Discrete-time Markov chains

January 27, 2012

This problem set is designed to cover the material on discrete time Markov chains that you have met in the lecture this morning, as well as introducing some additional ideas. Questions marked with a ∗ are for the more mathematically inclined. Worked answers will be made available at the end of the morning.

1 Markov chains

Recall that a discrete-time Markov chain is characterised by a transition matrix $P$, whose entries $P_{ij}$ give the probability of transitioning from state $i$ to state $j$, with $i, j \in \{1, \ldots, K\}$ indexing a finite state space\(^1\) of some kind, such that $\sum_j P_{ij} = 1$. If $X$ is a random variable that evolves according to the Markov chain specified by $P$, then $P_{ij} = p(X(n+1) = j | X(n) = i)$. After $n$ steps, the probability of being in the various possible states is given by a row vector $\pi(n)$, where $\pi_i(n) = p(X(n) = i)$. We can then write the distribution after $n$ steps in terms of the initial distribution as follows

$$\pi(n) = \pi(n-1)P$$
$$= (\pi(n-2)P)P$$
$$= ((\ldots (\pi(0)P)\ldots P)P$$
$$= \pi(0)P^n$$

In the case where we consider the evolution of one single instance $X$, which begins in a particular state $i$, then $\pi(0)$ is a vector with $\pi_j(0) = \delta_{ij}$, i.e. a vector of zeros with a one at position $i$. If, rather than just a single instance, we begin with a population $\{X^{(1)}(0), X^{(2)}(0), \ldots, X^{(M)}(0)\}$ all evolving according to the same Markov chain, then we can regard $\pi(0)$ as the initial proportion of the population in each state.

2 Stationary distributions

Under certain conditions, the distribution $\pi(n)$ will tend towards a unique stationary state $\pi^{(eq)}$ as $n$ tends towards infinity. More specifically, convergence requires that the Markov chain is

\(^1\)We will not discuss infinite state spaces in these exercises, but these will be mentioned in the lectures.
- Irreducible: the state space cannot be partitioned into non-communicating subsets. This means that it is possible to move from each state to every other state by some sequence of moves.

- Positive recurrent: if the chain is run for long enough, it will return to its initial state with probability 1.

- Aperiodic: a Markov chain is periodic if it is only possible to transition from state $i$ to itself in multiples of $N$ steps, where $N > 1$.

For finite state spaces, irreducibility implies positive recurrence, and it is usually easy to ensure that the state space is irreducible. Periodicity can be slightly more subtle, and we will examine this in more detail later.

### 3 Eigenvector decomposition

Since it is by definition unchanging, the stationary (or equilibrium) distribution satisfies $\pi^{(eq)} P = \pi^{(eq)}$, such that $\pi^{(eq)}$ is a left eigenvector of $P$ with eigenvalue 1. (Recall that left eigenvectors of a matrix $A$ satisfy the equation $v_k A = \lambda_k v_k$.)

In the case that the rest of the eigenvalues are distinct, we can write $P$ as an eigenvector decomposition

$$ P = V^{-1} \Lambda V $$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$ is a diagonal matrix of eigenvalues, and $V$ contains the left eigenvectors on its rows (convince yourself of this by looking at what happens if you form the matrix $VP$).

If this eigenvector decomposition can be formed, it enables us to do several things very easily. Firstly, we can determine the stationary distribution as the eigenvector corresponding to the eigenvalue 1. Furthermore, we can then write

$$ P^n = (V^{-1} \Lambda V)^n $$

$$ = V^{-1} \Lambda V V^{-1} \Lambda V \ldots V^{-1} \Lambda V V^{-1} \Lambda V $$

$$ = V^{-1} \Lambda^n V $$

where the last step uses the fact that a matrix multiplied by its inverse is the identity matrix. Since $\Lambda$ is a diagonal matrix, $\Lambda^n = \text{diag}(\lambda_1^n, \ldots, \lambda_K^n)$, and this enables us to easily compute powers of $P$ given the eigenvalues and eigenvectors. In the special case that the matrix $P$ is symmetric, we have $V^{-1} = V^T$, such that $P = V^T \Lambda V$.

We can also use the eigendecomposition to express the distribution at any particular step in terms of a sum of contributions from the different eigenvectors

$$ \pi^{(n)} = \pi^{(0)} V^{-1} \Lambda^n V $$

$$ = a \Lambda^n V $$

$$ = \sum_k a_k \lambda_k^n v_k $$

where $a = \pi^{(0)} V^{-1}$ is a projection of the starting distribution onto the right eigenvectors.
Reminder: Assuming that the eigendecomposition exists, we can find the eigenvalues and eigenvectors by solving

\[ VP = \Lambda V \]
\[ \Rightarrow (P - \Lambda)V = 0 \]

This gives a set of equations of the form

\[ (P - \lambda_k I)v_k = 0 \]

where \( I \) is the identity matrix. These have non-trivial solutions if \( det(P - \lambda_k I) = 0 \), where \( det \) is the matrix determinant.

### 3.1 Example

Let \( X(0) \) represent the allele of a particular gene present in a founding individual, and \( X(n) \) be the allele present in the (asexually produced) progeny of this individual after \( n \) generations. Suppose that this allele does not confer any effect on fitness, but that at each generation there is a certain probability of mutation to the other form of the allele.

We will represent this system by a discrete-time Markov chain with state space \( \{-1, 1\} \) (corresponding to the two alleles), evolving according to the transition matrix

\[
P = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \alpha & \alpha \end{pmatrix}
\]  

(2)

which is shown schematically by the diagram below.

![Diagram](image)

**Question 1.** For the \( 2 \times 2 \) matrix \( P \) shown above, compute the two eigenvalues as the solutions to the equation \( det(P - \lambda I) = 0 \). (Recall that the determinant of a \( 2 \times 2 \) matrix \( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) is given by \( ad - bc \).)

Substitute these eigenvalues back into the equation \( (P - \lambda_k I)v_k = 0 \) in order to compute the corresponding eigenvectors. Compute the \( k \)-step transition matrix \( P^k \), and hence (or otherwise) compute the distribution of the alleles after 4 generations, given that the founding individual has the \(-1\) copy, and that the probability of mutation is 0.25 (i.e. \( \alpha = 0.75 \)).

### 4 Reversibility

At stationarity/equilibrium, the distribution \( \pi^{(eq)} \) satisfies the equation \( \pi^{(eq)} P = \pi^{(eq)} \). This implies a series of equalities of the form

\[
\sum_j \pi_j^{(eq)} P_{ji} = \pi_i^{(eq)} \quad \text{for all } i
\]
Since each row of $P$ sums to one, we can multiply the right hand side by $\sum_j P_{ij}$, which yields

$$\sum_j \pi_j^{(eq)} P_{ji} = \pi_i^{(eq)} \sum_j P_{ij} = \sum_j \pi_i^{(eq)} P_{ij} \text{ for all } i$$

This equation defines what is known as global balance, whereby the probability of moving into a state equals the probability of leaving the state, which is an obvious requirement at equilibrium.

One way of ensuring global balance is to stipulate the stronger condition of detailed balance, also known as reversibility, which states that

$$\pi_j^{(eq)} P_{ji} = \pi_i^{(eq)} P_{ij}$$

(3)

Summing each side over $j$, we see that detailed balance implies global balance, although it is not a necessary condition (except as discussed below). The subset of transition matrices for which reversibility holds at equilibrium define a class of reversible Markov chains. One very useful consequence of reversibility is that it allows us to easily design a transition matrix that converges towards a particular stationary distribution, by choosing the elements of $P$ such that $P_{ij}/P_{ji} = \pi_i^{(eq)}/\pi_j^{(eq)}$ for all $i, j$. This is particularly important in the context of Markov chain Monte Carlo, which you will cover with Gil tomorrow.

**Question 2.** Show that for a two-state discrete Markov chain global balance implies detailed balance.

### 5 Autocorrelation

For a first-order Markov chain, we can say that, conditional on the value of $X(n-1)$, $X(n)$ is independent of all the previous observations. However, since each $X(n)$ depends on its predecessor, this induces a non-zero correlation between $X(n)$ and $X(n+d)$, even when the distance $d$ is greater than 1. (Typically conditional independence between two variables given a third variable does not imply that the first two are uncorrelated.) The correlation between $X(n)$ and $X(n+d)$ is termed the autocorrelation at lag $d$, and for a Markov chain that converges to a unique stationary distribution we expect the autocorrelation to decrease as the lag is increased.

* **Question 3.** Using your expression for the diagonal elements of $P^k$, show that the autocorrelation at lag $d$ for the Markov chain defined by the matrix in equation (2) is given by

$$\mathbb{E}[X(n)X(n+d)] = (2\alpha - 1)^d$$

**Question 4.**

a) Starting at $X(0) = -1$, generate a random sequence of length 100 according to the Markov chain defined in Question 1, with $\alpha = 0.75$. 4
b) Using the MATLAB function `acf.m` provided below, compare the observed autocorrelation to the theoretical value of \((2\alpha - 1)^d\). Does the agreement improve if you increase the length of the sampled sequence?

c) What happens when you set \(\alpha < 0.5\)?

```
function [ autocorr ] = acf( X )

n = length(X); % Number of steps in the trajectory
J = min(n-1,20); % Maximum lag
autocorr = zeros(1,J+1); % A vector to contain the results

for (i = 1:(n-1))
    for (j = 0:min(J,n-i))
        autocorr(j+1) = autocorr(j+1) + X(i) * X(i+j) / (n-j);
    end
end
plot(0:J,autocorr)
ylim([-1,1])
xlabel('Lag')
ylabel('Autocorrelation')
end
```

6 Periodicity

We mentioned earlier that one of the conditions required for a Markov chain to have a unique stationary distribution is for it to be aperiodic. However, we were somewhat vague about exactly what this entails. In this section we will examine this in more detail.

The figure below shows a simple example of a periodic Markov chain, where transitions from state \(i\) to state \(i+1\) occur with probability 1.

```
beginning in any state, this Markov chain guarantees that we will end up in the same state after 6 steps, and we say that it has period 6.
```
**Question 5.** Formulate the transition matrix corresponding to this system.

**Question 6.** Suppose we are interested in studying the traffic flow at different sections of the M25. Assuming for the moment that everyone is driving at the same speed, and that the number of vehicles entering at each junction is the same as the number exiting, we can represent the distribution of cars around the motorway by a cyclic Markov chain, where each transition represents a fixed time interval. Dividing the road into six sectors, we have a system analogous to the one above.

Starting with a random distribution of the cars at all sectors (for example setting \( \pi^{(0)} = \text{rand}(1,6) \)), compute the proportion of vehicles in region 6 at each iteration for 100 iterations. What do you notice?

* **Question 7.** Compute the eigenvalues of the transition matrix for this cyclic system. Using equation (1), show that each transition step corresponds to a rotation of the eigenvectors in the complex plane, and hence a rotation of the probability distribution around the cyclic system.

Our initial model of the M25 is clearly somewhat over-simplified, not least because we have neglected the fact that there may be reasons for vehicles to take longer to move through particular sectors. For example, perhaps there is a service station in sector 5, such that there is a certain probability for drivers to remain in this sector for a while.

Supposing that the probability of a driver feeling hungry and taking a break at the service station while in sector 5 is \( \alpha \), we can incorporate this into the model by adding in a ‘waiting state’ with self transition probability \( \alpha \), as shown below.

**Question 8.** Plot the proportion of cars at in section 6 at each iteration again, taking into account the effect of the service station. What happens this time?

**Question 9.** Does the system reach a stationary distribution? What does this imply about the eigenvalues?