CHL 5223H Applied Bayesian Methods
Winter 2019
Revised January 5, 2019

Instructor:	Michael Escobar
Office:	Room 658, Health Sciences Building
Phone:	(416) 978-7942
Email:	`m.escobar@utoronto.ca`
Website:	`http://individual.utoronto.ca/Escobar/`
Lecture time:	Monday, 2-5, in Health Science Building 106 (155 College Street)
Office hours:	TBA

1) Overview:

In the last 20 years, Bayesian methods have become an important tool for applied statisticians, biostatisticians, and data scientist. Previously, most Bayesian analyses were restricted to simple, limited applications. With the development of Markov Chain Monte Carlo methods, Bayesian inference has become an important applied technique and has been able to handle complex problems. Some problems are now easier to compute with Bayesian methods than with frequentist methods. This course will first explain the basics of Bayesian inference and Markov chain Monte Carlo methods. From there, this course will show how to compute and make inferences on complex data problems using these methods.

2) Teaching objectives:

a) Gain an understanding of basic Bayesian inference.

b) Gain an understanding of the basic theory of Markov chain Monte Carlo methods.

c) Gain proficiency in performing Bayesian data analysis on complex data problems.

3) Course prerequisites:

CHL 5202 (Introductory Biostatistics for Students in Biological Sciences II) or the equivalent. It is assumed that graduate students in either the Biostatistics or Statistics program meet this prerequisite. Please note that CHL 5202 contains an introduction to linear regression and logistic regression. It will be assumed that students are familiar with these methods.

4) Evaluation:

4) Evaluation: The grading will be based on Three (3) take home assignments and a final exam.

a) The first assignment is worth 20% and the other two assignments will be worth 25% each. For the assignments, there will be a 3 point per day penalty for late work (based on 100 total points for an assignments).

b) **The final exam will be held on April 8 from 2-5pm.** The final exam is worth 30%.

5) Format of instruction:

There will be a 3 hours of lecture each week. On most weeks, I will be available after class to talk to students one-on-one. There will be three assignments. Each problem set will include methodological and computational exercises. The final exam will be comprehensive over the entire course.

5.1 Communication with students– Quercus I am planning to use the university licensed product Quercus to communicate with the class. Students will find some of the lecture notes and homework there. One can access Quercus at `https:q.utoronto.ca`). The applied bayesian course will be accessible to any student who is officially signed up in the course. You will need your University of Toronto email (i.e. your UTORid account) to log in to the system.

5.2 Role of problem sets The only way to learn a new statistical and data analytical techniques is to do these techniques. Therefore, doing the homework problem sets is crucial to mastering these techniques. With this in mind, the usual sequence of instruction will to first have a lecture presentation on the basic material, methods and theory. Also a detailed example will be worked out, and a homework problem is passed out. The students are then expected to try the homework, using the lecture material, class notes, and the presented examples. I will usually begin my lecture sessions by answering and discussion difficulties in the homework. At this time the students should feel free to ask any questions about the assignment or course, especially computational issues.

Note 1: If you email me your assignment, DO NOT EMAIL ME A LARGE FILE (eg >10MB). If your file is that large, then either I misspecified what I want or perhaps you misinterpreted. If your email is that large, lets assume that I misspecified what I want and email me first to clarify what I want.

Note 2: You must justify all your answers and explain why what you are saying is correct. I or the grader have the right to assume that you are guessing and can mark an answer wrong if it is not clearly justified. That means you should show the code that was used or point out exactly where in the output the answer appears and what the answer is.

Assignment/Final schedule

	Assigned	Due	Weight
Assignment 1	January 14	February 4	20%
Assignment 2	February 4	March 4	25%
Assignment 3	March 4	Apr 1	25%
Final Exam	April 8, from 2-5pm		30%

Note: I need to submit grades for graduating students early. (? by April 16?)

5.3 Computing Computing is a vital part of this course. Most of the course will use a public domain software packages Bugs/WinBugs and R. These programs maybe freely obtained over the Web and runs on a variety of platforms including Window based machines. The program OpenBugs can be obtained at: *http://www.openbugs.net/w/FrontPage.* The statistical program/language, R, is public domain advanced statistical package and may be obtained at the website: *http://cran.r-project.org/.* It assumed that everyone has access to their own computing equipment (eg: laptop). If there is a problem, please contact me.

5.4 Supplemental Text book The supplemental text book for the course is: *The Bugs Book: A Practical Introduction to Bayesian Analysis* by Lunn, Jackson, Best, Thomas, and Spiegelhalter.

6) Timetable of Curriculum Topics:

This is a rough guide to the topics presented and dates. They might change a bit.

Week 1 (Jan 7): Basic Bayesian methods: Quantifying uncertainty as a probability measure. Priors: informed, reference/uninformative, skeptical and enthusiastic priors. Conjugate priors for the binomial model. Examples. Inferences from posterior: Credible regions versus Confidence intervals, Predictive inference. Plotting functions via the computer.

Week 2 (Jan 14): Basic Normal model: basic distribution. Conjugate priors. Posterior means and variance. Examples. Learning about distributions and functions of distributions via simulation.

Week 3 (Jan 21): Basic Markov chain convergence. Markov chain sampling: basic 2x2 Markov chain sampling and convergence. Singular value decomposition, first and second eigenvalues, and limiting distribution. Conditions needed for convergence and when are they "nearly" violated. The Markov Chain in Markov Chain Monte Carlo: What are the "states" for the Markov chain and extending from discrete states to infinite states.

Week 4 (Jan 28): Basic Gibbs sampling and metropolis sampling. Looking at the simple Gaussian model. Identifying the "states" of the Markov chain. Stationarity of the posterior distribution. Some basic convergence results. The conditionally independent hierarchical model. Some basic concerns for convergence: poor mixing ("nearly" separate), and slow convergence. Prototype "bad posterior surfaces" for MCMC: witches hat, ridges, and multimodes in the posterior. Introduction to Bugs and WinBugs.

Week 5 (Feb 4): MCMC applications with the simple Gaussian model. Making inferences from the output of the MCMC. Estimating posterior distribution. Basic Convergence diagnostics (graphical methods).

Week 6 (Feb 11): Analysis of Variance model using WinBugs. Two-way and multiway ANOVA models. Simple random effects models. Linear Regression. Trouble with parametrization. Inferences.

NO CLASS on Feb 18 due to Family Day/Reading Week

Week 7 (Feb 25): *[Note: Feb 25 is the drop date for PHS courses]* Categorical models and general linear models. Complex Gaussian models. Analysis of Variance and regression models. Growth curve, multilevel, and hierarchical models.

Week 8 (Mar 4): More convergence diagnostics including: Estimating MCMC accuracy (estimating standard error: time series methods, overlapping series, non-overlapping series). Gelman and Rubin (analysis of variance method) and Geweke (Time series methods). Cross correlations and lag statistics. CODA.

Week 9 (Mar 12): Model diagnostics: Model criticism. Cross-validation methods. Predictive ordinates. Global measures of goodness-of-fit. DIC statistic.

Week 10 (Mar 19): Model selection methods. Bayes factors.

Week 11 (Mar 26): Review for exam.

Week 12 (Apr 1): Advanced algorithm designs. Designing your own sampler. Samplers of the uniform. Samplers of other distributions. Acceptance rejection methods. General samplers for log concave distributions. Griddy Gibbs samplers. Reversible jump MCMC algorithms. Improved MCMC algorithms. Metropolis-Hasting algorithms

Week 13 (Apr 8): Final Exam. (location to be announced)

7) Course Material:

The material for the coures will be based on handouts and lecture notes. There is a supplemental textbook which is mentioned in section 5.4 above.