Teaching Bayesian Data Analysis to Social Scientists

Mark Andrews

Psychology, Nottingham Trent University

mark.andrews@ntu.ac.uk

@xmjandrews, @seriousstats, @priorexposure

https://github.com/lawsofthought/bayeslund2017

Background

- The Advanced Training Initiative (ATI) by the Economic and Social Research Council (ESRC) provided grants to support training in advanced social science topics.
- Professor Thom Baguley (`@seriousstats`) and I were funded to provide a series of workshops on Bayesian data analysis each year for the years 2015, 2016, & 2017:

`http://www.priorexposure.org.uk`

- Each workshop was limited to around 25 attendees, and could be attended by any UK based social science researchers (post-graduate students and above).
- There was a minimal fee of £20 per workshop, with a £10 fee for students. There were bursaries to support travel and accommodation expenses by students.
- We taught 4 workshops in 2015, 6 (or 9) in 2016, and will teach 6 (or 9) in 2017[1].

[1] The reason for and details of the extra workshops in 2016 and 2017 will be explained.

Workshops: Overview

- Each workshop was planned to be a combination of lecture style teaching and practical exercises.
- All practical exercises were computer based and used R and Jags[2].
- Most lecture teaching involved R and Jags based demonstrations, which could be followed along step by step by attendees.
- Attendees were required to use their laptops, and details of how to install the required software were provided in advance.
- Source code and (most) other teaching materials are available at: `https://github.com/lawsofthought/priorexposure`.

[2] Why Jags and not its alternatives? See below for discussion.

Workshop 1: Bayes for beginners

- This workshop aimed to be a general introduction to Bayesian data analysis and how it differs from the more familiar classical approaches to data analysis.
- Here, we provided a gentle introduction to Bayesian methods. Topics included:
  - Examples of Bayesian inference and using prior information in simple statistics problems.
  - Understanding the likelihood function.
  - Hypothesis testing using Bayes factors.

Workshop 2: Doing Bayesian data analysis

- This workshop aimed to provide a solid theoretical and practical foundation for real-world Bayesian data analysis in psychology and social sciences.
- Topics included:
  - Some detailed examples of analytically tractable Bayesian inference (e.g. inference of Bernoulli random variables, inference of Poisson random variables, inference of means of univariate Normal models, etc.)
  - Introduction to probabilistic modelling with Jags.
  - Linear models with Jags.

Workshop 3: Introduction to advanced Bayesian data analysis and Bayesian multilevel modelling

- This workshop focused on advanced probabilistic modelling in Bayesian data analysis, and in particular, Bayesian data analysis using multilevel regression models.
- Topics included:
  - Multilevel linear models.
  - Multilevel generalized linear models, e.g. logistic regression, Poisson regression.
  - Examples included models with categorical predictors, interactions, random slope and random intercept models, crossed and nested structures.

Workshop 4: Nonlinear and latent variable models

- This final workshop focused on Bayesian latent variable modelling, particularly using mixture models, and nonlinear regression.
- Topics included:
  - Nonlinear regression modelling using radial basis functions.
  - Nonlinear regression modelling using Gaussian processes.
  - Finite mixture modelling.
  - Nonparametric mixture modelling using Dirichlet processes.

The extra workshops

- Workshops 1 & 2 proved very popular, and to meet demand, in 2016, we provided both workshops in April and again in June. We will repeat this this year.
- In addition, initially, we assumed a basic proficiency in R on the part of the workshop attendees. This was not generally true, and those who were less familiar with R struggled to keep up with exercises. As such, we put on one extra R workshop before each regular workshop pair.
- As such, we had 3×3 workshops in 2016 and will again this year.

Participants

- Attendees were students and researchers from psychology, sociology, criminology, geography, linguistics, neuroscience, economics, epidemiology, education, business studies, etc.
- A more detailed survey of attendees of this month's (April, 2017) workshops (workshops 1 & 2) showed:
  - About 50% of attendees are from psychology (usually experimental, cognitive).
  - About 50% are PhD students.
  - In terms of general statistical knowledge, attendees rate themselves as around $\frac{5.5}{10}$ on average.
  - In terms of statistical computing skill, they rate themselves as around $\frac{3.5}{10}$ on average.
  - In terms of knowledge of Bayesian methods, they rate themselves as around $\frac{2.2}{10}$ on average.
  - In terms of motivation, about $\frac{2}{3}$ said they were attending to learn more about hypothesis testing and Bayes factors.

Some lessons learned

- Delving into mathematical details, e.g. derivations of formulae for posterior distributions, did not prove to be very effective.
- Learning by building and running Jags models proved much more effective.
- Being comfortable with R is vital. Pre-workshop R bootcamps were popular and effective.
- Software installation problems can stymie progress.
- For many attendees, Bayesian data analysis means Bayesian hypothesis testing (with Bayes factors). While for us, Bayesian data analysis is more about flexible probabilistic modelling.
- The age of Bugs/Jags has (probably) passed, Stan is now the preferred choice as a probabilistic modelling language.

Recommendations for the future syllabus

- There seem to be three levels of Bayesian data analysis that people want to learn:
  - *Introductory*: The fundamentals of Bayesian data analysis and Bayesian hypothesis testing.
  - *Intermediate*: Regression modelling, i.e. linear, general linear, generalized linear, multilevel regression, including robust regression, model checking, model evaluation.
  - *Advanced*: A wide set of topics including: Nonlinear regression, latent variable modelling, mixture modelling, time series modelling, and possibly also causal modelling, structural equation modelling, Bayesian machine learning (Bayesian deep learning), etc.