STAT40850
Bayesian Analysis (online)

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

January 2016

About the course:

Content — The Bayesian approach to statistical inference uses probability to model the uncertainty about all unknown quantities.

This allows for uncertainty to be accounted for in a coherent manner. The emphasis of this course will be on applying this approach in statistical inference.

Lectures and Labs

- The module runs over the 12 weeks of Semester 2.
  - Week 1 starts on Monday 25th January.
  - There is a two week break between Week 7 and Week 8 (14th Mar-27th Mar) where there will not be lecture material posted.
  - Week 12 ends on Friday the 29th April.

- There will be lecture material posted every week in the form of lecture videos ans screencasts.

- Tutorials and Labs begin in Week 3. Each assignment will be due by 23:59 on the following Monday night/Tuesday morning e.g. if an assignment or Lab is posted at the start of Week 4, it will be due on Monday night the 22nd Feb before 23:59.

Assessment

- 20% of your final grade comes from Continuous Assessment (lab and tutorial sheets)
- The remaining 80% of your final grade comes from a 2-hour exam taken at the end of the module.
- We will use the UCD School of Mathematics and Statistics grading scale: `mathsci.ucd.ie/tl/grading/en06`.
- However, this is only a guide and the extern examiners may shift the bands at exam correction time.
- If you have any queries, please send me an email.

Disability Support Service

- If you are registered with the Disability Support Service in UCD please let me know.

Plagiarism

- The work you submit for this module must be your own.
- Make sure you are familiar with the UCD plagiarism policy: `www.ucd.ie/t4cms/Plagiarism_Policy_Academic_Policy_2005.pdf`
- Students caught plagiarising or making their answers available to others will be given zero for the continuous assessment part of this module.

Textbooks

- **Bayesian Statistics: An Introduction** by Peter M. Lee published by Wiley.
- **Bayesian Data Analysis** by Andrew Gelman, John B. Carlin, Hal S. Stern, David B.Dunson, Aki Vehtari, Donald B Rubin
  published by Chapman and Hall.

Lecturing format

- Best to print out slides, have notes in front of you.
- Please write additional notes.
- Extra notes and examples will be filled in on the board - keep these together with your notes.
- Mixture of theory, practical examples and computer sessions.
- Always have a calculator ready.
- Lecture notes/labs/tutorial sheets will be on Blackboard.
- Always bring the notes for the relevant class and the classes before.
- Please ask questions.

Any questions before we start?

What is this course about?

- Bayesian statistics is concerned with making *inferences* from data.
- It differs from standard *frequentist* inference in that external or *prior* information can be explicitly included in the analysis.
- It also differs in its definition of probability.
- In this course, we will study
    - the philosophy behind Bayesian statistics
    - how to analyse data in a Bayesian fashion
    - the differences between frequentist and Bayesian statistics
    - How to fit Bayesian models on the computer.

Course objectives

- By the end of this course, you should be able to create and fit Bayesian models in a wide variety of situations.
- Hopefully you will move away from thinking about statistics as a set of hypothesis tests and more as a principled way of comparing statistical models.

Thanks

- Thanks are due in no small part to Dr Andrew Parnell and Prof. Nial Friel who taught this module in previous years.
- I will be making use of a lot of their material.
- Errors and omissions my own!

Why Bayes?

Figure: Rev. Thomas Bayes (1701-61)

Bayes' Theorem

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

Thomas Bayes (1763), *An essay towards solving a problem in the doctrine of chances*, Philosophical Transactions of the Royal Society, 370-418.

Conditional probability

- Being comfortable with conditional probabilities is vital to being good at Bayesian statistics!
- From Bayes' Theorem, we can show the following for events A, B and C.

Results following from Bayes' Theorem

$$P(A|B,C) = \frac{P(A,B|C)}{P(B|C)} \tag{1}$$

$$P(A,B,C) = P(A|B,C)P(B|C)P(C) \tag{2}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

Equation 3 allows us to invert the conditioning.

Further extension of Bayes' Theorem

Let A_i $i = 1, \ldots, n$ be a series of exclusive and exhaustive events
i.e. $P(A_i, A_j) = 0$ if $i \neq j$ and $\sum_{i=1}^{n} P(A_i) = 1$.
Now:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Substituting this result into Equation 3 we have:

Bayes Theorem for multi-outcome events

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Thus we can use Bayes' theorem to reverse the conditioning for events where there are lots of different outcomes.

Adapted from Lee P9

Example

Bayesian statistics and the law In the case of Connecticut vs Teal, a case of alleged discrimination on the basis of race was considered. The case was based on a test to determine eligibility of promotion. 307 people took the test, 48 of whom were black, and 259 were white. Of the black applicants taking the test, 38 passed. Of the white applicants taking the test, 206 passed. Is there evidence of discrimination?