Bayesian Data Analysis

Course details

Richard D. Morey — richarddmorey@gmail.com (mailto:richarddmorey@gmail.com)

Course description

In recent decades, there has been an explosion of interest in Bayesian methodologies in the sciences. There are several reasons: first, Bayesian methods allow for flexible and complex modeling where classical methods are unwieldy. Second, Bayesian methods often yield easier-to-interpret answers to statistical questions than classical methods. In this course, you will learn the basics of practical Bayesian data analysis.

The first week of the course introduces students to the philosophy behind Bayesian statistics with applications to common models in the social sciences, like t tests, ANOVA, and regression. The foundational ideas in the first week are built upon in the second week, where students learn about more complicated (multilevel) Bayesian models. Special attention will be paid to Markov Chain Monte Carlo (MCMC) methods, which give Bayesian methods their flexibility and power. Using free software (e.g. JAGS, stan) with R, MCMC methods are practical for researchers who are not specialists in Bayesian methods.

Learning objectives

After taking this course, students should be able to:

- Describe the difference between Bayesian and frequentist statistics
- Describe the roles of the prior, likelihood, and posterior in a Bayesian data analysis
- Build a model and obtain posterior inferences for relevant parameters
- Interpret the output of a MCMC sampler
- Diagnose problems with MCMC samplers

Software

- R (latest version, available on CRAN (https://cran.r-project.org))
- Rstudio (latest version, available from Rstudio (https://www.rstudio.com/products/rstudio/download2/))
- JAGS (latest version, available on sourceforge (https://sourceforge.net/projects/mcmc-jags/))
- stan (latest version, available on mc-stan.org (https://mc-stan.org/users/interfaces/rstan))
- R packages to be installed as needed

Prerequisites

In order to perform all but the basic analyses we will discuss in this course, some working knowledge of R will be needed. Those without basic fluency in R but with knowledge of another similar language will probably be able to get by asking questions and using online resources. If you do not have any experience with programming, the course will be a struggle because you will be learning several new concepts — Bayesian statistics and programming — at once. You may, however, still benefit from the course through the readings and learning from the solutions to the lab exercises.

The course also assumes a basic knowledge of standard statistical techniques, such as t tests, linear models, generalized linear models, and others that one would typically encounter in undergraduate

statistics courses.

Text and readings

Each day except the first there will be two sets of readings assigned: one covering the theory of Bayesian inference, and one reading covering the practice of Bayesian inference. Most of the theoretical readings are taken from the text assigned for the course.

The text for the course is McElreath (2015), "Statistical Rethinking" (ISBN: 9781482253443).

Schedule

Day	Hour 1 Topic	Hour 2 Topic	Lab Topic	Theory reading	Practical reading	Notes
1	Course overview and introduction	Why Bayes?	A first Bayesian analysis	—	—	—
2	Modeling philosophy	Bayesian statistics	Summarizing posteriors	McElreath, Ch. 1-3	Gelman (2012): Ethics and the statistical use of prior information (http://www.stat.columbia.edu/~gelman/research/published/ChanceEthics5.pdf)	—
3	Linear models	Multivariable linear models	Fitting a simple linear model	McElreath, Ch. 4-5	—	—
4	Parameters and data	Overparametrization and regularization	Priors and overparametrization	McElreath, Ch. 6	Gelman & Shalizi, 2013, "Philosophy and the Practice of Statistics" ((http://www.stat.columbia.edu/~gelman/research/published/philosophy.pdf))	—
5	Interactions I	Interactions II	Building models with interactions	McElreath, Ch. 7	—	—
6	Markov Chain Monte Carlo	Entropy and information	Working with MCMC chains	McElreath, Ch. 8-9	Hamra et al. (2013): Markov Chain Monte Carlo: an introduction for epidemiologists (http://ije.oxfordjournals.org/content/42/2/627.full.pdf+html)	—
7	Categorical outcomes	Count outcomes	MCMC chain diagnostics	McElreath, Ch. 10	Hartig (2011): MCMC chain analysis and convergence diagnostics with coda in R (https://theoreticalecology.wordpress.com/2011/12/09/mcmc-chain-analysis-and-convergence-diagnostics-with-coda-in-r/)	Install the shinystan (https://mc-stan.org/users/interfaces/shinystan) package.
8	Models of orderings	Zero-inflation and over-dispersion	Working with stan	McElreath, Ch. 11	—	—
9	Multilevel models I	Multilevel models II	Multilevel models in stan	McElreath, Ch. 12-13	Gelman et al (2008): Rich State, Poor State, Red State, Blue State (http://www.stat.columbia.edu/~gelman/research/published/rb_qjps.pdf)	—
10	Measurement error	Missing data	Missing data in stan	McElreath, Ch. 14	—	—

Draft compiled Tue Aug 06 13:16:32 2019 (Europe/London)