

HARVARD UNIVERSITY
DEPARTMENT OF STATISTICS



Tel. 617-495-5496
Fax. 617-496-8057
statdept@stat.harvard.edu
<http://www.stat.harvard.edu/>

Science Center 7th Fl.
One Oxford Street
Cambridge, MA 02138

STATISTICS COLLOQUIUM

MONDAY, SEPTEMBER 14, 2015
TALK: 4:15 PM — SCIENCE CENTER 300H
RECEPTION: 3:50 PM

“Hypothesis Testing for Detecting Dense and Sparse Signals in Analysis of Massive Whole Genome Sequencing Data”

[Xihong Lin](#)

**Departments of Biostatistics and Statistics
Harvard University**

ABSTRACT

Massive genetic and genomic data present many exciting opportunities as well as challenges in data analysis and result interpretation, e.g., how to develop effective strategies for signal detection using massive genetic and genomic data when signals are weak and sparse. Many variable selection methods have been developed for analysis of high-dimensional data in the statistical literature. However limited work has been done on statistical inference for massive data. In this talk, I will discuss hypothesis testing for analysis of high-dimensional data motivated by gene, pathway/network based analysis in genome-wide association studies using arrays and sequencing data. I will focus on signal detection when signals are weak and sparse, which is the case in genetic and genomic association studies. I will discuss hypothesis testing for signal detection using variable selection based penalized likelihood based methods and aggregated marginal test statistics based methods using the Generalized Higher Criticism (GHC) test. The results are illustrated using data from genome-wide association studies and sequencing studies.