STOR 890 [Spring 2014]: Concentration inequalities and Combinatorial optimization.

Times: Monday, Wednesday 9:30-10:45.

Description:

Concentration inequalities, namely quantifying fluctuations of functions of independent (and in some cases dependent) variables from their mean, now play a fundamental role in many different fields including probabilistic combinatorial optimization, model selection and empirical processes in statistics, statistical physics (spin glass) and computer science (performance of randomized algorithms, machine learning) and have connections to a diverse number of disciplines in mathematics including convex analysis and information theory. The aim of this course is to give researchers a rigorous grounding in the basics of this subject and expose them to a select few of the diverse applications including understanding the behavior of simple optimization algorithms over random data. See the planned list of topics to be covered below for more details.

Prerequisites: Upper division undergraduate probability theory. In one lecture we might use the definition and some properties of martingales.

Requirements for students:

1. Read and present in class a recent research paper. The length of each presentations will be determined once we know the size of the class. The presentations will be towards the end of the semester (last 2 weeks).

2. Every week I will assign a few problems as homework. These will not be collected but are just to make sure you understand the material.

3. Scribe a few lectures. You can work in groups of two. Writing up lectures including filling in ideas of the proof that I give in class.

Why you should take this course: Concentration inequalities have now evolved into a fundamental topic in probability theory. The ideas involved are intricate and deep. In my humble opinion, learning this in a course as opposed to trying to master them by yourself later is far more efficient. The mathematical ideas used to prove many of the results are intrinsically beautiful.

Format: We will mainly follow parts of the following two books:

a) *Probability Theory and Combinatorial Optimization by Michael Steele [S].*

b) *Concentration Inequalities: A non asymptotic theory of independence by Boucheron, Lugosi and Massart [BLM].*

Towards the end of the course, when trying to understand such techniques in more dependent settings, we will look into

c) *Papers by Sourav Chatterjee [SC].*

<u>Planned list of topics:</u> I will try to introduce the main ideas underlying the following topics.

1. Chapter 1 of [S]: Simple models of combinatorial optimization over random data. Longest common subsequences, subadditivity, Azuma's inequality. Dynamic programming.

2. Chapter 2 of [S]: Why we need concentration of measure in combinatorial optimization.

3. Chapter 1 and 2 of [BLM]: Basics of concentration inequalities.

4. Chapter 3 of [BLM]: Bounding the variance. Efron-Stein inequality. Gaussian Poincare inequality.

5. Chapter 4 of [BLM]: Information inequalities. Shanon Entropy and Relative entropy. Han's inequality. Duality and variational formulas.

6. Chapter 5 [BLM]: Log Sobolev inequalities for Bernoulli and Gaussian. Gaussian random projections. Performance bound for Lasso. Hypercontractivity. Largest eigen value of Random matrices.

7. Chapter 6 of [BLM]: Entropy method. Weakly self-bounding functions.

8. Chapters 7 and 8 [BLM]: Concentration and Isoperimetry and Talagrand's famous transportation inequality.

8. Chapter 9 of [BLM]: Influences and Threshold phenomena. Discrete Fourier analysis.

9. Concentration inequalities in the dependent setting [SC]: We will then study some of Chatterjees techniques for establishing concentration inequalities when one has dependence.