MAS2317/3317: Introduction to Bayesian Statistics

Dr. Lee Fawcett

Semester 2, 2014—2015

Contact details

My name: Dr. Lee Fawcett

Office: Room 2.07 Herschel Building

Phone: 0191 208 7228

Email: `lee.fawcett@ncl.ac.uk`

www: `www.mas.ncl.ac.uk/∼nlf8`

Timetable and Administrative arrangements

- **Lectures** are on Mondays at 11 (LT3) and Tuesdays at 12 (LT2).
- **Problems classes** are in ODD teaching weeks starting in teaching week 3. These take place on Thursdays at 9 in LT1. We will use this slot in week 1 as a lecture to get ahead in the notes.
- **Drop–in sessions** are at the same time and place in EVEN weeks starting in teaching week 4.
- **Computer practicals**: Wednesdays at 12 (Herschel full PC cluster). These happen in some EVEN weeks, and you will be notified of these well in advance.

Timetable and Administrative arrangements

So to summarise:

Activity	Day	Time	Room	Frequency
Lecture	Mon	11–12	Herschel: LT3	Every week
	Tues	12–1	Herschel: LT2	Every week
PC/ DI	Thurs	9–10	Herschel: LT1	PC odd weeks DI even weeks
Practical	Wed	12–1	Herschel: PC cluster	Some even weeks

Assessment

Assessment is by:

- End of semester exam in May/June (**80%**)
- In course assessment, including written solution to problems and computer practical work (**10%**)
- Mid-semester test (**10%**)
- No CBAs!

There will be four assignments, each one having a mixture of written and practical questions.

Solutions to the starred questions should be submitted by the dates given in your notes.

We will work through the unstarred questions in problems classes and practicals.

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.
  - **Listen and learn**
  - **Write down**
  - **Announcements**
- These notes and slides will be posted on the **course website** and/or BlackBoard after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- The course website can be found via the “Additional teaching information” link on the School’s webpage, or directly via:

  http://www.mas.ncl.ac.uk/∼nlf8

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

- I have scheduled an office hour for this course (Fridays from 1pm) where I will always be available to see students who need extra help/support with the course material.

  I also have an office hour for MAS1343 (most Thursdays at 3pm) that you are welcome to drop in to.

Preface to lecture notes:

The Bayesian paradigm

Statistics: an alternative approach

- Up until now, you have been taught Probability and Statistics according to a particular way of thinking
- The **Bayesian paradigm** offers another way of seeing things
- Your ideas about Probability and Statistics are deeply entrenched...perhaps so much so that at first, *Bayesian Statistics* might take you a while to grasp!
- Not because it's hard! Just because you have become so conditioned to think in a particular way!
- I want to broaden your horizons a bit!

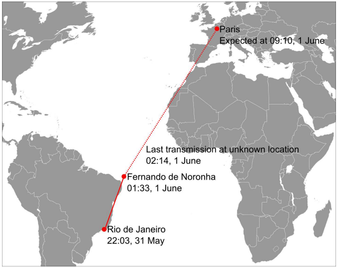
Bayes goes to war!

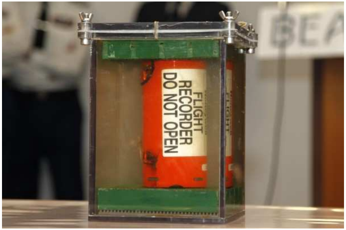
- The work of Thomas Bayes (see later) was published in 1764, 3 years after his death
- Over the course of the next 100–150 years, it received little attention
- In fact, some key figures in Statistics – e.g. **R.A. Fisher** – outrightly rejected the idea of Bayesian statistics
- By the start of WW2, Bayes' rule was **virtually taboo** in the world of Statistics!
- During WW2, some of the world's leading mathematicians resurrected Bayes' rule in deepest secrecy to crack the coded messages of the Germans

Bayes goes to war!

- **Alan Turing** – mathematician working at Bletchley Park
- Designed the 'bombe' – an electro–mechanical machine for testing every possible permutation of a message produced on the Enigma machine — could take up to **4 days** to decode a message
- New system: **Banburismus** – where Bayesian methods were used to quantify the belief in guesses of a stretch of letters in an Enigma message
- Certain permutations – unlikely to be the original message – were 'thrown out' before they were even tested
- Greatly reduced the time it took to crack Enigma codes

Air France 447





Air France 447

- Black boxes could be anywhere within an area of the South Atlantic the size of Switzerland (**6,500 square miles**)
- Mid–Atlantic ridge – between two tectonic plates – just as Mountainous as Switzerland!
- So remote – scientists have not yet charted the sea–bed
- Search method: sonar – detectors emitting sound waves which would bounce back once they hit something
- After two years of meticulously searching an area north of the plane's trajectory (after analysing debris drift) $\longrightarrow$ **nothing**

Air France 447

- April 2011: *Metron, Inc.*, of Retson, Virginia, hired to launch a **Bayesian review** of the entire search effort
- Included in the analysis were:
  - Data from 9 previous airline crashes involving loss of pilot control – reduced search area to **1,600 square miles**
  - **Expert opinions** on the credibility of the flight data
  - **Expert opinions** about whether or not the black box 'pingers' might have become damaged on impact
  - Positions/recovery times of bodies found drifting – **expert opinions** assigned to the reliability of this data because of the turbulent equatorial waters
  - **Expert information** from oceanographers: Sea state, visibility, underwater geography,...
- All information combined through **Bayes Theorem**: After one week, black boxes found!

Motivating example

School A and School B each have a football team. They have two matches against each other coming up in the next few weeks: a league match and a cup match.

What is the probability that School A wins both matches?

Equally likely outcomes?

One (probably not very realistic) approach might be to assume that, in each game, all outcomes are **equally likely**, i.e.

$$\begin{aligned}
\Pr(\text{School A wins}) &= \frac{1}{3} \qquad (= p) \\
\Pr(\text{School B wins}) &= \frac{1}{3} \qquad \text{and} \\
\Pr(\text{Draw}) &= \frac{1}{3}
\end{aligned}$$

This is often referred to as the **classical** interpretation of probability, and in this case would give

$$\Pr(\text{School A wins both}) = p \times p = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} = 0.111.$$

Similar examples: flipping (fair) coins, rolling (unloaded) dice, taking a random sample,...

Relative frequency approach?

Perhaps a better approach would be to consider what has happened historically when these two teams have played each other.

In fact, in their last 10 matches against each other, School A won 7 times.

Then we might use

$$\text{Pr(School A wins)} \quad = \quad \frac{7}{10} \qquad (= p).$$

This is known as the relative frequency or **frequentist** interpretation of probability, and would give

$$\text{Pr(School A wins both)} = p \times p = \frac{7}{10} \times \frac{7}{10} = \frac{49}{100} = 0.49.$$

Subjectivity?

You might argue that the second approach is more realistic.

However, you should notice that in both examples, once we have figured out how likely it is that School A wins a match against School B (p), this probability is assumed constant.

- What if the team that wins the first match gains confidence before the second meeting?
- What if they lost the last three consecutive matches?
- What if School A think the cup match is more important?
- What if we have some inside information: School A's top scorer will be missing from one or both matches?
- Enthusiastic supporters of School B might have their own (subjective) probabilities that their team will win, no matter what previous form suggests.

A third approach to probability

We have the **classical** approach to probability (equally likely outcomes).

We have the **frequentist** approach to probability (relative frequency).

And then we have the **subjective** approach to probability.

When we work within the Bayesian framework, we try to account for subjective notions of probability, instead of (artificially) assuming equally likely outcomes or solely looking at the outcomes of past ‘trials’ or ‘experiments’.

A third approach to probability

One way of doing this is to acknowledge that parameters in a statistical model (i.e. p in the above example) are not fixed constants.

Perhaps it is okay to assume that our parameter of interest can take on a whole range of values, reflecting the (perhaps subjective) beliefs of various parties?

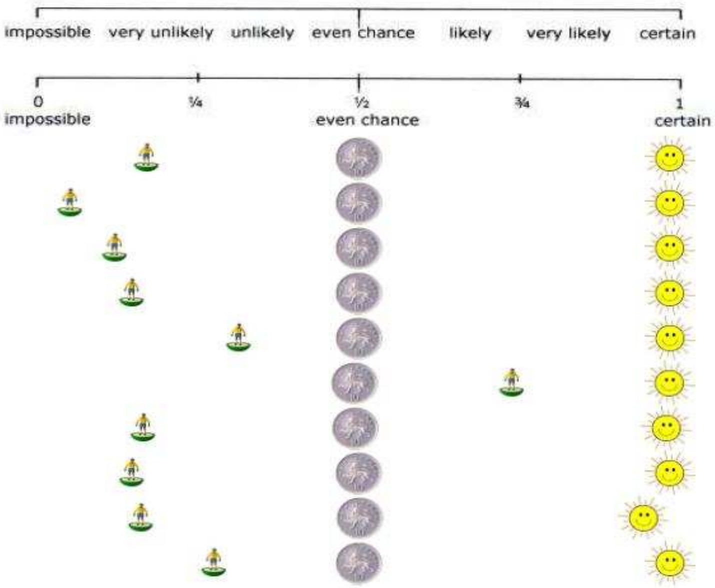
Year 7 – The probability scale

I was introduced to the concept of probability in year 7.

We were asked to place a sunshine counter, a ten pence piece and a Subbuteo figure on a ‘probability scale’ according to how likely we thought the following events were to occur:

- A: the sun will rise tomorrow;
- B: I get a head when I flip a coin;
- C: our school football team will win the cup this year.

"Results"



impossible
very unlikely
unlikely
even chance
likely
very likely
certain
0
¼
½
¾
1
impossible
even chance
certain

“Results”

Most of us understood there was an evens chance of getting a head.

Most of us also believed the sun was certain to rise tomorrow, although you can see that one pupil has expressed some doubt here!

However, we all had different opinions about where to place the footballer.

- Pupil 2 was on the school football team
- Most of us (thought we) knew our team were rubbish!
- Pupil 6 had inside information!

A Bayesian analysis would take into account the variability of the pupils’ beliefs. The Probability and Statistics that *you* are used to would not do this.

Probability and Statistics at School

The National Curriculum probability syllabus steers clear of events like C – we were never told how to think about interpreting subjective probabilities.

Teachers then have to focus on the **Classical** or **frequentist** interpretations of probability – hence all those examples about

- flipping (fair) coins
- rolling (unloaded) dice
- picking coloured balls at random from a bag
- drawing cards from a pack

In short: pretty uninspiring!

Probability and Statistics at School

In fairness, you *are* taught about

- the **language** of probability
- correlation and regression
- the Normal distribution

amongst other things, but let's face it: most applications of Probability and/or Statistics seem, at best, *contrived*, and *never* take into account **subjectivity**!

Probability and Statistics at University

First year: Not much better, but the aim of MAS1341/1342 is to get everyone up to the same level – basically A level standard.

You are introduced to **Statistical inference**: the process by which we attempt to *infer* things about the population based on information available in our sample.

- Estimation
- Confidence intervals
- Hypothesis tests

All taught within the **Classical**/**frequentist** framework!

Example: Body Mass Index

Suppose we are interested in whether or not there is any real difference between the Body Mass Index (BMI) of Maths & Stats students and Food & Human Nutrition students.

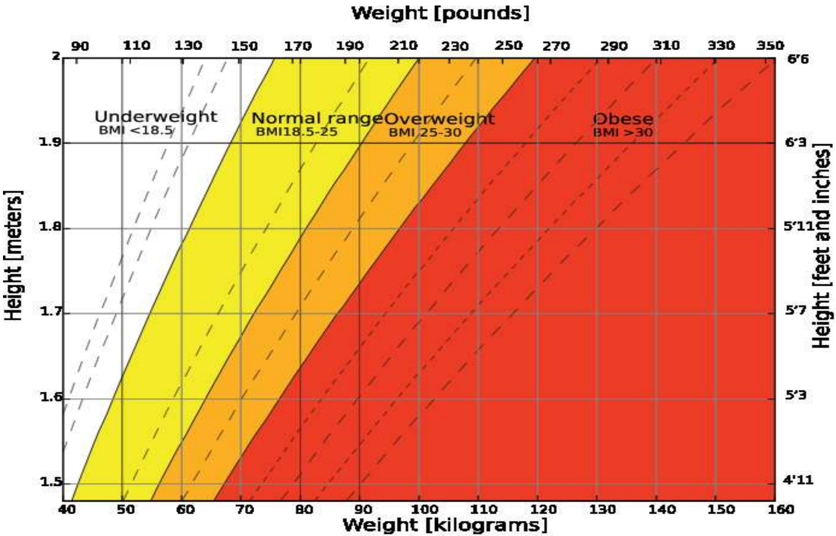
BMI is often used as a measure of 'fatness' or 'thinness': it is the ratio of a person's weight (measured in kg) to their squared height (measured in metres).

In a two–sample *t*–test, we often test the null hypothesis that there is no difference between the population means of the two groups:

$$H_0 : \mu_M = \mu_F,$$

How can we test this hypothesis?

Example: Body Mass Index



Example: Body Mass Index

Suppose we obtain the following BMI values for a random sample of students from both groups:

Maths & Stats	23.1	28.3	35.3	24.0	32.4	30.5	30.1	21.9	22.1
	29.9	27.2	33.2						
Food & Human Nutrition	25.5	27.3	21.0	23.3	19.1	22.2	29.5	27.5	23.2

From this we can obtain the summaries:

$$n_M = 12 \qquad \bar{x}_M = 28.17 \qquad s_M = 4.54$$

$$n_F = 9 \qquad \bar{x}_F = 24.29 \qquad s_F = 3.40$$

Example: Body Mass Index

The pooled standard deviation is

$$\begin{aligned} s &= \sqrt{\frac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}} \\ &= \sqrt{\frac{11 \times 4.54^2 + 8 \times 3.4^2}{12 + 9 - 2}} \\ &= 4.099 \end{aligned}$$

Then the test statistic is

$$\begin{aligned} t &= \frac{|\bar{x}_M - \bar{x}_F|}{s \times \sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \\ &= 2.147. \end{aligned}$$

Example: Body Mass Index

From t_{19}–tables, we get:

p–value	10%	5%	1%
critical value	1.729	2.093	2.861

Our test statistic $t = 2.147$ indicates that our p–value is less than 5%:

- We have a significant result at the 5% level of significance
- We **reject** H_0
- We conclude that there *is* a significant difference in BMI between M&S and FHN students

Notice from R that the *exact* p–value is 0.004488 = 4.488%.

- As we have already concluded, we would **reject** H_0 at the 5% level;
- However, if we work at the 1% level of significance, we would **retain** H_0!

What are we to do? Convention tells us to work at the 5% level, but why? Some practitioners work at the 1% level!

Example: Body Mass Index

In MAS1341/1342 and MAS2304 you were introduced to **Bayes' Theorem**.

As we shall see, Bayes' Theorem gives us a way of combining subjective assessments with observed data.

For example, what if – prior to observing the data – we believed that Food & Human Nutrition students were likely to have a **considerably lower BMI** than Maths & Stats students?

What if, from a previous study, we knew that Maths & Stats students were quite likely to have a **BMI somewhere between 25 and 33**?

Example: Body Mass Index

Would it not be sensible to build this information into our analysis before forming our conclusions using the data alone?

This could be a good idea – we have relatively small samples, which could be biased!

A **Bayesian analysis** can do this!

Some people argue against this on the grounds that it is subjective... But as we have just seen, the usual approach to hypothesis testing is also subjective!

Not only are hypothesis tests tricky to interpret... but so are **confidence intervals**.

Example: Body Mass Index

Obtain a 95% confidence interval for the population mean BMI for Maths & Stats students.

We use

$$\bar{x}_M \pm t\,\frac{s_M}{\sqrt{n_M}}, \qquad \text{giving}$$

$$28.17 \quad \pm \quad 2.201 \times \frac{4.54}{\sqrt{12}} \qquad \text{i.e.}$$

$$28.17 \quad \pm \quad 2.743,$$

giving (25.28, 31.05).

Example: Body Mass Index

What is $\Pr(25.28 < \mu_M < 31.05)$?

NOT 0.95!! In the frequentist approach, population parameters (in this case μ_M) are NOT random variables!

This means μ_M is a fixed (but unknown) quantity. So it's either *in* the interval, or it's *not*, i.e.

$$\begin{aligned} \Pr(25.28 < \mu_M < 31.05) &= 1 \qquad \text{or} \\ \Pr(25.28 < \mu_M < 31.05) &= 0. \end{aligned}$$

Nothing else!

Equivalent intervals in the **Bayesian framework** have more natural interpretations!

Frequentist/Subjectivist debate

One of the main arguments *against* working within the Bayesian paradigm is that it is **subjective**.

Although we have not yet thought about *how* the Bayesian framework combines subjective assessments with the data, we have said that this is what it does.

Surely we should strive to be as *objective* as we can?

Actually, to be able to incorporate a person's beliefs into an analysis of sample data is surely the **right thing to do** — especially if that person is an expert!

This is in line with how scientific experiments are carried out:

- The experimenter usually **knows something**;
- She then carries out the experiment in which **data are collected**;
- the experimenter then **updates her beliefs** from these results.

Frequentist/Subjectivist debate

In other words, the data are used to refine what the experimenter knows.

And that is what MAS2317/3317 is all about – we will consider how

- **prior beliefs** can be combined with
- **experimental data** to form
- **posterior beliefs** – which include both the prior knowledge and what we have learnt from the data

Bayesian Statistics: leading players

Thomas Bayes (1702–1761)

- Presbyterian minister in Tunbridge Wells, Kent
- Bayes' solution to a problem of "inverse probability" was presented in the *Essay towards Solving a Problem in the Doctrine of Chances* (1764)
- Work published posthumously by his friend Richard Price in the *Philosophical Transactions of the Royal Society of London*.
- This work gives us **Bayes' Theorem**, a key result on which most of the work in this course rests.

Bayesian Statistics: leading players

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-

Bruno di Finetti (1906–1985)

- Italian mathematical probabilist, developed his ideas on subjective probability in the 1920s.
- Famed for saying

  "Probability does not exist"

- By this, he meant it has no *objective* existence – i.e. it is not a feature of the world around us.
- It is a measure of degree–of–belief – your belief, my belief, someone else's belief – all of which could be different.

Dennis Lindley (1923—2013)

- Worked hard to find a mathematical basis for the subject of Statistics
- With Leonard Savage, he found a deeper justification for Statistics in **Bayesian theory**
- Turned into a critic of the classical statistical inference he had hoped to justify
- Quoted as saying:

"Uncertainty is a personal matter. It is not the uncertainty, but your uncertainty"

Recent developments

One of the difficulties in early applications of Bayesian methodology was the maths:

Combining prior beliefs with a probability model for the data often resulted in maths that was just **too difficult/impossible to solve by hand**.

Recent developments

Then – in the early 1990s – a technique called "**Markov chain Monte Carlo**" (MCMC for short) was developed.

MCMC is a computer–intensive simulation–based procedure that gets around the problem of hard maths.

If you choose to take **MAS3321: Bayesian Inference** next year, you will be introduced to this technique.

MCMC has **revolutionised the use of Bayesian Statistics**, to the extent that Bayesian data analyses are now routinely used by non–Statisticians in fields as diverse as biology, civil engineering and sociology.

Newcastle’s contribution

Number of academic staff publishing advanced research using Bayesian methods:

- 1985: **1** (Professor Boys)
- 2012: **9**
  - *Systems Biology:* (Boys, Farrow, Gillespie, Golightly, Henderson and Wilkinson)
  - *Environmental Extremes:* (Fawcett, Walshaw)

Newcastle's contribution: Environmental Extremes

Use Statistical methods to help plan for environmental extremes:

- **Hurricane–induced waves**: how high should we build a sea–wall to protect New Orleans against another *Hurricane Katrina*?
- **Extreme wind speeds**: how strong should we design buildings so they will not fail in storm–strength wind speeds?
- **Extreme cold spells**: How much fuel should we stockpile to cater for an extremely cold winter?

Newcastle's contribution: Environmental Extremes

We work within the Bayesian framework to combine

- expert information from hydrologists/oceanographers, with
- extreme rainfall data/hurricane–induced sea surge data,

to help estimate how likely an **extreme flooding event is to occur**, or to aid the **design of sea wall defences** in storm–prone regions.

If you're interested, you can see my web–page for more details.

Quick quiz

Suppose A, θ and μ_M are constants and Z is a random variable, such that $A = 5$, $\theta = 10$, $\mu_M = 33.5$ and $Z \sim N(0, 1)$.

Write down

1. $\Pr(Z > 0)$
2. $\Pr(-1.96 < Z < 1.96)$
3. $\Pr(2 < A < 4)$
4. $\Pr(7 < \theta < 12)$
5. $\Pr(25.28 < \mu_M < 31.05)$