BCHM 145a Spring 2018

How to decide: Bayesian inference and computational statistics

Syllabus and course organization

Instructors: Douglas Theobald, Kosow 308, 781-736-2303, dtheobald@brandeis.edu
Jeff Gelles, Kosow 208, 781-736-2377, gelles@brandeis.edu
TA: Chie Ueda, caueda@brandeis.edu

Meeting time and location: Tues, Fri 9:30 AM–10:50 AM; location TBA

What is the course about?

The last several decades have seen a revolution in the development of new statistical methods used in experimental science (as well as in a wide range of other fields). Bayesian methods (which are based on a rigorous probability theory) together with other computationally intensive methods are supplanting traditional 19th- and early 20th century parametric statistical methods. The development of the traditional methods was dictated by the need to make statistical calculations computationally simple. There is no such need in the 21st century because of the ready availability of computers. Computational and Bayesian inference are now used in a wide variety of scientific fields, from single-molecule biophysics, to structural biology, to neurobiology, to political polling. Because of the rising prominence of these methods, they have become an essential part of the education of scientists. **The goal of the course to give the students the ability to use modern statistical methods in their own scientific research and to allow them to read and critically evaluate scientific literature that makes use of these methods.**

Topics to be covered

The course will cover the theory of the statistical methods, the implementation of the computational approaches, and the application of both to selected problems in biochemistry and molecular biophysics. A rough list (may change!) of the topics to be covered includes:

- Basic probability theory; distributions
- Descriptive statistics
- Bayes' Theorem
- Prior distributions
- Bayesian interpretation of conventional parametric statistics
- Maximum likelihood methods
- Expectation maximization algorithm
- Monte Carlo simulation
- Markov chain analysis; Metropolis-Hastings algorithm; Gibbs algorithm
- Error propagation by simulation
- Model selection
- Significance testing by bootstrap
- Jackknife and computational cross-validation
- Bayesian interpretation of computational statistics
- Applications
    - R-free in crystallography
    - Image classification in light and electron microscopy
    - Phylogenetics
    - Objective analysis of noisy data in single-molecule experiments

- Brownian dynamics simulations

Intended audience

This course is for undergraduate and graduate students at all levels. No previous formal training in statistics is required. The only pre-requisite is that **students must have high proficiency with both differential and integral calculus**. That means:

- You must be familiar with and have good intuition about the usual analytic functions — polynomials, exponentials, logarithms, trigs, etc.
- You must be able to differentiate and integrate these functions in your sleep.
- You must know basic calculus tricks (e.g., chain rule, integration by substitution, integration by parts, differentiation under the integral sign, Taylor expansions, finding maximums and minimums, etc).
- You will have to perform calculus derivations on exams under time constraints.

If you do not feel comfortable with these basic manipulations of univariate calculus, **you should not be in this course**.

No prior knowledge of computer programming is required; the course is designed to give a working knowledge of Matlab to all students regardless of previous computer experience. Students without prior knowledge should expect to spend additional time at the beginning of the course completing tutorials designed to familiarize them with MATLAB and/or programming.

Course meetings

Each course meeting will have one of two possible formats:

- Traditional lecture presentations of the course material.
- In-class computational exercise in which the students will work on their own laptops, in pairs or small groups, supported by the instructors and TA, to tackle a problem in data analysis to which they will apply the methods learned in the course.

Both the lectures and computational exercises will begin at 9:30 am sharp! Late arrivals are disruptive and may cause you to miss key points, so please make sure you arrive by 9:25. Attendance is required and class participation is part of your grade. The course is very "lecture-heavy" -- assigned readings supplement but do not duplicate the materials presented during the course meetings. You will need to attend every class to learn the material.

Questions

Both instructors welcome your questions and discussion both in and out of the lectures. We try to arrange our schedules so that we are free right after lectures, so that is often a convenient time to talk. You can also call or email either instructor or the TA to set up an appointment to talk with us outside of class. You are also welcome to come by the instructor's offices any afternoon if you would like to talk to one of us. If we cannot talk to you immediately when you come by, we will set up an appointment to help you later.

Recitation section

Chie will hold weekly meetings in which she will challenge students with additional problems, which they will work on individually or in small groups and then immediately discuss solutions. She is also available to provide help on problems sets and clarification of lecture material either during the recitation (if time permits) or at other times. We will try to schedule the recitation meetings at a time that does not conflict with other academic commitments of any registered students.

Readings

Most of the readings will be from handouts prepared by the instructors and from journal articles and selected book chapters. You need not purchase any textbooks for the course.

Computers

You will need a laptop computer that you can bring to class in order to do the computational exercises. You won't be able to participate if you forget your computer! Therefore, we strongly recommend that you get into the habit of bringing it to every course meeting, so that you will be sure to have it when you need it. Please try to have your laptop fully charged when you come to class, as it may be hard to access an electrical outlet in the classroom. You will need to install the current version of Matlab (some exercises use features that may not be available in older versions). Matlab is available free of charge through the Brandeis site license.

Problem sets

The material covered in this course cannot be understood just by listening to someone talk about it; **you have to try it out for yourself**. For that reason, the problem sets are the most important part of the course. There will be problem sets roughly every two weeks over the course of the semester.

There are two deadlines for each problem set. Problem sets must be turned in by the first deadline to receive for credit for effort (up to 50%). After the first deadline, an solution set will be posted on Latte. Students should then either revise each answer (if needed) or add a note that the initial answer was correct (if it was). The revised/annotated problem set must be turned in again by the second deadline to receive credit for correctness (up to 50%). Problem sets should be uploaded to Latte as a single PDF-format document. Scanners to convert hardcopy documents into PDFs are available outside of the Biochemistry office, in the library, and at other locations on campus.

Students are welcome to collaborate with other students on the problem sets if you feel that this will help your understanding of the material. However, please do not turn in any work unless **you personally understand** how the answer was obtained.

Exams and quizzes

There will be two 50-minute in-class exams held during regular class hours. In addition, occasional short, in-class quizzes may be given during the semester.

Unless otherwise specified on the exam paper, exams and quizzes are "closed book", meaning you may not consult any notes or other reference materials during the exam.

Final project

In lieu of a final exam, students will complete a final project in which they analyze experimental data (taken by the student or by someone else) using techniques taught in the course and submit a written report on the results.

Grades

Final grades will be determined based on class participation (~10%), problem sets (~20%), and final project (~20%), and exams and quizzes (~50%).

Expectations and Warnings

WORKLOAD: This is a challenging course taught at a level appropriate for graduate and advanced undergraduate students. It is a four-credit course (with three hours of class-time per week). Success in the course is based on the expectation that students will spend a **minimum** of nine hours of study time per week in preparation for class (readings, problem sets, discussion section, preparation for exams, etc.).

MATHEMATICAL EXPECTATIONS: We emphasize again (see “Intended Audience” above): this entire course uses abundant math/calculus throughout. You have been warned.

MISSED WORK AND MAKEUP EXAMS: If you miss an exam, you will receive a failing grade. As it is essentially impossible to prepare two equivalent exams in this subject, no written makeup exams will be given for any reason. In exceptional circumstances in which a student is absent from an exam due to a documentable emergency beyond the student’s control, an oral exam will be given. **Pre-purchased travel arrangements are not considered a valid reason for missing an exam.** Students with interviews for graduate programs that conflict with exam dates should inform the instructors immediately upon learning of the conflict.

POLICY ON ACADEMIC HONESTY: All students (both undergraduate and graduate) are expected to be familiar with, and to follow, the University’s policies on academic integrity. Please consult the Brandeis University *Rights and Responsibilities* document (https://www.brandeis.edu/studentlife/srcs/rightsresponsibilities/2017-18%20RR.pdf) for all policies and procedures. All policies related to academic integrity apply to all projects, assignments, exams, and quizzes. Students may only collaborate on assignments with permission from the instructor. Instances of suspected academic dishonesty will be forwarded to the Director of Academic Integrity. Sanctions for academic dishonesty can include failing grades and/or suspension from the university.

STUDENTS WITH DISABILITIES: If you are a student with a documented disability on record at Brandeis University and wish to have a reasonable accommodation made for you in this class, please see one of the instructors immediately. Accommodations cannot be made retroactively.