Introduction to Applied Bayesian Modeling

Ryan Bakker
Department of Political Science
University of Georgia

May 12, 2015

Office: TBD.
Office hours: TBD. Feel free to stop by the office any time and come in if our door is open. We're also happy to schedule meetings at most other times during the day.

Course description

This course introduces the basic theoretical and applied principles of Bayesian statistical analysis in a manner geared toward students and researchers in the social sciences. The Bayesian paradigm is particularly useful for the type of data that social scientists encounter given its recognition of the mobility of population parameters, its ability to incorporate information from prior research, and its ability to update estimates as new data are observed. The course begins with a discussion of the strengths of the Bayesian approach for social science data and the philosophical differences between Bayesian and frequentist analyses. Next, the course covers the theoretical underpinnings of Bayesian modeling and provides a brief introduction to the primary estimation algorithms. The bulk of the course focuses on estimating and interpreting Bayesian models from an applied perspective. Participants are introduced to the Bayesian forms of the standard statistical models taught in regression and MLE courses (i.e., normal, logit/probit, Poisson, etc.) as well as a variety of measurement and multilevel models. This course assumes a solid understanding of the linear model and matrix algebra and some exposure to models with limited dependent variables. The course relies mostly on R and WinBUGS/JAGS for estimation. Prior experience with R is preferred but not assumed; we offer lab sessions to familiarize participants with WinBUGS and JAGS (no prior experience necessary).

Goals. Upon conclusion of this course, we aim for participants to be able to:

- appreciate the fundamental differences and similarities between frequentist and Bayesian approaches to inference
- apply Bayes' rule to the regression context
- formulate linear and generalized linear models in the Bayesian framework
- estimate linear and generalized linear models in the Bayesian framework using flexible code
- exploit the advantages of Bayesian estimation with regard to
  - incorporating prior information
  - incorporating uncertainty in parameter estimates
  - dealing with missing data
  - measuring latent concepts
  - incorporating variance at multiple levels of observation
- present and communicate results from Bayesian (and frequentist!) estimation in an efficient manner
- have fun learning new methods!

Level of difficulty. Although this course will cover some of the basics of MCMC and the Gibbs Sampler (among other sampling algorithms), application and interpretation will be the primary focus. For this reason, students already familiar with the basics of Bayesian modeling using WinBUGS, MCMCpack, JAGS or some other software for Bayesian estimation may find the course in *Advanced Bayesian Models for the Social Sciences* offered in the second session more appropriate.

A note on computing. This course uses WinBUGS and JAGS as the preferred software options to fit Bayesian models. Some lectures may rely on WinBUGS for demonstration purposes, but the languages used by WinBUGS and JAGS for model specification are nearly identical. WinBUGS and its sibling OpenBUGS run on Macs only with the appropriate "make your Mac run Windows" software, but can be a bit buggy. JAGS runs on all platforms, including Macs. We offer special Mac-friendly lab sessions and support both JAGS and WinBUGS. JAGS code for all models encountered in this course and other JAGS-specific code and examples are provided.

Course resources

Z-Drive. All slides, code used in course sessions, and problem sets will be posted on the Z-Drive (`Z:/bakker/applied.bayes.2014`). Participants can access the Z-Drive from any computer in the three computer labs in the Helen Newberry building.

Course website with additional materials: Additional code, a JAGS tutorial, and other materials for weeks 3–4 are posted on Johannes' website: `http://www.jkarreth.net`.

Reading materials

Books

The main texts used in this course are:

- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, New York, NY.
- Gill, J. (2008). *Bayesian Methods: A Social And Behavioral Sciences Approach, Second Edition.* Chapman and Hall/CRC, Boca Raton, FL.

You may also find the following titles useful for many of the topics discussed in this course. They are available in the ICPSR Summer Program Library for borrowing:

- Congdon, P. D. (2003). *Applied Bayesian Modelling.* Wiley, Chichester.
- Congdon, P. D. (2010). *Applied Bayesian Hierarchical Methods.* Chapman and Hall/CRC, Boca Raton, FL.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition.* Chapman and Hall/CRC, Boca Raton, FL.
- Jackman, S. (2009). *Bayesian Analysis for the Social Sciences.* Wiley, Chichester.
- Kruschke, J. (2011). *Doing Bayesian Data Analysis: A Tutorial Introduction with R.* Academic Press / Elsevier, Oxford.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS.* Wiley, Hoboken, NJ.

As a general primer for R, we recommend:

- Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression, Second Edition.* Sage, Thousand Oaks.

Articles

All articles listed in the syllabus are available on the Z-Drive (`Z:/bakker/applied.bayes.2014`).

Software

This course relies mostly on R and WinBUGS/JAGS, but may also discuss Stata as an alternative for some applications. We provide assistance installing R and WinBUGS/JAGS on your computers in the first week of the course. The labs at the Helen Newberry building have all necessary software as well. R, WinBUGS and JAGS are available at no cost from:

- `http://www.cran.r-project.org`

- `http://www.mrc-bsu.ca.ac.uk/bugs`
- `http://mcmc-jags.sourceforge.net`

Each website links to relevant documentation and user manuals. There is a learning curve for these programs, but you need not have any computer programming background to learn them rather easily—just patience and desire. Our goal is to make you as comfortable as possible with these programs by the end of this course so that you will be able to use them with ease at your home institutions and in your own work.

Mac and JAGS users: See Johannes' website for more information on JAGS.

Homework assignments

Homework exercises are assigned in class. We will have a 'no child left behind' policy, that is: our goal is to make sure participants receive sufficient feedback to complete all assignments successfully. There will be between 2 and 4 assignments per week. They will be mostly computer-based with the exception of the first assignment. Please email your assignments to both TAs (`chare@uga.edu`, `racooper@uga.edu`) as PDF files and include [Bayes2014] in the subject line. Also always include all code you used to complete your assignments. They will aim to return graded assignments to you within 2 days with comments via email. TAs and instructors are happy to provide help with assignments during office hours: don't be afraid to come by and ask.

Labs

We offer several labs with guided hands–on exercises. The lab sessions will be held in the computer labs at the Helen Newberry building (exact times and locations TBC). Topics (see the schedule for dates):

1 Installing and using R
2a Installing and accessing JAGS from R
2b Accessing WinBUGS from R
3 Obtaining convergence diagnostics using R
4 Model presentation
5 Advanced measurement models

Course content and schedule

The following dates and topics may be modified as the course proceeds. The most current version of the syllabus will always be at `www.jkarreth.net/files/bayes2014.pdf`.

Monday, July 23: No course meeting

Recommended: Dave Armstrong's Introduction to the LaTeX Text Processing System, 5:30pm–7:30pm.

Day 1: Introduction: Background and Basics of Bayesian Inference

Please read:

- Gill: Chapter 1.
- Siegfried, T. (2010). Odds are, it's wrong: Science Fails to Face the Shortcomings of Statistics. *Science News*, 177(7):26–29.
- Senn, S. (2003). Bayesian, Likelihood, and Frequentist Approaches to Statistics. *Applied Clinical Trials*, 12(8):35–38.

Day 2: Review of Generalized Linear Models

Refresher:

- Gill: Section 2.2.
- Gelman & Hill: Chapter 6.

Day 3: Probability and Bayes' Rule

Please read:

- Gill: Chapter 2.
- Western, B. and Jackman, S. (1994). Bayesian Inference for Comparative Research. *American Political Science Review*, 88(2):412–423.

Day 4: Priors

Please read:

- Gill: Chapter 5.
- Gill, J. and Walker, L. D. (2005). Elicited Priors for Bayesian Model Specifications in Political Science Research. *Journal of Politics*, 67(3):841–872.

HW 1 assigned: Prior and posterior distributions.
Lab 1: Installing and using R.

Day 5: Sampling Methods and Introduction to the BUGS/JAGS Language

Please read:

- Gill: Chapters 8 & 9.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2003). WinBUGS Version 1.4 User Manual.
- Plummer, M. (2013). JAGS Version 3.4.0 User Manual.

Lab 2a: Installing and accessing JAGS from R.
Lab 2b: Accessing WinBUGS from R.

Day 6: Convergence Diagnostics

Please read either one of:

- Smith, B. J. (2007). boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, 21(11):1–37.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

HW 2 assigned: Becoming familiar with WinBUGS/JAGS.
Lab 3: Obtaining convergence diagnostics using R.

Day 7: The Normal Distribution; Priors (ctd.)

Please read:

- Gill: Chapter 3
- Kerman, J. (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics*, 5:1450–1470 (if you want to know more about noninformative priors).

Day 8: The Bayesian Linear Model

Please read:

- Gill: Chapter 4.
- Efron, B. (1986). Why Isn't Everyone a Bayesian? *American Statistician*, 40(1):1–5.

HW 3 assigned: Linear model.

Day 9: Missing Data in Bayesian Models

Please read:

- Jackman, S. (2000). Estimation and Inference Are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation. *Political Analysis*, 8(4):307–332.

HW 4 assigned: Debugging BUGS/JAGS code.
Lab 4: Model presentation.

Day 10 : Binary Outcomes

If you'd like a refresher, please read:

- Gelman & Hill, Chapter 5.

HW 5 assigned: Logistic regression model.

Day 11: Measurement and IRT Models

Please read one of:

- Fox, J.-P. and Glas, C. (2001). Bayesian Estimation of a Multilevel IRT Model Using Gibbs Sampling. *Psychometrika*, 66(2):271–288.
- Treier, S. and Jackman, S. (2008). Democracy as a Latent Variable. *American Journal of Political Science*, 52(1):201–217.
- Gray, J. and Slapin, J. B. (2012). How Effective are Preferential Trade Agreements? Ask the Experts. *Review of International Organizations*, 7(3):309–333.

HW 6 assigned: Factor or IRT model.

Day 12: Measurement Models and Identification

Please read one of:

- Bakker, R. (2009). Re-measuring Left–Right: A Comparison of SEM and Bayesian Measurement Models for Extracting Left–Right Party Placements. *Electoral Studies*, 28(3):413–421.
- Bakker, R. and Poole, K. T. (2013). Bayesian Metric Multidimensional Scaling. *Political Analysis*, 21(1):125–140.

- Fariss, C. J. (2014). Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability . *American Political Science Review*, 108(2):297–318.
- Linzer, D. A. and Staton, J. K. (ND). A Measurement Model for Synthesizing Multiple Comparative Indicators: The Case of Judicial Independence. *Working paper*.

Lab 5: Advanced measurement models.

Day 13: Ordered and Categorical Outcomes

Please read one of:

- Duch, R. M., May, J., and Armstrong, D. A. (2010). Coalition-directed Voting in Multiparty Democracies. *American Political Science Review*, 104(4):698–719.
- Stegmueller, D. (2013b). Modeling Dynamic Preferences: A Bayesian Robust Dynamic Latent Ordered Probit Model. *Political Analysis*, 21(3):314–333.
- Stegmueller, D., Scheepers, P., Roßteutscher, S., and de Jong, E. (2012). Support for Redistribution in Western Europe: Assessing the Role of Religion. *European Sociological Review*, 28(4):482–497.

HW 7 assigned: Ordered or multinomial logit model.

Day 14: Model Checking and Presentation

Please read:

- Gill: Chapters 6 & 7.
- Gelman, A., Goegebeur, Y., Tuerlinckx, F., and Mechelen, I. V. (2000). Diagnostic Checks for Discrete Data Regression Models Using Posterior Predictive Simulations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 49(2):247–268.
- Quinn, K. M., Martin, A. D., and Whitford, A. B. (1999). Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models. *American Journal of Political Science*, 43(4):1231–1247 (if you are interested in model comparison).

Day 15: Multilevel Models (Intro)

Please read:

- Montgomery, J. M. and Nyhan, B. (2010). Bayesian Model Averaging: Theoretical Developments and Practical Applications. *Political Analysis*, 18(2):245–270.
- Warren, T. C. (2014). Not by the Sword Alone: Soft Power, Mass Media, and the Production of State Sovereignty. *International Organization*, forthcoming (skim as an example of an application of BMA).
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25:111–163 (Background on BMA, read if you're interested)
- Gelman, A. and Rubin, D. B. (1995). Avoiding Model Selection in Bayesian Social Research. *Sociological Methodology*, 25:165–173 (Background on BMA, read if you're interested)
- Bartels, L. M. (1997). Specification Uncertainty and Model Averaging. *American Journal of Political Science*, 41(2):641–674 (Background on BMA, read if you're interested)
- Gelman & Hill: Chapter 16 or/and Gill: Chapter 10
- Gelman & Hill: Chapter 11 (for a refresher on multilevel models).

Day 16: Multilevel Models (Continuous Outcomes)

Please continue to read:

- Gelman & Hill: Chapter 16.

HW 8 assigned: Multilevel model.

Day 17: Multilevel Models (Other Outcomes)

Please read:

- Gelman and Hill: Chapters 15 and 17 (see Chapter 15 for a refresher).

as well as any of these empirical articles that is in your area of interest:

- Shor, B., Bafumi, J., Keele, L., and Park, D. (2007). A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data. *Political Analysis*, 15(2):165–181.
- Ward, M. D., Siverson, R. M., and Cao, X. (2007). Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace. *American Journal of Political Science*, 51(3):583–601.

· Blaydes, L. and Linzer, D. A. (2012). Elite Competition, Religiosity and Anti-Americanism in the Islamic World. *American Political Science Review*, 106(2):225–243.

· Lock, K. and Gelman, A. (2010). Bayesian Combination of State Polls and Election Forecasts. *Political Analysis*, 18(3):337–348.

· Stegmueller, D. (2013a). How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science*, 57(3):748–761.

· Chaudoin, S., Milner, H. V., and Pang, X. (2014). International Systems and Domestic Politics: Linking Complex Theories with Empirical Models in International Relations. *International Organization*, forthcoming.

· Bakker, R., Hill, D. W., and Moore, W. H. (2014). Modeling Terror Attacks: A Cross-National, Out-of-Sample Study. In Caruso, R. and Locatelli, A., editors, *Understanding Terrorism: A Socio-economic Perspective (Contributions to Conflict Management, Peace Economics and Development, Volume 22)*, pages 51–68. Emerald Group.
