# *Teaching Bayesian Data Analysis to Social Scientists*

Mark Andrews & Thom Baguley

Psychology, Nottingham Trent University

mark.andrews@ntu.ac.uk, @xmjandrews
thomas.baguley@ntu.ac.uk, @seriousstats

https://github.com/lawsofthought/priorexposure-esrc-conf

# *Background: Advanced Training Initiative*

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

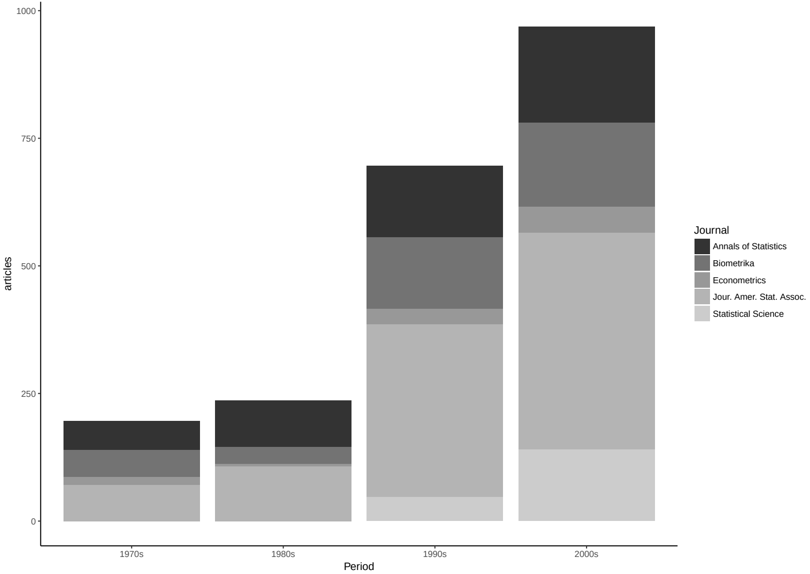
  `http://www.priorexposure.org.uk`

- 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 6 (or 9) in 2017[1].

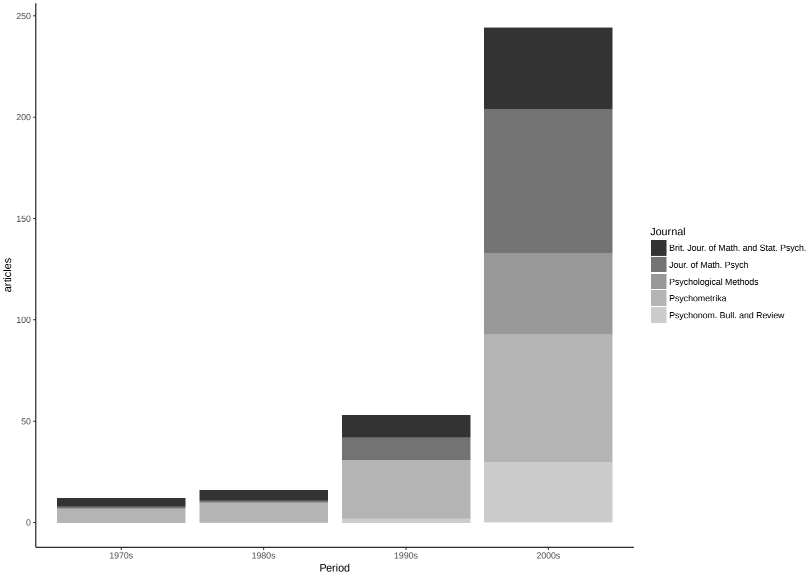
[1] The reason for and details of the extra workshops in 2016 and 2017 will be explained.

# Our case for support

Bayesian methods are growing in popularity, but are not yet part of the social science curriculum.



# *Our case for support*



# *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`.

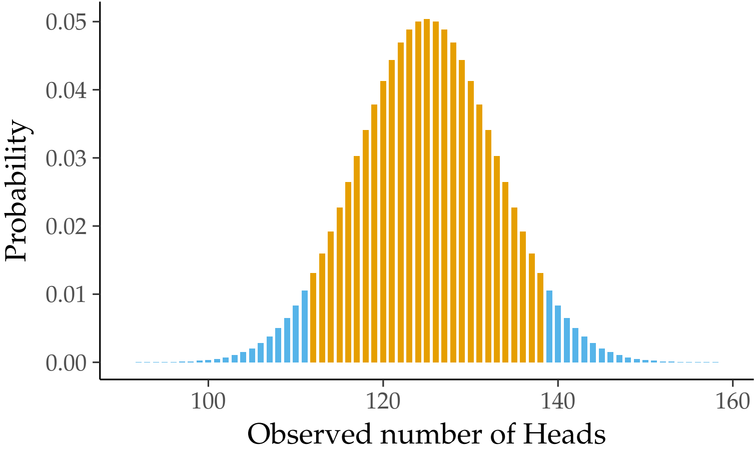
[2]Why Jags and not its alternatives? See below for discussion.

# *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.

# *Sampling theory based inference*

*A binomial distribution with* $n = 250$ *and* $\theta = 0.5$.




- Areas shaded in blue are values *as or more extreme* than $m = 139$.

# *The likelihood function versus sampling distribution*

- The sampling distribution is the binomial distribution:

$$P(m|\theta, n) = \binom{n}{m} \theta^m (1-\theta)^{n-m}.$$

  This is always a function of $m$, with $\theta$ and $n$ fixed.

- The corresponding likelihood function treats the same function

$$\binom{n}{m} \theta^m (1-\theta)^{n-m},$$

  as a function of $\theta$, but now with $m$ and $n$ fixed.

- As such, the binomial likelihood is

$$L(\theta|n, m) \propto \theta^m (1-\theta)^{n-m}.$$

# 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 $3 \times 3$ 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.

# *Conjugate prior for $\mu$ and $\sigma$*

- A common choice of conjugate prior is the normal/inverse-gamma distribution, also known as the *normal* $\times$ *scaled inverse-*$\chi^2$ distribution.
- Thus, our full model is

$$
\begin{aligned}
x_i &\sim N(\mu, \sigma^2), \quad \text{for } i \in 1, 2 \cdots n, \\
\mu &\sim N(\mu_0, \sigma^2/\kappa_0), \\
\sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2).
\end{aligned}
$$

- This corresponds to the joint prior density

$$
\begin{aligned}
P(\mu, \sigma^2) &= \text{N-Inv-}\chi^2(\mu, \sigma^2 | \mu_0, \kappa_0, \nu_0, \sigma_0^2), \\
&= N(\mu | \mu_0, \sigma^2/\kappa_0) \times \text{Inv-}\chi^2(\sigma^2 | \nu_0, \sigma_0^2), \\
&\propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} e^{\left(-\frac{1}{2\sigma^2}[\nu_0 \sigma_0^2 + \kappa_0 (\mu_0 - \mu)^2]\right)}
\end{aligned}
$$

# *Joint posterior on μ and σ*

- With the normal likelihood and normal/scaled inverse-$\chi^2$ distribution, the joint posterior is:

$$
\begin{aligned}
P(\mu, \sigma^2|D) &\propto P(D|\mu, \sigma^2)P(\mu, \sigma^2), \\
&\propto \sigma^{-n} e^{\left(-\frac{1}{2\sigma^2}[ns^2+n(\bar{x}-\mu)^2]\right)} \\
&\quad \times \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} e^{\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2]\right)} \\
&= \text{N-Inv-}\chi^2(\mu, \sigma^2|\mu_n, \kappa_n, \nu_n, \sigma_n^2)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_n &= \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_n}, \\
\kappa_n &= \kappa_0 + n, \\
\nu_n &= \nu_0 + n, \\
\sigma_n^2 &= \frac{1}{\nu_n}\left(\nu_0\sigma_0^2 + ns^2 + \frac{n\kappa_0}{\kappa_0+n}(\mu_0-\bar{x})^2\right).
\end{aligned}
$$

# *Recommendations for future workshops*

## *Syllabus*

- *Introductory*: The fundamentals of Bayesian data analysis. Bayesian inference and Bayesian model evaluation. Introduction to sampling.
- *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.

# *Recommendations for future workshops*

## *Format*

- Supplement workshops with video tutorial, and extensive notes (ideally a textbook).
- Especially for advanced topics, more workshops, emphasizing depth rather than breadth.
- Continue R workshops in parallel.
- Strive for completion elimination of software etc. installation issues.

# *Bayes in the core curriculum?*

- Bayesian methods must coexist with sampling theory approaches.
- Can (should) Bayesian methods be taught at introductory level?
- Multilevel models as the great Bayesian bait-and-switch.

# *Removing barriers to Bayes*

- Curriculum barriers, e.g. QAA benchmarks, BPS guidance.
- Course level barriers, e.g. module specifications, validation documents.
- Hidden barriers, e.g., terminology (*statistical tests*, *significance*, etc.), software (courses teaching SPSS).

# BPS curriculum 2004

## *curriculum defined by BPS qualifying exam*

- Descriptive and summary statistics . . .
- Probability theory . . .
- The normal distribution: z scores and areas under the curve; the sampling distribution of the sample mean.
- Statistical inference: significance testing (including the null and alternative hypothesis, type 1 and type 2 errors, significance level, power and sample size);
- Effect size and confidence intervals. z-tests and t-tests of means for single sample, independent samples and related samples designs.
- Confidence intervals: for the population mean; for the difference between two population means. Mean and error bar graphs. Non-parametric alternatives to t-tests: the sign test; Wilcoxon matched-pairs signed ranks test; Mann-Whitney test. Tests of proportions: chi-squared tests for goodness of fit and for contingency tables. Cramer's Phi as a measure of association in contingency tables. McNemar's test of change.
- Bivariate correlation and linear regression: scatterplots; Pearson's correlation coefficient; partial correlation; the significance of a correlation coefficient; the linear regression equation and its use in prediction; the accuracy of prediction; Spearman's and Kendall's rank order correlation coefficients.
- The analysis of variance: one factor independent and repeated measures designs; two factor independent, repeated measures and mixed designs; main effects and interaction effects (including graphical presentation); planned (including trend) comparisons; the Bonferroni correction; post hoc comparisons (including the choice between methods); the analysis of simple effects.
- Non-parametric alternatives to one factor analyses of variance: Kruskal-Wallis, Friedman and Cochran's Q tests. The choice of an appropriate statistical analysis: the issue of level of measurement (nominal, ordinal, interval and ratio scales); test assumptions (e.g. normality, homogeneity of variance, linearity); transformations of the dependent variable in an attempt to meet assumptions; robustness; power efficiency.

# *BPS curriculum 2017*

*curriculum defined by QAA benchmarks, BPS supplementary guidance.*

- Students will need to know how to conduct qualitative and quantitative research.
- This requires an awareness of different types of statistical inference such as hypothesis testing and interval estimation, whether through frequentist approaches (e.g., significance testing, confidence intervals) or recognised alternatives (e.g., Bayesian inference).
- Students will require knowledge of how to plan a study including how to select an appropriate sample and sample size (e.g., statistical power), to detect differences in sample means (e.g., t tests, ANOVA) and relationships between variables (e.g., Chi square, correlation, regression).
- This should include an appreciation of the philosophy and assumptions underpinning statistical procedures that they use and familiarity with robust alternatives when those assumptions are not met (e.g. non parametric alternatives).
- They will need to understand issues relating to scale construction (e.g. reliability, factor structure).