

Projects

Projects are due Friday March 18 by 12 noon. There are no strict requirements, but it seems hard that you would be able to document an adequate amount of work in less than 10 pages, and going over 15 pages will make your paper harder to read. As you will see the prompts below are open ended: we hope you will enjoy your explorations!

Take your own Bayesian peek at Feller vol. 1

Find a copy of the article *A Bayesian peek at Feller volume 1*, by Diaconis and Holmes. Take one of the examples and understand it in your own language: when the authors say “standard computations show” or “it is easy to see...” put in some detail. *Then* find a copy of W. Feller, *An introduction to probability and its applications vol. 1 (3rd ed.)*. Take any other topic (eg. a homework problem or a section) and make a Bayesian version.

BDA3

Take one of the latest chapters in Gelman, Carlin, Stern Dunson, Vehtari and Rubin (2013) *Bayesian Data Analysis*, Chapman Hall, 3rd edition. We suggest one of section 19-22. Take one example in the section (preferably with data—all the datasets are on the web-site for the book). Explain clearly what the example is and detail the procedures used. *Then* “do something”: change the assumptions; vary the prior; where the authors say “by the usual arguments” put in the details; try to prove something; find something wrong with the analysis; or go get some more data... (You do not have to do all of this, of course: these are just suggestions.

Disputed Authorship

For this project, you will recreate a classic Bayesian analysis of disputed authorship. Read the paper “Inference in an Authorship Problem” by Mosteller & Wallace (*Journal of the American Statistical Association*, 1963). Choose two contemporary authors who write articles/papers/blog posts for roughly similar audiences (e.g., Larry Summers and Paul Krugman). Their writing styles should not be so obviously different that distinguishing between them is trivial. Collect 50 works from each author and sample 5 from each to be “disputed papers.” Finally, apply the Bayesian methodology of Mosteller & Wallace to infer authorship of the disputed papers.

Suggestions:

- As in the original analysis, choose “non-contextual” words on the basis of intuition and/or features of the data.

- The original paper was limited by the computational tools and resources of the time. For example, the authors choose to ignore uncertainty in the estimated rates μ_M and μ_H . Instead, you might want to simulate from the posterior to evaluate the posterior probability of authorship for the disputed papers. You may also find that selecting a very small set of words, as the authors did, is not necessary.
- Feel free to modify parts the original analysis, e.g., if you think a different prior distribution would be appropriate.
- If you use R, the package `tm` might be useful for parsing and analyzing text files.

Applied Bayesian Analysis

The papers below present examples of Bayesian statistics in action. Choose one and study it carefully. Provide a clear summary of one of the models they consider; discuss the assumptions; how does the Bayesian approach translates in conclusions that might be different from the ones obtained within a classical framework? what do you think are the limitations of the analysis? Extend the work presented in one direction, either using new data, new model or priors, new computational methods, or studying some properties of the estimates that is not analyzed in the original paper.

- Efron and Thisted (1976) “Estimating the Number of Unseen Species: How Many Words Did Shakespeare Know?” *Biometrika* **63**:435–47.
- Geman and Geman (1984) “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.
- Gelman and King (1990) “Estimating the Electoral Consequences of Legislative Redistricting,” *JASA* **85**: 274–282.
- James Albert and Siddhartha Chib (1993) “Bayesian Analysis of Binary and Polychotomous Response Data,” *Journal of the American Statistical Association* **88**: 669–679.