.$$

In words, the ratio of the two limiting state probabilities represents the expected number of visits to state j between two successive visits to i.

When a chain is irreducible, positive recurrent, and periodic of period d, we call it a *periodic Markov chain*. The solution of $\pi = \pi P$ can be interpreted as the the long-run fraction of time that the process will be visiting state j. To show this is the case, let $I_j(k)$ be the indicator function of state j and define the time-average probability as

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} E\left[\sum_{k=1}^{n} I_j(k)\right].$$

Conditioning on the possible states leading into state j in one step, we write:

$$\begin{aligned}
\pi_j &= \lim_{n\to\infty} \frac{1}{n} E\left[\sum_{k=1}^{n}\sum_{i=0}^{\infty} I_i(k-1)I_j(k)\right] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n}\sum_{i=0}^{\infty} E\left[I_i(k-1)I_j(k)\right] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n}\sum_{i=0}^{\infty} E\left[I_i(k-1)p_{ij}\right] \\
&= \sum_{i=0}^{\infty} p_{ij} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E\left[I_i(k-1)\right] \\
&= \sum_{i=0}^{\infty} \pi_i p_{ij}.
\end{aligned}$$

6 Censored Markov chain

Let $X_0, X_1, X_2, \ldots$ be an ergodic Markov chain with state space S and transition probability matrix P. Let $A \subseteq S$ and $B = A^c$ the complement of A. We form a stochastic process $Y_0, Y_1, Y_2, \ldots$ by stopping X_n at the random times $T_A^{(n)}$, where

$$T_A^{(0)} = \min\{m \geq 0 | X_m \in A\} \quad \text{and} \quad T_A^{(n+1)} = \min\{m > T_A^{(n)} | X_m \in A\}.$$

We also write $T_A = T_A^{(1)}$ for the first return time to A.

So $Y_n = X_{T_A^{(n)}}$ for $n \geq 0$. As the chain is ergodic, $P(T_A^{(n)} < \infty) = 1$. By the strong Markov property $Y_0, Y_1, \ldots$ is a Markov chain, called the *censored* Markov chain. Thus the states of the censored Markov chain are elements of A, and correspond to the states of the original chain at the return times to A.

We want to describe the transition probabilities matrix, Q, for the Y_n. To do so we first write P in block form:

$$P = \begin{pmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{pmatrix},$$

where P_{AA} contains the probabilities of transitions from a state in A to another in A, P_{AB} contains the probabilities of transitions from a state in A to a state in B and so on. For $i, j \in A$,

$$\begin{aligned} Q(i,j) &= P(i,j) + \sum_{n=1}^{\infty} P(X_0 = i, X_1 \in B, \ldots, X_n \in B, X_{n+1} = j) \\ &= P(i,j) + \sum_{n=1}^{\infty} \sum_{i_1 \in B} \cdots \sum_{i_n \in B} P(i, i_1) P(i_1, i_2) \ldots P(i_n, j) \\ &= P_{AA}(i,j) + \sum_{n=1}^{\infty} (P_{AB} P_{BB}^{n-1} P_{BA})(i,j), \end{aligned}$$

Therefore,

$$Q = P_{AA} + P_{AB} \left(\sum_{n=0}^{\infty} P_{BB}^n \right) P_{BA}.$$

Notice that the matrix P_{BB} has the property that the sum of the entries in each row is strictly less than 1. Call this number a. If we define the norm $\|R\|$ of a matrix R to be the maximum of $|R(i,j)|$ over all the entries, then $\|P_{BB}^n\| \leq Ca^n$ for some constant $C > 0$. This remark can be used to show that the matrix series in the expression of Q is convergent, $I - P_{BB}$ is invertible, and

$$\sum_{n=0}^{\infty} P_{BB}^n = (I - P_{BB})^{-1}.$$

Therefore,

$$Q = P_{AA} + P_{AB}(I - P_{BB})^{-1} P_{BA}.$$

It is not difficult to show, using ergodicity, that Q is a stochastic matrix.

We now wish to find the stationary probability distribution for Q. Let $\pi = (\pi_1, \ldots, \pi_N)$ be the stationary distribution for P, and $\eta = (\eta_1, \ldots, \eta_K)$ the stationary distribution for Q, where we use $\{1, \ldots, K\}$ to designate the elements of A. We claim that

$$\eta_i = \frac{\pi_i}{\sum_{j \in A} \pi_j}.$$

This can be seen as follows. Write $\pi = (\pi_A, \pi_B)$, where π_A and π_B are the restrictions of π to indices in A and B, respectively. The chain Y_k is also ergodic, so it has a unique stationary distribution η. If we shown that $\pi_A Q = \pi_A$, it will follow that $\eta = c\pi_A$, where $c > 0$ is the normalization constant $\sum_{i \in A} \pi_i$. From $\pi P = \pi$ we obtain

$$\begin{aligned} \pi_A &= \pi_A P_{AA} + \pi_B P_{BA} \\ \pi_B &= \pi_A P_{AB} + \pi_B P_{BB}. \end{aligned}$$

Recursively replacing π_B, as given in the second equation, into the first gives

$$\pi_A = \pi_A \left(P_{AA} + \sum_{k=0}^{n} P_{AB} P_{BB}^k P_{BA} \right) + \pi_B P_{BB}^{n+1} P_{BA}.$$

Now P_{BB}^n converges to the zero matrix as $n \to \infty$, so we obtain

$$\pi_A = \pi_A Q.$$

This proves the claim.

7 Computation of the stationary probabilities

The ideas of the previous section can be used to derive a numerically stable algorithm for computing the stationary distribution of an ergodic Markov chain, called the method of *state space reduction*. Throughout this section, we indicate vector and matrix components using function notation rather than indices. Thus the (i, j)-entry of a matrix P will be written $P(i, j)$.

Consider such a chain with state space $S = \{1, \dots, N\}$ and transition probabilities matrix P. The method of state space reduction consists of first deriving from P, inductively, the stochastic matrices: $P_N, P_{N-1}, \dots P_1$, where $P_N = P$ and P_n is the transition probabilities matrix for the return process to the subset $\{1, \dots, n\}$ of S. For each n, write the matrix P_n in block form as

$$P_n = \begin{pmatrix} T_n & u_n \\ r_n & \lambda_n \end{pmatrix}$$

As we saw in the previous section, P_n is obtained from P_{n+1} as follows:

$$P_n = T_{n+1} + (1 - \lambda_{n+1})^{-1} u_{n+1} r_{n+1}$$

where $u_{n+1} r_{n+1}$ denotes matrix multiplication of a row and a column vector. At each step, from $N - 1$ to 1, we store the value of the vector

$$a_n = u_{n+1}/(1 - \lambda_{n+1}) = u_{n+1}/(r_{n+1}(1) + \cdots + r_{n+1}(n)).$$

Denote by π_n the stationary distribution of P_n. Then π_n satisfies the equation $\pi_n = \pi_n P_n$. In particular,

$$\pi_n(n) = \pi_n(1) P_n(1, n) + \cdots + \pi_n(n) P_n(n, n).$$

Using the definition of a_n and isolating $\pi_n(n)$ on the left-hand side, this equation can be written as

$$\pi_n(n) = \pi_n(1) a_n(1) + \cdots + \pi_n(n - 1) a_n(n - 1).$$

Therefore, using this equation with the stored values of the vectors a_n, $n = N, \dots, 2$, obtained from the backward recursion, we can obtain π_{n+1} from π_n. In fact, recall from the previous section that $\pi_{n+1}(j) = c\pi_n(j)$ for a constant independent of $j = 1, \dots, n$. The last component, $\pi_{n+1}(n + 1)$ is then obtained as described above. The arbitrary constant c is then obtained by normalizing the vector. This algorithm is implemented by the following program.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function p=limitdist(P)
%Obtain the stationary probability distribution
%vector p of an irreducible, recurrent Markov
%chain by state reduction. P is the transition
%probabilities matrix of a discrete-time Markov
%chain or the generator matrix Q.
[ns ms]=size(P);
n=ns;
while n>1
    n1=n-1;
    s=sum(P(n,1:n1));
    P(1:n1,n)=P(1:n1,n)/s;
    n2=n1;
    while n2>0
        P(1:n1,n2)=P(1:n1,n2)+P(1:n1,n)*P(n,n2);
        n2=n2-1;
    end
    n=n-1;
end
%backtracking
p(1)=1;
j=2;
while j<=ns
    j1=j-1;
    p(j)=sum(p(1:j1).*(P(1:j1,j))');
    j=j+1;
end
p=p/(sum(p));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

References

[Bré] Pierre Brémaud. *Markov Chains – Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer-Verlag, 1999.

[Kao] Edward P.C. Kao. *An Introduction to Stochastic Processes.*