BIOSTATISTICS SEMINAR

‘ Harnessing the unseen for next generation population genomics and epigenomics’

James Zou, PhD.

Microsoft Research New England
Cambridge, MA.

ABSTRACT:

Sequencing of large human populations has the potential to transform disease diagnosis and treatment. In order to harness the power of this data avalanche, it is crucial to model and leverage the data and covariates that we do not see. I will illustrate this concept with two examples in genomics and epigenomics, where I developed scalable statistical algorithms with strong mathematical guarantees. I will first discuss my close collaboration with the largest exome sequencing consortium (ExAC) to infer statistical properties of rare and unseen human genetic variations. This work provides a unified framework to quantify the natural selection acting on our genome, annotate disease variants, and predict the discovery rate of future sequencing projects. In the second part, I will describe complementary work to identify changes in the packing and chemical modifications of DNA—i.e., epigenomic variation—that are associated with diseases. This work requires flexible models of unseen covariates, especially cell-type composition. I will conclude by discussing the general statistical lessons we have learned and new research directions.

12:00 noon Monday, February 1, 2016
LEPH 216, 60 College Street
Lunch