STAT 721: Stochastic Processes

Spring 2022

Instructor:	Ray Bai	**Time:**	1:10-2:00 pm MWF
Email:	RBAI@mailbox.sc.edu	**Place:**	Carolina Coliseum 2006

Course Page:

`https://blackboard.sc.edu/` (Check regularly for announcements and homework assignments)

Office Hours: By appointment. I am also very accessible by e-mail and will typically reply to e-mails within one business day of receiving them.

Course Description: Stochastic processes are probabilistic (non-deterministic) systems indexed by time or space. They are very useful for modeling a variety of phenomena that vary in a random manner, such as the volatility of financial assets or the spatial distribution of crimes in a city. Stochastic processes are also very useful for many machine learning tasks such as regression, classification, and clustering.

The course takes a modeling-based approach to the study of stochastic processes; that is, applications and practical implementation of these models are stressed, while any theory that is introduced is used to motivate the applications. The course will also introduce students to state-of-the-art, modern scalable computing techniques (e.g. global approximation methods and distributed computing), so that these methods may be implemented efficiently on "big data." The tentative schedule of topics is:

- **Weeks 1-2**: mathematical finance (Brownian motion, martingales, Bachelier's model and Black-Scholes model for asset pricing)
- **Weeks 3-4**: point processes (Poisson processes, Hawkes processes)
- **Weeks 5-8**: Markov chain Monte Carlo (Bayesian inference, Gibbs sampling, Metropolis-Hastings, slice sampling, Hamiltonian Monte Carlo, embarrassingly parallel MCMC)
- **Weeks 9-11**: Gaussian processes (GPs) for regression and classification, global approximation methods for GPs in big data
- **Week 12**: clustering with Bayesian nonparametrics (Dirichlet processes, Dirichlet process mixtures)
- **Week 13**: Kalman filter and extended Kalman filter
- **Week 14**: Group project presentations

Learning Outcomes:

1. Be able to apply stochastic processes to a variety of tasks, including regression, classification, and clustering.
2. Be able to implement the algorithms and models covered in the course.
3. Be able to design scalable algorithms for efficient computation on large data sets.
4. Be able to communicate effectively through writing scientific reports and making/delivering presentations.

Prerequisites: STAT 512 and MATH 544 or equivalent. Students should also be comfortable with a programming language such as R, Python, MATLAB, or C++.

Main References: We will use typed handouts prepared by the instructor. Parts of these lecture notes are *not* complete and will be filled in during lecture. Thus, it is in your best interest to attend lectures.

Computing: This course involves programming. Students may use any "standard" programming language of their choice, e.g. Python, R, MATLAB, C++, etc. Using a more obscure language is discouraged because I will not be as familiar with it.

Homework: There will be several homework assignments. Students will work in small groups, and each group will submit a single, typed report for each homework assignment. The homework may consist of both conceptual/theoretical exercises and questions that involve programming.

All homework reports should be typed, including answers to math exercises and data analysis portions that may include figures, tables, etc. Your code should be e-mailed to the instructor in a Zip file. The first few weeks of the class, groups should scour the Internet for a homework template to use. It is expected that students will be able to learn Latex on their own by following existing templates and using the Internet.

Project: Students will work in small groups to research a topic of their choosing, prepare a 15-20 minute presentation, and write a short report in the style of a journal article: abstract, introduction, method, data analysis, and a bibliography. The project must include an application to a real data set. The last week of the semester will be devoted to project presentations. Some potential examples of projects include:

- local approximation methods for Gaussian processes (e.g. mixture-of-experts)
- a scalable MCMC or filtering method that was not discussed in class
- self-exciting point processes for modeling crime
- nested Dirichlet processes
- queueing networks, applications to telecommunications
- Gillespie algorithm for simulating stochastic equations of molecular reactions
- other applications, including economics, finance, physics, biology, social sciences

Students are encouraged to pursue projects that are relevant to their current research or their research interests. Projects must be approved in advance by the instructor, and no two groups may do the same topic for their project. If you have an idea of what you want to do for your project, please "claim" it early. Detailed instructions for the presentation and the report will be given at a later date.

Grading: Your grade will be determined by homework (70%) and the project (30%). The tentative grading scale is as follows: 90-100 for an A, 80-89 for a B+, 70-79 for a B, 60-69 for a C+, 0-59 for a C.

Honor Code: See the Carolinian Creed in the *Carolina Community: Student Handbook and Policy Guide*. The *minimum* punishment for violations of the USC Honor Code is a grade of zero for the work in question. In accordance with university policy, there may be other punishments, including an automatic F in the class and/or expulsion from the university.

Accommodation: If you need special accommodations for any aspects of the course, please contact me before or during the first week of the semester.

Note that reasonable accommodations are available for students with a documented disability. If you have a disability and may need accommodations to fully participate in this class, contact the Office of Student Disability Services by phone (803-777-6142) or e-mail sasds@mailbox.sc.edu. All accommodations must be approved through the Office of Student Disability Services.