University of California, Irvine
Statistics Seminar

Hypothesis Testing for Large-Scale Data: Enhancing Reliability and Efficiency

Yinqiu He
PhD candidate, University of Michigan

4 p.m., Thursday, January 14, 2021

In scientific research that involves large-scale data, researchers often start with questions regarding the global properties of a large set of measurements. For instance, are a group of related genes in the same functional pathway jointly associated with a trait of interest? Such questions can be formulated as hypothesis testing problems that globally examine a large number of parameters in a high-dimensional joint distribution. Examples include hypothesis testing on mean vectors, covariance matrices and regression coefficients. To extract informative scientific knowledge from abundant data, reliability and efficiency are among the major concerns in statistical inference.

In this talk, I will address particular reliability and efficiency issues arising from jointly testing a large number of parameters. First, I will discuss how reliable the popular likelihood ratio tests (LRTs) are in terms of the type I error control for high-dimensional data. I will provide theoretical insights into the reliability of the LRTs in a variety of problems, which are based on phase transition results of the foundational Wilk's theorem. Next, to improve efficiency of the existing testing procedures under high-dimensional settings, I will introduce a new adaptive testing framework that can maintain high statistical power against a wide range of alternative hypotheses. The proposed framework is based on a family of U-statistics that are constructed to capture the information in different directions in high-dimensional spaces. For a broad class of problems, we establish high-dimensional asymptotic theory for the U-statistics and develop adaptive testing procedures that are statistically powerful in a wide variety of scenarios.