Why we don't teach, and why we should and could teach, Bayesian methods

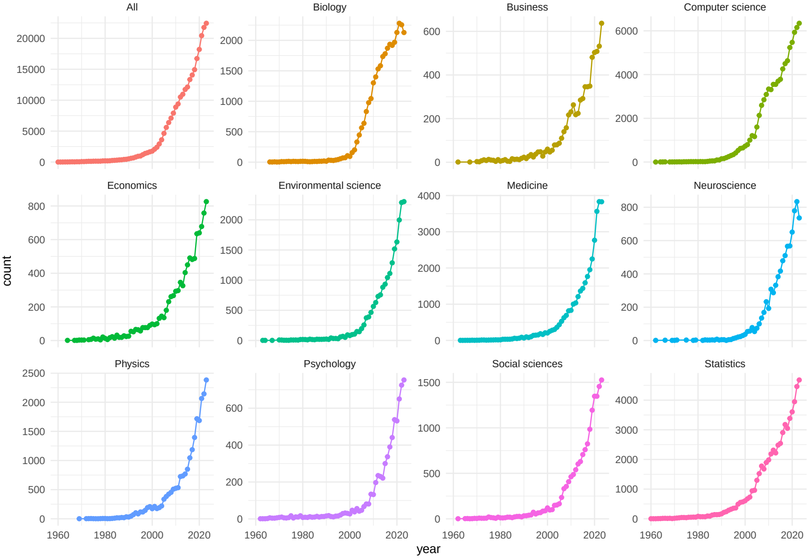
Mark Andrews

mark.andrews@ntu.ac.uk

Department of Psychology, Nottingham Trent University
@xmjandrews

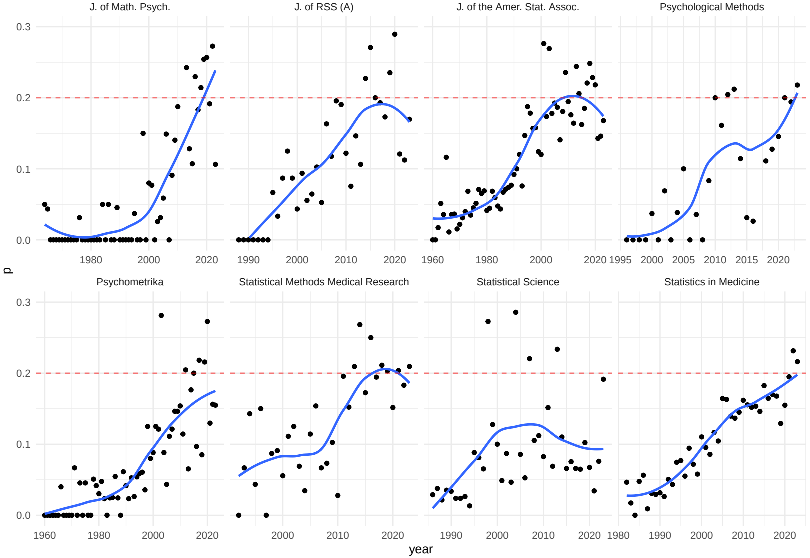
The Rise of Bayesian Data Analysis

Number of publications per year 1960-2023 with "Bayesian" in title, abstract, or keywords (Scopus)



The Rise of Bayesian Data Analysis

Proportion of publications in selected journals with "Bayesian" in title, abstract, or keywords (Scopus)



Why Bayes? Why now?

- Bayesian methods can be automatically applied to (almost) any statistical model.
- For any statistical model, if we can evaluate the function

$$f(\theta) \triangleq \underbrace{P(\text{data}|\theta)}_{\text{likelihood}} \underbrace{P(\theta)}_{\text{prior}},$$

  then we can use Markov Chain Monte Carlo (MCMC) to draw samples from the posterior distribution $P(\theta|\text{data})$.
- Exponentially increasing computational power have exponentially decreased the cost of using Bayesian methods.

Typical statistics teaching curriculum

Very rough approximation

- Across many fields, the core or foundational statistics topics are usually approximately:
  - Descriptive statistics, exploring data
  - Populations, samples, normal distributions
  - Hypothesis testing, p-values, significance, confidence intervals
  - Regression etc
  - Anova etc
- This is almost always exclusively based on frequentist inference.

Why not teach Bayes? Possible Reason 1

- Bayesian and frequentist approaches seem fundamentally incompatible:
  - *…(Bayesian inference) is founded upon an error, and must be wholly rejected* (Fisher, 1925, p. 10)
  - *…the only good statistics statistics is Bayesian statistics …* (Lindley, 1975, p. 106).
- Bayesian methods seem to require the complete rejection of p-values, significance, confidence intervals etc., and vice versa.
- *Possible rebuttal*: Frequentist and Bayesian inference are both reasonable methods of statistical inference.

Why not teach Bayes? Possible Reason 2

- Bayesian methods are traditionally seen as requiring a *subjectivist* interpretation of probability and statistics.
- Accordingly, probabilities represent degrees of belief and Bayes' theorem is used to update degrees of belief in light of new evidence.
- *Possible rebuttal*: Priors are just model assumptions. Both frequentist and Bayesian inference is based on deductions from assumptions and data.

Why not teach Bayes? Possible Reason 3

- Bayesian inference is too technical, e.g. Bayesian linear regression:

$$
\begin{aligned}
\rho(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X}) &\propto \rho(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2)\rho(\boldsymbol{\beta} \mid \sigma^2)\rho(\sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\mathsf{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \\
&\quad \times (\sigma^2)^{-k/2} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^\mathsf{T}\boldsymbol{\Lambda}_0(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right) \\
&\quad \times (\sigma^2)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right)
\end{aligned}
$$

- *Possible rebuttal*: Deriving formulas for even a t-test is just as technical.

Why not teach Bayes? Possible Reason 4

- Bayesian methods are still the minority approach.
- *Possible rebuttal*: See George Cobb's remark about the circularity of teaching and practice: *We teach it because it's what we do; we do it because it's what we teach.* (see Wasserstein & Lazar, 2016).

Why not teach Bayes? Possible Reason 5

- Bayesian and frequentist methods lead to same(ish) results:

```
M <- lm(dist ~ speed, data = cars)
Mb <- brm(dist ~ speed, data = cars)
```

```
tidy(M, conf.int = T)
#> # A tibble: 2 x 7
#>   term        estimate std.error statistic  p.value conf.low conf.high
#>   <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
#> 1 (Intercept)   -17.6      6.76      -2.60 1.23e- 2   -31.2      -3.99
#> 2 speed           3.93     0.416      9.46 1.49e-12     3.10      4.77
fixef(Mb) %>% round(2)
#>           Estimate Est.Error   Q2.5 Q97.5
#> Intercept   -17.64      7.01 -31.62 -3.86
#> speed         3.94      0.44   3.09  4.83
```

- *Possible rebuttal*: Bayesian methods can be used where there are no frequentist options.

Should we teach Bayesian methods?

If so, how?

- It depends on course topic, learning outcomes, available time, etc.
  - *Personal example 1*: The core statistics modules for BSc Psychology degree ($\approx$ 60 hrs). Here, covering Bayesian methods exclusively, or covering both approaches, would be impractical.
  - *Personal example 2*: An optional advanced statistical module in a BSc degree (40 hrs). Here, we begin with Bayesian and frequentist statistical inference, and then cover general, generalized and mixed effects models using both approaches.
  - *Personal example 3*: A foundational statistics module in a data science MSc degree (40 hrs). Same approach as example 2.

References

Fisher, R. A. (1925). *Statistical Methods For Research Workers*. Oliver; Boyd.

Lindley, D. V. (1975). The future of statistics: A Bayesian 21st century. *Advances in Applied Probability*, *7*, 106–115. http://www.jstor.org/stable/1426315

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. In *The American Statistician* (2; Vol. 70, pp. 129–133). Taylor & Francis.