Class 1: An introduction to Bayesian Hierarchical Modelling

Andrew Parnell
andrew.parnell@ucd.ie

Learning outcomes

- Understand the terms prior, likelihood and posterior
- Know what a posterior probability distribution is, and why we take samples from it
- Know how to formulate of a linear regression model in a Bayesian format
- Be able to suggest appropriate prior distributions for different situations

How this course works

- This course lives on GitHub, which means anyone can see the slides, code, etc, and make comments on it
- The timetable html document provides links to all the pdf slides and practicals
- The slides and the practicals are all written in `Rmarkdown` format, which means you can load them up in Rstudio and see how everything was created
- Let me know if you spot mistakes, as these can be easily updated on the GitHub page

Course format and other details

- Lectures will take place in the morning, practical classes in the afternoon
- We will finish earlier on Thursday for a mini-trip
- Please ask lots of questions

Who was Bayes?

An essay towards solving a problem on the doctrine of chances (1763)

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

What is Bayesian statistics?

- Bayesian statistics is based on an interpretation of Bayes' theorem
- All quantities are divided up into *data* (i.e. things which have been observed) and *parameters* (i.e. things which haven't been observed)
- We use Bayes' interpretation of the theorem to get the *posterior probability distribution*, the probability of the unobserved given the observed
- Used now in almost all areas of statistical application (finance, medicine, environmetrics, gambling, etc, etc)

Why Bayes?

The Bayesian approach has numerous advantages:

- It's easier to build complex models and to analyse the parameters you want directly
- We automatically obtain the best parameter estimates and their uncertainty from the posterior samples
- It allows us to get away from (terrible) null hypothesis testing and *p*-values

Bayes theorem in english

Bayes' theorem can be written in words as:

posterior is proportional to likelihood times prior

. . . or . . .

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Each of the three terms *posterior*, *likelihood*, and *prior* are *probability distributions* (pdfs).
In a Bayesian model, every item of interest is either data (which we will write as x) or parameters (which we will write as θ). Often the parameters are divided up into those of interest, and other *nuisance parameters*

Bayes theorem in maths

Bayes' equation is usually written mathematically as:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta)$$

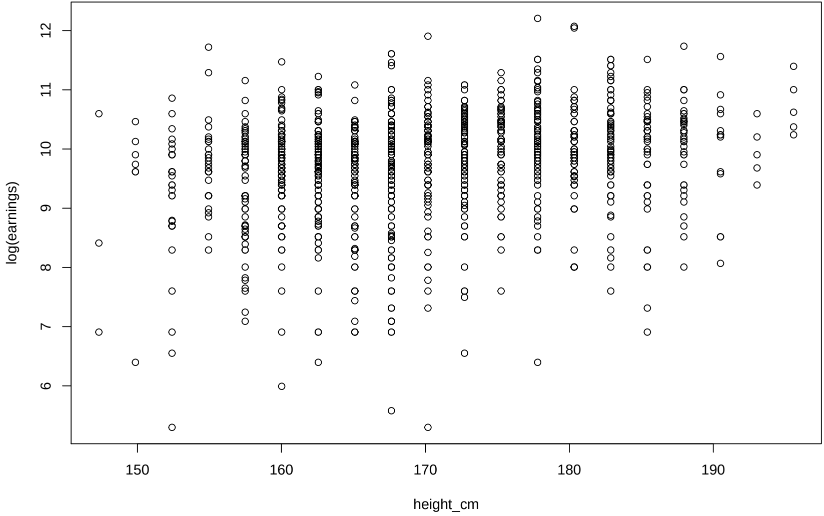
or, more fully:

$$p(\theta|x) = \frac{p(x|\theta) \times p(\theta)}{p(x)}$$

- The *posterior* is the probability of the parameters given the data
- The *likelihood* is the probability of observing the data given the parameters (unknowns)
- The *prior* represents external knowledge about the parameters

A very simple linear regression example

Suppose you had some data that looked like this:



What you are used to doing

```
model = lm(log(earnings) ~ height_cm, data = dat)
summary(model)
```

```
##
## Call:
## lm(formula = log(earnings) ~ height_cm, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4209 -0.3975  0.1394  0.5833  2.3536
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.778506   0.450927  12.815   <2e-16 ***
## height_cm   0.023156   0.002649   8.743   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8931 on 1190 degrees of freedom
## Multiple R-squared:  0.06035,    Adjusted R-squared:  0.05957
## F-statistic: 76.44 on 1 and 1190 DF,  p-value: < 2.2e-16
```

From parameters to data

Using prior information

A very basic Bayesian model with R code

Understanding the different parts of a Bayesian model

Lots of probability distributions

Choosing a likelihood and a prior

An example: linear regression

Simulating from the prior and the likelihood

Posterior computation in JAGS

Posterior computation in Stan

Stan vs JAGS

Calculating the posterior vs sampling from it

Things you can do with posterior samples

Summary so far: for and against Bayes

How to create Bayesian models: a general recipe (start with the data, fit it into a framework, LR, GLM, TS, then look at the parameters, and think of what priors are suitable)

Checking assumptions (e.g. residuals)

The danger of vague priors

Replication in Science and the horror of p-values

Example 1: 8 Schools

Example 2: Earnings data - https://raw.githubusercontent.com/stan-dev/example-models/master/ARM/Ch.6/earnings2.data.R A linear hierarchical model with Gaussian errors

Example 3: Swiss Willow Tit data - `http://www.mbr-pwrc.usgs.gov/pubanalysis/roylebook/wtmatrix.csv`. A logistic regression model with non-linear covariates

General tips: build one model for all the data, use informative priors, check your model