Prior Exposure 1: Bayes for beginners: **Introduction and overview**

Thom Baguley and Mark Andrews
Nottingham Trent University

Summary

This workshop aims to be a general introduction to Bayesian data analysis and how it differs from the more familiar classical approaches to data analysis. We will start by providing a brief overview of the topic. The fundamental concepts of Bayesian statistical inference will follow, contrasted with frequentist methods of inference. To provide a bridge between Bayesian and classical methods, we will describe likelihood function approaches to inference and introduce both the likelihood principle and the law of the likelihood as the general precepts of likelihood based inference. During this workshop, there will also be practical exercises to illustrate the concepts and introduce participants to the practical application of Bayesian data analysis - with more to follow in workshop 2.

Psychology is embracing Bayesian data analysis (again)

On taking up the post of editor of the *Journal of Experimental Psychology: General* - arguably the most prestigious outlet for experimental work in psychology - the editor announced that the journal

> *will seek to attract work in specialized areas critical to the development of our science, for example [...] the adoption of Bayesian methods in data analysis.*

[Gauthier, 2012]

What is wrong with classical (frequentist) methods?

Leaving aside philosophical problems (e.g., that inference is based on unobserved data) here are a few of the things that have made me reconsider frequentist methods ...

- p values (and related quantities) are of limited informational value

What is wrong with classical (frequentist) methods?

Leaving aside philosophical problems (e.g., that inference is based on unobserved data) here are a few of the things that have made me reconsider frequentist methods ...

- p values (and related quantities) are of limited informational value
- null hypothesis tests are ill-suited to certain problems

What is wrong with classical (frequentist) methods?

Leaving aside philosophical problems (e.g., that inference is based on unobserved data) here are a few of the things that have made me reconsider frequentist methods ...

- *p* values (and related quantities) are of limited informational value
- null hypothesis tests are ill-suited to certain problems

  e.g., quantifying support *for* a null hypothesis

What is wrong with classical (frequentist) methods?

Leaving aside philosophical problems (e.g., that inference is based on unobserved data) here are a few of the things that have made me reconsider frequentist methods ...

- p values (and related quantities) are of limited informational value
- null hypothesis tests are ill-suited to certain problems

  e.g., quantifying support *for* a null hypothesis
- relevant external information enters a frequentist analysis in an *ad hoc* way

So ... how does Bayesian data analysis work?

A Bayesian analysis involves three things:

1 A probability model for the data (the likelihood)

So ... how does Bayesian data analysis work?

A Bayesian analysis involves three things:

1 A probability model for the data (the likelihood)

2 A probability model of relevant prior information (the prior)

So … how does Bayesian data analysis work?

A Bayesian analysis involves three things:

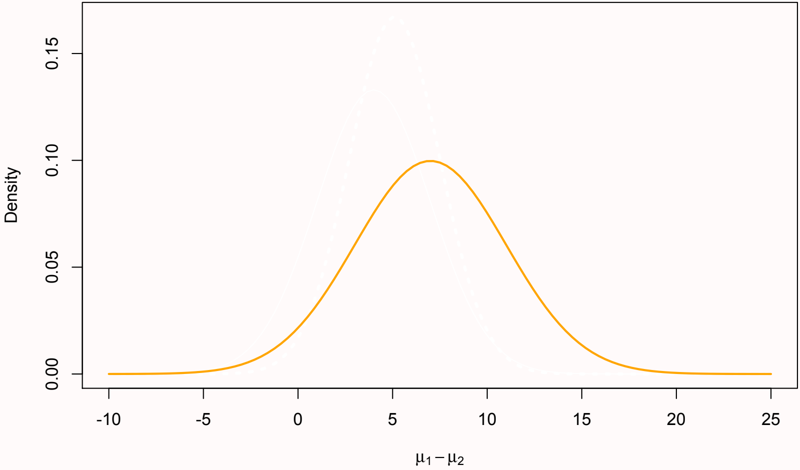
1. A probability model for the data (the likelihood)
2. A probability model of relevant prior information (the prior)
3. A probability model that *combines* the data and the prior information (the posterior)

The likelihood

The likelihood is a mathematical function that models the probability of the observed data as a function of a parameter (e.g., a population mean) or a set of parameters

e.g., imagine that a new teaching method is observed to increase reading age by 7 months with a standard error of 4 months

Assuming a normal distribution for the effect, the likelihood would look like this ...

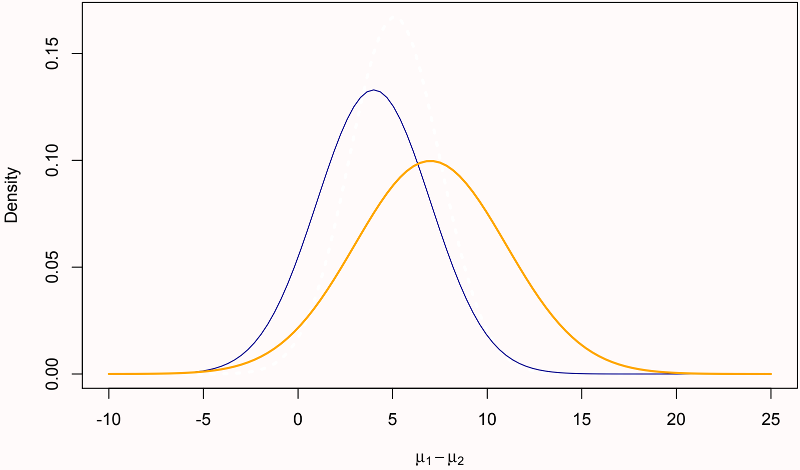

0.00
0.05
0.10
0.15
Density
-10
-5
0
5
10
15
20
25
μ1 − μ2

The prior

The prior is a probability distribution that reflects relevant information about a parameter (e.g., a population mean)

e.g., a previous meta-analysis found a mean reading age reduction of 4 months with a pooled standard error of 3

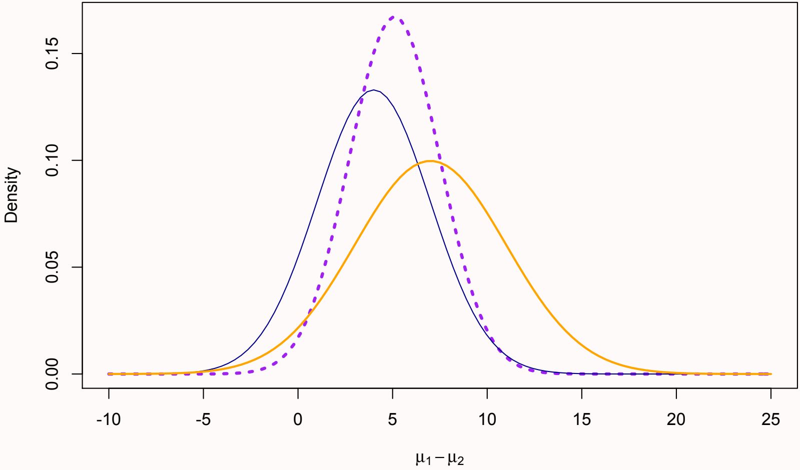
Again, assuming a normal distribution for the prior ...



Density
0.00
0.05
0.10
0.15
-10
-5
0
5
10
15
20
25
$\mu_1 - \mu_2$

Obtaining the posterior using Bayes' theorem ...

$$\overbrace{\mathrm{P}(\theta|\mathcal{D},\Omega)}^{\text{Posterior}} = \frac{\overbrace{\mathrm{P}(\mathcal{D}|\theta)}^{\text{Likelihood}}\ \overbrace{\mathrm{P}(\theta|\Omega)}^{\text{Prior}}}{\int \mathrm{P}(\mathcal{D}|\theta)\mathrm{P}(\theta|\Omega)\,d\theta}$$


Density
0.00
0.05
0.10
0.15
-10
-5
0
5
10
15
20
25
μ1 − μ2

Outcome of analysis influenced by the prior ...

Frequentist confidence interval

$\hat{\mu}_1 - \hat{\mu}_2 = 7$, 95% CI [-0.9, 14.9] N.B. $p > .05$

Outcome of analysis influenced by the prior ...

Frequentist confidence interval

$\hat{\mu}_1 - \hat{\mu}_2 = 7$, 95% CI [-0.9, 14.9] N.B. $p > .05$

Bayesian posterior probability interval and Bayes factor

$\hat{\mu}_1 - \hat{\mu}_2 = 5.1$, 95% probability interval [0.4, 9.8], BF $= 3.23$

Why Bayesian data analysis (BDA)?

- efficient and *coherent* use of available information

Why Bayesian data analysis (BDA)?

- efficient and *coherent* use of available information
- The right tool for certain problems

  (e.g., assessing evidence in favour of H_0)

- less reliance on *ad hoc* procedures

  (e.g., adjustment of α)

Why Bayesian data analysis (BDA)?

- efficient and *coherent* use of available information
- The right tool for certain problems

(e.g., assessing evidence in favour of H_0)

- less reliance on *ad hoc* procedures

(e.g., adjustment of α)

- greater emphasis on transparency in modelling

Why Bayesian data analysis (BDA)?

- efficient and *coherent* use of available information
- The right tool for certain problems

  (e.g., assessing evidence in favour of H_0)

- less reliance on *ad hoc* procedures

  (e.g., adjustment of α)

- greater emphasis on transparency in modelling
- more flexibility in modelling complex, messy (real world) data sets