B.Sc./Grad. Dip.: Probability Models and Time Series
MAS programmes: Stochastic Models and Time Series

1 Markov Chains: Part I

1.1 Introduction

This set of lectures will be concerned with the study of a certain type of *stochastic process* that evolves in discrete time, takes values within a discrete state space, and which adheres to a characteristic known as the *Markov condition* or *Markov property*.

There are plenty of applications for Markov chain models including: position after a certain number of goes in a board game like "snakes and ladders"; the amount of water in a dam at 12 noon each day; the number of people waiting in a queue at a supermarket checkout just before/after a customer has been served/has arrived.

The key characteristic of Markov chain models is that given the present state of the process, the future evolution of the process is independent of the past!

1.2 Notation and Model Framework

Let $\{X_0, X_1, X_2, \ldots\}$ be a sequence of random variables which take values in some discrete set $\mathbb{S}$ called the *state space*. Call the sequence $\mathbf{X}$ for short. $\mathbb{S}$ may consist of either a finite set of values or may be countably infinite.

Let $N = |\mathbb{S}|$ be the cardinality or the size of the state space; $\mathbb{Z}^+$ be the positive integers; and $\mathbb{N}$ be the natural numbers which are here defined as $\{0\} \cup \mathbb{Z}^+$.

So here is the mathematical characterization of the kinds of processes we intend to study in these lectures.

Definition 1.2.1 (Markov chain)
The process $\{X_n : n \in \mathbb{N}\}$ is a Markov chain if it satisfies the Markov condition

$$\mathbb{P}(X_n = s | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s | X_{n-1} = x_{n-1})$$

for all $n \geq 1$ and $s, x_0, x_1, \ldots, x_{n-1} \in \mathbb{S}$.

Remarks 1.2.2 (language)
If $X_n = i$, then we say that the chain is in the i-th state at the n-th step or stage.

The evolution of the chain depends on the so-called transition probabilities given by

$$\mathbb{P}(X_{n+1} = j | X_n = i).$$

This depends on states i, j, and stage n. For simplicity, we will restrict attention and only consider processes whose transition probabilities do not depend on n.

Definition 1.2.3 (homogeneous chain)
• The chain $\mathbf{X}$ is called *homogeneous* if

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

for all n, i, j.

• The corresponding *transition matrix* $\mathbf{P}$ is the $N \times N$ matrix whose (i, j)-th element, p_{ij} is given by

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

KEY ASSUMPTIONS
Assume that:
• all Markov chains are homogeneous unless otherwise stated;
• $\mathbf{X}$ is a Markov chain;
• $\mathbf{P}$ is the transition matrix of $\mathbf{X}$

By virtue of the fact that the elements of the transition matrix are (conditional) probabilities, then the matrix itself will have some properties which are inherent to probability mass functions.

Theorem 1.2.4
$\mathbf{P}$ is a *stochastic matrix* i.e.
(a) $p_{ij} \geq 0$ for all i, j indexing the matrix;
(b) each row of $\mathbf{P}$ sums to 1, i.e. $\sum_j p_{ij} = 1$.

Just as the transition matrix $\mathbf{P}$ characterizes the probabilistic behaviour of the chain from one stage to the next, we could also consider the probabilistic behaviour of the chain between epochs that are seperated by n stages.

Definition 1.2.5 (n-step transition matrix)
The n-step transition matrix $\mathbf{P}^{(n)}$ is the matrix whose (i, j)-th element is given by the n-step transition probability

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

Remarks 1.2.6

• Clearly $\mathbf{P}^{(1)} = \mathbf{P}$.
• $\mathbf{P}$ characterizes the short term evolution of $\mathbf{X}$ whereas $\mathbf{P}^{(n)}$ characterizes it over a longer time scale (n stages in fact).

It is not hard to be believe that we could study the behaviour of the evolution of the chain between epochs separated by n stages by looking at the problem over smaller chunks of time, and then combining the results in an appropriate manner. The next theorem tells us how to do this.

Theorem 1.2.7 (Chapman-Kolmogorov equations)

$$p_{ij}(m+n) = \sum_k p_{ik}(m)p_{kj}(n).$$

In other words,

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

It follows that $\mathbf{P}^{(n)} = \mathbf{P}^n$ (i.e. the n-th power of $\mathbf{P}$).

Proof

$$p_{ij}(m+n) = \mathbb{P}(X_{m+n} = j | X_0 = i)$$
$$= \frac{\mathbb{P}(X_{m+n} = j, X_0 = i)}{\mathbb{P}(X_0 = i)} = \frac{\sum_k \mathbb{P}(X_{m+n} = j, X_m = k, X_0 = i)}{\mathbb{P}(X_0 = i)}$$
$$= \sum_k \frac{\mathbb{P}(X_m = k, X_0 = i)}{\mathbb{P}(X_0 = i)} \mathbb{P}(X_{m+n} = j | X_m = k, X_0 = i)$$
$$= \sum_k \mathbb{P}(X_m = k | X_0 = i)\mathbb{P}(X_{m+n} = j | X_m = k)$$
$$= \sum_k \mathbb{P}(X_m = k | X_0 = i)\mathbb{P}(X_n = j | X_0 = k)$$

where the penultimate line follows from the Markov property and the final line follows from time-homogeneity. □

So, in fact, the transition matrix over a long time scale can be formulated entirely from the transition matrix $\mathbf{P}$ for evolution through a single step by taking appropriate powers of $\mathbf{P}$.

Furthermore, if we know the mass function of the chain at a particular stage, and we know the transition matrix $\mathbf{P}$, then it is straightforward to deduce the mass function for any later stage. Thus, again, this relates long-term behaviour to that of the short term.

Definition 1.2.8 (mass function at stage n)

Let $\pi_i^{(n)} = \mathbb{P}(X_n = i)$ and set $\boldsymbol{\pi}^{(n)}$ to be the row vector whose i-th component, $i \in \mathbb{S}$, is $\pi_i^{(n)}$. Thus $\boldsymbol{\pi}^{(n)}$ represents the mass function of $\mathbf{X}$ at stage n.

Lemma 1.2.9

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

and hence

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)}\mathbf{P}^n.$$

Proof

$$\pi_j^{(m+n)} = \mathbb{P}(X_{m+n} = j) = \sum_i \mathbb{P}(X_{m+n} = j | X_m = i)\mathbb{P}(X_m = i)$$
$$= \sum_i \pi_i^{(m)} p_{ij}(n) = (\boldsymbol{\pi}^{(m)}\mathbf{P}^{(n)})_j.$$

Thus the evolution of the chain can be characterized in terms of the initial mass function $\boldsymbol{\pi}^{(0)}$ and the transition matrix $\mathbf{P}$. □

It is perhaps worth considering what could be said about the mass functions at any stage if, in some sense, the chain has settled down to some kind of *equilibrium behaviour*. If the chain were known to exhibit its equilibrium behaviour at stage 0, then in what sense would the mass function at stage 0 have any kind of similarity to the mass function at stage n? We'll look into this in a later chapter.

1.3 Examples

Example 1.3.1 (simple random walk)
Imagine a 'random walker' who, at each stage, can either take one step to the right (or in the positive direction) with probability p, or one step to the left (or in the negative direction) with probability q. The walk could represent, for example, money in the bank or value of an economic entity which can take movements up or down over a period of discrete epochs in time.

This scenario can be modelled as a Markov chain, whereby the state space is given by
• $\mathbb{S} = \{\ldots, \ldots, -2, -1, 0, 1, 2, \ldots, \ldots\} = \mathbb{Z}$
and the one step transition probabilities are given by
•

$$p_{ij} = \begin{cases} p & \text{if} \quad j = i+1 \\ q & \text{if} \quad j = i-1 \\ 0 & \quad \text{otherwise} \end{cases}$$

where $0 < p < 1$ and $q = 1 - p$.

It can be shown[1] (see Grimmett & Stirzaker for e.g.) that the n-step transition probability $p_{ij}(n)$ is given by

$$p_{ij}(n) = \begin{cases} \binom{n}{\frac{1}{2}(n+j-i)} p^{\frac{1}{2}(n+j-i)} q^{\frac{1}{2}(n-j+i)} & \text{if } n+j-i = 2r;\ r = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}.$$

To get some intuition into how these random walks behave, we can simulate some realizations of the Markov chain $\mathbf{X}$, i.e. the *sample paths* (which we can think of as being typical trajectories).

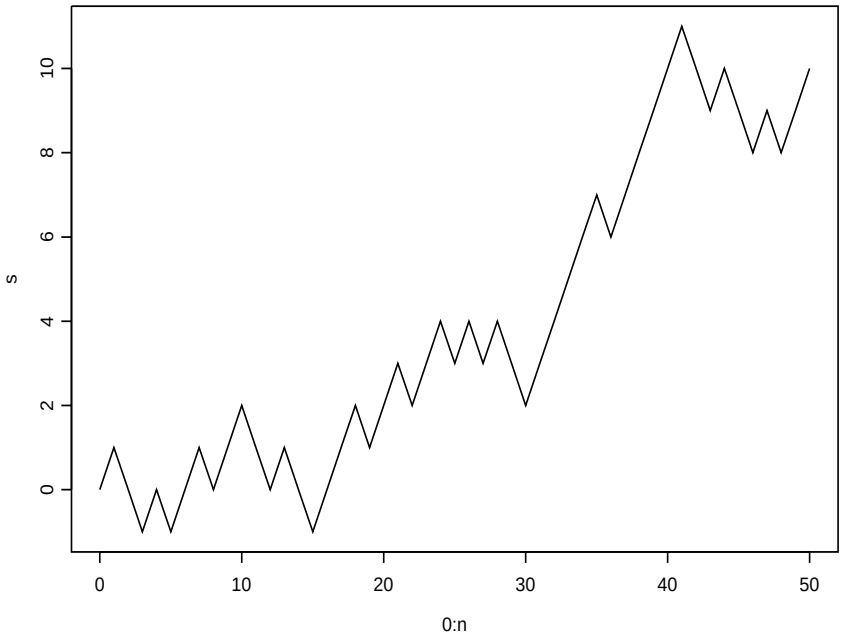
```
n=50
p=0.55
s=0
u.values=runif(0:n)
jumps=1*(u.values<=p)-1*(u.values>p)
for (i in 1:n)
  s=c(s,s[length(s)]+jumps[i])
plot(0:n,s, type="l")
```

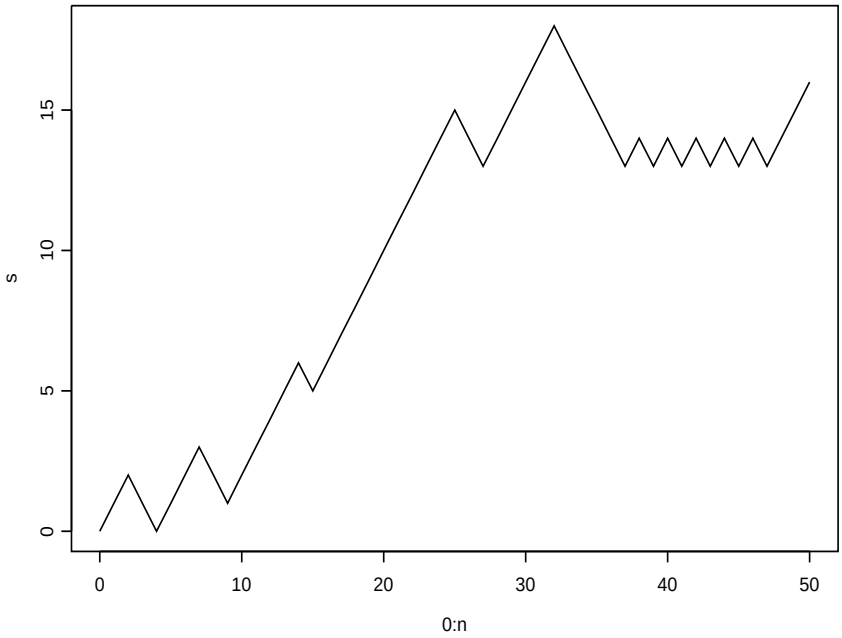
[1] Calculation is not examinable!

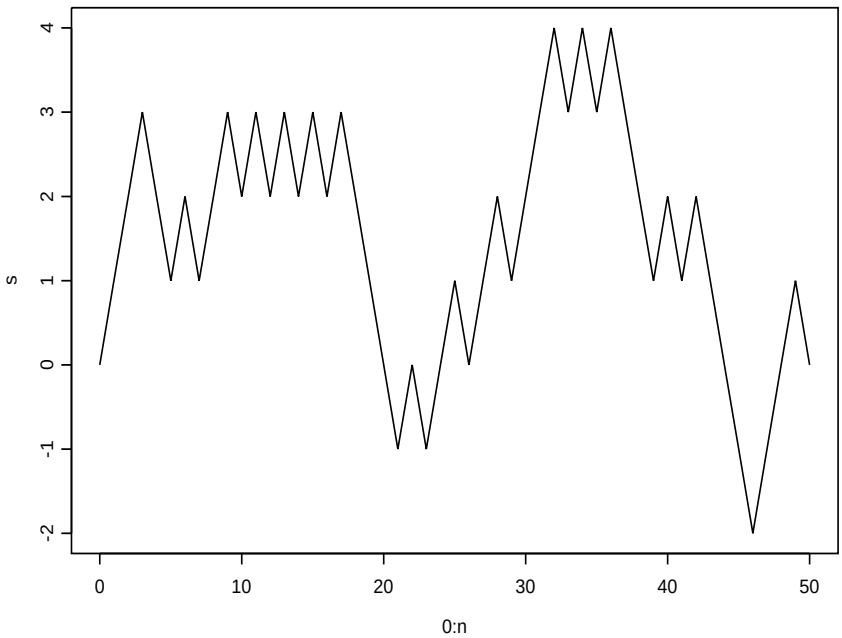
In the above code, the realization is constructed from stage 0 up to stage n. The starting value is zero at stage 0, i.e. $s_0 = 0$. The jump size is simulated such that it is equal to $+1$ with probability p and -1 with probability $1 - p$. The new position of the walker is appended to the vector containing all other previous positions of the walker [somewhat in the same way that one would update a *singly linked list* in a computer language like Pascal].

In the examples below, $p = 0.55$, thus there is a slight bias towards positive or rightward jumps rather than negative or leftward jumps.

[If we were to set $p = 0.5$, then this would correspond to the *symmetric random walk*].




s
0
5
10
15
0
10
20
30
40
50
0:n


s
-2
-1
0
1
2
3
4
0
10
20
30
40
50
0:n

Example 1.3.2 (inventory model)
• Consider a supermarket stocking certain sized boxes of washing powder.
• Let ξ_n be the demand for the n-th week;
• Weekly demands for boxes of the washing powder, $\{\xi_n\}$, are independent and identically distributed such that

$$\mathbb{P}(\xi_n = j) = a_j,\ j \geq 0 \quad \text{for all } n.$$

• Suppose that the supermarket uses an (s, S) ordering policy. If at start of the week, current supply is s boxes or higher, then no order is made. If number of boxes is less than s, then supermarket orders enough to bring the supply back up to S boxes.
Hence, if supply is x, then order

$$\begin{cases} 0 & \text{if} \quad x \geq s \\ S - x & \text{if} \quad x < s \end{cases}$$

i.e. $(S - x)\mathbf{1}\{x < s\}$ where $\mathbf{1}\{\cdot\}$ is the indicator function.
• Assume that all unfulfilled customer demands are lost.
• Let X_n be the inventory supply at the end of the n-th week.

Then $\{X_n\}$ is a Markov chain with

$$p_{ij} = \begin{cases} \sum_{k=i}^{\infty} a_k & \text{if} \quad j = 0, & i \geq s \\ a_{i-j} & \text{if} \quad 0 < j \leq i, & i \geq s \\ \sum_{k=S}^{\infty} a_k & \text{if} \quad j = 0, & i < s \\ a_{S-j} & \text{if} \quad 0 < j \leq S, & i < s \\ 0 & \text{otherwise} \end{cases}.$$

Example 1.3.3 (Bernoulli process and circular buffering)
• A circular buffer is a data structure that uses a single, fixed-size buffer as if it were connected end to end;
• Used in the buffering of data streams; e.g. producer-consumer problem in which the producer could be an audio generator and the consumer could be the sound card.
• At each stage n, producer provides an item of data or does not. If it does, then it is added to the current location in the buffer, then moves to the next location; otherwise it stays in the same position in the buffer.
• Let $\mathbb{S} = \{0, 1, 2, \ldots\}$ and define the Markov chain $\{Y_n\}$ by

$$Y_1 = 0, \quad \mathbb{P}(Y_{n+1} = s + 1 | Y_n = s) = p, \quad \mathbb{P}(Y_{n+1} = s | Y_n = s) = 1 - p$$

for $n = 1, 2, \ldots$ and $0 < p < 1$.

Clearly

$$p_{ij}(m) = \mathbb{P}(Y_{n+m} = j | Y_n = i) = \binom{m}{j-i} p^{j-i}(1-p)^{m-(j-i)}, \ \ i, j \in \mathbb{N}, \ \ 0 \leq j - i \leq m.$$

• Suppose that the buffer has exactly B positions. Then the position of the "producer" in the buffer at start of the n-th time slot is

$$X_n = (Y_n \text{ modulo } B) + 1.$$

Thus $\{X_n : n \geq 0\}$ is a Markov chain on the state space $\mathbb{S}' = \{1, 2, \ldots, B\}$ with transition matrix

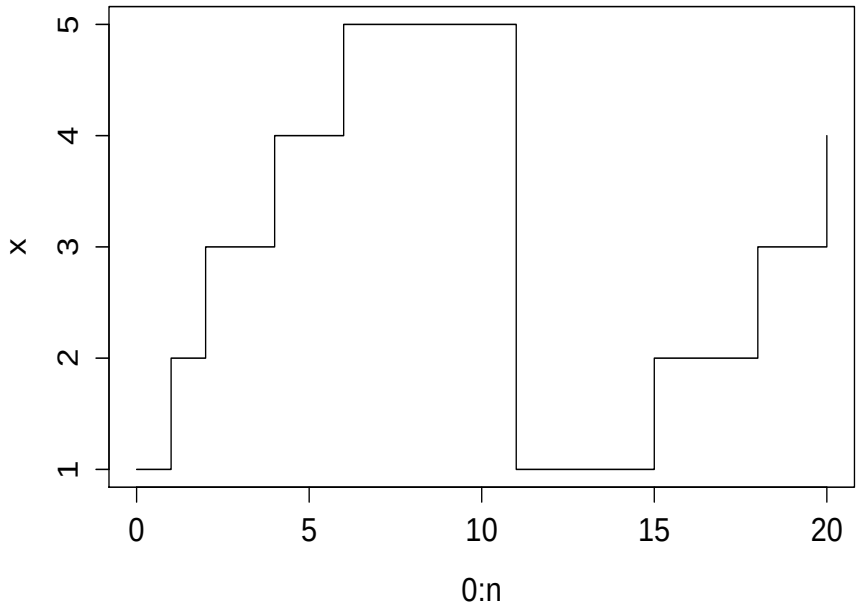
$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 & 0 & \ldots & \ldots & \ldots & \ldots & 0 & 0 \\ 0 & 1-p & p & 0 & \ldots & \ldots & \ldots & \ldots & 0 & 0 \\ 0 & 0 & 1-p & p & \ldots & \ldots & \ldots & \ldots & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & & & & \vdots & \vdots \\ \vdots & \vdots & & & \ddots & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & & & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & & & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & & & & & \ddots & \ddots & \\ 0 & 0 & \ldots & \ldots & \ldots & 0 & 0 & 0 & 1-p & p \\ p & 0 & \ldots & \ldots & \ldots & \ldots & 0 & 0 & 0 & 1-p \end{pmatrix}$$

Let us consider a simulation of this system over 20 time steps, with a buffer of size 5, which is initially empty, and where there is a 50% chance that an item of data is placed at the current position of the provider within the buffer at each stage.

```
n=20
p=0.5
y=0
B=5
u.values=runif(1:n)
jumps=1 * (u.values <= p) + 0 * (u.values > p)
for(i in 1:n)
  y = c(y, y[length(y)] + jumps[i])
x=(y - B * trunc(y/B)) + 1
plot(0:n, x, type = "s")
```

Just by way of corroboration, we note the following sets of numbers for `y` and `x`:

```
> y
 [1] 0 1 2 2 3 3 4 4 4 4 4 5 5 5 5 6 6 6 7 7 8
> x
 [1] 1 2 3 3 4 4 5 5 5 5 5 1 1 1 1 2 2 2 3 3 4
```


x
1
2
3
4
5
0
5
10
15
20
0:n