WELCOME!

Instructor: [Pierre E. Jacob] (pjacob@fas.harvard.edu, call me "Pierre", pronounced like "Pea Air"), tenure-track faculty in the Department of Statistics, Harvard. Please send me an email with any question about this course.

Lectures: MW 9am Eastern US time.

Course Webpage: https://canvas.harvard.edu/courses/85301

Teaching Fellow: Rohit Chaudhuri (PhD student in Statistics, 3rd year).

Calendar: here's what's coming! All times are US Eastern Time.

Coming next:

- Wednesday, May 5: 2pm-3pm: office hours.
- Project due: May 5. **If you are working as a group, it's fine to upload your final project only once. Make sure the names of all group participants are written on the report.**
- **The end**! Your final grades will be communicated to you by email, before May 15.

Grading & Assessment:

The final grade will be 70% assignments and 30% final project. Participation in lectures and sections is strongly encouraged but we won't grade it directly. The assignments are provided as three lists of exercises and projects,each related to a different chapter of the lecture notes. The first list is already available in Files/Exercises/. The other two lists will be provided in due time. We require that each student works on a total of 10 exercises or projects (outside of the final project) and that there should be at least two taken from each list.

- ~~The assignments related to Chapter 1 will be due on February 26.~~
- ~~The assignments related to Chapter 2 will be due on March 26.~~
- ~~The assignments related to Chapter 3 will be due on April 21.~~

The students will work on their exercises, projects & final projects either on their own or in small groups of two or three, which can be different groups each time. Students can reach out to the staff if they are looking for groups. The students are encouraged to propose their own ideas of projects or exercises and check with the staff to validate their ideas. Also, we encourage the students to try the more "open-ended", "research-y" projects, which might be more intimidating but also certainly more rewarding.

The final, bigger project should keep you busy during the month of April. Its due date is May 5. There are details and a list of suggested projects here.

Description:

This new course is about Monte Carlo methods and the method of coupling in probability. Monte Carlo methods are algorithms that generate samples distributed according to probability distributions and have found countless applications in physics, chemistry, statistics, machine learning to name a few fields. The course includes a self-contained introduction to Monte Carlo methods, followed by in-depth coverage of the coupling method for both theoretical and practical purposes. On the theoretical side, the course covers the celebrated "coupling inequality" and its ramifications, and coupling interpretations of different notions of distances between probability distributions, with connections to optimal transport. Concretely this allows studying the performance of Monte Carlo algorithms. On the methodological side, the course covers implementable coupling methods to monitor the convergence of Markov chain Monte Carlo algorithms, to parallelize computation, and for variance reduction. The course is intended for students interested in learning about computation in statistics, applied probability, numerical methods, and algorithms that are routinely used to compute volumes and integrals across many scientific fields.

Materials: lecture notes, pieces of code in R, articles, put in the Files/ tab (on the left).

An important point: we are all affected by the on-going pandemic in different but significant ways. Please keep that in mind when you interact with your peers and with the instructional staff. Let's make this course an inspiring, stimulating and peaceful experience for all.

Prerequisites:

Basic notions of mathematics, probability and algorithms. For example Stat 210 or at least 110, and, at least as important, a good motivation to learn more on these topics. Prospective students can reach out to the instructor to discuss prerequisites. Stat 221 is not required. Prospective students are invited to have a look at the first two chapters of the lecture notes, available in Files/, before registering.

Outline:

The course involves a mix of conceptual/theoretical components and more algorithmic/methodological components. A rough outline goes as follows (which is also the structure of the lecture notes).

1. Introduction to Monte Carlo methods, rejection sampling, importance sampling, MCMC methods such as Metropolisâ€“Hastings, Langevin, Hamiltonian Monte Carlo. These methods will be motivated by problems arising in statistics and elsewhere. General references include the books of Jun S. Liu, or Robert & Casella.

2. Couplings of random variables, the celebrated "coupling inequality". The relation between couplings and distances between probability distributions. Some connections to optimal transport (as in Monge/Kantorovich/Wasserstein distances, see PeyrÃ© & Cuturi 2018). General references include Lindvall, or Thorisson.

3. Coupling of Markov chains to analyze their convergence, ergodicity of MCMC algorithms. References include the articles of Jerrum 1998, Roberts & Rosenthal 2004.

4. Monte Carlo methods employing coupled Markov chains, including Coupling From The Past (Propp & Wilson 1996), debiasing techniques (Glynn & Rhee 2014, Jacob, O'Leary & AtchadÃ© 2020), diagnostics of convergence (Johnson 1996, Biswas, Jacob & Vanetti 2019), variance reduction (Pinto & Neal 2001). There is no comprehensive reference on this topic yet, but the lecture notes will try to fill this gap.

Goals:

- Foundations: probability, metrics on the space of probability distributions, convergence analysis, algorithmic complexity, stochastic process

- Skills: proof techniques, numerical integration, coding in R, python or similar, design of algorithms, design of numerical experiments, visualization