

STAT2003 Introduction to Applied Probability

STAT3102 Stochastic Processes

Lecture notes

Course details

- Course organiser: Dr. Elinor Jones (elinor.jones@ucl.ac.uk).
- Lectures: 9-11am TUESDAYS and 3-4pm FRIDAYS.
- Office hours: 1.30-2.30pm TUESDAYS and 1.30-2.30pm FRIDAYS. If neither time is convenient, please email me to arrange another slot. You can find me in Room 140, 1-19 Torrington Place.
- Useful books:
 - ‘Introduction to Probability Models’ by S. Ross
 - ‘Probability and Random Processes’ by Grimmett & Stirzaker (more mathematical).

Tutorials

- Exercises will be uploaded to Moodle each Friday afternoon.
- You must hand in your work by 12 noon the following THURSDAY. Please post your work in your tutor's box in the Undergraduate Common Room. Make sure that your name and tutor group is clearly printed on your work. There is no need to attach a copy of the homework sheet.
- Late submissions will not be marked, and **scanned or emailed copies will not be considered**.

Assessment

- ICA : Friday 4th of March 3-4pm (provisional)- open book, 45 minutes. This will be 10% of your final mark.
- Final exam in Term 3 - closed book, 2 or 2.5 hours (for STAT3102 and STAT2003 respectively). This will be 90% of your final mark.
- There is no choice of questions in the ICA or final written exam.
- College approved calculators only (but you will need one!).

Contents

1	Introduction (Important!)	4
1.1	A course overview	4
1.2	Mathematical/ Statistical Writing	4
1.3	Random variables	5
1.3.1	Definitions (optional)	5
1.3.2	Discrete random variables	7
1.3.3	Continuous random variables	7
1.3.4	Important distributions	8
1.3.5	Generating functions	8
1.3.6	Probability generating functions	9
1.3.7	Moment generating functions	10
1.3.8	What else can we do with generating functions?	12
1.4	Conditioning on events	13
1.4.1	Conditional probability	15
1.4.2	Conditional expectation	15
1.4.3	Useful conditioning formulae	17
1.4.4	The idea of ‘first step decomposition’	18
1.5	Exercises	22
2	What is a stochastic process?	23
2.1	Definitions and basic properties	23
2.2	The Markov property	24
2.3	Why is the Markov property useful?	25
2.4	How can I tell if a process is Markov?	25
3	Discrete-time Markov processes	29
3.1	Introduction to discrete time Markov chains	29
3.1.1	Time homogeneity and transition probabilities	29
3.1.2	Transition matrices	29
3.1.3	The n-step transition probabilities and the Chapman-Kolmogorov equations	30
3.1.4	Marginal probabilities	32
3.1.5	Duration of stay in a state	33
3.2	Classification of Markov chains	34
3.2.1	Irreducible classes of intercommunicating states	34
3.2.2	Recurrence and transience	36
3.2.3	Positive recurrence and null recurrence	38
3.2.4	Periodicity	38
3.2.5	Class properties	38
3.2.6	Further properties of discrete time Markov chains	41
3.2.7	Closed classes and absorption	42
3.2.8	Decomposition of the states of a Markov chain	44
3.3	Limiting behaviour	46
3.3.1	Invariant distributions	46
3.3.2	Equilibrium distributions	47
3.3.3	Invariant and equilibrium distributions	48
3.3.4	Ergodic Markov chains and the equilibrium distribution	49
3.3.5	Extending the Main Limit Theorem	51
3.3.6	When does an equilibrium distribution exist?	52

4	Continuous-time Markov processes	55
4.1	Continuous-time Markov chains	55
4.2	The importance of the exponential distribution and ‘little-o’ notation	56
4.2.1	Order notation: $o(h)$	56
4.2.2	The lack-of-memory property of the exponential distribution	57
4.3	Breaking down the definition of a continuous time Markov process	58
4.3.1	Holding times	58
4.3.2	The jump chain	59
4.4	Analysis of transition probabilities	61
4.4.1	Transition probabilities over a small interval	62
4.4.2	Transitions over longer periods: the Chapman-Kolmogorov equations	63
4.4.3	Kolmogorov’s forward equations	64
4.4.4	Kolmogorov’s backward equations	66
4.4.5	Solving the KFDEs and KBDEs	67
4.4.6	The generator matrix, Q	68
4.5	Limiting behaviour	69
4.5.1	Invariant distributions	69
4.5.2	Equilibrium distributions	69
4.5.3	A limit theorem for continuous time Markov chains	70
5	Important types of continuous-time processes	72
5.1	The Poisson process	72
5.1.1	The generator matrix for the Poisson process	73
5.1.2	The distribution of the number of events by time t	74
5.1.3	Other properties of the Poisson process	75
5.1.4	Superposition and thinning of a Poisson process	78
5.2	Birth and death processes	81
5.2.1	Transition rates and the jump chain	82
5.2.2	The generator matrix	83
5.2.3	Important types of birth-death processes	84
5.2.4	Transition probabilities (over a small interval)	85
5.2.5	The equilibrium distribution of the (general) birth and death process	86
5.2.6	The probability of extinction for linear birth-death process	88
5.3	Exercises	90

1 Introduction (Important!)

1.1 A course overview

A **stochastic process** is a process which evolves randomly over time (or space, or both):

- ‘stochastic’: equivalent to ‘random’.
- ‘process’: occurring over time (or space, or both).

We will consider some simple stochastic processes which are used to model real-life situations. Our main focus will be on developing mathematical tools to study their properties and learn about long-term behaviour of such processes. Though this is a mathematical course, the applications of stochastic processes are wide-ranging, e.g. in physics, engineering, the environment, social science, epidemics, genetics and finance. We’ll see examples from some of these areas during the course.

1.2 Mathematical/ Statistical Writing

Mathematics is a language for which you must learn its grammar. Whether you are handing in a tutorial sheet or writing an exam, you need to present your work in an appropriate manner. You could start by thinking of the following:

- Think carefully about precisely what you are trying to calculate/ solve and what information you have available to you in order to do so.
- Have you clearly defined all your notation? Be *precise*: are your definitions watertight?
- Is your notation appropriate?
- Does your solution flow logically, that is, does each step follow on logically from the previous step? Again, be *precise*. The examiner is not a mind reader. YOU need to guide the reader through your thought process.
- Have you explained your reasoning, not just written down some formulae?
- It is important to *understand* what you are doing; don’t just repeat a list of steps. Always ask yourself: ‘do I know what I’m doing?’ and ‘how would I explain this to a fellow student?’.

Being able to write clearly will help you to **think clearly**. It forces you to understand the material: if you can’t write it well, you don’t understand it well enough. Learning mathematics is not about repetition, as it was at A-Level, and so past papers are NOT A GOOD REVISION TOOL. In particular, I write ICA and exam questions that probe your understanding, and not your ability to regurgitate methods. This reflects the fact that proper mathematics is not about learning to apply a set of methods, but using logical reasoning to find your way through a problem.

Some tips for writing in STAT2003/STAT3102 (based on previous cohorts of students):

- Random variables form the basis for most, if not every, question you will encounter. Make sure you define each and every random variable you use. All random variables should be assigned to CAPITAL letters.
- Probabilities should be written out in full - if you want to calculate $P(X = 1)$ for some random variable X , then write it as such. NEVER use notation such as $P(1)$ to denote this. And, as the previous bullet point states, writing $P(x = 1)$ is plain wrong.
- The ‘equals sign’ has a very specific meaning. Please don’t abuse it.

- Do a reality check on your answer. For example, a probability should be between 0 and 1. An expectation of how many time units it takes to reach a particular point should be at least 1 (assuming moving between locations takes 1 unit of time, and that you're not already there!).
- If your answer doesn't contain written explanation, then it is not a complete solution. Think about guiding the marker through your thought process, rather than expecting the marker to second guess what you meant to say.
- Use tutorial sheets to practice good mathematical writing.

1.3 Random variables

Informally speaking, a random variable is a variable [i.e. can take different values], whose numerical values are randomly chosen from a set of possible values, governed by a probability distribution. For example, the time that a passenger must wait until the next southbound tube train at Goodge street station is a variable. Now suppose you arrive at Goodge street station - the amount of time you must wait until the next southbound tube arrives is a realisation of the random variable 'waiting time'.

It is important that you can spot a random variable, and can also define appropriate random variables for a given scenario.

1.3.1 Definitions (optional)

Ideally, we should think about random variables as functions. In order to do so, we must first build up some terminology. We'll use a coin tossing example to illustrate: suppose you toss a coin twice and record whether it lands heads or tails at each toss. Note that the following is a simplification of ideas from Probability Theory.

- An **experiment** is any situation where there is a set of possible outcomes.
 - Our coin-tossing example is an experiment.
 - The possible outcomes are HH, HT, TH or TT.
- The set of possible outcomes is called the **sample space**, Ω .
 - In our case, $\Omega = \{HH, HT, TH, TT\}$
- An **outcome** is denoted ω and is an element of the sample space.
 - Suppose we conduct the experiment, and the outcome is TH. This is our ω .
- An **event**, A , is a subset of the sample space.
 - e.g. define an event as 'at least one head'.
 - This is satisfied by HH, HT and TH.
 - This is a subset of Ω .

Probability model	Real world
Sample space, Ω	‘Set of all possible outcomes of the experiment.’
Sample point, ω	‘Outcome of the experiment.’
Event $A \subseteq \Omega$	‘Property that the outcome of the experiment may possess.’

With a finite sample space Ω , we can assign probabilities to individual sample points ω via a ‘weight function’, $p : \Omega \rightarrow [0, 1]$. This allocates a value $p(\omega)$ for each possible outcome ω , which we interpret as the probability that ω occurs. We need

$$\sum_{\omega \in \Omega} p(\omega) = 1,$$

then we may calculate the probability of an event $A \subseteq \Omega$ via

$$P(A) = \sum_{\omega \in A} p(\omega) = 1.$$

Caution: this doesn’t always work with infinite sample spaces, but we won’t consider this further.

Notice that:

1. $P(A) \geq 0$ for all events A ,
2. if A_1, A_2, A_3, \dots are events with $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ i.e. P is countably additive,
3. $P(\Omega) = 1$.

As the outcome of an experiment corresponds to a sample point $\omega \in \Omega$, we can make some numerical measurements whose values depend on the outcome ω . This gives rise to a *random variable*, let’s call it X , which is a function $X : \Omega \rightarrow \mathbf{R}$. Its value at a sample point, $X(\omega)$, represents the numerical value of the measurement when the outcome of the experiment is ω .

In our set-up, there are different random variables you could define. Let’s define X to be the number of heads in the two tosses of a coin. In this case X maps Ω to 0 (if we see TT), 1 (if we see TH or HT) or 2 (if we see HH).

$$\begin{aligned} X(TT) &= 0 \\ X(TH) &= X(HT) = 1 \\ X(HH) &= 2 \end{aligned}$$

All outcomes are equally likely here, and so $p(\omega) = 1/4$ for all outcomes ω . If we define our event $A = \text{“two heads”}$, then since the only sample point in A is HH,

$$\sum_{\omega \in A} p(\omega) = 1/4,$$

and we say that $P(A) = 1/4$.

Similarly, we can also define another event $B = \text{“at least one tail”}$, and this is given by $P(X \leq 1)$, which we can calculate as:

$$P(B) = P(X \leq 1) = P(\{\omega \in \Omega : X(\omega) \leq 1\}) \quad (1)$$

Which ω satisfy (1) in this case? Now you should be able to calculate the probability required.

Note: we have defined the **distribution function** of X above. In general, the **distribution function** of X of a random variable is given, for $x \in \mathbf{R}$, by:

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

1.3.2 Discrete random variables

X is a **discrete random variable** if it takes only finitely many or countably infinitely many values.

The **probability mass function** of a discrete random variable is

$$p_X(x) = P(X = x) = P(\{\omega : X(\omega) = x\}).$$

The **expectation** of X is

$$E(X) = \sum_x x P(X = x).$$

For a function g we have

$$E(g(X)) = \sum_x g(x) P(X = x).$$

The variance of X is

$$\text{var}(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2.$$

1.3.3 Continuous random variables

X is a **continuous random variable** if it has a **probability density function**, i.e., if there exists a non-negative function f_X with $\int_{-\infty}^{\infty} f_X(x) dx = 1$, such that for all x ,

$$F_X(x) = \int_{-\infty}^x f_X(u) du.$$

Then

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

The expectation of X is

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

and the expectation of $g(X)$, for some function g , is

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

provided that these integrals exist. As before, the variance of X is

$$\text{var}(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2.$$

1.3.4 Important distributions

There are several probability distributions which we will use repeatedly during the course. Make sure you are very familiar with the following:

Discrete distributions					
Distribution	Notation	Support	PMF	Expectation	Variance
Binomial	$X \sim \text{Bin}(n, p)$	$0, 1, \dots, n$	$P(X = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$ $k = 0, 1, \dots, n; q = 1 - p$	np	npq $q = 1 - p$
Poisson	$X \sim \text{Po}(\lambda)$ where $\lambda \geq 0$	$0, 1, 2, 3, \dots$	$P(X = k) = \exp(-\lambda) \frac{\lambda^k}{k!}$ $k = 0, 1, 2, \dots$	λ	λ
Geometric	$X \sim \text{Geom}(p)$ where $p \in [0, 1]$	$1, 2, 3, 4, \dots$	$P(X = k) = (1 - p)^{k-1} p$ $k = 1, 2, 3, \dots$	$1/p$	q/p^2 where $q = 1 - p$

Continuous distributions					
Distribution	Notation	Support	PDF	Expectation	Variance
Exponential	$X \sim \text{Exp}(\lambda)$ where $\lambda > 0$	\mathbf{R}^+ (positive reals)	$f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$	$1/\lambda$	$1/\lambda^2$
Uniform	$X \sim \text{Unif}(a, b)$	$[a, b]$	$f_X(x) = \begin{cases} 1 & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{1}{2}(b - a)$	$\frac{1}{12}(b - a)^2$
Gamma	$X \sim \text{Gamma}(\alpha, \beta)$	\mathbf{R}^+ (positive reals)	$f_X(x) = \frac{\beta^\alpha \exp(-\beta x) x^{\alpha-1}}{\Gamma(\alpha)}$ $x > 0$	α/β	α/β^2

1.3.5 Generating functions

Generating functions are USEFUL. Don't think of them as baffling mathematical constructs; instead think of them as marvelous time-saving devices.

Imagine you have a sequence of real numbers, z_1, z_2, z_3, \dots . It can be hard to keep track of such numbers and all the information they contain. A generating function provides a way of wrapping up the entire sequence into one expression. When we want to extract information about the sequence, we simply interrogate this one expression in a particular way in order to get what we want.

There are many different types of generating function. These are different in the sense that we wrap up the sequence in different ways, and so extracting particular pieces of information from them requires different techniques. The two types of generating function that you should be familiar with are:

1. The probability generating function (PGF);
2. The moment generating function (MGF)

1.3.6 Probability generating functions

The **probability generating function (p.g.f)** of X is defined to be

$$G_X(s) = E(s^X) = \sum_{k=0}^{\infty} s^k P(X = k) \quad \text{for } |s| \leq 1.$$

Basic requirement(s)	The random variable X takes values in $\mathbf{N}_0 = \{0, 1, 2, \dots\}$
Sequence to ‘wrap up’	$P(X = 0), P(X = 1), P(X = 2), P(X = 3), \dots$
Method	$\sum_{k=0}^{\infty} s^k P(X = k)$ (for any $s \in [-1, 1]$)
Information stored	<ul style="list-style-type: none"> · The probabilities $P(X = 0), P(X = 1), P(X = 2), \dots$ · Some moments (expectations).

The information stored in probability generating functions can be accessed as follows.

1. **Calculating moments.**

$$E(X) = \left| \frac{dG_X(s)}{ds} \right|_{s=1} = G'_X(1),$$

$$E[X(X-1)] = \left| \frac{d^2 G_X(s)}{ds^2} \right|_{s=1} = G''_X(1), \text{ and so on.}$$

2. **Calculating probabilities.** Suppose that X takes values in $\mathbf{N}_0 = \{0, 1, 2, \dots\}$.

(a)

$$G_X(0) = \sum_{k=0}^{\infty} 0^k P(X = k) = P(X = 0).$$

- (b) If we expand $G_X(s)$ in powers of s the coefficient of s^k is equal to $P(X = k)$, so we can also find $P(X = k)$ for all k .

Example 1.1

You can also work ‘backwards’: given a PGF for a random variable X , what is the distribution of X ?

For example, suppose you are told that $G_X(s) = ps/[1 - (1 - p)s]$. We can convert this expression into the summation format of a PGF and spot its distribution:

$$\begin{aligned} G_X(s) &= \frac{ps}{1 - (1 - p)s}, \\ &= ps \sum_{k=0}^{\infty} (1 - p)^k s^k \\ &= \sum_{j=1}^{\infty} s^j (1 - p)^{j-1} p, \quad \text{where } j = k + 1. \end{aligned}$$

Therefore, $P(X = j) = (1 - p)^{j-1}p$, so $X \sim \text{Geometric}(p)$.

In fact, different distributions have different PGF formats but random variables with the same distribution have a PGF of the same format. You can therefore spot the distribution of a random variable instantly from the format of its PGF.

Exercise 1.2

Find the format of the following PGFs, either by calculating the PGF directly or looking in a text book!

Distribution of X	Probability generating function, $G_X(s)$
$X \sim \text{Geometric}(p)$	
$X \sim \text{Poisson}(\lambda)$	
$X \sim \text{Binomial}(n, p)$	
$X \sim \text{Bernoulli}(p)$	

1.3.7 Moment generating functions

The **moment generating function (m.g.f)** of X is defined to be

$$M_X(s) = E[\exp(tX)] = \int_{\mathbf{R}} \exp(tx) f_X(x) dx \quad \text{for any } t \in \mathbf{R} \text{ such that the integral converges.}$$

Basic requirement(s)	A real valued random variable
Quantity to ‘wrap up’	$f_X(x)$
Method	$\int_{\mathbf{R}} \exp(tx) f_X(x) dx$ (for any $t \in \mathbf{R}$ such that the integral converges)
Information stored	Moments

We can access all moments of X from $M_X(t)$:

$$E[X^n] = M_X^{(n)}(0) = \left. \frac{d^n M_X}{dt^n} \right|_{t=0}$$

The same as for PGFs, you can spot the distribution of a random variable instantly from the format of its MGF.

Exercise 1.3

Find the format of the following PGFs, either by calculating the PGF directly or looking in a text book!

Distribution of X	Probability generating function, $G_X(s)$
$X \sim \text{Normal}(\mu, \sigma^2)$	
$X \sim \text{Exponential}(\lambda)$	
$X \sim \text{Binomial}(n, p)$	
$X \sim \text{Bernoulli}(p)$	
$X \sim \text{Geometric}(p)$	

Notice that a discrete or continuous random variable can have a MGF, whereas a PGF is for discrete random variables only (i.e. those with countable state space).

1.3.8 What else can we do with generating functions?

We'll concentrate here on (useful) things you can do with **probability** generating functions, though similar conclusions can be reached using moment generating functions too.

1. Calculating the distribution of sums of random variables.

Suppose that X_1, \dots, X_n are i.i.d. random variables with common PGF $G_X(s)$. The PGF $G_Y(s)$ of $Y = X_1 + \dots + X_n$ is given by

$$\begin{aligned} G_Y(s) &= \mathbf{E}(s^{X_1 + \dots + X_n}), \\ &= \prod_{i=1}^n \mathbf{E}(s^{X_i}), \\ &= [G_X(s)]^n, \end{aligned}$$

and by looking at $[G_X(s)]^n$, we can spot the distribution of Y , which may otherwise be difficult. Also, if $Z_1 \perp\!\!\!\perp Z_2$ (that is, Z_1 and Z_2 are independent but do not necessarily have the same distribution) then

$$G_{Z_1+Z_2}(s) = G_{Z_1}(s)G_{Z_2}(s).$$

By looking at the PGF of $Z_1 + Z_2$, $G_{Z_1+Z_2}(s)$, we can then deduce the distribution $Z_1 + Z_2$. This can be an easier strategy than trying to deduce its distribution directly.

Example 1.4

X and Y are both Poisson random variables, with parameters λ_1 and λ_2 , respectively. Assume that X and Y are also independent of each other. What is the distribution of $X + Y$?

Solution: Use probability generating functions!

The PGF of a Poisson random variable with parameter λ is

$$G(s) = e^{\lambda(s-1)}.$$

Therefore, the PGF of $X + Y$ is

$$G_{X+Y}(s) = e^{\lambda_1(s-1)}e^{\lambda_2(s-1)} = e^{(\lambda_1+\lambda_2)(s-1)}.$$

We recognise the format of this PGF as the PGF of a Poisson distribution with parameter $(\lambda_1 + \lambda_2)$. Therefore,

$$X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

2. Calculating the PGF of a random sum (that is, the sum of a *random* number of random variables).

Suppose that X_1, \dots, X_N are i.i.d. random variables with common PGF $G_X(s)$ and that N has PGF $G_N(s)$.

The PGF $G_Y(s)$ of $Y = X_1 + \dots + X_N$ is given by

$$G_Y(s) = \mathbf{E}\left[\mathbf{E}(s^Y \mid N)\right].$$

We'll see this technique of expanding an expectation again later (you have already seen it if you studied STAT2001 or STAT3101), and it will prove to be a very useful technique in solving problems. Using this, we can deduce that

$$\mathbf{E}\left[\mathbf{E}(s^Y \mid N)\right] = \mathbf{E}\left[G_X(s)^N\right] = G_N[G_X(s)].$$

Similarly, it can be shown (challenge!) that the m.g.f. $M_Y(t)$ of Y is given by

$$M_Y(t) = \mathbf{E}(\exp^{tY}) = G_N[M_X(t)].$$

Example 1.5

The number of people who enter a particular bookstore on Gower Street, per day, follows a Poisson distribution with parameter 300. Of those who visit, 60% will make a purchase. What is the distribution of the number of people who make a purchase at the bookstore in one day?

Solution: Use probability generating functions!

Notice that whether a visitor makes a purchase is a Bernoulli trial, with probability of success 0.6. Let

$$X_i = \begin{cases} 1 & \text{if } i\text{th customer makes a purchase} \\ 0 & \text{otherwise} \end{cases}$$

so that we are looking to find the distribution of $Y = X_1 + \dots + X_N$ where $N \sim \text{Poisson}(300)$. Notice that the PGF of the Bernoulli random variable X is $G_X(s) = (0.4 + 0.6s)$, and the PGF of N is $e^{300(s-1)}$.

By the result above, we know that the PGF of Y is $G_Y(s) = G_N(G_X(s))$. Think of this as using $G_X(s)$ as the 'in place of' s in the PGF of N :

$$\begin{aligned} G_Y(s) &= G_N(G_X(s)) \\ &= e^{300(0.4+0.6s-1)} = e^{300(0.6s-0.6)} = e^{180(s-1)} \end{aligned}$$

We recognise this as the PGF of a Poisson random variable with mean 180. Therefore, the distribution of those who make a purchase is Poisson(180).

1.4 Conditioning on events

Calculating probabilities that are conditional on a particular event or occurrence is a natural requirement. Make sure you understand the rationale behind conditional probability, don't just learn the formulae!

Easy exercise

You have a deck of playing cards (52 cards in total, split equally into four suits: hearts, diamonds, spades and clubs, with the first two considered 'red' cards, and the others considered 'black' cards).

- I pick a card at random. What is the probability that the chosen card is from the diamond suit?
- I pick a card at random and tell you that it is a red card (i.e. either hearts or diamonds). What is the probability that the chosen card is from the diamond suit?

The second question requires a conditional probability (it is conditional on knowing that the card is red), but you most likely calculated the probability intuitively without using any formulae. How did you do that? You applied Bayes's theorem intuitively, without even realising it.

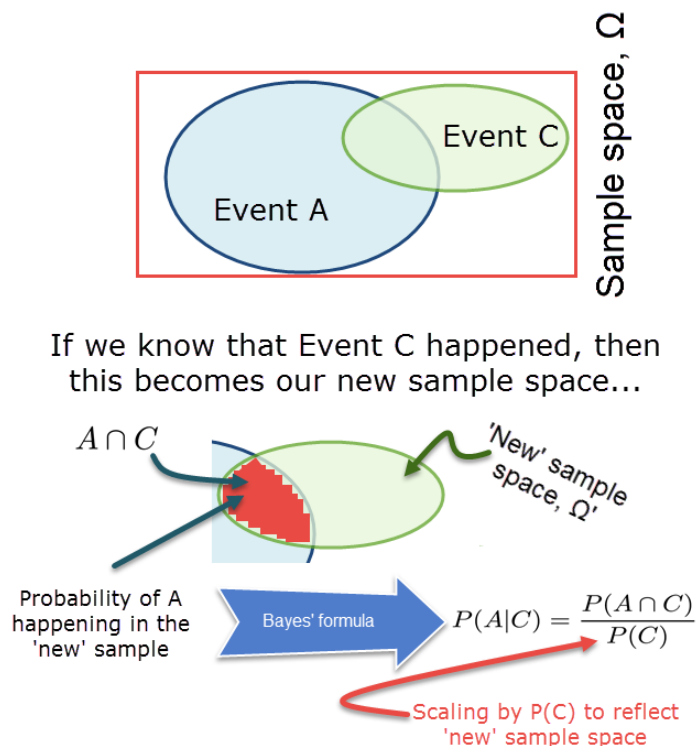
In fact, Bayes's theorem for calculating conditional probabilities is perfectly intuitive. Suppose we are calculating $P(A|C)$ for some events A and C :

- This is asking for the probability of A occurring, given that C occurs.
- Assuming C occurs, then the contents of the statespace Ω which relate to C becomes the 'new' statespace, Ω' (which is just the event C for our purposes).
- What we are asking here, in effect, is to calculate the probability of A in the 'new' sample space Ω' (i.e. $P(A \cap C)$), scaled to Ω' (i.e. divided by $P(C)$).
- The final equation is therefore $P(A|C) = P(A \cap C)/P(C)$

Why scale by $P(C)$?

Simply because the probability of all disjoint events in any sample space must sum to 1. Our sample space is now $\Omega' = C$, so the new (probability) weight function we apply to all events in Ω' must sum to 1. We scale all probabilities by $P(C)$ to ensure that this is the case.

The following graphic may help you to visualise why Bayes's theorem 'works'.



1.4.1 Conditional probability

Let A and C be events with $P(C) > 0$. The **conditional probability of A given C** is

$$P(A|C) = \frac{P(A \cap C)}{P(C)}.$$

It is easy to verify that

1. $P(A|C) \geq 0$,
2. if A_1, A_2, \dots are mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \dots | C) = \sum_i P(A_i | C),$$

(try drawing diagrams like the above to visualise this)

3. $P(C|C) = 1$.

Conditional probability for random variables instead of events works is the same way.

Discrete RVs X and Y	Continuous RVs X and Y
Joint probability mass function: $p_{X,Y}(x, y) = P(X = x, Y = y).$	Joint density function: $f_{X,Y}(x, y)$
X and Y independent if and only if for all x, y : $p_{X,Y}(x, y) = p_X(x)p_Y(y) = P(X = x)P(Y = y).$	X and Y independent if and only if for all x, y : $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
Marginal probability mass function (of X): $p_X(x) = P(X = x) = \sum_y P(X = x, Y = y).$	Marginal density (of Y): $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx.$
Conditional PMF of X given $Y = y$ $p_{X Y=y}(x) = P(X = x Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$	Conditional density of X given $Y = y$ ($f_Y(y) > 0$) $f_{X Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$
Conditional distribution function of X given $Y = y$ $F_{X Y=y}(x) = P(X \leq x Y = y)$	Conditional distribution function of X given $Y = y$ $F_{X Y=y}(x) = P(X \leq x Y = y)$

1.4.2 Conditional expectation

Just like calculating an expectation of, say, X , requires knowing the distribution of X , the conditional expectation of X , given that we know $Y = y$ requires knowing the distribution of $(X|Y = y)$.

The **conditional expectation of X given $Y = y$** is

$$E(X|Y = y) = \begin{cases} \sum_x xP(X = x | Y = y) & \text{for discrete RVs} \\ \int_{-\infty}^{\infty} x f_{X|Y=y}(x) \, dx & \text{for continuous RVs} \end{cases}$$

The **conditional expectation of $g(X)$ given $Y = y$, for some function g** , is

$$E(g(X)|Y = y) = \begin{cases} \sum_x g(x)P(X = x | Y = y) & \text{for discrete RVs} \\ \int_{-\infty}^{\infty} g(x)f_{X|Y=y}(x) \, dx & \text{for continuous RVs} \end{cases}$$

Example 1.6

Suppose that $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $P(\omega_i) = 1/3$ for $i = 1, 2, 3$. Suppose also that X and Y are random variables with $X(\omega_1) = 2$, $X(\omega_2) = 3$, $X(\omega_3) = 1$, and $Y(\omega_1) = 2$, $Y(\omega_2) = 2$, $Y(\omega_3) = 1$. Find the conditional pmf $p_{X|Y=2}(x)$ and the conditional expectation $E(X | Y = 2)$.

Possible values of X are 1, 2, 3.

$$p_{X|Y=2}(1) = P(X = 1 | Y = 2) = 0.$$

$$p_{X|Y=2}(2) = P(X = 2 | Y = 2) = \frac{P(X=2, Y=2)}{P(Y=2)} = \frac{P(\omega_1)}{P(\omega_1)+P(\omega_2)} = \frac{1}{2}.$$

$$p_{X|Y=2}(3) = P(X = 3 | Y = 2) = \frac{P(X=3, Y=2)}{P(Y=2)} = \frac{P(\omega_2)}{P(\omega_1)+P(\omega_2)} = \frac{1}{2}.$$

Thus the conditional pmf of X given $Y = 2$ equals $\frac{1}{2}$ for $x = 2$ and for $x = 3$ and is 0 otherwise. The conditional expectation is

$$E(X | Y = 2) = \sum_x x P(X = x | Y = 2) = (1 \times 0) + (2 \times 1/2) + (3 \times 1/2) = 5/2.$$

An important concept

So far, we have only calculated conditional expectations when conditioning on *specific values*, e.g. $E[X|Y = 2]$. We can extend this idea to conditioning on random variables without equating them to specific values.

The notation $E(X | Y)$ is used to denote a **random variable** that takes the value $E(X | Y = y)$ with probability $P(Y = y)$.

Note that $E(X | Y)$ is a **function** of the random variable Y , and is itself a random variable.

Example 1.7

In Example 1.6 we found that $E(X | Y = 2) = \frac{5}{2}$. It can similarly be shown that $E(X | Y = 1) = 1$. You can also check that $P(Y = 1) = 1/3$ and $P(Y = 2) = 2/3$. Thus, the **random variable** $E(X | Y)$ has two possible values 1 and $\frac{5}{2}$, with probabilities $\frac{1}{3}$ and $\frac{2}{3}$, respectively. That is,

$$E[X|Y] = \begin{cases} 1 & \text{with probability } 1/3 \\ 5/2 & \text{with probability } 2/3 \end{cases}$$

1.4.3 Useful conditioning formulae

Three important formulae:

- (i) **Law of total probability** Let A be an event and Y be ANY random variable. Then

$$P(A) = \begin{cases} \sum_y P(A|Y=y)P(Y=y) & \text{if } Y \text{ discrete} \\ \int_{-\infty}^{\infty} P(A|Y=y)f_Y(y) \, dy & \text{if } Y \text{ continuous} \end{cases}$$

- (ii) **Modification of (i):** if Z is a third discrete random variable,

$$\begin{aligned} P(X=x | Z=z) &= \sum_y P(X=x, Y=y | Z=z) \\ &= \sum_y \frac{P(X=x, Y=y, Z=z)}{P(Z=z)} \\ &= \sum_y \frac{P(X=x, Y=y, Z=z)}{P(Y=y, Z=z)} \frac{P(Y=y, Z=z)}{P(Z=z)} \\ &= \sum_y P(X=x | Y=y, Z=z) P(Y=y | Z=z), \end{aligned}$$

assuming the conditional probabilities are all defined. Note the similarity with (i).

- (iii) **Law of conditional (iterated) expectations**

$$E[X] = E[E(X|Y)] = \begin{cases} \sum_y E(X|Y=y)P(Y=y) & \text{if } Y \text{ discrete} \\ \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y) \, dy & \text{if } Y \text{ continuous} \end{cases}$$

and, more generally,

$$E[g(X)] = E(E[g(X)|Y])$$

Note in particular:

- (i) The law of total probability applies to ANY event A and ANY random variable Y .
- (ii) The law of the conditional (iterated) expectation, $E[X] = E[E[X|Y]]$ applies to ANY random variables X and Y , and we will be using this identity heavily during the course.

Example 1.8

In Example 1.7 we found the distribution of the random variable $E(X | Y)$ was

$$E[X|Y] = \begin{cases} 1 & \text{w.p. } 1/3 \\ 5/2 & \text{w.p. } 2/3 \end{cases}$$

The expectation of this random variable is

$$E[E(X|Y)] = 1 \times \frac{1}{3} + \frac{5}{2} \times \frac{2}{3} = 2.$$

Also, the expectation of X is given by

$$E(X) = 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2,$$

so that we see that the law of the conditional (iterated) expectation, i.e. that $E[X] = E[E[X|Y]]$, holds.

Example 1.9

Let X and Y have joint density

$$f_{X,Y}(x,y) = \frac{1}{y} e^{-x/y} e^{-y}, \quad 0 < x, y < \infty.$$

Find $f_Y(y)$, $E(X | Y)$ and hence $E(X)$.

Solution

For any $y > 0$,

$$f_Y(y) = \int_0^\infty \frac{1}{y} e^{-x/y} e^{-y} \, dx = e^{-y} [-e^{-x/y}]_0^\infty = e^{-y},$$

so Y has an exponential distribution with parameter 1.

Also, for any fixed $y > 0$ and any $x > 0$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{e^{-x/y} e^{-y}}{y e^{-y}} = \frac{1}{y} e^{-x/y},$$

so $X | Y = y \sim \exp(1/y)$.

It follows that $E(X | Y = y) = y$ and so $E(X | Y) = Y$. Using the Law of Conditional Expectations we then have

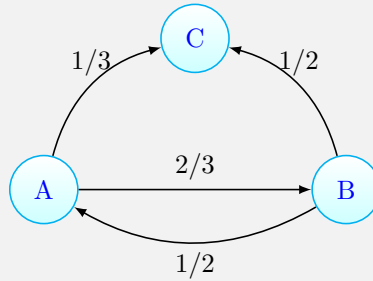
$$E[X] = E[E(X | Y)] = E(Y) = 1.$$

1.4.4 The idea of ‘first step decomposition’

The ‘first step decomposition’ technique is a very useful method for computing certain types of expectations and probabilities. We will be using this method extensively throughout the course.

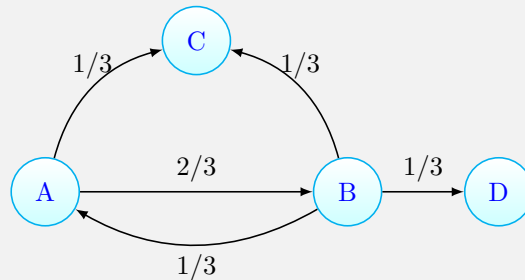
Example 1.10

Roads connecting three locations in a city are shown in the following diagram, the time taken to cycle along any of the roads being 1 minute. Geraint is a bicycle courier whose office is located at location A. He has a parcel that he needs to deliver to location C. However, he is a law abiding cyclist who will only follow the direction of travel permitted on each road, as indicated by the arrows. Whenever he comes to a junction, he selects any of the routes available to him with probabilities as given in the diagram below.



- (i) How long, on average, will it take Geraint to deliver the parcel?
- (ii) How many times, on average, will Geraint visit location B before delivering the parcel?

Now suppose that another location, D, is added to the mix, but that all couriers arriving at D are stuck there indefinitely.



- (iii) With what probability does the parcel never get delivered?

Some initial comments on these type of questions:

- (i) How long, on average, will it take Geraint to deliver the parcel?
 - If there is a positive probability of never getting to location C, then the expectation will be infinite.
 - In the first diagram, the parcel will be delivered eventually as in this case $P(\text{Parcel not delivered by time } n) \rightarrow 0$ as $n \rightarrow \infty$.
 - For this question, the minimum possible time taken is 1 minute (if he goes directly from A to C). Therefore the expectation should be greater than 1 (this is a handy reality check at the end of such questions).
- (ii) How many times, on average, will Geraint visit location B before delivering the parcel?
 - Such expectations must be positive as Geraint will visit location B with positive probability.
 - This expectation cannot be negative! Again, this is a good reality check.

- This expectation can be anything in the interval $(0, \infty)$. Compare this with the previous question.
- (iii) For the updated map, with what probability does the parcel never get delivered?
- This must be between 0 and 1. This may be obvious, but I have known (good) students to forget that they are calculating a probability and give either a negative answer or something greater than 1.

There are many ways of solving these questions, but a convenient method is to use ‘first step decomposition’: think about ‘where do we go first’?

First step decomposition: what is it, and why is it useful?

This technique can be adapted to solve many different problems, and it will be used frequently in this course. In fact, all three questions posed about Geraint’s cycling trip can be solved using the general idea of first step decomposition.

The idea is to decompose the process on the basis of where the process goes first (or its first step, hence the name). This, combined with the iterated law of conditional expectation (when we want to compute an expectation), or the law of total probability (for computing probabilities), proves to be a rather powerful method.

Typical questions which can be solved via first step decomposition include:

1. How long does it take, on average, to reach a particular point for the first time?
2. How often, on average, do we visit a certain point before a particular event occurs?
3. What’s the probability that we ever reach/ never reach a particular point?

Example 1.10 *continued*

How long, on average, will it take Geraint to deliver the parcel?

We'll apply first step decomposition to solve this question. Let X_n denote Geraint's location at time n , and let T denote the time it takes for Geraint to deliver the parcel to location C.

We need to calculate $E[T]$, and we will do this by conditioning on Geraint's next move. In the first instance, he starts in location A, so $X_0 = A$. Using the iterated law of conditional expectation, we get:

$$E[T] = E[E[T|X_1]] = E[T|X_1 = B]P(X_1 = B) + E[T|X_1 = C]P(X_1 = C).$$

Since $X_1 = B$ with probability $2/3$ and $X_1 = C$ with probability $1/3$, we get:

$$E[T] = \frac{2}{3}E[T|X_1 = B] + \frac{1}{3}E[T|X_1 = C].$$

Now notice that $E[T|X_1 = C] = 1$ because we are conditioning on Geraint reaching location C for the first time at time 1, so this simplifies to

$$E[T] = \frac{2}{3}E[T|X_1 = B] + \frac{1}{3}. \quad (2)$$

Now compute $E[T|X_1 = B]$ by applying first step decomposition again: from B, where can Geraint go next?

$$E[T|X_1 = B] = E[E[T|X_2, X_1 = B]] = \frac{1}{2}E[T|X_2 = C, X_1 = B] + \frac{1}{2}E[T|X_2 = A, X_1 = B].$$

Now notice that $E[T|X_2 = C, X_1 = B] = 2$ and $E[T|X_2 = A, X_1 = B] = 2 + E[T]$ (the latter because Geraint used 2 minutes going to B and then back to A, and once back in A it is as if the process starts over from scratch).

Therefore $E[T|X_1 = B] = 1 + \frac{1}{2}(2 + E[T])$. Substituting this back in to (2), we have:

$$E[T] = \frac{2}{3}E[T|X_1 = B] + \frac{1}{3} = \frac{1}{3}E[T] + \frac{5}{3}.$$

Solving gives $E[T] = 2.5$ minutes.

You may want to consider simplifying the notation of the solution to Example 1.8 above, by defining new quantities. For example, let L_i denote the expected time to reach C, given that we start in state i . Then, the solution above simplifies to

$$L_A = 1 + \frac{2}{3}L_B + \frac{1}{3}L_C. \quad (3)$$

We must add 1 since we are using one minute to move from state A to another state. Now we know that $L_C = 0$ (if we start in C, it takes no time to get there!), and $L_B = 1 + L_A/2 + L_C/2 = 1 + L_A/2$. Therefore, substituting into (3), we need to solve

$$L_A = 1 + \frac{2}{3}(1 + L_A/2) = \frac{1}{3}L_A + \frac{5}{3}.$$

Solving gives $L_A = 2.5$ minutes, as before.

Exercise 1.11

Use a similar method to solve parts (ii) and (iii) of Example 1.10.

1.5 Exercises

The following problems can be solved by a variety of methods. Try to solve them using conditioning arguments.

Exercise 1.12

A coin, with probability of heads equal to p , is tossed repeatedly until the first head is obtained. What is the expected number of throws required?

Exercise 1.13

Two UCL students are trapped in a corridor in Torrington Place after 7 p.m. The corridor has three exits. The first exit leads out of the building in 2 minutes, the second leads back to the original corridor after 5 minutes and the third is a dead end that leads back after 1 minute. Assume that the students are trying to get out of the building, but are so engrossed in a statistical discussion that they fail to learn from experience, and they choose an exit at random each time they return to the corridor.

- (a) Show that the students are certain to get out of the building eventually.
- (b) Find the expected length of time until they get out.
- (c) Find the variance of the time until they get out.
- (d) Find the probability generating function of the time until they get out.

Exercise 1.14

The number of claims arriving at an insurance company in a week has a Poisson distribution with mean λ . For any claim form, there is a probability p that it is completed incorrectly, independently of other claim forms and of the number of claims.

- (a) Find the distribution of the number of forms completed incorrectly in a week.
- (b) Deduce the distribution of the number of forms completed incorrectly in a year.
- (c) The sizes of the claims are independent random variables X_1, X_2, X_3, \dots that have the same distribution, and that are independent of the number of claims. Let T be the total amount claimed in one week (including claims on incorrect forms). Obtain formulae for the expectation and variance of T .

2 What is a stochastic process?

2.1 Definitions and basic properties

Definition: Stochastic process

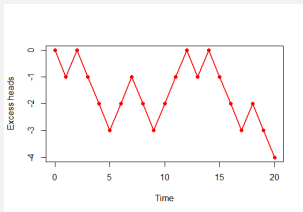
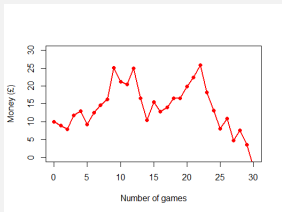
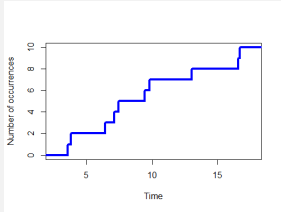
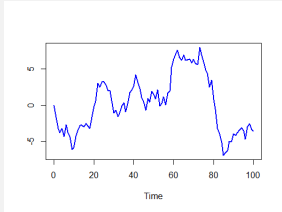
A **stochastic process** is a collection of random variables $\{X_t, t \in T\}$ taking values in the *state space* S . The parameter, or index, set T often represents time.

Notation

- Discrete-time processes: $\{X_t, t \in T\}$, where $T = \{0, 1, 2, \dots\}$. This is often written in terms of the natural numbers (including zero) rather than T , i.e. $\{X_n, n \in \mathbf{N}\}$, where $\mathbf{N} = \{0, 1, 2, \dots\}$
- Continuous-time processes: $\{X(t), t \in T\}$, where $T = \mathbf{R}_+$ (i.e. $t \geq 0$).
- Notice that the *state space* can be discrete or continuous. In this course we will only consider discrete state spaces.

Examples of stochastic processes.

Below are examples of stochastic processes, categorised according to whether they are continuous or discrete time processes, and whether the state space is discrete or continuous. We only consider discrete state space processes in this course, in both discrete and continuous time.

		STATE SPACE	
		Discrete	Continuous
TIME	Discrete	<p>How many more heads than tails in total?</p>  <p>‘Time’ here is coin tosses</p>	<p>A gambler’s winnings and losses after each bet</p>  <p>‘Time’ here is bet number</p>
	Continuous	<p>Arrival times</p>  <p>Process increases by 1 at each arrival</p>	<p>Brownian motion</p>  <p>Used to model stock prices</p>

NOTE: Continuous-time processes can change their value/state ('jump') at **any** instant of time; discrete-time processes can only do this at a discrete set of time points. For discrete-time processes, when we use the word 'time' we mean 'number of transitions/steps'.

2.2 The Markov property

Definition: Markov process

A stochastic process $\{X_t, t \in T\}$ is a *Markov* process if, for any sequence of times $0 < t_1 < \dots < t_n < t_{n+1}$ and any $n \geq 0$,

$$P(X_{t_{n+1}} = j \mid X_{t_n} = i_n, \dots, X_{t_0} = i_0) = P(X_{t_{n+1}} = j \mid X_{t_n} = i_n),$$

for any j, i_0, \dots, i_n in S .

The Markov property - an intuitive explanation.

In other words, the Markov property states that the value (or *state*) of the process at time t_{n+1} depends only on its value (state) at time t_n , and not on its behaviour previous to time t_n . That is, the probability of any future event depends only on the most recent information we have about the process.

Let

- X_n be the present state;
- A be a past event (involving X_0, \dots, X_{n-1});
- B be a future event (involving X_{n+1}, X_{n+2}, \dots)

The Markov property (MP) states that given the present state (X_n), the future B is conditionally independent of the past A , denoted

$$B \perp\!\!\!\perp A \mid X_n.$$

Therefore,

$$P(A \cap B \mid X_n = i) = P(A \mid X_n = i) P(B \mid X_n = i).$$

Similarly, on applying the Markov property,

$$P(B \mid X_n = i, A) = P(B \mid X_n = i),$$

for any $n \geq 0$, any i and any A and B . Notice here that we can use the Markov property because the most recent information about the chain, its location at time n , is given exactly as $X_n = i$.

An important note about the Markov property.

Think carefully about the definition of the Markov property. It states that

$$P(X_{t_{n+1}} = j \mid \mathbf{X}_{\mathbf{t}_n} = \mathbf{i}_n, \dots, X_{t_0} = i_0) = P(X_{t_{n+1}} = j \mid X_{t_n} = i_n).$$

Notice the term I highlight above. We can only use the Markov property if the most recent information (highlighted in bold) gives us specific information about the location of the process. We could not apply the Markov property if, for example, we wanted to simplify

$$P(X_{t_{n+1}} = j \mid \mathbf{X}_{\mathbf{t}_n} = \mathbf{i}_n \text{ or } \mathbf{j}_n, \dots, X_{t_0} = i_0),$$

since the most recent information (the location of the process at time t_n) does not tell us where exactly the Markov chain is at time t_n .

2.3 Why is the Markov property useful?

We'll be using the Markov property repeatedly throughout the course, so make sure you are familiar with it! The Markov property is a strong independence assumption. *It is useful because it simplifies probability calculations.* It enables joint probabilities to be expressed as a product of simple conditional probabilities.

Example 2.1

Suppose that X is a (discrete time) Markov process taking values 0, 1 and 2. Write

$$P(X_0 = 1, X_1 = 2, X_2 = 0)$$

as a product of conditional probabilities, simplifying your answer as much as possible, justifying each step of your argument.

$$\begin{aligned} P(X_0 = 1, X_1 = 2, X_2 = 0) &= P(X_2 = 0 \mid X_1 = 2, X_0 = 1) P(X_1 = 2, X_0 = 1) && \text{Bayes's theorem} \\ &= P(X_2 = 0 \mid X_1 = 2) P(X_1 = 2, X_0 = 1) && \text{Markov property} \\ &= P(X_2 = 0 \mid X_1 = 2) P(X_1 = 2 \mid X_0 = 1) P(X_0 = 1), && \text{Bayes's theorem} \end{aligned}$$

Therefore, $P(X_0 = 1, X_1 = 2, X_2 = 0) = P(X_2 = 0 \mid X_1 = 2) P(X_1 = 2 \mid X_0 = 1) P(X_0 = 1)$.

Exercise 2.2

Make sure you understand how to apply the Markov property (repeatedly!) by answering the following.

Suppose that X is a (discrete time) Markov process taking values in some statespace S . Show that

$$P(X_0 = i_0, \dots, X_n = i_n) = \prod_{k=1}^n P(X_k = i_k \mid X_{k-1} = i_{k-1}) P(X_0 = i_0).$$

This is a very useful identity which we'll be making use of throughout the course.

2.4 How can I tell if a process is Markov?

You will come across many ICA and exam questions asking you to explain whether or not a particular process is Markov. There are two ways that you could establish whether a process is Markov.

1. Without using any mathematics, deduce that the process is/ isn't Markov logically.
2. Checking (mathematically) that the Markov property holds for all states in the statespace at all times.

The second strategy is particularly hard, so we mainly use this technique when we suspect that the Markov property does not hold by finding a counter example which does not satisfy the Markov property. To show that the Markov property does not hold, find one set of states a, b, c and perhaps d such that **one** of the following holds (you do not need to show both hold - one will do!):

- $P(Y_{n+1} = a \mid Y_n = b, Y_{n-1} = c) \neq P(Y_{n+1} = a \mid Y_n = b)$ for some a, b, c ;

- $P(Y_{n+1} = a | Y_n = b, Y_{n-1} = c) \neq P(Y_{n+1} = a | Y_n = b, Y_{n-1} = d)$ for some a, b, c, d with $c \neq d$. This shows that the state of the process at time $(n - 1)$ affects the state of the process at time $(n + 1)$, violating the Markov property.

Exploiting features of the process to decide whether it is Markov

Many questions that you will encounter will provide you with a process $\{X_n, n = 0, 1, 2, \dots\}$ which is known to be Markov, and will define a further process $\{Y_n, n = 0, 1, 2, \dots\}$ as a function of the original Markov process. The question is usually whether the ‘new’ process is also a Markov process. That is, if $\{X_n, n = 0, 1, 2, \dots\}$ is a Markov process, and $\{Y_n, n = 0, 1, 2, \dots\}$ is a function of this Markov process, then does the Markov property hold for $\{Y_n, n = 0, 1, 2, \dots\}$ too?

In this kind of situation, it is often expected (or even necessary) to take a mathematical approach to proving or disproving that the Markov property holds in the ‘new’ process. Two strategies to keep in mind:

- Is the function in question bijective (one-to-one)? If it is, then the new process will also be Markov. (Think carefully as to why this has to be the case!).
- If the function in question is not bijective, then the new process may, or may not, be Markov. In this case, if you are trying to prove that the new process is *not* Markov, a good strategy is to exploit the fact that you do not have uniqueness.

We’ll see these strategies being used in the next example.

Example 2.3.

Let X_0, X_1, X_2, \dots be a sequence of independent random variables such that for $n = 0, 1, 2, \dots$, we have $P(X_n = +1) = P(X_n = -1) = \frac{1}{2}$. For $n = 1, 2, 3, \dots$ define:

$$Y_n = \frac{X_{n-1} + X_n}{2} \quad ; \quad Z_n = \frac{X_{n-1} + 2X_n}{3}.$$

- (i) Establish that X_0, X_1, \dots is a Markov process.
- (ii) Is $\{Y_n, n = 0, 1, 2, \dots\}$ a Markov process?
- (ii) Is $\{Z_n, n = 0, 1, 2, \dots\}$ a Markov process?

Solution:

- (i) It is easy to show that ANY sequence of i.i.d. random variables satisfies the Markov property (try it), therefore part (i) is complete.
- (ii) The state space for $\{Y_n, n = 0, 1, 2, \dots\}$ is $S = \{-1, 0, 1\}$. Notice that
 - $Y_n = -1$ only if $X_{n-1} = -1$ and $X_n = -1$;
 - $Y_n = 1$ only if $X_{n-1} = 1$ and $X_n = 1$;
 - $Y_n = 0$ if $X_{n-1} = -1$ and $X_n = 1$ OR $X_{n-1} = 1$ and $X_n = -1$;

The transformation function from X to Y is not bijective because there is ambiguity in the value of the pair (X_n, X_{n-1}) when $Y_n = 0$. Let's try to show that the Markov property does not hold by finding a counter example. After a trial-and-error approach I find a counter example, which proves that the Markov property does not hold (note: this is not the only counter example in this case):

$$\begin{aligned} P(Y_{n+1} = 1 | Y_n = 0, Y_{n-1} = -1) &= 0.5 \\ P(Y_{n+1} = 1 | Y_n = 0, Y_{n-1} = 1) &= 0 \end{aligned}$$

Therefore, $\{Y_n, n = 0, 1, 2, \dots\}$ is not a Markov process.

Notice that my strategy here was to let $Y_n = 0$: this prompts ambiguity in the value of the original process at time n , and hence lets me construct a counter example.

- (iii) The state space for $\{Z_n, n = 0, 1, 2, \dots\}$ is $S = \{-1, -1/3, 1/3, 1\}$. Notice that
 - $Z_n = -1$ only if $X_{n-1} = -1$ and $X_n = -1$;
 - $Z_n = -1/3$ only if $X_{n-1} = 1$ and $X_n = -1$;
 - $Z_n = 1/3$ only if $X_{n-1} = -1$ and $X_n = 1$;
 - $Z_n = 1$ only if $X_{n-1} = 1$ and $X_n = 1$.

Therefore, whatever the value of Z , we can deduce exactly the values of the relevant random variables X : we have a one-to-one relationship between pairs of X variables and values for Z . We must therefore conclude that $\{Z_n, n = 0, 1, 2, \dots\}$ is a Markov process.

Exercise 2.4.

Try the following exercises to make sure that you are comfortable in establishing whether the Markov property holds for the following processes.

Let X_0, X_1, X_2, \dots be a sequence of independent random variables such that for $n = 0, 1, 2, \dots$, we have $P(X_n = 0) = 0.75$ and $P(X_n = 1) = 0.25$. For $n = 1, 2, 3, \dots$ define

$$Y_n = (X_n + X_{n-1})^2$$

and $Y_0 = 0$. State whether or not the stochastic process $\{Y_n, n = 0, 1, 2, \dots\}$ is a Markov chain. If it is a Markov chain, state why it must be so. If the process is not a Markov chain, then give an example where the Markov property breaks down.

DNA consists of four nucleotide bases - A, G, C, and T.

Suppose that at each loci (location) along a DNA strand:

- Given the base at loci n , the probability that it will not change at loci $n + 1$ is 0.5.
- If the base at loci $n + 1$ is different from loci n , the base at loci $n + 1$ is chosen randomly from the remaining bases.

Let Z_n be the base at loci n . Is $\{Z_n\}$ a Markov chain? [HINT: No maths required! Deduce that the process must satisfy the Markov property logically.]

3 Discrete-time Markov processes

3.1 Introduction to discrete time Markov chains

Definition: Discrete-time Markov chain

A discrete-time *Markov chain*, often abbreviated to *Markov chain*, is a sequence of random variables X_0, X_1, X_2, \dots taking values in a finite or countable state space S such that, for all $n, i, j, i_0, i_1, \dots, i_{n-1}$

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i),$$

that is, a discrete-time stochastic process satisfying the Markov property.

3.1.1 Time homogeneity and transition probabilities

Definition: Time homogeneous Markov chain

A Markov chain is *time-homogeneous* if

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

does not depend on n for all $i, j \in S$.

For a time homogeneous Markov chain, we have

$$P(X_{n+1} = j \mid X_n = i) = P(X_{m+1} = j \mid X_m = i)$$

even if $m \neq n$. In effect, this allows us to forget about **when** the chain moved from i to j , all that matters is that this occurred in **one time step**. When a chain is time homogeneous, we also have that for any integer $r \geq 1$,

$$P(X_{n+r} = j \mid X_n = i) = P(X_{m+r} = j \mid X_m = i).$$

That is, the probability of moving from state i to state j in r time steps is the same *regardless of when this happened*. In this course, we will assume that all discrete time Markov processes are time homogeneous unless otherwise stated.

The probabilities

$$p_{ij} = P(X_{n+1} = j \mid X_n = i), \quad \text{for } i, j \in S,$$

are called the (1-step) *transition probabilities* of the Markov chain. These are very important and will help us to solve many interesting questions about the behaviour of Markov chains. Notice that p_{ij} does not depend on n as we are assuming that X is time homogeneous.

3.1.2 Transition matrices

The p_{ij} s are often collected into a matrix P called the *transition matrix*. The transition probability p_{ij} is the element of P in the i^{th} row and the j^{th} column, that is, the $(i, j)^{\text{th}}$ element of P .

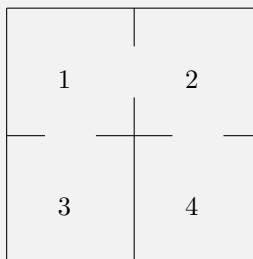
Important properties of a transition matrix include:

- (i) All entries are non-negative (and, because they are probabilities, are ≤ 1).
- (ii) Each of the rows sum to 1, that is, $\sum_j p_{ij} = 1$, for all $j \in S$.

Why must each row sum to 1? Why needn't each column sum to 1?

Exercise 3.1

A rat is put in room 1 in the maze illustrated.



Every minute the rat changes rooms, choosing its exit at random from the exits available. Let Z_n be the number of the room occupied just after the n th transition.

Justify that $\{Z_n\}$ is a Markov chain, and find its transition matrix.

3.1.3 The n-step transition probabilities and the Chapman-Kolmogorov equations

Definition: n-step transition probabilities

The n -step transition probabilities are

$$\begin{aligned} p_{ij}^{(n)} &= P(X_n = j \mid X_0 = i) \\ &= P(X_{m+n} = j \mid X_m = i) \text{ for all } m \text{ if } \{X_n\} \text{ is time-homogeneous.} \end{aligned}$$

So $p_{ij}^{(n)}$ is the probability that a process in state i will be in state j after n 'steps'. Note that due to time homogeneity, this does not depend on m .

Definition: n-step transition matrix

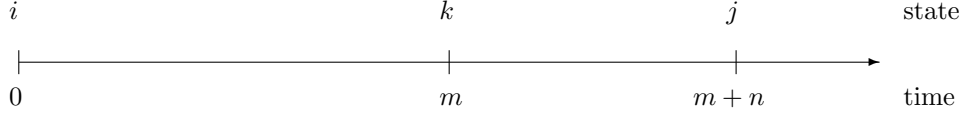
The n -step transition matrix $P^{(n)}$ has $p_{ij}^{(n)}$ as its (i, j) th element.

The Chapman-Kolmogorov (C-K) equations show us how to compute $p_{ij}^{(n+m)}$ or $P^{(n+m)}$ for $m \geq 0, n \geq 0$. These are useful equations which will enable us to calculate these quantities quickly and easily.

Consider time points $0, m$ and $m + n$, where $m \geq 0, n \geq 0$ and suppose that we want to compute the probability that we go from state i to state j in $(n + m)$ steps,

$$p_{ij}^{(n+m)} = P(X_{n+m} = j | X_0 = i).$$

The C-K equations are derived using the simple observation that at time m , we must be in some state (let's call it k). By summing over all possibilities for k , we derive the C-K equations.



By conditioning on the state k at time m we derive $p_{ij}^{(m+n)}$:

$$\begin{aligned}
 p_{ij}^{(n+m)} &= P(X_{m+n} = j \mid X_0 = i) \\
 &= \sum_{k \in S} P(X_{m+n} = j, X_m = k \mid X_0 = i) \\
 &\quad \downarrow \text{see (ii) in 'Useful conditioning formulae'} \\
 &= \sum_{k \in S} P(X_{m+n} = j \mid X_m = k, X_0 = i) P(X_m = k \mid X_0 = i) \\
 &\quad \downarrow \text{Markov property and time homogeneity} \\
 &= \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}, \quad \text{for all } n, m \geq 0, \text{ all } i, j \in S.
 \end{aligned}$$

Definition: Chapman-Kolmogorov equations

The Chapman-Kolmogorov equations are

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}, \quad \text{for all } n, m \geq 0, \text{ all } i, j \in S.$$

Note the plural - there is one equation for each (i, j) pair.

Since

$$\{p_{ik}^{(m)}, k \in S\} \text{ is the } i\text{th row of } P^{(m)}$$

and

$$\{p_{kj}^{(n)}, k \in S\} \text{ is the } j\text{th column of } P^{(n)},$$

the Chapman-Kolmogorov equations can be written in matrix form as

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

where \cdot denotes matrix multiplication. Note that this single equation encodes all the individual C-K equations above.

This may not seem particularly helpful, but note the following:

1. $P^{(1)}$ is the transition matrix P .
2. $P^{(0)}$ is the identity matrix I .

3. Note

$$\begin{aligned}P^{(2)} &= P^{(1+1)} = P \cdot P = P^2. \\P^{(3)} &= P^{(1+2)} = P \cdot P^{(2)} = P^3, \quad \text{etc.}\end{aligned}$$

By using the C-K equations repeatedly, we find that:

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

That is, the n -step transition matrix $P^{(n)}$ is equal to the n th matrix power of P .

This is enormously helpful - given a Markov chain with (one-step) transition matrix P , we can now compute its n -step transition matrix **by simply multiplying P with itself n times**.

Exercise 3.2

The weather on a particular day is either ‘rainy’ (R) or ‘fine’ (F). We assume that a Markov chain model is appropriate. Let X_n be the state of weather on day n . The state space is $\{R, F\}$.

The Markov chain assumption means that, *given information on whether today is rainy or not*, the probability that tomorrow is a rainy day does not depend on the weather on days prior to today. That is,

$$X_{n+1} \perp\!\!\!\perp \{X_{n-1}, \dots, X_0\} \mid X_n.$$

If it rains today, then it will rain tomorrow with probability 0.6.

If it is fine today then it will be fine tomorrow with probability 0.5.

- (a) Find the 2-step transition matrix. [HINT: Find the (1-step) transition matrix first, then apply Chapman-Kolmogorov.]
- (b) Find $P(X_4 = R, X_3 = R \mid X_0 = R)$.

3.1.4 Marginal probabilities

So far, we have only considered *conditional* probabilities. However, we may also be interested in *marginal* probabilities such as $P(X_n = j)$.

Let $p_j^{(n)} = P(X_n = j)$. If, for example, $S = \mathbf{N}_0 = \{0, 1, 2, \dots\}$, then

$$\underline{p}^{(n)} = (p_0^{(n)}, p_1^{(n)}, \dots) = (P(X_n = 0), P(X_n = 1), \dots)$$

is a *probability row vector* (i.e. a row vector with non-negative entries summing to 1) specifying the distribution of X_n .

Do not confuse $p_{ij}^{(n)}$ and $p_j^{(n)}$. The former is a **conditional** probability, and the latter a **marginal** probability.

Definition: Initial distribution

When $n = 0$, $\underline{p}^{(0)} = (p_0^{(0)}, p_1^{(0)}, \dots) = (P(X_0 = 0), P(X_0 = 1), \dots)$ is the distribution of X_0 . This is the *initial distribution*: the distribution of states in which we start the Markov chain. This tells us the probability of the chain starting in any particular state.

If we look at $p_j^{(n)}$, the probability that the chain is in state j at time n , then

$$p_j^{(n)} = P(X_n = j) = \sum_i P(X_n = j \mid X_0 = i) P(X_0 = i) = \sum_i p_i^{(0)} p_{ij}^{(n)}.$$

Equivalently, in matrix notation,

$$\underline{p}^{(n)} = \underline{p}^{(0)} \cdot P^{(n)} = \underline{p}^{(0)} \cdot P^n.$$

Thus, the **initial distribution**, $\underline{p}^{(0)}$, and the **transition matrix**, P , together contain **all the probabilistic information about the Markov chain**. That is, this is all we need in order to compute or derive any quantity of interest related to the Markov chain.

For example,

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p_{i_0}^{(0)} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

is a useful identity to compute a joint probability using only the (one-step) transition probabilities and initial distribution.

Exercise 3.3

In Exercise 3.2, suppose $p_R^{(0)} = 0.7$ and $p_F^{(0)} = 0.3$.

1. Find $\underline{p}^{(0)}$, $\underline{p}^{(1)}$ and $\underline{p}^{(2)}$.
2. Compute $P(X_0 = R, X_1 = F)$.
3. Compute $P(X_0 = R, X_1 = F, X_2 = F, X_3 = R, X_4 = F)$.

3.1.5 Duration of stay in a state

Remember that a Markov chain is *memoryless* in the sense that its future is independent of its past, conditional on its current state (or the most recent information that we have about its location). Therefore, if the chain is currently in state i , the further amount of time that it will remain there does not depend on how long it has already been in state i .

For any Markov chain the duration of stay in state i satisfies

$$\begin{aligned} & P(\text{stay in state } i \text{ for } n \text{ steps} \mid \text{chain is currently in state } i) \\ &= P(\text{stay in } i \text{ for a further } n-1 \text{ steps, then leave } i) \\ &= p_{ii}^{n-1} (1 - p_{ii}). \end{aligned}$$

Therefore, the duration of stay in state i has a geometric distribution with parameter $(1 - p_{ii})$, from which we can deduce that the expected amount of time that the process remains in state i is $\frac{1}{1 - p_{ii}}$.

Where will the Markov chain go when it leaves i ?

For $j \neq i$,

$$\begin{aligned} P(\text{MC moves to } j \mid \text{MC leaves state } i) &= P(X_{n+1} = j \mid X_n = i, X_{n+1} \neq i) \\ &= \frac{P(X_{n+1} = j, X_{n+1} \neq i \mid X_n = i)}{P(X_{n+1} \neq i \mid X_n = i)} \\ &= \frac{P(X_{n+1} = j \mid X_n = i)}{P(X_{n+1} \neq i \mid X_n = i)} \\ &= \frac{p_{ij}}{1 - p_{ii}}. \end{aligned}$$

This answer makes sense: the probability is proportional to p_{ij} and is scaled so that

$$\sum_{j \neq i} \frac{p_{ij}}{1 - p_{ii}} = 1.$$

Summary:

The duration of stay of the Markov chain in state i has a **geometric distribution** with expected value $\frac{1}{1-p_{ii}}$ steps.

When the chain **moves from state** i , it goes to state j ($\neq i$) with probability $\frac{p_{ij}}{1-p_{ii}}$.

3.2 Classification of Markov chains

Suppose we wish to study a particular Markov chain, summarised by its transition matrix P and an initial distribution $\underline{p}^{(0)}$. We may be interested in the **long-run probabilistic behaviour** of the chain or in the **probability that some future event of interest happens**.

Even if nobody has studied this particular Markov chain before, we are able to use some general results on the behaviour of Markov chains to help us. In order to use this theory we need to **classify each state of the chain** as being one of a number types, and hence **classify the type of Markov chain** that we have.

The first step in classifying Markov chains is to split the state space into non-overlapping groups (classes) of states. We will see later that states in the same class are of the same type. This simplifies the problem of classifying all the states in the chain because we only need to classify *one* state in each class.

3.2.1 Irreducible classes of intercommunicating states

Definition: Communicating and intercommunicating classes

State i *communicates* with state j (we write $i \rightarrow j$) if

$$p_{ij}^{(n)} > 0 \quad \text{for some } n \geq 0,$$

that is, starting from state i , it is possible that the chain will eventually enter state j .

States i and j *intercommunicate* (we write $i \leftrightarrow j$) if $i \rightarrow j$ and $j \rightarrow i$.

Exercise 3.5

A Markov chain has the following state-space and transition matrix.

$$S = \{1, 2, 3, 4\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/3 & 1/6 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Draw a diagram to summarise the possible moves which can be made in one step. Which states intercommunicate?

Properties of \leftrightarrow

- (i) $i \leftrightarrow i$ for all states i .
 - (ii) if $i \leftrightarrow j$ then $j \leftrightarrow i$.
 - (iii) if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.
- (i)+(ii)+(iii) mean that \leftrightarrow is an *equivalence relation* on S .

Definition: Irreducible classes and irreducible chains

We can partition the state space S into non-overlapping classes of intercommunicating states. Within a particular class, every pair of states intercommunicate; states in different classes do not intercommunicate.

Each class is called an *irreducible class* of states.

If there is only one irreducible class, then the Markov chain is said to be *irreducible*.

Example 3.6

For the Markov chain with state space $S = \{0, 1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix},$$

the state space can be partitioned into the irreducible classes $\{0, 1\}$ and $\{2\}$. In particular, the Markov chain is not irreducible because there are two irreducible classes.

Exercise 3.6

Find the irreducible classes for each of the Markov chains given below.

$$(a) \quad S = \{0, 1, 2\}, \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}.$$

$$(b) \quad S = \{0, 1, 2, 3\}, \quad P = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

3.2.2 Recurrence and transience

Let $\{X_n\}$ be a Markov chain with state space S and let $i \in S$. Let

$$f_i = P(\text{ever return to } i \mid X_0 = i) = P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i).$$

Definition: Recurrence and transience

If $f_i = 1$ then i is *recurrent* (eventual return to state i is certain).

If $f_i < 1$ then i is *transient* (eventual return to state i is uncertain).

Exercise 3.7

1. For the Markov chain defined in Exercise 3.6(a), deduce whether $f_i = 1$ or $f_i < 1$ for $i = 0, 1, 2$, and hence classify each state as either recurrent or transient.
2. For the Markov chain defined in Exercise 3.6(b), deduce whether $f_i = 1$ or $f_i < 1$ for $i = 0, 1, 2, 3$, and hence classify each state as either recurrent or transient.

Studying recurrence and transience enables us to ask further interesting questions of our Markov chain, for example, **how many times will a Markov chain hit state i , given it starts in state i ?**

(i) If i is **transient**

Let N be the number of hits on i (including the hit from the fact that $X_0 = i$). Then

$$\begin{aligned} P(N=n \mid X_0=i) &= P(\text{return } n-1 \text{ times to } i, \text{ then never return} \mid X_0=i) \\ &= f_i^{n-1}(1-f_i), \end{aligned}$$

So, given $X_0 = i$, N has a geometric distribution with parameter $1 - f_i$. So N is finite with probability 1 and

$$\mathbf{E}(N \mid X_0 = i) = \frac{1}{1 - f_i} \quad (< \infty).$$

(ii) If i is **recurrent**

With probability 1 the chain will eventually return to state i .

By time homogeneity and the Markov property, the chain ‘starts afresh’ on return to the initial state, so that state i will eventually be visited again (with probability 1).

Repeating the argument shows that the chain will return to i infinitely often (with probability 1).

We also have

$$\mathbf{E}(N \mid X_0 = i) \text{ is infinite.}$$

Another idea connected with recurrence and transience is that of *first passage times*. This describes the time it takes for a Markov chain to return to state i , given that it started there.

Definition: First passage time

Let

$$T_{ii} = \min\{n \geq 1 : X_n = i \mid X_0 = i\}.$$

T_{ii} is the *first passage time* from state i to itself, that is, the number of steps until the chain first returns to state i given that it starts in state i .

If the chain never returns to state i we set $T_{ii} = \infty$.

Note that:

$$f_i = \mathbf{P}(\text{ever return to } i \mid X_0 = i) = \mathbf{P}(T_{ii} < \infty)$$

and

$$1 - f_i = \mathbf{P}(\text{never return to } i \mid X_0 = i) = \mathbf{P}(T_{ii} = \infty).$$

Connecting this with the idea of transience and recurrence yields:

- if i is recurrent then $\mathbf{P}(T_{ii} < \infty) = 1$, that is, T_{ii} is finite with probability 1.
- if i is transient then $\mathbf{P}(T_{ii} < \infty) < 1$, or equivalently $\mathbf{P}(T_{ii} = \infty) = 1 - \mathbf{P}(T_{ii} < \infty) > 0$, that is, T_{ii} is infinite with positive probability. (Therefore, $\mu_i = \mathbf{E}[T_{ii}]$ is infinite.)

	i recurrent	i transient
$\mathbf{P}(\text{ever return to } i \mid X_0 = i)$	$f_i = 1$	$f_i < 1$
$\mathbf{P}(\text{never return to } i \mid X_0 = i)$	$1 - f_i = 0$	$1 - f_i > 0$
(number of hits on $i \mid X_0 = i$)	∞ (w.p. 1)	$< \infty$ (w.p. 1)
$\mathbf{E}[\text{number of hits on } i \mid X_0 = i]$	∞	$\frac{1}{1-f_i} < \infty$
$\mathbf{P}(T_{ii} < \infty)$	$f_i = 1$	$f_i < 1$
$\mathbf{P}(T_{ii} = \infty)$	$1 - f_i = 0$	$1 - f_i > 0$

3.2.3 Positive recurrence and null recurrence

In fact, there are two types of recurrent state. Recall that the first passage time (also called the *recurrence time*) T_{ii} of a state i is the number of steps until the chain first returns to state i given that it starts in state i .

Let $\mu_i = \mathbf{E}(T_{ii})$ be the *mean recurrence time* of state i . We have already seen that if i is *transient* then $\mu_i = \infty$. We have not computed μ_i when i is recurrent. In fact, the value of μ_i when i is recurrent enables us to distinguish between two types of recurrence: **positive recurrence** and **null recurrence**.

Definition: Positive and null recurrence

A recurrent state i is

positive recurrent if $\mu_i < \infty$ (that is, μ_i is finite),
null recurrent if $\mu_i = \infty$ (that is, μ_i is infinite).

At first glance it may seem strange that a state i can be null recurrent: return to state i is certain, but the expected time (number of steps) to return to i is infinite. In other words, the random variable T_{ii} is finite with probability 1, but its mean, $\mathbf{E}(T_{ii})$, is infinite. Such states *do* exist and we will see examples later in the course.

3.2.4 Periodicity

Definition: Periodicity

The *period* of a state i is the greatest common divisor of the set of integers $n \geq 1$ such that $p_{ii}^{(n)} > 0$, that is,

$$d_i = \mathbf{gcd} \left\{ n \geq 1 : p_{ii}^{(n)} > 0 \right\}.$$

If $d_i = 1$ then state i is *aperiodic*.

If $p_{ii}^{(n)} = 0$ for all $n \geq 1$ (i.e. return to state i is impossible) then, by convention, state i is also said to be aperiodic.

3.2.5 Class properties

Definition: Class property

A *class property* is a property which is **shared by all the states in an irreducible class**. If it can be shown that one of the states in a class has a certain class property then all the states in the irreducible class must also have that same property.

I give sketch proofs for some of the following results relating to class properties. The other proofs can be found in any good textbook on Markov processes.

Result: Recurrence/transience is a class property

A pair of intercommunicating states are either both recurrent or both transient. That is, recurrence/transience is a class property. It is not possible to have a mixture of recurrent and transient states in an irreducible class.

Proof: Look in ‘Introduction to Probability Models’ (Sheldon Ross).

Result

Null recurrence is a class property.

Proof:

Recall that $\mu_i = E(T_{ii})$ is the expected first passage time or the *mean recurrence time* of state i , and that

$$\mu_i \text{ is } \begin{cases} \text{finite} & \text{if } i \text{ is positive recurrent} \\ \text{infinite} & \text{if } i \text{ is null recurrent or transient.} \end{cases}$$

It can be shown (proof omitted) that

$$\begin{aligned} i \text{ aperiodic} &\Rightarrow p_{ii}^{(n)} \rightarrow \frac{1}{\mu_i} \text{ as } n \rightarrow \infty, \\ i \text{ periodic, period } d &\Rightarrow p_{ii}^{(nd)} \rightarrow \frac{d}{\mu_i} \text{ as } n \rightarrow \infty \text{ (and } p_{ii}^{(m)} = 0 \text{ if } m \text{ is not a multiple of } d). \end{aligned}$$

So, as $n \rightarrow \infty$

- if i is aperiodic, positive recurrent then $p_{ii}^{(n)} \rightarrow \frac{1}{\mu_i}$ (positive, finite),
- if i is periodic, positive recurrent then $p_{ii}^{(nd)} \rightarrow \frac{d}{\mu_i}$ (positive, finite),
- if i is null recurrent then $p_{ii}^{(n)} \rightarrow 0$,
- if i is transient then (proof omitted) $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$, and $p_{ii}^{(n)} \rightarrow 0$.

Now let $i \leftrightarrow j$, with $p_{ij}^{(n)} > 0$, $p_{ji}^{(m)} > 0$. Suppose that i is null recurrent. Then

$$p_{ii}^{(m+k+n)} \geq p_{ij}^{(n)} p_{jj}^{(k)} p_{ji}^{(m)}.$$

As $k \rightarrow \infty$

$$p_{ii}^{(m+k+n)} \rightarrow 0$$

(because i is null recurrent), and so

$$p_{jj}^{(k)} \rightarrow 0$$

(because $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$). So j is either transient or null recurrent. But i is recurrent and $i \leftrightarrow j$, so j must be recurrent as well (because recurrence/transience is a class property). Hence j must be null recurrent. \square

Result

Positive recurrence is a class property.

Proof:

Suppose that $i \leftrightarrow j$ and that i is positive recurrent. Then j must be recurrent (because recurrence is a class property). If j is null recurrent then i is also null recurrent. This is a contradiction. Hence j must be positive recurrent. \square

Result

Intercommunicating states have the same period. That is, periodicity is a class property.

Proof:

Let i and j be states with $i \leftrightarrow j$. Let d_i be the period of i and d_j be the period of j . Let n and m be such that $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$. Then

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

because there is *at least* one way of returning to j in $n + m$ steps: j to i in m steps and then i to j in a further n steps. Therefore,

$$d_j \mid n + m, \quad (d_j \text{ is a divisor of } n + m).$$

Let k be such that $p_{ii}^{(k)} > 0$. Then

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

so

$$d_j \mid n + k + m, \quad (d_j \text{ is a divisor of } n + k + m).$$

together imply that

$$d_j \mid k, \quad (d_j \text{ is a divisor of } k).$$

That is, d_j is a factor of any $\{k \geq 1 : p_{ii}^{(k)} > 0\}$.

However, by definition d_i is the greatest common divisor of such k . Therefore, $d_j \leq d_i$. Reversing the roles of i and j gives $d_i \leq d_j$. Hence $d_i = d_j$. \square

Summary

The following are class properties:

- Null recurrence;
- Positive recurrence;
- Transience;
- Periodicity.

Exercise 3.8

Find and classify the irreducible classes of these Markov chains.

(a)

$$S = \{1, 2, 3, 4\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/3 & 1/6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(b)

$$S = \{0, 1, 2\}, \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

3.2.6 Further properties of discrete time Markov chains**Result**

A finite irreducible Markov chain cannot be null recurrent.

Proof: Recall that $p_j^{(n)} = P(X_n = j)$. Suppose that the chain is null recurrent. **It can be shown** that $p_j^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ for all $j \in S$.

We know that

$$\sum_{j \in S} p_j^{(n)} = 1 \text{ for all } n,$$

so

$$\lim_{n \rightarrow \infty} \sum_{j \in S} p_j^{(n)} = 1.$$

But S is finite so if the chain is null recurrent

$$\lim_{n \rightarrow \infty} \sum_{j \in S} p_j^{(n)} = \sum_{j \in S} \lim_{n \rightarrow \infty} p_j^{(n)} = 0.$$

This is a contradiction, so the Markov chain cannot be null recurrent. □

Result

A finite Markov chain cannot contain any null recurrent states.

Proof:

Suppose that the chain contains a state i which is null recurrent. Then, since null recurrence is a class property, state i is in some irreducible finite closed class of null recurrent states. This is not possible (see above). □

Result: Finite Markov chains

It is not possible for all states in a finite state space Markov chain to be transient.

Note that this implies that *all* states in a finite irreducible Markov chain are recurrent (because recurrence is a class property). That is, finite and irreducible \Rightarrow recurrent Markov chain.

Proof:

Suppose all states are transient and $S = \{0, 1, \dots, M\}$. A transient state will be visited only a finite number of times. Therefore for each state i there exists a time T_i after which i will never be visited again. Therefore, after time $T = \max\{T_0, \dots, T_M\}$ no state will be visited again. BUT the Markov chain must be in some state. Therefore we have a contradiction. Therefore, at least one state must be recurrent. \square

3.2.7 Closed classes and absorption

Definition: Closed classes and absorbing states

A class C of intercommunicating states is *closed* if $p_{ij} = 0$ for all $i \in C$ and $j \notin C$, that is, once the chain gets into C , it never leaves C .

If a single state forms a closed class then it is called an *absorbing state*.

Exercise 3.9

Find all closed classes and absorbing states for the following Markov chains.

(a)

$$S = \{1, 2, 3, 4\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/3 & 1/6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(b)

$$S = \{0, 1, 2\}, \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

Note the following important observations:

- (i) A finite closed class must be recurrent: once entered, the class behaves like a finite irreducible Markov chain.
- (ii) An absorbing state is positive recurrent: its mean recurrence time is 1.

Result

An irreducible class C of recurrent states is closed.

Proof:

Suppose that C is not closed. Then there exist $i \in C$ and $j \notin C$ such that $i \rightarrow j$. But $j \not\rightarrow i$, otherwise i and j would intercommunicate and j would be in C . Thus, there is positive probability that the chain leaves i and never returns to i . This means that i is transient. This is a contradiction, as $i \in C$ and C is recurrent. \square

The next three exercises ask you to find quantities relating to closed classes and absorption. Try to use first step decomposition to solve these questions (see section 1.4.4).

Exercise 3.10

Suppose that X is a discrete time Markov chain with state-space $S = \{0, 1, 2, 3, 4\}$ and transition matrix P given by

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

- (a) Suppose that $X_0 = 1$. Find the probability that the chain is eventually absorbed in $C_3 = \{3, 4\}$.
- (b) If $\underline{p}^{(0)} = (0, 1/3, 2/3, 0, 0)$, calculate the probability that the chain is absorbed in C_3 .
- (c) If the chain is absorbed in $C_3 = \{3, 4\}$, what is the probability that $X_0 = 2$?

Exercise 3.11

Now suppose that X is a discrete time Markov chain with state-space $S = \{0, 1, 2, 3\}$ and transition matrix P given by

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

- (a) Suppose that $X_0 = 2$. Find the expected time until absorption into $\{0, 1\}$.
- (b) Find the expected time to absorption in one or other of the two closed classes, starting from state 1.

Exercise 3.12

Now suppose that X is a discrete time Markov chain with state-space $S = \{0, 1, 2, 3\}$ and transition matrix P given by

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 0 & 2/3 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Find the expected first passage time from state 0 to state 3.

3.2.8 Decomposition of the states of a Markov chain

Putting the previous results together shows that the state space S of a Markov chain can be decomposed as

$$S = T \cup C_{m_1} \cup C_{m_2} \cup \dots,$$

where

- T is a set of transient states;
- C_{m_1}, C_{m_2}, \dots are irreducible closed classes of recurrent states.

Each C_{m_i} is either null recurrent or positive recurrent, and all states in a particular C_{m_i} have the same period (different C_{m_i} 's can have different periods). The following observations apply:

- (a) If $X_0 \in C_{m_i}$ then the chain stays in C_{m_i} forever.
- (b) If $X_0 \in T$ then either:
 - (i) the chain always stays in T (only possible if the set T of states is infinite); or
 - (ii) the chain is eventually absorbed in one of the C_{m_i} 's (where it stays forever).
- (c) If S is finite then it is impossible for the chain to remain forever in the (finite) set T of transient states.
- (d) If S is finite then at least one state must be visited infinitely often and so there must be at least one recurrent state.

We also know that if S is finite then there are no null recurrent states. The consequences this are:

- (a) If a Markov chain has a finite state space then there must be at least one positive recurrent state.
- (b) If a Markov chain has a finite state space and is irreducible, then it must be positive recurrent.
- (c) A finite closed irreducible class must be positive recurrent.

Exercise 3.13

Find and classify the irreducible classes of the following Markov chains.

(a)

$$S = \{0, 1, 2, 3, 4\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/2 \end{pmatrix}.$$

(b)

$$S = \{0, 1, 2, \dots\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & \cdots \\ 1/3 & 0 & 2/3 & 0 & 0 & \cdots \\ 1/4 & 0 & 0 & 3/4 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

that is,

$$p_{i0} = \frac{1}{i+2} \quad \text{and} \quad p_{i,i+1} = \frac{i+1}{i+2}, \quad \text{for } i = 0, 1, 2, \dots$$

(c)

$$S = \{1, 2, 3, 4\}, \quad P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

(d) For $p + q = 1$, $0 < p < 1$,

$$S = \{0, 1, 2, \dots\}, \quad P = \begin{pmatrix} q & p & 0 & 0 & 0 & \cdots \\ 0 & q & p & 0 & 0 & \cdots \\ 0 & 0 & q & p & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

that is,

$$p_{ii} = q \quad \text{and} \quad p_{i,i+1} = p \quad \text{for } i = 0, 1, 2, \dots$$

3.3 Limiting behaviour

This is the crux of the first half of the course: understanding a Markov chain's behaviour in the long run (*limiting* behaviour).

Our goal here is to find whether the distribution of states, $\underline{p}^{(n)}$, settles down (i.e. converges) as $n \rightarrow \infty$. If it does, what is this limit?

We now have enough tools at our disposal. It will turn out that to fully determine its long term behaviour we must classify the Markov chain in terms of:

- (a) Irreducible classes;
- (b) Transience and positive/null recurrence;
- (c) Periodicity;

3.3.1 Invariant distributions

We will start with a somewhat simpler question than that of limiting behaviour: are there particular initial distributions for a Markov chain, such that when the Markov chain starts in any of these distributions, we can determine the limit of $\underline{p}^{(n)}$ as $n \rightarrow \infty$?

The answer to this question is 'yes', and these distributions are known as 'invariant'.

Definition: Invariant distribution

The probability row vector $\underline{\pi}$ is an invariant distribution of a discrete-time process $\{X_t, t \in T\}$ if

$$\underline{\pi} = \underline{\pi} P.$$

If $\underline{\pi}$ is an invariant distribution, and the initial distribution of the Markov chain is $\underline{\pi}$, why does this imply that we can study its long term behaviour?

The idea is that if the chain starts in $\underline{\pi}$ (or equivalently, we find at some later time that the distribution of states is $\underline{\pi}$), then this tells us that the distribution of states is $\underline{\pi}$ forever: we've uncovered the long term behaviour of the chain **in this specific case**.

This works since, suppose that an invariant distribution $\underline{\pi}$ exists and that $\underline{p}^{(0)} = \underline{\pi}$, then

$$\begin{aligned}\underline{p}^{(1)} &= \underline{p}^{(0)} P = \underline{\pi} P = \underline{\pi}; \\ \underline{p}^{(2)} &= \underline{p}^{(1)} P = \underline{\pi} P = \underline{\pi}; \\ \underline{p}^{(3)} &= \underline{p}^{(2)} P = \underline{\pi} P = \underline{\pi}; \quad \text{etc.}\end{aligned}$$

In fact $\underline{p}^{(n)} = \underline{\pi}$, for $n = 1, 2, \dots$ and we have found the limit of $\underline{p}^{(n)}$ as $n \rightarrow \infty$, as required!

If our initial distribution is $\underline{\pi}$, with $\underline{\pi} = \underline{\pi} P$, then the distribution of the states at any given future time n is also $\underline{\pi}$. This is why it is called an *invariant* distribution.

An invariant distribution **always exists if the state space S is finite**, but it need not be unique. If S is infinite then an invariant distribution need not exist.

Assuming that the transition matrix P is a $k \times k$ matrix, and the state space is $S = \{1, 2, \dots, k\}$, solving $\underline{\pi} = \underline{\pi}P$ is a matter of solving a set of k simultaneous equations, each of the form

$$\pi_i = \pi_1 p_{1i} + \pi_2 p_{2i} + \dots + \pi_k p_{ki}.$$

If you try this you will find that one equation is redundant (try it!): we are therefore trying to solve for k unknowns (π_1, \dots, π_k) using $(k - 1)$ equations. This is clearly impossible, but we have the advantage that since $\underline{\pi}$ is itself a probability distribution, we must have $\pi_1 + \dots + \pi_k = 1$. This added requirement means that we can now solve for π_1, \dots, π_k .

Exercise 3.14

Find an invariant distribution for the Markov chain in Exercise 3.2.

3.3.2 Equilibrium distributions

Simply looking at invariant distributions in order to determine a Markov chain's long term behaviour is very restrictive. We want a far more general idea of long term behaviour. Can we determine whether $\underline{p}^{(n)}$ converges as $n \rightarrow \infty$, **regardless of the initial distribution**? If this limit exists, what is it?

Example 3.15

Recall the weather example (Exercise 3.2), with state space $S = \{R, F\}$ and transition matrix P ,

$$P = \begin{pmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{pmatrix}.$$

We examine how $P^{(n)} = P^n$, the n -step transition matrix changes as n increases.

$$P^2 = \begin{pmatrix} 0.56 & 0.44 \\ 0.55 & 0.45 \end{pmatrix}, \quad P^3 = \begin{pmatrix} 0.556 & 0.444 \\ 0.555 & 0.445 \end{pmatrix}, \quad P^6 = P^3 \cdot P^3 = \begin{pmatrix} 0.555556 & 0.444444 \\ 0.555555 & 0.444445 \end{pmatrix}.$$

Looking at different initial distributions for this Markov chain and what happens in the 'long run' from these starting distributions:

Initial distribution	Distribution of X_1	Distribution of X_2	...	Distribution of X_6
(0,1)	(0.5,0.5)	(0.55, 0.45)	...	(0.555556, 0.444444)
(1,0)	(0.6,0.4)	(0.56, 0.44)	...	(0.555556, 0.444444)
(0.7,0.3)	(0.57,0.43)	(0.557, 0.443)	...	(0.555556, 0.444444)
(0.5,0.5)	(0.55,0.45)	(0.555, 0.445)	...	(0.555556, 0.444444)

Notice that as $n \rightarrow \infty$, $P^{(n)}$ is tending to a limiting matrix whose rows are identical to each other, and $p^{(n)}$ seems to be tending to a limiting probability distribution which is the same **regardless of the initial distribution**. This is a very important phenomena which we describe in the next definition.

Definition: Equilibrium distribution

A probability row vector $\underline{\pi} = \{\pi_j, j \in S\}$ is an equilibrium distribution for the discrete-time Markov chain $\{X_n\}$ if

$$\underline{p}^{(n)} \longrightarrow \underline{\pi} \quad \text{as } n \longrightarrow \infty,$$

independently of the initial distribution $\underline{p}^{(0)}$ (i.e. for any initial distribution, the long-run distribution of X_n as $n \rightarrow \infty$ is the same).

When an equilibrium distribution exists,

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)},$$

for all $i, j \in S$.

Notice what this says: $\underline{\pi}$ is an equilibrium distribution if $\underline{p}^{(n)} \rightarrow \underline{\pi}$ as $n \rightarrow \infty$ **regardless** of the start distribution, $\underline{p}^{(0)}$.

It is important to note that there cannot be two or more of these limits: either such a limit does not exist, or exactly one exists. Think about this carefully!

3.3.3 Invariant and equilibrium distributions

We know that $\underline{p}^{(n)} = \underline{p}^{(n-1)}P$. If this chain has an equilibrium distribution then $\underline{p}^{(n)} \rightarrow \underline{\pi}$ as $n \rightarrow \infty$, and $\underline{\pi}$ is the equilibrium distribution. Therefore, letting $n \rightarrow \infty$ we have $\underline{\pi} = \underline{\pi}P$, so that $\underline{\pi}$ is also an *invariant distribution*.

Though the concepts of an invariant distribution and an equilibrium distribution are therefore closely related, they are NOT the same thing: an equilibrium distribution must also be an invariant distribution, BUT an invariant distribution is not necessarily an equilibrium distribution.

$$\begin{array}{ccc} & \Rightarrow & \\ \text{EQUILIBRIUM} & & \text{INVARIANT} \\ & \Leftarrow & \end{array}$$

Invariant distribution $\underline{\pi}$:

- If the chain starts in (or equivalently, gets to) $\underline{\pi}$, the distribution of states at all further times is also $\underline{\pi}$.

Equilibrium distribution π :

- If the chain is run for long enough (i.e. forever) its probabilistic behaviour settles down to that of π .
- **This is regardless of the state in which the chain starts.**

Notice:

- No invariant distribution \Rightarrow no equilibrium distribution.
- Exactly 1 invariant distribution \Rightarrow there may or may not be an equilibrium distribution.
- Two or more invariant distributions \Rightarrow no equilibrium distribution.

Some other important things to note:

- Not all Markov chains have equilibrium distributions.
- If a Markov chain has an equilibrium distribution, then it is unique (cannot have more than one equilibrium distribution).
- If a Markov chain has an equilibrium distribution, then it also has an invariant distribution and these distributions are identical.
- If a Markov chain does not have an equilibrium distribution, then it may or may not have an invariant distribution.
- A Markov chain can have any number of invariant distributions (including no invariant distribution).
- It is possible for a Markov chain to have more than one invariant distribution but not have an equilibrium distribution.
- It is possible for a Markov chain to have neither an invariant distribution nor an equilibrium distribution.

3.3.4 Ergodic Markov chains and the equilibrium distribution

Definition: Ergodic state

An *ergodic* state is one that is both positive recurrent and aperiodic.

Result: Main limit theorem

Let $\{X_n\}$ be an irreducible ergodic Markov chain. Without loss of generality we assume that $S = \{0, 1, 2, \dots\}$. Then

- (a) there is a unique probability row vector $\underline{\pi} = (\pi_0, \pi_1, \pi_2, \dots)$ satisfying

$$\underline{\pi} = \underline{\pi} P,$$

- (b) $p_{ij}^{(n)} \rightarrow \pi_j$ as $n \rightarrow \infty$, for all $i, j \in S$,

- (c) $\underline{\pi}$ satisfies

$$\pi_j = \frac{1}{\mu_j}, \quad \text{for all } j \in S,$$

where μ_j is the mean recurrence time of state j .

Note that this theorem tells us (among other things) that
an irreducible, ergodic Markov chain has an equilibrium distribution.

Some remarks on the Main Limit Theorem

1. Proof omitted.
2. Part (b) of the result shows that the n -step transition matrix $P^{(n)}$ tends to

$$\begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \cdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix},$$

as $n \rightarrow \infty$.

3. Part (b) of the result implies that $\underline{\pi}$ is the equilibrium distribution of the chain (since $\underline{p}^{(n)} = \underline{p}^{(0)} P^{(n)}$ tends to $\underline{\pi}$ as $n \rightarrow \infty$).
4. The following interpretations of $\underline{\pi}$ are important.
 - (i) Observation 3 above implies that π_j is the limiting probability that the chain is in state j at time n .
 - (ii) It can be shown that π_j is also equal to the long-run proportion of time that the chain spends in state j .
5. Note that part (a) of the result can be strengthened to:

An irreducible aperiodic Markov chain has an invariant distribution *if and only if* the chain is positive recurrent.

This implies that π is unique and satisfies $\pi_j = 1/\mu_j$, $j \in S$.

This provides an alternative way to show that a Markov chain is positive recurrent: if the chain is irreducible, aperiodic and has an invariant distribution (that is, there exists a π such that $\pi = \pi P$) then it must be positive recurrent.

Exercise 3.16

Determine whether the Markov chains with the following features have an equilibrium distribution and/or invariant distribution. [HINT: are the following irreducible, ergodic Markov chains? How many invariant distributions does each one have?]

$$(a) \ S = \{R, F\}, \quad P = \begin{pmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{pmatrix}.$$

$$(b) \ S = \{0, 1\}, \quad P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

3.3.5 Extending the Main Limit Theorem

Though the Main Limit Theorem (MLT) tells us that an equilibrium distribution exists for an irreducible and ergodic Markov chain, we want to be able to deduce whether *any* discrete time Markov chain has an equilibrium distribution. We will do this by relaxing each assumption of the Main Limit Theorem and assess whether an equilibrium distribution is still guaranteed to exist.

Periodicity

The MLT assumed that the chain was ergodic, and therefore aperiodic. What happens if we relax this assumption, so that we have an irreducible positive recurrent Markov chain $\{X_n, n = 0, 1, 2, \dots\}$ with period $d > 1$?

Let state j have mean recurrence time μ_j . We construct a new process $\{Y_n, n = 0, 1, 2, \dots\}$ with $Y_n = X_{nd}$. Then $\{Y_n, n = 0, 1, 2, \dots\}$ is an ergodic Markov chain and in this new chain, state j has mean recurrence time μ_j/d . From the Main Limit Theorem it follows that

$$P(Y_n = j \mid Y_0 = j) \rightarrow \frac{d}{\mu_j} \text{ as } n \rightarrow \infty,$$

that is,

$$P(X_{nd} = j \mid X_0 = j) = p_{jj}^{(nd)} \rightarrow \frac{d}{\mu_j} \text{ as } n \rightarrow \infty.$$

However,

$$p_{jj}^{(n)} = 0 \quad \text{for } n \notin \{0, d, 2d, 3d, \dots\}.$$

Hence $\{X_n\}$ has no equilibrium distribution.

Positive recurrence

The MLT assumed that the chain was ergodic, and therefore positive recurrent. What happens if we relax

this assumption so that we have an irreducible aperiodic Markov chain $\{X_n, n = 0, 1, 2, \dots\}$ with transient or null-recurrent states only?

Suppose that a Markov chain consists only of transient and/ or null-recurrent states. Then all states have $\mu_j = \infty$. It follows from the proof that null recurrence is a class property that for all i, j

$$p_{ij}^{(n)} \rightarrow \frac{1}{\mu_j} = 0 \text{ as } n \rightarrow \infty,$$

that is, all limits exist but are zero. *This limit is not a proper probability distribution so no equilibrium distribution exists.*

Irreducibility

The MLT assumed that the chain was irreducible. What happens if we relax this assumption so that the Markov chain has more than one class? We can split this into two possibilities:

- (a) Suppose that a Markov chain contains 2 or more closed classes, C_{m_1}, C_{m_2}, \dots

If, for example, class C_{m_1} is ergodic then for all $j \in C_{m_1}$

$$\begin{cases} p_{ij}^{(n)} \rightarrow \frac{1}{\mu_j}, \text{ as } n \rightarrow \infty, & \text{if } i \in C_{m_1} \\ p_{ij}^{(n)} = 0, \text{ for all } n, & \text{if } i \in C_{m_2}. \end{cases}$$

Hence there is no equilibrium distribution (the limit depends on the initial state).

- (b) Suppose that a Markov chain consists of a finite number of transient states and a **single** closed class C . Then eventually the chain will be absorbed into C .

- If C is **ergodic** then for all states $j \in C$ **there is an equilibrium distribution** since:

$$\pi_j = \begin{cases} 0 & \text{if } j \text{ is transient} \\ \frac{1}{\mu_j} & \text{otherwise.} \end{cases}$$

- For a transient state j , $\mu_j = \infty$ so

$$\pi_j = \frac{1}{\mu_j} \text{ for all } j.$$

This does not give a proper probability distribution, and so *if C is not ergodic then no equilibrium distribution exists.*

3.3.6 When does an equilibrium distribution exist?

To summarise these results:

- (a) More than 1 closed class \Rightarrow no equilibrium distribution exists (the long-term behaviour of the chain depends on $p^{(0)}$).
- (b) No closed classes (i.e. all states are transient) \Rightarrow no equilibrium distribution exists.
- (c) If there exists exactly one closed class C , is it certain that the Markov chain will eventually be absorbed into C ?

- If “no” then no equilibrium distribution exists (the long-term behaviour of the chain depends on $\underline{p}^{(0)}$).
- If “yes” then an equilibrium distribution exists if and only if C is ergodic.

An equilibrium distribution exists if and only if there is an ergodic class C and the Markov chain is certain to be absorbed into C eventually, wherever it starts.

Exercise 3.17

Three cards, labelled A,B and C, are placed in a row and their order changed as follows. A random choice is made of either the left hand card or the right hand card, either card having probability $1/2$ of being chosen, and this card is then placed between the other two. This process is repeated, the successive random choices being independent.

- Find the transition matrix for the six-state Markov chain of successive orders X_1, X_2, \dots of the cards.
- Is the chain irreducible, aperiodic?
- Find the two-step transition matrix.
- If the initial order is ABC, find approximately the probability that after $2n$ changes, where n is large, the order is ABC.

Exercise 3.18

For each of these Markov chains state, with a reason, whether an equilibrium distribution exists. If it does exist then find it.

(a)

$$S = \{0, 1, 2, 3, 4\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/2 \end{pmatrix}.$$

(b)

$$S = \{0, 1, 2, \dots\}, \quad P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & \cdots \\ 1/3 & 0 & 2/3 & 0 & 0 & \cdots \\ 1/4 & 0 & 0 & 3/4 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

(c)

$$S = \{1, 2, 3, 4\}, \quad P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

(d)

$$S = \{0, 1, 2, 3\}, \quad P = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

(e)

$$S = \{0, 1, 2, \dots\}, \quad P = \begin{pmatrix} q & p & 0 & 0 & 0 & \cdots \\ 0 & q & p & 0 & 0 & \cdots \\ 0 & 0 & q & p & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

4 Continuous-time Markov processes

Notation

Discrete-time processes: $\{X_t, t \in T\}$, where $T = \{0, 1, 2, \dots\}$.

Continuous-time processes: $\{X(t), t \in T\}$, where $T = \mathbf{R}_+$ (i.e. $t \geq 0$).

Continuous-time processes can change their value/state ('jump') at *ANY* instant of time.

Remember that in this course we only consider stochastic processes $\{X(t), t \geq 0\}$ with **finite or countable state space** S .

4.1 Continuous-time Markov chains

Definition: Continuous time Markov process

$\{X(t), t \geq 0\}$ is a *continuous-time Markov process* if, for all $t \geq 0$, $0 \leq t_0 < t_1 < \dots < t_n < s$ and for all states $i, j, i_0, i_1, \dots, i_n \in S$

$$P(X(s+t) = j \mid X(s) = i, X(t_n) = i_n, \dots, X(t_0) = i_0) = P(X(s+t) = j \mid X(s) = i),$$

that is, the continuous-time process $\{X(t), t \geq 0\}$ satisfies the *Markov property*.

As in the discrete time case, we only consider *time homogeneous* processes, which implies that:

$$P(X(s+t) = j \mid X(s) = i) = P(X(t) = j \mid X(0) = i)$$

for all s, t and all $i, j \in S$.

The difficulty with continuous time processes is that there is no 'smallest' unit of time. In the discrete time case, we talked about 1-step transition probabilities and extended these to more general n -step transition probabilities. Here, we have no such 'minimum time unit' and so our notation for the transition probabilities must clearly show the amount of time taken to reach one state from another. We will use the notation:

$$p_{ij}(t) = P(X(t) = j \mid X(0) = i)$$

for transition probabilities, and store these in a *transition matrix* $P(t)$ (note again here that we are being careful with our unit of time). If $S = \{0, 1, 2, \dots\}$ this will be of the form

$$P(t) = \begin{pmatrix} p_{00}(t) & p_{01}(t) & p_{02}(t) & \cdots \\ p_{10}(t) & p_{11}(t) & p_{12}(t) & \cdots \\ p_{20}(t) & p_{21}(t) & p_{22}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The same 'rules' apply here as for the discrete time case: the elements of $P(t)$ must satisfy, for all $t \geq 0$,

$$p_{ij}(t) \geq 0 \quad \text{and} \quad \sum_j p_{ij}(t) = 1.$$

In particular, when $t = 0$,

$$p_{ij}(0) = P(X(0) = j \mid X(0) = i) = \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{if } j \neq i, \end{cases}$$

so $P(0) = I$ (the identity matrix). This makes sense - the process stays in the same state if we allow no time to pass!

4.2 The importance of the exponential distribution and ‘little-o’ notation

The exponential distribution will prove to be fundamental for continuous time Markov processes, as it is the only probability distribution to display the *lack of memory property*. Before going any further, we need more notation (this will make calculations simpler).

4.2.1 Order notation: $o(h)$

Definition: An $o(h)$ function

Let $f(x)$ be a function. We define f to be $o(h)$ (we say “ f is order h ”) if

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0,$$

that is,

as $h \rightarrow 0$, $f(h) \rightarrow 0$ faster than h does.

Example 4.1

- $f(x) = x^2$.

$$\frac{f(h)}{h} = \frac{h^2}{h} = h \rightarrow 0 \text{ as } h \rightarrow 0.$$

So x^2 is $o(h)$.

- $f(x) = \sqrt{x}$.

$$\frac{f(h)}{h} = \frac{\sqrt{h}}{h} = \frac{1}{\sqrt{h}} \rightarrow \infty \text{ as } h \rightarrow 0.$$

So \sqrt{x} is not $o(h)$.

Example 4.2 Suppose that $X \sim \text{exponential}(\lambda)$. Then for $h > 0$,

$$\begin{aligned}
 P(X \in (t, t+h] \mid X > t) &= \frac{P(X \in (t, t+h], X > t)}{P(X > t)} \\
 &= \frac{P(X \in (t, t+h])}{P(X > t)} \\
 &= \frac{P(X > t) - P(X > t+h)}{P(X > t)} \\
 &= \frac{\exp(-\lambda t) - \exp(-\lambda(t+h))}{\exp(-\lambda t)} \\
 &= 1 - \exp(-\lambda h) \\
 &= 1 - \left(1 - \lambda h + \frac{(\lambda h)^2}{2!} - \frac{(\lambda h)^3}{3!} + \dots \right) \\
 &= \lambda h - \frac{(\lambda h)^2}{2!} + \frac{(\lambda h)^3}{3!} - \dots \\
 &= \lambda h + o(h).
 \end{aligned} \tag{4}$$

4.2.2 The lack-of-memory property of the exponential distribution

Definition: Memoryless distribution

A random variable X is *memoryless* if for all $s, t \geq 0$,

$$P(X > s+t \mid X > t) = P(X > s)$$

Exercise 4.3

Show that if a random variable X is memoryless, then

$$P(X > s+t) = P(X > s)P(X > t)$$

It turns out that the exponential distribution is memoryless (in fact, it is the only memoryless distribution). Example 4.2 demonstrates this clearly: we showed that if $X \sim \text{exponential}(\lambda)$, then, according to (4) above,

$$P(X \in (t, t+h] \mid X > t) = 1 - \exp(-\lambda h), \quad \text{for } h > 0,$$

or equivalently,

$$P(X - t \in (0, h] \mid X > t) = 1 - \exp(-\lambda h), \quad \text{for } h > 0. \tag{5}$$

The right hand side of (5) is the distribution function of an $\text{exponential}(\lambda)$ distribution. This implies that

$$\{(X - t) \mid X > t\} \sim \text{exponential}(\lambda).$$

That is, if $X \sim \text{exponential}(\lambda)$ then $\{(X - t) \mid X > t\} \sim \text{exponential}(\lambda)$ too. Check back to the definition of a memoryless distribution to see that the exponential distribution is therefore memoryless.

The identity in exercise 4.3 shows the impact of the memoryless property. Suppose that we are waiting for an event to occur (e.g. a train arriving at Goodge Street station), the time we must wait being exponentially distributed with parameter λ . If we wait s minutes without a train arriving, the remaining time until the event occurs is still exponential(λ). That is, the fact that we have already waited s minutes without an event is ‘forgotten’ and we essentially start our wait again.

Example 4.4

Question: Service times at a bank are exponentially and independently distributed with parameter μ . One queue feeds two clerks. When you arrive both clerks are busy, but the queue is empty. You will be next to be served; what is the probability that, of the three customers present, you will be last to leave?

Answer: When the first customer leaves you are served. Your service time T_1 has an exponential(μ) distribution and by the lack-of-memory property of the exponential distribution so does the further service time T_2 of the other customer. T_1 and T_2 are independent. Therefore

$$\begin{aligned} P(\text{you are last to leave}) &= P(T_1 > T_2) \\ &= \int_0^\infty P(T_1 > T_2 \mid T_2 = t) f_{T_2}(t) dt \\ &= \int_0^\infty P(T_1 > t) f_{T_2}(t) dt \\ &= \int_0^\infty \exp(-\mu t) \mu \exp(-\mu t) dt = \frac{\mu}{2\mu} = \frac{1}{2}. \end{aligned}$$

4.3 Breaking down the definition of a continuous time Markov process

A continuous-time Markov chain can be described by the distribution of *holding times* for each state (i.e. the distribution of the length of time the chain stays in a particular state), together with the *jump chain* (i.e. when the continuous-time Markov chain does leave a state, to which other state does it go?).

4.3.1 Holding times

Remember that for a continuous time Markov process with discrete state-space, the chain can ‘jump’ to another state at any point in time. We’ll first look at how long a chain stays in a state before moving. It can be shown that if the chain starts in state i then it stays there for a random length of time, the *holding time*, that is exponentially distributed with parameter, say, q_i (i.e. with mean $1/q_i$).

Why is the distribution of the holding time exponential?

Suppose that the process starts in state i , and let $T(i)$ be the holding time in state i . We will show that $P(T(i) > u + s \mid T(i) > u) = P(T(i) > s)$ for $u, s \geq 0$, and from this we can deduce that the distribution of $T(i)$ is memoryless (and hence exponential).

Notice that the event $\{T(i) > u\}$ is identical to the event $\{X(t) = i \text{ for all } 0 \leq t \leq u\}$, for any $u \geq 0$. Now:

$$\begin{aligned}
P(T(i) > s + t | T(i) > s) &= P(X(u) = i \text{ for all } 0 \leq u \leq t + s | X(u) = i \text{ for all } 0 \leq u \leq s) \\
&\quad \downarrow \text{Using information from the conditioning} \\
&= P(X(u) = i \text{ for all } s \leq u \leq t + s | X(u) = i \text{ for all } 0 \leq u \leq s) \\
&\quad \downarrow \text{Markov property} \\
&= P(X(u) = i \text{ for all } s \leq u \leq t + s | X(s) = i) \\
&\quad \downarrow \text{Time homogeneity} \\
&= P(X(u) = i \text{ for all } 0 \leq u \leq t | X(0) = i) \\
&\quad \downarrow \text{Equivalent statements} \\
&= P(T(i) > t)
\end{aligned}$$

So $T(i)$ is memoryless, and therefore exponentially distributed. Supposing that we let q_i denote the parameter of this exponential distribution, then q_i can also be thought of as the *rate at which the chain leaves state i* .

4.3.2 The jump chain

The sequence of states visited by a continuous-time Markov chain forms a discrete-time Markov chain called the embedded *jump chain*.

In the previous section we defined the rate at which the chain leaves state i to be q_i . Notice that this makes no assumptions about the state which the chain goes to, simply that the chain moves from state i . We will use the notation q_{ij} to denote the rate at which the chain moves from state i to state j .

A simple way of thinking of about these q_{ij} is with an ‘alarm clock’ analogy. Suppose that there are k states that our chain can visit, and that the chain is currently in state i . For each of the $k - 1$ states that the chain can visit next, there is an exponential alarm clock. For example, the alarm clock for state j will ring at an exponentially distributed time, with parameter q_{ij} . From state i , we next visit the state whose alarm clock rings next. Once the chain moves to this state, we reset all alarm clocks again, one for each of the states, noting again that there is no alarm clock for the state that the chain is currently in.

Notice also that we must have

$$q_i = \sum_{j \neq i} q_{ij}. \quad (6)$$

That is, the rate at which we leave state i is the sum of the rates at which we enter another state.

These ideas give rise to some interesting problems. For example, if the chain is in state i , what’s the probability that the alarm clock for state j will be the first to ring? This is equivalent to asking ‘what’s the probability that when we move from state i , we go to state j ?’.

It is helpful to note the following result before we attempt this problem.

Result: Distribution of the minimum of exponential random variables

Let Y_1, \dots, Y_n be independent random variables such that $Y_i \sim \text{exponential}(\lambda_i)$. Then:

- The distribution of the minimum of these random variables is also exponential, with parameter $(\lambda_1 + \dots + \lambda_n)$.
- The probability that Y_k is the minimum of these random variables can be shown to be

$$\frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}$$

Challenge

Prove these assertions about the minimum of exponential random variables!

Using these results, can we find the probability that the alarm clock for state j will be the first to ring, given that the process is currently in state i ? This is equivalent to asking ‘what’s the probability that when we move from state i , we go to state j ?’ Since the time until the alarm clock for state j rings is exponential with parameter q_{ij} , then

$$\begin{aligned} p_{ij} &:= P(\text{from state } i, \text{ move to state } j) \\ &= P(\text{alarm for state } j \text{ rings first} \mid \text{currently in state } i) \\ &= P(T(j) < T(k) \text{ for } k \neq j \mid \text{currently in state } i) \\ &= P(T(j) = \min\{T(k)\} \mid \text{currently in state } i) \\ &= \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} =: \frac{q_{ij}}{q_i} \end{aligned}$$

Therefore, the probability that from state i , the process goes to state j is given by $p_{ij} = q_{ij}/q_i$.

Note: Suppose that the continuous-time Markov chain has an absorbing state, state i . If the chain enters state i then it will stay there forever. In this case we would set $p_{ii} = 1$, rather than using the equation given above.

Notice that the jump chain is itself a *discrete-time Markov chain*, and therefore we can construct its transition matrix.

Example 4.5

Suppose that a continuous time Markov chain has state-space $S = \{1, 2, 3, 4\}$, and that q_{ij} denotes the rate at which the chain moves from state i to state j (for $j \neq i$). Let

$$q_i = \sum_{j \neq i} q_{ij} \quad \text{for } i = 1, 2, 3, 4.$$

The transition matrix of its jump chain is

$$P = \begin{pmatrix} 0 & q_{12}/q_1 & q_{13}/q_1 & q_{14}/q_1 \\ q_{21}/q_2 & 0 & q_{23}/q_2 & q_{24}/q_2 \\ q_{31}/q_3 & q_{32}/q_3 & 0 & q_{34}/q_3 \\ q_{41}/q_4 & q_{42}/q_4 & q_{43}/q_4 & 0 \end{pmatrix}$$

Notice that the diagonal of the transition matrix for a jump chain is ALWAYS ZERO, since we are assuming that the chain moves to *another state*. Check that all rows sum to 1!

The jump chain can be used to answer questions which involve the states that the chain enters but not the times at which they are entered:

- What is the probability that $\{X(t), t \geq 0\}$ ever enters a particular state?
- If a chain states are thought of as the number of individuals in a population (e.g. if the process is in state 5, then there are 5 individuals in the population), what is the probability that a population ever becomes extinct? That is, what is the probability that the chain ever enters state 0?

4.4 Analysis of transition probabilities

So far, we have only considered the rate at which a continuous time Markov chain moves from one state to another. For discrete time processes, our interest was initially in transition probabilities. We will now explore transition probabilities for continuous time Markov chains.

The obvious ‘problem’ with continuous time Markov chains is that, unlike in the discrete time setting, there is no “smallest time” until the next transition. Instead, we can jump to another state at *any* time. The transition probability $p_{ij}(t)$, for states i and j ($j \neq i$), is a function of t which in principle can be studied. This is far more complex than in the discrete time case, and our aim is to *dispose of the need to work with transition probabilities to study the long term behaviour of the chain*.

It turns out that this is possible, and we will later encounter the *generator matrix* of a continuous time Markov chain, which will be much easier to work with than transition probabilities $p_{ij}(t)$. To understand where the generator matrix comes from, we will first investigate the nature of transition probabilities for continuous time Markov processes.

We will make use of $T(i)$, the holding time in state i , throughout this section. First we will consider transition probabilities $p_{ij}(h)$ over a *very small* period of time of length h . This will help us in studying transition probabilities over longer periods of time, and we will be able to develop further tools, namely:

1. The continuous time version of the Chapman-Kolmogorov equations;

2. Kolmogorov's forward equations;
3. Kolmogorov's backward equations

Don't forget that this is all in the hope of understanding the long term behaviour of continuous time Markov chains!

4.4.1 Transition probabilities over a small interval

We'll first look at transition probabilities $p_{ij}(h)$ for very small $h > 0$. To get from state i to state j in a time interval of length h , there are two broad options:

1. Go directly from state i to state j , with the transition occurring sometime in the interval $[0, h]$.
2. Go from state i to state j via another state k , with $k \neq i, j$, (or via more than one other state) with all transitions occurring within the interval $[0, h]$.

In fact, we will show that only the first option is possible, that is, the probability of making more than one transition in a very short time interval is negligible.

Suppose that we start in state i . The probability of one transition in $[0, h]$ is

$$P(T(i) < h) = 1 - \exp(-q_i h) = q_i h + o(h),$$

while the probability of one transition in $[0, h]$ and that the transition is to state $j \neq i$ is

$$\begin{aligned} p_{ij}(h) &= P(T(i) < h \mid \text{move from } i \text{ to } j \text{ in the interval } [0, h]) P(\text{move from } i \text{ to } j \text{ in the interval } [0, h]) \\ &= P(T(i) < h) P(\text{move from } i \text{ to } j \text{ in the interval } [0, h]) \\ &= (1 - \exp(-q_i h)) \frac{q_{ij}}{q_i} \\ &= (q_i h + o(h)) \frac{q_{ij}}{q_i} = q_{ij} h + o(h) \end{aligned}$$

Probability of two transitions in $[0, h]$ is

$$\sum_{k \neq i} P(T(i) + T(k) < h \mid \text{move from } i \text{ to } k \text{ first}) P(\text{move from } i \text{ to } k \text{ first}) = \sum_{k \neq i} o(h) q_{ik} = o(h),$$

while the probability of two transitions in $[0, h]$ and that the chain at time h is in state j is

$$P(\text{two transitions in } [0, h] \text{ and that the chain at time } h \text{ is in state } j) \leq P(\text{two transitions in } [0, h]) = o(h)$$

Similarly, we can show that for more than two transitions, $P(\text{more than two transition in } [0, h]) = o(h)$.

Putting this together,

$$\begin{aligned} p_{ii}(h) &= P(\text{no transitions in } [0, h] \mid X(0) = i) \\ &= 1 - P(\text{one transition in } [0, h] \mid X(0) = i) - P(\text{two transitions in } [0, h] \mid X(0) = i) - \dots \\ &= 1 - q_i h + o(h) \end{aligned}$$

In summary:

- In a very small time interval of length h , the probability of making more than one transition is $o(h)$. This is *negligible*, and so we consider that AT MOST one transition will occur in such a small time interval.

- The probability of moving from state i to state j in the small time interval $[0, h]$ is

$$p_{ij}(h) = q_{ij}h + o(h)$$

and that this transition happens directly from i to j .

- The probability being in state i at time h , given that the chain starts in state i , is

$$p_{ii}(h) = 1 - q_i h + o(h)$$

and that no transitions occur during this interval (otherwise, to get back to state i , we would require at least two transitions which has negligible probability).

Putting this together we have that

$$p_{ij}(h) = \begin{cases} 1 - q_i h + o(h) & \text{for } i = j \\ q_{ij} h + o(h) & \text{for } i \neq j. \end{cases}$$

4.4.2 Transitions over longer periods: the Chapman-Kolmogorov equations

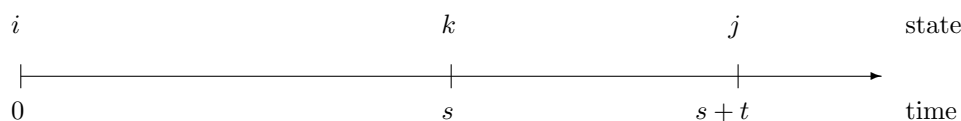
It turns out that the Chapman-Kolmogorov equations are just as important in continuous time as they were in discrete time. They will become our main tool in proving two crucial results on transition probabilities later on: the Kolmogorov forward and backward equations.

Result: Chapman-Kolmogorov equations in continuous time

The Chapman-Kolmogorov equations in continuous time are

$$P(s + t) = P(s) P(t).$$

The proof is as in discrete time: condition on the state at time s .



Individual conditional probabilities:

$$p_{ij}(s + t) = \sum_{k \in S} p_{ik}(s) p_{kj}(t)$$

Matrix format:

$$P(s + t) = P(s) P(t)$$

We also have

$$\underline{p}^{(t)} = \underline{p}^{(0)} P(t),$$

where $\underline{p}^{(0)}$ is the distribution of X_0 (the *initial* distribution) and $\underline{p}^{(t)} = (p_0^{(t)}, p_1^{(t)}, \dots)$ (the distribution of the possible location of process at time t).

4.4.3 Kolmogorov's forward equations

The next results, Kolmogorov's equations, show us how to compute the transition matrix $P(t)$ for any $t > 0$. Their more useful role, however, is in proving later results on the long-term behaviour of continuous time processes. There are two versions of Kolmogorov's equations: the *forward* and the *backward* equations.

Result: Kolmogorov forward differential equations (KFDEs)

The Kolmogorov forward differential equations (KFDEs) are

$$p'_{ij}(t) = -p_{ij}(t) q_j + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj}$$

for all pairs (i, j) , or, in matrix notation,

$$P'(t) = P(t) Q,$$

where, for a continuous time Markov chain with state space $S = \{1, 2, 3, \dots\}$,

$$Q = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \dots \\ q_{21} & -q_2 & q_{23} & \dots \\ q_{31} & q_{32} & -q_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where q_{ij} is the transition rate from state i to state j and $q_i = \sum_{j \neq i} q_{ij}$ (the rate at which the process leaves state i , as before).

The KFDEs are rather easy to prove, the main tool being the Chapman-Kolmogorov equations for continuous time processes. We take an initial time point (0) followed later by two time points (t and $t+h$) with h small. Notice that the times t and $t+h$ are therefore close together.



By the Chapman-Kolmogorov equations, we know that

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(t) p_{kj}(h)$$

and recalling that for small $h > 0$,

$$p_{ij}(h) = \begin{cases} 1 - q_i h + o(h) & \text{for } i = j \\ q_{ij} h + o(h) & \text{for } i \neq j. \end{cases}$$

Therefore, substituting this in and separating the case $k = j$ from $k \neq j$,

$$\begin{aligned} p_{ij}(t+h) &= (1 - q_j h + o(h)) p_{ij}(t) + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) (q_{kj} h + o(h)) \\ &= p_{ij}(t) - h p_{ij}(t) q_j + h \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj} + o(h). \end{aligned}$$

Now subtract $p_{ij}(t)$ from each side, and divide each side by h :

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = -p_{ij}(t) q_j + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj} + \frac{o(h)}{h}.$$

Letting $h \downarrow 0$ gives

$$p'_{ij}(t) = -p_{ij}(t) q_j + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj}, \quad \text{for all } i, j \in S,$$

which shows that the KFDEs hold.

In matrix notation

$$P'(t) = P(t) Q.$$

Exercise 4.6

Convince yourself that the matrix version of the KFDEs contains all the individual differential equations

$$p'_{ij}(t) = -p_{ij}(t) q_j + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj},$$

for all i, j in S .

The importance of the KFDEs

If we wish to find $p_{ij}(t) = P(X(t) = j \mid X(0) = i)$, for $j = 0, 1, 2, \dots$ then we have only *one set of equations to solve*.

If we are able to do this we will get a *general* solution. Then we use the initial distribution $\underline{p}^{(0)}$ (often of the form $X(0) = i$ with probability 1) to find the *particular* solution which satisfies this initial condition.

4.4.4 Kolmogorov's backward equations

Result: Kolmogorov backward differential equations (KBDEs)

The Kolmogorov backward differential equations (KBDEs) are

$$p'_{ij}(t) = -p_{ij}(t) q_j + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj}$$

or, in matrix notation,

$$P'(t) = Q P(t),$$

where, for a continuous time Markov chain with state space $S = \{1, 2, 3, \dots\}$,

$$Q = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \dots \\ q_{21} & -q_2 & q_{23} & \dots \\ q_{31} & q_{32} & -q_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where q_{ij} is the transition rate from state i to state j and $q_i = \sum_{j \neq i} q_{ij}$ (the rate at which the process leaves state i , as before).

Exercise 4.7

Compare Kolmogorov's forward and backward equations.

The KBDEs can be shown to hold by conditioning on the state of the process at time h (compare this with the forward equations). We take two initial time points (0 and h) close together followed later by a single time point ($t + h$).



By the Chapman-Kolmogorov equations, we know that

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(h) p_{kj}(t)$$

and recalling that for small $h > 0$,

$$p_{ik}(h) = \begin{cases} 1 - q_i h + o(h) & \text{for } i = k \\ q_{ik} h + o(h) & \text{for } i \neq k. \end{cases}$$

Therefore, substituting this in and separating the case $i = k$ from $i \neq k$,

$$\begin{aligned} p_{ij}(t+h) &= (1 - q_i h + o(h)) p_{ij}(t) + \sum_{\substack{k \in S \\ k \neq j}} (q_{ik} h + o(h)) p_{kj}(t) \\ &= p_{ij}(t) - h p_{ij}(t) q_i + h \sum_{\substack{k \in S \\ k \neq j}} q_{ik} p_{kj}(t) + o(h). \end{aligned}$$

Now subtract $p_{ij}(t)$ from each side, and divide each side by h :

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = -p_{ij}(t) q_i + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj} + \frac{o(h)}{h}.$$

Letting $h \downarrow 0$ gives

$$p'_{ij}(t) = -p_{ij}(t) q_i + \sum_{\substack{k \in S \\ k \neq j}} p_{ik}(t) q_{kj}, \quad \text{for all } i, j \in S,$$

which shows that the KBDEs hold. The KBDEs in matrix notation are written as

$$P'(t) = Q P(t).$$

4.4.5 Solving the KFDEs and KBDEs

The KFDEs are $P'(t) = P(t)Q$ and the KBDEs are $P'(t) = QP(t)$. As these are differential equations, we can solve them to find an expression for $P(t)$.

An initial condition is required to obtain a specific as opposed to general solution. It makes sense to use $P(0) = I$ as the initial condition (why?). In doing so, both the KFDEs and KBDEs give the same solution for $P(t)$ (this is a good thing, otherwise something very strange is happening!).

Result: Solution to the KFDEs and KBDEs

Subject to the initial condition $P(0) = I$, both KFDEs and KBDEs have the same solution for $P(t)$, namely

$$P(t) = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!} = \exp(tQ).$$

Exercise 4.8

Using the solution to the differential equations, i.e.

$$P(t) = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!},$$

show that for a small time $h > 0$, we have

$$P(h) = I + hQ + o(h),$$

where I denotes the identity matrix.

4.4.6 The generator matrix, Q

We've seen that the transition *rates* that we introduced in section 4.3 play a significant role for continuous time Markov processes, appearing in the KFDEs and KBDEs. The solution to these differential equations show us that the transition matrix $P(t)$ is entirely dependent on these rates through the matrix Q .

In fact, the matrix Q is so important that it has a name - the **generator matrix**. We'll now have a look at some properties of this matrix Q in preparation for the next section, where we will use it extensively.

For a continuous time Markov process with state-space $S = \{1, 2, 3, \dots\}$, the generator matrix Q is defined as

$$Q = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \dots \\ q_{21} & -q_2 & q_{23} & \dots \\ q_{31} & q_{32} & -q_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Since the q_{ik} are rates, they are strictly non-negative, and therefore q_i must be so too. Also recall that in section 4.2, we saw

$$q_i = \sum_{k \neq i} q_{ik}.$$

Consider the rows of Q , and you will notice that they **sum to zero**. Therefore, the diagonal of the matrix Q is negative (can be zero too), while all the off-diagonal entries are positive (can be zero).

In fact, the matrix Q is so important that together with the initial distribution of the continuous time Markov chain, $\underline{p}^{(0)}$, we can specify the process exactly and can compute all that we're interested in, including the long term behaviour of the chain. Compare this to the discrete time chain, where we required the initial distribution together with the transition matrix:

Discrete time process specified by:	Continuous time process specified by:
Initial distribution, $\underline{p}^{(0)}$	Initial distribution, $\underline{p}^{(0)}$
Transition matrix, P	Generator matrix, Q

In effect, for continuous time Markov processes, the generator matrix Q takes the place of the transition matrix P in the discrete time case.

Note: throughout this chapter, we assume that the q_{ij} s satisfy certain technical conditions which are sufficient to prevent a Markov chain 'behaving badly', for example, the elements in the matrix Q (the rates) exist and are finite. These conditions will be met in all our examples.

So far, we have thought about the generator matrix, Q , as a matrix of rates. In fact, we could also show that

$$Q := \left. \frac{dP(t)}{dt} \right|_{t=0} = P'(0).$$

That is, Q contains the rates of change of the transition probabilities $P(X(t) = j \mid X(0) = i)$ at $t = 0$: just as we already know. In order for this differentiation to make sense, we assume that $P(t)$ is

continuous at 0, that is,

$$p_{ij}(t) \rightarrow \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \quad \text{as } t \downarrow 0.$$

4.5 Limiting behaviour

As we did for the discrete time processes, we will study both invariant and equilibrium distributions. We interpret invariant and equilibrium distributions in the same way as we did for discrete time processes.

4.5.1 Invariant distributions

Definition: Invariant distribution for continuous time Markov chains

The probability row vector $\underline{\pi}$ is an invariant distribution of a continuous-time Markov chain $\{X(t), t \geq 0\}$ with transition matrix $P(t)$ if

$$\underline{\pi} = \underline{\pi} P(t), \quad \text{for all } t \geq 0.$$

Using $P(t) = \exp(tQ)$ gives

$$\begin{aligned} \underline{\pi} \exp tQ &= \underline{\pi} \\ \underline{\pi} \left(1 + tQ + \frac{t^2 Q^2}{2!} + \dots\right) &= \underline{\pi} \\ t\underline{\pi}Q + \frac{t^2 \underline{\pi}Q^2}{2!} + \dots &= \underline{0}, \end{aligned}$$

which holds for all $t \geq 0$ if and only if $\underline{\pi}Q = \underline{0}$.

$$\underline{\pi} = \underline{\pi} P(t) \text{ is equivalent to } \underline{\pi} Q = \underline{0}.$$

The consequence of this is as follows. If $\underline{p}^{(0)} = \underline{\pi}$ then $\underline{p}^{(t)} = \underline{\pi}$ for all $t \geq 0$, that is, if we start the chain using the distribution $\underline{\pi}$, the distribution of the chain at any future time t is also $\underline{\pi}$.

More generally, if the distribution of states of our continuous Markov chain is $\underline{\pi}$ eventually, then the distribution of states of our continuous Markov chain will be $\underline{\pi}$ for all future times.

Finding the invariant distribution of a continuous time Markov chain requires finding a vector $\underline{\pi}$ which satisfies $\underline{\pi}Q = \underline{0}$. This is much easier than solving $\underline{\pi} = \underline{\pi} P(t)$ as $P(t)$ itself is difficult to compute!

4.5.2 Equilibrium distributions

Definition: Equilibrium distribution for continuous time Markov chains

An equilibrium distribution $\underline{\pi} = \{\pi_j, j \in S\}$ exists if $p_{ij}(t) \rightarrow \pi_j$ as $t \rightarrow \infty$ for each $j \in S$, where $\underline{\pi}$ is a probability distribution that does not depend on the initial state i .

Suppose that such a distribution exists. Then since $p_{ij}(t) \rightarrow \pi_j$ (a constant) as $t \rightarrow \infty$ we expect

$$p'_{ij}(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

The KFDEs also give

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t) q_{kj}.$$

Letting $t \rightarrow \infty$ in the KFDE above gives

$$\lim_{t \rightarrow \infty} p'_{ij}(t) = \lim_{t \rightarrow \infty} \sum_{k \in S} p_{ik}(t) q_{kj} = \lim_{t \rightarrow \infty} \sum_{k \in S} \pi_k q_{kj}.$$

Therefore, since $p'_{ij}(t) \rightarrow 0$ as $t \rightarrow \infty$, this is equivalent to

$$0 = \sum_{k \in S} \pi_k q_{kj} \quad \text{or} \quad \underline{\pi} Q = \underline{0}.$$

What does this tell us?

The fact that an equilibrium distribution $\underline{\pi}$ satisfies $\underline{\pi} Q = \underline{0}$ implies that an equilibrium distribution is also an invariant distribution (just as was the case in discrete time). Therefore an equilibrium distribution is an invariant distribution of $\{X(t), t \geq 0\}$.

Note that it is not always true the other way around: an invariant distribution of $\{X(t), t \geq 0\}$ is not necessarily an equilibrium distribution.

We used the KFDEs to establish this result. Since the forward and backward equations give the same solution for $P(t)$, could we have used the KBDEs to do the same?

In this case, the answer is ‘no’. Though the KFDEs and KBDEs look very similar, it is sometimes the case that only one of these is useful in solving a particular problem. Here, using the KFDEs allowed us to see the connection between the invariant and equilibrium distributions. However, if we had tried to use the KBDEs instead, we would get the following:

$$\underline{0} = \sum_{k \in S} q_{ik} \pi_j = \pi_j \sum_{k \in S} q_{ik} = \underline{0}.$$

Obviously, the equation is perfectly valid and correct, but $\underline{0} = \underline{0}$ is of no use whatsoever!

In other cases, which we will not consider in this course, the backward equations prove to be useful while the forward equations are not.

4.5.3 A limit theorem for continuous time Markov chains

Think back to the discrete time case: we were able to establish a theorem that told us exactly when an equilibrium distribution existed for a particular discrete time Markov chain. We would like to do the same here, too, for continuous time Markov chains. In the continuous time case, things are more difficult, and we end up with a result that is a little more restrictive than we did in the discrete time case.

The theorem will depend on the Markov chain being *irreducible*, which we define next.

Definition: Irreducible continuous time Markov chains

A continuous-time Markov chain is *irreducible* if, for every i and j in S , $p_{ij}(t) > 0$ for some t . That is, all states in the chain intercommunicate.

Result: Existence of an equilibrium distribution for irreducible chains

Suppose that $\{X(t), t \geq 0\}$ is an **irreducible** continuous-time Markov chain.

- (i) If there exists an invariant distribution π then it is unique and $p_{ij}(t) \rightarrow \pi_j$ as $t \rightarrow \infty$ for all $i, j \in S$. (i.e. π is the equilibrium distribution of the chain.)
- (ii) If there is no invariant distribution then $p_{ij}(t) \rightarrow 0$ as $t \rightarrow \infty$ for all $i, j \in S$. (i.e. no equilibrium distribution exists.)

Note:

- For *continuous-time* Markov chains we do not need to worry about periodicity.
- If the chain is positive recurrent then such an invariant distribution does exist. Therefore, if an irreducible Markov chain has no invariant distribution it must be either null recurrent or transient.

It does not follow that only irreducible processes can have an equilibrium distribution. In the next section we will show that a process which is not irreducible does have an equilibrium distribution under certain conditions. However, to prove this, we cannot rely on the result above: we do the maths ‘brute force’.

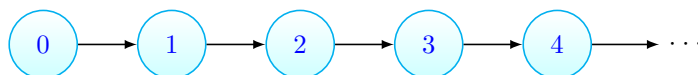
5 Important types of continuous-time processes

5.1 The Poisson process

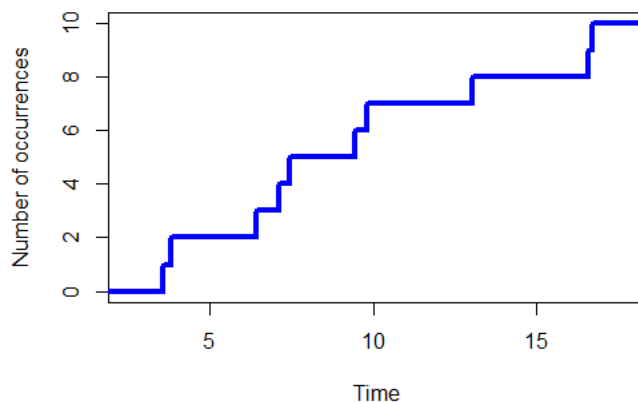
This is a very simple but important **continuous time Markov process** which is the building block for many complex processes. A Poisson process, $\{N(t), t \geq 0\}$, counts the number of events in $(0, t]$. These events occur one at a time and independently of each other.

- The state space of a Poisson process is infinite and consists of the natural numbers together with $\{0\}$: $S = \{0, 1, 2, 3, \dots\}$.
- The Poisson process always starts in state 0: $N(0) = 0$.
- As it is a counting process, when it leaves state 0 (no events) it moves to state 1 (one event has occurred).
- From state 1 (one event has occurred), it proceeds to state 2 (two events have occurred).
- And so on...
- These 'events' occur uniformly, i.e. at a constant rate, λ , per unit time.

You may find it helpful to visualise a Poisson process using the following state space diagram,



Notice that there are no probabilities on the arrows here, unlike the discrete time case. Remember that moving to the next state can occur at any time. Equivalently, we can visualise a Poisson process using a graph which plots the realisation of the process; each 'step' upward denotes the occurrence of an event.



Formally, a Poisson process is defined as follows.

Definition: Poisson process

Let $N(t)$ denote the number of occurrences of some event in $(0, t]$ for which there exists $\lambda > 0$ such that for $h > 0$:

- (i) $P(1 \text{ event in } (t, t + h]) = \lambda h + o(h)$;
- (ii) $P(\text{no events in } (t, t + h]) = 1 - \lambda h + o(h)$;
- (iii) The number of events in $(t, t + h]$ is independent of the process in $(0, t]$.

Then $\{N(t), t \geq 0\}$ is a **Poisson process of rate λ** .

Notes

1. Statement (i) can be written as $P(N(t + h) - N(t) = 1) = \lambda h + o(h)$.
Statement (ii) can be written as $P(N(t + h) - N(t) = 0) = 1 - \lambda h + o(h)$.
2. Statement (i) says that $P(1 \text{ event in } (t, t + h])$ is approximately λh (that is, approximately proportional to the length h of the time interval $(t, t + h]$ for small h .
Statement (ii) says that $P(0 \text{ events in } (t, t + h]) \approx 1 - \lambda h$, for small h . Therefore, $P(\text{more than 1 event in } (t, t + h])$ is $o(h)$.
3. Statement (iii) says that the process has *independent increments*.
4. Note that the Poisson process satisfies the Markov property: consider times $t_1 < t_2 < \dots < t_n < t_{n+1}$. Then

$$\begin{aligned} & P(N(t_{n+1}) = i_{n+1} \mid N(t_n) = i_n, \dots, N(t_1) = i_1) \\ &= P(i_{n+1} - i_n \text{ events in } (t_n, t_{n+1}] \mid N(t_n) = i_n, \dots, N(t_1) = i_1) \\ &= P(i_{n+1} - i_n \text{ events in } (t_n, t_{n+1})) \quad \text{by (iii)} \\ &= P(N(t_{n+1}) = i_{n+1} \mid N(t_n) = i_n). \end{aligned}$$

5.1.1 The generator matrix for the Poisson process

Remember that a Poisson process currently in state i must next move to state $i + 1$: there is no choice of states. How long does the Poisson process stay in state i ? This is the holding time, and we already know this is exponentially distributed with some parameter q_i . What is q_i in this case? Recall that for small $h > 0$,

$$p_{ii}(h) = 1 - q_i h + o(h).$$

Compare this to the definition of the Poisson process, which states that

$$p_{ii}(h) = P(\text{no events in } (t, t + h]) = 1 - \lambda h + o(h),$$

and that this holds for all $i = 0, 1, 2, \dots$. Together, these imply that $q_i = \lambda$: the rate of leaving state i is the rate of the Poisson process.

Notice also that:

- The rate of leaving state i is equivalent to the rate at which we go from state i to state $i + 1$, therefore $q_{i,i+1} = q_i = \lambda$.

- For $k < i$, $q_{ik} = 0$ since the Poisson process cannot go down in value - it is an increasing process!
- For $k > (i + 1)$, $q_{ik} = 0$ since the Poisson process cannot jump directly from state i to state $k > i + 1$: it must jump to state $i + 1$ first.

Putting this together, the generator matrix for a Poisson process is

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & 0 & -\lambda & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Notice that since we now have the generator matrix, Q , for a Poisson process, and we know that its initial distribution is always $\underline{p}^{(0)} = (1, 0, 0, \dots)$, we can use the results from Chapter 4 on continuous time processes. In particular, test your understanding of these results by answering the following.

Exercise 5.1

Does a Poisson process have an invariant distribution? An equilibrium distribution?

Exercise 5.2

Write down the forward and backward equations for the Poisson process (you need not solve them).

Exercise 5.3

Find the embedded jump chain of the Poisson process.

5.1.2 The distribution of the number of events by time t

Now we'll look at further properties of the Poisson process.

What is the probability that, by time t , k events have occurred? In other words, can we calculate $p_{0k}(t) =: P(N(t) = k)$? Note that this is the same as asking for the probability that the Poisson process has moved from state 0 to state k in a time interval of length t .

We can use the forward equations for a Poisson process to establish the answer to this question. The forward

equations are

$$\begin{aligned}
 p'_{i,i+1}(t) &= -p_{i,i+1}(t) q_{i+1} + \sum_{\substack{k \in S \\ k \neq i}} p_{ik}(t) q_{k,i+1} \\
 &= -p_{i,i+1}(t) q_{i+1} + p_{ii}(t) q_{i,i+1} \\
 &= -p_{i,i+1}(t) \lambda + p_{ii}(t) \lambda \\
 &= \lambda(p_{ii}(t) - p_{i,i+1}(t)),
 \end{aligned}$$

for all $i = 0, 1, 2, \dots$. We can solve these differential equations, as before, and we use

$$p_k(0) = P(N(0) = k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{for } k = 1, 2, 3, \dots, \end{cases}$$

as the initial condition, i.e. $N(0) = 0$. The solution is

$$p_{0,k}(t) = P(N(t) = k) = \frac{(\lambda t)^k \exp(-\lambda t)}{k!}, \quad k = 0, 1, 2, \dots,$$

i.e.

$$N(t) \sim \text{Poisson}(\lambda t).$$

Exercise 5.4

Show that $P(N(t) = k) = \frac{(\lambda t)^k \exp(-\lambda t)}{k!}$ is indeed the solution to the forward equations, by substituting this into both sides of the equation.

Exercise 5.5

Tourists arrive at the departure point of a sightseeing bus according to a Poisson process of rate λ per minute. Denote by $N(t)$ the number of tourists that arrive at the departure point of the bus during a time interval of length t minutes.

- (i) For $t > 0$, name the distribution of $N(t)$ and state its mean and variance.
- (ii) The bus driver gets bored and drives off after 10 minutes. If $\lambda = 3$, what is the probability that there will be 32 tourists on the bus? What is the expected number of passengers that will be on the bus?
- (iii) Given that the *first* tourist arrived during the first 5 minutes, find the probability that he or she arrived during the first 2 minutes.

5.1.3 Other properties of the Poisson process

Result: Number of events

The number of events in $(s, s + t]$ has a $\text{Poisson}(\lambda t)$ distribution (for all s).

Statements (i), (ii) and (iii) in the definition of the Poisson process are constant over time, so

$$P(k \text{ events in } (s, s+t]) = P(k \text{ events in } (0, t]) = \frac{(\lambda t)^k \exp(-\lambda t)}{k!}.$$

[The Poisson process has **stationary increments**: the distribution of the number of events in a time interval does not depend on when the interval starts. This is equivalent to time homogeneity.]

Result: Number of events in a small interval

In a small interval of length h , the probability of one event is $\lambda h + o(h)$. The probability of two or more of these events in a small interval of length h is $o(h)$.

$$P(0 \text{ events in } (t, t+h]) = P(T_1 > h) = \exp(-\lambda h) = 1 - \lambda h + o(h).$$

$$P(\geq 1 \text{ event in } (t, t+h]) = P(T_1 < h) = 1 - \exp(-\lambda h) = \lambda h + o(h).$$

$$P(\geq 2 \text{ events in } (t, t+h]) = P(T_1 + T_2 < h) = \int_0^h \lambda^2 x \exp(-\lambda x) \, dx = o(h).$$

Result: Disjoint intervals

If $(a, b]$ and $(c, d]$ are non-overlapping intervals, then the number of events in $(a, b]$ is independent of the numbers of events in $(c, d]$.

This follows from condition (iii).

Result: Time to first event

The time T_1 to the first event is exponentially distributed with parameter λ .

$$P(T_1 > t) = P(N(t) = 0) = \exp(-\lambda t), \quad t > 0.$$

Therefore,

$$F_{T_1}(t) = P(T_1 \leq t) = 1 - \exp(-\lambda t) \quad \text{and} \quad f_{T_1}(t) = \lambda \exp(-\lambda t), t > 0.$$

The random variable T_1 has an exponential distribution with rate λ .

Result: Time between successive events

The times between successive events are i.i.d. $\text{exponential}(\lambda)$ random variables.

Let T_2 be the time between the first and second events. Firstly, we find the marginal distribution of T_2 :

$$\begin{aligned}
P(T_2 > t) &= \int_0^\infty P(T_2 > t \mid T_1 = s) f_{T_1}(s) \, ds \\
&= \int_0^\infty P(\text{no events in } (s, s+t]) \lambda \exp(-\lambda s) \, ds \\
&= \int_0^\infty \exp(-\lambda t) \lambda \exp(-\lambda s) \, ds \\
&= \exp(-\lambda t).
\end{aligned}$$

Therefore, $T_2 \sim \exp(\lambda)$. Now we show that T_1 and T_2 are independent.

$$\begin{aligned}
P(T_1 > v, T_2 > t) &= \int_v^\infty P(T_2 > t \mid T_1 = u) f_{T_1}(u) \, du \\
&= \int_v^\infty P(\text{no events in } (u, u+t]) \lambda \exp(-\lambda u) \, du \\
&= \int_v^\infty \exp(-\lambda t) \lambda \exp(-\lambda u) \, du \\
&= \exp(-\lambda t) \exp(-\lambda v) = P(T_2 > t) P(T_1 > v).
\end{aligned}$$

Therefore, T_1 and T_2 are independent. Repeating the argument for (T_2, T_3) , (T_3, T_4) , \dots gives the result.

Result: Time to r^{th} event

The time to the r^{th} event has a $\text{Gamma}(r, \lambda)$ distribution.

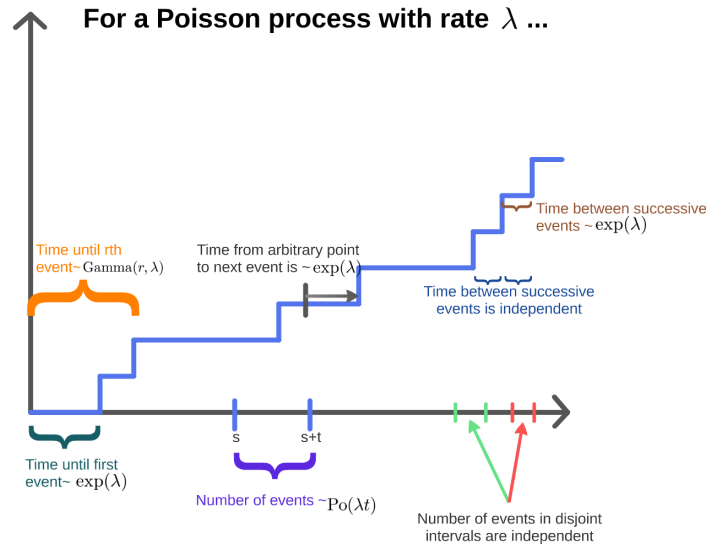
Let S_r be the time to the r th event. Then $S_r = T_1 + \dots + T_r$. This is the sum of r i.i.d. $\exp(\lambda)$ random variables, which has a $\text{Gamma}(r, \lambda)$ distribution.

Result: Time to next event from arbitrary time point

The time from an arbitrary time point t to the next event is an $\exp(\lambda)$ random variable.

This follows directly from the lack-of-memory property of the exponential distribution and is consistent with the definition (i), (ii), (iii) of a Poisson process.

These properties are summarised in the following diagram.



A further **very useful** property is the distribution of the times at which events occur *when we know exactly the number of events that have happened*. Notice that this is a result which is *conditional* on knowing the number of events that have occurred.

Result: Distribution of times of events when k events have occurred

Given that exactly k events occur in $(0, t]$, the members of the set (U_1, \dots, U_k) of k **(unordered)** arrival times are i.i.d. random variables, each of which has distribution $U(0, t)$. That is,

$$U_1, \dots, U_{N(t)} \mid N(t) = k \stackrel{\text{i.i.d.}}{\sim} U(0, t).$$

This makes sense: the k events are randomly scattered in the interval $(0, t]$. You do not need to be able to prove this result, but you will need to use it.

5.1.4 Superposition and thinning of a Poisson process

Result: Superposition of Poisson processes

Suppose that $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ are independent Poisson processes with rates λ_1 and λ_2 respectively, and $N(t) = N_1(t) + N_2(t)$ for all $t \geq 0$ (so that $\{N(t)\}$ counts the total number of events).

Then $\{N(t), t \geq 0\}$ is a Poisson process of rate $\lambda_1 + \lambda_2$.

We'll show that the three conditions, (i), (ii) and (iii), required of a Poisson process, hold.

(i)

$$\begin{aligned}
& P(1 \text{ event in } (t, t+h]) \\
&= P(1 \text{ event from process 1})P(0 \text{ events from process 2}) \\
&\quad + P(0 \text{ events from process 1})P(1 \text{ event from process 2}) \\
&= (\lambda_1 h + o(h))(1 - \lambda_2 h + o(h)) + (1 - \lambda_1 h + o(h))\{\lambda_2 h + o(h)\} \\
&= \lambda_1 h - \lambda_1 \lambda_2 h^2 + o(h) + \lambda_2 h - \lambda_1 \lambda_2 h^2 + o(h) \\
&= (\lambda_1 + \lambda_2)h + o(h).
\end{aligned}$$

(ii)

$$\begin{aligned}
& P(0 \text{ events in } (t, t+h]) \\
&= P(0 \text{ events in process 1})P(0 \text{ events in process 2}) \\
&= (1 - \lambda_1 h + o(h))(1 - \lambda_2 h + o(h)) \\
&= 1 - (\lambda_1 + \lambda_2)h + o(h).
\end{aligned}$$

(iii) Independent increments. This holds because it holds for each process individually and the two process are independent.

Result: Thinning of a Poisson process

$\{N(t), t \geq 0\}$ is a Poisson process of rate λ . Each event of the process is deleted with probability $(1 - p)$ and kept with probability p , independently of all other events in the process.

Let $\{M(t), t \geq 0\}$ be the process containing only the events which are kept, then $\{M(t), t \geq 0\}$ is a Poisson process of rate $p\lambda$.

We'll show that the three conditions, (i), (ii) and (iii), required of a Poisson process, hold.

(i)

$$\begin{aligned}
& P(1 \text{ event in } (t, t+h]) \\
&= P(1 \text{ event in } \{N(t)\} \text{ in } (t, t+h], \text{event not deleted}) + o(h) \\
&= (\lambda h + o(h)) \times p + o(h) = p\lambda h + o(h).
\end{aligned}$$

(ii)

$$\begin{aligned}
& P(0 \text{ events in } (t, t+h]) \\
&= P(0 \text{ events in } \{N(t)\} \text{ in } (t, t+h]) \\
&\quad + P(1 \text{ event in } \{N(t)\} \text{ in } (t, t+h], \text{event deleted}) + o(h) \\
&= 1 - \lambda h + o(h) + (\lambda h + o(h)) \times (1 - p) + o(h) \\
&= 1 - p\lambda h + o(h).
\end{aligned}$$

(iii) Deleting events doesn't affect the independent increments property because events are deleted independently of each other.

Example 5.6

Outside Heals on Tottenham Court Road a charity worker attempts to stop people in order to have a conversation. People walking along the street arrive at Heals in a Poisson process of rate 40 per minute. On average only 1 in every 20 people stop to talk. Each person's decision is taken independently of all other people. Each conversation takes a time which is exponentially distributed with mean 1 minute, and is independent of other conversations. If the charity worker is busy, people passing Heals do not stop.

Let $X(t)$ denote whether the charity worker is busy at time t , with $S = \{0, 1\}$. State Q and the forward equation for $p_{01}(t)$. If the charity worker is free at time $t = 0$ minutes, show that the probability that they are busy at time $t > 0$ is

$$\frac{2}{3} - \frac{2}{3} \exp(-3t).$$

Hence, or otherwise, find the long-run proportion of the time that the charity worker is busy.

Solution

Firstly, $p_n(t) = P(X(t) = n \mid X(0) = 0)$ with $X(0) = 0$ so that $p_0(0) = 1$. People who stop to talk arrive in a Poisson process of rate $40 \times \frac{1}{20} = 2$ per minute. [Thinning of a Poisson process.]

The generator matrix is therefore

$$Q = \begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix},$$

and forward equation for $p_{01}(t)$ is

$$p'_{01}(t) = 2p_{00}(t) - p_{01}(t).$$

Notice that $p_{00}(t) = 1 - p_{01}(t)$, and substituting this into the forward equation for $p_{01}(t)$ yields

$$p'_{01}(t) = 2(1 - p_{01}(t)) - p_{01}(t) = 2 - 3p_{01}(t).$$

We now need to show that

$$p_{01}(t) = \frac{2}{3} - \frac{2}{3} \exp(-3t).$$

Using this identity, the left hand side of the forward equation above is

$$p'_{01}(t) = \frac{6}{3} \exp(-3t) = 2 \exp(-3t),$$

and the right hand side of the forward equation above is also $2 \exp(-3t)$. Therefore, $\frac{2}{3} - \frac{2}{3} \exp(-3t)$ satisfies the forward equation for $p_{01}(t)$, and so

$$p_{01}(t) = \frac{2}{3} - \frac{2}{3} \exp(-3t),$$

as required.

We get the long-run proportion of time that the charity worker is busy by letting $t \rightarrow \infty$, giving the answer $2/3$. Don't forget that we could have obtained this answer more easily by solving $\pi Q = 0$, yielding $\pi_1 = 2/3$, as before.

Exercise 5.7

A bicycle factory has three independent production lines, each of which produces bicycles according to a Poisson Process. The first, for racing bikes, has a time in minutes between completed machines that is exponential $(1/5)$. The interarrival times for the second, for BMX bikes, are exponential $(1/4)$ and for the third, for mountain bikes, are exponential $(1/6)$.

- (a) What is the expected number of bikes produced in an 8 hour shift?
- (b) What is the probability that exactly fourteen BMX bikes are made in the first hour of production?
- (c) What is the variance of the time in minutes until the third mountain bike is made from the beginning of the shift?
- (d) Sir Bradley Wiggins arrives during a shift. How long would he expect to wait until he sees the first BMX bicycle roll off the production line?
- (e) Given that 20 bikes are produced in an hour, what is the mean number that will be produced in the period $[21, 36)$ minutes?

Exercise 5.8

On a single telephone line, calls arrive according to a Poisson process of rate λ per minute.

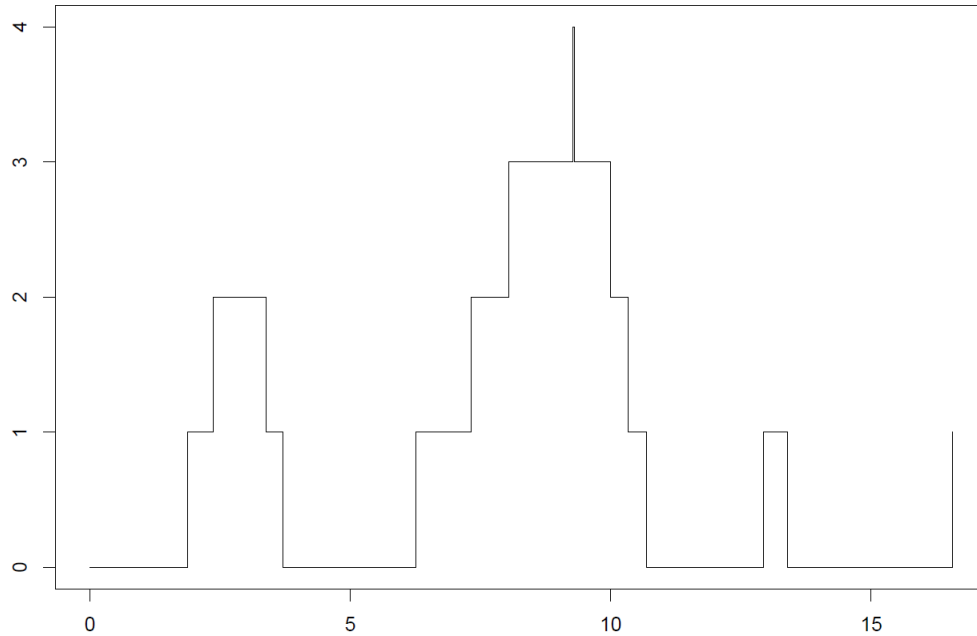
- (a) Name the distribution of the number of phone calls that arrive during a time interval of length t minutes ($t > 0$), and state its mean.
- (b) Name the distribution of the time (in minutes) between the arrivals of two successive phone calls, and state its mean.

If a caller finds the line free then his/her call is ‘effective’; effective calls have durations that are independent, exponentially distributed random variables with a mean of $1/\mu$ minutes, independently of the arrival process. If a caller finds the line busy then this call is lost.

- (c) Let Y be the time (in minutes) from the beginning of an effective phone call to the arrival of the next effective phone call. Write down the mean and the variance of Y .
- (d) State, with a reason, whether or not the arrivals of the *effective* phone calls follow a Poisson process.

5.2 Birth and death processes

This is another important class of continuous-time Markov processes in which jumps are always one step up (‘birth’) or one step down (‘death’). Births and deaths occur **singly** and **independently** of each other. The state space is $S = \{0, 1, 2, \dots\}$. A graph showing the path of a birth-death process is shown below. A step upward corresponds to a birth, while a step downward corresponds to a death. Compare this to the Poisson process!



We think about the ‘states’ of a birth death process as the size of the population: if the process is ‘in state n ’ then ‘the size of the population is n ’. The state identifies the number of individuals in the population (though note that a birth-death process doesn’t *have* to describe the size of a population - it can describe any process which goes up or down by one unit at each transition).

Important questions to ask of birth-death processes include:

- What is the probability of a birth (death) in a small time interval?
- What is the probability that the next event is a birth? A death?
- How long must we wait until the next event?
- How do these processes behave in the long run?
- Will the population ever die out (become extinct)?

5.2.1 Transition rates and the jump chain

Let $N(t)$ denote the size of a population at time $t \geq 0$ for a birth-death process. If at time t the size of the population is n , then next transition is:

- to state $(n + 1)$ (**a birth**) OR
- to state $(n - 1)$ (**a death**).

Notation: if the size of the population is n , then, using the same notation as in Chapter 4,

$$\begin{aligned} \text{Birth rate: } q_{n,n+1} &= \lambda_n \\ \text{Death rate: } q_{n,n-1} &= \mu_n \end{aligned}$$

- Obviously we must have $\mu_0 = 0$. Why?
- The birth and death rates can depend on the population size. That is, they *can* change as the population size changes which is why the general birth and death rates are indicated with a subscript n , the population size.

How long does the process stay in state n ? We know from Chapter 4 that the distribution of this time is exponential, and since the only transitions allowable are to $(n+1)$ (with rate $q_{n,n+1} = \lambda_n$) or to $(n-1)$ (with rate $q_{n,n-1} = \mu_n$) then we have that the rate at which the process leaves state n is $q_n = q_{n,n+1} + q_{n,n-1} = (\lambda_n + \mu_n)$. Therefore, the time T until the next ‘event’ will be exponential with rate $(\lambda_n + \mu_n)$,

$$T \sim \text{exponential}(\lambda_n + \mu_n).$$

From this we can construct the jump chain of a birth-death process. Using the birth and death rates λ_n and μ_n , the next event will be a:

- **birth**, with probability $\lambda_n/(\lambda_n + \mu_n)$;
- **death**, with probability $\mu_n/(\lambda_n + \mu_n)$.

The transition matrix of the embedded jump chain is therefore

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{\mu_1}{\lambda_1 + \mu_1} & 0 & \frac{\lambda_1}{\lambda_1 + \mu_1} & 0 & 0 & 0 \\ 0 & \frac{\mu_2}{\lambda_2 + \mu_2} & 0 & \frac{\lambda_2}{\lambda_2 + \mu_2} & 0 & 0 \\ 0 & 0 & \frac{\mu_3}{\lambda_3 + \mu_3} & 0 & \frac{\lambda_3}{\lambda_3 + \mu_3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

5.2.2 The generator matrix

We can immediately construct the generator matrix, Q , for the birth-death process using the transition rates λ_n and μ_n , and the following observations:

- For $k < (i - 1)$, $q_{ik} = 0$ since the birth-death process cannot jump down by more than 1.
- For $k > (i + 1)$, $q_{ik} = 0$ since the birth-death process cannot jump directly from state i to state $k > i + 1$: it must jump to state $i + 1$ first.

Definition: Generator matrix for the birth-death process

The generator matrix for a birth-death process is

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & 0 \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

We could also have constructed the generator matrix by computing the transition probabilities $p_{ij}(h)$ for small h , and reading off the associated transition rates. Let $p_{ij}(h) = P(N(t+h) = j \mid N(t) = i)$. Then

$$\begin{aligned} p_{n,n+1}(h) &= \{\lambda_n h + o(h)\}\{1 - \lambda h + o(h)\} + o(h) \\ &= \lambda_n h + o(h) \\ p_{n,n-1}(h) &= \{1 - \mu h + o(h)\}\{\mu h + o(h)\} + o(h) \\ &= \mu h + o(h) \\ p_{nn}(h) &= \{1 - \mu h + o(h)\}\{1 - \lambda h + o(h)\} + o(h) \\ &= 1 - (\mu + \lambda)h + o(h). \\ p_{nm}(h) &= o(h), \quad \text{if } m \notin \{n-1, n, n+1\}. \end{aligned}$$

Now compare this to the general result for any continuous time Markov process with discrete state space:

$$\begin{aligned} p_{ij}(h) &= q_{ij}h + o(h) \\ p_{ii}(h) &= 1 - q_i h + o(h) \end{aligned}$$

Match up the rates with the q_i and q_{ij} , to construct the generator matrix. Of course, this is a far more cumbersome way of extracting the generator matrix!

5.2.3 Important types of birth-death processes

- (i) IMMIGRATION at rate α : $\lambda_n = \alpha$, $n = 0, 1, 2, \dots$
- (ii) LINEAR BIRTH at rate λ per individual (that is, every individual alive at time t has probability $\lambda h + o(h)$ of giving birth to a new individual in $(t, t+h]$):

$$\lambda_n = n\lambda, \quad n = 0, 1, 2, \dots$$

- (iii) EMIGRATION at rate β : $\lambda_n = \beta$, $n = 1, 2, \dots$

- (iv) LINEAR DEATH at rate μ per individual (that is, every individual alive at time t has probability $\mu h + o(h)$ of dying in $(t, t+h]$):

$$\mu_n = n\mu, \quad n = 1, 2, \dots$$

Easy exercise

Which types of birth-death processes (listed above) have birth or death rates which depend on the current population size?

We could define a process that has (i), (ii), (iii) and (iv), that is,

$$\lambda_n = \alpha + n\lambda, \quad n = 0, 1, 2, \dots, \quad \text{and} \quad \mu_n = \beta + n\mu, \quad n = 1, 2, \dots,$$

although it would be difficult to study this process algebraically.

Some comments on particular types of birth-death process.

- A Poisson process of rate λ is a pure immigration process, i.e.

$$\lambda_n = \lambda, \quad n = 0, 1, \dots \quad \text{and} \quad \mu_n = 0, \quad n = 1, 2, \dots$$

Therefore, an immigration process of rate α is equivalent to a Poisson process of rate α .

- A pure death process has $\lambda_0 = 0$, for $n = 0, 1, 2, \dots$

Recall that the time until the next jump is $\text{exponential}(q_i)$, so that under pure linear death the time until the next jump is $\text{exponential}(n\mu)$.

This makes sense because the time until a given individual dies is $\text{exponential}(\mu)$ and all individuals are independent of each other.

[Recall from section 4.3.2 that if $T_i \stackrel{\text{i.i.d.}}{\sim} \text{exponential}(\mu)$, $i = 1, \dots, n$ then $\min(T_1, \dots, T_n) \sim \text{exponential}(n\mu)$]

- A pure birth process has $\mu_n = 0$, for $n = 1, 2, \dots$
(... and we can make similar statements to pure linear death with ‘death’ replaced by ‘birth’).

5.2.4 Transition probabilities (over a small interval)

We already know from Chapter 4 that for any continuous time, discrete state space, Markov process we have:

$$p_{ij}(h) = q_{ij}h + o(h)$$

for **small** h . For the birth-death process,

$$q_{n,n+1} = \lambda_n, \quad q_{n,n-1} = \mu_n \quad \text{and} \quad q_{ij} = 0 \text{ for } j \neq (i-1), (i+1),$$

therefore, supposing that the process is currently in state n (n individuals in the population):

- the probability of a birth in $(t, t+h]$, is $p_{n,n+1}(h) = \lambda_n h + o(h)$.
- the probability of a death in $(t, t+h]$, is $p_{n,n-1}(h) = \mu_n h + o(h)$.

Exercise 5.9

Fill in the table below, assuming that the population size at time t is n .

	Immigration-Emigration process	Immigration-Emigration w. linear birth
Birth rate		
Death rate		
$P(\text{Birth} \in (t, t+h])$ (small h)		
$P(\text{Death} \in (t, t+h])$ (small h)		
Q		

5.2.5 The equilibrium distribution of the (general) birth and death process

Do all birth-death processes have an equilibrium distribution? Our result in Chapter 4 only specified whether an irreducible continuous time Markov chain had an equilibrium distribution:

Result from Chapter 4

Suppose that $\{X(t), t \geq 0\}$ is an **irreducible** continuous-time Markov chain.

- (i) If there exists an invariant distribution $\underline{\pi}$ then it is unique and $p_{ij}(t) \rightarrow \pi_j$ as $t \rightarrow \infty$ for all $i, j \in S$.
 - That is, $\underline{\pi}$ is the equilibrium distribution of the chain.
- (ii) If there is no invariant distribution then $p_{ij}(t) \rightarrow 0$ as $t \rightarrow \infty$ for all $i, j \in S$.
 - No equilibrium distribution exists.

Are all birth-death processes irreducible?

No (consider, for example, a Poisson process). When a process is not irreducible, we cannot guarantee that an invariant distribution is an equilibrium distribution. Which birth-death processes are irreducible (and so we can apply the result above)?

Exercise 5.10

Fill in the table to show which birth-death processes are irreducible, and which are not.

	Linear birth	Immigration	Linear birth with Immigration
Linear death			
Emigration			
Linear death with emigration			

Note that for a process to have an equilibrium distribution, it must have exactly one invariant distribution. Recall that the existence of exactly one invariant distribution does not necessarily imply that the process has an equilibrium distribution, though! Furthermore, on applying the result in Chapter 4,

- If the process is irreducible - existence of invariant distribution implies existence of equilibrium distribution.
- If the process is NOT irreducible - cannot yet tell if existence of an invariant distribution implies the existence of an equilibrium distribution.

First of all, we'll search for conditions under which an invariant distribution exists for a birth-death process. Once this has been established, we will show that for one particular non-irreducible birth-death process (for which exactly one invariant distribution exists) the invariant distribution is in fact an equilibrium distribution.

So when does a birth-death process have an invariant distribution?

Solving $\pi Q = 0$ gives

$$\begin{aligned} -\lambda_0 \pi_0 + \mu_1 \pi_1 &= 0 \\ \lambda_{j-1} \pi_{j-1} - (\lambda_j + \mu_j) \pi_j + \mu_{j+1} \pi_{j+1} &= 0, \quad j = 1, 2, \dots \end{aligned}$$

The first equation gives

$$\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0.$$

Solving the second set of equations iteratively gives

$$\pi_j = \frac{\lambda_{j-1} \cdots \lambda_0}{\mu_j \cdots \mu_1} \pi_0. \quad j = 2, 3, \dots \quad (7)$$

Therefore, $\underline{\pi}$ defined by:

$$\begin{aligned}\pi_1 &= \frac{\lambda_0}{\mu_1} \pi_0; \\ \pi_j &= \frac{\lambda_{j-1} \cdots \lambda_0}{\mu_j \cdots \mu_1} \pi_0. \quad \text{for } j = 2, 3, \dots\end{aligned}$$

satisfies $\underline{\pi}Q = \underline{0}$.

For $\underline{\pi}$ to be a **probability distribution** it must also satisfy $\sum_{j=0}^{\infty} \pi_j = 1$.

$$\begin{aligned}\pi_0 \left(1 + \sum_{j=1}^{\infty} \frac{\lambda_{j-1} \cdots \lambda_0}{\mu_j \cdots \mu_1} \right) &= 1, \\ \left(1 + \sum_{j=1}^{\infty} \frac{\lambda_{j-1} \cdots \lambda_0}{\mu_j \cdots \mu_1} \right)^{-1} &= \pi_0.\end{aligned}\tag{8}$$

These equations define a probability distribution if and only if the sum in (8) is finite. *If the state space S is finite, then the sum in (8) must be finite.*

In summary, if (8) is finite, then an invariant distribution exists. This is guaranteed for birth-death processes with finite state-space.

If an invariant distribution exists, is it an equilibrium distribution?

Notice that if (8) is divergent an invariant distribution (and therefore an equilibrium distribution) **cannot exist**.

If the process is **irreducible**, the equilibrium and invariant distributions coincide (if the latter exists).

What if the process is not irreducible? Can an equilibrium distribution exist? The answer is yes. We will now look for conditions under which an equilibrium distribution exists for a **linear birth-death process** only (we will not look at other types of non-irreducible birth-death processes). This equilibrium distribution will be linked to the *probability of extinction* for the process, which we look at next.

5.2.6 The probability of extinction for linear birth-death process

The linear birth-death process is not irreducible because state 0 does not communicate with any other state. Note that for a linear birth-death process, 0 is an absorbing state: once the process reaches 0, there are no individuals in the population to give birth and the population is therefore ‘extinct’. It is interesting in its own right to study this probability, but we’ll be interested in it because of its link with the invariant and equilibrium distribution for this process.

How are ‘extinction probability’ and invariant distributions linked in this case?

For linear birth-death processes, once extinction (reaching state 0) occurs, we stay in this state forever. Therefore, $\underline{\pi} = (1, 0, 0, \dots)$ is an invariant distribution for the process.

Under what circumstances is $\underline{\pi} = (1, 0, 0, \dots)$ also an equilibrium distribution? This is equivalent to asking ‘under which conditions does the population die out (become extinct) with probability 1?’. Alternatively, ‘under what circumstances is it certain that the chain will reach the state 0?’.

An elegant way of solving this is via **generating functions**. Let $N(t)$ denote the state of the linear birth-death process at time t (i.e. the number of individuals in the population at time t). Then

$$G(s, t) = E \left[s^{N(t)} \right] = \sum_{n=0}^{\infty} s^n P(N(t) = n).$$

It can be shown that

$$G(s, t) = \begin{cases} \left(\frac{\mu(1-s) - (\mu - \lambda s)e^{-(\lambda - \mu)t}}{\lambda(1-s) - (\mu - \lambda s)e^{-(\lambda - \mu)t}} \right)^{N(0)} & \text{if } \lambda \neq \mu \\ \left(\frac{\lambda t(1-s) + s}{\lambda t(1-s) + 1} \right)^{N(0)} & \text{if } \lambda = \mu \end{cases}$$

Now, the probability that extinction occurs at or before time t is given by $P(N(t) = 0)$, which can be extracted from the generating function via

$$G(0, t) = \sum_{n=0}^{\infty} 0^n P(N(t) = n) = P(N(t) = 0),$$

so that

$$G(0, t) = \begin{cases} \left(\frac{\mu - \mu \exp(-(\lambda - \mu)t)}{\lambda - \mu \exp(-(\lambda - \mu)t)} \right)^{N(0)} & \text{if } \lambda \neq \mu \\ \left(\frac{\lambda t}{1 + \lambda t} \right)^{N(0)} & \text{if } \lambda = \mu \end{cases}$$

If we let $t \rightarrow \infty$ we get

$$P(\text{eventual extinction}) = \begin{cases} 1 & \text{if } \lambda \leq \mu \\ (\mu/\lambda)^{N(0)} & \text{if } \lambda > \mu. \end{cases}$$

What does all this mean?

If the birth rate is no larger than the death rate, the population is certain to become extinct eventually.

This implies that $\underline{\pi} = (1, 0, 0, \dots)$ is an **equilibrium distribution** if $\lambda \leq \mu$.

If, on the other hand, $\lambda > \mu$, then expected population size increases to ∞ as $t \rightarrow \infty$. You can see this by using $G(s, t)$ to show that

$$E[N(t)] = N(0) \exp(-(\lambda - \mu)t).$$

However, even under these conditions, extinction is still possible and so the population size will either increase without limit, or become extinct. As there is uncertainty as to how the process behaves in the future, there is **no equilibrium distribution**.

5.3 Exercises

The next questions are from previous exam papers. They are hard, but will test your understanding of birth-death processes.

Exercise 5.11

Suppose that arrivals to a utopian island occur as a Poisson process with rate λ per year. Those who arrive on the island can never leave, but each person on the island at time t has probability $\mu h + o(h)$ of dying in the interval $(t, t + h]$ years. Any individual on the island at time t has probability $\theta h + o(h)$ of giving birth in the interval $(t, t + h]$ years.

Arrivals, births and deaths are independent and the probability of two or more of these occurring in $(t, t + h]$ is $o(h)$. The parameters λ, μ, θ are all positive. Let X_t denote the population size of the island by time t .

- (a) Give the generator matrix for the process X .
- (b) Given that $X_t = 0$, state the distribution of the time until the next immigration, stating its mean.
- (c) Given that $X_t = i$, consider T_i , the time that it takes until the population size is $(i + 1)$ for the first time since time t . Justify that

$$E[T_i | \text{next event is a birth or immigration}] = \frac{1}{\lambda + i\theta + i\mu};$$

$$E[T_i | \text{next event is a death}] = \frac{1}{\lambda + i\theta + i\mu} + E[T_{i-1}] + E[T_i].$$

- (d) Using the expressions given in part (c), calculate the expected time until the population reaches size $(i + 1)$ for the first time, for $i = 0, 1, 2$.
- (e) Given that $X_t = 0$, give an expression for how long, on average, it takes for the population to reach size 2 for the first time.

Exercise 5.12

The population of an island evolves according to an immigration-death process. Specifically, the population evolves according to the following rules.

People immigrate onto the island in a Poisson process of rate α per year. Each person on the island at time t has probability $\mu h + o(h)$ of dying in the time interval $(t, t + h]$ years. There is no emigration from the island and there are no births.

Immigrations and deaths are independent and the probability of two or more events (immigrations or deaths) in $(t, t + h]$ is $o(h)$. The parameters α and μ are positive.

Let $N(t)$ denote the number of people on the island at time t .

- (a) Write down the state space S and the generator matrix Q of transition rates of the continuous-time Markov chain $\{N(t), t \geq 0\}$.
- (b) Let $\{Y_n, n = 0, 1, 2, \dots\}$ be the embedded discrete-time jump chain of $\{N(t); t \geq 0\}$.
 - (i) Write down the transition matrix P of $\{Y_n; n = 0, 1, 2, \dots\}$.
 - (ii) Let T_i be the holding time in state i , that is, the time spent in state i . State the distribution of T_i , for $i = 0, 1, 2, \dots$.
- (c) Suppose that $N(0) = 0$. Calculate the expected time until there are two people on the island.
- (d) You are given that, if $N(0) = 0$, the probability generating function $G(s, t) = E(s^{N(t)})$ of $N(t)$ is given by

$$G(s, t) = \exp \left\{ \frac{\alpha}{\mu} (1 - e^{-\mu t}) (s - 1) \right\}.$$

- (i) Use $G(s, t)$ to infer the distribution of $N(t) \mid N(0) = 0$.
 - (ii) Hence, or otherwise, infer the equilibrium distribution of $\{N(t); t \geq 0\}$.
- (e) Suppose now that the island has an active volcano. When the volcano erupts the whole population of the island is killed, reducing the population size to zero. Eruptions occur in a Poisson process of rate ϕ per year. Otherwise the population of the island evolves according to the immigration-death process described above.

Using your answer to (d)(i), or otherwise, calculate the expected number of people killed by the next eruption to occur after time 0, given that $N(0) = 0$.