THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

SAYAN MUKHERJEE

Department of Statistical Science
Duke University

Stochastic Topology: It May Be Useful?

MONDAY, March 9, 2015, at 4:00 PM
Eckhart 133, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110

ABSTRACT

I will give three vignettes where topological ideas are explored in statistical models of complex traits, machine learning, and probability.

The (quantitative) genetics of morphological traits: In biological and medical applications, data are being collected in the form of meshes, for example, CT scans of bones or organs. Modeling these surfaces is an interesting challenge. I describe how topological summaries of the data can be used to measure distances between surfaces and state likelihood models for surfaces, all without the use of landmarks. Topics such as sufficient statistics and dictionary learning will be touched on. I will describe an application in evolutionary morphology where the objective is to compute the distances between the calcanei of 106 extinct and extant primates.

Machine learning with higher-order interactions: The graph Laplacian and random walks on graphs has impacted machine learning. Examples include label propagation and Laplacian eigenmaps for semi-supervised learning and dimension reduction. I will motivate why it is of interest to extend these graph based algorithms to simplicial complexes, which capture higher-order relations. I will describe recent efforts to define random walks on simplicial complexes with stationary distributions related to the combinatorial (Hodge) Laplacian. This work will touch on higher-order Cheeger inequalities, an extension of label propagation to edges or higher-order complexes, and a generalization of results for near linear time solutions for linear systems.

Percolation: Given n points down from a point process on a manifold, consider the random set which consists of the union of balls of radius r around the points. As n goes to infinity, r is sent to

---

zero at varying rates. For this stochastic process, I will provide scaling limits and phase transitions on the counts Betti numbers and critical points.

Joint work with: Kate Turner, Doug Boyer, Lek-Heng Lim, John Steenbergen, Carly Klivans, Omer Bobrowski