MATH2750 Introduction to Markov Processes

Matthew Aldridge

University of Leeds, 2020–2021

Contents

Schedule

Week 11 (4–7 May):

- **Section 21**: Queues
- **Section 22**: End of Part II – Continuous time Markov jump processes
- **Problem Sheet 1**
- **Assessment 4**: due Thursday 6 May at 1400 (this week)
- **Lecture**: Tuesday at 1400 (Zoom)
- **Workshops** on Problem Sheet 7: Tuesday or Thursday (Zoom) – note some changes of dates
- **Drop-in sessions**: Tuesday or Wednesday (Teams)

Week 10 (26–30 April):

- **Section 19**: Class structure and hitting times
- **Section 20**: Long-term behaviour of Markov jump processes
- **Problem Sheet 10**
- **Assessment 4**: due Thursday 6 May (next week)

Easter break (29 March–23 April)

Week 9 (22–26 March):

- **Section 17**: Continuous time Markov jump processes
- **Section 18**: Forward and backward equations
- **Problem Sheet 9**
- **Assessment 3**: due Thursday 25 March (this week)

Week 8 (15–19 March):

- **Section 15**: Poisson process in infinitesimal time periods
- **Section 16**: Counting processes
- **Problem Sheet 8**
- **Computational Worksheet 2 / Assessment 2**: due Thursday 18 March at 1400 (this week)
- **Assessment 3**: due Thursday 25 March (next week)

Week 7 (8–12 March):

- **Section 13**: Poisson process with Poisson increments
- **Section 14**: Poisson process with exponential holding times
- **Problem Sheet 7**
- **Computational Worksheet 2 / Assessment 2**: computational drop-in sessions this week, due Thursday 18 March (next week)
- **Assessment 3**: due Thursday 25 March (week after next)

Week 6 (1–5 March):

- **Section 11**: Long-term behaviour of Markov chians
- **Section 12**: End of of Part I: Discrete time Markov chains
- **Problem Sheet 6**
- **Computational Worksheet 2 / Assessment 2**: computational drop-in sessions next week, due Thursday 18 March (week after next)
- **Assessment 3**: due Thursday 25 March (three weeks' time)

Week 5 (22–26 February):

- **Section 9**: Recurrence and transience
- **Section 10**: Stationary distributions
- **Problem Sheet 5**

Week 4 (15–19 February):

- **Section 7**: Class Structure
- **Section 8**: Hitting times
- **Problem Sheet 4**
- **Computational Worksheet 1**, with computational drop-in sessions

Week 3 (8–12 February):

- **Section 5**: Discrete time Markov chains
- **Section 6**: Examples from actuarial science
- **Problem Sheet 3**
- **Assessment 1** due Thursday 11 February (this week)
- **Computational Worksheet 1**: computational drop-in sessions next week

Week 2 (1–5 February):

- **Section 3**: Gambler's ruin
- **Section 4**: Linear difference equations
- **Problem Sheet 2**
- **Assessment 1** due Thursday 11 February (next week)

Week 1 (25–29 January):

- About the module
- **Section 1**: Stochastic processes and the Markov property
- **Section 2**: Random walks
- **Problem Sheet 1**

About MATH2750

This module is **MATH2750 Introduction to Markov Processes**. The module manager and lecturer is Dr Matthew Aldridge, and my email address is m.aldridge@leeds.ac.uk.

Organisation of MATH2750

This module lasts for 11 weeks. The first nine weeks run from 25 January to 26 March, then we break for Easter, and then the final two weeks run from 26 April to 7 May.

Notes and videos

The main way I expect you to learn the material for this course is by reading these notes and by watching the accompanying videos. I will set two sections of notes each week, for a total of 22 sections.

Reading mathematics is a slow process. Each section roughly corresponds to one lecture last year, which would have been 50 minutes. If you find yourself regularly getting through sections in much less than an hour, you're probably not reading carefully enough through each sentence of explanation and each line of mathematics, including understanding the motivation as well as checking the accuracy.

It is possible (but not recommended) to learn the material by only reading the notes and not watching the videos. It is not possible to learn the material by only watching the videos and not reading the notes.

You are probably reading the web version of the notes. If you want a PDF copy (to read offline or to print out), then click the PDF button in the top ribbon of the page. (Warning: I have not made as much effort to make the PDF neat and tidy as I have the web version.)

Since we will all be relying heavily on these notes, I'm even more keen than usual to hear about errors mathematical, typographical or otherwise. Please, please email me if think you may have found any.

Problem sheets

There will be 10 problem sheets; Problem Sheet n covers the material from the two sections from week n (Sections $2n - 1$ and $2n$), and will be discussed in your workshop in week $n + 1$.

Lectures

There will be one online synchronous "lecture" session each week, on Tuesdays at 1400, with me, run through Zoom.

This will not be a "lecture" in the traditional sense of the term, but will be an opportunity to re-emphasise material you have already learned from notes and videos, to give extra examples, and to answer common student questions, with some degree of interactivity.

I will assume you have completed all the work for the previous week by the time of the lecture, but I will not assume you've started the work for that week itself.

I am very keen to hear about things you'd like to go through in the lectures; please email me with your suggestions.

Workshops

There will be 10 workshops, starting in the second week. The main goal of the workshops will be to go over your answers to the problems sheets in smaller classes. You will have been assigned to one of three workshop groups, meeting on Mondays or Tuesdays, led by Jason Klebes, Dr Jochen Voss, or me. Your workshop will be run through Zoom or Microsoft Teams; your workshop leader will contact you before the end of this week with arrangements.

My recommended approach to problem sheets and workshops is the following:

- Work through the problem sheet before the workshop, spending plenty of time on it, and making multiple efforts at questions you get stuck on. I recommend spending *at least three hours* on each problem sheet, in more than one block. Collaboration is encouraged when working through the problems, but I recommend writing up your work on your own.
- Take advantage of the smaller group setting of the workshop to ask for help or clarification on questions you weren't able to complete.
- After the workshop, attempt again the questions you were previously stuck on.
- If you're still unable to complete a question after this second round of attempts, *then* consult the solutions.

Assessments

There will be four pieces of assessed coursework, making up a total of 15% of your mark for the module. Assessments 1, 3 and 4 will involve writing up answers to a few problems, in a similar style to the problem sheets, and are worth 4% each. (In response to previous student feedback, there are fewer questions per assessment.) Assessment 2 will be a report on some computational work (see below) and is worth 3%.

Copying, plagiarism and other types of cheating are not allowed and will be dealt with in accordance with University procedures.

The assessments deadlines are:

- Assessment 1: Thursday 11 February 1400 (week 3)
- Assessment 2 (Computational Worksheet 2): Thursday 18 March 1400 (week 8)
- Assessment 3: Thursday 25 March 1400 (week 9)
- Assessment 4: Thursday 6 May 1400 (week 11)

Work will be submitted via Gradescope.

Your markers are Jason Klebes, Macauley Locke, and Muyang Zhang – but you should contact the module leader if you have marking queries, not the markers directly.

Computing worksheets

There will be two computing worksheets, which will look at the material in the course through simulations in R. This material is examinable. You should be able to work through the worksheets in your own time, but if you need help, there will be optional online drop-in sessions in the weeks 4 and 7 with Muyang Zhang through Microsoft Teams. (Your computing drop-in session may be listed as "Practical" on your timetable.)

The first computing worksheet will be a practice run, while a report on the second computing worksheet will be the second assessed piece of work.

Drop-in sessions

If you there is something in the course you wish to discuss in detail, the place for the is the optional weekly drop-in session. The drop-in sessions are an optional opportunity for you to ask questions you have to a member of staff – nothing will happen unless you being your questions.

You will have been assigned to one of three groups on Tuesdays or Wednesdays with Nikita Merkulov or me. The drop-in sessions will be run the Microsoft Teams. Your drop-in session would be an excellent place to go if you are having trouble understanding something in the written notes, or if you're still struggling on a problem sheet question after your workshop.

Microsoft Team

I have set up a Microsoft Team for the course. I propose to use the "Q and A" channel there as a discussion board. This is a good place to post questions about material from the course, and – even better! – to help answer you colleagues' questions. The idea is that you all as a group should help each other out. I will visit a couple of times a week to clarify if everybody is stumped by a question, or if there is disagreement.

Time management

It is, of course, up to you how you choose to spend your time on this module. But, if you're interested, my recommendations would be something like this:

- **Every week:** 7.5 hours per week
  - **Notes and videos:** 2 sections, 1 hour each
  - **Problem sheet:** 3.5 hours per week
  - **Lecture:** 1 hour per week
  - **Workshop:** 1 hour per week
- **When required:**
  - **Assessments 1, 2 and 4:** 2 hours each
  - **Computer worksheets:** 2 hours each
  - **Revision**: 12 hours
- **Total:** 100 hours

Exam

There will be an exam – or, rather, a final "online time-limited assessment" – after the end of the module, making up the remaining 85% of your mark. The exam will consist of four questions, and you are expected to answer all of them. You will have 48 hours to complete the exam, although the exam itself should represent half a day to a day's work. Further details to follow nearer the time.

Who should I ask about...?

- *I don't understand something in the notes or on a problem sheet*: Go to your weekly drop-in session, or post a question on the Teams Q and A board. (If you email me, I am likely to respond, "That would be an excellent question for your drop-in session or the Q and A board.")
- *I don't understand something in on a computational worksheet:* Go to your computing drop-in session in weeks 4 or 7.
- *I have an admin question about general arrangements for the module:* Email me.
- *I have an admin question about arrangements for my workshop:* Email your workshop leader.
- *I have suggestion for something to cover in the lectures:* Email me.
- *I need an extension on or exemption from an assessment:* Email the Maths Taught Students Office.

Content of MATH2750

Prerequisites

Some students have asked what background you'll be expected to know for this course.

It's essential that you're very comfortable with the basics of probability theory: events, probability, discrete and continuous random variables, expectation, variance, approximations with the normal distribution, etc. Conditional probability and independence are particularly important concepts in this course. This course will use the binomial, geometric, Poisson, normal and exponential distributions, although the notes will usually remind you about them first, in case you've forgotten.

Many students on the module will have studied these topics in MATH1710 Probability and Statistics 1; others will have covered these in different modules.

Syllabus

The course has two major parts: the first part will cover processes in discrete time and the second part processes in continuous time.

An outline plan of the topics covered is the following. (Remember that one week's work is two sections of notes.)

- **Discrete time Markov chains** [12 sections]
    - Introduction to stochastic processes [1 section]
    - Important examples: Random walk, gambler's ruin, linear difference equations, examples from actuarial science [4 sections]
    - General theory: transition probabilities, n-step transition probabilities, class structure, periodicity, hitting times, recurrence and transience, stationary distributions, long-term behaviour [6 sections]
    - Revision [1 section]
- **Continuous time Markov jump processes** [10 sections]
    - Important examples: Poisson process, counting processes, queues [5 sections]
    - General theory: holding times and jump chains, forward and backward equations, class structure, hitting times, stationary distributions, long-term behaviour [4 sections]
    - Revision [1 section]

Books

You can do well on this module by reading the notes and watching the videos, attending the lectures and workshops, and working on the problem sheets, assignments and practicals, without any further reading. However, students can benefit from optional extra background reading or an alternative view on the material.

My favourite book on Markov chains, which I used a lot while planning this course and writing these notes, is:

- J.R. Norris, *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1997. Chapters 1-3.

This a whole book just on Markov processes, including some more detailed material that goes beyond this module. Its coverage of of both discrete and continuous time Markov processes is very thorough. Chapter 1 on discrete time Markov chains is available online.

Other good books with sections on Markov processes that I have used include:

- G.R. Grimmett and D.R. Stirzaker, *Probability and Random Processes*, 4th edition, Oxford University Press, 2020. Chapter 6.
- G. Grimmet and D. Walsh, *Probability: an introduction*, 2nd edition, Oxford University Press, 2014. Chapter 12.
- D.R. Stirzaker, *Elementary Probability*, 2nd edition, Cambridge University Press, 2003. Chapter 9.

Grimmett and Stirzaker is an excellent handbook that covers most of undergraduate probability – I bought a copy when I was a second-year undergraduate and still keep it next to my desk. The whole of Stirzaker is available online.

A gentler introduction with plenty of examples is provided by:

- P.W. Jones and P. Smith, *Stochastic Processes: an introduction*, 3nd edition, Texts in Statistical Science, CRC Press, 2018. Chapters 2-7.

although it doesn't cover everything in this module. The whole book is available online.

(I've listed the newest editions of these books, but older editions will usually be fine too.)

And finally...

These notes were mostly written by Matthew Aldridge in 2018–19, and have received updates (mostly in Sections 9–11) and reformatting this year. Some of the material (especially Section 1, Section 6, and numerous diagrams) follows closely previous notes by Dr Graham Murphy, and I also benefited from reading earlier notes by Dr Robert Aykroyd and Prof Alexander Veretennikov. Dr Murphy's general help and advice was also very valuable. Many thanks to students in previous runnings of the module for spotting errors and suggesting improvements.

Part I: Discrete time Markov chains

1 Stochastic processes and the Markov property

- Stochastic processes with discrete or continuous state space and discrete or continuous time
- The Markov "memoryless" property

1.1 Deterministic and random models

A **model** is an imitation of a real-world system. For example, you might want to have a model to imitate the world's population, the level of water in a reservoir, cashflows of a pension scheme, or the price of a stock. Models allow us to try to understand and predict what might happen in the real world in a low risk, cost effective and fast way.

To design a model requires a set of assumptions about how it will work and suitable parameters need to be determined, perhaps based on past collected data.

An important distinction is between **deterministic** models and **random** models. Another word for a random model is a **stochastic** ("*sto-KASS-tik*") model. Deterministic models do not contain any random components, so the output is completely determined by the inputs and any parameters. Random models have variable outcomes to account for uncertainty and unpredictability, so they can be run many times to give a sense of the range of possible outcomes.

Consider models for:

- the future position of the Moon as it orbits the Earth,
- the future price of shares in Apple.

For the moon, the random components – for example, the effect of small meteorites striking the Moon's surface – are not very significant and a deterministic model based on physical laws is good enough for most purposes. For Apple shares, the price changes from day to day are highly uncertain, so a random model can account for the variability and unpredictability in a useful way.

In this module we will see many examples of stochastic models. Lots of the applications we will consider come from financial mathematics and actuarial science where the use of models that take into account uncertainty is very important, but the principles apply in many areas.

1.2 Stochastic processes

If we want to model, for example, the total number of claims to an insurance company in the whole of 2020, we can use a random variable X to model this – perhaps a Poisson distribution with an appropriate mean. However, if we want to track how the number of claims changes over the course of the year 2021, we will need to use a **stochastic process** (or "random process").

A stochastic process, which we will usually write as (X_n), is an indexed sequence of random variables that are (usually) dependent on each other.

Each random variable X_n takes a value in a **state space** $\mathcal{S}$ which is the set of possible values for the process. As with usual random variables, the state space $\mathcal{S}$ can be **discrete** or **continuous**. A discrete state space denotes a set of distinct possible outcomes, which can be finite or countably infinite. For example, $\mathcal{S} = \{\text{Heads}, \text{Tails}\}$ is the state space for a single coin flip, while in the case of counting insurance claims, the state space would be the nonnegative integers $\mathcal{S} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$. A continuous state space denotes an uncountably infinite continuum of gradually varying outcomes. For example, the nonnegative real line $\mathcal{S} = \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ is the state space for the amount of rainfall on a given day, while some bounded subset of $\mathbb{R}^3$ would be the state space for the position of a gas particle in a box.

Further, the process has an **index set** that puts the random variables that make up the process in order. The index set is usually interpreted as a **time** variable, telling us when the process will be measured. The index set for time can also be discrete or continuous. Discrete time denotes a process sampled at distinct points, often denoted by $n = 0, 1, 2, \dots$, while continuous time denotes a process monitored constantly over time, often denoted by $t \in \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. In the insurance example, we might count up the number of claims each day – then the discrete index set will be the days of the year, which we could denote $\{1, 2, \dots, 365\}$. Alternatively, we might want to keep a constant tally that we update after every claim, requiring a continuous time index t representing time across the whole year. In discrete time, we can write down the first few steps of the process as $(X_0, X_1, X_2, \dots)$.

This gives us four possibilities in total:

- **Discrete time, discrete space**
    - Example: Number of students attending each lecture of maths module.
    - **Markov chains** – discrete time, discrete space stochastic processes with a certain "Markov property" – are the main topic of the first half of this module.

- **Discrete time, continuous space**
    - Example: Daily maximum temperature in Leeds.
    - We will briefly mention continuous space Markov chains in the first half of the course, but these are not as important.

- **Continuous time, discrete space**
    - Example: Number of visitors to a webpage over time.
    - **Markov jump processes** – continuous time, discrete space stochastic processes with the "Markov property" – are the main topic of the second half of this module.

- **Continuous time, continuous space**
    - Example: Level of the FTSE 100 share index over time.
    - Such processes, especially the famous Brownian motion – another process with the Markov property – are very important, but outside the scope of this course. See MATH3734 Stochastic Calculus for Finance next year, for example.

1.3 Markov property

Because stochastic processes consist of a large number – even infinitely many; even uncountably infinitely many – random variables that could all be dependent on each other, they can get extremely complicated. The Markov property is a crucial property that restricts the type of dependencies in a process, to make the process easier to study, yet still leaves most of the useful and interesting examples intact. (Although

particular examples of Markov processes go back further, the first general study was by the Russian mathematician Andrey Andreyevich Markov, published in 1906.)

Think of a simple board game where we roll a dice and move that many squares forward on the board. Suppose we are currently on the square X_n. Then what can we say about which square X_{n+1} we move to on our next turn?

- X_{n+1} is random, since it depends on the roll of the dice.
- X_{n+1} depends on where we are now X_n, since the score of dice will be added onto the number our current square,
- Given the square X_n we are now, X_{n+1} doesn't depend any further on which sequence of squares $X_0, X_1, ..., X_{n-1}$ we used to get here.

It is this third point that is the crucial property of the stochastic processes we will study in this course, and it is called the **Markov property** or **memoryless property**. We say "memoryless", because it's as if the process forgot how it got here – we only need to remember what square we've reached, not which squares we used to get here. The stochastic process before this moment has no bearing on the future, given where we are now. A mathematical way to say this is that "the past and the future are conditionally independent given the present."

To write this down formally, we need to recall **conditional probability**: the conditional probability of an event A given another event B is written $\mathbb{P}(A \mid B)$, and is the probability that A occurs *given* that B definitely occurs. You may remember the definition

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

although is often more useful to reason directly about conditional probabilities than use this formula.

(You may also remember that the definition of conditional probability requires that $\mathbb{P}(B) > 0$. Whenever we write down a conditional probability, we implicitly assume the conditioning event has strictly positive probability without explicitly saying so.)

Definition 1.1. Let $(X_n) = (X_0, X_1, X_2, ...)$ be a stochastic process in discrete time $n = 0, 1, 2, ...$ and discrete space $\mathcal{S}$. Then we say that (X_n) has the **Markov property** if, for all times n and all states $x_0, x_1, ..., x_n, x_{n+1} \in \mathcal{S}$ we have

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

Here, the left hand side is the probability we go to state x_{n+1} next conditioned on the entire history of the process, while the right hand side is the probability we go to state x_{n+1} next conditioned only on where we are now x_n. So the Markov property tells us that it only matters where we are now and not how we got here.

(There's also a similar definition for continuous time processes, which we'll come to later in the course.)

Stochastic processes that have the Markov property are much easier to study than general processes, as we only have to keep track of where we are now and we can forget about the entire history that came before.

In the next section, we'll see the first, and most important, example of a discrete time discrete space Markov chain: the "random walk".

2 Random walk

- Definition of the simple random walk and the exact binomial distribution
- Expectation and variance of general random walks

2.1 Simple random walk

Consider the following **simple random walk** on the integers $\mathbb{Z}$: We start at 0, then at each time step, we go up by one with probability p and down by one with probability $q = 1 - p$. When $p = q = \frac{1}{2}$, we're equally as likely to go up as down, and we call this the **simple symmetric random walk**.

The simple random walk is a simple but very useful model for lots of processes, like stock prices, sizes of populations, or positions of gas particles. (In many modern models, however, these have been replaced by more complicated continuous time and space models.) The simple random walk is sometimes called the "drunkard's walk", suggesting it could model a drunk person trying to stagger home.
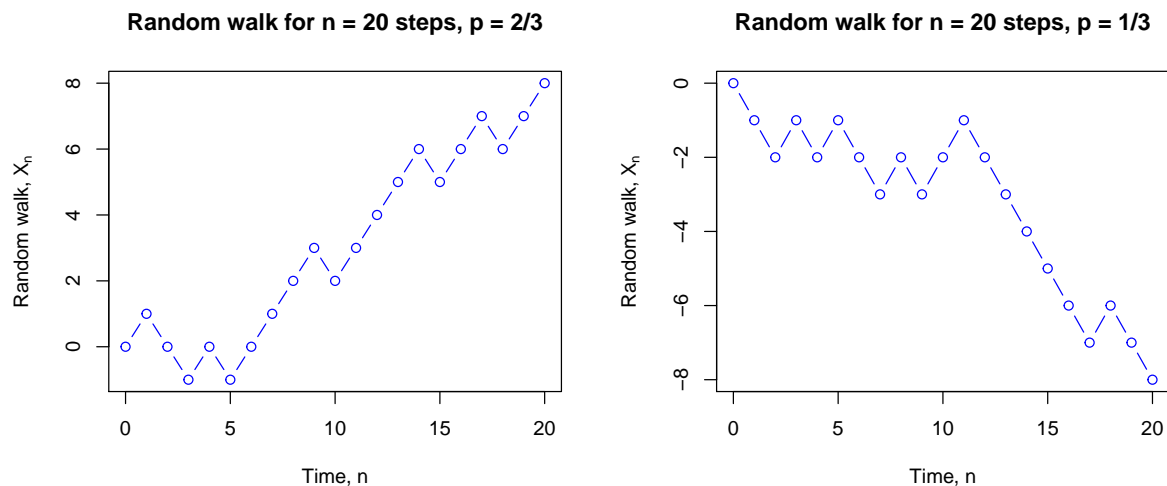
Figure 1: Two simulations of random walks.

We can write this as a stochastic process (X_n) with discrete time $n = \{0, 1, 2, \ldots\} = \mathbb{Z}_+$ and discrete state space $\mathcal{S} = \mathbb{Z}$, where $X_0 = 0$ and, for $n \geq 0$, we have

$$X_{n+1} = \begin{cases} X_n + 1 & \text{with probability } p, \\ X_n - 1 & \text{with probability } q. \end{cases}$$

It's clear from this definition that X_{n+1} (the future) depends on X_n (the present), but, given X_n, does not depend on $X_{n-1}, \ldots, X_1, X_0$ (the past). Thus the Markov property holds, and the simple random walk is a **discrete time Markov process** or **Markov chain**.

Example 2.1. *What is the probability that after two steps a simple random walk has reached $X_2 = 2$?*

To achieve this, the walk must go upwards in both time steps, so $\mathbb{P}(X_2 = 2) = pp = p^2$.

Example 2.2. *What is the probability that after three steps a simple random walk has reached $X_3 = -1$?*

There are three ways to reach -1 after three steps: up–down–down, down–up–down, or down–down–up. So

$$\mathbb{P}(X_3 = -1) = pqq + qpq + qqp = 3pq^2.$$

2.2 General random walks

An alternative way to write the simple random walk is to put

$$X_n = X_0 + \sum_{i=1}^{n} Z_i, \tag{1}$$

where the starting point is $X_0 = 0$ and the **increments** $Z_1, Z_2, \dots$ are independent and identically distributed (IID) random variables with distribution given by $\mathbb{P}(Z_i = 1) = p$ and $\mathbb{P}(Z_i = -1) = q$. You can check that (1) means that $X_{n+1} = X_n + Z_{n+1}$, and that this property defines the simple random walk.

Any stochastic process with the form (1) for some X_0 and some distribution for the IID Z_is is called a **random walk** (without the word "simple").

Random walks often have state space $\mathcal{S} = \mathbb{Z}$, like the simple random walk, but they could be defined on other state spaces. We could look at higher dimensional simple random walks: in $\mathbb{Z}^2$, for example, we could step up, down, left or right each with probability $\frac{1}{4}$. We could even have a continuous state space like $\mathbb{R}$, if, for example, the Z_is had a normal distribution.

We can use this structure to calculate the expectation or variance of any random walk (including the simple random walk).

Let's start with the expectation. For a random walk (X_n) we have

$$\mathbb{E}X_n = \mathbb{E}\left(X_0 + \sum_{i=1}^{n} Z_i\right) = \mathbb{E}X_0 + \sum_{i=1}^{n} \mathbb{E}Z_i = \mathbb{E}X_0 + n\mathbb{E}Z_1,$$

where we've used the linearity of expectation, and that the Z_is are identically distributed.

In the case of the simple random walk, we have $\mathbb{E}X_0 = 0$, since we start from 0 with certainty, and

$$\mathbb{E}Z_1 = \sum_{z \in \mathbb{Z}} z\mathbb{P}(Z_1 = z) = 1 \times p + (-1) \times q = p - q.$$

Hence, for the simple random walk, $\mathbb{E}X_n = n(p - q)$.

If $p > \frac{1}{2}$, then $p > q$, so $\mathbb{E}X_n$ grows ever bigger over time, while if $p < \frac{1}{2}$, then $\mathbb{E}X_n$ grows ever smaller (that is, negative with larger absolute value) over time. If $p = \frac{1}{2} = q$, which is the case of the simple symmetric random walk, then then the expectation $\mathbb{E}X_n = 0$ is zero for all time.

Now the variance of a random walk. We have

$$\mathrm{Var}(X_n) = \mathrm{Var}\left(X_0 + \sum_{i=1}^{n} Z_i\right) = \mathrm{Var}\,X_0 + \sum_{i=1}^{n} \mathrm{Var}\,Z_i = \mathrm{Var}\,X_0 + n\,\mathrm{Var}\,Z_1,$$

where it was crucial that X_0 and all the Z_is were independent (so we had no covariance terms).

For the simple random walk we have $\mathrm{Var}\,X_0 = 0$, since we always start from 0 with certainty. To calculate the variance of the increments, we write

$$\begin{aligned}
\mathrm{Var}(Z_1) &= \mathbb{E}(Z_1 - \mathbb{E}Z_1)^2 \\
&= p\big(1 - (p - q)\big)^2 + q\big(-1 - (p - q)\big)^2 \\
&= p(2q)^2 + q(-2p)^2 \\
&= 4pq^2 + 4p^2q \\
&= 4pq(p + q) \\
&= 4pq.
\end{aligned}$$

Here we've used that $1 - p = q$, $1 - q = p$, and $p + q = 1$; you should take a few moments to check you've followed the algebra here. Hence the variance of the simple random walk is $4pqn$. We see that (unless p is 0 or 1) the variance grows over time, so it becomes harder and harder to predict where the random walk will be.

The variance of the simple symmetric random walk is $4\frac{1}{2}\frac{1}{2}n = n$.

For large n, we can use a normal approximation for a random walk. Suppose the increments process (Z_n) has mean μ and variance σ^2, and that the walk starts from $X_0 = 0$. Then we have $\mathbb{E}X_n = \mu n$ and $\text{Var}(X_n) = \sigma^2 n$, so for large n we can use the normal approximation $X_n \approx \text{N}(\mu n, \sigma^2 n)$. (Remember, of course, that the X_n are not independent.) To be more formal, the central limit theorem tells us that, as $n \to \infty$, we have

$$\frac{X_n - n\mu}{\sigma\sqrt{n}} \to \text{N}(0, 1).$$

2.3   Exact distribution of the simple random walk

We have calculated the expectation and variance of any random walk. But for the simple random walk, we can in fact give the exact distribution, by writing down an exact formula for $\mathbb{P}(X_n = i)$ for any time n and any state i.

Recall that, at each of the first n times, we independently take an upward step with probability p, and otherwise take a downward step. So if we let Y_n be the number of upward steps over the first n time periods, we see that Y_n has a binomial distribution $Y_n \sim \text{Bin}(n, p)$.

Recall that the binomial distribution has probability

$$\mathbb{P}(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k},$$

for $k = 0, 1, \ldots, n$, where $\binom{n}{k}$ is the binomial coefficient "n choose k".

If $Y_n = k$, that means we've taken k upward steps and $n - k$ downward steps, leaving us at position $k - (n - k) = 2k - n$. Thus we have that

$$\mathbb{P}(X_n = 2k - n) = \mathbb{P}(Y_n = k) = \binom{n}{k} p^k q^{n-k}. \tag{2}$$

Note that after an odd number of time steps n we're always at an odd-numbered state, since $2k -$ odd = odd, while after an even number of time steps n we're always at an even-numbered state, since $2k -$ even = even.

Writing $i = 2k - n$ gives $k = (n + i)/2$ and $n - k = (n - i)/2$. So we can rearrange (2) to see that the distribution for the simple random walk is

$$\mathbb{P}(X_n = i) = \binom{n}{(n+i)/2} p^{(n+i)/2} q^{(n-i)/2},$$

when n and i have the same parity with $-n \le i \le n$, and is 0 otherwise.

In the special case of the simple symmetric random walk, we have

$$\mathbb{P}(X_n = i) = \binom{n}{(n+i)/2} \left(\frac{1}{2}\right)^{(n+i)/2} \left(\frac{1}{2}\right)^{(n-i)/2} = \binom{n}{(n+i)/2} 2^{-n}.$$

In the next section, we look at a gambling problem based on the simple random walk.

Problem Sheet 1

You should attempt all these questions and write up your solutions in advance of your workshop in week 2 (Monday 1 or Tuesday 2 February) where the answers will be discussed.

1. When designing a model for a quantity that changes over time, one has many decisions to make:

- Discrete or continuous state space?
- Discrete or continuous index set for time?
- Deterministic or stochastic model?
- If a stochastic model is chosen, is it reasonable to assume that the Markov property holds?

What would you decide for the following scenarios:

(a) The percentage of UK voters with a positive opinion of Boris Johnson in a weekly tracking poll.

(b) The number of points won by a football league club throughout the season.

(c) The temperature of a bowl of water placed in an oven.

(d) The number of people inside the University of Leeds library.

2. A fair six-sided dice is rolled twice, resulting in the values $X_1, X_2 \in \{1, 2, \ldots, 6\}$. Let $Y = X_1 + X_2$ be the total score. Calculate:

(a) the probability $\mathbb{P}(Y = 10)$;

(b) the conditional probability $\mathbb{P}(Y = 10 \mid X_1 = x)$ for $x = 1, 2, \ldots, 6$;

(c) the conditional probability $\mathbb{P}(X_1 = x \mid Y = 10)$ for $x = 1, 2, \ldots, 6$.

3. Let (X_n) be a simple random walk starting from $X_0 = 0$ and that at each step goes up one with probability p or down one with probability $q = 1 - p$. What are:

(a) $\mathbb{P}(X_5 = 3)$,

(b) $\mathbb{P}(X_5 = 3 \mid X_2 = 2)$,

(c) $\mathbb{P}(X_n = n - 2)$,

(d) $\mathbb{E}X_4$,

(e) $\mathbb{E}(X_6 \mid X_4 = 2)$,

4. The price X_n of a stock at the close of day n is modelled as a Gaussian random walk, where the increments (Z_n) have a normal distribution $Z_n \sim \mathrm{N}(\mu, \sigma^2)$. The model assumes a drift of $\mu = 0.7$ and a volatility of $\sigma = 2.2$. The initial price is $X_0 = 42.3$.

(a) Calculate the mean and variance of the price of the stock at the close of day 5.

(b) Give a 95% prediction interval for the price at the close of day 5. (You might find it useful to recall that, if $W \sim \mathrm{N}(0, 1)$ is a standard normal random variable, then $\mathbb{P}(W \leq 1.96) = 0.975$.)

(c) After day 4, the prices at the end of each of the first four days have been recorded as $X_1 = 44.4, X_2 = 44.0, X_3 = 47.1, X_4 = 47.8$. Update your prediction interval for the price at the close of day 5, and comment on how it differs from the earlier prediction interval.

5. A gambler decides to model her total winnings as a simple random walk starting from $X_0 = 0$ that at each time goes up one with probability p and down one with probability $1 - p$, but where p is unknown. The first 10 recordings, X_1 to X_{10}, are

$$(1, 2, 1, 2, 3, 4, 5, 6, 5, 6).$$

(a) What would you guess for the value of p, given this data?

(b) More generally, how would you estimate p from the data $X_0 = 0, X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$?

(c) Show that your estimate is in fact the maximum likelihood estimate of p.

3   Gambler's ruin

- The gambler's ruin Markov chain
- Equations for probability of ruin and expected duration of the game by conditioning on the first step

3.1   Gambler's ruin Markov chain

Consider the following gambling problem. Alice is gambling against Bob. Alice starts with £a and Bob starts with £b. It will be convenient to write $m = a + b$ for the total amount of money, so Bob starts with £$(m - a)$. At each step of the game, both players bet £1; Alice wins £1 off Bob with probability p, or Bob wins £1 off Alice with probability q. The game continues until one player is out of money (or is "ruined").

Let X_n denote how much money Alice has after n steps of the game. We can write this as a stochastic process with discrete time $n \in \{0, 1, 2, \dots\} = \mathbb{Z}_+$ and discrete state space $\mathcal{S} = \{0, 1, \dots, m\}$. Then $X_0 = a$, and, for $n \geq 0$, we have

$$X_{n+1} = \begin{cases} X_n + 1 & \text{with probability } p \text{ if } 1 \leq X_n \leq m - 1, \\ X_n - 1 & \text{with probability } q \text{ if } 1 \leq X_n \leq m - 1, \\ 0 & \text{if } X_n = 0, \\ m & \text{if } X_n = m. \end{cases}$$

So Alice's money goes up one with probability p or down one with probability q, unless the game is already over with $X_n = 0$ (Alice is ruined) or $X_n = m$ (Alice has won all Bob's money, so Bob in ruined).

We see that the gambler's ruin process (X_n) clearly satisfies the Markov property: the next step X_{n+1} depends on where we are now X_n, but, given that, does not depend on how we got here.

The gambler's ruin process is exactly like a simple random walk started from $X_0 = a$ except that we have **absorbing barriers** and 0 and m, where the random walk stops because one of the players has ruined. (One could also consider random walks with **reflecting barriers**, that bounce the random walk back into the state space, or **mixed barriers** that are absorbing or reflecting at random.)

There are two questions about the gambler's ruin that we'll try to answer in this section:

1. What is the probability that the game ends by Alice ruining?
2. How long does the game last on average?

3.2   Probability of ruin

The gambling game continues until either Alice is ruined ($X_n = 0$) or Bob is ruined ($X_n = m$). A natural question to ask is: What is the probability that the game ends in Alice's ruin?

Let us write r_i for the probability Alice ends up ruined if she currently has £i. Then the probability of ruin for the whole game is r_a, since Alice initially starts with £a. The probability Bob will end up ruined is $1 - r_a$, since one of the players must lose.

What can we say about r_i? Clearly we have $r_0 = 1$, since $X_n = 0$ means that Alice has run out of money and is ruined, and $r_m = 0$, since $X_n = m$ means that Alice has won all the money and Bob is ruined. What about when $1 \leq i \leq m - 1$?

The key is to *condition on the first step*. That is, we can write

$$\begin{aligned} \mathbb{P}(\text{ruin}) &= \mathbb{P}(\text{win first round})\,\mathbb{P}(\text{ruin} \mid \text{win first round}) \\ &\quad + \mathbb{P}(\text{lose first round})\,\mathbb{P}(\text{ruin} \mid \text{lose first round}) \\ &= p\,\mathbb{P}(\text{ruin} \mid \text{win first round}) + q\,\mathbb{P}(\text{ruin} \mid \text{lose first round}). \end{aligned}$$

Here we have conditioned on whether Alice wins or loses the first round. More formally, we have used the **law of total probability**, which says that if the events $B_1, \dots, B_k$ are disjoint and cover the whole sample space, then

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(B_i)\,\mathbb{P}(A \mid B_i).$$

Here, {Alice wins the first round} and {Alice loses the first round} are indeed disjoint events that cover the whole sample space. This idea of "conditioning on the first step" will be the most crucial tool throughout this whole module.

If Alice wins the first round from having £i, she now has £$(i+1)$. Her probability of ruin is now r_{i+1}, because, by the Markov property, it's as if the game were starting again with Alice having £$(i+1)$ to start with. The Markov property tells us that it doesn't matter *how* Alice got to having £$(i+1)$, it only matters how much she has now. Similarly, if Alice loses the first round, she now has £$(i-1)$, and the ruin probability is r_{i-1}. Hence we have

$$r_i = pr_{i+1} + qr_{i-1}.$$

Rearranging, and including the "boundary conditions", we see that the equation we want to solve is

$$pr_{i+1} - r_i + qr_{i-1} = 0 \qquad \text{subject to} \qquad r_0 = 1,\ r_m = 0.$$

This is a **linear difference equation** – and, because the left-hand side is 0, we call it a **homogeneous** linear difference equation.

We will see how to solve this equation in the next lecture. We will see that, if we set $\rho = q/p$, then the ruin probability is given by

$$r_a = \begin{cases} \dfrac{\rho^a - \rho^m}{1 - \rho^m} & \text{if } \rho \neq 1, \\[2mm] 1 - \dfrac{a}{m} & \text{if } \rho = 1. \end{cases}$$

Note that $\rho = 1$ is the same as the symmetric condition $p = q = \frac{1}{2}$.

Imagine Alice is not playing against a similar opponent Bob, but rather is up against a large casino. In this case, the casino's capital £$(m - a)$ is typically much bigger than Alice's £a. We can model this by keeping a fixed taking a limit $m \to \infty$. Typically, the casino has an "edge", meaning they have a better than $50:50$ chance of winning; this means that $q > p$, so $\rho > 1$. In this case, we see that the ruin probability is

$$\lim_{m\to\infty} r_a = \lim_{m\to\infty} \frac{\rho^a - \rho^m}{1 - \rho^m} = \lim_{m\to\infty} \frac{\rho^a/\rho^m - 1}{1/\rho^m - 1} = \frac{0-1}{0-1} = 1,$$

so Alice will be ruined with certainty.

Even with a generous casino that offered an exactly fair game with $p = q = \frac{1}{2}$, so $\rho = 1$, we would have

$$\lim_{m\to\infty} r_a = \lim_{m\to\infty} \left(1 - \frac{a}{m}\right) = 1 - 0 = 1,$$

so, even with this fair game, Alice would still be ruined with certainty.

(The official advice of the University of Leeds module MATH2750 is that you shouldn't gamble against a casino if you can't afford to lose.)

3.3 Expected duration of the game

We could also ask for how long we expect the game to last.

We approach this like before. Let d_i be the expected duration of the game from a point when Alice has £i. Our boundary conditions are $d_0 = d_m = 0$, because $X_n = 0$ or $X_n = m$ means that the game is over with Alice or Bob ruined. Again, we proceed by conditioning on the first step, so

$$\begin{aligned}
\mathbb{E}(\text{duration}) = {}& \mathbb{P}(\text{win first round})\,\mathbb{E}(\text{duration} \mid \text{win first round}) \\
& + \mathbb{P}(\text{lose first round})\,\mathbb{E}(\text{duration} \mid \text{lose first round}) \\
= {}& p\,\mathbb{E}(\text{duration} \mid \text{win first round}) + q\,\mathbb{E}(\text{duration} \mid \text{lose first round}).
\end{aligned}$$

More formally, we've used another version of the law of total probability,

$$\mathbb{E}(X) = \sum_{i=1}^{k} \mathbb{P}(B_i)\,\mathbb{E}(X \mid B_i),$$

or, alternatively, the **tower law** for expectations

$$\mathbb{E}(X) = \mathbb{E}_Y\mathbb{E}(X \mid Y) = \sum_y \mathbb{P}(Y = y)\,E(X \mid Y = y),$$

where, in our case, Y was the outcome of the first round.

Now, the expected duration given we win the first round is $1 + d_{i+1}$. This is because the round itself takes 1 time step, and then, by the Markov property, it's as if we are starting again from $i+1$. Similarly, the expected duration given we lose the first round is $1 + d_{i-1}$. Thus we have

$$d_i = p(1 + d_{i+1}) + q(1 + d_{i-1}) = 1 + pd_{i+1} + qd_{i-1}.$$

Don't forget the 1 that counts the current round!

Rearranging, and including the boundary conditions, we have another linear difference equation:

$$pd_{i+1} - d_i + qd_{i-1} = -1 \qquad \text{subject to} \qquad d_0 = 0,\ d_m = 0.$$

Because the right-hand side, -1, is nonzero, we call this an **inhomogeneous** linear difference equation.

Again, we'll see how to solve this in the next lecture, and will find that the solution is given by

$$d_a = \begin{cases} \dfrac{1}{q - p}\left(a - m\dfrac{1 - \rho^a}{1 - \rho^m}\right) & \text{if } \rho \neq 1, \\ a(m - a) & \text{if } \rho = 1. \end{cases}$$

Thinking again of playing against the casino, with $q > p$, $\rho > 1$, and $m \to \infty$, we see that the expected duration is

$$\lim_{m \to \infty} d_a = \lim_{m \to \infty} \frac{1}{q - p}\left(a - m\frac{1 - \rho^a}{1 - \rho^m}\right) = \frac{1}{q - p}(a - 0) = \frac{a}{q - p},$$

since ρ^m grows much quicker than m. So Alice ruins with certainty, and it will take time $a/(q - p)$, on average.

In the case of the generous casino, though, with $q = p$, so $\rho = 1$, we have

$$\lim_{m \to \infty} d_a = \lim_{m \to \infty} a(m - a) = \infty.$$

So here, Alice will ruin with certainty, but it may take a very long time until the ruin occurs, since the expected duration is infinite.

In the next section, we see how to solve linear difference equations, in order to find the ruin probability and expected duration of the gambler's ruin.

4   Linear difference equations

- How to solve homogeneous and inhomogeneous linear difference equations
- Solving for probability of ruin and expected duration of the gambler's ruin

In the previous section, we looked at the probability of ruin and expected duration of the gambler's ruin process. We set up linear difference equations to find these. In this section, we'll learn how to solve these equations.

A **linear difference equation** is an equation that looks like

$$a_k x_{n+k} + a_{k-1} x_{n+k-1} + \cdots + a_1 x_{n+1} + a_0 x_n = f(n) \tag{3}$$

for $n = 0, 1, \ldots$, where the a_i are given constants, $f(n)$ is a given function, and we want to solve for the sequence (x_n). The equation normally comes with some extra conditions, such as the value of the first few x_ns.

When the right-hand side of (3) is zero, so $f(n) = 0$, we say the equation is **homogeneous**; when the right-hand side is nonzero, it is **inhomogeneous**. The number k, where there are $k + 1$ terms on the left-hand side, is called the **degree** of the equation; we are mostly interested in second-degree linear difference equations, which have three terms on the left-hand side.

4.1   Homogeneous linear difference equations

We start with the homogeneous case, which is simpler.

Consider a homogeneous linear difference equation. We shall use the second-degree example

$$x_{n+2} - 5x_{n+1} + 6x_n = 0 \qquad \text{subject to } x_0 = 4, x_1 = 9.$$

Here, the conditions on x_0 and x_1 are **initial conditions**, because they tell us how the sequence (x_n) starts.

For the moment, we shall put the initial conditions to the side and just worry about the equation

$$x_{n+2} - 5x_{n+1} + 6x_n = 0.$$

We start by guessing there might be a solution of the form $x_n = \lambda^n$ for some constant λ. We can find out if there is such a solution by substituting in $x_n = \lambda^n$, and seeing if there's a λ that solves the equation. For our example, we get

$$\lambda^{n+2} - 5\lambda^{n+1} + 6\lambda^n = 0.$$

After cancelling off a common factor of λ^n, we get

$$\lambda^2 - 5\lambda + 6 = 0.$$

This is called the **characteristic equation**. For a general homogeneous linear difference equation (3), the characteristic equation is

$$a_k \lambda^k + a_{k-1} \lambda^{k-1} + \cdots + a_1 \lambda + a_0 = 0. \tag{4}$$

When writing out answers to questions, you can jump straight to the characteristic equation.

We can now solve the characteristic equation for λ. In our example, we can factor the left-hand side to get $(\lambda - 3)(\lambda - 2) = 0$, which has solutions $\lambda = 2$ and $\lambda = 3$. Thus $x_n = 2^n$ and $x_n = 3^n$ both solve our equation. In fact, since the right-hand side of the equation is 0, any linear combination of these two solutions is a solution also, thus we get the **general solution**

$$x_n = A2^n + B3^n,$$

which is a solution for any values of the constants A and B.

For a general characteristic equation with distinct roots $\lambda_1, \lambda_2, \dots, \lambda_k$, the general solution is

$$x_n = C_1 \lambda_1^n + C_2 \lambda_2^n + \cdots + C_k \lambda_k^n.$$

If we have a repeated root – say, $\lambda_1 = \lambda_2 = \cdots = \lambda_r$ is repeated r times – than you can check that a solution is given by

$$x_n = (D_0 + D_1 n + \cdots + D_{r-1} n^{r-1}) \lambda_1^n,$$

which should take its place in the general solution.

Once we have the general solution, we can use the extra conditions to find the values of the constants. In our example, we can use the initial conditions to find out the values of A and B. We see that

$$x_0 = A2^0 + B3^0 = A + B = 4,$$
$$x_1 = A2^1 + B3^1 = 2A + 3B = 9.$$

We can now solve this pair of simultaneous equations to solve for A and B. By subtracting twice the first equation from the second we get $B = 1$, and substituting this into the first equation we get $A = 3$. Thus the solution is

$$x_n = 3 \cdot 2^n + 3^n.$$

In conclusion, the process here was:

1. Find the general solution by writing down and solving the characteristic equation.
2. Use the extra conditions to find the values of the constants in the general solution.

Here are two more examples. These also give an idea of how I would expect you to set out your own answers to similar problems.

Example 4.1. *Solve the homogeneous linear difference equation*

$$x_{n+2} - x_{n+1} - 6x_n = 0 \qquad \text{subject to} \quad x_0 = 3, \quad x_1 = 4.$$

Step 1. The characteristic equation is
$$\lambda^2 - \lambda - 6 = 0.$$
We can solve this by factorising it as $(\lambda - 3)(\lambda + 2) = 0$, to find the solutions $\lambda_1 = -2$ and $\lambda_2 = 3$. Thus the general solution is
$$x_n = A(-2)^n + B3^n.$$

Step 2. Substituting the initial conditions into the general solution, we have

$$x_0 = A(-2)^0 + B3^0 = A + B = 3$$
$$x_1 = A(-2)^1 + B3^1 = -2A + 3B = 4.$$

We can add twice the first equation to the second to get $5B = 10$, so $B = 2$. We can substitute this into the first equation to get $A = 1$.

The solution is therefore
$$x_n = 1 \cdot (-2)^n + 2 \cdot 3^n = (-2)^n + 2 \cdot 3^n.$$

Example 4.2. *Solve the homogeneous linear difference equation*

$$x_{n+2} + 4x_{n+1} + 4x_n = 0 \qquad \text{subject to} \quad x_0 = 2, \quad x_1 = -6.$$

Step 1. The characteristic equation is
$$\lambda^2 + 4\lambda + 4 = 0.$$
We can solve this by factorising it as $(\lambda + 2)^2 = 0$, to find a repeated root $\lambda_1 = \lambda_2 = -2$. Thus the general solution is
$$x_n = (A + Bn)(-2)^n.$$

Step 2. Substituting the initial conditions into the general solution, we have

$$x_0 = (A + B0)(-2)^0 = A = 2$$
$$x_1 = (A + B1)(-2)^1 = -2A - 2B = -6.$$

The first immediately gives $A = 2$, and substituting this into the second equation gives $B = 1$.

The solution is therefore

$$x_n = (2 + n)(-2)^n.$$

4.2 Probability of ruin for the gambler's ruin

In the last lecture we saw that probability of ruin for the gambler's ruin process is the solution to

$$pr_{i+1} - r_i + qr_{i-1} = 0 \qquad \text{subject to} \qquad r_0 = 1, \ r_m = 0,$$

where the extra conditions here are **boundary conditions**, because they tell us what happens at the boundaries of the state space.

The characteristic equation is

$$p\lambda^2 - \lambda + q = 0.$$

We can solve the characteristic equation by factorising it as $(p\lambda - q)(\lambda - 1) = 0$. (It might take a moment to check this really is a factorisation of the characteristic equation. Hint: we've used that $p + q = 1$.) So the characteristic equation has roots $\lambda = q/p$, which we called ρ last time, and $\lambda = 1$. Now, if $\rho = 1$ (so $p = q = \frac{1}{2}$) we have a repeated root, while if $\rho \neq 1$ we have distinct roots, so we'll need to deal with the two cases separately.

First, the case $\rho \neq 1$. Since the two roots are distinct, we have the general solution

$$r_i = A\rho^i + B1^i = A\rho^i + B.$$

We can now use the boundary conditions to find A and B. We have

$$r_0 = A\rho^0 + B = A + B = 1,$$
$$r_m = A\rho^m + B = 0.$$

From the first we get $B = 1 - A$, which we substitute into the second to get

$$A\rho^m + 1 - A = 0 \quad \Rightarrow \quad A = \frac{1}{1 - \rho^m},$$

and hence

$$B = 1 - A = 1 - \frac{1}{1 - \rho^m} = -\frac{\rho^m}{1 - \rho^m}.$$

Thus the solution is

$$r_i = \frac{1}{1 - \rho^m}\rho^i - \frac{\rho^m}{1 - \rho^m} = \frac{\rho^i - \rho^m}{1 - \rho^m},$$

as we claimed last time.

Second, the case $\rho = 1$. Now we have a repeated root $\lambda = 1$, so the general solution is

$$r_i = (A + Bi)1^i = A + Bi.$$

Again, we use the boundary conditions, to get

$$r_0 = A + B \cdot 0 = A = 1,$$
$$r_m = A + Bm = 0,$$

and we immediately see that $A = 1$ and $B = -1/m$. Thus the solution is

$$r_i = 1 - \frac{1}{m}i = 1 - \frac{i}{m},$$

as claimed last time.

4.3 Inhomogeneous linear difference equations

Solving inhomogeneous linear difference equations requires three steps:

1. Find the general solution to the *homogeneous* equation by writing down and solving the characteristic equation.
2. By making an "educated guess", find a solution (a "particular solution") to the inhomogeneous equation. The general solution to the inhomogeneous equation is a particular solution plus the general solution to the homogeneous equation.
3. Use the extra conditions to find the values of the constants in the general solution.

This idea works because, once you have a particular solution, adding a solution to the *homogeneous* equation to the left-hand side adds zero to the right-hand side, so maintains a solution to the inhomogeneous equation.

Let's work through the example

$$x_{n+2} - 5x_{n+1} + 6x_n = 2 \qquad \text{subject to } x_0 = 4, x_1 = 9.$$

We already know from earlier that the general solution to the homogeneous equation $x_{n+2} - 5x_{n+1} + 6x_n = 0$ (with a zero on the right-hand side) is

$$x_n = A2^n + B3^n.$$

We now need to find a **particular solution** – that is, any solution – to our new inhomogeneous equation. The usual process here is to guess a solution with the same "shape" as the right-hand side. For example, if the right-hand side is a constant, try a constant for the particular solution. Here our right-hand side is the constant 2, so we should try a constant $x_n = C$. Substituting this into the inhomogeneous equation gives us $C - 5C + 6C = 2$, thus $2C = 2$ and $C = 1$, giving a particular solution $x_n = 1$. The general solution to the inhomogeneous equation is therefore

$$x_n = 1 + A2^n + B3^n,$$

the sum of the particular solution $x_n = 1$ and the general solution to the homogeneous equation.

Because the right-hand side was a constant, we guessed a constant – this is the main case we will deal with. Other cases you could come across include:

- If the right-hand side is a polynomial of degree d, try a polynomial of degree d.
- If the right-hand side is α^n for some α, try $C\alpha^n$.
- If the right-hand side is a constant, but a constant C doesn't work, try Cn. If that still doesn't work, try Cn^2, and so on. A general rule is that is 1 is a root of the characteristic equation with multiplicity m, you need to try Cn^m. We discuss this case further in the next subsection.

Continuing with the example, we use the initial conditions to get the constants A and B. We have

$$x_0 = 1 + A2^0 + B3^0 = 1 + A + B = 4,$$
$$x_1 = 1 + A2^1 + B3^1 = 1 + 2A + 3B = 9.$$

The second equation minus twice the first gives $-1 + B = 1$, so $B = 2$, and substituting that back into the first gives $A = 1$. Thus the solution is

$$x_n = 1 + 1 \cdot 2^n + 2 \cdot 3^n = 1 + 2^n + 2 \cdot 3^n.$$

Here's another example.

Example 4.3. *Solve the inhomogeneous linear difference equation*

$$10x_{n+2} - 7x_{n+1} + x_n = 8 \qquad \text{subject to} \quad x_0 = 0, \quad x_1 = \tfrac{13}{10}.$$

Step 1. The characteristic equation is

$$10\lambda^2 - 7\lambda + 1 = 0.$$

We can solve this by factorising it as

$$(2\lambda - 1)(5\lambda - 1) = 0,$$

to find the solutions $\lambda_1 = \tfrac{1}{2}$ and $\lambda_2 = \tfrac{1}{5}$. Thus the general solution of the homogeneous equation is

$$x_n = A\left(\frac{1}{2}\right)^n + B\left(\frac{1}{5}\right)^n.$$

Step 2. Since the right hand side of the inhomogeneous equation is a constant, we guess a constant particular solution with shape $x_n = C$. Substituting in this guess, we get

$$10C - 7C + C = 4C = 8$$

with solution $C = 2$. Thus a particular solution is $x_n = 2$, and the general solution to the inhomogeneous equation is

$$x_n = 2 + A\left(\frac{1}{2}\right)^n + B\left(\frac{1}{5}\right)^n.$$

Step 3. Substituting the initial conditions into the general solution, we have

$$x_0 = 2 + A\left(\frac{1}{2}\right)^0 + B\left(\frac{1}{5}\right)^0 = 2 + A + B = 0 \quad \Rightarrow \quad A + B = -2$$

$$x_1 = 2 + A\left(\frac{1}{2}\right)^1 + B\left(\frac{1}{5}\right)^1 = 2 + A\frac{1}{2} + B\frac{1}{5} = \frac{13}{10} \quad \Rightarrow \quad 5A + 2B = -7.$$

We can take twice the first equation abd subtract the second to get $-3A = 3$, so $A = -1$. We can substitute this into the second equation to get $B = -1$.

The solution is therefore

$$x_n = 2 - \left(\frac{1}{2}\right)^n - \left(\frac{1}{5}\right)^n.$$

4.4 Expected duration for the gambler's ruin

From last time, the expected duration of the gambler's ruin game solves

$$pd_{i+1} - d_i + qd_{i-1} = -1 \qquad \text{subject to} \qquad d_0 = 0, \ d_m = 0.$$

As before, we divide cases based on whether or not $\rho = 1$.

First, the case $\rho \neq 1$. We already know that the general solution to the homogeneous equation is

$$d_i = A\rho^i + B.$$

Now we need a particular solution. It's tempting to guess a constant C for a particular solution, but we know that constants solve the homogeneous equation, since $d_i = B$ is a solution, so a constant will give right-hand side 0, not -1. (We could try out $x_i = C$ if we wanted; we would get $(p - 1 + q)C = -1$, but $p - 1 + q = 0$, and $0 \times C = -1$ has no solution.) The next best try is to go one degree up: let's guess $x_i = Ci$ instead. This gives

$$-1 = pC(i + 1) - Ci + qC(i - 1)$$
$$= C(pi + p - i + qi - q)$$
$$= C((p + q - 1)i + (p - q))$$
$$= C(p - q),$$

since $p + q - 1 = 1 - 1 = 0$. This $C = -1/(p - q) = 1/(q - p)$. Finding a solution for C shows that our guess worked. The general solution to the inhomogeneous equation is

$$d_i = \frac{i}{q - p} + A\rho^i + B.$$

Then to find the constants, we have

$$d_0 = \frac{0}{q - p} + A\rho^0 + B = A + B = 0,$$
$$d_m = \frac{m}{q - p} + A\rho^m + B = 0,$$

which you can check gives

$$A = -B = \frac{m}{q - p} \cdot \frac{1}{1 - \rho^m}.$$

Hence, the solution is

$$d_i = \frac{i}{q - p} + \frac{m}{q - p}\frac{1}{1 - \rho^m}\rho^i - \frac{m}{q - p}\frac{1}{1 - \rho^m} = \frac{1}{q - p}\left(i - m\frac{1 - \rho^i}{1 - \rho^m}\right).$$

Second, the case $\rho = 1$, so $p = q = \frac{1}{2}$. We already know that the general solution to the homogeneous equation is

$$d_i = A + Bi.$$

We need a particular solution. Since 1 is a double root of the characteristic equation, both constants $x_i = A$ and linear $x_i = Bi$ terms solve the homogeneous equation. (You can check that guessing $x_i = C$ or $x_i = Ci$ doesn't work, if you like.) So we'll have to go up another degree and try $x_i = Ci^2$. This gives

$$
\begin{aligned}
-1 &= \tfrac{1}{2}C(i + 1)^2 - Ci^2 + \tfrac{1}{2}C(i - 1)^2 \\
&= \tfrac{1}{2}C(i^2 + 2i + 1 - 2i^2 + i^2 - 2i + 1) \\
&= \tfrac{1}{2}C((1 - 2 + 1)i^2 + (2 - 2)i + (1 + 1)) \\
&= C,
\end{aligned}
$$

so the general solution to the inhomogeneous equation is

$$d_i = -i^2 + A + Bi.$$

Then to find the constants, we have

$$d_0 = -0^2 + A + B \cdot 0 = A = 0,$$
$$d_m = -m^2 + A + Bm = 0,$$

giving $A = 0, B = m$. The solution is

$$d_i = -i^2 + 0 + mi = i(m - i).$$

In the next section, we move on from the specific cases we've looked at so far to the general theory of discrete time Markov chains.

Problem sheet 2

You should attempt all these questions and write up your solutions in advance of your workshop in week 3 (Monday 8 or Tuesday 9 February) where the answers will be discussed.

1. Solve the following linear difference equations:

(a) $x_{n+2} - 4x_{n+1} + 3x_n = 0$, subject to $x_0 = 0$, $x_1 = 2$.

(b) $4x_{n+1} = 4x_n - x_{n-1}$, subject to $x_0 = 1$, $x_1 = 0$.

(c) $x_n - 5x_{n-1} + 6x_{n-2} = 1$, subject to $x_0 = 1$, $x_1 = 2$.

(d) $x_{n+2} - 2x_{n+1} + x_n = -1$, subject to $x_0 = 0$, $x_1 = 2$.

2. Consider a simple symmetric random walk on the state space $\mathcal{S} = \{0, 1, \ldots, m\}$ with an absorbing barrier at 0 and a reflecting barrier at m. In other words,

$$\mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = 1 \quad \text{and} \quad \mathbb{P}(X_{n+1} = m - 1 \mid X_n = m) = 1.$$

Let η_i be the expected time until the the walk hits 0 when starting from $i \in \mathcal{S}$.

(a) Explain why (η_i) satisfies

$$\eta_i = 1 + \tfrac{1}{2}\eta_{i+1} + \tfrac{1}{2}\eta_{i-1}$$

for $i \in \{1, 2, \ldots, m - 1\}$.

(b) Give a similar equation for η_m, and state the value of η_0.

(c) Hence, find the value of η_i for all $i \in \mathcal{S}$.

(d) You should notice that your answer is the same as the expected duration of the gambler's ruin for $p = \frac{1}{2}$, except with m replaced by $2m$. Can you explain why this might be?

3. Consider the gambler's ruin problem with draws: at each step, Alice wins £1 with probability p, loses £1 with probability q, and neither wins nor loses any money with probability s, where $p + q + s = 1$, and $0 < p, q, s < 1$. Alice starts with £a and Bob with £$(m - a)$.

(a) Let r_i be Alice's probability of ruin given that she has £i.

(i) Write down a linear difference equation for (r_i), remembering to include appropriate boundary conditions.

(ii) Solve the linear difference equation, to find r_a, Alice's probability of ruin. You may assume that $p \neq q$.

(b) Let d_i be be the expected duration of the game from the point that Alice has £i.

(i) Write down a linear difference equation for (d_i), remembering to include appropriate boundary conditions.

(ii) Solve the linear difference equation, to find d_a, the expected duration of the game. You may assume that $p \neq q$.

(c) Alice starts playing against Bob in a standard gambler's ruin game with probabilities $p \neq q$ and $s = 0$. A draw probability $s > 0$ is then introduced in such a way that the ratio $\rho = q/p$ remains constant. Comment on how this changes Alice's ruin probability and the expected duration of the game.

4. The Fibonacci numbers are 1, 1, 2, 3, 5, 8, 13, 21, 34, ..., where each number in the sequence is the sum of the two previous numbers. Show that the ratio of consecutive Fibonacci numbers tends to the "golden ratio" $\phi = (1 + \sqrt{5})/2$.

Assessment 1

A solutions sheet for this assessment is available on Minerva. Marks and feedback are available on Minerva and Gradescope respectively.

1. Let (X_n) be a simple random walk that starts from $X_0 = 0$ and on each step goes up one with probability p and down one with probability $q = 1 - p$.

Calculate:

(a) $\mathbb{P}(X_6 = 0)$, [1 mark]

(b) $\mathbb{E}X_6$, [1]

(c) $\text{Var}(X_6)$, [1]

(d) $\mathbb{E}(X_{10} \mid X_4 = 4)$, [1]

(e) $\mathbb{P}(X_{10} = 0 \mid X_6 = 2)$, [1]

(f) $\mathbb{P}(X_4 = 2 \mid X_{10} = 6)$. [1]

Consider the case $p = 0.6$, so $q = 0.4$.

(g) What are $\mathbb{E}X_{100}$ and $\text{Var}(X_{100})$? [1]

(h) Using a normal approximation, estimate $\mathbb{P}(16 \leq X_{100} \leq 26)$. You should use an appropriate "continuity correction", and explain why you chose it. (Bear in mind the possible values X_{100} can take.) [3]

2. Consider the gambler's ruin with draws: Alice starts with £a and Bob with £$(m - a)$, and at each time step Alice wins £1 off Bob with probability p, loses £1 to Bob with probability q, and no money is exchanged with probability s, where $p + q + s = 1$. We consider the case where Bob and Alice are equally matched, so $p = q$ and $s = 1 - 2p$. (We assume $0 < p < 1/2$.)

Let r_i be Alice's ruin probability from the point she has £i.

(a) By conditioning on the first step, explain why $pr_{i+1} - (1 - s)r_i + pr_{i-1} = 0$, and give appropriate boundary conditions. [2]

(b) Solve this linear difference equation to find an expression for r_i. [2]

Let d_i be the expected duration of the game from the point Alice has £i.

(c) Explain why $pd_{i+1} - (1 - s)d_i + pd_{i-1} = -1$, and give appropriate boundary conditions. [2]

(d) Solve this linear difference equation to find an expression for d_i. [2]

(e) Compare your answer to parts (b) and (d) with those for the standard gambler's ruin problem with $p = 1/2$, and give reasons for the similarities or differences. [2]

5 Discrete time Markov chains

- Definition of time homogeneous discrete time Markov chains
- Calculating n-step transition properties
- The Chapman–Kolomogorov equations

5.1 Time homogeneous discrete time Markov chains

So far we've seen a a few examples of stochastic processes in discrete time and discrete space with the Markov memoryless property. Now we will develop the theory more generally.

To define a so-called "Markov chain", we first need to say where we start from, and second what the probabilities of transitions from one state to another are.

In our examples of the simple random walk and gambler's ruin, we specified the start point $X_0 = i$ exactly, but we could pick the start point at random according to some distribution $\lambda_i = \mathbb{P}(X_0 = i)$.

After that, we want to know the **transition probabilities** $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ for $i, j \in \mathcal{S}$. Here, because of the Markov property, the transition probability only needs to condition on the state we're in now $X_n = i$, and not on the whole history of the process.

In the case of the simple random walk, for example, we had initial distribution

$$\lambda_i = \mathbb{P}(X_0 = i) = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{otherwise} \end{cases}$$

and transition probabilities

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} p & \text{if } j = i + 1 \\ q & \text{if } j = i - 1 \\ 0 & \text{otherwise.} \end{cases}$$

For the random walk (and also the gambler's ruin), the transition probabilities $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ don't depend on n; in other words, the transition probabilities stay the same over time. A Markov process with this property is called **time homogeneous**. We will always consider time homogeneous processes from now on (unless we say otherwise).

Let's write $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ for the transition probabilities, which are independent of n. We must have $p_{ij} \geq 0$, since it is a probability, and we must also have $\sum_j p_{ij} = 1$ for all states i, as this is the sum of the probabilities of all the places you can move to from state i.

Definition 5.1. Let (λ_i) be a probability distribution on a sample space $\mathcal{S}$. Let p_{ij}, where $i, j \in \mathcal{S}$, be such that $p_{ij} \geq 0$ for all i, j, and $\sum_j p_{ij} = 1$ for all i. Let the time index be $n = 0, 1, 2, \dots$. Then the **time homogeneous discrete time Markov process** or **Markov chain** (X_n) with initial distribution (λ_i) and transition probabilities (p_{ij}) is defined by

$$\mathbb{P}(X_0 = i) = \lambda_i,$$
$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{ij}.$$

When the state space is finite (and even sometimes when it's not), it's convenient to write the transition probabilities (p_{ij}) as a matrix P, called the **transition matrix**, whose (i, j)th entry is p_{ij}. Then the condition that $\sum_j p_{ij} = 1$ is the condition that each of the rows of P add up to 1.

Example 5.1. *In this notation, what is $\mathbb{P}(X_0 = i$ and $X_1 = j)$?*

First we must start from i, and then we must move from i to j, so

$$\mathbb{P}(X_0 = i \text{ and } X_1 = j) = \mathbb{P}(X_0 = i)\mathbb{P}(X_1 = j \mid X_0 = i) = \lambda_i p_{ij}.$$

In this notation, what is $\mathbb{P}(X_{n+2} = j$ and $X_{n+1} = k \mid X_n = i)$?

First we must move from i to k, then we must move from k to j, so

$$\mathbb{P}(X_{n+2} = j \text{ and } X_{n+1} = k \mid X_n = i) = \mathbb{P}(X_{n+1} = k \mid X_n = i)\mathbb{P}(X_{n+2} = j \mid X_{n+1} = k)$$
$$= p_{ik}p_{kj}.$$

Note that the term $\mathbb{P}(X_{n+2} = j \mid X_{n+1} = k)$ did not have to depend on X_n, thanks to the Markov property.

5.2 A two-state example

Consider a simple two-state Markov chain with state space $\mathcal{S} = \{0, 1\}$ and transition matrix

$$\mathsf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

for some $0 < \alpha, \beta < 1$. Note that the rows of P add up to 1, as they must.

We can illustrate P by a **transition diagram**, where the blobs are the states and the arrows give the transition probabilities. (We don't draw the arrow if $p_{ij} = 0$.) In this case, our transition diagram looks like this:
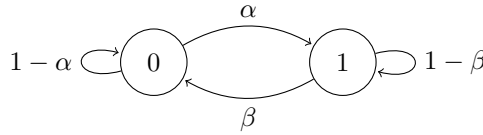
Figure 2: Transition diagram for the two-state Markov chain

We can use this as a simple model of a broken printer, for example. If the printer is broken (state 0) on one day, then with probability α it will be fixed (state 1) by the next day; while if it is working (state 1), then with probability β it will have broken down (state 0) by the next day.

Example 5.2. *If the printer is working on Monday, what's the probability that it also is working on Wednesday?*

If we call Monday day n, then Wednesday is day $n + 2$, and we want to find the two-step transition probability.

$$p_{11}(2) = \mathbb{P}(X_{n+2} = 1 \mid X_n = 1).$$

The key to calculating this is to *condition on the first step* again – that is, on whether the printer is working on Tuesday. We have

$$\begin{aligned} p_{11}(2) &= \mathbb{P}(X_{n+1} = 0 \mid X_n = 1)\,\mathbb{P}(X_{n+2} = 1 \mid X_{n+1} = 0, X_n = 1) \\ &\quad + \mathbb{P}(X_{n+1} = 1 \mid X_n = 1)\,\mathbb{P}(X_{n+2} = 1 \mid X_{n+1} = 1, X_n = 1) \\ &= \mathbb{P}(X_{n+1} = 0 \mid X_n = 1)\,\mathbb{P}(X_{n+2} = 1 \mid X_{n+1} = 0) \\ &\quad + \mathbb{P}(X_{n+1} = 1 \mid X_n = 1)\,\mathbb{P}(X_{n+2} = 1 \mid X_{n+1} = 1) \\ &= p_{10}p_{01} + p_{11}p_{11} \\ &= \beta\alpha + (1 - \beta)^2. \end{aligned}$$

In the second equality, we used the Markov property to mean conditional probabilities like $\mathbb{P}(X_{n+2} = 1 \mid X_{n+1} = k)$ did not have to depend on X_n.

Another way to think of this as we summing the probabilities of all length-2 paths from 1 to 1, which are $1 \to 0 \to 1$ with probability $\beta\alpha$ and $1 \to 1 \to 1$ with probability $(1 - \beta)^2$

5.3 n-step transition probabilities

In the above example, we calculated a two-step transition probability $p_{ij}(2) = \mathbb{P}(X_{n+2} = j \mid X_n = i)$ by conditioning on the first step. That is, by considering all the possible intermediate steps k, we have

$$p_{ij}(2) = \sum_{k \in \mathcal{S}} \mathbb{P}(X_{n+1} = k \mid X_n = i)\mathbb{P}(X_{n+2} = j \mid X_{n+1} = k) = \sum_{k \in \mathcal{S}} p_{ik}p_{kj}.$$

But this is exactly the formula for multiplying the matrix P with itself! In other words, $p_{ij}(2) = \sum_k p_{ik}p_{kj}$ is the (i,j)th entry of the matrix square $\mathsf{P}^2 = \mathsf{PP}$. If we write $\mathsf{P}(2) = (p_{ij}(2))$ for the matrix of two-step transition probabilities, we have $\mathsf{P}(2) = \mathsf{P}^2$.

More generally, we see that this rule holds over multiple steps, provided we sum over all the possible paths $i \to k_1 \to k_2 \to \cdots \to k_{n-1} \to j$ of length n from i to j.

Theorem 5.1. *Let (X_n) be a Markov chain with state space $\mathcal{S}$ and transition matrix $\mathsf{P} = (p_{ij})$. For $i, j \in \mathcal{S}$, write*

$$p_{ij}(n) = \mathbb{P}(X_n = j \mid X_0 = i)$$

for the n-step transition probability. Then

$$p_{ij}(n) = \sum_{k_1, k_2, \ldots, k_{n-1} \in \mathcal{S}} p_{ik_1} p_{k_1 k_2} \cdots p_{k_{n-2}k_{n-1}} p_{k_{n-1}j}.$$

In particular, $p_{ij}(n)$ is the (i,j)th element of the matrix power P^n, and the matrix of n-step transition probabilities is given by $\mathsf{P}(n) = \mathsf{P}^n$.

The so-called **Chapman–Kolmogorov equations** follow immediately from this.

Theorem 5.2 (Chapman–Kolmogorov equations)**.** *Let (X_n) be a Markov chain with state space $\mathcal{S}$ and transition matrix $\mathsf{P} = (p_{ij})$. Then, for non-negative integers n, m, we have*

$$p_{ij}(n + m) = \sum_{k \in \mathcal{S}} p_{ik}(n)p_{kj}(m),$$

or, in matrix notation, $\mathsf{P}(n + m) = \mathsf{P}(n)\mathsf{P}(m)$.

In other words, a trip of length $n + m$ from i to j is a trip of length n from i to some other state k, then a trip of length m from k back to j, and this intermediate stop k can be any state, so we have to sum the probabilities.

Of course, once we know that $\mathsf{P}(n) = \mathsf{P}^n$ is given by the matrix power, it's clear to see that $\mathsf{P}(n + m) = \mathsf{P}^{n+m} = \mathsf{P}^n\mathsf{P}^m = \mathsf{P}(n)\mathsf{P}(m)$.

Sidney Chapman (1888–1970) was a British applied mathematician and physicist, who studied applications of Markov processes. Andrey Nikolaevich Kolmogorov (1903–1987) was a Russian mathematician who did very important work in many different areas of mathematics, is considered the "father of modern probability theory", and studied the theory of Markov processes. (Kolmogorov is also my academic great-great-grandfather.)

Example 5.3. In our two-state broken printer example above, the matrix of two-state transition probabilities is given by

$$\begin{aligned} \mathsf{P}(2) = \mathsf{P}^2 &= \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \\ &= \begin{pmatrix} (1 - \alpha)^2 + \alpha\beta & (1 - \alpha)\alpha + \alpha(1 - \beta) \\ \beta(1 - \alpha) + (1 - \beta)\beta & \beta\alpha + (1 - \beta)^2 \end{pmatrix}, \end{aligned}$$

where the bottom right entry $p_{11}(2)$ is what we calculated earlier.

One final comment. It's also convenient to consider the initial distribution $\lambda = (\lambda_i)$ as a *row* vector. The first-step distribution is given by

$$\mathbb{P}(X_1 = j) = \sum_{i \in \mathcal{S}} \lambda_i p_{ij},$$

by conditioning on the start point. This is exactly the jth element of the vector–matrix multiplication $\lambda\mathsf{P}$. More generally, the row vector of of probabilities after n steps is given by $\lambda\mathsf{P}^n$.

In the next section, we look at how to model some actuarial problems using Markov chains.

6 Examples from actuarial science

- Three Markov chain models for insurance problems

In this lecture we'll set up three simple models for an insurance company that can be analysed using ideas about Markov chains. The first example has a direct Markov chain model. For the second and third examples, we will have to be clever to find a Markov chain associated to the situation.

6.1 A simple no-claims discount model

A motor insurance company puts policy holders into three categories:

- no discount on premiums (state 1)
- 25% discount on premiums (state 2)
- 50% discount on premiums (state 3)

New policy holders start with no discount (state 1). Following a year with no insurance claims, policy holders move up one level of discount. If they start the year in state 3 and make no claim, they remain in state 3. Following a year with at least one claim, they move down one level of discount. If they start the year in state 1 and make at least one claim, they remain in state 1. The insurance company believes that probability that a motorist has a claim free year is $\frac{3}{4}$.

We can model this directly as a Markov chain:

- the state space $\mathcal{S} = \{1, 2, 3\}$ is discrete;
- the time index is discrete, as we have one discount level each year;
- the probability of being in a certain state at a future time is completely determined by the present state (the Markov property);
- the one-step transition probabilities are not time dependent (time homogeneous).

The transition probability and transition diagram of the Markov chain are:

$$\mathsf{P} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$
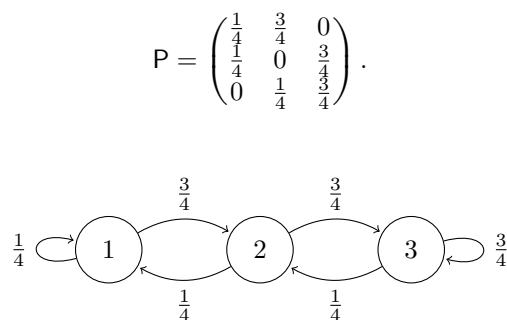
Figure 3: Transition diagram for the simple no-claims discount model

Example 6.1. *What is the probability of having a 50% reduction to your premium three years from now, given that you currently have no reduction on the premium?*

We want to find the three-step transition probability

$$p_{13}(3) = \mathbb{P}(X_3 = 3 \mid X_0 = 1).$$

We can find this by summing over all paths $1 \to k_1 \to k_2 \to 3$. There are two such paths, $1 \to 1 \to 2 \to 3$ and $1 \to 2 \to 3 \to 3$. Thus

$$p_{13}(3) = p_{11}p_{12}p_{23} + p_{12}p_{23}p_{33} = \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{36}{64} = \frac{9}{16}.$$

Alternatively, we could directly calculate all the three-step transition probabilities by the matrix method, to get

$$\mathsf{P}(3) = \mathsf{P}^3 = \mathsf{PPP} = \frac{1}{64} \begin{pmatrix} 7 & 21 & 36 \\ 7 & 12 & 45 \\ 4 & 15 & 45 \end{pmatrix}.$$

(You can check this yourself, if you want.) The desired $p_{13}(3)$ is the top right entry $36/64 = 9/16$.

6.2   An accident model with memory

Sometimes, we are presented with a situation where the "obvious" stochastic process is not a Markov chain. But sometimes we can find a related process that *is* a Markov chain, and study that instead. As an example of this, we consider at a different accident model.

According to a different model, a motorist's nth year of driving is either accident free, or has exactly one accident. (The model does not allow for more than one accident in a year.) Let Y_n be a random variable so that,

$$Y_n = \begin{cases} 0 & \text{if the motorist has no accident in year } n, \\ 1 & \text{if the motorist has one accident in year } n. \end{cases}$$

This defines a stochastic process (Y_n) with finite state space $\mathcal{S} = \{0, 1\}$ and discrete time $n = 1, 2, 3, \ldots$.

The probability of an accident in year $n + 1$ is modelled as a function of the total number of previous accidents over a function of the number of years in the policy; that is,

$$\mathbb{P}(Y_{n+1} = 1 \mid Y_n = y_n, \ldots, Y_2 = y_2, Y_1 = y_1) = \frac{f(y_1 + y_2 + \cdots + y_n)}{g(n)},$$

and $Y_{n+1} = 0$ otherwise, where f and g are non-negative increasing functions with $0 \leq f(m) \leq g(m)$ for all m. (We'll come back to these conditions in a moment.)

Unfortunately (Y_n) is *not* a Markov chain – it's clear that Y_{n+1} depends not only on Y_n, the number accidents this year, but the entire history $Y_1, Y_2, \ldots, Y_n$.

However, we have a cunning work-around. Define $X_n = \sum_{i=1}^{n} Y_i$ to be the total number of accidents up to year n. Then (X_n) *is* a Markov chain. In fact, we have

$$\begin{aligned} \mathbb{P}(X_{n+1} &= x_n + 1 \mid X_n = x_n, \ldots, X_2 = x_2, X_1 = x_1) \\ &= \mathbb{P}(Y_{n+1} = 1 \mid Y_n = x_n - x_{n-1}, \ldots Y_2 = x_2 - x_1, Y_1 = x_1) \\ &= \frac{f((x_n - x_{n-1}) + \cdots + (x_2 - x_1) + x_1)}{g(n)} \\ &= \frac{f(x_n)}{g(n)}, \end{aligned}$$

and $X_{n+1} = x_n$ otherwise. This clearly depends only on x_n. Thus we can use Markov chain techniques on (X_n) to lean about the non-Markov process (Y_n).

Note that the probability that $X_{n+1} = x_n$ or $x_n + 1$ depends not only on x_n but also on the time n. So this is a rare example of a time *inhomogeneous* Markov process, where the transition probabilities do depend on the time n.

Before we move on, let's think about the conditions we placed on this model. First, the condition that f is increasing means that between drivers who have been driving the same number of years, we think the more accident-prone in the past is more likely to have an accident in the future. Second, the condition that g is increasing means that between drivers who have had the same number of accidents, we think the one who has spread those accidents over a longer period of time is less likely to have accidents in the future. Third, the transition probabilities should lie in the range $[0, 1]$; but since $\sum_{i=1}^{m} y_i \leq m$, our condition $0 \leq f(m) \leq g(m)$ guaranteed that this is the case.

6.3   A no-claims discount model with memory

Sometimes, we are presented with a stochastic process which is not a Markov chain, but where by altering the state space $\mathcal{S}$ we *can* end up with a process which *is* a Markov chain. As such, when making a model, it is important to think carefully about choice of state space. To see this we will return to the no-claims discount example.

Suppose now we have an model with four levels of discount:

- no discount (state 1)
- 20% discount (state 2)
- 40% discount (state 3)
- 60% discount (state 4)

If a year is accident free, then the discount increases one level, to a maximum of 60%. This time, if the year has an accident, then the discount decreases by one level if the year previous to that was accident free, but decreases by *two* levels if the previous year had an accident as well, both to a minimum of no discount.

As before, the insurance company believes that probability that a motorist has a claim-free year is $\frac{3}{4} = 0.75$.

We might consider the most natural choice of a state space, where the states are discount levels; say, $\mathcal{S} = \{1, 2, 3, 4\}$. But this is not a Markov chain, since if a policy holder has an accident, we may need to know about the past in order to determine probabilities for future states, which violates the Markov property. In particular, if a motorist is in state 3 (40% discount) and has an accident, they will either move down to level 2 (if they had not crashed the previous year, so had previously been in state 2) or to level 1 (if they had crashed the previous year, so had previously been in state 4) – but that depends on their previous level too, which the Markov property doesn't allow. (You can check this is the only violation of the Markov property.)

However, we can be clever again, this time in the choice of our state space. Instead, we can split the 40% level into two different states: state "3a" if there was no accident the previous year, and state "3b" if there was an accident the previous year. Our states are now:

- no discount (state 1)
- 20% discount (state 2)
- 40% discount, no claim in previous year (state 3a)
- 40% discount, claim in previous year (state 3b)
- 60% discount (state 4)

Now this *is* a Markov chain, because the new states 3s carry with them the memory of the previous year, to ensure the Markov property is preserved. Under the assumption of 25% of drivers having an accident each year, the transition matrix is

$$\mathsf{P} = \begin{pmatrix} 0.25 & 0.75 & 0 & 0 & 0 \\ 0.25 & 0 & 0.75 & 0 & 0 \\ 0 & 0.25 & 0 & 0 & 0.75 \\ 0.25 & 0 & 0 & 0 & 0.75 \\ 0 & 0 & 0 & 0.25 & 0.75 \end{pmatrix}.$$

The transition diagram is shown below. (Recall that we don't draw arrows with probability 0.)

Note that when we move up from state 2, we go to 3a (no accident in the previous year); but when we move down from state 4, we go to 3b (accident in the previous year).

In the next section, we look at how to study big Markov chains by splitting them into smaller pieces called "classes".
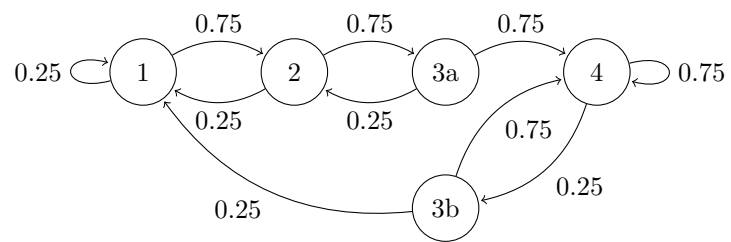
Figure 4: Transition diagram for the no-claims discount model with memory

Problem sheet 3

You should attempt all these questions and write up your solutions in advance of your workshop in week 4 (Monday 15 or Tuesday 16 February) where the answers will be discussed.

1. Consider a Markov chain with state space $\mathcal{S} = \{1, 2, 3\}$, and transition matrix partially given by

$$\mathsf{P} = \begin{pmatrix} ? & 0.3 & 0.3 \\ 0.2 & 0.4 & ? \\ ? & ? & 1 \end{pmatrix}.$$

(a) Replace the four question marks by the appropriate transition probabilities.

(b) Draw a transition diagram for this Markov chain.

(c) Find the matrix $\mathsf{P}(2)$ of two-step transition probabilities.

(d) By summing the probabilities of all relevant paths, find the three-step transition probability $p_{13}(3)$.

2. Consider a Markov chain (X_n) which moves between the vertices of a tetrahedron.

At each time step, the process randomly chooses one of the edges connected to the current vertex and follows it to a new vertex. The edge to follow is selected randomly with all options having equal probability and each selection is independent of the past movements. Let X_n be the vertex the process is in after step n.

(a) Write down the transition matrix P of this Markov chain.

(b) By summing over all relevant paths of length two, calculate the two-step transition probabilities $p_{11}(2)$ and $p_{12}(2)$. Hence, write down the two-step transition matrix $\mathsf{P}(2)$.

(c) Check your answer by calculating the matrix square P^2.
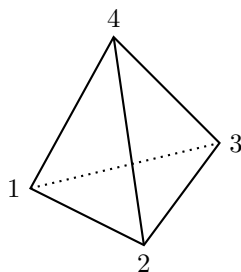
Figure 5: A tetrahedron

3. Consider the two-state "broken printer" Markov chain, with state space $\mathcal{S} = \{0, 1\}$, transition matrix

$$\mathsf{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

with $0 < \alpha, \beta < 1$, and initial distribution $\lambda = (\lambda_0, \lambda_1)$. Write $\mu_n = \mathbb{P}(X_n = 0)$.

(a) By writing μ_{n+1} in terms of μ_n, show that we have

$$\mu_{n+1} - (1 - (\alpha + \beta))\mu_n = \beta.$$

(b) By solving this linear difference equation using the initial condition $\mu_0 = \lambda_0$, or otherwise, show that

$$\mu_n = \frac{\beta}{\alpha + \beta} + \left(\lambda_0 - \frac{\beta}{\alpha + \beta}\right)(1 - (\alpha + \beta))^n.$$

(c) What, therefore, are $\lim_{n \to \infty} \mathbb{P}(X_n = 0)$ and $\lim_{n \to \infty} \mathbb{P}(X_n = 1)$?

(d) Explain what happens if the Markov chain is started in the distribution

$$\lambda_0 = \frac{\beta}{\alpha + \beta}, \qquad \lambda_1 = \frac{\alpha}{\alpha + \beta}.$$

4. Let (X_n) be a Markov chain. Show that, for any $m \geq 1$, we have

$$\mathbb{P}(X_{n+m} = x_{n+m} \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_{n+m} = x_{n+m} \mid X_n = x_n).$$

5. A car insurance company operates a no-claims discount system for existing policy holders. The possible discounts on premiums are $\{0\%, 25\%, 40\%, 50\%\}$. Following a claim-free year, a policyholder's discount level increases by one level (or remains at 50% discount). If the policyholder makes one or more claims in a year, the discount level decreases by one level (or remains at 0% discount).

The insurer believes that the probability of of making at least one claim in a year is 0.1 if the previous year was claim-free and 0.25 if the previous year was not claim-free.

(a) Explain why we cannot use $\{0\%, 25\%, 40\%, 50\%\}$ as the state space of a Markov chain to model discount levels for policyholders.

(b) By considering additional states, show that a Markov chain can be used to model the discount level.

(c) Draw the transition diagram and write down the transition matrix.

6. The credit rating of a company can be modelled as a Markov chain. Assume the rating is assessed once per year at the end of the year and possible ratings are A (good), B (fair) and D (in default). The transition matrix is

$$P = \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}.$$

(a) Calculate the two-step transition probabilities, and hence find the expected number of defaults in the next two years from 100 companies all rated A at the start of the period.

(b) What is the probability that a company rated A will at some point default without ever having been rated B in the meantime?

A corporate bond portfolio manager follows an investment strategy which means that bonds which fall from A-rated to B-rated are sold and replaced with an A-rated bond. The manager believes this will improve the returns on the portfolio because it will reduce the number of defaults experienced.

(c) Calculate the expected number of defaults in the manager's portfolio over the next two years given there are initially 100 A-rated bonds.

7 Class structure

- Communicating classes and irreducibility
- Period of a state (and class)

7.1 Communicating classes

If we have a large complicated Markov chain, it can be useful to split the state space up into smaller pieces that can be studied separately. The idea is that states i and j should definitely be in the same piece (or "class) if we can get from i to j and then back to i again after some number of steps.

Definition 7.1. Consider a Markov chain on a state space $\mathcal{S}$ with transition matrix P. We say that state $j \in \mathcal{S}$ is **accessible** from state $i \in \mathcal{S}$ and write $i \to j$ if, for some n, $p_{ij}(n) > 0$.

If $i \to j$ and $j \to i$, we say that i **communicates with** j and write $i \leftrightarrow j$.

Here, the condition $p_{ij}(n) > 0$ means that, starting from i, there's a positive chance that we'll get to j at some point in the future – hence the term "accessible".

Theorem 7.1. *Consider a Markov chain on a state space $\mathcal{S}$ with transition matrix P. Then the "communicates with" relation $\leftrightarrow$ is an **equivalence relation**; that is, it has the following properties:*

- ***reflexive**: $i \leftrightarrow i$ for all i;*
- ***symmetric**: if $i \leftrightarrow j$ then $j \leftrightarrow i$;*
- ***transitive**: if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.*

Proof. Reflexivity: Clearly $p_{ii}(0) = 1 > 0$, because in "zero steps" we stay where we are. So $i \leftrightarrow i$ for all i.

Symmetry: The definition of $i \leftrightarrow j$ is symmetric under swapping i and j.

Transitivity. If we can get from i to j and we can get from j to k, then we can get from i to k by going via j. We just need to write that out formally.

Since $i \to j$, we have $p_{ij}(n) > 0$ for some n, and since $j \to k$, we also have $p_{jk}(m) > 0$ for some m. Then, by the Chapman–Kolmogorov equations, we have

$$p_{ik}(n + m) = \sum_{l \in \mathcal{S}} p_{il}(n) p_{lk}(m) \geq p_{ij}(n) p_{jk}(m) > 0,$$

from just picking out the $l = j$ term in the sum. So $i \to k$ too.

The same argument with k and i swapped gives $k \to i$ also, so $i \leftrightarrow k$.

$\square$

A fact you may remember about equivalence relations is that an equivalence relation, like $\leftrightarrow$, partitions the space $\mathcal{S}$ into **equivalence classes**. This means that each state i is in exactly one equivalence class, and that class is the set of states j such that $i \leftrightarrow j$. In this context, we call these **communicating classes**.

Example 7.1. In the simple random walk, provided p is not 0 or 1, every state communicates with every other state, because from i when can get to $j > i$ by going up $j - i$ times, and we can get to $j < i$ by going down $i - j$ times. Therefore the whole state space $\mathcal{S} = \mathbb{Z}$ is one communicating class.

Example 7.2. Consider the gambler's ruin Markov chain on $\{0, 1, \dots, m\}$. There are three communicating classes. The ruin states $\{0\}$ and $\{m\}$ each don't communicate with any other states, so each are a class by themselves. The remaining states $\{1, 2, \dots, m - 1\}$ are all in the same class, like the simple random walk.

Example 7.3. Consider the following simple model for an epidemic. We have three states: healthy (H), sick (S), and dead (D). This transition matrix is

$$\mathsf{P} = \begin{pmatrix} p_{\text{HH}} & p_{\text{HS}} & 0 \\ p_{\text{SH}} & p_{\text{SS}} & p_{\text{SD}} \\ 0 & 0 & 1 \end{pmatrix},$$

and the transition diagram is:
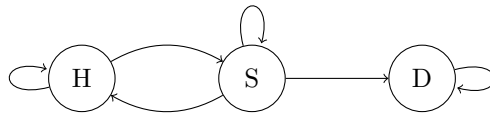
Figure 6: Transition diagram for the healthy–sick–dead chain.

Clearly H and S communicate with each other (you can become infected or recover), while D only communicates with itself (the dead do not recover). Hence, the state space $\mathcal{S} = \{\text{H}, \text{S}, \text{D}\}$ partitions into two communicating classes: $\{\text{H}, \text{S}\}$ and $\{\text{D}\}$.

A few more definitions that will be important later.

Definition 7.2. If the entire state space $\mathcal{S}$ is one communicating class, we say that the Markov chain is **irreducible**.

We say that a communicating class is **closed** if no state outside the class is accessible from any state within the class. That is, class $C \subset \mathcal{S}$ is closed if whenever there exist $i \in C$ and $j \in \mathcal{S}$ with $i \to j$, then $j \in C$ also. If a class is not closed, we say it is **open**.

If a state i is in a communicating class $\{i\}$ by itself and that class is closed, then we say state i is **absorbing**.

In non-maths language:

- An irreducible Markov chain can't be broken down into smaller pieces.
- Once you enter a closed class, you can't leave that class.
- Once you reach an absorbing state, you can't leave that state.

How do these work for our earlier examples?

Example 7.4. Going back to the previous examples:

- In the simple random walk, the whole state space is one communicating class which must therefore be closed. The Markov chain has only one class, so is irreducible.
- In the gambler's ruin, classes $\{0\}$ and $\{m\}$ are closed, because the Markov chain stays there forever, and because these closed classes consist of only one state each, 0 and m are absorbing states. The class $\{1, 2, \ldots, m-1\}$ is open, as we can escape the class by going to 0 or m. The gambler's ruin chain has multiple classes, so is not irreducible.
- In the "healthy–sick–dead" chain, the class $\{D\}$ is closed, so D is an absorbing state, while the class $\{H, S\}$ is open, as one can leave it by dying. The Markov chain is not irreducible.

7.2 Periodicity

When we discussed the simple random walk, we noted that it alternates between even-numbered and odd-numbered states. This "periodic" behaviour is important to understand if we want to know which state we will be in at some point in the future.

The idea is this: List the number of steps for all possible paths starting and ending in the state. Then the period is the greatest common divisor (or "highest common factor") of the integers in this list.

Definition 7.3. Consider a Markov chain with transition matrix P. We say that a state $i \in \mathcal{S}$ has **period** d_i, where

$$d_i = \gcd\{n \in \{1, 2, \dots, \} : p_{ii}(n) > 0\},$$

where gcd denotes the greatest common divisor.

If $d_i > 1$, then the state i is called **periodic**; if $d_i = 1$, then i is called **aperiodic**.

Example 7.5. Consider the simple random walk with $p \neq 0, 1$. We have $p_{ii}(n) = 0$ for odd n, since we swap from odd to even each step. But $p_{ii}(2) = 2pq > 0$. Therefore, all states are periodic with period $\gcd\{2, 4, 6, \dots\} = 2$.

Example 7.6. For the gambler's ruin, states 0 and m are aperiodic (have period 1), since they are absorbing states. The remaining states states $1, 2, \dots, m-1$ are periodic with period 2, because we swap between odd and even states, as in the simple random walk.

Example 7.7. Consider the Markov chain with transition diagram as shown:
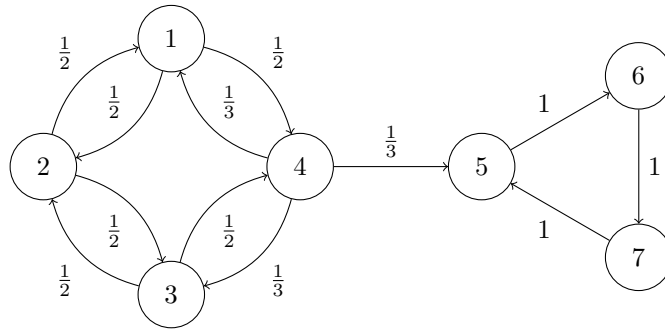
Figure 7: Transition diagram for an aperiodic irreducible Markov chain.

Importantly, we can't return from the triangle side back to the circle side. We thus see there are two communicating classes: $\{1, 2, 3, 4\}$, which is open, and $\{5, 6, 7\}$, which is closed. The Markov chain is not irreducible, and there are no absorbing states.

The circle side swaps between odd and even states (until exiting from 4 to 5), so states 1,2, 3 and 4 all have period 2. The triangle side cycles around with certainty, meaning that states 5, 6, and 7 all have period 3.

You may have noticed in these examples that, within a communicating class, every state has the same period. In fact, it's always the case that states in the same class have the same period.

Theorem 7.2. *All states in a communicating class have the same period.*

Formally: Consider a Markov chain on a state space $\mathcal{S}$ with transition matrix P. If $i, j \in \mathcal{S}$ are such that $i \leftrightarrow j$, then $d_i = d_j$.

In particular, in an irreducible Markov chain, all states have the same period d. We say that an irreducible Markov chain is **periodic** if $d > 1$ and **aperiodic** if $d = 1$.

Proof. Let i, j be such that $i \leftrightarrow j$. We want to show that $d_i = d_j$. First we'll show that $d_i \leq d_j$, and then we'll show that $d_j \leq d_i$, and thus conclude that they're equal.

Since $i \leftrightarrow j$, there exist n, m such that $p_{ij}(n) > 0$ and $p_{ji}(m) > 0$. Then, by the Chapman–Kolmogorov equations,

$$p_{ii}(n + m) = \sum_{k \in \mathcal{S}} p_{ik}(n) p_{ki}(m) \geq p_{ij}(n) p_{ji}(m) > 0.$$

So d_i divides $n + m$.

Let r be such that $p_{jj}(r) > 0$. Then, by the same Chapman–Kolmogorov argument,

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

because we can get from i to i by going $i \to j \to j \to i$. Hence d_i divides $n + m + r$.

But if d_i divides both $n + m$ and $n + m + r$, it must be that d_i divides r also. So whenever $p_{jj}(r) > 0$, we have that d_i divides r. Since d_i is a common divisor of all the rs with $p_{jj}(r) > 0$, it can't be any bigger that the *greatest* common divisor of all those rs. But that greatest common divisor is by definition d_j, the period of j. So $d_i \leq d_j$.

Repeating the same argument but with i and j swapped over, we get $d_j \leq d_i$ too, and we're done.

$\square$

In the next section, we look at two problems to do with "hitting times": What is the probability we reach a certain state, and how long on average does it take us to get there?

8 Hitting times

- Definitions: Hitting probability, expected hitting time, return probability, expected return time
- Finding these by conditioning on the first step
- Return probability and expected return time for the simple random walk

8.1 Hitting probabilities and expected hitting times

In Section 3 and Section 4, we used conditioning on the first step to find the ruin probability and expected duration for the gambler's ruin problem. Here, we develop those ideas for general Markov chains.

Definition 8.1. Let (X_n) be a Markov chain on state space $\mathcal{S}$. Let H_A be a random variable representing the **hitting time** to hit the set $A \subset \mathcal{S}$, given by

$$H_A = \min\left\{n \in \{0, 1, 2, ... \} : X_n \in A\right\}.$$

The most common case we'll be interested in will be when $A = \{j\}$ is just a single state, so

$$H_j = \min\left\{n \in \{0, 1, 2, ... \} : X_n = j\right\}.$$

(We use the convention that $H_A = \infty$ if $X_n \notin A$ for all n.)

The **hitting probability** h_{iA} of the set A starting from state i is

$$h_{iA} = \mathbb{P}(X_n \in A \text{ for some } n \geq 0 \mid X_0 = i) = \mathbb{P}(H_A < \infty \mid X_0 = i).$$

Again, the most common case of interest is h_{ij} the hitting probability of state j starting from state i.

The **expected hitting time** η_{iA} of the set A starting from state i is

$$\eta_{iA} = \mathbb{E}(H_A \mid X_0 = i).$$

Clearly η_{iA} can only be finite if $h_{iA} = 1$.

The short summary of this is:

- The **hitting probability** h_{ij} is the probability we hit state j starting from state i.
- The **expected hitting time** η_{ij} is the expected time until we hit state j starting from state i.

The good news us that we already know how to find hitting probabilities and expected hitting times, because we already did it for the gambler's ruin problem! The way we did it then is that we first found equations for hitting probabilities or expected hitting times by conditioning on the first step, and then we solved those equations. We do the same here for other Markov chains.

Let's see an example of how to find a hitting probability.

Example 8.1. *Consider a Markov chain with transition matrix*

$$\begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \\ 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Calculate the probability that the chain is absorbed at state 2 when started from state 1.

This is asking for the hitting probability h_{12}. We use – as ever – a conditioning on the first step argument. Specifically, we have

$$h_{12} = p_{11}h_{12} + p_{12}h_{22} + p_{13}h_{32} + p_{14}h_{42} \tag{5}$$
$$= \tfrac{1}{5}h_{12} + \tfrac{1}{5}h_{22} + \tfrac{1}{5}h_{32} + \tfrac{2}{5}h_{42} \tag{6}$$

This is because, starting from 1, there's a $p_{11} = \frac{1}{5}$ probability of staying at 1; then by the Markov property it's like we're starting again from 1, so the hitting probability is still h_{12}. Similarly, there's a $p_{12} = \frac{1}{5}$ probability of moving 2; then by the Markov property it's like we're starting again from 2, so the hitting probability is still h_{22}. And so on.

Since (6) includes not just h_{12} but also h_{22}, h_{32} and h_{42}, we need to find equations for those too. Clearly we have $h_{22} = 1$, as we are "already there". Also, $h_{42} = 0$, since 4 is an absorbing state. For h_{32}, another "condition on the first step" argument gives

$$h_{32} = p_{32}h_{22} + p_{34}h_{42} = \tfrac{1}{2}h_{22} + \tfrac{1}{2}h_{42} = \tfrac{1}{2},$$

where we substituted in $h_{22} = 1$ and $h_{42} = 0$.

Substituting $h_{22} = 1$, $h_{32} = \frac{1}{2}$ and $h_{42} = 0$ all into (6), we get

$$h_{12} = \tfrac{1}{5}h_{12} + \tfrac{1}{5} + \tfrac{1}{5}\tfrac{1}{2} = \tfrac{1}{5}h_{12} + \tfrac{3}{10}.$$

Hence $\frac{4}{5}h_{12} = \frac{3}{10}$, so the answer we were after is $h_{12} = \frac{3}{8}$.

So the technique to find hitting probability h_{ij} from i to j is:

1. Set up equations for all the hitting probabilities h_{kj} to j by conditioning on the first step.
2. Solve the resulting simultaneous equations.

It is recommended to derive equations for hitting probabilities from first principles by conditioning on the first step, as we did in the example above. However, we can state what the general formula is: by the same conditioning method, we get

$$h_{iA} = \begin{cases} \displaystyle\sum_{j \in \mathcal{S}} p_{ij}h_{jA} & \text{if } i \notin A \\ 1 & \text{if } i \in A. \end{cases}$$

It can be shown that, if these equations have multiple solutions, then the hitting probabilities are in fact the smallest non-negative solutions.

Now an example for expected hitting times.

Example 8.2. *Consider the simple no-claims discount chain from Lecture 6, which had transition matrix*

$$\mathsf{P} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

Given we start in state 1 (no discount), find the expected amount of time until we reach state 3 (50% discount).

This question asks us to find η_{13}. We follow the general method, and start by writing down equations for all the η_{i3}s.

Clearly we have $\eta_{33} = 0$. For the others, we condition on the first step to get

$$\eta_{13} = 1 + \tfrac{1}{4}\eta_{13} + \tfrac{3}{4}\eta_{23} \qquad \Rightarrow \qquad \tfrac{3}{4}\eta_{13} - \tfrac{3}{4}\eta_{23} = 1,$$
$$\eta_{23} = 1 + \tfrac{1}{4}\eta_{13} + \tfrac{3}{4}\eta_{33} \qquad \Rightarrow \qquad -\tfrac{1}{4}\eta_{13} + \eta_{23} = 1.$$

This is because the first step takes time 1, then we condition on what happens next – just like we did for the gambler's ruin.

The first equation plus three-quarters times the second gives

$$(\tfrac{3}{4} - \tfrac{3}{4}\tfrac{1}{4})\eta_{13} = \tfrac{9}{16}\eta_{13} = 1 + \tfrac{3}{4} = \tfrac{7}{4} = \tfrac{28}{16},$$

which has solution $\eta_{13} = \frac{28}{9} = 3.11$.

Similarly, if we need to, we can give a general formula

$$\eta_{iA} = \begin{cases} 1 + \displaystyle\sum_{j \in \mathcal{S}} p_{ij}\eta_{jA} & \text{if } i \notin A \\ 0 & \text{if } i \in A. \end{cases}$$

Again, if we have multiple solutions, the expected hitting times are the smallest non-negative solutions.

8.2 Return times

Under the definitions above, the hitting probability and expected hitting time to a state from itself are always $h_{ii} = 1$ and $\eta_{ii} = 0$, as we're "already there".

In this case, it can be interesting to look instead at the random variable representing the **return time**,

$$M_i = \min \left\{ n \in \{1, 2, \dots\} : X_n = i \right\}.$$

Note that this only considers times $n = 1, 2, \dots$ not including $n = 0$, so is the *next* time we come back, after $n = 0$.

We then have the **return probability** and **expected return time**

$$m_i = \mathbb{P}(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) = \mathbb{P}(M_i < \infty \mid X_0 = i),$$
$$\mu_i = \mathbb{E}(M_i \mid X_0 = i).$$

Just as before, we can find these by conditioning on the first step. The general equations are

$$m_i = \sum_{j \in \mathcal{S}} p_{ij} h_{ji}, \qquad \mu_i = 1 + \sum_{j \in \mathcal{S}} p_{ij} \eta_{ji},$$

where, if necessary, we take the minimal non-negative solution again.

8.3 Hitting and return times for the simple random walk

We now turn to hitting and return times for the simple random walk, which goes up with probability p and down with probability $q = 1 - p$. This material is mandatory and is examinable, but is a bit technical; students who are struggling or have fallen behind might make a tactical decision just to read the two summary theorems below and come back to the details at a later date.

Let's start with hitting probabilities. Without loss of generality, we look at h_{i0}, the probability the random walk hits 0 starting from i. We will assume that $i > 0$. For $i < 0$, we can get the desired result by looking at the random walk "in the mirror" – that is, by swapping the role of p and q, and treating the positive value $-i$.

For an initial condition, it's clear we have $h_{00} = 1$.

For general $i > 0$, we condition on the first step, to get

$$h_{i0} = p h_{i+1\,0} + q h_{i-1\,0}.$$

We recognise this equation from the gambler's ruin problem, and recall that we have to treat the cases $p \neq \frac{1}{2}$ and $p = \frac{1}{2}$ separately.

When $p \neq \frac{1}{2}$, the general solution is $h_{i0} = A + B\rho^i$, where $\rho = q/p \neq 1$, as before. The initial condition $h_{00} = 1$ gives $A = 1 - B$, so we have a family of solutions

$$h_{i0} = (1 - B) + B\rho^i = 1 + B(\rho^i - 1).$$

In the gambler's ruin problem we had another boundary condition to find B. Here we have no other conditions, but we can use the minimality condition that the hitting probabilities are the smallest non-negative solution to the equation. This minimal non-negative solution will depend on whether $\rho > 1$ or $\rho < 1$.

When $\rho > 1$, so $p < \frac{1}{2}$, the term ρ^i tends to infinity. Thus the minimal non-negative solution will have to take $B = 0$, because taking $B < 0$ would eventually give a negative solution, while $B = 0$ gives the smallest of the non-negative solutions. Hence the solution is $h_{i0} = 1 + 0(\rho^i - 1) = 1$, meaning we hit 0 with certainty. This makes sense, because for $p < \frac{1}{2}$ the random walk "drifts" to the left, and eventually gets back to 0.

When $\rho < 1$, so $p > \frac{1}{2}$, we have that $\rho^i - 1$ is negative and tends to -1. So to get a small solution we want B large and positive, but keeping the solution non-negative limits us to $B \leq 1$, so the minimality

condition is achieved at $B = 1$. The solution is $h_{i0} = 1 + 1(\rho^i - 1) = \rho^i$. This is strictly less than 1 as expected, because for $p > 1/2$, the random walk drifts to the right, so might drift further and further away from 0 and not come back.

For $p = \frac{1}{2}$, so $\rho = 1$, we recall the general solution $h_{i0} = A + Bi$. The condition $h_{00} = 1$ gives $A = 1$, so $h_{i0} = 1 + Bi$. Because i is positive, to get the answer to be small we want B to be small, but non-negativity limits us to B positive, so the minimal non-negative solution takes $B = 0$. This gives $h_{i0} = 1 + 0i = 1$, so we will always get back to 0.

In conclusion, we have the following. (Recall that we get the result for $i < 0$ by swapping the role of p and q, and treating the positive value $-i$.)

Theorem 8.1. *Consider a random walk with up probability $p \neq 0, 1$. Then the hitting probabilities to 0 are givin by, for $i > 0$,*

$$h_{i0} = \begin{cases} \left(\dfrac{q}{p}\right)^i < 1 & \text{if } p > \frac{1}{2} \\ 1 & \text{if } p \leq \frac{1}{2}, \end{cases}$$

and for $i < 0$ by

$$h_{i0} = \begin{cases} \left(\dfrac{p}{q}\right)^{-i} < 1 & \text{if } p < \frac{1}{2} \\ 1 & \text{if } p \geq \frac{1}{2}. \end{cases}$$

Now we look at return times. What is the return probability to 0 (or, by symmetry, to any state $i \in \mathbb{Z}$)? By conditioning on the first step, $m_0 = ph_{10} + qh_{-10}$, and we've calculated those hitting times above. We have

$$m_0 = \begin{cases} p \times 1 \ + q \times 1 \ = 1 & \text{if } p = \frac{1}{2} \\ p \times \dfrac{q}{p} + q \times 1 \ = 2q < 1 & \text{if } p > \frac{1}{2} \\ p \times 1 \ + q \times \dfrac{p}{q} = 2p < 1 & \text{if } p < \frac{1}{2}. \end{cases}$$

So for the simple symmetric random walk ($p = \frac{1}{2}$) we have $m_i = 1$ and are certain to return to the initial state again and again, while for $p \neq \frac{1}{2}$, we have $m_i < 1$ and we might never return.

What about the expected return times μ_0? For $p \neq \frac{1}{2}$, we have $\mu_0 = \infty$, since $m_0 < 1$ and we might never come back. For $p = \frac{1}{2}$, though, $m_0 = 1$, so we have work to do.

To get the expected return time for $p = \frac{1}{2}$, we'll need the expected hitting times for for $p = \frac{1}{2}$ too. Conditioning on the first step gives the equation

$$\eta_{i0} = 1 + \tfrac{1}{2}\eta_{i+10} + \tfrac{1}{2}\eta_{i-10},$$

with initial condition $\eta_{00} = 0$. We again recognise from the gambler's ruin problem the general solution $\eta_{i0} = A + Bi - i^2$. The initial condition gives $A = 0$, so $\eta_{i0} = Bi - i^2 = i(B - i)$. Then non-negativity demands $B = \infty$, as any other value would get drowned out by the large negative value $-i^2$, making the whole expression negative for big enough i. Thus we have $\eta_{i0} = \infty$.

Then we see easily that the return time is

$$\mu_0 = 1 + \tfrac{1}{2}\eta_{10} + \tfrac{1}{2}\eta_{-10} = 1 + \tfrac{1}{2}\infty + \tfrac{1}{2}\infty = \infty.$$

So for $p = \frac{1}{2}$, while the random walk always return, it may take a very long time to do so.

In conclusion, this is what we found out:

Theorem 8.2. *Consider the simple random walk with $p \neq 0, 1$. Then for all states i we have the following:*

- *For $p \neq \frac{1}{2}$, the return probability m_i is strictly less than 1, so the expected return time μ_i is infinite.*
- *For the simple symmetric random walk with $p = \frac{1}{2}$, the return probability m_i is equal to 1, but the expected return time μ_i is still infinite.*

In the next section, we look at how the return probabilities and expected return times characterise which states continue to be visited by a Markov chain in the long run.

Problem sheet 4

You should attempt all these questions and write up your solutions in advance of your workshop in week 5 (Monday 22 or Tuesday 23 February) where the answers will be discussed.

1. Consider the Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4\}$ and transition matrix

$$
\mathsf{P} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1-\alpha & 0 & \alpha & 0 \\
0 & \beta & 0 & 1-\beta \\
0 & 0 & 0 & 1
\end{pmatrix}
$$

where $0 < \alpha, \beta < 1$.

(a) Draw a transition diagram for this Markov chain.

(b) What are the communicating classes for this Markov chain? Is the chain irreducible? Which classes are closed? Which states are absorbing?

(c) Find the hitting probability h_{21} that, starting from state 2, the chain hits state 1.

(d) What is the expected time, starting from state 2, to reach an absorbing state?

2. Consider the Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4, 5\}$ and transition matrix

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

(a) Draw a transition diagram for this Markov chain.

(b) Show that the chain is irreducible.

(c) What are the periods of the states?

(d) Find the expected hitting times η_{i1} from each state i to 1, and the expected return time μ_i to 1.

3.

(a) Show that every Markov chain on a finite state space $\mathcal{S}$ has at least one closed communicating class.

(b) Give an example of a Markov chain that has no closed communicating classes.

4. Prove or give a counterexample: The period of a state i is the smallest $d > 0$ such that $p_{ii}(d) > 0$.

5. Consider the simple random walk with $p < \frac{1}{2}$, and let $i > 0$. Show that $\eta_{i0} = i/(q-p)$.

6. Consider the Markov chain with the following transition matrix and transition diagram:

Figure 8: Transition diagram for Question 6.
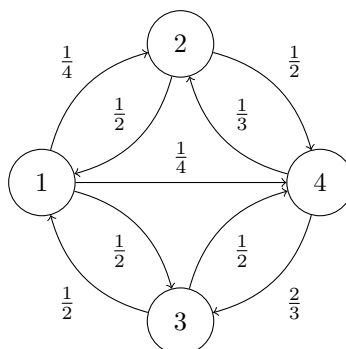
$$P = \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix}$$

The Markov chain starts from state 1.

(a) What is the probability that we hit state 2 before we hit state 3?

(b) What is the expected time until we hit a state in the set $\{2, 3\}$?

9 Recurrence and transience

- Definition of properties of recurrence and transience
- Recurrence and transience as class properties
- Positive and null recurrence

9.1 Recurrent and transient states

When thinking about the long-run behaviour of Markov chains, it's useful to classify two different types of states: "recurrent" states and "transient" states.

Recurrent states	Transient states
If we ever visit i, then we keep returning to i again and again	We might visit i a few times, but eventually we leave i and never come back
Starting from i, the expected number of visits to i is infinite	Starting from i, the expected number of visits to i is finite
Starting from i, the number of visits to i is certain to be infinite	Starting from i, the number of visits to i is certain to be finite
The return probability m_i equals 1	The return probability m_i is strictly less than 1

We'll take the last one of these, about the return probability, as the definition, then prove that the others follow.

Definition 9.1. Let (X_n) be a Markov chain on a state space $\mathcal{S}$. For $i \in \mathcal{S}$, let m_i be the return probability
$$m_i = \mathbb{P}(X_n = i \text{ for some } n \geq 1 \mid X_0 = i).$$
If $m_i = 1$, we say that state i is **recurrent**; if $m_i < 1$, we say that state i is **transient**.

Before stating this theorem, let us note that, from the point we're at state i, the expected number of visits to i is
$$\mathbb{E}(\# \text{ visits to } i \mid X_0 = i) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = i \mid X_0 = i) = \sum_{n=1}^{\infty} p_{ii}(n).$$

Theorem 9.1. *Consider a Markov chain with transition matrix* P.

- *If the state i is recurrent, then $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$, and we return to state i infinitely many times with probability 1.*
- *If the state i is transient, then $\sum_{n=1}^{\infty} p_{ii}(n) < \infty$, and we return to state i infinitely many times with probability 0.*

We'll come to the proof in a moment, but first some examples.

Example 9.1. Consider the simple random walk. In the last section we saw that for the simple symmetric random walk with $p = \frac{1}{2}$ we have $m_i = 1$, so the simple symmetric random walk is recurrent, but for $p \neq \frac{1}{2}$ we have $m_i < 1$, so all the other simple random walks are transient.

Example 9.2. We saw this chain previously as Example 7.7:

For states 5, 6 and 7, it's clear that the return probability is 1, since the Markov chain cycles around the triangle, so these states are recurrent.

States 1, 2, 3 and 4 are transient. In a moment we'll see a very quick way to show this, but in the meantime we can prove it directly by getting our hands dirty.

From state 4, we might go straight to state 5, in which case we can't come back, so $m_4 \leq 1 - p_{45} = \frac{2}{3}$, and state 4 is transient. Similarly, if we move from 1 to 4 to 5, we definitely won't come back to 1, so $m_1 \leq 1 - p_{14}p_{45} = \frac{5}{6}$, and state 1 is transient. By the similar arguments, $m_3 \leq 1 - p_{34}p_{45} = \frac{5}{6}$, and $m_2 \leq 1 - p_{21}p_{14}p_{45} = \frac{11}{12}$, so these states are both transient too.
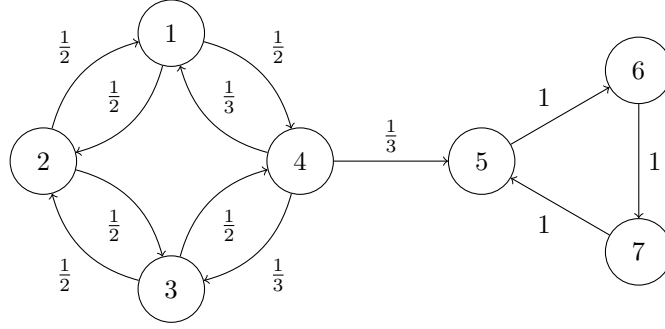
Figure 9: Transition diagram from Subsection 7.2.

Notice that all states in the communicating class $\{1, 2, 3, 4\}$ are transient, while all states in the communicating class $\{5, 6, 7\}$ are recurrent. We shall return to this point shortly. But first, we've put off the proof for too long.

Proof of Theorem 9.1. Suppose state i is recurrent. So starting from i, the probability we return is $m_i = 1$. After that return, it's as if we restart the chain from i, because of the Markov property – so the probability we return to i is again still $m_i = 1$. Repeating this, we keep on returning, definitely visit infinitely often (with probability 1). In particular, since the number of visits to i starting from i is always infinite, its expectation is infinite too, and this expectation is $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$.

Suppose, on the other hand, that state i is transient. So starting from i the probability we return is $m_i < 1$. Then the probability we return to i exactly r times before never coming back is

$$\mathbb{P}\big((\text{number of returns to } i) = r\big) = m_i^r (1 - m_i),$$

since we must return on the first r occasions, but then fail to return on the next occasion.. This is a geometric distribution $\text{Geom}(1 - m_i)$ (the version with support $\{0, 1, 2, \dots\}$). Since the expectation of this type of $\text{Geom}(p)$ random variable is $(1 - p)/p$, the expected number of returns is

$$\mathbb{E}(\text{number of returns to } i) = \sum_{n=1}^{\infty} p_{ii}(n) = \frac{1 - (1 - m_i)}{1 - m_i} = \frac{m_i}{1 - m_i}.$$

This is finite, since $m_i < 1$. Since the expected number of returns is finite, the probability we return infinitely many times must be 0.

$\square$

9.2 Recurrent and transient classes

We could find whether each state is transient or recurrent by calculating (or bounding) all the return probabilities m_i, using the methods in the previous section. But the following two theorems will give some highly convenient short-cuts.

Theorem 9.2. *Within a communicating class, either every state is transient or every state is recurrent.*

Formally: Let $i, j \in \mathcal{S}$ be such that $i \leftrightarrow j$. If i is recurrent, then j is recurrent also; while if i is transient, then j transient also.

For this reason, we can refer to a communicating class as a "recurrent class" or a "transient class". If a Markov chain is irreducible, we can refer to it as a "recurrent Markov chain" or a "transient Markov chain".

Proof. First part. Suppose $i \leftrightarrow j$ and i is recurrent. Then, for some n, m we have $p_{ij}(n), p_{ji}(m) > 0$.

Then, by the Chapman–Kolmogorov equations,

$$\sum_{r=1}^{\infty} p_{jj}(n+m+r) \geq \sum_{r=1}^{\infty} p_{ji}(m)p_{ii}(r)p_{ij}(n)$$

$$= p_{ji}^{(m)} \left(\sum_{r=1}^{\infty} p_{ii}(r) \right) p_{ij}(n).$$

If i is recurrent, then $\sum_r p_{ii}(r) = \infty$. Then from the above equation, we also have $\sum_r p_{jj}(n+m+r) = \infty$, meaning $\sum_s p_{jj}(s) = \infty$, and j is recurrent.

Second part. Suppose i is transient. Then j cannot be recurrent, because if it were, the previous argument with i and j swapped over would force i to in fact be recurrent also. So j must be transient.

$\square$

Theorem 9.3.

- *Every non-closed communicating class is transient.*
- *Every finite closed communicating class is recurrent.*

This theorem completely classifies the transience and recurrence of classes, with rare exception of infinite closed classes, which can require further examination.

Proof. First part. Suppose i is in a non-closed communicating class, so for some j we have $i \to j$, meaning $p_{ij}(n) > 0$ for some n, but $j \not\to i$, meaning that once we reach j we cannot return to i. We need to show that i is transient.

Consider the probability we return to i after time n. We condition on whether $X_n = j$ or not. This gives

$$\mathbb{P}(\text{return to } i \text{ after time } n \mid X_0 = i)$$
$$= p_{ij}(n)\, \mathbb{P}(\text{return to } i \text{ after time } n \mid X_n = j, X_0 = i)$$
$$\quad + (1 - p_{ij}(n))\, \mathbb{P}(\text{return to } i \text{ after time } n \mid X_n \neq j, X_0 = i)$$
$$\leq \mathbb{P}(\text{return to } i \text{ after time } n \mid X_n = j, X_0 = i) + (1 - p_{ij}(n))$$
$$= 0 + (1 - p_{ij}(n))$$
$$< 1,$$

since we can't get from j to i, and since $p_{ij}(n) > 0$. If i were recurrent we would certainly return infinitely often, and in particular certainly return after time n. So i must be transient instead.

Second part. Suppose the class C is finite and closed. Then there must be an $i \in C$ such that, once we visit i, the probability that we return to i infinitely many times is strictly positive; this is because we are going to stay in the finitely many states of C for infinitely many time steps. Then that state i is not transient, so it must be recurrent, which means that the whole class is recurrent.

$\square$

Going back to the earlier example, we see that the class $\{5, 6, 7\}$ is closed and finite, and therefore recurrent, while class $\{1, 2, 3, 4\}$ is not closed and therefore transient. This is much less effort than the previous method!

9.3 Positive and null recurrence

It will be useful later to further divide recurrent classes, where the return probability $m_i = 1$, by whether the expected return time μ_i is finite or not. (Note that transient states always have $\mu_i = \infty$.)

Definition 9.2. Let (X_n) be a Markov chain on a state space $\mathcal{S}$. Let $i \in \mathcal{S}$ be a recurrent state, so $m_i = 1$, and let μ_i be the expected return time. If $\mu_i < \infty$, we say that state i is **positive recurrent**; if $\mu_i = \infty$, we say that state i is **null recurrent**.

The following facts can be proven in a similar way to the previous results:

1. In a recurrent class, either all states are positive recurrent or all states are null recurrent.
2. All finite closed classes are positive recurrent.

The first result means we can refer to a "positive recurrent class" or a "null recurrent class", and an irreducible Markov chain can be a "positive recurrent Markov chain" or a "null recurrent Markov chain".

Putting everything so far together, we have the following classification:

- non-closed classes are transient;
- finite closed classes are positive recurrent;
- infinite closed classes can be positive recurrent, null recurrent, or transient.

We know that the simple symmetric random walk is recurrent. We also saw in the last section that $\mu_i = \infty$, so it is null recurrent.

We can also consider the simple symmetric random walk in d-dimensions, on $\mathbb{Z}^d$. At each step we pick one of the coordinates and increase or decrease it by one; each of the $2d$ possibilities having probability $1/(2d)$. We have seen that for $d = 1$ this is null recurrent. A famous result by the Hungarian mathematician George Pólya from 1921 states the simple symmetric random walk is null recurrent for $d = 1$ and $d = 2$, but is transient for $d \geq 3$. (Perhaps this is why cars often crash into each other, but aeroplanes very rarely do?)

9.4   Strong Markov property

This subsection is optional and nonexaminable.

There was a cheat somewhere earlier in this section – did you notice it?

The Markov property says that, if at some fixed time n we have $X_n = i$, then the Markov chain from that point on is just like starting all over again from the state i. When we applied this in the proof of Theorem 9.1, we were using as n the first return to state i. But that's not a fixed time – it's a random time! Did we cheat?

Actually we're fine. The reason is that the first return to i isn't just any old random time, it's a "stopping time", and the Markov property also applies to stopping times too. Roughly speaking, a stopping time is a random time which has the property that "you know when you get there".

Definition 9.3. Let (X_n) be a stochastic process in discrete time, and let T be a random time. Then T is a **stopping time** if for all n, whether or not the event $\{T = n\}$ occurs is completely determined by the random variables $X_0, X_1, \ldots, X_n$.

So, for example:

- "The first visit to state i" is stopping time, because as soon as we reach i, we know the value of T.
- "Three time-steps after the second visit to j" is a stopping time, because after our second visit we count on three more steps and have T.
- "The time-step *before* the first visit to i" is not a stopping time, because we still need to go one step further on to know whether we had just been at time T or not.
- "The final visit to j" is not a stopping time, because at the time of the visit we don't yet know whether we'll come back again or not.

There are lots of places in probability theory and finance when something that is true about a fixed time is also true about a random stopping time. When we use the Markov property with a stopping time, we call it the "strong Markov property".

Theorem 9.4 (Strong Markov property). *Let (X_n) be a Markov chain on a state space $\mathcal{S}$, and let T be a stopping time that is finite with probability 1. Then all states $x_0, \ldots, x_{T-1}, i, j \in \mathcal{S}$ we have*

$$\mathbb{P}(X_{T+1} = j \mid X_T = i, X_{T-1} = x_{T-1} \ldots, X_0 = x_0) = p_{ij}.$$

Proof. We have

$$\mathbb{P}(X_{T+1} = x_j \mid X_T = i, X_{T-1} = x_{T-1} \dots, X_0 = x_0)$$

$$= \sum_{n=0}^{\infty} \mathbb{P}(T = n)\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1} \dots, X_0 = x_0, T = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{P}(T = n)\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1} \dots, X_0 = x_0)$$

$$= \sum_{n=0}^{\infty} \mathbb{P}(T = n)\mathbb{P}(X_{n+1} = j \mid X_n = i)$$

$$= \sum_{n=0}^{\infty} \mathbb{P}(T = n)p_{ij}$$

$$= p_{ij} \sum_{n=0}^{\infty} \mathbb{P}(T = n)$$

$$= p_{ij},$$

as desired. The second line was by conditioning on the value of T; in the third line we deleted the superfluous conditioning $T = n$, because T is a stopping time, so the event $T = n$ is entirely decided by $X_n, X_{n-1}, \dots, X_0$; the fourth line used the (usual non-strong) Markov property; the fifth line is just the definition of p_{ij}; the sixth line took p_{ij} out of the sum; and the seventh line is because T is finite with probability 1, so $\mathbb{P}(T = n)$ sums to 1.

$\square$

9.5   A useful lemma

This subsection is optional and nonexaminable.

The following lemma will be used in some later optional and nonexaminable proofs.

Lemma 9.1. *Let (X_n) be an irreducible and recurrent Markov chain. Then for any initial distribution and any state j, we will certainly hit j, so the hitting time H_j is finite with probability 1.*

Proof. It suffices to prove the lemma when the initial distribution is "start at i". (We can repeat for all i, then build any initial distribution from a weighted sum of "start at i"s.)

Since the chain is irreducible, we have $j \to i$, so pick m with $p_{ji}(m) > 0$. Since the chain is recurrent, we know the return probability from j to j is 1, and we return infinitely many times with probability 1. We just need to glue these two facts together.

We have

$$1 = \mathbb{P}(X_n = j \text{ for infinitely many } n \mid X_0 = j)$$

$$= \mathbb{P}(X_n = j \text{ for some } n > m \mid X_0 = j)$$

$$= \sum_k \mathbb{P}(X_m = k \mid X_0 = j)\, \mathbb{P}(X_n = j \text{ for some } n > m \mid X_m = k, X_0 = j)$$

$$= \sum_k p_{jk}(m)\, \mathbb{P}(H_j < \infty \mid X_0 = k),$$

where the last line used the Markov property to treat the chain as starting over again when it reaches some state k at time m. Note that $\sum_k p_{jk}(m) = 1$, since that's the sum of the probabilities of going anywhere in m steps. This means we must have $\mathbb{P}(H_j < \infty \mid X_0 = k)$ whenever $p_{jk}(m) > 0$, to ensure the final line does indeed sum to 1. But we stated earlier that $p_{ji}(m) > 0$, so we indeed have $\mathbb{P}(H_j < \infty \mid X_0 = i)$, as required.

$\square$

In the next section, we look at how positive recurrent Markov chains can settle into a "stationary distribution" and experience long-term stability.

10 Stationary distributions

- Stationary distributions and how to find them
- Conditions for existence and uniqueness of the stationary distribution

10.1 Definition of stationary distribution

Consider the two-state "broken printer" Markov chain from Lecture 5.
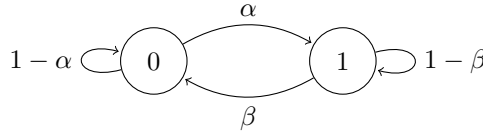
Figure 10: Transition diagram for the two-state broken printer chain.

Suppose we start the chain from the initial distribution

$$\lambda_0 = \mathbb{P}(X_0 = 0) = \frac{\beta}{\alpha + \beta} \qquad \lambda_1 = \mathbb{P}(X_0 = 1) = \frac{\alpha}{\alpha + \beta}.$$

(You may recognise this from Question 3 on Problem Sheet 3 and the associated video.) What's the distribution after step 1? By conditioning on the initial state, we have

$$\mathbb{P}(X_1 = 0) = \lambda_0 p_{00} + \lambda_1 p_{10} = \frac{\beta}{\alpha + \beta}(1 - \alpha) + \frac{\alpha}{\alpha + \beta}\beta = \frac{\beta}{\alpha + \beta},$$

$$\mathbb{P}(X_1 = 1) = \lambda_0 p_{01} + \lambda_1 p_{11} = \frac{\beta}{\alpha + \beta}\alpha + \frac{\alpha}{\alpha + \beta}(1 - \beta) = \frac{\alpha}{\alpha + \beta}.$$

So we're still in the same distribution we started in. By repeating the same calculation, we're still going to be in this distribution after step 2, and step 3, and forever.

More generally, if we start from a state given by a distribution $\pi = (\pi_i)$, then after step 1 the probability we're in state j is $\sum_i \pi_i p_{ij}$. So if $\pi_j = \sum_i \pi_i p_{ij}$, we stay in this distribution forever. We call such a distribution a stationary distribution. We again recodnise this formula as a matrix–vector multiplication, so this is $\pi = \pi \mathsf{P}$, where π is a *row* vector.

Definition 10.1. Let (X_n) be a Markov chain on a state space $\mathcal{S}$ with transition matrix P. Let $\pi = (\pi_i)$ be a distribution on $\mathcal{S}$, in that $\pi_i \geq 0$ for all $i \in \mathcal{S}$ and $\sum_{i \in \mathcal{S}} \pi_i = 1$. We call π a **stationary distribution** if

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_i p_{ij} \quad \text{for all } j \in \mathcal{S},$$

or, equivalently, if $\pi = \pi \mathsf{P}$.

Note that we're saying the *distribution* $\mathbb{P}(X_n = i)$ stays the same; the Markov chain (X_n) itself will keep moving. One way to think is that if we started off a thousand Markov chains, choosing each starting position to be i with probability π_i, then (roughly) $1000\pi_j$ of them would be in state j at any time in the future – but not necessarily the same ones each time.

10.2 Finding a stationary distribution

Let's try an example. Consider the no-claims discount Markov chain from Lecture 6 with state space $\mathcal{S} = \{1, 2, 3\}$ and transition matrix

$$\mathsf{P} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

We want to find a stationary distribution π, which must solve the equation $\pi = \pi\mathsf{P}$, which is

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

Writing out the equations coordinate at a time, we have

$$\pi_1 = \tfrac{1}{4}\pi_1 + \tfrac{1}{4}\pi_2,$$
$$\pi_2 = \tfrac{3}{4}\pi_1 + \tfrac{1}{4}\pi_3,$$
$$\pi_3 = \tfrac{3}{4}\pi_2 + \tfrac{3}{4}\pi_3.$$

Since π must be a distribution, we also have the "normalising condition"

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

The way to solve these equations is first to solve for all the variables π_i in terms of a convenient π_j (called the "working variable") and then substitute all of these expressions into the normalising condition to find a value for π_j.

Let's choose π_2 as our working variable. It turns out that $\pi = \pi\mathsf{P}$ always gives one more equation than we actually need, so we can discard one of them for free. Let's get rid of the second equation, and the solve the first and third equations in terms of our working variable π_2, to get

$$\pi_1 = \tfrac{1}{3}\pi_2 \qquad \pi_3 = 3\pi_2. \tag{7}$$

Now let's turn to the normalising condition. That gives

$$\pi_1 + \pi_2 + \pi_3 = \tfrac{1}{3}\pi_2 + \pi_2 + 3\pi_2 = \tfrac{13}{3}\pi_2 = 1.$$

So the working variable is solved to be $\pi_2 = \frac{3}{13}$. Substituting this back into (7), we have $\pi_1 = \tfrac{1}{3}\pi_2 = \frac{1}{13}$ and $\pi_3 = 3\pi_2 = \frac{9}{13}$. So the full solution is

$$\pi = (\pi_1, \pi_2, \pi_3) = \left(\tfrac{1}{13}, \tfrac{3}{13}, \tfrac{9}{13} \right).$$

The method we used here can be summarised as follows:

1. Write out $\pi = \pi\mathsf{P}$ coordinate by coordinate. Discard one of the equations.
2. Select one of the π_i as a working variable and treat it as a parameter. Solve the equations in terms of the working variable.
3. Substitute the solution into the normalising condition to find the working variable, and hence the full solution.

It can be good practice to use the equation discarded earlier to check that the calculated solution is indeed correct.

One extra example for further practice and to show how you should present your solutions to such problems:

Example 10.1. *Consider a Markov chain on state space $\mathcal{S} = \{1, 2, 3\}$ with transition matrix*

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

Find a stationary distribution for this Markov chain.

Step 1. Writing out $\pi = \pi\mathsf{P}$ coordinate-wise, we have

$$\pi_1 = \tfrac{1}{2}\pi_1 + \tfrac{1}{4}\pi_2$$
$$\pi_2 = \tfrac{1}{4}\pi_1 + \tfrac{1}{2}\pi_2 + \tfrac{1}{4}\pi_3$$
$$\pi_3 = \tfrac{1}{4}\pi_1 + \tfrac{1}{4}\pi_2 + \tfrac{3}{4}\pi_3.$$

We choose to discard the third equation.

Step 2. We choose π_1 as our working variable. From the first equation we get $\pi_2 = 2\pi_1$. From the second equation we get $\pi_3 = 2\pi_2 - \pi_1$, and substituting the previous $\pi_2 = 2\pi_1$ into this, we get $\pi_3 = 3\pi_1$.

Step 3. The normalising condition is

$$\pi_1 + \pi_2 + \pi_3 = \pi_1 + 2\pi_1 + 3\pi_1 = 6\pi_1 = 1.$$

Therefore $\pi_1 = \frac{1}{6}$. Substituting this into our previous expressions, we get $\pi_2 = 2\pi_1 = \frac{1}{3}$ and $\pi_3 = 3\pi_1 = \frac{1}{2}$. Thus the solution is $\pi = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$.

We can check our answer with the discarded third equation, just to make sure we didn't make any mistakes. We get

$$\pi_3 = \tfrac{1}{4}\pi_1 + \tfrac{1}{4}\pi_2 + \tfrac{3}{4}\pi_3 = \tfrac{1}{4}\tfrac{1}{6} + \tfrac{1}{4}\tfrac{1}{3} + \tfrac{3}{4}\tfrac{1}{2} = \tfrac{1}{24} + \tfrac{2}{24} + \tfrac{9}{24} = \tfrac{1}{2},$$

which is as it should be.

10.3 Existence and uniqueness

Given a Markov chain its natural to ask:

1. Does a stationary distribution exist?
2. If a stationary distribution does exists, is there only one, or are there be many stationary distributions?

The answer is given by the following very important theorem.

Theorem 10.1. *Consider an irreducible Markov chain.*

- *If the Markov chain is positive recurrent, then a stationary distribution π exists, is unique, and is given by $\pi_i = 1/\mu_i$, where μ_i is the expected return time to state i.*
- *If the Markov chain is null recurrent or transient, then no stationary distribution exists.*

We give an optional and nonexaminable proof to the first part below.

In our no-claims discount example, the chain is irreducible and, like all finite state irreducible chains, it is positive recurrent. Thus the stationary distribution $\pi = (\frac{1}{13}, \frac{3}{13}, \frac{9}{13})$ we found is the unique stationary distribution for that chain. Once we have the stationary distribution π, we get the expected return times $\mu_i = 1/\pi_i$ for free: the expected return times are $\mu_1 = 13$, $\mu_2 = \frac{13}{3} = 4.33$, and $\mu_3 = \frac{13}{9} = 1.44$.

Note the condition in Theorem 10.1 that the Markov chain is irreducible. What if the Markov chain is not irreducible, so has more than one communicating class? We can work out what must happen from the theorem:

- If none of the classes are positive recurrent, then no stationary distribution exists.
- If exactly one of the classes is positive recurrent (and therefore closed), then there exists a unique stationary distribution, supported only on that closed class.
- If more the one of the classes are positive recurrent, then many stationary distributions will exist.

Example 10.2. Consider the simple random walk with $p \neq 0, 1$. This Markov chain is irreducible, and is null recurrent for $p = \frac{1}{2}$ and transient for $p \neq \frac{1}{2}$. Either way, the theorem tells us that no stationary distribution exists.

Example 10.3. Consider the Markov chain with transition matrix

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

This chain has two closed positive recurrent classes, $\{1, 2\}$ and $\{3, 4\}$.

Solving $\pi = \pi\mathsf{P}$ gives

$$
\begin{aligned}
\pi_1 &= \tfrac{1}{2}\pi_1 + \tfrac{1}{2}\pi_2 &\Rightarrow& &\pi_1 &= \pi_2 \\
\pi_2 &= \tfrac{1}{2}\pi_1 + \tfrac{1}{2}\pi_2 &\Rightarrow& &\pi_1 &= \pi_2 \\
\pi_3 &= \tfrac{1}{4}\pi_3 + \tfrac{1}{2}\pi_4 &\Rightarrow& &3\pi_3 &= 2\pi_4 \\
\pi_4 &= \tfrac{3}{4}\pi_3 + \tfrac{1}{2}\pi_4 &\Rightarrow& &3\pi_3 &= 2\pi_4,
\end{aligned}
$$

giving us the same two constraints twice each. We also have the normalising condition $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$. If we let $\pi_1 + \pi_2 = \alpha$ and $\pi_3 + \pi_4 = 1 - \alpha$, we see that

$$
\pi = \begin{pmatrix} \tfrac{1}{2}\alpha & \tfrac{1}{2}\alpha & \tfrac{2}{5}(1-\alpha) & \tfrac{3}{5}(1-\alpha) \end{pmatrix}
$$

is a stationary distribution for any $0 \le \alpha \le 1$, so we have infinitely many stationary distributions.

10.4   Proof of existence and uniqueness

This subsection is optional and nonexaminable.

It's very important to be able to find the stationary distribution(s) of a Markov chain – you can reasonably expect a question on this to turn up on the exam. You should also know the conditions for existence and uniqueness of the stationary distribution. Being able to *prove* existence and uniqueness is less important, although for completeness we will do so here.

Theorem 10.1 had two points. The more important point was that irreducible, positive recurrent Markov chains have a stationary distribution, that it is unique, and that it is given by $\pi_i = 1/\mu_i$. We give a proof of that below, doing the existence and uniqueness parts separately. The less important point was that null recurrent and transitive Markov chains do not have a stationary distribution, and this is more fiddly. You can find a proof (usually in multiple parts) in books such as Norris, *Markov Chains*, Section 1.7.

Existence: *Every positive recurrent Markov chain has a stationary distribution.*

Before we start, one last definition. Let us call a vector ν a **stationary vector** if $\nu\mathsf{P} = \nu$. This is exactly like a stationary distribution, except without the normalisation condition that it has to sum to 1.

Proof. Suppose that (X_n) is recurrent (either positive or null, for the moment).

Our first task will be to find a stationary vector. Fix an initial state k, and let ν_i be the expected number of visits to i before we return back to k. That is,

$$
\begin{aligned}
\nu_i &= \mathbb{E}(\# \text{ visits to } i \text{ before returning to } k \mid X_0 = k) \\
&= \mathbb{E}\sum_{n=1}^{M_k} \mathbb{P}(X_n = i \mid X_0 = k) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \text{ and } n \le M_k \mid X_0 = k),
\end{aligned}
$$

where M_k is the return time, as in Section 8. Let us note for later use that, under this definition, $\nu_k = 1$, because the only visit to k is the return to k itself.

Since ν is counting the number of visits to different states in a certain (random) time, it seems plausible that ν suitably normalised could be a stationary distribution, meaning that ν itself could be a stationary vector. Let's check.

We want to show that $\sum_i \nu_i p_{ij} = \nu_j$. Let's see what we have:

$$\sum_{i \in \mathcal{S}} \nu_i p_{ij} = \sum_{i \in \mathcal{S}} \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \text{ and } n \leq M_k \mid X_0 = k) \, p_{ij}$$

$$= \sum_{n=1}^{\infty} \sum_{i \in \mathcal{S}} \mathbb{P}(X_n = i \text{ and } X_{n+1} = j \text{ and } n \leq M_k \mid X_0 = k)$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(X_{n+1} = j \text{ and } n \leq M_k \mid X_0 = k).$$

(Exchanging the order of the sums is legitimate, because recurrence of the chain means that M_k is finite with probability 1.) We can now do a cheeky bit of monkeying around with the index n, by swapping out the visit to k at time M_k with the visit to k at time 0. This means instead of counting the visits from 1 to M_k, we can count the visits from 0 to $M_k - 1$. Shuffling the index about, we get

$$\sum_{i \in \mathcal{S}} \nu_i p_{ij} = \sum_{n=0}^{\infty} \mathbb{P}(X_{n+1} = j \text{ and } n \leq M_k - 1 \mid X_0 = k)$$

$$= \sum_{n+1=1}^{\infty} \mathbb{P}(X_{n+1} = j \text{ and } n + 1 \leq M_k \mid X_0 = k)$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(X_n = j \text{ and } n \leq M_k \mid X_0 = k)$$

$$= \nu_j.$$

So ν is indeed a stationary vector.

We now want normalise ν into a stationary distribution by dividing through by $\sum_i \nu_i$. We can do this if $\sum_i \nu_i$ is finite. But $\sum_i \nu_i$ is the expected total number of visits to all states before return to k, which is precisely the expected return time μ_k. Now we use the assumption that (X_n) is *positive* recurrent. This means that μ_k is finite, so $\pi = (1/\mu_k)\nu$ is a stationary distribution.

$\square$

Uniqueness: *For an irreducible, positive recurrent Markov chain, the stationary distribution is unique and is given by $\pi_i = 1/\mu_i$.*

I read the following proof in Stirzaker, *Elementary Probability*, Section 9.5.

Proof. Suppose the Markov chain is irreducible and positive recurrent, and suppose π is a stationary distribution. We want to show that $\pi_i = 1/\mu_i$ for all i.

The only equation we have for μ_k is this one from Section 8:

$$\mu_k = 1 + \sum_j p_{kj} \eta_{jk}. \tag{8}$$

Since that involves the expected hitting times η_{ik}, let's write down the equation for them too:

$$\eta_{ik} = 1 + \sum_j p_{ij} \eta_{jk} \qquad \text{for all } i \neq k. \tag{9}$$

In order to apply the fact that π is a stationary distribution, we'd like to get these into an equation with $\sum_i \pi_i p_{ij}$ in it. Here's a way we can do that: Take (9), multiply it by π_i and sum over all $i \neq k$, to get

$$\sum_i \pi_i \eta_{ik} = \sum_{i \neq k} \pi_i + \sum_j \sum_{i \neq k} \pi_i p_{ij} \eta_{jk}. \tag{10}$$

(The sum on the left can be over all i, since $\eta_{kk} = 0$.) Also, take (8) and multiply it by π_k to get

$$\pi_k \mu_k = \pi_k + \sum_j \pi_k p_{kj} \eta_{jk} \tag{11}$$

Now add (10) and (11) together to get

$$\sum_i \pi_i \eta_{ik} + \pi_k \mu_k = \sum_i \pi_i + \sum_j \sum_i \pi_i p_{ij} \eta_{jk}.$$

We can now use $\sum_i \pi_i p_{ij} = \pi_j$, along with $\sum_i \pi_i = 1$, to get

$$\sum_i \pi_i \eta_{ik} + \pi_k \mu_k = 1 + \sum_j \pi_j \eta_{jk}.$$

But the first term on the left and the last term on the right are equal, and because the Markov chain is irreducible and positive recurrent, they are finite. (That was our lemma in the previous section.) Thus we're allowed to subtract them, and we get $\pi_k \mu_k = 1$, which is indeed $\pi_k = 1/\mu_k$. We can repeat the argument for every choice of k.

$\square$

In the next section, we see how the stationary distribution tells us very important things about the long-term behaviour of a Markov chain.

Problem sheet 5

You should attempt all these questions and write up your solutions in advance of your workshop in week 6 (Monday 1 or Tuesday 2 March) where the answers will be discussed.

1. Find a stationary distribution for the Markov chain with transition matrix

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

2. Consider a Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4\}$ and transition matrix

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

(a) Draw a transition diagram for this Markov chain.

(b) Identify the communicating classes. State whether each class closed or not. State whether each class is positive recurrent, null recurrent, or transient.

(c) Find a stationary distribution for this Markov chain.

(d) Is this the only stationary distribution?

3. Consider a Markov chain with state space $\mathcal{S} = \{1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{1}{5} & \frac{1}{5} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

(a) Draw a transition diagram for this Markov chain.

(b) Identify the communicating classes. State whether each class closed or not. State if each class is positive recurrent, null recurrent, or transient.

(c) Find all of the stationary distributions for this Markov chain.

4. Consider the simple random walk (X_n) on $\mathcal{S} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ with up probability p and down probability $q = 1 - p$, and a mixed barrier at 0, where $p_{01} = p$, $p_{00} = q$, and $p_{0i} = 0$ otherwise. We seek a stationary distribution for this Markov chain.

(a) Suppose $p \neq \frac{1}{2}$. Show that the general solution to

$$\pi_j = \sum_i \pi_i p_{ij} = p\pi_{j-1} + q\pi_{j+1}$$

is $\pi_i = A + B\tau^i$, where $\tau = p/q$.

(b) Show that the initial condition $\pi_0 = q\pi_0 + q\pi_1$ gives $A = 0$.

(c) By considering the normalising condition $\sum_i \pi_i = 1$, work out for what values of $p \neq \frac{1}{2}$ there exists a stationary distribution for (X_n). What is the stationary distribution (when it exists)?

(d) Does there exist a stationary distribution when $p = \frac{1}{2}$?

5. The infinite rooted binary tree is a graph with no cycles. There is one special vertex, the root 0, that has two edges, and every other edge has three edges. A Markov chain (X_n) starts from 0, then at each time step, takes one of the edges coming out of the current vertex and moves along it to the neighbouring vertex. Note that a step can go away from the root or towards the root.

By considering the distance of (X_n) from the root, or otherwise, show that (X_n) is transient.

6. *"Every Markov chain on a finite state space has at least one stationary distribution."* Explain carefully why this is true. You may use facts from the notes or previous example sheets, provided that you state them clearly.
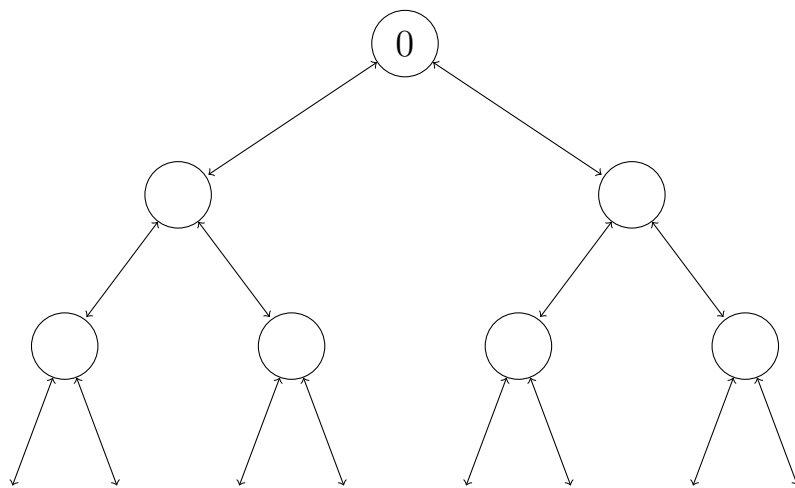
Figure 11: The first three-and-a-bit levels of the rooted binary tree.

11 Long-term behaviour of Markov chains

- The limit theorem: convergence to the stationary distribution for irreducible, aperiodic, positive recurrent Markov chains
- The ergodic theorem for the long-run proportion of time spent in each state

11.1 Convergence to equilibrium

In this section we're interested in what happens to a Markov chain (X_n) in the long-run – that is, when n tends to infinity.

One thing that *could* happen over time is that the distribution $\mathbb{P}(X_n = i)$ of the Markov chain could gradually settle down towards some "equilibrium" distribution. Further, perhaps that long-term equilibrium might not depend on the initial distribution, but the effects of the initial distribution might eventually almost disappear, exhibiting a "lack of memory" of the start of the process.

Just in case that does happen, let's give it a name.

Definition 11.1. Let (X_n) be a Markov chain on a state space $\mathcal{S}$ with transition matrix P. Suppose there exists a distribution $\mathbf{p}^* = (p_i^*)$ on $\mathcal{S}$ (so $p_i^* \geq 0$ and $\sum_i p_i^* = 1$) such that, whatever the initial distribution $\lambda = (\lambda_i)$, we have $\mathbb{P}(X_n = j) \to p_j^*$ as $n \to \infty$ for all $j \in \mathcal{S}$. Then we say that $\mathbf{p}^*$ is an **equilibrium distribution**.

It's clear there can only be at most one equilibrium distribution – but will there be one at all? The following is the most important result in this course. (Recall that an irreducible Markov chain is aperiodic if it has period 1.)

Theorem 11.1 (Limit theorem). *Let (X_n) be an irreducible and aperiodic Markov chain. Then for any initial distribution λ, we have that $\mathbb{P}(X_n = j) \to 1/\mu_j$ as $n \to \infty$, where μ_j is the expected return time to state j.*

In particular:

- *Suppose (X_n) is positive recurrent. Then the unique stationary distribution π given by $\pi_j = 1/\mu_j$ is the equilibrium distribution, so $\mathbb{P}(X_n = j) \to \pi_j$ for all j.*
- *Suppose (X_n) is null recurrent or transient. Then $\mathbb{P}(X_n = j) \to 0$ for all j, and there is no equilibrium distribution.*

I particularly like how this one theorem gathers together all the ideas from the course in one result (Markov chains, irreducibility, periodicity, recurrence/transience, positive/null recurrence, return times, stationary distribution...).

Note the three conditions for convergence to an equilibrium distribution: irreducibility, aperiodicity, and positive recurrence.

Consider a irreducible, aperiodic, positive recurrent Markov chain. Taking the initial distribution to be starting in state i with certainty, the limit theorem tells us that $p_{ij}(n) \to \pi_j$ for all i and j. This means that the n-step transition matrix will have the limiting value

$$\lim_{n \to \infty} \mathsf{P}(n) = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix},$$

where each row is identical.

We give an full proof of the limit theorem below (optional and nonexaminable). However, this easier result gets part way there.

Theorem 11.2. *If an equilibrium distribution $\mathbf{p}^*$ does exist, then $\mathbf{p}^*$ is a stationary distribution.*

Given this result it's clear that an irreducible Markov chain cannot have an equilibrium distribution if it is null recurrent or transient, as it doesn't even have a stationary distribution. So the positive recurrent case is the hard (nonexaminable) one.

Proof. We need to verify that $\mathbf{p}^*\mathsf{P} = \mathbf{p}^*$. We have

$$\sum_i p_i^* p_{ij} = \sum_i \left(\lim_{n\to\infty} p_{ki}(n) \right) p_{ij} = \lim_{n\to\infty} \sum_i p_{ki}(n) p_{ij} = \lim_{n\to\infty} p_{kj}(n+1) = p_j^*,$$

as desired.

$\square$

(Strictly speaking, swapping the sum and the limit is only formally justified when the state space is finite, although the theorem is true universally.)

11.2 Examples of convergence and non-convergence

Example 11.1. The two-state "broken printer" Markov chain is irreducible, aperiodic, and positive recurrent, so its stationary distribution is also the equilibrium distribution. We proved this from first principles in Question 3 on Problem Sheet 3.

Example 11.2. Recall the simple no-claims discount Markov chain from Lecture 6, which is irreducible, aperiodic, and positive recurrent. We saw last time that is has the unique stationary distribution

$$\pi = \begin{pmatrix} \frac{1}{13} & \frac{3}{13} & \frac{9}{13} \end{pmatrix} = (0.0769 \quad 0.2308 \quad 0.6923).$$

From the limit theorem, we see that the n-step transition probability tends to a limit where every row is equal to π. We can check using a computer: for $n = 12$, say,

$$\mathsf{P}(12) = \mathsf{P}^{12} = \begin{pmatrix} 0.0770 & 0.2308 & 0.6923 \\ 0.0769 & 0.2308 & 0.6923 \\ 0.0769 & 0.2308 & 0.6923 \end{pmatrix},$$

where $p_{ij}(12)$ is equal to π_j up to at least 3 decimal places for all i, j. As n gets bigger, the matrix gets closer and closer to the limiting form.

Example 11.3. The simple random walk is null recurrent for $p = \frac{1}{2}$ and transient otherwise. Either way, we have $\mathbb{P}(X_n = i) \to 0$ for all states i, and there is no equilibrium distribution.

Example 11.4. Consider a Markov chain (X_n) on state space $\mathcal{S} = \{0, 1\}$ and transition matrix

$$\mathsf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

So at each stage we swap from state 0 to state 1 and back again. This chain is irreducible and positive recurrent, so it has a unique stationary distribution, which is clearly $\pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

However, we don't have convergence to equilibrium. If we start from initial distribution (λ_0, λ_1), then $\mathbb{P}(X_n = 0) = \lambda_0$ for even n and $\mathbb{P}(X_n = 0) = \lambda_1$ for odd n. When $\lambda_0 \neq \frac{1}{2}$, this does not converge.

The point is that this chain is *not* aperiodic: it has period 2, so the limit theorem does not apply.

Example 11.5. Consider a Markov chain with state space $\mathcal{S} = \{1, 2, 3\}$ and transition diagram as shown below.

This chain is not irreducible, but has two aperiodic and positive recurrent communicating classes. In particular, it has many stationary distributions including $(1, 0, 0)$ and $(0, \frac{8}{17}, \frac{9}{17})$ (optional exercise for the reader). If we start in state 1, then the limiting distribution is the former, while if we start in states 2 or 3, the limiting distribution is the latter.

In particular, as $n \to \infty$, we have

$$\mathsf{P}^{(n)} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{8}{17} & \frac{9}{17} \\ 0 & \frac{8}{17} & \frac{9}{17} \end{pmatrix}.$$
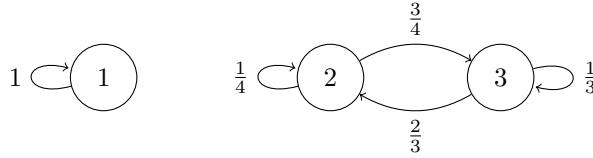
Figure 12: Transition diagram for a Markov chain with two positive recurrent classes.

11.3 Ergodic theorem

The limit theorem looked at the limit of $\mathbb{P}(X_n = j)$, the probability that the Markov chain is in state j at some specific point in time n a long time in the future. We could also look at the *long-run amount of time* spent in state j; that is, averaging the behaviour over a long time period. (The word "ergodic" is used in mathematics to refer to concepts to do with the long-term proportion of time.)

Let us write

$$V_j(N) := \#\{n < N : X_n = j\}$$

for the total number of visits to state j up to time N. Then we can interpret $V_j(n)/n$ as the proportion of time up to time n spent in state j, and its limiting value (if it exists) to be the **long-run proportion of time** spent in state j.

Theorem 11.3 (Ergodic theorem)**.** *Let (X_n) be an irreducible Markov chain. Then for any initial distribution λ we have that $V_j(n)/n \to 1/\mu_j$ almost surely as $n \to \infty$, where μ_j is the expected return time to state j.*

In particular:

- *Suppose (X_n) is positive recurrent. Then there is a unique stationary distribution π given by $\pi_j = 1/\mu_j$, and $V_j(n)/n \to \pi_j$ almost surely for all j.*
- *Suppose (X_n) is null recurrent or transient. Then $V_j(n)/n \to 0$ almost surely for all j.*

For completeness, we should note that "almost sure" convergence means that $\mathbb{P}(V_j(n)/n \to 1/\mu_j) = 1$, although the precise definition is not important for us in this module.

Note that, because we are averaging over a long-time period, we no longer need the condition that the Markov chain is aperiodic; for convergence of the long-term proportion of time to the stationary distribution we just need irreducibility and positive recurrence.

Again, we give an optional and nonexaminable proof below.

Example 11.6. Recall the simple no-claims discount Markov chain. Since this chain is irreducible and positive recurrent, we now see that the long-term proportion of time spent in each state corresponds to the stationary distribution $\pi = (\frac{1}{13} \quad \frac{3}{13} \quad \frac{9}{13})$. Therefore, over the lifetime of an insurance policy held for a long period of time, the average discount is approximately

$$\frac{1}{13}(0\%) + \frac{3}{13}(25\%) + \frac{9}{13}(50\%) = \frac{21}{52} = 40.4\%.$$

Example 11.7. The two-state "swapping" chain we saw earlier did have a unique stationary distribution $(\frac{1}{2}, \frac{1}{2})$, but did not have an equilibrium distribution, because it was periodic. But it is true that $V_0(n)/n \to \pi_0 = \frac{1}{2}$ and $V_1(n)/n \to \pi_1 = \frac{1}{2}$, due to the ergodic theorem. So although where we are at some specific point in the future depends on where we started from, in the long run we always spend half our time in each state.

11.4 Proofs of the limit and ergodic theorems

This subsection is optional and nonexaminable.

Again, it's important to be able to use the limit and ergodic theorems, but less important to be able to prove them.

First, the limit theorem. The only bit left is the first part: that for an irreducible, aperiodic, positive recurrent Markov chain, the stationary distribution π is an equilibrium distribution.

This cunning proof uses a technique called "coupling". When looking at two different random objects X and Y (like random variables or stochastic processes), it seems natural to prefer X and Y to be independent. However, **coupling** is the idea that it can sometimes be beneficial to let X and Y actually be dependent on each other.

Proof of Theorem 11.1. Let (X_n) be our irreducible, aperiodic, positive recurrent Markov chain with transition matrix P and initial distribution λ. Let (Y_n) be a Markov chain also with transition matrix P but "in equilibrium" – that is, started from the stationary distribution π, and thus staying in that distribution for ever.

Pick a state $s \in \mathcal{S}$, and let T be the first time the $X_n = Y_n = s$ (or $T = \infty$, if that never happens). Now here's the coupling: after T, when (X_n) and (Y_n) collide at s, then make (X_n) stick to (Y_n), so $X_n = Y_n$ for $n \geq T$. Since a Markov chain has no memory, $(X_{T+n}) = (Y_{T+n})$ is still just a Markov chain with the same transition probabilities from that point on. (Readers of a previous optional subsection will recognise T as a stopping time and will notice we're using the strong Markov property.) Most importantly, thanks to the coupling, from the time T onwards, (X_n) will also always have distribution π, a fact that will obviously be very useful in this proof.

It will be important that T is finite with probability 1. Define $(\mathbf{Z}_n)$ by $\mathbf{Z}_n = (X_n, Y_n)$. So $(\mathbf{Z}_n)$ is a Markov chain on $\mathcal{S} \times \mathcal{S}$, and T is the expected hitting time of $(\mathbf{Z}_n)$ to the state $(s, s) \in \mathcal{S} \times \mathcal{S}$. The transition probabilities for $(\mathbf{Z}_n)$ are $\tilde{\mathsf{P}} = (\tilde{p}_{(i,k)(j,l)})$ where

$$\tilde{p}_{(i,k)(j,l)} = p_{ij} p_{kl}.$$

This is the probability that the joint chain goes from $\mathbf{Z}_n = (X_n, Y_n) = (i, k)$ to $\mathbf{Z}_{n+1} = (X_{n+1}, Y_{n+1}) = (j, l)$.

Since the original Markov chain is irreducible and aperiodic, this means that $p_{ij}(n), p_{kl}(n) > 0$ for all n sufficiently large, so $\tilde{p}_{(i,k)(j,l)}(n) > 0$ for all n sufficiently large also, meaning that $(\mathbf{Z}_n)$ is irreducible (and, although this isn't required, aperiodic). Further, $(\mathbf{Z}_n)$ has a stationary distribution $\tilde{} = (\tilde{\pi}_{(i,k)})$ where

$$\tilde{\pi}_{(i,k)} = \pi_i \pi_k,$$

which means that $(\mathbf{Z}_n)$ is positive recurrent. Thus T is finite with probability 1.

So we can finally prove the limit theorem. We want to show that $\mathbb{P}(X_n = i)$ tends to π_i. The difference between them is

$$\begin{aligned}
\big|\mathbb{P}(X_i = i) - \pi_i\big| &= \mathbb{P}(n \leq T) \times \big|\mathbb{P}(X_i = i \mid n \leq T) - \pi_i\big| + P(n > T) \times |\pi_i - \pi_i| \\
&= \mathbb{P}(n \leq T) \times \big|\mathbb{P}(X_i = i \mid n \leq T) \\
&\leq \mathbb{P}(n \leq T).
\end{aligned}$$

Here, the equality on the first line is because (X_n) follows the stationary distribution exactly once it sticks to (Y_n) after time T, and the inequality on the third line is because the absolute difference between two probabilities is between 0 and 1. But we've already shown that T is finite with probability 1, so $\mathbb{P}(n \leq T) = \mathbb{P}(T \geq n) \to 0$, and we're done.

$\square$

The proof of the ergodic theorem uses the law of large numbers. Recall that the law of large numbers states that if $Y_1, Y_2, \dots$ are IID random variables with mean μ, then

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{n} \to \mu \qquad \text{as } n \to \infty.$$

This means it's also true that for any sequence (a_n) with $a_n \to \infty$, we also have

$$\frac{Y_1 + Y_2 + \cdots + Y_{a_n}}{a_n} \to \mu \qquad \text{as } n \to \infty.$$

Proof of Theorem 11.3. If (X_n) is transient, then the number of visits to state i is finite with probability 1, so $V_i(n)/n \to 0$, as required.

Suppose instead that (X_n) is recurrent. By our useful lemma we know we will hit i in finite time, so we can ignore that negligible "burn-in" period, and (by the strong Markov property) assume we start from i. Let $M_i^{(r)}$ be the time between the rth and $(r+1)$th visits to i. Note that the $M_i^{(r)}$ are IID with mean μ_i.

The time of the last visit to i before time n is

$$M_i^{(1)} + M_i^{(2)} + \cdots + M_i^{(V_i(n)-1)} < n,$$

and the time of the first visit to i after time n is

$$M_i^{(1)} + M_i^{(2)} + \cdots + M_i^{(V_i(n))} \geq n.$$

Hence

$$\frac{M_i^{(1)} + M_i^{(2)} + \cdots + M_i^{(V_i(n)-1)}}{V_i(n)} < \frac{n}{V_i(n)} \leq \frac{M_i^{(1)} + M_i^{(2)} + \cdots + M_i^{(V_i(n))}}{V_i(n)}. \tag{12}$$

Because (X_n) is recurrent, we keep returning to i, so $V_i(n) \to \infty$ with probability 1. Hence, by the law of large numbers, both the left- and right-hand sides of (12) tend to $\mathbb{E}M_i^{(r)} = \mu_i$. So $n/V_i(n)$ is sandwiched between them, and tends to μ_i too. Finally $n/V_i(n) \to \mu_i$ is equivalent to $V_i(n)/n \to 1/\mu_i$, so we are done.

$\square$

This completes the material on discrete time Markov chains. **In the next section**, we recap what we have learned, and have a little time for some revision.

12 End of of Part I: Discrete time Markov chains

- No new material in this section, but a half-week break to catch up and take stock on what we've learned

12.1 Things to do

We've now finished the material of the Part 1 of the module, on discrete time Markov chains. So this is a good time to take stock, revise what we've learned, and make sure we're completely up to date before starting Part 2 of the module on continuous time processes.

Some things you may want to do in lieu of reading a section of notes:

- Make sure you've completed **Problem Sheets 1 to 6**, and go back to any questions that stumped you before.
- Start working on **Computational Worksheet 2** (which doubles as **Assessment 2**). There are optional computer drop-in sessions in Week 7, and the work is due on Thursday 18 March 1400 (week 8).
- Start working on **Assessment 3** which is due on Thursday 25 March 1400 (week 9).
- Re-read any sections of notes you struggled with earlier.
- Take the opportunity to look at the optional nonexaminable subsections, if you opted out the first time around. (Section 9, Section 10, Section 11)
- Let me know if there's anything from this part of the course you'd like me to go through in next week's lecture.

12.2 Summary of Part I

The following list is not exhaustive, but if you can do most of the things on this list, you're well placed for the exam.

- Define the simple random walk and other random walks.
- Perform elementary calculations for the simple random walk, including by referring to the exact binomial distribution.
- Calculate the expectation and variance of general random walks.
- Define the gambler's ruin Markov chain.
- Find the ruin probability and expected duration for the gambler's ruin by (i) setting up equations by conditioning on the first step and (ii) solving the resulting linear difference equation.
- Draw a transition diagram of a Markov chain given the transition matrix.
- Calculate n-step transitions probabilities by (a) summing the probabilities over all relevant paths or (b) calculating the matrix power.
- Find the communicating classes in a Markov chain.
- Calculate the period of communicating class.
- Calculate hitting probabilities and expected hitting times by (i) setting up equations by conditioning on the first step and (ii) solving the resulting simultaneous equations.
- Define positive recurrence, null recurrence and transience, and explain their properties.
- Find the positive recurrence, null recurrence or transience of communicating classes.
- Find the stationary distribution of Markov chain.
- Give conditions for a stationary distribution to exist and be unique.
- Give conditions for convergence to an equilibrium distribution.
- Calculate long-term proportions of time using the ergodic theorem.

In the next section, we begin our study of continuous time Markov processes by looking at the most important example: the Poisson process.

Problem sheet 6

You should attempt all these questions and write up your solutions in advance of your workshop in week 7 (Monday 8 or Tuesday 9 March) where the answers will be discussed.

Remember that the **mid-semester survey** is still open.

1. Consider a Markov chain (X_n) with state space $\mathcal{S} = \{1, 2, 3\}$ and transition matrix

$$\mathsf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

(a) Draw a transition diagram for this Markov chain. Is it irreducible? Is each state periodic or aperiodic?

(b) What is m_i, the return probability, for each state? What is μ_i, the expected return time for each state.

(c) By solving $\pi = \pi\mathsf{P}$, find the stationary distribution. Use this to confirm the values of μ_i.

(d) For what initial distributions λ do the limits $\lim_{n \to \infty} \mathbb{P}(X_n = i)$ exist?

(e) What is the long-run proportion of time spent in each state?

2. Every person has two chromosomes; each chromosome is a copy of a chromosome from one of the person's parents. There are two types of chromosome, which are conventionally labelled X and Y. A child born with a Y chromosome is male, while a child with two X chromosomes is female.

Haemophilia is a blood-clotting disorder caused by a defective X chromosome (we will label this as X^*). Females with the defective chromosome ($\mathrm{X}^*\mathrm{X}$) will not typically show symptoms of the disease but can pass it on to children – they are "carriers". Males with the defective chromosome ($\mathrm{X}^*\mathrm{Y}$) have the disease and its symptoms.

A medical statistician is studying the progress of the disease through first-born children, starting with a female carrier. The statistician makes the following assumptions: First, each parent has an equal probability of passing either of their chromosomes to their children. Second, the partner of each person in the study does not have a defective X chromosome. Third, no new genetic disorders occur.

(a) Show that we can use a Markov chain to model the progress of the disease under the above assumptions. What is the state space? Draw a transition diagram.

(b) What are the communicating classes in the chain? Is each class positive recurrent, null recurrent, or transient?

(c) Calculate a stationary distribution. Is this the only stationary distribution?

(d) Under this model, what is the limiting probability that, in many generations' time, a child has haemophilia?

3. An airline operates a frequent flyer scheme with four classes of membership; Ordinary, Bronze, Silver and Gold. Scheme members get benefits according to their membership class. Changing membership class operates as follows:

- If a member books two or more flights in a given year, they are moved up a class of membership for the next year (or remain at Gold).
- If a member books a single flight, they remain in their current class in the following year.
- If a member books no flights, they move down a class (or remain at Ordinary).

The airline's research has shown that in a given year 40% of members book no flights, 40% book exactly one flight and the remaining 20% book two or more flights, independent of their history. Moreover, the cost of running the scheme per member is estimated as £0 for Ordinary members, £10 for Bronze members, £20 for Silver members, and £30 for Gold members.

(a) Show that this system can be modelled using a Markov chain. Write down the transition probabilities and draw a transition diagram.

(b) Explain why a unique stationary distribution exists and calculate it.

(c) The airline makes a profit of £10 per passenger per flight, before the cost of the frequent flyer scheme. In the long run, does the airline expect to remain in profit after the cost of the scheme?

4. We have N balls, each of which is placed into one of two urns. At each time step, a ball is chosen uniformly at random and moved to the other urn.

(a) Show that the stationary probability the first urn contains i balls is

$$\frac{1}{2^N}\binom{N}{i}.$$

(b) Is this an equilibrium distribution?

(c) In the long run, for what proportion of time are all of the balls in the same urn?

Assessment 3

This assessment counts as 4% of your final module grade. You should attempt both questions. You must show your working, and there are marks for the quality of your mathematical writing.

The deadline for submission is **Thursday 25 March at 2pm**. Submission will be to Gradescope via Minerva, from Monday 21 March. It would be helpful to start your solution to Question 2 on a new page. If you hand-write your solutions and scan them using your phone, please convert to PDF using a scanning app (I like Microsoft Office Lens or Adobe Scan) rather than submit images.

Late submissions up to Thursday 1 April at 2pm will still be marked, but the total mark will be reduced by 10% per day or part-day for which the work is late. Submissions are not permitted after Thursday 1 April.

Your solutions to this assessment should be your own work. Copying, collaboration or plagiarism are not permitted. Asking others to do your work, including via the internet, is not permitted. Transgressions are considered to be a very serious matter, and will be dealt with according to the University's disciplinary procedures.

1. A firm rents out cars and operates from three locations – the Airport, the Beach and the City. Customers may return vehicles to any of the three locations. The company estimates that the probability of a car being returned to each location is as follows:

Car hired from	Car returned to		
	Airport	Beach	City
Airport	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
Beach	$\frac{1}{4}$	$\frac{3}{4}$	0
City	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Figure 13: Table of car-hire data.

(a) A car is currently parked at the beach. What is the probability that, after being hired twice, it ends up at the airport? [2 marks]

(b) A car is parked in the city. On average, how many times will it need to be hired until it is left at the beach? [2]

(c) The firm has been running for many years with a fleet of 25 cars, of which 7 are currently on hire. How many of the other cars would you expect to be parked at each of the three locations? Explain your answer clearly. [4]

2. Consider a Markov chain (X_n) with state space $\mathcal{S} = \{1, 2, 3, 4, 5\}$ and transition matrix

$$
\mathsf{P} = \begin{pmatrix}
0.2 & 0.8 & 0 & 0 & 0 \\
0.4 & 0.6 & 0 & 0 & 0 \\
0.2 & 0.1 & 0.2 & 0.5 & 0 \\
0 & 0 & 0 & 0.4 & 0.6 \\
0 & 0 & 0 & 0.8 & 0.2
\end{pmatrix}.
$$

(a) What are the communicating classes in the Markov chain? State, giving brief reasons, whether each class is positive recurrent, null recurrent, or transient. [3]

(b) Calculate the hitting probability h_{34}. [2]

(c) Find two different stationary distributions for this Markov chain. [3]

(d) What is the limit as $n \to \infty$ of the n-step transition matrix $\mathsf{P}(n)$? Explain your answer. [4]

Part II: Continuous time Markov jump processes

13 Poisson process with Poisson increments

- Reminder: the Poisson distribution
- The Poisson process has independent Poisson increments
- Summed and marked Poisson processes

13.1 Poisson distribution

In the next three sections we'll be considering the Poisson process, a continuous time discrete space process with the Markov property. Given its name, it's not surprising to hear the Poisson *process* is related to the Poisson *distribution*. Let's start with a reminder of that.

Recall that a discrete random variable X has a **Poisson distribution** with rate λ, written $X \sim \mathrm{Po}(\lambda)$, if its probability mass function is

$$\mathbb{P}(X = n) = \mathrm{e}^{-\lambda}\frac{\lambda^n}{n!} \qquad n = 0, 1, 2, \dots.$$

The Poisson distribution is often used to model the number of "arrivals" in a fixed amount of time – for example, the number of calls to a call centre in one hour, the number of claims to an insurance company in one year, or the number of particles decaying from a large amount of radioactive material in one second.

The Poisson distribution is named after the French mathematician and physicist Siméon Denis Poisson, who studied it in 1837, although it was used by another French mathematician, Abraham de Moivre, more than 100 years earlier.

Recall the following facts about of the Poisson distribution:

1. Its expectation is $\mathbb{E}X = \lambda$ and the variance is $\mathrm{Var}(X) = \lambda$.
2. If $X \sim \mathrm{Po}(\lambda)$ and $Y \sim \mathrm{Po}(\mu)$ are independent, then $X + Y \sim \mathrm{Po}(\lambda + \mu)$.
3. Let $X \sim \mathrm{Po}(\lambda)$ represent some arrivals, and independently "mark" each arrival with probability p. Then the number of marked arrivals Y has distribution $Y = \mathrm{Po}(p\lambda)$, the number of unmarked arrivals Z has distribution $Z \sim \mathrm{Po}((1-p)\lambda)$, and Y and Z are independent.

(I say "recall", but it's OK if the third one is new to you.)

13.2 Definition 1: Poisson increments

Suppose that, instead of just modelling the number of arrivals in a *fixed* amount of time, we want to continually model the total number of arrivals as it changes over time. This will be a stochastic process with discrete state space $\mathcal{S} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ and continuous time $\mathbb{R}_+ = [0, \infty)$. In continuous time, we will normally write stochastic processes as $(X(t))$, with the time variable being a t in brackets, rather than a subscript n as we had in discrete time.

Suppose calls arrive at a call centre at a rate of $\lambda = 100$ an hour. The following assumptions seem reasonable:

- We begin counting with $X(0) = 0$ calls.
- The number of calls in the first hour is $X(1) \sim \mathrm{Po}(100)$. The number of calls in the second hour $X(2) - X(1)$ will also be $\mathrm{Po}(100)$, and will be independent of the number of calls in the first hour.

- The number of calls in a two hour period will be $X(t+2) - X(t) \sim \text{Po}(200)$, while the number of calls in a half-hour period will be $X(t + \frac{1}{2}) - X(t) \sim \text{Po}(50)$.

These properties will define the Poisson process.

Definition 13.1. The **Poisson process** with rate λ is defined as followed. It is a stochastic process $(X(t))$ with continuous time $t \in [0, \infty)$ and discrete state space $\mathcal{S} = \mathbb{Z}_+$ with the following properties:

1. $X(0) = 0$;
2. Poisson increments: $X(t+s) - X(t) \sim \text{Po}(\lambda s)$ for all $s, t > 0$;
3. independent increments: $X(t_2) - X(t_1)$ and $X(t_4) - X(t_3)$ are independent for all $t_1 \leq t_2 \leq t_3 \leq t_4$.

Note that the condition $t_1 \leq t_2 \leq t_3 \leq t_4$ means that the time interval from t_1 to t_2 and the time interval from t_3 to t_4 don't overlap. (Overlapping time intervals will not have independent increments, as arrivals in the overlap will count for both.)

The Poisson process was discovered in the first decade 20th century, and the process was named after the distribution. Many people were working on similar things, so it's difficult to say who discovered it first, but important work was done by the Swedish actuary and mathematician Filip Lundberg and the Danish engineer and mathematician AK Erlang.

Example 13.1. *Claims arrive at insurance company at a rate of $\lambda = 8$ per hour, modelled as a Poisson process. What is the probability there are no claims in a given 15 minute period?*

By property 2, the number of claims in 15 minutes is a Poisson distribution with mean $\frac{15}{60}\lambda = 2$. The probability there are no claims is

$$\mathrm{e}^{-2}\frac{2^0}{0!} = \mathrm{e}^{-2} = 0.135.$$

Example 13.2. *A professor receives visitors to her office at a rate of $\lambda = 2.5$ per day, modelled as a Poisson process. What is the probability she gets at least one visitor every day this (5-day) week?*

The probability she gets at least one visitor on any given day is

$$1 - \mathrm{e}^{-2.5}\frac{2.5^0}{0!} = 1 - \mathrm{e}^{-2.5} = 0.918.$$

By property 3, the numbers of visitors on different days are independent, so the probability of getting at least one visitor each day this week is $0.918^5 = 0.652$.

13.3 Summed and marked Poisson processes

The following theorem shows that the sum of two Poisson processes is itself a Poisson process.

Theorem 13.1. *Let $(X(t))$ and $(Y(t))$ be independent Poisson processes with rates λ and μ respectively. Then the process $(Z(t))$ given by $Z(t) = X(t) + Y(t)$ is a Poisson process with rate $\lambda + \mu$.*

Proof. The proof of this is a question on Problem Sheet 7.

$\square$

Example 13.3. *A student receives email to her university mail address at a rate of $\lambda = 4$ emails per hour, and to her personal email address at a rate of $\mu = 2$ per hour. Using a Poisson process model, what is the probability the student receives 3 or fewer emails in a 30 minute period?*

The total number of emails is a sum of Poisson processes with rate $\lambda + \mu = 6$. The total number of emails received in half an hour is Poisson with rate $(\lambda + \mu)/2 = 3$. Thus the probability that 3 or fewer emails are received is

$$\mathrm{e}^{-3}\frac{3^0}{0!} + \mathrm{e}^{-3}\frac{3^1}{1!} + \mathrm{e}^{-3}\frac{3^2}{2!} + \mathrm{e}^{-3}\frac{3^3}{3!} = 13\mathrm{e}^{-3} = 0.647.$$

We also have the **marked Poisson process**, which can be thought of as the opposite to the summed process: the summed process combines two processes together, while the marked process splits one process into two.

Theorem 13.2. *Let $(X(t))$ be a Poisson process with rate λ. Each arrival is independently marked with probability p. Then the marked process $(Y(t))$ is a Poisson process with rate $p\lambda$, the unmarked process $(Z(t))$ is a Poisson process with rate $(1-p)\lambda$, and $(Y(t))$ and $(Z(t))$ are independent.*

Proof. Given the third fact that you were "reminded" of, it's easy to check the necessary properties.

$\square$

Example 13.4. *In the 2019/20 English Premier League football season, an average of $\lambda = 2.72$ goals were scored per game, with a proportion $p = 0.56$ of them scored by the home team. If we model this as a Poisson process, what is the probability a match ends in a 1–1 draw?*

The number of home goals scored is Poisson with rate $p\lambda = 0.56 \times 2.72 = 1.52$, and the number of away goals scored is Poisson with rate $(1-p)\lambda = 1.20$. Under the Poisson process assumption, these are independent, so the probability the home and the away team both score 1 goal is

$$\mathrm{e}^{-1.52}\frac{1.52^1}{1!} \times \mathrm{e}^{-1.20}\frac{1.20^1}{1!} = 1.52\mathrm{e}^{-1.52} \times 1.20\mathrm{e}^{-1.20} = 0.12,$$

or 12%.

In the next section, we look at the Poisson process in a different way: the times in between arrivals have an exponential distribution.

14 Poisson process with exponential holding times

- Reminder: the exponential distribution
- The Poisson process has exponential holding times
- The Markov property in continuous time

14.1 Exponential distribution

Last time, we introduced the Poisson process by looking at the random number of arrivals in fixed amount of time, which follows a Poisson distribution. Another way of looking at the Poisson process is to look at the random amount of of time required for a fixed number of arrivals. For this, we'll need the exponential distribution.

We start by recalling the exponential distribution. Recall that we say a continuous random variable T has the **exponential distribution** with rate λ, and write $T \sim \mathrm{Exp}(\lambda)$, if it has the probability distribution function $f(t) = \lambda \mathrm{e}^{-\lambda t}$ for $t \geq 0$.

Exponential distributions are often used to model "waiting times" – for example, the amount of time until a light bulb breaks, the times between buses arriving, or the time between withdrawals from a bank account.

You are reminded of the following facts about the exponential distribution:

- The cumulative distribution function is $F(t) = \mathbb{P}(T \leq t) = 1 - \mathrm{e}^{-\lambda t}$; although it is usually more convenient to deal with the tail probability $\mathbb{P}(T > t) = 1 - F(t) = \mathrm{e}^{-\lambda t}$.
- The expectation is $\mathbb{E}T = 1/\lambda$ and the variance is $\mathrm{Var}(T) = 1/\lambda^2$.

The following memoryless property will of course be important in a course about memoryless Markov processes.

Theorem 14.1. *Let $T \sim \mathrm{Exp}(\lambda)$. Then, for any $s, t \geq 0$,*

$$\mathbb{P}(T > t + s \mid T > t) = \mathbb{P}(T > s).$$

Suppose we are waiting an exponentially distributed time for an alarm to go off. No matter how long we've been waiting for, the remaining time to wait is still exponentially distributed, with the same parameter – hence "memoryless".

Proof. By standard use of conditional probability, we have

$$\mathbb{P}(T > t + s \mid T > t) = \frac{\mathbb{P}(T > t + s \text{ and } T > t)}{\mathbb{P}(T > t)} = \frac{\mathbb{P}(T > t + s)}{\mathbb{P}(T > t)} = \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda s},$$

which is still the tail probability of the exponential distribution.

$\square$

The following property will be important later on in the course.

Theorem 14.2. *Let $T_1 \sim \mathrm{Exp}(\lambda_1)$, $T_2 \sim \mathrm{Exp}(\lambda_2)$, ..., $T_n \sim \mathrm{Exp}(\lambda_n)$ be independent exponential distributions, and let T be the minimum of the T_is, so $T = \min\{T_1, T_2, ..., T_n\}$. Then*

$$T \sim \mathrm{Exp}(\lambda_1 + \lambda_2 + \cdots + \lambda_n).$$

Further, the probability that T_j is the smallest of all the T_is is

$$\mathbb{P}(T = T_j) = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}.$$

Proof. The proof of this is an question on Problem Sheet 7.

$\square$

14.2 Definition 2: exponential holding times

We mentioned before that exponential distributions are often used to model "waiting times". When modelling a process $(X(t))$ counting many arrivals at rate λ, we might model the process like this: after waiting an $\mathrm{Exp}(\lambda)$ amount of time, an arrival appears. After another $\mathrm{Exp}(\lambda)$ amount of time, another arrival appears. And so on. We often use the term "holding time" to refer to the time between consecutive arrivals.

This suggests a process with the following properties:

- We start with $X(0) = 0$.
- Let $T_1, T_2, \cdots \sim \mathrm{Exp}(\lambda)$ be the holding times, all independent. Then

$$X(t) = \begin{cases} 0 & \text{for } 0 \le t < T_1 \\ 1 & \text{for } T_1 \le t < T_1 + T_2 \\ 2 & \text{for } T_1 + T_2 \le t < T_1 + T_2 + T_3 \\ \text{and so on.} \end{cases}$$

A process described like this is also the Poisson process!
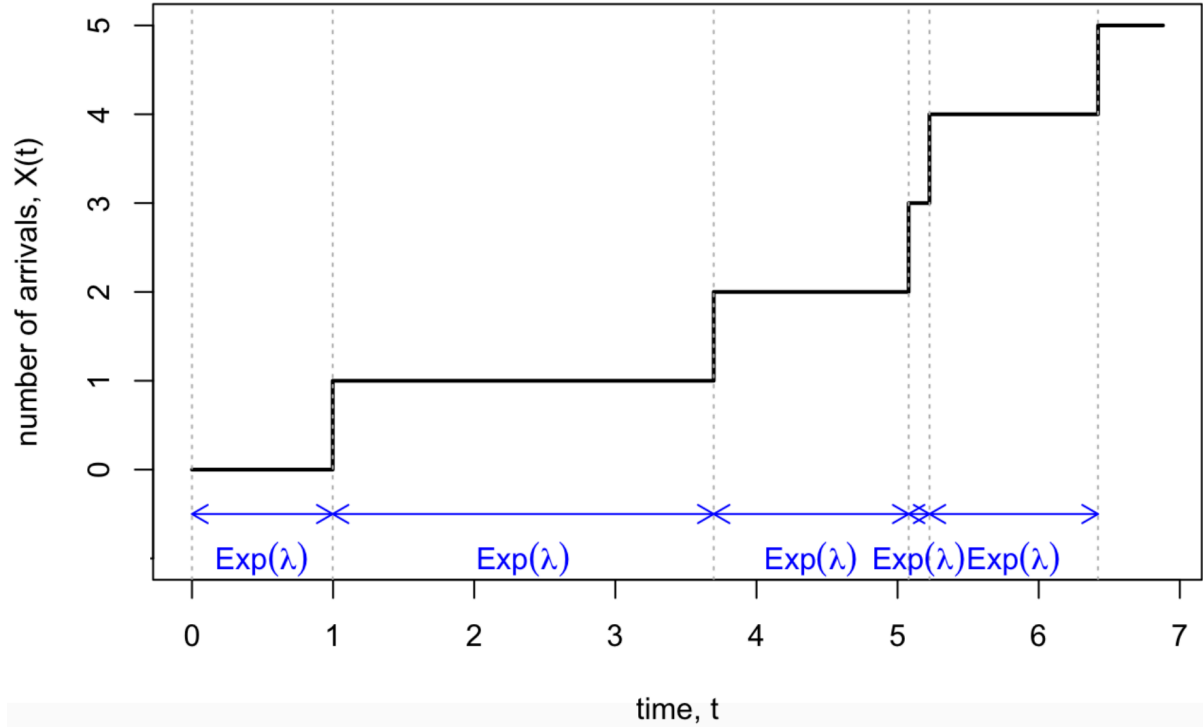
Figure 14: A Poisson process with exponentially distributed holding times.

Theorem 14.3. *Let $(X(t))$ be a Poisson process with rate λ, as defined by its independent Poisson increments (see Subsection 13.2). Then $(X(t))$ has the exponential holding times structure described above.*

Proof. Let $(X(t))$ be a Poisson process with rate λ. We seek the distribution of the first arrival time T_1. We have

$$\mathbb{P}(T_1 > t_1) = \mathbb{P}(X(t_1) - X(0) = 0) = \mathrm{e}^{-\lambda t_1} \frac{(\lambda t_1)^0}{0!} = \mathrm{e}^{-\lambda t_1},$$

since the arrival comes after t_1 if there are no arrivals in the interval $[0, t_1]$. But this is precisely the tail probability of an exponential distribution with rate λ.

Now consider the second holding time. We have

$$\mathbb{P}(T_2 > t_2 \mid T_1 = t_1) = \mathbb{P}(X(t_1 + t_2) - X(t_1) = 0) = \mathrm{e}^{-\lambda((t_1 + t_2) - t_1)} = \mathrm{e}^{-\lambda t_2}$$

by the same argument as before. This is the tail distribution for another $\mathrm{Exp}(\lambda)$ distribution, independent of T_1.

Repeating this argument for all n gives the desired exponential holding time structure.

$\square$

Example 14.1. *Customers visit a second-hand bookshop at a rate of $\lambda = 5$ per hour. Each customer a book with probability $p = 0.4$. What is the expected time to make ten sales, and what is the standard deviation?*

The count of books sold marked Poisson process, which we saw last time is itself a Poisson process with rate $p\lambda = 2$.

The expected time of the tenth sale is

$$\mathbb{E}(T_1 + \cdots + T_{10}) = 10 \, \mathbb{E}T_1 = 10 \times \tfrac{1}{2} = 5 \text{ hours.}$$

The variance is

$$\mathrm{Var}(T_1 + \cdots + T_{10}) = 10 \, \mathrm{Var}(T_1) = 10 \times \frac{1}{2^2} = 2.5,$$

where we used that the holding times are independent and identically distributed, so the standard deviation is $\sqrt{2.5} = 1.58$ hours.

What is the probability it takes more than an hour to sell the first book?

This is the probability the first holding time is longer than an hour, which is the tail of the exponentital distribution,

$$\mathbb{P}(T_1 > 1) = \mathrm{e}^{-2 \times 1} = \mathrm{e}^{-2} = 0.135.$$

An alternative way to solve this would be to say that the first book is is sold later than 1 hour in if $X(1) = 0$. So using the Poisson increments definition from last time, we have

$$\mathbb{P}(X(1) = 0) = \mathrm{e}^{-2 \times 1} \frac{(2 \times 1)^0}{0!} = \mathrm{e}^{-2} = 0.135.$$

We get the same answer either way.

14.3 Markov property in continuous time

We previously saw the Markov "memoryless" property in discrete time. The equivalent definition in continuous time is the following.

Definition 14.1. Let $(X(t))$ be a stochastic process on a discrete state space $\mathcal{S}$ and continuous time $t \in [0, \infty)$. We say that $(X(t))$ has the **Markov property** if

$$\mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, \dots, X(t_1) = x_1, X(t_0) = x_0\big)$$
$$= \mathbb{P}(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n)$$

for all times $t_0 < t_1 < \cdots < t_n < t_{n+1}$ and all states $x_0, x_1, \dots, x_n, x_{n+1} \in \mathcal{S}$.

In other words, the state of the process at some point t_{n+1} in the future depends on where we are now t_n, but, given that, does not depend on any collection of previous times $t_0, t_1, \dots, t_{n-1}$.

The Poisson process does indeed have the Markov property, since, by the property of independent increments, we have that

$$\mathbb{P}(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, \dots, X(t_0) = x_0) = \mathbb{P}(X(t_{n+1}) - X(t_n) = x),$$

where $x = x_{n+1} - x_n$. By Poisson increments, this has probability

$$e^{-\lambda(t_{n+1}-t_n)} \frac{\left(\lambda(t_{n+1}-t_n)\right)^x}{x!}.$$

Alternatively, by the memoryless property of the exponential distribution, we see that we can restart the holding time from t_n and still have this and all future holding times exponentially distributed.

In the next section, we look at the Poisson process in a third way, by looking at what happens in a very small time period.

Problem sheet 7

You should attempt all these questions and write up your solutions in advance of your workshop in week 8 (Monday 15 or Tuesday 16 March) where the answers will be discussed.

1. Let $(X(t))$ be a Poisson process with rate $\lambda = 5$. Calculate:

(a) $\mathbb{P}(X(0.4) \le 2)$;

(b) $\mathbb{E}X(6.4)$;

(c) $\mathbb{P}(X(0.5) = 0$ and $X(1) = 1)$.

Let T_n be the nth holding time, and let $J_n = T_1 + \cdots + T_n$ be the nth arrival time. Calculate:

(d) $\mathbb{P}(0.1 \le T_2 < 0.3)$;

(e) $\mathbb{E}J_{100}$;

(f) $\mathrm{Var}(J_{100})$.

(g) Using a normal approximation, approximate $\mathbb{P}(18 \le J_{100} \le 22)$.

2. Suppose that telephone calls arrive at a call centre according to a Poisson process with rate $\lambda = 100$ per hour, and are answered with probability 0.6.

(a) What is the probability that there are no answered calls in the next minute?

(b) Use a suitable normal approximation, with a continuity correction, to find the probability that there will be at least 25 answered calls in the next 30 minutes.

3.

(a) Let $X \sim \mathrm{Po}(\lambda)$ and $Y \sim \mathrm{Po}(\mu)$ be two independent Poisson distributions. Show that $X + Y \sim \mathrm{Po}(\lambda + \mu)$. One way to start would be to write

$$\mathbb{P}(X + Y = z) = \sum_{x=0}^{z} \mathbb{P}(X = x)\,\mathbb{P}(Y = z - x).$$

(b) Let $(X(t))$ and $(Y(t))$ be independent Poisson processes with rate λ and μ respectively. Use part (a) to show that $(X(t) + Y(t))$ is a Poisson process with rate $\lambda + \mu$.

(c) Number 1 buses arrive at a bus stop at a rate of $\lambda_1 = 4$ per hour, and Number 6 buses arrive at the rate $\lambda_6 = 2$ per hour. I've been waiting at the bus stop for 5 minutes for either bus to arrive; how much longer do I have to wait, on average?

4. Let $T_1 \sim \mathrm{Exp}(\lambda_1), T_2 \sim \mathrm{Exp}(\lambda_2), \ldots, T_n \sim \mathrm{Exp}(\lambda_n)$ be independent exponential distributions, and let T be the minimum $T = \min\{T_1, T_2 \ldots, T_n\}$.

(a) Show that $T \sim \mathrm{Exp}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$. (You may use the fact that

$$\mathbb{P}(T > t) = \mathbb{P}(T_1 > t)\,\mathbb{P}(T_2 > t) \cdots \mathbb{P}(T_n > t),$$

provided you explain why it's true.)

(b) Show that the probability that the minimum is T_j is given by

$$\mathbb{P}(T_j = T) = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}.$$

(You could choose to begin by proving the $n = 2$ case, if you want.)

5. Let $(X(t))$ be a Poisson process with rate λ. Conditional on there being exactly 1 arrival before time t, find the distribution of the time of that arrival.

15 Poisson process in infinitesimal time periods

- Poisson process in terms of increments in infinitesimal time
- Forward equations for the Poisson process

15.1 Definition 3: increments in infinitesimal time

We have seen two definitions of the Poisson process so far: one in terms of increments having a Poisson distribution, and one in therms of holding times having an exponential distribution. Here we will see another definition, by looking at what happens in a very small time period of length τ. Advantages of this approach include that does not require direct assumptions on the distributions involved, so may seem more "natural" or "inevitable". It also shows links between Markov processes and differential equations. A disadvantage is that it's normally easier to do calculations with the Poisson or exponential definitions.

Let $(X(t))$ be a stochastic processes counting some arrivals over time. Consider the number of arrivals in a very small time period of length τ; that number of arrivals is $X(t+\tau) - X(t)$. The following (seemingly) weak assumptions seem justified:

- It's quite likely that there will be no arrivals in the very small time period.
- It is possible there will be one arrival, and that probability will be roughly proportional to the length τ of the time period.
- It's extremely unlikely there will be two or more arrivals in such a short time period.

To write this down formally in maths, we will consider $\mathbb{P}(X(t + \tau) - X(t) = j)$ as the length τ of the time period tends to zero.

It will be helpful for us to use "little-o" notation. Here $o(\tau)$ means a term that is of "lower order" compared to τ. These are terms like τ^2 or τ^3 that are so tiny that we're safe to ignore them. Formally, $o(\tau)$ stands for a function of τ that tends to 0 as $\tau \to 0$ even when divided by τ. That is,

$$f(\tau) = o(\tau) \qquad \Longleftrightarrow \qquad \frac{f(\tau)}{\tau} \to 0 \quad \text{as } \tau \to 0.$$

Then our suggestions above translate to

$$\mathbb{P}(X(t + \tau) - X(t) = j) = \begin{cases} 1 - \lambda\tau + o(\tau) & \text{if } j = 0, \\ \lambda\tau + o(\tau) & \text{if } j = 1, \\ o(\tau) & \text{if } j \geq 2. \end{cases} \tag{13}$$

as $\tau \to 0$.

It turns out that we have once again defined the Poisson process with rate λ.

Theorem 15.1. *Let $(X(t))$ be a stochastic process with the following properties:*

1. *$X(0) = 0$;*
2. *infinitesimal increments: $X(t + \tau) - X(t)$ has the structure in (13) above, as $\tau \to 0$;*
3. *independent increments: $X(t_2) - X(t_1)$ and $X(t_4) - X(t_3)$ are independent for all $t_1 \leq t_2 \leq t_3 \leq t_4$.*

Then $(X(t))$ is a Poisson process with rate λ.

We will prove this in a moment. First let's see an example of its use.

15.2 Example: sum of two Poisson processes

We mentioned before the sum of two Poisson processes: that if $(X(t))$ is a Poisson process with rate λ and $(Y(t))$ is an independent Poisson process with rate μ, then the sum $(Z(t))$ where $Z(t) = X(t) + Y(t)$ is a Poisson process with rate $\lambda + \mu$.

On Problem Sheet 7, Question 3, you proved this from the Poisson increments definition. But it's perhaps even easier to prove using the infinitesimals definition.

We have

$$\begin{aligned}
\mathbb{P}(Z(t+\tau) - Z(t) = 0) &= \mathbb{P}(X(t+\tau) - X(t) = 0)\,\mathbb{P}(Y(t+\tau) - Y(t) = 0) \\
&= (1 - \lambda\tau + o(\tau))(1 - \mu\tau + o(\tau)) \\
&= 1 - (\lambda + \mu)\tau + o(\tau),
\end{aligned}$$

since Z does not increase provided neither X nor Y increases.

We also have

$$\begin{aligned}
\mathbb{P}(Z(t+\tau) &- Z(t) = 1) \\
&= \mathbb{P}(X(t+\tau) - X(t) = 1)\,\mathbb{P}(Y(t+\tau) - Y(t) = 0) \\
&\quad + \mathbb{P}(X(t+\tau) - X(t) = 0)\,\mathbb{P}(Y(t+\tau) - Y(t) = 1) \\
&= (\lambda\tau + o(\tau))(1 - \mu\tau + o(\tau)) + (1 - \lambda\tau + o(\tau))(\mu\tau + o(\tau)) \\
&= (\lambda + \mu)\tau + o(\tau),
\end{aligned}$$

since Z increases by 1 if either X increases by 1 and Y stays fixed or vice versa. We used here that the probability that *both* X and Y increase is of order $\tau^2 = o(\tau)$.

Finally, since probabilities must add up to 1, we must have

$$\mathbb{P}(Z(t+\tau) - Z(t) \geq 2) = o(\tau).$$

But what we have written down is precisely the infinitesimals definition of a Poisson process with rate $\lambda + \mu$, so we are done.

15.3 Forward equations and proof of equivalence

If we compare Theorem 15.1 to the definition of the Poisson process in terms of Poisson increments, we see that properties 1 and 3 are the same. We only need to check property 2, that $X(t+s) - X(s)$ is a Poisson distribution with rate λt.

Since the increments as defined in (13) are time homogeneous, it will suffice to consider $s = 0$. So we need to show that $X(t) \sim \text{Po}(\lambda t)$.

Write $p_j(t) = \mathbb{P}(X(t) = j)$. Then for $j \geq 1$ we have

$$p_j(t + \tau) = (1 - \lambda\tau + o(\tau))p_j(t) + (\lambda\tau + o(\tau))p_{j-1}(t) + o(\tau),$$

since we get to j either by staying at j or by moving up from $j - 1$, with all other highly unlikely possibilities being absorbed into the $o(\tau)$. In order to deal with increments, it will be convenient to take a $p_j(t)$ to the left-hand side and rearrange, to get the increment

$$p_j(t + \tau) - p_j(t) = -\lambda\tau p_j(t) + \lambda\tau p_{j-1}(\tau) + o(\tau),$$

where "constant times $o(\tau)$" is itself $o(\tau)$.

The only way to deal with the $o(\tau)$ term, given its definition, is to divide everything by τ and send $\tau \to 0$. Dividing by τ gives

$$\frac{p_j(t + \tau) - p_j(t)}{\tau} = -\lambda p_j(t) + \lambda p_{j-1}(t) + \frac{o(\tau)}{\tau}.$$

Sending τ to 0, the term $o(\tau)/\tau$ tends to 0 and vanishes, while we recognise the limit of the left-hand side as being the derivative $\mathrm{d}p_j(t)/\mathrm{d}t = p_j'(t)$. This leaves us with the differential equation

$$p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1}(t).$$

We also have the initial condition $p_j(0) = 0$ for $j \geq 1$, since we start at 0, not at any $j \geq 1$.

We have to deal with the case $j = 0$ separately. This is similar but a bit easier. We have

$$p_0(t + \tau) = (1 - \lambda\tau + o(\tau))p_0(t),$$

since we must previously been at 0, then seen no arrivals. After rearranging to get an increment and sending $\tau \to 0$, as before, we get the differential equation

$$p_0'(t) = -\lambda p_0(t)$$

with the initial condition $p_0(0) = 1$, since we always start at 0.

In summary, we have derived the following equations:

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) & p_0(0) &= 1, \\ p_j'(t) &= -\lambda p_j(t) + \lambda p_{j-1}(t) & p_j(0) &= 0 \quad \text{for } j \geq 1. \end{aligned}$$

These are called the **Kolmogorov forward equations**.

We claim that the solution to these equations is the Poisson distribution

$$p_j(t) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!}.$$

This would prove the theorem.

We must check that the claim holds. For $j = 0$, the Poisson probability is

$$p_0(t) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^0}{0!} = \mathrm{e}^{-\lambda t}.$$

This has

$$p_0'(t) = -\lambda \mathrm{e}^{-\lambda t} = -\lambda p_0(t),$$

as desired, and we also have the correct initial condition $p_0(0) = \mathrm{e}^0 = 1$.

For $j \geq 1$, the left-hand side is

$$\begin{aligned} p_j'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!} \\ &= \frac{1}{j!}\left(-\lambda \mathrm{e}^{-\lambda t}(\lambda t)^j + \mathrm{e}^{-\lambda t}\lambda j(\lambda t)^{j-1}\right) \\ &= -\lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!} + \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-1}}{(j-1)!} \\ &= -\lambda p_j(t) + \lambda p_{j-1}(t) \end{aligned}$$

as desired. We also have the initial condition $p_j(0) = \mathrm{e}^0 0^j/j! = 0$. The claim is proven, and we are (finally) done.

In the next section, we two new types of process that are similar to the Poisson process.

16 Counting processes

- Birth processes, including the simple birth process
- Time inhomogeneous Poisson processes

16.1 Birth processes

The Poisson process was made easier to understand because it was a **counting process**. That is, because it was counting arrivals, we always knew that the next change would be an increase by 1; the only question was when that increase would happen.

In this section, we look at two other types of counting process; the only transitions will still only ever be from i to $i + 1$, but the holding times will not just be IID exponential distributions any more.

We start with the **simple birth process**, which can model, for example, the division of cells in a biological experiment. We suppose that each individual has an offspring (eg the cell divides) after an $\text{Exp}(\lambda)$ period of time, and continues to have more offspring after another $\text{Exp}(\lambda)$ period of time, and so on.

We start with $X(0) = 1$ individual. After an $\text{Exp}(\lambda)$ time, the individual has an offspring, and we have $X(t) = 2$.

Now each of these two individuals will have an offspring after an $\text{Exp}(\lambda)$ time each. So how much longer will it be until the first offspring appears and we get $X(t) = 3$? Well, the earliest offspring will appear at the minimum of these two $\text{Exp}(\lambda)$ times, and we saw in the Section 14 that this has an $\text{Exp}(2\lambda)$ distribution. (Remember that the mean of an exponential distribution is the reciprocal of the rate parameter, so the holding time for the second offspring is half that of the first, on average.)

Now suppose we have n individuals. By the memoryless property of the exponential distribution, they each still have an $\text{Exp}(\lambda)$ time until producing an offspring, so the time until the next offspring is the minimum of these, which is $\text{Exp}(n\lambda)$.

In general, we have built a counting process defined entirely by its starting point $X(0) = 1$, that the nth holding time T_n is exponential with rate $n\lambda$, and that all transitions are increases by 1.

On Problem Sheet 8, you will show that the expected size of the population of the simple birth process at time t is $\mathbb{E}X(t) = e^{\lambda t}$, so the population grows exponentially quickly (on average).

The simple birth process is an example of the more general class of birth processes.

Definition 16.1. A **birth process** (X_n) with rates (λ_n) is defined by its starting population $X(0) = x_0$ and that the holding times are $T_n \sim \text{Exp}(\lambda_n)$. So we have

$$X(t) = \begin{cases} x_0 & 0 \leq t < J_1 \\ x_0 + n & J_n \leq t < J_{n+1}, \text{ for } n = 1, 2, \dots, \end{cases}$$

where $J_n = T_1 + T_2 + \cdots + T_n$ are the **jump times**.

Example 16.1. We have the following examples:

- Setting $x_0 = 1$ and $\lambda_n = n\lambda$ gives the simple birth process.
- Setting $x_0 = 0$ and $\lambda_n = \lambda$ constant gives the Poisson process.
- Setting $x_0 = 1$ and $\lambda = n\lambda + \mu$ gives a birth process with immigration from outside the system at rate μ.

We can also give an equivalent definition using infinitesimal time periods, as we did for the Poisson process. Suppose we start from $X(0) = 1$. We have

$$\begin{aligned} \mathbb{P}(X(t + \tau) = j \quad | X(t) = j) &= 1 - \lambda_j \tau + o(\tau), \\ \mathbb{P}(X(t + \tau) = j + 1 \mid X(t) = j) &= \lambda_j \tau + o(\tau), \\ \mathbb{P}(X(t + \tau) \geq j + 2 \mid X(t) = j) &= o(\tau), \end{aligned}$$

which is exactly the same as the Poisson process, except with λ replaced by λ_j.

We can continue as for the Poisson process. Write $p_j(t) = \mathbb{P}(X(t) = j)$. Then, for $j \geq 2$,

$$p_j(t + \tau) = (1 - \lambda_j \tau + o(\tau))p_j(t) + (\lambda_{j-1}\tau + o(\tau))p_{j-1}(t) + o(\tau),$$

since the two ways to get to j are either we're already at j and the jth arrival doesn't occur, or we're at $j-1$ and the $(j-1)$th arrival does occur; other possibilities have $o(\tau)$ probability. As before, take $p_j(t)$ over to the left-hand side, divide by τ to get

$$\frac{p_j(t + \tau) - p_j(t)}{\tau} = -\lambda_j p_j(t) + \lambda_{j-1} p_{j-1}(t) + \frac{o(\tau)}{\tau},$$

and then send $\tau \to 0$. We end up with the forward equation

$$p_j'(t) = -\lambda_j p_j(t) + \lambda_{j-1} p_{j-1}(t).$$

The initial condition is $p_j(0) = 0$, since we start at 1, not at any $j \geq 2$.

Following the similar process for $j = 1$, we get

$$p_1'(t) = -\lambda_1 p_1(t),$$

with initial condition $p_1(0) = 1$, since we start from 1.

In some cases the forward equations can be explicitly solved to give an expression for $p_j(t) = \mathbb{P}(X(t) = j)$. We saw the solution for the Poisson process in Lecture 14, and you will find a solution for the simple birth process on Problem Sheet 8.

16.2   Time inhomogeneous Poisson process

When we dealt with the Poisson process, the rate of arrivals λ was the same all the time – so our call centre always received calls at the same rate, or the insurance company always received claims at the same rate. In real life, though, the rate of arrivals might be lower at some times and higher at some other times, so the rate $\lambda = \lambda(t)$ should perhaps depend on the time t.

This gives a process that is **time inhomogeneous**. The normal Poisson process and the birth processes were all **time homogeneous**; that is, the transition probabilities $\mathbb{P}(X(t + s) = j \mid X(t) = i)$ depended on the state i, j and on the length of time period t but not on the current time s. Most of the processes we consider in the rest of this course will be time homogeneous, but this subsection is the exception, because here the transition probabilities will change over time.

Definition 16.2. The **time inhomogeneous Poisson process** with rate function $\lambda = \lambda(t) \geq 0$ is defined as followed. It is a stochastic process $(X(t))$ with continuous time $t \in [0, \infty)$ a state space $\mathcal{S} = \mathbb{Z}_+$ with the following properties:

1. $X(0) = 0$;
2. Poisson increments: $X(t + s) - X(t) \sim \text{Po}(\int_t^{t+s} \lambda(u)\,du)$ for all $s, t > 0$;
3. independent increments: $X(t_2) - X(t_1)$ and $X(t_4) - X(t_3)$ are independent for all $t_1 \leq t_2 \leq t_3 \leq t_4$.

The difference here is in point 2, where we integrate the rate function $\lambda(t)$ over the interval of interest. In particular, if $\lambda(t) = \lambda$ is constant, then $\int_t^{t+s} \lambda(u)\,du = \lambda s$, and we get back the definition of the Poisson process in terms of Poisson increments.

Example 16.2. *A call centre notes that, when it opens the phone lines in the morning, phone calls arrive slowly at first, gradually becoming more common over the first hour. The owners of the centre model this is a time inhomogeneous Poisson process with rate function*

$$\lambda(t) = \begin{cases} 20t & 0 \leq t < 1 \\ 20 & t \geq 1 \end{cases}$$

calls per hour. What is the probability they receive no calls in the first 10 minutes?

The number of calls in the first 10 minutes, or $\frac{1}{6}$ of an hour, is Poisson with rate

$$\int_0^{1/6} \lambda(t)\,\mathrm{d}t = \int_0^{1/6} 20t\,\mathrm{d}t = \left[10t^2\right]_0^{1/6} = \frac{10}{36} = 0.278.$$

The probability no calls are received is

$$\mathrm{e}^{-0.278}\frac{0.278^0}{0!} = \mathrm{e}^{-0.278} = 0.757.$$

What is the expected number of calls in the first 2 hours?

The number of calls in the first 2 hours is Poisson with rate

$$\int_0^2 \lambda(t)\,\mathrm{d}t = \int_0^1 20t\,\mathrm{d}t + \int_1^2 20\,\mathrm{d}t = \left[10t^2\right]_0^1 + \left[20t\right]_1^2 = 10 + (40 - 20) = 30.$$

So the expected number of calls is 30.

The time inhomogeneous Poisson process can also be given a definition in terms of inifinitesimal incre-ments. We have

$$\mathbb{P}(X(t+\tau) - X(t) = j) = \begin{cases} 1 - \lambda(t)\tau + o(\tau) & \text{if } j = 0, \\ \lambda(t)\tau + o(\tau) & \text{if } j = 1, \\ o(\tau) & \text{if } j \geq 2. \end{cases}$$

The only difference here is that we have replaced the rate λ with the "current rate" $\lambda(t)$.

In the next section, we begin to look at the general theory of continuous time Markov jump processes.

Problem Sheet 8

You should attempt all these questions and write up your solutions in advance of your workshop in week 9 (Monday 22 or Tuesday 23 March) where the answers will be discussed.

1. Let $(X(t))$ be a Poisson process with rate λ.

(a) Fix n. What is the expected time between the nth arrival and the $(n+1)$th arrival?

(b) Fix t. What is the expected time between the previous arrival before t and the next arrival after t?

(c) Your answers to the previous two questions should be different. Explain why one should expect the second answer to be bigger than the first.

2. Let $X(t)$ be a Poisson process with rate λ, and mark each arrival independently with probability p. Use the infinitesimals definition to show that the marked process is a Poisson process with rate $p\lambda$.

3. Let $(X(t))$ be a simple birth process with rates $\lambda_j = \lambda j$ starting from $X(0) = 1$. Let $p_j(t) = \mathbb{P}(X(t) = j)$.

(a) Write down the Kolmogorov forward equations for $p_j(t)$. You should have separate equations for $j = 1$ and $j \geq 2$. Remember to include the initial conditions $p_j(0)$.

(b) Show that $X(t)$ follows a geometric distribution $X(t) \sim \mathrm{Geom}(e^{-\lambda t})$. That is, show that

$$p_j(t) = (1 - e^{-\lambda t})^{j-1} e^{-\lambda t}$$

satisfies the forward equation.

(c) Hence, calculate $\mathbb{E}X(t)$, the expected population size at time t.

4. Let $(X(t))$ be a simple birth process with rates $\lambda_n = \lambda n$ starting from $X(0) = 1$. Let $T_n \sim \mathrm{Exp}(\lambda n)$ be the nth holding time, and let $J_n = T_1 + T_2 + \cdots + T_n$ be the time of the nth birth.

(a) Write down $\mathbb{E}T_n$ and $\mathrm{Var}(T_n)$.

(b) Show that, as $n \to \infty$, the expectation $\mathbb{E}J_n$ tends to infinity, but the variance $\mathrm{Var}(J_n)$ is bounded.

5. The number of phonecalls my office receives in a three hour period is modelled as a time inhomogeneous Poisson process with rate function

$$\lambda(t) = \begin{cases} 3t & 0 \leq t < 1 \\ 3 & 1 \leq t < 2 \\ 9 - 3t & 2 \leq t \leq 3 \end{cases}.$$

(a) Calculate the probability I receive exactly one phonecall **(i)** in the first hour; **(ii)** in the second hour; **(iii)** in the third hour.

(b) Calculate the probability I receive exactly 3 phonecalls over the three hour period.

17 Continuous time Markov jump processes

- Time homogeneous continuous time Markov jump processes in terms of the jump Markov chain and the holding times
- Explosion of Markov jump processes

17.1 Jump chain and holding times

In this section, we will start looking at general Markov processes in continuous time and discrete space. All our examples will be time homogeneous, in that the transition probabilities and transition rates will remain constant over time.

We call these **Markov jump processes**, as we will wait in a state for a random holding time, and then we will suddenly "jump" to another state.

The idea will be treat two matters separately:

- **where we jump too:** This will be studied through the "jump chain", a discrete time Markov chain that tells us *what* transitions are made (but not *when* they are made).
- **how long we wait until we jump:** we call these the "holding times". To preserve the Markov property, these holding times must have an exponential distribution, since this is the only random variable that has the memoryless property.

Let us consider a Markov jump process $(X(t))$ on a state space $\mathcal{S}$. Suppose we are at a state $i \in \mathcal{S}$. The **transition rate** at which we wish to jump to a state $j \neq i$ will be written $q_{ij} \geq 0$. This means that, after a time with an exponential distribution $\text{Exp}(q_{ij})$, if the process has not jumped yet, it will jump to state j.

(We use the convention that if $q_{ij} = 0$ for some $j \neq i$, this means we will never jump from i to j. If for some i we have $q_{ij} = 0$ for *all* $j \neq i$, then we stay in state i forever, and i is an abaorbing state.)

So from state i, there are many other states j we could jump to, each waiting for a time $\text{Exp}(q_{ij})$. Which of these times will be up first, leading the process to jump to that state? And how long will it be until that first time is up and we move? The answer is given by Theorem 14.2, which showed us the distribution of the minimum of the $\text{Exp}(q_{ij})$s is itself exponential, with rate

$$q_i := \sum_{j \neq i} q_{ij}.$$

Further, and also from Theorem 14.2, the probability that we move to some state $j \neq i$ is

$$r_{ij} := \frac{q_{ij}}{\sum_{j \neq i} q_{ij}} = \frac{q_{ij}}{q_i}.$$

These last two displayed equations define the quantities q_i and r_{ij}. (If i is an absorbing state i, then by convenetion we put $r_{ii} = 1$ and $r_{ij} = 0$ for $j \neq i$.

So, this is the crucial idea: from state i, we wait for a holding time with distribution $\text{Exp}(q_i)$, then move to another state, choosing state j with probability r_{ij}.

It will be convenient to write all the transition rates q_{ij} down in a **generator matrix Q** defined as follows: the off-diagonal entries are q_{ij}, for $i \neq j$, and the diagonal entries are

$$q_{ii} = -q_i = -\sum_{j \neq i} q_{ij}.$$

In particular, the off-diagonal entries are positive (or 0), the diagonal entries are negative (or 0) and each row adds up to 0.

One convenient way to keep track of the movement of a continuous time Markov jump process will be to look at which states it jumps to, leaving the holding times for later consideration. That is, we can

consider the discrete time **jump Markov chain** (or just **jump chain**) (Y_n) associated with $(X(t))$ given by $Y_0 = X(0)$ and

$$Y_n = \text{state of } X(t) \text{ just after the } n\text{th jump.}$$

This is a discrete time Markov chain that starts from the same place $Y_0 = X(0)$ as $(X(t))$ does, and has transitions given by $r_{ij} = q_{ij}/q_i$. (The jump chain cannot move from a state to itself.)

Once we know where the jump process will move, we then know what the next holding time will be: from state j, the holding time will be $\text{Exp}(q_j)$

We can some all this up in the following (rather long-winded) formal definition.

Definition 17.1.

- Let $\mathcal{S}$ be a set, and λ a distribution on $\mathcal{S}$.
- Let $\mathsf{Q} = (q_{ij} : i, j \in \mathcal{S})$ be a matrix where $q_{ij} \geq 0$ for $i \neq j$ and $\sum_j q_{ij} = 0$ for all i, and write $q_i = -q_{ii} = -q_i = \sum_{j \neq i} q_{ij}$.
- Define $\mathsf{R} = (r_{ij} : i, j \in \mathcal{S})$ as follows: for i such that $q_i \neq 0$, write $r_{ij} = q_{ij}/q_i$ for $j \neq i$, and $r_{ii} = 0$; and for i such that $q_i = 0$, write $r_{ij} = 0$ for $j \neq i$, and $r_{ii} = 1$.

We wish to define the **Markov jump process** on a state space $\mathcal{S}$ with generator matrix Q and initial distribution λ.

- The **jump chain** (Y_n) is the discrete time Markov chain on $\mathcal{S}$ with initial distribution λ and transition matrix R.
- The **holding times** $T_1, T_2, \ldots$ have distribution $T_n \sim \text{Exp}(q_{Y_{n-1}})$, and are conditionally independent given (Y_n).
- The **jump times** are $J_n = T_1 + T_2 + \cdots + T_n$.

Then (at last!) the Markov jump process $(X(t))$ is defined by

$$X(t) = \begin{cases} Y_0 & \text{for } t < J_1 \\ Y_n & \text{for } J_n \leq t < J_{n+1}. \end{cases}$$

17.2 Examples

Example 17.1. Consider the Markov jump process on a state space $\mathcal{S} = \{1, 2, 3\}$ with transition rates as illustrated in the following transition rate diagram. (Note that a transition rate diagram never has arrows from a state to itself.)
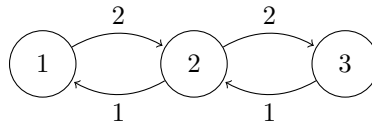
Figure 15: Transition diagram for a continuous Markov jump process.

The generator matrix is

$$\mathsf{Q} = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 1 & -1 \end{pmatrix},$$

where the off-diagonal entries are the rates from the diagram, and the diagonal entries are chosen to make sure the rows add up to 0.

The transition matrix of the jump chain is

$$\mathsf{R} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 1 & 0 \end{pmatrix},$$

where we set the diagonal elements equal to 0, then normalize each row so that it adds up to 1.

So, for example, starting from state 2, we wait for a holding time $T_1 \sim \text{Exp}(q_2) = \text{Exp}(-q_{22}) = \text{Exp}(3)$. We then move to state 1 with probability $r_{21} = q_{21}/q_2 = \frac{1}{3}$ and to state 3 with probability $r_{23} = q_{23}/q_2 = \frac{2}{3}$.

Suppose we move to state 1. Then we stay in state 1 for a time $\text{Exp}(q_1) = \text{Exp}(2)$, before moving with certainty back to state 2. And so on.

Example 17.2. Consider the Markov jump process with state space $\mathcal{S} = \{A, B, C\}$ and this transition rate diagram.
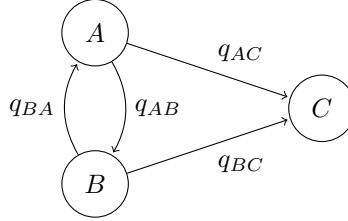

Figure 16: Transition diagram for a continuous Markov jump process with an absorbing state.

The generator matrix is

$$Q = \begin{pmatrix} -q_A & q_{AB} & q_{AC} \\ q_{BA} & -q_B & q_{BC} \\ 0 & 0 & 0 \end{pmatrix},$$

where $q_A = q_{AB} + q_{AC}$ and $q_B = q_{BA} + q_{BC}$. Here, state C is an absorbing state, so if we ever reach it, we stay there forever. So the jump chain has transition rates

$$R = \begin{pmatrix} 0 & q_{AB}/q_A & q_{AC}/q_A \\ q_{BA}/q_B & 0 & q_{BC}/q_B \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 17.3. The Poisson process with rate λ has state space $\mathcal{S} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$. The transition rates from i to $i+1$ are $q_{i,i+1} = \lambda$, and all other transition rates are 0. So the generator matrix is

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 \\ 0 & -\lambda & \lambda & 0 \\ 0 & 0 & -\lambda & \lambda \\ & & \ddots & \ddots \end{pmatrix}.$$

The jump chain is very boring: it starts from 0 and moves with certainty to 1, then with certainty to 2, then to 3, and so on.

17.3 A brief note on explosion

There is one point we have to be a little careful about with when dealing with continuous time processes with an infinite state space – the potential of "explosion". **Explosion** is when the process can get arbitrarily large in finite time. Think, for example, of a birth process with birth rates $\lambda_j = \lambda j^2$. The nth birth is expected to be occur at time

$$\sum_{j=1}^{n} \frac{1}{\lambda_j} = \frac{1}{\lambda} \sum_{j=1}^{n} \frac{1}{n^2}.$$

But this has a finite limit as $n \to \infty$, meaning we expect to get infinitely many births in a finite amount of time.

When explosion happens, it can then be rather difficult to define what's going on. In that birth process, for example, the state space is $\mathcal{S} = \mathbb{Z}_+$; but after some finite time, we have $X(t) > j$ for all $j \in \mathcal{S}$ – so what state *is* the process in?

Although there are often technical fixes that can get around this problem (doing things like creating an "infinity" state, or restarting the process from $X(0)$ again), we will not concern ourselves with them here. We shall normally deal with finite state spaces, where explosion is not a concern. When we do use infinite state space models (such as birth processes, for example), we shall only look at examples that explode with probability 0.

In the next section, we look at how to find the transition probilities $p_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i)$, which are the equivalents to the n-step transition probabilities in discrete time.

18 Forward and backward equations

- Markov jump process in infinitesimal time periods
- Transition semigroup, Chapman–Kolmogorov equations, and forward and backward equations
- Matrix exponential as a solution to the forward and backward equations

18.1 Transitions in infinitesimal time periods

In the last section, we defined the continuous time Markov jump process with generator matrix Q in terms of the holding times and the jump chain: from state i, we wait for a holding time exponentially distributed with rate $q_i = -q_{ii}$, then jump to state j with probability $r_{ij} = q_{ij}/q_i$.

So what happens starting from state i in a very small amount of time τ? The remaining time until a move is still $T \sim \operatorname{Exp}(q_i)$, by the memoryless property. So the probability we don't move in the small amount of time τ is

$$\mathbb{P}(T > \tau) = \mathrm{e}^{-q_i \tau} = 1 - q_i \tau + o(\tau).$$

Here we have used the tail probability of the exponential distribution and the first two terms of the Taylor series.

$$\mathrm{e}^x = 1 + x + o(x) \qquad \text{as } x \to 0.$$

The probability that we do move in the small amount of time τ, and that the move is to state j is

$$\mathbb{P}(T \leq \tau) r_{ij} = (q_i \tau + o(\tau)) \frac{q_{ij}}{q_i} = q_{ij} \tau + o(\tau).$$

The probability we make two or more jumps is a lower order term $o(\tau)$. Thus we have

$$\mathbb{P}\big(X(t + \tau) = j \mid X(t) = i\big) = \begin{cases} 1 - q_i \tau + o(\tau) & \text{for } i = j \\ q_{ij} \tau + o(\tau) & \text{for } i \neq j. \end{cases}$$

This is an equivalent definition of the Markov jump process.

18.2 Transition semigroup and the forward and backward equations

Let us write $p_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i)$ for the transition probability over a time t. This is the continuous-time equivalent to the n-step transition probability $p_{ij}(n)$ we had before in discrete time, but now t can be any positive real value. It can be convenient to use the matrix form $\mathsf{P}(t) = (p_{ij}(t) : i, j \in \mathcal{S})$.

In discrete time, we were given a transition matrix P, and it was easy to find $\mathsf{P}(n) = \mathsf{P}^n$ as a matrix power. In continuous time, we are given the generator matrix Q, so how can we find $\mathsf{P}(t)$ from Q?

First, let's consider $p_{ij}(s + t)$. To get from i to j in time $s + t$, we could first go from i to some k in time s, then from that k to j in time t, and the intermediate state k can be any state. So we have

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

You should recognise this as the **Chapman–Kolmogorov equations** again. In matrix form, we can write this as

$$\mathsf{P}(s + t) = \mathsf{P}(s)\mathsf{P}(t).$$

Pure mathematicians sometimes call this equation the "semigroup property", so we sometimes call the matrices $(\mathsf{P}(t) : t \geq 0)$ the **transition semigroup**.

Second, as in previous examples, we can try to get a differential equation for $p_{ij}(t)$ by looking at an infinitesimal increment $p_{ij}(t + \tau)$.

We can start with the Chapman–Kolmogorov equations. We have

$$p_{ij}(t + \tau) = \sum_k p_{ik}(t)p_{kj}(\tau)$$

$$= p_{ij}(t)(1 - q_j \tau) + \sum_{k \neq j} p_{ik}(t)q_{kj}\tau + o(\tau)$$

$$= p_{ij}(t) + \sum_k p_{ik}(t)q_{kj}\tau + o(\tau),$$

where we have treated the $k = j$ term of the sum separately, and taken advantage of the fact that $q_{jj} = -q_j$.

As we have done many times before, we take a $p_{ij}(t)$ to the left hand side and divide through by τ to get

$$\frac{p_{ij}(t + \tau) - p_{ij}(t)}{\tau} = \sum_k p_{ik}(t)q_{kj} + \frac{o(\tau)}{\tau}.$$

Sending τ to 0 gives us the (**Kolmogorov**) **forward equation**

$$p'_{ij}(t) = \sum_k p_{ik}(t)q_{kj}.$$

The initial condition is, of course, $p_{ii}(0) = 1$ and $p_{ij}(0) = 0$ for $j \neq i$.

Writing $\mathsf{P}(t) = (p_{ij}(t))$, and recognising the right hand side as a matrix multiplication, we get the convenient matrix form

$$\mathsf{P}'(t) = \mathsf{P}(t)\mathsf{Q} \qquad \mathsf{P}(0) = \mathsf{I},$$

where I is the identity matrix.

Alternatively, we could have started with the Chapman–Kolmogorov equations the other way round as

$$p_{ij}(t + \tau) = \sum_k p_{ik}(\tau)p_{kj}(t),$$

with the τ in the first term rather than the second. Following through the same argument would have given the **Kolmogorov backward equation**

$$\mathsf{P}'(t) = \mathsf{Q}\mathsf{P}(t) \qquad \mathsf{P}(0) = \mathsf{I},$$

where the Q and P are also the other way round.

The forward and backward equations both define the transition semigroup $(\mathsf{P}(t))$ in terms of the generator matrix Q.

18.3 Matrix exponential

In discrete time, when we are given the transition matrix P, we could find the n-step transition probabilities $\mathsf{P}(n)$ using the matrix power P^n. (At least when the state space was finite – one has to be careful with what the matrix power means if the state space is infinite.) In continuous time (and finite state space), when we are given the generator matrix Q, it will turn out that a similar role is played by the **matrix exponential** $\mathsf{P}(t) = e^{t\mathsf{Q}}$, which solves the forward and backward equations. In this way, the generator matrix Q is a bit like "the logarithm of the transition matrix P" – although this is only a metaphor, rather than a careful mathematical statement.

Let's be more formal about this. You may remember that the usual exponential function is defined by the Taylor series

$$e^x = \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots.$$

Similarly, we can define the matrix exponential for any square matrix by the same series

$$e^{\mathsf{A}} = \exp(\mathsf{A}) = \sum_{n=0}^{\infty} \frac{\mathsf{A}^n}{n!} = \mathsf{I} + \mathsf{A} + \tfrac{1}{2}\mathsf{A}^2 + \tfrac{1}{6}\mathsf{A}^3 + \cdots,$$

where we interpret $\mathsf{A}^0 = \mathsf{I}$ to be the identity matrix.

In the same way that a computer can easily calculate a matrix power P^n very quickly, a computer can also calculate matrix powers very quickly and very accurately – see Problem Sheet 9 Question 5 for more details about how.

Since a matrix commutes with itself, we have $\mathsf{A}^n\mathsf{A} = \mathsf{A}\mathsf{A}^n$, so a matrix commutes with its own exponential, meaning that $\mathsf{A}\mathrm{e}^{\mathsf{A}} = \mathrm{e}^{\mathsf{A}}\mathsf{A}$. This will be useful later.

Some properties of the standard exponential include

$$(\mathrm{e}^x)^n = \mathrm{e}^{nx} \qquad \mathrm{e}^{x+y} = \mathrm{e}^x\mathrm{e}^y \qquad \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{ax} = a\mathrm{e}^{ax}.$$

Similarly, we have for the matrix exponential

$$\left(\mathrm{e}^{\mathsf{A}}\right)^t = \mathrm{e}^{t\mathsf{A}} \qquad \mathrm{e}^{\mathsf{A}+\mathsf{B}} = \mathrm{e}^{\mathsf{A}}\mathrm{e}^{\mathsf{B}} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{t\mathsf{A}} = \mathsf{A}\mathrm{e}^{t\mathsf{A}} = \mathrm{e}^{t\mathsf{A}}\mathsf{A}.$$

From this last expression, we get that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{t\mathsf{Q}} = \mathrm{e}^{t\mathsf{Q}}\mathsf{Q} = \mathsf{Q}\mathrm{e}^{t\mathsf{Q}}.$$

Comparing this to the forward equation $\mathsf{P}'(t) = \mathsf{P}(t)\mathsf{Q}$ and backward equation $\mathsf{P}'(t) = \mathsf{Q}\mathsf{P}(t)$, we see that we have a solution to the forward and backward equations of $\mathsf{P}(t) = \mathrm{e}^{t\mathsf{Q}}$, which also satisfies the common initial condition $\mathsf{P}(0) = \mathrm{e}^{0\mathsf{Q}} = \mathsf{I}$.

We also see that we have the semigroup property

$$\mathsf{P}(s+t) = \mathrm{e}^{(s+t)\mathsf{Q}} = \mathrm{e}^{s\mathsf{Q}+t\mathsf{Q}} = \mathrm{e}^{s\mathsf{Q}}\mathrm{e}^{t\mathsf{Q}} = \mathsf{P}(s)\mathsf{P}(t).$$

As a formal summary, we have the following.

Theorem 18.1. *Let $(X(t))$ be a time homogeneous continuous time Markov jump process with generator matrix Q. Let $(\mathsf{P}(t))$ be the transition semigroup, where $\mathsf{P}(t) = (p_{ij}(t))$ is defined by $p_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i)$.*

Then $(\mathsf{P}(t))$ is the minimal nonnegative solution to the forward equation

$$\mathsf{P}'(t) = \mathsf{P}(t)\mathsf{Q} \qquad \mathsf{P}(0) = \mathsf{I},$$

and is also the minimal nonnegative solution to the backward equation

$$\mathsf{P}'(t) = \mathsf{Q}\mathsf{P}(t) \qquad \mathsf{P}(0) = \mathsf{I}.$$

When the state space $\mathcal{S}$ is finite, the forward and backward equations both have a unique solution given by the matrix exponential $\mathsf{P}(t) = \mathrm{e}^{t\mathsf{Q}}$.

In the next section, we develop the theory we already know in discrete time: communicating classes, hitting times, recurrence and transience.

Problem Sheet 9

You should attempt all these questions and write up your solutions in advance of your workshop in week 10 (Monday 26 or Tuesday 27 April) where the answers will be discussed.

1. Consider a Markov jump process with state space $\mathcal{S} = \{0, 1, 2, \dots, N\}$ and generator matrix

$$
Q = \begin{pmatrix}
-q_0 & q_{01} & q_{02} & \cdots & q_{0N} \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}.
$$

(a) Draw a transition rate diagram for this jump process.

This process is a "multiple decrement model": there is one "active state" 0 and a number of "exit states" $1, 2, \dots, N$.

(b) What is the probability that the process exits at state i?

(c) Give a 95% prediction interval for the amount of time spent in the active state. (Your answer will be in terms of q_0.)

2. Consider the Markov jump process $(X(t))$ with state space $\mathcal{S} = \{1, 2, 3\}$ and generator matrix

$$
Q = \begin{pmatrix}
-3 & 2 & 1 \\
2 & -6 & 4 \\
1 & 3 & -4
\end{pmatrix}.
$$

The process begins from the state $X(0) = 1$. Let (Y_n) be the associated Markov jump chain.

(a) Write down the transition matrix R of the jump chain.

(b) What is the expected time of the first jump J_1?

(c) What is the probability the first jump is to state 2?

(d) By conditioning on the first jump, calculate the expected time of the second jump time $J_2 = T_1 + T_2$.

(e) What is the probability that the second jump is to state 2?

(f) What is the probability that the third jump is to state 2?

3. Consider the Markov jump process on $\mathcal{S} = \{1, 2\}$ with generator matrix

$$
Q = \begin{pmatrix}
-\alpha & \alpha \\
\beta & -\beta
\end{pmatrix},
$$

where $\alpha, \beta > 0$.

(a) Write down the Kolmogorov forward equation for $p_{11}(t)$, including the initial condition.

(b) Hence, show that
$$
p'_{11}(t) + (\alpha + \beta)p_{11}(t) = \beta.
$$

(c) Solve this differential equation to find $p_{11}(t)$. (You may wish to first find a general solution the homogeneous differential equation by guessing a solution of the form $e^{\lambda t}$, then find a particular solution to the inhomogeneous differential equation, and then use the initial condition.)

4. We continue with the setup of Question 3.

(a) Show that $Q^2 = -(\alpha + \beta)Q$.

(b) Hence, write down Q^n for $n \geq 1$ in terms of Q.

(c) Show that
$$
P(t) = e^{tQ} = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!} = I + \frac{Q}{\alpha + \beta}\left(1 - e^{-(\alpha+\beta)t}\right).
$$

(Take care with $n = 0$ term in the sum.)

(d) What, therefore, is $p_{11}(t)$? Check you answer agrees with Question 3(c).

5. Let

$$\mathsf{D} = \begin{pmatrix} d_1 & 0 & & \\ 0 & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}$$

be a diagonal matrix.

(a) What is D^2?

(b) What is D^n for $n \geq 2$?

(c) What is e^{D}?

(d) Explain how you could use eigenvalue methods and diagonalisation to find an explicit formula for the entries of the exponential of a matrix A.

19 Class structure and hitting times

- Communicating classes
- Hitting probabilities and expected hitting times
- Recurrence and transience

When we studied discrete time Markov chains, we considered matters including communicating classes, periodicity, hitting probabilities and expected hitting times, recurrence and transience, stationary distributions, convergence to equilibrium, and the ergodic theorem. In this section and the next, we develop this theory for continuous time Markov jump processes. Luckily, things will be simpler this time round, as we will often be able to use exactly the same techniques as the discrete time case, often by looking at the discrete time jump chain.

19.1 Communicating classes

In discrete time, we said that j is accessible from i, and wrote $i \to j$ if $p_{ij}(n) > 0$ for some n. This allowed us to split the state space into communicating classes. We can do exactly the same in discrete time.

Definition 19.1. Let $(X(t))$ be a Markov jump process on a state space $\mathcal{S}$ with transition semigroup $(\mathsf{P}(t))$. We say that a state $j \in \mathcal{S}$ is **accessible from** another state $i \in \mathcal{S}$, and write $i \to j$ if $p_{ij}(t) > 0$ for some $t \geq 0$. If $i \to j$ and $j \to i$, we say that i **communicates with** j and write $i \leftrightarrow j$.

The equivalence relation $\leftrightarrow$ partitions $\mathcal{S}$ into equivalence classes, which we call **communicating classes**.

If all states are in the same class, the jump process is **irreducible**. A class C is **closed** if $i \to j$ for $i \in C$ means that $j \in C$ also. If i is a closed class $\{i\}$ be itself, then i is an **absorbing state**.

Recall that each state i is in exactly one communicating class, and that communicating class is the set of all j such that $i \leftrightarrow j$.

If $p_{ij}(t) > 0$, then this means there is some sequence of jumps that can take us from i to j. This means that, letting $\mathsf{R} = (r_{ij})$ be the transition matrix for the jump chain (Y_n), we have that $p_{ij}(t) > 0$ for some t if and only if $r_{ij}(n) > 0$ for some n. So the communicating classes for a Markov jump process are exactly the same as the communicating classes in its discrete time jump chain.

Further, since $r_{ij} > 0$ if and only if $q_{ij} > 0$, we can tell what the communicating classes are directly from the transition rate diagram.

Example 19.1. *Write down the communicating classes for the Markov jump process with generator matrix*

$$\mathsf{Q} = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$
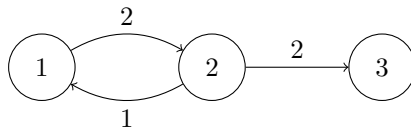
The transition rate diagram is

Figure 17: A Markov jump process with three states and two communicating classes.

We immediately see that the communicating classes are $\{1, 2\}$, which is not closed, and $\{3\}$ which is closed, so 3 is an absorbing state. The process is not irreducible.

19.2 A brief note on periodicity

In discrete time we had to worry about periodic behaviour, especially when considering long-term behaviour like the existence of an equilibrium distribution. However, for a continuous time jump process, it's possible stay in a state for any positive real-number amount of time. Thus (with probability 1) we never see periodic behaviour, and we don't have to worry about this.

This is one way in which continuous time processes are actually more pleasant to deal with than discrete time processes.

19.3 Hitting probabilities and expected hitting times

Recall that the **hitting probability** h_{iA} is the probability of reaching some state $j \in A$ at any point in the future starting from i, and the **expected hitting time** η_{iA} is the expected amount of time to get there. We found these in discrete time by forming simultaneous equations by conditioning on the first step.

For hitting probabilities, we just care about where we jump to, and not how long we wait. Thus we are free to consider only the jump chain (Y_n) and its transition matrix R. Everything works exactly as in discrete time.

Example 19.2. *Continuing with the previous example, calculate h_{21}, the probability of hitting state 1 starting from 2.*

The jump chain has transition matrix
$$\mathsf{R} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 1 \end{pmatrix}.$$

By conditioning on the first step, we have
$$h_{21} = \tfrac{1}{3} h_{11} + \tfrac{2}{3} h_{31} = \tfrac{1}{3} \times 1 + \tfrac{2}{3} \times 0 = \tfrac{1}{3},$$

where $h_{31} = 0$ since 3 is an absorbing state. Hence, $h_{21} = \frac{1}{3}$.

For expected hitting times, we have to more careful, as we will spend a *random* amount of time in the current state before moving on. In particular, the time we spend in state i is exponential with rate $q_i = -q_{ii}$, so has expected value $1/q_i$.

We illustrate the approach with an example.

Example 19.3. *Continuing with the same example, find the expected hitting time η_{13}.*

We condition on the first step. Starting from 1, we spend an expected time $1/q_1 = \frac{1}{2}$ in state 1 before moving on with certainty to state 2, so
$$\eta_{13} = \tfrac{1}{2} + \eta_{23}.$$

From state 2, we have a holding time with expectation $1/q_2 = \frac{1}{3}$, before moving to 1 with probability $\frac{1}{3}$ and 2 with probability $\frac{2}{3}$. So
$$\eta_{23} = \tfrac{1}{3} + \tfrac{1}{3} \eta_{12} + \tfrac{2}{3} \eta_{23} = \tfrac{1}{3} + \tfrac{1}{3} \eta_{12},$$

since η_{33} is of course 0. Substituting the second equation into the first, we get
$$\eta_{13} = \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{3} \eta_{13},$$

which rearranges to $\frac{2}{3} \eta_{13} = \frac{5}{6}$, giving $\eta_{13} = \frac{5}{4}$.

We have to be a little bit careful with return times and return probabilities. The return time is now the first time we come back to a state *after having left it*. In particular, if we begin in state i, we first leave at time $T_1 \sim \mathrm{Exp}(q_i)$, and we are looking for the first return after that.

So, specifically, the definitions of **return time**, **return probability**, and **expected return time** are

$$M_i = \min \{t > T_1 : X(t) = i\},$$
$$m_i = \mathbb{P}(X(t) = i \text{ for some } t > T_1 \mid X(0) = i) = \mathbb{P}(M_i < \infty \mid X(0) = i),$$
$$\mu_i = \mathbb{E}(M_i \mid X(0) = i).$$

In particular, if $q_i = 0$, so $T_1 = \infty$ and we never leave state i, the return time, return probability and expected return time are not defined. (We will have no use for them.)

By conditioning on the first step, it's clear we have

$$m_i = \sum_j r_{ij} h_{ji} = \sum_{j \neq i} \frac{q_{ij}}{q_i} h_{ji}, \qquad \mu_i = \frac{1}{q_i} + \sum_j r_{ij} \eta_{ji} = \frac{1}{q_i} \left(1 + \sum_{j \neq i} q_{ij} \eta_{ji} \right).$$

We see (except for the $q_i = 0$ case) that the return probability is the same in the jump chain as it is in the original Markov jump process, as was the case for the hitting probability, but for the expected return time we have to remember the random amount of time spent in each state.

19.4 Recurrence and transience

In discrete time, we said a state i was recurrent if the return probability $m_i = 1$ equalled 1 and transient if the return probability $m_i < 1$ was strictly less than 1. We take almost the same definition in continuous time – the difference is that if we never leave a state, because $q_i = 0$, we should call state i recurrent.

Definition 19.2. If a state $i \in \mathcal{S}$ has return probability $m_i = 1$ or has $q_i = 0$, we say that i is **recurrent**. If $m_i < 1$, we say that i is **transient**.

Suppose i is recurrent. If the expected return time $\mu_i < \infty$ is finite or $q_i = 0$, we say that i is **positive recurrent**. If the expected return time $\mu_i = \infty$ is infinite, we say that i is **null recurrent**.

We see that the return probability is totally determined the jump chain, so whether a state is recurrent or transient can be decided exactly as in the discrete case. In particular, in a communicating class either all states are transient or all states are recurrent. As before, finite closed classes are recurrent, and non-closed classes are transient.

Example 19.4. Picking up our example again, $\{1, 2\}$ is not closed, so is transient, while the absorbing state 3 is positive recurrent.

In the next section, we look at the long-term behaviour of continuous time Markov jump processes in a similar way to discrete time Markov chains: stationary distributions, convergence to equilibrium, and an ergodic theorem.

20 Long-term behaviour of Markov jump processes

- Stationary distributions, which solve $\pi\mathsf{Q} = \mathbf{0}$
- The limit theorem and the ergodic theorem

Our goal here is to develop the theory of the long-term behaviour of continuous time Markov jump processes in the same way as we did for discrete time Markov chains.

In discrete time, we defined **stationary distributions** as solving $\pi\mathsf{P} = \pi$. We then had the **limit theorem**, which told us about equilibrium distributions and the limit of $\mathbb{P}(X_n = i)$, and the **ergodic theorem**, which told us about the long term proportion of time spent in each state. We develop the same results here.

20.1 Stationary distributions

We start by defining a stationary distribution as before.

Definition 20.1. Let $(X(t))$ be a Markov jump process on a state space $\mathcal{S}$ with generator matrix Q and transition semigroup $(\mathsf{P}(t))$. Let $\pi = (\pi_i)$ be a distribution on $\mathcal{S}$, in that $\pi_i \geq 0$ for all $i \in \mathcal{S}$ and $\sum_{i \in \mathcal{S}} \pi_i = 1$. We call π a **stationary distribution** if

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_i p_{ij}(t) \quad \text{for all } j \in \mathcal{S} \text{ and } t \geq 0,$$

or, equivalently, if $\pi = \pi\mathsf{P}(t)$ for all $t \geq 0$.

This initially looks rather awkward: we're usually given a generator matrix Q, and then we might have to first find all the $\mathsf{P}(t)$s and then simultaneously solve infinitely many equations. Luckily it's actually much easier than that: we just have to solve $\pi\mathsf{Q} = \mathbf{0}$ (where $\mathbf{0}$ is the *row* vector of all zeros).

Theorem 20.1. *Let $(X(t))$ be a Markov jump process on a state with generator matrix Q. If $\pi = (\pi_i)$ is a distribution with $\sum_i \pi_i q_{ij} = 0$ for all j, then π is a stationary distribution. In matrix form, this condition is $\pi\mathsf{Q} = \mathbf{0}$.*

Proof. Suppose $\pi\mathsf{Q} = \mathbf{0}$. We need to show that $\pi\mathsf{P}(t) = \pi$ for all t. We'll first show that $\pi\mathsf{P}(t)$ is constant, then show that that constant is indeed π.

First, we can show that $\pi\mathsf{P}(t)$ is constant by showing that its derivative is zero. Using the backward equation $\mathsf{P}'(t) = \mathsf{Q}\mathsf{P}(t)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\pi\mathsf{P}(t) = \pi\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{P}(t) = \pi\mathsf{Q}\mathsf{P}(t) = 0\mathsf{P}(t) = 0,$$

as desired.

Second, we can find that constant value by setting t to any value: we pick $t = 0$. Since $\mathsf{P}(0) = \mathsf{I}$, the identity matrix, we have
$$\pi\mathsf{P}(t) = \pi\mathsf{P}(0) = \pi\mathsf{I} = \pi \qquad \text{for all } t.$$

$\square$

(Strictly speaking, taking the π outside the derivative in the first step is only formally justified when the state space is finite, but the result is true for general processes.)

Example 20.1. *In Section 17 we considered the Markov jump process with generator matrix*

$$\mathsf{Q} = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 1 & -1 \end{pmatrix}.$$

Find the stationary distribution.

We approach this in the same way as we would in the discrete time case.

First, we write out the equations of

$$
(\pi_1 \quad \pi_2 \quad \pi_3) \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 1 & -1 \end{pmatrix} = (0 \quad 0 \quad 0)
$$

coordinate by coordinate, to get

$$
\begin{aligned}
-2\pi_1 + \pi_2 &= 0 \\
2\pi_1 - 3\pi_2 + \pi_3 &= 0 \\
2\pi_2 - \pi_3 &= 0.
\end{aligned}
$$

We discard one of the equations – I'll discard the second one, as it's the most complicated.

Second, we pick a working variable – I'll pick π_2 since it's in both remaining equations – and solve the equations in terms of the working variable. This gives

$$
\pi_1 = \tfrac{1}{2}\pi_2 \qquad \pi_3 = 2\pi_2.
$$

Third, we use the normalising condition

$$
\pi_1 + \pi_2 + \pi_3 = \left(\tfrac{1}{2} + 1 + 2\right)\pi_2 = \tfrac{7}{2}\pi_2 = 1.
$$

This gives $\pi_2 = \tfrac{2}{7}$. Hence $\pi_1 = \tfrac{1}{7}$ and $\pi_3 = \tfrac{4}{7}$. Therefore, the stationary distribution is $\pi = (\tfrac{1}{7}, \tfrac{2}{7}, \tfrac{4}{7})$.

As before, we have a result on the existence and uniqueness of the stationary distribution.

Theorem 20.2. *Consider an irreducible Markov jump process with generator matrix* Q.

- *If the Markov jump process is positive recurrent, then a stationary distribution π exists, is unique, and is given by $\pi_i = 1/q_i\mu_i$, where μ_i is the expected return time to state i (unless the state space is a single absorbing state i, in which case $\pi_i = 1$).*
- *If the Markov jump process is null recurrent or transient, then no stationary distribution exists.*

As with the discrete time case, two conditions guarantee existence and uniqueness of the stationary distribution: irreducibility and positive recurrence. Compared to the discrete case, there's can extra factor of $1/q_i$ in the expression $\pi_i = 1/q_i\mu_i$, representing the expected amount of time spent in state i until jumping.

20.2   Convergence to equilibrium

The limit theorem tells us about the limit of $\mathbb{P}(X(t) = j)$ as $t \to \infty$. (We assume throughout that infinite-state processes are non-explosive – see Subsection 17.3.)

Recall from the discrete case that sometimes we have a distribution $\mathbf{p}^*$ such that $\mathbb{P}(X(t) = j) \to p_j^*$ for any initial distribution λ, and that such a distribution is called an **equilibrium distribution**. There can be at most one equilibrium distribution.

Theorem 20.3 (Limit theorem)**.** *Let $(X(t))$ be an irreducible and Markov jump process with generator martix Q. Then for any initial distribution λ, we have that that $\mathbb{P}(X(t) = j) \to 1/q_j\mu_j$ as $t \to \infty$ for all j, where μ_j is the expected return time to state j (unless the state space is a single absorbing state j, in which case $\mathbb{P}(X(t) = j) \to 1$).*

- *Suppose $(X(t))$ is positive recurrent. Then there is an equilibrium distribution which is the unique stationary distribution π given by $\pi_j = 1/q_j\mu_j$ (unless the state space is a single absorbing state j, in which case $\pi_j = 1$), and $\mathbb{P}(X(t) = j) \to \pi_j$ for all j.*
- *Suppose $(X(t))$ is null recurrent or transient. Then $\mathbb{P}(X(t) = j) \to 0$ for all j, and there is no equilibrium distribution.*

Note here that the conditions for convergence to an equilibrium distribution are irreducible and positive recurrent – we do not need a periodicity condition here as we did in the discrete time case.

For the positive recurrent case, if we take the initial distribution to be "starting in state i with certainty", we see that

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

and hence $\mathsf{P}(t)$ has the limiting value

$$\lim_{t \to \infty} \mathsf{P}(t) = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix},$$

with each row the same.

20.3   Ergodic theorem

As in the discrete time case, the ergodic theorem refers to the long run proportion of time spent in each state.

We write

$$V_j(t) = \int_0^t \mathbb{I}[X(s) = j] \, \mathrm{d}s$$

for the total amount of time spent in state j up to time t. Here, the indicator function $\mathbb{I}[X(s) = j]$ is 1 when $X(s) = j$ and 0 otherwise, so the integral is measuring the time we spend at j. So $V_j(t)/t$ is the proportion of time up to time t spent in state j, and we interpret its limiting value (if it exists) to be the long-run proportion of time spent in state j.

Theorem 20.4 (Ergodic theorem). *Let $(X(t))$ be an irreducible Markov jump process with generator matrix Q. Then for any initial distribution λ we have, for that $V_j(t)/t \to 1/q_j\mu_j$ almost surely as $t \to \infty$, where μ_j is the expected return time to state j (unless the state space is a single absorbing state j, in which case $V_j(t)/t \to 1$ almost surely).*

- *Suppose $(X(t))$ is positive recurrent. Then there is a unique stationary distribution π given by $\pi_j = 1/q_j\mu_j$ (unless the state space is a single absorbing state j, in which case $\pi_j = 1$), and $V_j(t)/t \to \pi_j$ almost surely for all j.*
- *Suppose $(X(t))$ is null recurrent or transient. Then $V_j(t)/t \to 0$ almost surely for all j.*

Example 20.2. In the previous example, the chain was irreducible and positive recurrent. Hence, by the limit theorem, $\pi = (\frac{1}{7}, \frac{2}{7}, \frac{4}{7})$ is the equilibrium distribution for the chain, and by the ergodic theorem, π also describes the long run proportion of time spent in each state.

In the next section, we apply our knowledge of continuous time Markov jump processes to two queuing models.

Problem Sheet 10

You should attempt all these questions and write up your solutions in advance of your workshop in week 11 (Tuesday 4 or Thursday 6 May) where the answers will be discussed.

1. Consider the Markov jump process on $\mathcal{S} = \{1, 2, 3, 4\}$ with generator matrix

$$Q = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & -\frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{6} & 0 & -\frac{1}{3} & \frac{1}{6} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

(a) Draw a transition rate diagram for this process.

(b) Write down the communicating classes for this process, and state whether they are recurrent or transient.

(c) Calculate the hitting probability h_{13}.

(d) Calculate the expected hitting time η_{14}.

2. Consider a Markov jump process $(X(t))$ on a triangle, with the vertices labelled 1, 2, 3 going clockwise. In a short time period τ, we move one step clockwise with probability $\alpha\tau + o(\tau)$, one step anticlockwise with probability $\beta\tau + o(\tau)$, or stay where we are.

(a) Write down a generator matrix for this Markov jump process, and draw a transition rate diagram.

(b) What is the probability, starting from state 1, that we hit state 3 before state 2?

(c) What is the expected time η_{13} to hit state 3 starting from state 1.

(d) Write down the transition matrix R for the jump chain (Y_n).

(e) What is the probability in the jump chain (Y_n) that, starting from state 1, that we hit state 3 before state 2?

(f) What is the expected number of steps in the jump chain (Y_n) to hit state 3 starting from state 1.

3.

(a) Calculate the stationary distribution for a Markov jump process with generator matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -3 & 2 \\ 0 & 2 & -2 \end{pmatrix}.$$

(b) Fill in the missing entries of the following generator matrix, and find two stationary distributions for the associated Markov jump process:

$$Q = \begin{pmatrix} ? & 2 & 0 \\ ? & -3 & 0 \\ ? & ? & 0 \end{pmatrix}.$$

4. Consider a Markov jump process $(X(t))$ with state space $\mathcal{S} = \{1, 2, 3, 4\}$ and generator matrix

$$Q = \begin{pmatrix} -2 & 2 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 2 & -4 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

(a) Draw a transition rate diagram for $(X(t))$.

(b) Write down the communicating classes for $(X(t))$, and state if each one is positive recurrent, null recurrent, or transient. Are there any absorbing states?

(c) Calculate the hitting probability h_{31}.

(d) Does $(X(t))$ have an equilibrium distribution?

(d) Conditional on Markov jump process starting from $X(0) = 3$, calculate the limiting probabilities $\lim_{n \to \infty} \mathbb{P}(X(t) = j \mid X(0) = 3)$ for all $j \in \mathcal{S}$.

5. Suppose $(X(t))$ is a continuous time Markov jump process with a stationary distribution $\phi = (\phi_i)$. Show that $\pi = (\pi_i)$, where

$$\pi_i = \frac{q_i \phi_i}{\sum_j q_j \phi_j},$$

is a stationary distribution for the discrete time jump chain associated to $(X(t))$.

Assessment 4

This assessment counts as 4% of your final module grade. You should attempt both questions. You must show your working, and there are marks for the quality of your mathematical writing.

The deadline for submission is **Thursday 6 May at 2pm**. Submission will be to Gradescope via Minerva, from Tuesday 4 May. It would be helpful to start your solution to Question 2 on a new page. If you hand-write your solutions and scan them using your phone, please convert to PDF using a proper scanning app (I like Microsoft Office Lens or Adobe Scan); please don't submit images or photographs.

Late submissions up to Thursday 13 May at 2pm will still be marked, but the total mark will be reduced by 10% per day or part-day for which the work is late. Submissions are not permitted after Thursday 13 May.

Your solutions to this assessment should be your own work. Copying, collaboration or plagiarism are not permitted. Asking others to do your work, including via the internet, is not permitted. Transgressions are considered to be a very serious matter, and will be dealt with according to the University's disciplinary procedures.

1. The arrival of buses is modelled as a Poisson process with rate $\lambda = 4$ per hour.

(a) What is the expected number of buses that arrive in a two-hour period? [1 mark]

(b) What is the probability that at least two buses arrives in a half hour period. [1]

(c) I arrive at the bus stop and start my watch. What is the expected amount of time until the sixth bus arrives. What is the standard deviation of this time? [2]

(d) Given that exactly two buses arrive between 2pm and 3pm, what is the expected arrival time of the first bus? [2]

(e) I arrive at the bus stop at 4pm. What is the expected inter-arrival time between the previous bus that I missed and the next bus to arrive that I catch? [1]

(f) I want to get the number 6 bus. Each bus that arrives is a number 6 bus independently with probability 0.25. What is the expected amount of time to wait until a number 6 bus? [1]

(g) What is the probability that at least two buses arrive in a one hour period, at least one of which is a number 6 bus? [2]

2. Consider the Markov jump process with generator matrix

$$
Q = \begin{pmatrix}
-2 & 2 & 0 & 0 & 0 \\
3 & -3 & 0 & 0 & 0 \\
0 & 1 & -2 & 1 & 0 \\
0 & 0 & 3 & -5 & 2 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix},
$$

and let $(P(t))$ be the associated transition semigroup. Find $\lim_{t\to\infty} P(t)$. [10]

Note: The majority of the marks for this question are for quality of mathematical exposition, with only a few marks for getting the answer right. Make sure to clearly explain and justify the steps you take. You may use results from the notes provided that you state them clearly and explain why they apply.

21 Queues

- The M/M/∞ infinite server process
- The M/M/1 single server queue

21.1 M/M/∞ infinite server process

In a mathematical model for a queue, individuals arrive, receive a service, and leave. In our first example, individuals receive service as soon as they arrive – so there is no actual "queueing". In our second example, later, individuals will actually queue up and wait until the server is able to serve them.

Our first process as follows. Individuals arrive like a Poisson process of rate λ. Then they receive a service which takes time $\mathrm{Exp}(\mu)$, independent of everything else, and then they leave.

This can be a useful model for many applications: for example, visitors watching a livestream on a website could arrive at the website like a Poisson process of rate λ and watch the livestream for a $\mathrm{Exp}(\mu)$ amount of time. Other models could include the number of students in the University library, the number of insurance contracts held by an insurance broker, or the number of files being downloaded from a server.

The number of individuals $(X(t))$ in the process (that is, receiving the service) at time t is a Markov jump process with arrival rates $q_{i,i+1} = \lambda$. The rate of departures when i individuals are being serviced has the rate and $q_{i,i-1} = i\mu$, because, thanks to the memoryless property of the exponential distribution, the first service time to finish is the minimum of i remaining exponential service times, and this is exponential with rate $i\mu$. Since rows of a generator matrix sum to 0, we have $q_i = -q_{ii} = \lambda + i\mu$.
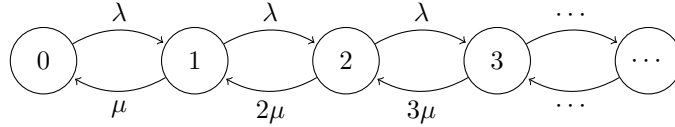

Figure 18: An M/M/∞ queue

Writing $\mathsf{Q} = (q_{ij})$ as a matrix, it is

$$\mathsf{Q} = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\lambda+\mu) & \lambda & & \\ & 2\mu & -(\lambda+2\mu) & \lambda & \\ & & 3\mu & -(\lambda+3\mu) & \lambda \\ & & & \ddots & & \ddots & \ddots \end{pmatrix},$$

where all the blank entries are 0.

This process is called the **M/M/∞** process.

- The first M stands for **memoryless arrivals**, which means the arrivals process is a Poisson process, which has exponential – and therefore memoryless – interarrival times.
- The second M stands for **memoryless service times**, which means the service times are exponentially distributed, and hence have the memoryless property.
- The ∞ stands for **infinite servers**, which means there is no limit on how many individuals can be serviced at once.

Two questions we might want to know the answer to are these:

1. In the long run, what's the average number of individuals being serviced?
2. In the long run, for what proportion of the time are all the servers idle?

To find out about this long run behaviour, we need to calculate the stationary distribution. The stationary distribution solves $\pi Q = 0$. Thus we have

$$\lambda \pi_{i-1} - (\lambda + i\mu)\pi_i + (i+1)\mu\pi_{i+1} = 0 \tag{14}$$

for $i \geq 1$, and

$$-\lambda \pi_0 + \mu\pi_1 = 0 \tag{15}$$

from the first column of Q.

Rearranging (15) gives $\pi_1 = \rho\pi_0$, where $\rho = \lambda/\mu$. Putting this into (14) for $i = 1$ gives (after some algebra) $\pi_2 = \pi_0\rho^2/2$. Putting that into for $i = 2$ gives $\pi_3 = \pi_0\rho^3/6$.vThis looks suspiciously like the stationary distribution might be the Poisson distribution with rate ρ; that is, $\pi_i = \pi_0\rho^i/i! = \mathrm{e}^{-\rho}\rho^i/i!$.

Theorem 21.1. *Let $(X(t))$ be an $M/M/\infty$ process with arrivals at rate λ and service at rate μ. The process has the stationary distribution $\pi_i = \mathrm{e}^{-\rho}\rho^i/i!$, corresponding to $X(t) \sim \mathrm{Po}(\rho)$, where $\rho = \lambda/\mu$.*

Proof. We know that this π is a distribution. We must check that π fulfills $\pi Q = 0$. We've already checked (15). For $i \geq 1$, we have from (14),

$$\lambda \pi_{i-1} - (\lambda + i\mu)\pi_i + (i+1)\mu\pi_{i+1}$$
$$= \lambda \mathrm{e}^{-\rho}\frac{\rho^{i-1}}{(i-1)!} - (\lambda + i\mu)\mathrm{e}^{-\rho}\frac{\rho^i}{i!} + (i+1)\mu\,\mathrm{e}^{-\rho}\frac{\rho^{i+1}}{(i+1)!}$$
$$= \lambda\mathrm{e}^{-\rho}\frac{i}{\rho}\frac{\rho^i}{i!} - (\lambda + i\mu)\mathrm{e}^{-\rho}\frac{\rho^i}{i!} + \mu\mathrm{e}^{-\rho}\rho\frac{\rho^i}{i!}$$
$$= \mathrm{e}^{-\rho}i\mu\frac{\rho^i}{i!} - (\lambda + i\mu)\mathrm{e}^{-\rho}\frac{\rho^i}{i!} + \mathrm{e}^{-\rho}\lambda\frac{\rho^i}{i!}$$
$$= \mathrm{e}^{-\rho}\frac{\rho^i}{i!}\big(i\mu - (\lambda + i\mu) + \lambda\big)$$
$$= 0,$$

as desired.

$\square$

This process is clearly irreducible, and since it has a stationary distribution it must be positive recurrent. So, by the limit theorem, we see that π is also the equilibrium distribution. By the ergodic theorem, the long-run amount of time spent in each state is governed by π.

In answer to question 1, the average number of individuals being serviced is $\mathbb{E}(\mathrm{Po}(\rho)) = \rho = \lambda/\mu$. In answer to question 2, the long-run proportion of time the process is empty is $\pi_0 = \mathrm{e}^{-\rho} = \mathrm{e}^{-\lambda/\mu}$.

21.2 M/M/1 single server queue

We now consider a slightly different process with only one server. This is called an **M/M/1** queue, where the "1" is the number of servers.

As before, individuals arrive as a Poisson process of rate λ. If there is no one currently being serviced, the individual goes straight to the server, and begins an $\mathrm{Exp}(\mu)$ service time. Otherwise, they have to join a queue waiting for the server to become free. When the server finishes serving one individual, then, provided the queue is nonempty, they immediately begin to service the individual who has been waiting in the queue for the longest time.

This is again a very useful model for many applications. For example, it could model a lecturer's office hours: students arrive at my office at rate λ; if I'm free, I can immediately answer their question, which will take an $\mathrm{Exp}(\mu)$ amount of time; but if I'm already dealing with a student, they must join a queue and wait until I am free. This could also model purchases at a shop with a single till, phone calls to a receptionist, or the work of a self-employed person on different tasks.

Let $(X(t))$ be the number of individuals in the process at time t. That could be $X(t) = 0$, if the server is free; $X(t) = 1$ if one individual is currently being serviced but no one else is waiting; or $X(t) = x > 1$

if one individual is being services and $x - 1$ individuals are waiting in the queue.. Individuals enter the queue at rate $q_{i,i+1} = \lambda$. When $i \geq 1$, meaning someone is being serviced, we have departures at rate $q_{i,i-1} = \mu$. Thus $q_i = -q_{ii} = \lambda + \mu$ for $i \geq 1$, and $q_0 = -q_{00} = \lambda$.
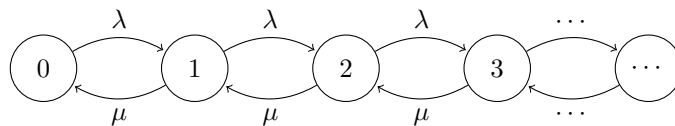

Figure 19: An M/M/1 queue

The matrix Q this time is

$$\mathsf{Q} = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\lambda + \mu) & \lambda & & \\ & \mu & -(\lambda + \mu) & \lambda & \\ & & \mu & -(\lambda + \mu) & \lambda \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

The first thing we would like to know is if the queue keeps getting longer and longer, or whether eventually every individual is serviced.

For this continuous time jump process $(X(t))$, the discrete time jump chain (Y_n) is a simple random walk with up probability $p = \lambda/(\lambda + \mu)$, down probability $q = \mu/(\lambda + \mu)$, and a reflecting barrier at 0. Recall that this Markov chain is transient for $p > q$, null recurrent for $p = q$, and positive recurrent for $p < q$. Thus we see that:

- For $\lambda > \mu$, the queue is transient, and eventually the length of the queue will grow longer and longer and never clear. This is clearly highly undesirable.
- For $\lambda = \mu$, the queue is null recurrent, so while the queue will eventually clear, the expected time for it to do so is infinite, so the queue will get very long, and individuals will may have to wait for an extremely long time. This is also undesirable.
- For $\lambda < \mu$, the queue is positive recurrent, so while individuals may sometimes have to wait, the queue will clear fairly regularly. This is what we want.

In what follows, we consider only the "good" case $\lambda < \mu$. Again we have two questions:

1. In the long run, what's the average number of individuals being serviced?
2. In the long run, for what proportion of the time is the server idle and for what proportion of time is the server busy?

Again, we look for a stationary distribution, which will tell us about the long term behaviour of the queue. The equation $\pi\mathsf{Q} = \mathbf{0}$ this time becomes

$$\mu\pi_{i+1} - (\lambda + \mu)\pi_i + \lambda\pi_{i-1} = 0 \qquad \text{for } i \geq 1,$$

together with the initial condition $\mu\pi_1 - \lambda\pi_0 = 0$ from the first column of Q, and the normalising condition $\sum_i \pi_i = 1$.

We recognise this as a linear difference equation. The characteristic equation (using α where we used λ previously, since λ is already in use) is

$$\mu\alpha^2 - (\lambda + \mu)\alpha + \lambda = (\alpha - 1)(\mu\alpha - \lambda) = 0.$$

This has solutions $\alpha = 1$ and $\alpha = \rho$, where $\rho = \lambda/\mu < 1$, so the general solution to the equation is $\pi_i = A + B\rho^i$.

The initial condition gives

$$\mu\pi_1 - \lambda\pi_0 = \mu(A + B\rho) - \lambda(A + B) = A\mu + B\lambda - A\lambda - B\lambda = A(\mu - \lambda) = 0.$$

Since $\lambda < \mu$, we must have $A = 0$. The normalising condition requires

$$\sum_{i=0}^{\infty} \pi_i = \sum_{i=0}^{\infty} B\rho^i = \frac{B}{1-\rho} = 1,$$

where we've used the formula for the sum of a geometric progression with $\rho < 1$, which gives $B = 1 - \rho$.

Thus we have the stationary distribution $\pi_i = (1-\rho)\rho^i$, which is a geometric distribution (the version that starts from $i = 0$).

Theorem 21.2. *Let $(X(t))$ be an M/M/1 queue with arrivals at rate λ and service at rate $\mu > \lambda$. The queue has the stationary distribution $\pi_i = (1-\rho)\rho^i$, corresponding to $X(t) \sim \mathrm{Geom}(\rho)$, where $\rho = \lambda/\mu < 1$.*

By the limit theorem, this is an equilibrium distribution, and the ergoic theorem also applied.

To answer question 1, the long-run average number of individuals in the process will be

$$\mathbb{E}(\mathrm{Geom}(\rho)) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

If λ is only slightly smaller than μ then this number can be quite large; in this case the queue can get very long, even though we know it will clear in a reasonable finite-expectation time. To answer question 2, the server will be free a proportion $\pi_0 = 1 - \rho = 1 - \lambda/\mu$ of the time and busy a proportion $1 - \pi_0 = \lambda/\mu$ of the time, in the long run. Again, if λ is only slightly smaller than μ, then the server will be busy most of the time.

This brings us to the end of the mathematical material in this module. **In the next section**, we summarise Part II on continuous time Markov jump processes and discuss the forthcoming exam.

22   End of Part II: Continuous time Markov jump processes

- Summary of Part II
- Exam questions answered

22.1   Summary of Part II

The following list is not exhaustive, but if you can do most of the things on this list, you're well placed for the exam. Recall that there was a similar summary of Part I previously.

- Define the Poisson process in terms of (a) independent Poisson increments, (b) independent exponential holding times, (c) transitions in an infinitesimal time period
- Perform basic calculations to do with a Poisson process, including with summed Poisson processes and marked Poisson processes
- Define and perform basic calculations with birth processes, including the simple birth process
- State the Markov property in continuous time
- Define and perform basic calculations with time inhomogeneous Poisson processes
- Define Markov jump processes in terms of (a) holding times and the jump chain, (b) transitions in an infinitesimal time period
- State and derive the forward and backward equations
- Draw a transition rate diagram
- Partition the state space into communicating classes, and identify if they are recurrent or transient
- Calculate hitting probabilities and expected hitting times
- Find a stationary distribution by solving $\pi Q = \mathbf{0}$
- Apply the limit and ergodic theorems on the long term behaviour of Markov jump chains
- Use techniques from the course to analyse simple models of queues

22.2   Exam FAQs

Some frequently asked questions about the exam. (I'll add to this list if there are more questions at Tuesday's lecture.)

- **How long is the exam?** The time between the paper being emailed to you and the deadline for submitting your work is 48 hours (plus submission time). However, you are *not* expected to spend all that time working on the exam. I was instructed that the paper should represent about 5 hours work, and that would seem a reasonable amount of time to spend on it to me.
- **How many questions are there in the exam?** There are four long-ish multi-part questions, of a similar length to those on the past papers, worth 20 marks each.
- **Is there R work on the exam?** There are two parts of one question that discuss R. In the first part, a few lines of R code are given, and you are asked to explain what is wrong with it; in the second part, you are asked to supply the correct commands. It is perfectly possible to answer this question without using R or RStudio, although you may optionally decide to use R to help give hints for the first part and check your answer to the second part.
- **Are there marks for showing my working and clearly explaining my answer?** Yes!
- **How will the exam compare to the 2020 past paper?** The exam will be very similar to the 2020 past paper, which was also a 48-hour take-home exam written by me. I can think of two differences. First, last year, the last quarter of the course (Sections 17–21) was disrupted due to the UCU strike and the start of the pandemic, so did not appear much on the exam; these sections *will* appear on your exam. Second, marks were a bit too high last year; I intend for a greater proportion of the 2021 paper to be as hard as the harder parts of the 2020 paper.

Problem Sheet 11

There are no workshops for this problem sheet, but you might like to use these two questions to test your understanding of Section 21.

1.

(a) Consider an M/M/1 queue with arrival rate λ and service rate $\mu < \lambda$. Justify the statement that the average time an individual spends waiting in the queue is

$$\frac{\lambda}{\mu(\mu - \lambda)}.$$

Solution. We saw in lectures that there are on average $\lambda/(\mu - \lambda)$ individuals in the queue. Each one will take an $\text{Exp}(\mu)$ time to be serviced, which has mean $1/\mu$. Multiplying these gives the result.

(b) Consider the queue for the single cash register in a shop. Customers arrive at a rate of 28 per hour, and that the average time at the till is 2 minutes. What is the average time, in minutes, that a customer waits in the queue?

Solution. If we work in time units of hours, we have $\lambda = 28$ and $\mu = 30$, and the average waiting time is

$$\frac{28}{30(30 - 28)} = \frac{28}{60} = 28 \text{ minutes.}$$

(c) The shop decides to open a second cash register. Each customer chooses one of the registers to queue for 50 : 50 at random. What is the average wait time now? (Remember to justify your answer carefully.)

Solution. The arrival process at each till is now a marked Poisson process, so has rate $\lambda/2 = 14$. This means the waiting time is

$$\frac{14}{30(30 - 14)} = \frac{7}{4 \times 60} = \frac{7}{4} \text{ minutes} = 1 \text{ minute } 45 \text{ seconds.}$$

(d) How might you improve the model of a shop with two cash registers to more accurately reflect true queueing behaviour? What effect would this have on the average wait time?

Solution. We might assume that customers join the shortest queue, rather than picking one at random. We might assume a customer will move to the other till if it becomes empty. We might even assume that customers will change queues if the other queue gets shorter than the one they are currently in. In all cases, the average waiting time will go down.

2. Telephone calls to a call centre are modelled as an M/M/s/s queue: call arrivals are a Poisson process with rate λ, call lengths are exponential with rate μ, and there are s workers available to answer the phone. The second s denotes that there is only enough phone lines for s callers at a time – if all workers are answering calls, any new calls are immediately dropped, and there is no technology for holding in a queue or waiting for a worker to become free.

(a) Let $X(t)$ be the number of calls being answered at time t. Write down the transition rates q_{ij} for $i, j \in \{0, 1, \ldots, s\}$, $i \neq j$.

Solution. The transition rate are $q_{i,i+1} = \lambda$ for $i = 0, 1, \ldots, s-1$, representing arrivals at rate λ when there is at least one free server, and $q_{i,i-1} = i\mu$ for $i = 1, 2, \ldots, s$ representing departures.

(b) *Erlang's formula* states that the long-run proportion of time that all s workers are answering calls is

$$\frac{\rho^s/s!}{\sum_{i=0}^{s} \rho^i/i!},$$

where $\rho = \lambda/\mu$. Prove Erlang's formula.

Solution. We claim that a stationary distribution is given by $\pi_j = C\rho^j/j!$ for $j = 0, 1, \ldots, s$, where the normalising constant will have to be

$$C = \frac{1}{\sum_{i=0}^{s} \rho^i/i!}.$$

By the ergodic theorem, the long-run proportion of time all s workers are answering calls is π_s.

It remains to prove the claim. The equations for $\pi Q = 0$ are

$$(i+1)\mu\pi_{i+1} - (\lambda + i\mu)\pi_i + \lambda\pi_{i-1} = 0.$$

for $i = 0, 1, \ldots, s-1$, and

$$-s\mu\pi_s + \lambda\pi_{s-1} = 0.$$

Proving this holds for $i = 0, 1, \ldots, s-1$ is exactly the same as for the M/M/∞ process in lectures, while for $i = s$ we have

$$
\begin{aligned}
-s\mu\pi_s + \lambda\pi_{s-1} &= -s\mu C\frac{\rho^s}{s!} + \lambda C\frac{\rho^{s-1}}{(s-1)!} \\
&= -C\mu\rho\frac{\rho^{s-1}}{(s-1)!} + C\lambda\frac{\rho^{s-1}}{(s-1)!} \\
&= -C\lambda\frac{\rho^{s-1}}{(s-1)!} + C\lambda\frac{\rho^{s-1}}{(s-1)!} \\
&= 0,
\end{aligned}
$$

thus proving the claim.

Computational worksheets

- **Computational Worksheet 1:** [HTML format] [Rmd format]
  - Example report: [HTML format] [Rmd format]
- **Computational Worksheet 2 (Assessment 2):** [HTML format] [Rmd format]
  - Marks and feedback are available on Minerva and Gradescope respectively. My example report is available on Minerva.

About the computational worksheets

These computational worksheets are an opportunity to learn more about Markov chains through simulation using R.

This first worksheet is not an assessed part of the course, but it is for you to learn and practice. The second worksheet *is* an assessed part of the course, and counts for 3% of your grade on this course. A report on Worksheet 2 will be due on **Thursday 18 March** (week 8) at 2pm. The material in these worksheets is examinable.

I recommend working on Computational Worksheet 1 in weeks 3 or 4, and on Computational Worksheet 2 in weeks 6 or 7. I estimate that each worksheet may take about 2 hours to work through.

You will have two computational drop-in sessions available with Muyang Zhang. These drop-in sessions are optional opportunities for you to come along to ask for help if you are stuck or want to know more. These sessions will happen on Teams. The sessions may appear in your timetable as "Practicals". It is important that you work through most of the worksheet *before* your drop-in session, as this will be your main opportunity to ask for help. (You can also use the module discussion board on Teams.) The dates of the drop-in sessions are:

- Computational Worksheet 1: Monday 15 – Wednesday 17 February (week 4)
- Computational Worksheet 2: Monday 8 – Wednesday 10 March (week 7)

Note that the Worksheet 2 practical sessions are the week before the deadline, so it's in your benefit to start working on that worksheet early.

The computational worksheets are available in two formats:

- First, as an easy-to-read **HTML file**. You should open this in a web browser.
- Second, as a file with the suffix **.Rmd**. This can be read as a plain text file. However, I recommend downloading this file and opening it in RStudio. This will make it easy to run the R code included in the file, by clicking the green "play" button by each chunk of R code. (These files is written in a language called "R Markdown", which you could choose to use for writing your report.)

How to access R

These worksheets use the statistical programming language **R**. Use of R is mandatory. I recommend interacting with R via the program **RStudio**, although this is optional. There are various ways to access R and RStudio.

- You may already have R and RStudio installed on your own computer.
- You can install R and RStudio on your own computer now if you haven't previously. You should first download R from the Comprehensive R Archive Network, and then download "RStudio Desktop" from rstudio.com. Remember to download and install R first, and only then to download and install RStudio.
- The RStudio Cloud is like "Google Docs for R". You can get 15 hours a month for free, which should be more than enough for these worksheets. Because this doesn't require installation, it's good for Chromebooks or computers where you don't have full installation rights.

- If you are in Leeds, all the university computers have R and RStudio installed. However, at the time of writing (5 February 2021), all IT clusters on campus are closed. You can see the latest news on cluster availability here.
- You can access R and RStudio via the university's virtual Windows desktop or (for those who have a Windows computer) via the university's AppsAnywhere system.

R background

These worksheets are mostly self-explanatory. However, they do assume a little background. For example, I assume you know how to operate R (for example by opening RStudio and typing commands in the "console"), and that you know that `x <- c(1, 2, 3)` creates a vector called `x` and that `mean(x)` calculates the mean of the entries of `x`.

Most students on this course will know R from MATH1710 and MATH1712 or other courses. If you are new to R, I recommend Dr Jochen Voss's "A Short Introduction to R" or Dr Arief Gusnanto's "Introduction to R", both available from Dr Gusnanto's MATH1712 homepage.