MAS2317/3317

Introduction to Bayesian Statistics

Semester 2, 2014—2015

Lecturer: Dr. Lee Fawcett

School of Mathematics & Statistics

MAS2317/3317: Introduction to Bayesian Statistics

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 11am (LT3) and Tuesdays at 12pm (LT2).
- *Problems classes* are in odd teaching weeks and will start in teaching week 3. These will take place on Thursdays at 9am in LT1.
- *Drop–in sessions* are at the same time and place as the problems classes and take place in even teaching weeks, starting in teaching week 4.
- *Computer practicals* will take place in teaching weeks 4, 6 (before Easter), 8 and 10 (after Easter). These will be on Wednesdays at 12pm in the Herschel PC cluster. A reminder will be given the week before!
- *Office hours:* I will be available to see MAS2317/3317 students on Fridays at 1pm (my office).

What?	When?		Where?	How often?
Lecture	Mon	11–12	Herschel: LT3	Every week
	Tues	12–1	Herschel: LT2	Every week
Problems class/ Drop–in session	Thurs	9–10	Herschel: LT1	PC odd weeks DI even weeks
Computer practical	Wed	12–1	Herschel: PC cluster	Some even weeks
Office hours	Fri	from 1pm	My office	Every week

Assessment

- **80%** exam in May/June
- **20%** in course assessment:
  - Mid–semester test (10%): **Thursday 12th March, 9am**
  - 4 assignments (10%), each one having a mixture of written questions and practical work. Deadlines – **Tuesdays** at **4pm** on the following dates:

    24th February, **10th March**, **21st April**, **5th May**.

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.
- 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 model solutions to assignment questions, supplementary handouts etc.
- The course website can be found at:

  http://www.mas.ncl.ac.uk/∼nlf8

- 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 handout 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

Motivating example

Much of the material in this course will show Statistics in a very different light to previous courses in the subject. 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. So deeply entrenched are your ideas about Probability and Statistics that, at first, it might take you a while to get to grips with some concepts in this course. That's not because the material in MAS2317/3317 is particularly difficult, it's just because you have become so conditioned to think in a particular way, and now we want to broaden your horizons a bit! We'll get things going with a very simple 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?

One (probably not very realistic) approach might be to assume that, in each game, all outcomes are equally likely: perhaps Pr(School A wins) $= \frac{1}{3}(= p)$, Pr(School B wins) $= \frac{1}{3}$ and Pr(Draw) $= \frac{1}{3}$. This is often referred to as the classical interpretation of probability, and in this case would give

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

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 Pr(School A wins) $= \frac{7}{10}(= p)$; this is known as the relative frequency interpretation of probability, and would give

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

Although you might argue that the second approach is more realistic, 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 ($\frac{1}{3}$ or $\frac{7}{10}$) for both matches that are about to be played in the near future.

- What if the team that wins the first match gains confidence before the second meeting?

- We know that School A have won seven out of the last ten matches, but 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 – for example, we know that 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 (and likewise supporters of School A), no matter what previous form suggests.

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*. 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. We do this in the Bayesian framework.

Your experience of probability and statistics so far?

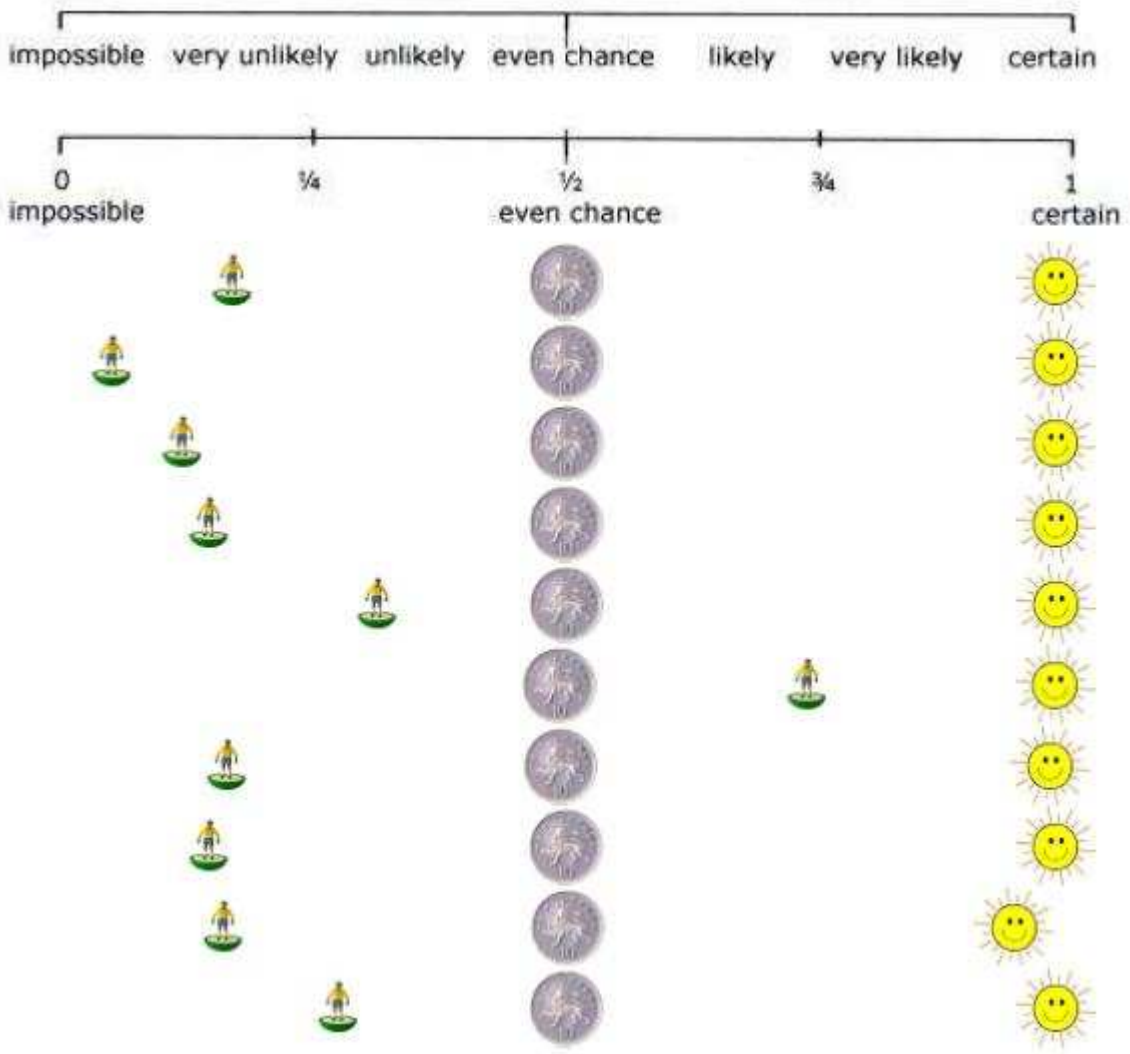
Year 7 – The probability scale

I was first introduced to the concept of probability at school 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.

Obviously, I can't remember the exact results (this was a long time ago!), but the diagram overleaf is fairly similar to what was observed. The teacher told us that we should assume the coin in B was 'fair', and so we all understood that 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 how likely it was that our School football team would win the cup. Although most of us believed it was unlikely, we all believed this to different degrees, and some had inside information: David Graham (pupil 2) was on the football team, and he knew more than most just how bad the team were! Conversely, Victoria Baker (pupil 6) actually thought the team were in with a decent chance. Later that term, Victoria's cousin joined our School. He was a brilliant football player, and made the first team – we made it to the final of the cup. Perhaps Victoria was the one with inside information! In

any calculations (think about those in the previous section), a Bayesian analysis would attempt to take into account the variability of the pupils' beliefs. The Probability and Statistics that you are used to would not do this.



Some results of the Year 7 experiment

Years 8–11

When I was at school, all probability classes in subsequent years seemed to steer clear of events like C given above, and why my friends and I placed our little plastic footballer at different points along the probability scale. The probability I studied from year 8 until about year 11 focused on classical or frequentist views of probability; hence came all of those examples about flipping coins, rolling dice, picking cards from a pack, picking different coloured balls from a bag, looking at defective items off a production line... I'm sure you're all sick of examples like these! And I think things like this may have led many of you to find probability dull by the time you came to University. During this time you would have been introduced to the notion of independence between two (or more) events,

and how this leads to the multiplication rule for probability, as well as the addition rule for events that are mutually exclusive. What had been covered in statistics during this time? Well, you'd been taught about taking samples; how to summarise sample data with graphs and tables (tally charts, pie charts, bar charts, time series plots, histograms, etc.); and how to summarise sample data numerically (mean, median, mode, standard deviation, etc.). Exciting stuff.

A Level courses: S1 and S2

As many of you know from S1, A level courses in probability and statistics can be quite dull. Although you are introduced to the notion of conditional probability, you continue to be taught probability in the classical or frequentist sense, with syllabuses obsessing over events that are equally likely or whose probabilities are obtained via long run relative frequencies. You are, of course, introduced to the Normal distribution and correlation and regression – both extremely important topics in Statistics – but on the whole, prospective students at our Maths & Stats open days tell me that they find Statistics really boring. I'm not surprised!

Probability and Statistics at University so far

To some extent, Stage 1 Statistics courses at University are not much better – much of the material here is no more than A Level standard. In first year and in some courses in second year the A level syllabus is expanded upon; new probability distributions are introduced and the method of maximum likelihood for estimating parameters in a statistical model is developed. Statistical inference (e.g. MAS2302) is largely based around using sample data (and nothing more) to draw conclusions about what happens in the population: here, you will have considered things like hypothesis tests and confidence intervals.

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. Here, we might write this as:

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

where μ_M and μ_F denote the population mean BMI of Maths & Stats students and Food & Human Nutrition students, respectively. How can we test this hypothesis? We would usually take a sample of students from both populations, find the mean BMI of students in both samples, and then compare the sample means with each other via

the test statistic. 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:

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

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

where n, $\bar{x}$ and s denote the sample size, the sample mean and sample standard deviation, respectively, and M and F are labels for Maths & Stats and Food & Human Nutrition, respectively (as before).

Of course, we can do this more 'automatically' using R. The command `t.test` can be used to perform a two–sample t–test. After storing the BMI values above in `MAS` and `FHN` for Maths & Stats and Food & Human Nutrition (respectively):

```
> MAS=c(23.1,28.3,35.3,...)
> FHN=c(25.5,27.3,21.0,...)
```

we would type:

```
> t.test(MAS,FHN,var.equal=TRUE)
```

This gives the following output:

```
Two Sample t-test
data: MAS and FHN
t = 2.1473, df = 19, p-value = 0.04488
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.09808992 7.65746564
sample estimates:
mean of x mean of y
28.16667 24.28889
```

Notice we have a p–value of 0.04488. What does this mean? This depends on the level of significance we choose to work at. The 'default' is often 5%; our p–value is 4.488%, and so we would reject H_0 at the 5% level of significance, concluding that there is sufficient evidence to suggest a real difference between the mean BMI of Maths & Stats and Food & Human Nutrition students (notice that it looks like Maths & Stats students have, on average, a higher BMI than Food & Human Nutrition students: 28.17 compared to 24.29, respectively). However, if we work at the 1% level of significance, we would actually retain H_0, concluding that there is insufficient evidence to suggest a difference.

Similarly, the 95%/99% confidence intervals for the mean differences are $(0.098, 7.657)$ and $(-1.289, 9.044)$ respectively: the second includes zero, the first does not. How should we proceed?

In MAS1341/1342 and MAS2304 you were introduced to Bayes' Theorem. As we shall see in this course, 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, on average, 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? A Bayesian analysis of the BMI data would use Bayes' Theorem to include such *prior information* about how likely it is that μ_M and μ_F take certain values, and would use this in conjunction with the observed sample data given above. This could be a good idea – we have relatively small samples, which could be biased! Those averse

to subjective notions of probability might argue against this on the grounds that what one person might think is a plausible range of values for μ might be different to what another person thinks. Notice, however, that the frequentist approach is also subjective: for example, the analyst chooses the significance level to work at and, as we have seen here, this can lead to opposing conclusions. In the classical setting, confidence intervals are also quite tricky to interpret. For example, we can find a 95% confidence interval for the population mean μ using

$$\bar{x} \pm t_{\nu,0.975} \times \frac{s}{\sqrt{n}},$$

where t is a critical value from Student's t distribution on ν degrees of freedom.

Obtain a 95% confidence interval for the population mean BMI for MAS students.

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

The frequentist/subjectivist debate

As we have already discussed, 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. In fact, this is in line with how scientific experiments are carried out: the experimenter usually knows something; s/he then carries out the experiment in which data are collected; the experimenter then learns from these results. 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. Leaving aside the scientific argument in favour of a Bayesian analysis, surely the assessment of a classical hypothesis test or confidence interval is just as subjective anyway? For example, we need to to decide on the level of significance we are going to work at and, as highlighted by the BMI example, any conclusions drawn can be sensitive to this choice.

Still, exactly how we quantify the beliefs of experts in a meaningful way, using the language of probability, can be difficult. Various tools and approaches have been developed to this end, some of which you will be introduced to in this course. This can often be a challenge for "Bayesians" (or "Subjectivists" as they are often called); indeed, this is probably still the main weapon used by "Frequentists" in their fight against the use of Bayesian methodology. However, although even Bayesians themselves might acknowledge it can sometimes be a challenge to incorporate people's personal beliefs into a statistical analysis, in this course you will learn that – with some thought – not only is it possible, but it is the right thing to do.

Philosophical debates between Bayesians and Frequentists rage on. However, Bayesians are gaining ground and, as we shall discuss in the next section, Bayesian techniques now permeate many fields. Some Bayesians argue that so natural is the Bayesian approach to statistical analysis that this is the way all statistics should be taught. All standard techniques – for example, hypothesis tests, confidence intervals – have a Bayesian analogue, but the Bayesian analogue is far more intuitive: for instance, a 95% Bayesian credible interval (analogous to a 95% confidence interval in frequentist statistics) does contain the true parameter value with probability 0.95. Some authors now even argue that Bayesian methods should be taught to non–mathematics undergraduates who need basic statistics training (e.g. biology students or business students who take basic statistics courses). It has also been argued that the more time we spend teaching out–dated statistics, the more likely it is that the discipline of Statistics will decline in the academic world: Statisticians don't have a lock on Bayesian ideas, and if there is a significant movement for other departments to teach Bayesian statistics "in–house" (e.g. Computing Science), Statistics will lose out big time. The inclusion of MAS2317/3317 to the Mathematics & Statistics undergraduate program at Newcastle is certainly a step in the right direction!

The development of Bayesian statistics

Leading players

The reason why we use the word "Bayesian" is because Bayes' Theorem (see MAS1342 and MAS2304) is crucial for many problems if we adopt the Bayesian philosophy. Thomas Bayes (1702–1761; see picture below) was a 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*, published posthumously by his friend Richard Price in the *Philosophical Transactions of the Royal Society of London*. This paper gave us Bayes' Theorem.

Bruno di Finetti (1906–1985) was an Italian probabilist who developed his ideas on subjective probability in the 1920s, drawing upon ideas from H. Jeffreys, I.J. Good and B.O. Koopman. In his classic book *The Theory of Probability* (1974) he is famed for saying that *"Probability does not exist"*. By this, di Finetti 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.

In his early career, Dennis Lindley (1923–2013) 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. Lindley is quoted as saying: *"Uncertainty is a personal matter. It is not the uncertainty, but your uncertainty"*.

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-

Portrait of Thomas Bayes, and letter from Richard Price to the Philosophical Transactions of the Royal Society of London

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. One way around this was to simplify the model formulations, but this often led to unrealistic modelling assumptions. 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 the hard maths, and 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

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: Professor Boys. Now, there are 9 Bayesians: Boys, Farrow, Fawcett, Gillespie, Golightly, Henderson, Shi, Walshaw and Wilkinson. Members of this group publish in two main areas: Systems Biology & Bioinformatics (Boys, Farrow, Gillespie, Golightly, Henderson and Wilkinson) and Environmental Extremes (Fawcett and Walshaw). In the area of Environmental Extremes, Dr. Walshaw and I have worked within the Bayesian framework to combine expert information from hydrologists with extreme rainfall data to help estimate how likely an extreme flooding event is to occur; we have also worked with oceanographers and wind engineers to aid the design of sea wall defences in storm–prone regions; those who are interested can see my web–page for more details (`www.mas.ncl.ac.uk/~nlf8/publications`).

Other recent and interesting applications

As we discussed earlier, recent advances in statistical computation have made it possible to apply Bayesian techniques to almost all areas of life:

- Internet shopping – e.g. Amazon recommendations
- Medical diagnoses: Gill, C.J., Sabin, L. and Schmid, C.H. (2005). Why Clinicians are Natural Bayesians. *British Medical Journal*, **330**, pp. 1080–1083
- Sports betting – e.g. Sports Betting Group: `http:sportsbettinggroup.com/`
- Law – weighing evidence, e.g. paternity tests: Vicard, P. and Dawid, A.P. (2004). Paternity analysis in special fatherless cases without direct testing of alleged father. *Forensic Science International*, **163**, pp. 158–160