STAT40380/STAT40390/STAT40850 Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

February 2016

Catch up

We already know how to do a number of Bayesian tasks:

- Specifying a likelihood and prior distribution
- Calculating a posterior distribution up to the constant of proportionality
- Creating a conjugate prior for distributions in the exponential family
- Creating a Jeffreys' prior for parameters where we have little or no information
- Summarising posterior distributions with credible intervals and highest density ranges

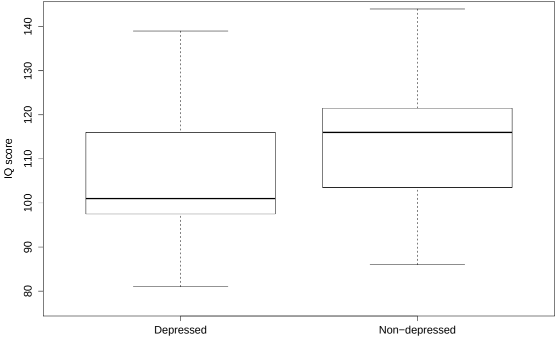
Today we will start to look at how to test hypotheses using Bayesian techniques

Traditional frequentist hypothesis testing

Example

Suppose we wish to run a simple 2-sample t-test on some data concerning the IQ scores of children of $n = 17$ depressed (x) mothers and $m = 78$ non-depressed (y) mothers.

$$\bar{x} = 107.6, \bar{y} = 113.6, \text{s.d.}(x) = 16.434, \ \text{s.d.}(x) = 12.364$$



Frequentist hypothesis testing 2

Usual steps

- Decide on null and alternative hypotheses, eg $H_0 : mu_x = \mu_y$ vs $H_1 : mu_y > \mu_x$
- Decide upon a significance level, eg $\alpha = 0.05$
- Assume the null hypothesis is true, choose a test and calculate a test statistic, eg the two-sample t-test with equal variances:

$$T = \frac{\bar{y} - \bar{x}}{S_{pooled}\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{113.6 - 107.6}{13.03\sqrt{\frac{1}{78} + \frac{1}{17}}} = 1.72 \sim t_{93}$$

- Find the appropriate tail value of the test statistics's distribution, eg $t_{93,0.95} = 1.661$
- Decide whether test statistic lies within critical region, and hence whether to reject H_0 or not. Here we do reject H_0
- Calculate a p-value as $P(T > t)$, eg here $p = 0.044$

Some criticisms of frequentist hypothesis testing

- The null and alternative hypotheses are often unrealistic, or strongly biased in favour of one over the other
- The significance level (almost always $\alpha = 0.05$) is somewhat arbitrary
- The p-value does not measure the probability that the null or alternative hypothesis is true. Recall the definition of a p-value:

  > *The probability of obtaining a test statistic at least as extrem as that observed, assuming the null hypothesis is true*

- Worse, the p-value can simply be changed (increased or decreased) with changes to the assumed model (eg normality in the t-test example)
- Even worse still, the p=value is *guaranteed* to be below α providing you collect enough data.

Some more criticisms of frequentist hypothesis testing

- The p-value is based around the probability distribution of the data given the parameters (eg the hypotheses). This is the same error of the transconditional we met in Lecture 1 and Tutorial sheet 1.
- The p-value is calculated, not just on observations we have made but on observations we might have made but did not.
- Finally a quote from Gelman et all (p176)

  "The relevant goal is not to ask the question 'Do the data come from the assumed model?' (to which the answer is almost always no), but to quantify the discrepencies between data and model, and assess whether they could have arisen by chance, under the model's own assumptions."

A short (defence) of the frequentist approach

- Since its discovery the classical approach has identified and confirmed many important scientific findings (see, eg smoking and lung cancer)
- In many situations, the error of the transconditional will be small, so different approaches will still yield similar results
- In some cases, the errors in hypothesis tests can be reduced by considering the power of the test
- Many eminent statisticians still prefer the classical approach (see eg Sir David Cox, John Nelder etc)

A Bayesian approach to hypothesis testing

- We want to directly answer the question of interest: *What is the probability that the hypothesis is true given the data we have observed?*
- To proceed, we will create posterior probabilities of the sort: $p(H_0 \text{ true } |\mathbf{x})$
- We will think of different hypotheses corresponding to different models, so $p(H_0 \text{ true } |\mathbf{x})$ is equivalent to *the probability that the model $\mathcal{M}_0$ (corresponding to hypothesis H_0) is true given the data*
- Key idea: *use Bayes' theorem to calculate the posterior probability of a model being true given the data:*

$$p(\mathcal{M}_0|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{M}_0)p(\mathcal{M}_0)}{p(\mathbf{x})}$$

More on Bayesian hypothesis testing

- We have already met $p(\mathbf{x}|\mathcal{M}_0)$, it is the normalising constant created from $\int p(\mathbf{x}|\theta)p(\theta)d\theta = \int p(\mathbf{x}|\theta, \mathcal{M}_0)p(\theta|\mathcal{M}_0)d\theta$
- To calculate the posterior, we need a prior probability that the model is true: $p(\mathcal{M}_0)$
- We can compare two hypotheses (ie two models) via the ratio:

$$\frac{p(\mathcal{M}_1|\mathbf{x})}{p(\mathcal{M}_0|\mathbf{x})} = \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_0)} \times BF$$

  where *BF* is the *Bayes Factor*:

$$BF = \frac{p(\mathbf{x}|\mathcal{M}_1)}{p(\mathbf{x}|(\mathcal{M}_0)}$$

- If the ratio is large (ie much bigger than 1) we will favour $\mathcal{M}_1$, if it's small (ie closer to 0) we will favour $\mathcal{M}_0$. If it is $\simeq 1$ then there is very little to choose between the two hypotheses

Notes on Bayesian hypothesis testing

- If we have no preference for one model or another *a priori* then the Bayes factor will represent the posterior odds of $\mathcal{M}_1$ over $\mathcal{M}_0$
- In many simple cases, the Bayes factor will be a useful comparison between two models
- In more complicated cases (eg when we are using certain improper prior distributions), the Bayes factor will be difficult to calculate because the normalising constand $p(\mathbf{x}|\mathcal{M}_i)$ is hard to obtain
- Whilst the Bayes factor is a useful tool, it is not the only means to perform hypothesis testing. Sometimes simply computing the posterior distribution for a certain model will suffice
- Later in the course we will study some shortcuts to the Bayes factor that are simpler to calculate

A Bayesian approach: IQ example

Example

Define the likelihood as:

$$x_i \sim N(\lambda, \phi), i = 1, \ldots, n \text{ and } y_j \sim N(\mu, \phi), j = 1, \ldots, m$$

Using the reference prior $p(\lambda, \mu, \phi) \propto 1/\phi$, find the posterior distribution of $\delta = \lambda - \mu$, and hence $p(\delta > 0|\mathbf{x}, \mathbf{y})$.
Norw: This is not an example that involves competing models