STA 6276 Statistical Computing II: Monte Carlo Methods
Spring 2023

Instructor Hani Doss—222 Griffin-Floyd; email: `doss@stat.ufl.edu` (email is primarily for administrative purposes, not for questions regarding the course material; for such questions, talk to me during office hours). Office Hours: MWF period 3, i.e. 9:35am–10:25am. If you want to talk to me during office hours you should come before 10:10am. Office hours will be in person most of the time, but occasionally will be on zoom. When office hours are in person, I allow only two students in my office at a time. Office hours will be on zoom if I think that many people will want to attend, for example just before an exam or before a homework assignment is due. I will send the zoom link via email. For short questions, you may talk to me in person right after class.

Teaching Assistant Cheng Zeng—email: `czeng1@ufl.edu`.

Course Description Monte Carlo methods are now used in virtually every scientific area, including statistical physics (where they originated), Bayesian and frequentist statistical inference, image reconstruction, and various parts of machine learning. The basic idea is to carry out a simulation to estimate quantities of interest that cannot be computed analytically. This course will begin with a brief discussion of some standard Monte Carlo schemes, before moving to Monte Carlo methods based on Markov chains.

Consider the situation where there is a distribution π on some space, and we are interested in estimating π or $\int f \, d\pi$ where f is some function, but π is analytically intractable. Markov chain Monte Carlo proceeds as follows. We set up a Markov chain with the property that its transition function has π as its stationary distribution. Then we run a chain $X_1, X_2, \ldots$ with this transition function. If the Markov chain converges to its stationary distribution (i.e. for large n, the distribution of X_n is approximately π), then by running the chain long enough, we can obtain a sample from π. This sample can be used to estimate π or some feature of it such as $\int f \, d\pi$.

In this course I will explain the method in detail, describe the main implementations, discuss some large classes of problems in statistics where it has had success (primarily in Bayesian inference), and illustrate the method on several real data sets. The method is not fool-proof. I will talk about some of the mathematical results pertaining to convergence issues, and also discuss some practical convergence diagnostics.

Course Web Page `http://users.stat.ufl.edu/~doss/Courses/mcmc`

Prerequisites/Corequisites STA 6326 (Introduction to Theoretical Statistics I) is a prerequisite and STA 6327 (Introduction to Theoretical Statistics II) is a corequisite. You also need to know some probability theory beyond what is covered in the Master's program, but I will go over the facts you need to know. Additionally, you need to be familiar with the statistical computing language R. I will not assume you know anything about Markov chains.

Grading Your course grade will be based on the four components below, with the stated weights.

Exam 1:	Friday February 17, 8:20pm.	25%
Exam 2:	Monday March 27, 8:20pm. Covers the material after Exam 1.	25%
Exam 3:	Wednesday May 3, 7:30am–9:30am. Covers the material after Exam 2.	25%
HW:	There will be about 7 homeworks assigned during the semester.	25%

Some of the homework assignments will be of a theoretical nature, and some will involve computer implementation of the methods we discuss on specific data sets. The solutions to the homework assignments must be entirely your own (this applies also to the R code).

Topics

- Issues in practical implementation of Bayesian statistics
- Illustrative example: censored data
- Basic Monte Carlo methods
- General idea of Markov chain Monte Carlo
- The Gibbs sampler (general properties; application to latent variable models, including hierarchical Bayesian models and censored data models; application to high-dimensional problems)
- Rao-Blackwellization and variants thereof
- Convergence diagnostics
- Application of the Gibbs sampler to nonparametric Bayes problems
- The Metropolis-Hastings algorithm (general properties; application to Ising model; random walk chains and independence chains; adaptive rejection Metropolis sampling)
- Hamiltonian Monte Carlo
- Theory of convergence (ergodic theorems and central limit theorems)
- There will be one additional in-depth application of MCMC to some nontrivial area in statistics, which is likely to be either Bayesian variable selection in regression or latent Dirichlet allocation in topic modelling.

General Course Policies

- All emails to me or the TA must have the string “mcmc” in the subject line (so I can retrieve emails quickly) and must be sent from your official UF mail account.
- Cell phones may not be used; they should be turned off (or set on silent). Laptops must be shut.
- If you have a disability and will request accommodations you should see me as early in the semester as possible.

Policies on Covid Safety

Illness If you are sick, stay home. If you are sick and need immediate care, call your primary care provider or the UF Student Health Care Center at 352-392-1161 to be evaluated.

Absences If you test positive, you should not come to class, and as with any excused absence, you will be given a reasonable amount of time to make up missed work.

Masks As of this writing, UF is recommending that we wear masks inside all UF facilities if we are close to other people.

During the lecture, I may or may not wear a mask, depending on whether or not the distance between me and the person closest to me is much greater than six feet. However, I will put on a mask immediately after the lecture, and if you want to talk to me, then you must wear a mask if you get within six feet of me. If you come to office hours, then you must wear a mask regardless of whether or not you are within six feet of me.