

MATH1710 Probability and Statistics I

Matthew Aldridge

University of Leeds, 2021–22

Contents

Schedule	5
About MATH2750	7
Organisation of MATH2750	7
Content of MATH2750	11
About these notes	13
Other stuff	17
R Worksheets	17
R worksheets	17
About R and RStudio	17
How to access R and RStudio	18
Installing R and RStudio	18
0.1 Drop-in sessions	19

Schedule

MATH1710 begins on Monday 27 September.

About MATH2750

Organisation of MATH2750

This module is **MATH1710 Probability and Statistics I**. A few students will be taking this module as half of **MATH2700 Probability and Statistics for Scientists**.

This module lasts for 11 weeks from 27 September to 10 December 2021. The exam will take place between 10 and 21 January 2022.

The core teaching team are:

- Dr Matthew Aldridge (you can call me “Matt” or “Dr Aldridge”): I am the module leader, the main lecturer, and the main author of these notes.
- A module assistant TBC.

The shared email address for the core teaching team is math1710@leeds.ac.uk; please use this address, rather than emailing our personal addresses; this will ensure your email is seen as soon as possible.

Notes and videos

The main way you will learn new material for this module is by reading these notes and by watching the accompanying pre-recorded videos. There will be one section of notes each week, for a total of 11 sections, with the final section being a summary and revision.

Reading mathematics is a slow process. Each section should take one and a half to two hours to work through; we recommend you split this into two or more sessions. If you find yourself regularly getting through sections in much less than that amount of time, you’re probably not reading carefully enough through each sentence of explanation and each line of mathematics, including understanding the motivation, checking the accuracy, and making your own notes.

You are probably reading the web version of the notes. If you want a PDF or ebook copy (to read offline or to print out), they can be downloaded via the top ribbon of the page. (Warning: I have not made as much effort to make the PDF and ebook as neat and tidy as I have the web version, and there may be formatting errors.)

We are very keen to hear about errors in the notes mathematical, typographical or otherwise. Please, please email us if think you may have found any.

Problem sheets

There will be 5 problem sheets. Each problem sheet has a number of short and long questions for you to cover in your own time to help you learn the material, and two assessed questions, which you should submit for marking. The assessed questions on each problem sheet make up 3% of your mark on this module, for a total of 15%.

Problem Sheet	Sections covered	Assessed work due
1	1	Friday 8 October (week 2)
2	2 and 3	Friday 22 October (week 4)
3	4 and 5	Friday 5 November (week 6)
4	6 and 7	Friday 19 November (week 8)
5	8, 9 and 10	Friday 3 December (week 10)

Assessed questions should be submitted in PDF format through Gradescope. (Further Gradescope details will follow.) Most students choose to hand-write their solutions and then scan them to PDF using their phone; you should use a proper scanning app – we recommend Microsoft Office Lens or Adobe Scan – and not just submit photographs.

Lectures

You will have one online synchronous (that is, live, not recorded) “lecture” session each week, with me, run through Zoom. Because this is a large cohort, we will split into two groups:

- Group 1: Mondays at 1200
- Group 2: Mondays at 1500

You should check your timetable to see which lecture group you are in.

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 via quizzes, polls, and the chat box.

We will assume you have completed all the work for the previous week by the time of the lecture.

We are very keen to hear about things you’d like to go through in the lectures; please email us with your suggestions.

Tutorials

Tutorials are small groups of about a dozen students. You have been assigned to one of 38 tutorial groups, each with a member of staff as the tutor. Your tutorial group will meet five times, in weeks 2, 4, 6, 8, and 10. Tutorial groups will meet in person on campus; you should check your timetable to see when and where your tutorial group meets. (For those not yet on campus, due to travel restrictions or health conditions, there will be an extra online tutorial group for the first few tutorials.)

The main goal of the tutorials will be to go over your answers to the non-assessed questions on the problems sheets in an interactive session. In this smaller group, you will be able to ask detailed questions of your tutor, and have the chance to discuss your answers to the problem sheet. Your tutor may ask you to present some of your work to your fellow students, or may give you the opportunity to work together with others during the tutorial. Your tutor may be willing to give you a hint on the assessed questions if you've made a first attempt but have got stuck.

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

- Work through the problem sheet before the tutorial, spending plenty of time on it, and making multiple efforts at questions you get stuck on. I recommend spending *at least 3 hours per week* on the problem sheets, which will usually mean a total of *at least 6 hours per problem sheet* (as most problem sheets cover two weeks). Collaboration is encouraged when working through the non-assessed problems, but I recommend writing up your work on your own; answers to assessed questions must be solely your own work.
- Take advantage of the small group setting of the tutorial to ask for help or clarification on questions you weren't able to complete.
- After the tutorial, 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.

Your tutor will also be the marker of your answers to the assessed questions on the problem sheets.

R worksheets

R is a programming language that is particularly good at working with probability and statistics. Learning to use R is an important part of this module, and is used in many other modules in the University, particularly in MATH1712 Probability and Statistics II. R is used by statisticians throughout academic and increasingly in industry too. Learning to program is a valuable skill for all students, and learning to use R is particularly valuable for students interested in statistics and related topics like actuarial science.

You will learn R by working through one R worksheet each week in your own time. Worksheets 3, 5, 7, 9 and 11 will also contain a couple of questions for assessment. Each of these is worth 3% of your mark for a total of 15%. I recommend spending one hour per week on the week's R worksheet, plus one extra hour if there are assessed questions that week.

You can read more about the language R, and about the program RStudio that we recommend you use to interact with R, in the R section of these notes.

To help you if you have problems with R, we have organised optional **R troubleshooting drop-in sessions**, where you can discuss any problems you have with an R expert, in weeks 2 and 3. Check your timetable for details – these will be listed on your timetable as “practicals”.

Office hours

If you there is something in the module you wish to discuss in detail with the module core teaching team, the place for the is the optional weekly “office hours”, which will operate as drop-in sessions. These sessions are an optional opportunity for you to ask questions you have to a member of staff; these are particularly useful if there's something on the module that you are stuck on or confused about, but we're happy to discuss any statistics-related issues or questions you have.

There will be two office hours per week: Wednesdays at 1000 and at 1200. (For boring reasons, the 1000 sessions appear on the timetable for MATH2700 students and the 1200 sessions appear on the timetable for MATH1710 students, but I'm happy for anyone to attend either hour.) The sessions will happen, until further notice, at least, in my office, PRD 9.320 on the the ninth floor of the Physics Research Deck.

Time management

It is, of course, up to you how you choose to spend your time on this module. But my recommendations for your weekly work would be something like this:

- **Notes and videos:** 2 hours per week/section
- **Problem sheet:** 3 hours per week (so 6 hours for most problem sheets) plus 1 extra hour for writing up and submitting answers to assessed questions
- **R worksheet:** 1 hour per week/worksheet, plus 1 extra hour if there are assessed questions
- **Lecture:** 1 hour per week
- **Tutorial:** 1 hour every other week
- **Revision:** 13 hours total at the end of the module

That's roughly 8 hours a week, and makes 100 hours in total. (MATH1710 is a 10 credit module, so is supposed to represent 100 hours work. MATH2700 students are expected to be able to use their greater experience to get through the material in just 75 hours, so should scale these recommendations accordingly.)

Exam

There will be an exam in January, which makes up the remaining 70% of your mark. The exam will consist of 20 short and 2 long questions, and will be time-limited to 2 hours. We'll talk more about the exam format near the end of the module.

Who should I ask about...?

Remember that the email address for the core module teaching team is math1710@leeds.ac.uk. Please don't email our personal addresses; it will take longer for us to reply, and we may miss your email all together.

- *I don't understand something in the notes or on a problem sheet:* Come to office hours, or (if the timing works) ask your tutor in your next tutorial.
- *I'm having difficulties with R:* In weeks 2 or 3, you should attend the trouble-shooting drop-in session; at other times, come to office hours.
- *I have an admin question about arrangements for the module:* Come to office hours or email the core module teaching team.
- *I have an admin question about arrangements for my tutorial:* Contact your tutor.
- *I have an admin question about general arrangements for my course as a whole:* Email the Maths Taught Students Office (Maths.Taught.Students@leeds.ac.uk) or speak to your personal academic tutor.
- *I have a question about the marking of my assessed work on the problem sheets:* First, check your feedback on Gradescope; if you still have questions, contact your tutor.
- *I have a question about the marking of my assessed work on the R worksheets:* Come to office hours or email the core module teaching team.
- *I have suggestion for something to cover in the lectures:* Email the core module teaching team.
- *Due to exceptional personal circumstances I require an extension on or exemption from assessed work:* Email the Maths Taught Students Office; neither the core module teaching team nor your tutor are able to offer extensions or exemptions. (Only exemptions, not extensions, are available for R worksheets.)

Content of MATH2750

Prerequisites

The formal prerequisite for MATH1710 is "Grade B in A-level Mathematics or equivalent". We'll assume you have some basic school-level maths knowledge, but we don't assume you've studied probability or statistics in detail before (although we recognise that many of you will have). If you have studied probability and/or statistics at A-level (or post-16 equivalent) level, you'll recognise

some of the material in this module; however you should find that we go deeper in some areas, and that we treat the material through with a greater deal of mathematical formality and rigour. “Rigour” here means precisely stating our assumptions, and carefully *proving* how other statements follow from those assumptions.

Syllabus

The module has three parts: a short first part on “exploratory data analysis”, a long middle part on probability theory, and a short final part on a statistical framework called “Bayesian statistics”. There’s also the weekly R worksheets, which you could count as a fourth part running in parallel, but which will connect with the other parts too.

An outline plan of the topics covered is the following. (Remember that one section is one week’s work.)

- **Exploratory data analysis** [1 section] Summary statistics, data visualisation
- **Probability** [8 sections]
 - Probability with events: Probability spaces, probability axioms, examples and properties of probability, “classical probability” of equally likely events, independence, conditional probability, Bayes’ theorem [3 sections]
 - Probability with random variables: Discrete random variables, expectation and variance, binomial distribution, geometric distribution, Poisson distribution, multiple random variables, law of large numbers, continuous random variables, exponential distribution, normal distribution, central limit theorem [5 sections]
- **Bayesian statistics** [1 section]: Bayesian framework, Beta prior, normal-normal model
- Summary and revision [1 section]

Books

You can do well on this module by reading the notes and watching the videos, attending the lectures and tutorials, and working on the problem sheets and R worksheets, without needing to do any further reading beyond this. However, students can benefit from optional extra background reading or an alternative view on the material, especially in the parts of the module on probability.

For exploratory data analysis, you can stick to Wikipedia, but if you really want a book, I’d recommend:

- GM Clarke and D Cooke, *A Basic Course in Statistics*, 5th edition, Edward Arnold, 2004.

For the probability section, any book with a title like “Introduction to Probability” would do. Some of my favourites are:

- JK Blitzstein and J Hwang, *Introduction to Probability*, 2nd edition, CRC Press, 2019.
- G Grimmett and D Welsh, *Probability: An Introduction*, 2nd edition, Oxford University Press, 2014. (The library has online access.)
- SM Ross, *A First Course in Probability*, 10th edition, Pearson, 2020.
- RL Scheaffer and LJ Young, *Introduction to Probability and Its Applications*, 3rd edition, Cengage, 2010.
- D Stirzaker, *Elementary Probability*, 2nd edition, Cambridge University Press, 2003. (The library has online access.)

On Bayesian statistics, I recommend:

- JV Stone, *Bayes’ Rule: A Tutorial Introduction to Bayesian Analysis*, Sebtel Press, 2013.

For R, there are many excellent resources online, and Google is your friend for finding them.

(For all these books I’ve listed the newest editions, but older editions are usually fine too.)

About these notes

These notes were written by Matthew Aldridge in 2021. Editing help was provided by XXX. They are based in part on previous notes by Dr Robert G Aykroyd and Prof Wally Gilks. Dr Jason Anquandah and Dr Aykroyd advised on the R worksheets. Dr Aykroyd’s help and advice on many aspects of the module was particularly valuable.

These notes (in the web format) should be accessible by screenreaders. The videos have (highly imperfect) automated subtitles. If you have accessibility difficulties with these notes, contact maths1710@leeds.ac.uk.

Other stuff

R Worksheets

R worksheets

Each week there will be an R worksheet to work through in your own time. We recommend spending about one hour on each worksheet, plus one extra hour for worksheets with assessed questions, for checking through and submitting your solutions.

Week	Worksheet	Deadline for assessed work
1	R basics	—
2	Working with vectors	—
3	Importing data into R	Friday 15 October
4	Plots I: Making plots	—
5	Plots II: Making plots nicer	Friday 29 October
6	RMarkdown (optional)	—
7	Discrete random variables	Friday 12 November
8	Discrete distributions	—
9	Normal distribution	Friday 26 November
10	Law of large numbers	—
11	Summary	Friday 10 December

About R and RStudio

- **R** is a *programming language* that is particularly good at working with probability and statistics. R is very widely used in universities and increasingly widely used in industry. Learning to use R is a mandatory part of this module, and exercises requiring use of R make up at least 15% of your module mark. Many other statistics-related course at the University also use R.
- **RStudio** is a *program* that gives a convenient way to work with the language R. RStudio is the most common way to use the language R, and learning to use RStudio is strongly recommended.

R and RStudio are free/open-source software.

How to access R and RStudio

There are a number of ways you can access R and RStudio:

- All **University computers** have R and RStudio already installed. Here is a directory of the University’s computer clusters.
- You can **install** R and RStudio on your own computer – see the instructions below.
- If you want to use R/RStudio on a non-University device for which you don’t have admin/installation rights (Chromebook, iPad, friend’s laptop, etc), you could try:
 - You can use the University’s copies of R/RStudio virtually through the Windows Virtual Desktop or AppsAnywhere client.
 - The RStudio Cloud is a cloud-hosted “Google Docs for R” that you can use through your web browser – you can get 15 hours per month for free (or pay for more).

Installing R and RStudio

Students who have their own computer usually find it most convenient to install R and RStudio on that computer. To do this, it’s important that you install R (the programming language) first, and only install RStudio (the program to use R) once R has already been installed.

1. *First*, install **R**. Go to the Comprehensive R Archive Network and follow the instructions:
 - Windows: Click “Download R for Windows”, then “Install R for the first time”. The main link at the top should be to download the most recent version of R.
 - Mac: Click Download R for macOS, and then download the relevant PKG file. (For pre-November 2020 Intel-based Macbooks, you must use the “Intel 64-bit build”; for post-November 2020 M1-based “Apple silicon” Macbooks, the “Apple silicon arm64 build” may be faster.)
2. *After* R is installed, *then* install **RStudio**. Go to the Download page at RStudio.com and follow the instructions. You want “RStudio Desktop”, and you want the free version.

If you have difficulty installing R, come along to the first computational drop-in session in Week ?? and bring your computer with you (if it’s sufficiently portable), and we’ll do our best to help.

0.1 Drop-in sessions

You will learn to use R by working through the R Worksheets. Learning to use a programming language is different from learning mathematics: you should expect to get frustrated and annoyed when the computer seems to refuse to do what you want it to. This is a normal part of learning.

However, many students find getting with started with R in the first few weeks particularly difficult. Also, sometimes students have problems installing R and RStudio on their own computers. To help with this XXXXXXXXx