Markov Chains in discrete time

4.1 The Markov Property

The Markov property is about independence, specifically about the future development of a process being independent of anything that has happened in the past, given its present value.

A random process X possesses the *Markov property*, and is called a *Markov Chain*, if

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, \mathcal{H}_n)$$

depends only on i and j, not on any past values.

For the simple random walk,

$$\mathbf{P}(X_{n+1} = i{+}1 \mid X_n = i, X_{n-r} = k)$$

is just equal to

$$\mathbf{P}(J_{n+1} = 1 \mid X_n = i, X_{n-r} = k) = p,$$

since J_{n+1} is independent of anything that happened previously.

The Markov property also holds for SRW with barriers. But there are many more examples.

Example: Weather. Sunny or Cloudy. Today's weather affects tomorrow's, but yesterday's is already irrelevant to tomorrow's forecast.

The probability $\mathbf{P}(X_{n+1} = 1 \mid X_n = i)$ is called a *transition probability* and we use the notation p_{ij}. In most of the examples we consider, the p_{ij} do not depend on n; such processes are termed *time-homogeneous*.

4.2 The transition matrix

Any random process has a fixed set of values which it can take. There could be finitely many or infinitely many; they could be numerical or descriptive (e.g. Sunny). The collection of possible values is called the *state space*, $\mathcal{S}$, and the possible values are termed *states*.

The transition probability p_{ij} is defined for all *i* and *j* in $\mathcal{S}$, though may be 0 in many cases.

We can assemble the p_{ij} into a matrix P, called the *transition matrix* of X.

Example: the transition matrix of a SSRW with a reflecting boundary at 0 and an absorbing one at 4 is

$$\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 \\ 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 \\ 0 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{array}$$

The elements of a transition matrix must lie between 0 and 1. In addition, since each transition must take the chain somewhere, the row sums must be equal to 1.

4.3 Multi-step transitions

Knowing today's weather, the transition probabilities give information about tomorrow's. But what of the day after tomorrow?

If today is sunny:

- the probability that tomorrow will be sunny is p_{SS} and the probability that both tomorrow and the day after will be sunny is $p_{SS}{}^2$;
- the probability that tomorrow will be cloudy and the next day sunny is $p_{SC}p_{CS}$

We can say, then, that

$$\mathbf{P}(X_2 = \mathrm{S} \mid X_0 = \mathrm{S}) = \sum_{j=\mathrm{S,C}} p_{\mathrm{S}j}\, p_{j\mathrm{S}}.$$

In general, the Law of Total Probability shows that the same holds for more than two states:

$$\mathbf{P}(X_2 = j \mid X_0 = i) = \sum_k p_{ik}\, p_{kj}.$$

This is the (i,j)th entry of the matrix P^2.

$\mathbf{P}(X_2 = j \mid X_0 = i)$ is a *two-step transition probability*; we denote it by $p_{ij}{}^{(2)}$. We see that the two-step transition matrix $P^{(2)}$ is equal to P^2.

The *Chapman-Kolmogorov equations* state that

$$P^{(m+n)} = P^{(m)}\, P^{(n)},$$

from which we deduce that

$$\mathbf{P}(X_n = j \mid X_0 = i) = (P^n)_{ij}.$$

In principle we only need to find the powers of P to evaluate the distribution of X_n for all n.

Example: in the weather model suppose $p_{\mathrm{SS}} = 0.6$ and $p_{\mathrm{CC}} = 0.7$. Then the 1-step, 2-step and 3-step transition matrices are

$$\begin{matrix} 0.6 & 0.4 \\ 0.3 & 0.4 \end{matrix}\,, \quad \begin{matrix} 0.48 & 0.52 \\ 0.39 & 0.61 \end{matrix} \quad \text{and} \quad \begin{matrix} 0.444 & 0.556 \\ 0.417 & 0.583 \end{matrix}$$

Within each column the values are getting closer.

4.4 Matrix algebra

Given a $k \times k$ square matrix P there exist k pairs $(\lambda, \underline{v})$, where λ is a number and $\underline{v}$ a vector, such that

$$P\underline{v} = \lambda\underline{v}.$$

We term λ an *eigenvalue*, $\underline{v}$ an *eigenvector* of P.

Any vector can be expressed as a linear combination of eigenvectors of P.

The eigenvalues of a transition matrix may be real or complex, but will all lie in or on the unit circle.

Find the eigenvalues of P by solving the equation

$$\det(P - \lambda I) = 0,$$

where I is the $k \times k$ identity matrix.

Then, for each solution λ, solve linear equations to calculate the corresponding $\underline{v}$.

Notice that the eigenvectors $\underline{v}$ are only unique up to a scalar multiple, since $c\underline{v}$ is also an eigenvector for any scalar c.

Eigenvectors represent directions rather than points. When an eigenvector of P is subjected to the linear transformation which P represents, the resulting vector is in the same direction as the original.

Construct a matrix Λ which has the eigenvalues of P down the major diagonal and zeroes everywhere else. (Such matrices are called *diagonal matrices*.) Construct also a matrix V by writing the eigenvectors in the columns, in the same order as the corresponding eigenvalues in Λ. Then

$$PV = V\Lambda \text{ and } V\Lambda V^{-1} = PVV^{-1} = P$$

Therefore

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

Λ^n is diagonal, with entries λ_j^n.

A transition matrix always has one eigenvalue equal to 1, with corresponding eigenvector $(1, 1, \ldots 1)^T$. Usually (see later for conditions) all the other eigenvalues will lie strictly inside the unit circle.

If $|\lambda_j| < 1$ then $|\lambda_j|^n \to 0$ as $n \to \infty$. For most transition matrices the limit of Λ^n is a matrix with a single 1 on the diagonal, making the limit of P^n easy to find.

4.5 Equilibrium

Luckily we only need the matrix algebra (a) to show that a limit exists, (b) to find $\mathbf{P}(X_n = j \mid X_0 = i)$ for all n. If all we need to do is find the limiting probabilities, there is a quicker way.

Using the Law of Total Probability,

$$\mathbf{P}(X_{n+1} = j) = \sum_i \mathbf{P}(X_{n+1} = j \mid X_n = i)\, \mathbf{P}(X_n = i).$$

Assuming we know that $\mathbf{P}(X_n = j)$ converges to some limit or other (call it π_j) as $n \to \infty$, we have

$$\pi_j = \sum_i \pi_i p_{ij}$$

In matrix notation,

$$\pi^T = \pi^T P$$

This is usually easy to solve. The solution π is called the *equilibrium probability vector* of X.

Example: for the weather model with $p_{SS} = 0.6, p_{CC} = 0.7$, we have

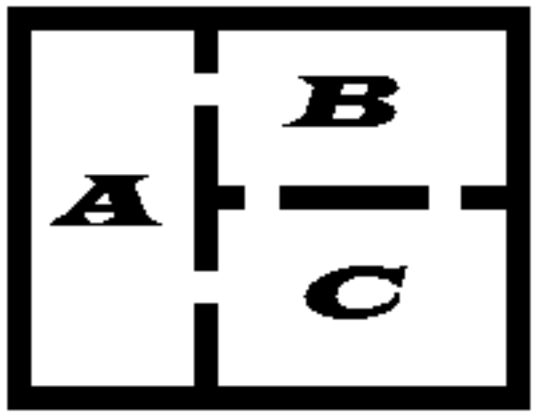
$$\pi_S = \pi_S p_{SS} + \pi_C p_{CS} = 0.6\pi_S + 0.3\pi_C$$
$$\pi_C = \pi_C p_{CC} + \pi_S p_{SC} = 0.4\pi_S + 0.7\pi_C$$

As may be seen, the solution is

$$\pi_S = 3/7,\ \pi_C = 4/7.$$

Note that the vector π is always scaled so that $\sum_i \pi_i = 1$, as it is a probability vector.

Exercise: A rat runs around a maze as shown. Each time it changes room it picks an exit from its current room entirely at random, independently of where it has been before. Let X_n be the room it occupies after the nth room change. Find the transition matrix of X and the limit of $\mathbf{P}(X_n = \mathrm{A})$.



4.6 Irreducibility

Example: take a simple random walk with absorbing barriers at 10 and 0. The walk is sooner or later absorbed, so $\mathbf{P}(X_n = i) \to 0$ unless i is one of the absorbing states.

If X_0 is close to 10, it is more likely that absorption takes place at 10 than at 0, whereas if X_0 is close to 0 the reverse is true. The limiting value of $\mathbf{P}(X_n = 0)$ therefore depends on the value of X_0, even for large n.

A Markov chain is called *irreducible* if, starting from any one of the states, it is possible to get to any other state (not necessarily in one jump). If there are any absorbing states, the chain is not irreducible. (The reverse is not true.)

4.7 Periodicity

Consider a simple random walk with reflecting barriers above and below. If the walk starts in state 0 it will always be in an even state at even times and an odd state at odd times. In other words, $\mathbf{P}(X_n = j)$ will not converge; the effect of the starting point does not die away as $n \to \infty$.

A state i has *period* d if, given that $X_0=i$, we can only have $X_n=i$ when n is a multiple of d. In the simple random walk example all states have period 2. We call i *periodic* if it has some period > 1.

If X is irreducible, either all states are periodic or none are.

The transition matrices of periodic Markov chains have eigenvalues on the unit circle.

Example: a gambler plays roulette, staking £1 each turn. Each turn either the £1 is lost or £35 is won. X_n is the amount of money the gambler has after the nth spin. Then any state (amount of money) is periodic with period 36, as can be checked by considering $X_n \pmod{36}$.

4.8 Ergodicity

Suppose *X* is a Markov chain with only finitely many possible states. If

- X is irreducible
- X is not periodic

then X is ergodic. We have:

Theorem (*Ergodic Theorem*) If the discrete-time Markov chain X is ergodic, with transition matrix P, then there is exactly one probability vector p which satisfies

$$\mathrm{p}^T P = \mathrm{p}^T$$

In addition, for each i and j,

$$\mathbf{P}(X_n = j \mid X_0 = i) \text{ ® } \mathrm{p}_j.$$

(The proof is too advanced for this course.)

The fact that $\mathrm{p}^T P = \mathrm{p}^T$ means that p is a *stationary distribution* for X: if X_0 is not fixed but random, and $\mathbf{P}(X_0 = i) = \mathrm{p}_i$ for each i, then $\mathbf{P}(X_1 = i) = \mathrm{p}_i$ for all i, and similarly for X_2, X_3, etc.

If we can find any probability vector p which satisfies the equation, we know it is the right answer. For example, if we can find p such that

$$\mathrm{p}_i \, p_{ij} = \mathrm{p}_j \, p_{ji}$$

for all i and j, then p is the required solution. But note that the reverse is not true: there are many cases in which the equilibrium distribution p does **not** satisfy these equations.

Example: frog on lilypads.

4.9 Reducible Markov Chains

The states of any Markov chain can be grouped together. We say that any two states i and j belong to the same *communicating class* if it is possible, starting from i, to get to j and, starting from j, to get back to i.

A communicating class is classified as *recurrent* if, having entered the class, the Markov chain can never leave it, *transient* otherwise. (Note: this definition is only valid when there are finitely many states. If there are infinitely many the situation can be more complicated.)

Each absorbing state forms a c.c. on its own.

When the state space is finite the chain must eventually enter a recurrent c.c., which it can then never leave, so we can apply the Ergodic Theorem to obtain limiting probabilities.

Example: children's playground.

4.10 Modelling with Markov Chains

We need to decide in advance whether the situation being modelled is suited to a Markov model and whether the model should be time-homogeneous.

Examples

- Weather: with a time span of a day, Markov models do not fit well.
- Marital status: (see next section). Prob of getting married depends on age.
- Company health plan: sickness rates are different for older people.

But when new entrants keep the age profile stable we can aggregate individuals so that age-dependent rates are smoothed out.

Once we accept that a MC model is appropriate, it is easy to fit. Having observed x_t, $t=1, 2, ..., n$ we define n_i as the number of times the chain was in state i, n_{ij} as the observed number of transitions from i to j, then use n_{ij}/n_i as the estimate of p_{ij}.

Testing goodness of fit:

- are transitions independent of past history?
- are transition probabilities time-independent?

Both are hard to test objectively.

4.11 Simulating Markov Chains

Quite easy in Excel using lookup tables. See second lab sheet.

This page is maintained by Russell Gerrard: R.J.Gerrard@city.ac.uk