BST 228: Applied Bayesian Analysis

Harvard T. H. Chan School of Public Health
Fall 2017

Instructor: Corwin Zigler

TAs: Georgia Papadogeorgou
Will Townes

Lectures: Tu, Th 9:45am - 11:15am, Kresge 200

Lab: TBD

Contact information:

Office: HSPH Building 2, Room 415
Phone: 617.432.5014

E-mail: czigler@hsph.harvard.edu
gpapadogeorgou@fas.harvard.edu
ftownes@g.harvard.edu

Office hours: Cory's: Thursdays 11:15am-12:15pm, or by appointment.
Note: **Please** let me know in advance if you are intending to attend my office hours

Will's: Tuesdays 11:30-12:30, Building 2, Room 434
Georgia's: By appointment

Webpage: `https://canvas.harvard.edu/courses/32119`

COURSE DESCRIPTION

BST 228 is a practical introduction to the Bayesian analysis of biomedical and public-health data. It is an intermediate Master's level course in the philosophy, analytic strategies, implementation, and interpretation of Bayesian data analysis. Topics that will be covered include:

- The Bayesian paradigm
- Bayesian analysis of basic models
- Bayesian computing: Markov Chain Monte Carlo
- R and STAN software for Bayesian data analysis
- Model checking and model uncertainty
- Hierarchical Models
- Meta-analysis
- Linear regression

- Generalized linear models
- Hierarchical regression models
- Models for missing data

In addition to instruction in the above (and other) topics, the course will also present several case studies of Bayesian analysis in various areas of biomedical and public health research. Computer labs will contain programming instruction to provide hands-on training in the concepts covered in lectures and case studies.

PREREQUISITES

The course is intended for masters and doctoral students in all departments from the School. BST 210 and BST 222 are core requirements. You must be very comfortable working with probability distributions, evaluating common integrals, and reasoning around basic statistical models (e.g., different types of regression models).

COURSE OBJECTIVES

The overall objective of the course is to provide students with the tools to:

- Articulate the differences between the core tenets and philosophies of Bayesian and classical frequentist inference
- Deploy the basic mechanics of Bayesian analysis in simple probability models such as specifying prior distributions, deriving posterior distributions, and calculating posterior summary quantities of unknown parameters
- Evaluate whether a given Bayesian model is well-formulated and be able to assess model fit when applied to a given dataset
- Use Stan statistical software to estimate parameters of complex statistical models
- Assess the convergence of simulation algorithms to sample from posterior distributions
- Navigate statistical literature on methods for Bayesian analysis
- Recognize research scenarios for which Bayesian methods offer advantages over alternatives
- Know where to start when deploying Bayesian methods in complex problems that arise in their own work and/or research
- Know how to interpret the results of their Bayesian statistical analysis and present such analyses in a comprehensive way

OUTCOME MEASURES

Evaluation and the course grade will be based on three components:

Homework	25%
Exams (midterm, final)	65%
Participation	10%

There will be approximately 4 homework assignments throughout the semester. All exams will be take-home exams.

Academic Integrity for Problem Sets: Much of statistical work is collaborative, therefore you may discuss your solutions with other students; in fact it is encouraged to work together in groups (2-4 students) for problem sets. While sharing ideas on how to demonstrate something analytically or with computer syntax is allowed, you should make a full effort on your own initially, before discussing matters on which you are stuck or comparing methods. After discussions with peers, make sure that you can work through the problem yourself and ensure that any answers you submit for evaluation are the result of your own efforts. **Your final writeup and any requested computer code MUST be your own, and no material may be copied from another student.** This means that for applied problems, each of you must do your own analyses, programming, and writeup. Similar guidelines apply to more theoretical problems. If you work with other students on the homework problem sets, **you must turn in your own completed assignment** and please be sure to include with your assignment the names of your collaborators.

Academic Integrity for Exams: Exams should be completed on your own, without direct collaboration with other students. While you are free to informally discuss exam questions with fellow students, **actually conducting work in collaboration with another student is not allowed**. Students will be encouraged to consult the TA and/or instructor for assistance with exam problems, rather than seek detailed help from other students. The standard for demonstrating your individual effort will be higher for exams than for problem sets.

COURSE MATERIALS

Electronic copies of course handouts, slides and notes, and homework assignments, as well as datasets, will be posted on the course website. The primary text for the class will be

- Bayesian Data Analysis (3rd Edition) by Gelman, Carlin, Stern, Dunson Vehtari, and Rubin. ISBN:9781439898208

Other excellent introductions to Bayesian methods that will be useful throughout the semester are:

- Andrew Gelman and Jennifer Hill Data Analysis using Regression and Multilevel/Hierarchical Models, Cambridge University Press
- Emmanuel Lesaffre and Andrew B. Lawson Bayesian Biostatistics, Wiley
- Peter D. Hoff A First Course in Bayesian Statistical Method, Springer.
- Peter M. Lee Bayesian Statistics: An Introduction, Wiley.
- Jean-Michel Marin and Christian Robert Bayesian Core: A Practical Approach To Computational Bayesian Statistics, Springer.
- Bradley P. Carlin and Thomas A. Louis Bayesian Methods for Data Analysis, Chapman & Hall/CRC.

Other useful references (in no particular order, some of which will be required reading and discussed during the course):

- Dumouchel, W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. The American Statistician. 1999
- Gelman, A. P values and statistical practice. Epidemiology. 2013.
- Dominici, F. Combining evidence on air pollution and and daily mortality from the 20 largest US cities: a hierarchical modeling strategy. Journal of the Royal Statistical Society, Series A. 2000.
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian Model Averaging: A Tutorial. Statistical Science 1999. 14(4):382-417.
- Greenland S. Prior data for non-normal priors. Statistics in Medicine 2007: 26:3578-3590.
- Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. International Journal of Epidemiology 2006. 35:765-775.
- McCandless L. Bayesian sensitivity analysis for unmeasured confounding in observational studies. Statistics in Medicine 2007. 26:2331-2347.
- Stingo F et al. Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. Annals of Applied Statistics 2011. 5(3): 1978-2002.
- Best N and Hansell AL. Geographic Variations in Risk: Adjusting for unmeasured confounders through joint modeling of multiple diseases. Epidemiology 2009. 20(3): 400-410.

OUTLINE OF CLASS SCHEDULE

Part 1. The Bayesian Paradigm and Basic Models ($\sim$ August 29 – October 5)

- Introduction to Bayesian Analysis and Motivating Examples
- The Bayesian Inferential Paradigm
- Prior Distributions, Posterior Inference, and Prediction
- Simple Bayesian Models
- Markov Chain Monte Carlo
- Regression Models
- Model Checking and Evaluation

Part 2. Intermediate Bayesian Models ($\sim$ October 10 – November 7)

- Basic Hierarchical Models
- Hierarchical Linear Models
- Multilevel Models
- Spatial Hierarchical Models
- Bayesian Regularization and Variable Selection
- Bayesian Model Averaging

Part 3. Other Topics ($\sim$ November 9 - December 14)

- Bayesian Adjustment for Confounding
- Bayesian Sensitivity Analysis
- Mixture Models
- Missing Data
- Causal Inference

Part 4. Case Studies*

- Combining Information in Air Pollution Studies
- Meta Analysis of Global Childhood Undernutrition
- Bayesian Assessments of Hospital Quality of Medical Care
- Analysis of Genetic Pathways
- Bayesian Inference of Epistatic Interactions
  * Note: Case studies in Part 4 of the course will be scheduled throughout the year and will **not** appear as a single block at the end of the lecture schedule