MAS2903

Introduction to Bayesian Methods

Semester 2, 2019—2020

Lecturer: Dr. Lee Fawcett

School of Mathematics, Statistics & Physics

MAS2903: Introduction to Bayesian Methods

Lecturer: Dr. Lee Fawcett

Office: Room 2.07, Herschel Building Phone: 208 7228

Email: lee.fawcett@ncl.ac.uk

Classes

- There will be two *lectures* each week: Mondays at 3pm and Thursdays at 1pm, both in Lecture Theatre 1 of the Bedson Building.
- In the first two weeks there will be a third lecture on Thursdays at 2pm, also in Lecture Theatre 1 of the Bedson Building.
- From the third week onwards *problems classes* and *drop-in sessions* take place in alternate weeks on Thursdays at 2pm in Lecture Theatre 1 of the Bedson Building.
- *Computer practicals* will take place in teaching weeks 4 and 9. These will be on Tuesdays at 12pm in the Herschel PC cluster. A reminder will be given the week before!
- *Office hours:* I will be available to see MAS2903 students on Mondays 2-3pm and Wednesdays 1-2pm (my office).

What?	**When?**		**Where?**	**How often?**
Lecture	Monday	3–4	Bedson: LT1	Every week
	Thursday	1–2	Bedson: LT1	Every week
Problems class/ Drop–in session	Thursday	2–3	Bedson: LT1	Every week has one or the other
Computer practical	Tuesday	12–1	Herschel: PC cluster	Occasionally
Office hours	Monday	2–3	My office	Every week
	Wednesday	1–2		

Assessment

- **85%** exam in May/June
- **15%** in course assessment:
  - Mid–semester test (10%): **Monday 9th March, 3pm** (split between several rooms – details to follow by email)
  - 2 assignments (5% in total), each one having a mixture of written questions and practical work. Deadlines – **Fridays** at **3pm** on the following dates:

    6th March, **1st May**.

Recommended textbooks

- "*Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*" - James Stone
- "*Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*" - John Krushke
- "*Bayesian Statistics: An Introduction*" - Peter Lee

"Bayes' rule" is a good introduction to the main concepts in Bayesian statistics but doesn't cover everything in this course. The other books are broader references which go well beyond the contents of this course.

Other stuff

- Notes (with gaps) will be handed out in lectures – you should fill in the gaps during lectures.
- A summarised version of the notes will be used in lectures as slides.
- The notes are available on Blackboard. The slides will be posted after each topic is finished, along with any other course material – such as model solutions to assignment questions, supplementary handouts etc.
- There is also a course webpage (just in case Blackboard breaks):

  www.mas.ncl.ac.uk/~nlf8/teaching/mas2903

- Please check your University email account regularly, as course announcements will often be made via email.

Before starting the course 'proper', we will spend the first lecture going over some background to the subject of Bayesian Statistics. Although the material in the rest of this preamble is non–examinable, it will help to 'set the scene' for the course as well as motivate the techniques we will develop. Let's make a start...

Preface to lecture notes: What is Bayesian Statistics?

There are two main approaches to statistics: *frequentist* (or classical) statistics and *Bayesian* statistics. All of the statistics teaching you've encountered so far is likely to be about frequentist methods. Bayesian methods are substantially different and can feel quite strange to start with. So, before starting the main course material, this *non-examinable* preface explains the concepts behind Bayesian statistics and how they differ from frequentist approaches.

One way of defining statistics is as a way to learn about the world from some data which is subject to random variation. For example, in *climate science* we want to learn about the climate given some imperfect measurements. Some statistical questions that we might ask in this field are:

- What is our best estimate of the world temperature this year?
  This is referred to as a *point estimation* problem.
- What is a plausible range of values?
  This is referred to as an *interval estimation* (or uncertainty quantification) problem.
- What will the temperature be in 100 years?
  This is a *prediction* problem.
- Is the climate warming?
  This is a *hypothesis testing* problem.

Frequentist and Bayesian methods address all these types of problem: point estimation, interval estimation, prediction and hypothesis testing. But they use different approaches to do so. Some typical frequentist approaches include:

- Least squares (point estimation)
- Maximum likelihood (point estimation)
- Confidence intervals (interval estimation)
- Test statistics and p-values (hypothesis testing)

Bayesian statistics doesn't use any of these familiar methods! However, there are some connections between them and their Bayesian alternatives which will be explored later in the course. For example we'll see that the idea of the *likelihood function* is very important in Bayesian statistics.

Motivating example

Air France Flight 447 disappeared over the Atlantic Ocean on June 1st 2009, during an overnight flight from Rio de Janeiro to Paris. A search and rescue mission was launched the following morning. After five days some floating wreckage was found, but there was no sign of the "black box" recorders which would reveal what had happened to the flight, and the search stalled.

The black boxes could be anywhere within an area of the South Atlantic the size of Switzerland (6,500 square miles). The "mid-Atlantic ridge" on the ocean floor lies between two tectonic plates and is just as mountainous as Switzerland! It's also so remote that scientists have not yet charted the sea-bed. The search method used was sonar-detectors, emitting sound waves which would bounce back once they hit something. However, after two years of meticulously searching an area north of the planes trajectory (after analysing debris drift), nothing had been found.

Metron, Inc., of Retson, Virginia had been performing Bayesian analyses to help the search effort. Included in their analysis were:

1. Data from 9 previous airline crashes involving loss of pilot control - reduced search area to 1,600 square miles.

2. Expert opinions on the credibility of the flight data.

3. Expert opinions about whether or not the black box 'pingers' might have become damaged on impact.

4. Positions/recovery times of bodies found drifting - expert opinions assigned to the reliability of this data because of the turbulent equatorial waters.

5. Expert information from oceanographers: Sea state, visibility, underwater geography,...

All were combined through Bayes Theorem to give a probability map of the most likely locations to search. In April 2011 they decided to try assuming that the 'pingers' in item 3 were probably damaged. One week later the black boxes were found![1]

[1] For more details see `https://www.technologyreview.com/s/527506/how-statisticians-found-air-france-flight-447-two-years-after-it-crashed-into-atlantic/`.

The frequentist/Bayesian debate

This example illustrates some key features of the Bayesian approach:

- It incorporates expert opinion.
- It outputs a probability distribution over the possible choices.
- The calculations required are all based on Bayes Theorem.

In contrast in frequentist statistics it is difficult (but not impossible) to incorporate opinions, and the output is not in the form of a probability distribution. For example a 95% frequentist confidence interval is a range which will contain the true value 95% of the time if the analysis is repeated a large number of times for different data obtained under the same process[2]. In contrast a Bayesian 95% interval estimate (see Chapter 4) is a range with a 95% probability of containing the true value. This Bayesian version is arguably much easier to interpret.

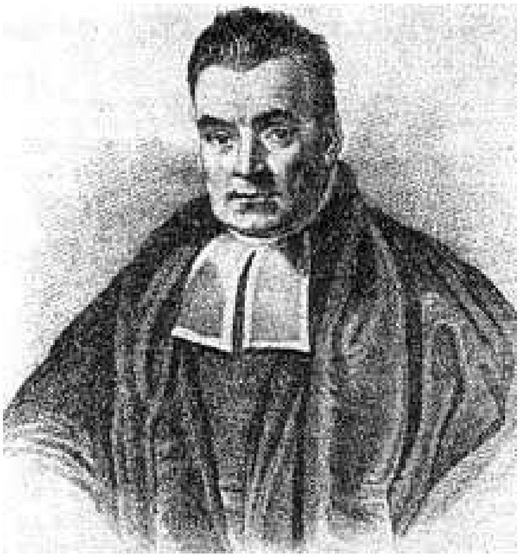
There has been a decades-long argument about whether Bayesian or frequentist methods are better. Both sides have some good points. For example, there is the argument we've just made about Bayesian interval estimates being easy to interpret. On the other hand one of the main arguments against the Bayesian approach is that it is *subjective*. This is because its results are based on expert opinions, and if we got a different opinion from another expert our results would change even though the data are the same! Bayesians have counter-arguments. For example many would argue that subjectivity is in fact a desirable feature, or that frequentist methods also have hidden subjective elements e.g. using 5% as a significance threshold rather than some other value. The philosophical debate seems unlikely to end any time soon!

In practice many problems are better suited to be tackled with methods from one approach or the other. For example Bayesian statistics was well suited to the Air France example as incorporating expert knowledge was particularly important in the solution. Therefore it's good to be familiar with both Bayesian and frequentist methods.

The history of Bayesian statistics

The reason why we use the word "Bayesian" is because Bayes' Theorem is crucial for statistical analysis if we adopt the Bayesian approach. Thomas Bayes (1702–1761; see picture below) was a Presbyterian minister in Tunbridge Wells, Kent. Bayes' solution to a problem in probability theory was presented in the *Essay towards Solving a Problem in the Doctrine of Chances*, published posthumously by his friend Richard Price in the *Philosophical Transactions of the Royal Society of London*. This paper gave us Bayes' Theorem. Its modern form is due to Pierre-Simon Laplace who generalised Bayes' work

[2]Here 95% is the *long-run frequency* that the interval contains the correct value. This kind of property is where the name "frequentist" comes from.



LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

Figure 1: Portrait believed to be of Thomas Bayes, and letter from Richard Price to the Philosophical Transactions of the Royal Society of London

and produced the first theory of Bayesian inference, which was referred to as "inverse probability". Laplace's theory had weaknesses[3] and in the early 20th century it was replaced by the more rigorous frequentist approach. R. A. Fisher was one of the leading creators of this approach. He was also the first to use the term "Bayesian statistics".

Nonetheless there was still some interest in Bayesian methods. For example, Alan Turing's group invented some new Bayesian methods to help break the Enigma code in World War II. Before and, increasingly, after the war a few researchers worked on developing a rigorous theory of Bayesian statistics.

Bruno de Finetti (1906–1985) was an Italian probabilist who developed his ideas on *subjective probability*[4] from the 1920s onwards, drawing upon ideas from H. Jeffreys, I.J. Good (one of the Enigma codebreakers) and B.O. Koopman. His classic book on the topic is *The Theory of Probability* (1974).

Dennis Lindley (1923–2013) was a leading British Bayesian. In his early career, he worked to find a mathematical basis for the subject of statistics. In 1954, Lindley met Leonard Savage and both found a deeper justification for statistics in Bayesian theory, turning into critics of the classical statistical inference they had hoped to justify.

Many of these theoretical advances are covered in this course. However, Bayesian statistics was still little used in practice. The difficulty was that applying Bayes theorem often

[3] For example it relied mainly on *flat priors*, which will be discussed and criticised in Chapter 3.
[4] Covered in Chapter 1.

Figure 2: Pierre-Simon Laplace (left) and R. A. Fisher (right)

became very mathematically challenging, resulting in problems that were too difficult or impossible to solve by hand.

Recent developments

In the 1990s a solution to these mathematical problems was developed, making use of the emergence of more powerful computers and "Markov chain Monte Carlo" (MCMC) numerical algorithms. If you take 3rd and 4th year Bayesian courses you will be introduced to these techniques. MCMC, and subsequent developments, have revolutionised the use of Bayesian Statistics, to the extent that today Bayesian data analyses are as popular as frequentist approaches and are routinely used in fields as diverse as artificial intelligence, biology, astrophysics and sociology.

Newcastle's contribution

Since 1980, the number of academic staff in Mathematics & Statistics at Newcastle publishing advanced research using Bayesian methods has increased dramatically. In the 1980s, there was only one Bayesian at Newcastle. Now there are at least 12. Our Bayesian research topics include environmental extremes, genetics, and modelling biological systems at a molecular level.