MAS275 Probability Modelling Chapter 1: Introduction and Markov chains

Dimitrios Kiagias

School of Mathematics and Statistics, University of Sheffield

Spring Semester, 2020

Stochastic processes

We study **stochastic processes**, which are families of random variables describing the evolution of a quantity with time.

In some situations, we can treat time as **discrete**.

I.e. we consider the non-negative integers $\mathbb{N}_0$, and for each $n \in \mathbb{N}_0$ we have a random variable X_n giving the value of the quantity at time n.

For example, when describing the evolution of a population, X_n could be the number of individuals in the nth generation.

Continuous time

Alternatively, it often makes more sense to treat time as **continuous**:

We represent time by the positive real numbers $\mathbb{R}^+$. For each $t \in \mathbb{R}^+$ we have a random variable X_t giving the value of the quantity of interest at time t. For example, X_t could be the price of a financial asset at time t.

For most of this course we will concentrate on discrete time, which is easier, but we will introduce some continuous time ideas towards the end of the course.

Markov property

A process going on over a period of time is said to have the **Markov property** if "given the present, the future is independent of the past", in the sense that . . .

. . . if at any time we know what the current state of the process is, then any information that we are given about what happened in the past will not affect the probability of any future event.

Markov property II

In mathematical terms, it may be written

$$P(C|A \cap B) = P(C|B)$$

whenever A represents a past event, B represents a statement of the present state of the process and C represents a future event.

The Markov property is a natural assumption in many situations, provided that the current state of the process can be appropriately described.

Markov chain

A **Markov chain** is a process with the Markov property where the set S of possible states at each time point is finite or countably infinite. The most obvious example is the integers or some subset of them, but in some contexts S is most naturally thought of as non-numeric.

S is known as the **state space**.

We will study Markov chains in discrete time only, although it is possible to develop a similar theory in continuous time.

Random variables

Let X_n and X_{n+1} be random variables, taking values in our state space S, representing the states of the process at times n and $n+1$ respectively.

Transition probabilities

To describe the behaviour of a Markov chain, we specify probabilities of the form

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

for $i, j \in S$.

Here p_{ij} is the probability, from state i, of moving to state j at the next time point, and is known as the **transition probability** from state i to state j.

Assumptions

In this course, we will assume that p_{ij} does not depend upon n: we say that the transition probabilities are **time-homogeneous**.

Note also that, because of the Markov property, we may write

$$P(X_{n+1} = j | X_n = i, \text{ previous history before time } n) = p_{ij}.$$

Transition matrix

For each ordered pair of states (i, j) there is a corresponding transition probability p_{ij}.

So we have a square two-dimensional array of numbers, known as the (one-step) **transition matrix** of the Markov chain.

Often helpful to label the rows and columns of the matrix, especially if the state space is non-numerical.

Finite and infinite matrices

Where S is a finite set, the transition matrix is a $|S| \times |S|$ matrix.

We will also consider some examples where S is an infinite set, in which case the transition matrix has to be thought of as an matrix of infinite extent.

Examples

Example

Wet and dry days

Example

Gambler's ruin

Example

Gambler's ruin with no target

Example

Ehrenfest model for diffusion

Properties of the transition matrix

Starting in state i, the process moves to exactly one state at the next time point.

So a transition matrix must have the property that each of its rows adds up to 1:

$$\sum_{j \in S} p_{ij} = 1 \text{ for each } i \in S.$$

Also, its elements are probabilities and therefore non-negative.

Stochastic matrices

Another way of saying this is that each row of the transition matrix represents a probability distribution, namely, the conditional distribution of the next state, given that the present state is i.

Any square matrix with these properties is called a **stochastic matrix**.

Eigenvector

The sum of each row being 1 implies that, if $\mathbf{1}$ is the column vector whose entries are all 1, then $P\mathbf{1} = \mathbf{1}$.

In other words a stochastic matrix always has 1 as an eigenvalue with (right) eigenvector $\mathbf{1}$.

Initial distribution

By their definition, transition probabilities are conditional probabilities.

In order to describe exactly how a process is behaving we need to specify some unconditional or absolute probabilities.

We can do this by specifying the **initial distribution** of the chain, probabilities of the form

$$\pi_i^{(0)} = P(X_0 = i)$$

for $i \in S$, where these numbers add up to 1 because they form a probability distribution.

Row vector

Note that we can write all these probabilities in a vector:

Let $\boldsymbol{\pi}^{(0)}$ be the vector (in $\mathbb{R}^{|S|}$) with entry i being $\pi_i^{(0)}$, for $i \in S$.

The usual convention in probability theory, fitting with the way we defined the transition matrix, is to treat this as a **row vector**.

Starting in a known state

Sometimes we will simply say that the chain starts at time 0 in a particular state.

To represent this, let $\boldsymbol{\pi}^{(0)}$ be the vector with entries given by

$$\pi_i^{(0)} = \begin{cases} 1 & i = i_0; \\ 0 & i \neq i_0 \end{cases}$$

which describes the chain being in i_0 with probability 1 at time 0.

Later states

To calculate probabilities involving later states, we then need to use the rules of conditional probability.

For example,

$$\begin{aligned} P(X_0 = i, X_1 = j) &= P(X_0 = i)P(X_1 = j | X_0 = i) \\ &= \pi_i^{(0)} p_{ij}. \end{aligned}$$

Markov property

To extend this to a further step, the key is to use the Markov property to simplify $P(X_2 = k | X_0 = i, X_1 = j)$: (on board)

General case

More generally, for $i_0, i_1 \ldots, i_n \in S$ we may write

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \pi^{(0)}_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

Row vectors

As above for time zero, we can write the distribution of X_n, the state of the chain at time n, as a row vector.

Let $\boldsymbol{\pi}^{(n)}$ be the vector with entry i being $\pi_i^{(n)}$, for $i \in S$.

Then $\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(0)} P$ (calculation on board).

n-step transition probabilities

For positive integer n, the n-step transition probabilities of a Markov chain are defined in the obvious way

$$p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i)$$

so that the transition probabilities which we have encountered are just the special case $n = 1$.

n-step transition matrix

As with the $n = 1$ case, we can gather these n-step transition probabilities into the form of a matrix, called the n-step transition matrix.

For now we denote it by $P^{(n)}$, but we will now see that it is related to the one-step transition matrix P in a simple way, by a set of equations known as the **Chapman-Kolmogorov equations**.

Chapman-Kolmogorov equations

Theorem

(Chapman-Kolmogorov equations)

(a) *For all positive integers m, n, we have*

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

(b) *For all $n = 1, 2, 3, \ldots,$ $P^{(n)} = P^n$.*

Corollary

We can now obtain a relationship between the row vectors representing the distribution of the state of the chain at different times.

Corollary

For all non-negative integers m and n, we have

$$\boldsymbol{\pi}^{(m+n)} = \boldsymbol{\pi}^{(m)} P^n.$$

Graphs

A **graph** here refers to a network consisting of a set of **vertices**, some pairs of which are linked by **edges**.

We assume for now that each edge can be traversed in either direction, and that there are no loops (that is, no edges where both ends are the same vertex).

Random walks on graphs

We construct a Markov chain by letting the state space S be the set of vertices, and assume that we have a particle moving from vertex to vertex.

At each step, the particle chooses one of the possible edges from its current vertex, each with equal probability, and travels along that edge.

This is called a **(symmetric) random walk on a graph**.

Example

Example

Example of a random walk on a graph

Diagonalisation

For n-step transition probabilities for large n, one approach we could consider is to try to diagonalise the transition matrix, namely to try to find an expression for P in the form

$$P = CDC^{-1}$$

where

- D is a matrix consisting of eigenvalues of P down the main diagonal and zeroes everywhere else,
- C is a matrix whose columns are corresponding right eigenvectors of P.

Diagonalisation II

If we can do this, then

$$P^n = CDC^{-1}.CDC^{-1}\ldots CDC^{-1} = CD^nC^{-1},$$

where D^n is easy to write down explicitly.

Diagonalisation of larger matrices is usually not easy...

...but it can be helped by the fact that a stochastic matrix always has 1 as an eigenvalue with $\mathbf{1}$ as the corresponding right eigenvector.

Example

Example

Diagonalisation

Left eigenvectors

A **left eigenvector** of P with eigenvalue λ is a row vector $\mathbf{x}$ such that

$$\mathbf{x}P = \lambda\mathbf{x}.$$

By thinking about the eigenvalues of a matrix and its transpose, it can be seen that the set of eigenvalues of a matrix is the same regardless of whether left or right eigenvectors are considered.

So a finite stochastic matrix always has a left eigenvector $\boldsymbol{\pi}$ with eigenvalue 1.

Stationary distributions

This means that if we choose our initial distribution $\boldsymbol{\pi}^{(0)} = \boldsymbol{\pi}$ ...

(which requires that the entries in $\boldsymbol{\pi}$ sum to 1 and that they are non-negative)

...then for any n

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi} P^n = \boldsymbol{\pi}$$

and so the (unconditional) distribution of X_n is the same as that of X_0, for any n.

Stationary distributions II

If this is the case, we say that the Markov chain is **in equilibrium**.

A distribution which, when chosen as initial distribution, causes a Markov chain to be in equilibrium is called a **stationary distribution**.

Finding stationary distributions

To find stationary distributions, we need to solve the (left) eigenvector equations

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}$$

for all $j \in S$, which are called the **equilibrium equations**.

Each equilibrium equation corresponds to a column of the transition matrix.

In the finite state space case, because we know that 1 is an eigenvalue, the equations are not linearly independent, and we may always discard one of them without losing any

Finding stationary distributions II

The eigenvector is only defined up to a constant multiple, but...

...as we are looking for a probability distribution we have the condition

$$\sum_{j \in S} \pi_j = 1,$$

which is sufficient to guarantee a unique solution in the case where the eigenvalue 1 has multiplicity 1.

Short cuts

In practice, the solution of these equations is often helped by the fact that (in many examples) there are a lot of zeros in the transition matrix.

When we know that the solution is unique (which, as we will see later in the course, is very often the case) it is also often valuable to appeal to symmetry to argue that some of the π_j's are equal to each other, thereby enabling us to eliminate some variables from the equations.

Example

Example

Stationary distribution for random walk on a graph

Infinite state space

In the infinite state space case, the same equations apply, but our vectors and matrices are now infinite in extent, so we have an infinite family of equations, each of which may have an infinite sum on the right hand side.

Such families of equations may not have a solution, and even if one does exist it may not be easy to find.

However there are some examples of infinite state Markov chains with stationary distributions which can be found by exploiting the structure of the equations.