University of California, Irvine
Statistics Seminar

Oracle P-values for High Dimensional Regression Models

Helen Zhang
Professor of Mathematics
University of Arizona

4 p.m., Thursday, May 20, 2021
Join via Zoom: https://uci.zoom.us/j/93841331474

The concept of P-value, proposed by Fisher to measure inconsistency of data with the specified null hypothesis, plays a central role in statistical inference. For classical linear regression models, it is a standard procedure to calculate P-values of regression coefficients based on their least squares estimators (LSEs) to determine their significance. However, for high dimensional data when the number of predictors exceeds the sample size, ordinary LSEs are no longer proper and there is not a valid definition for P-values based on OLS estimates. In this paper, we introduce the concept of oracle P-values to generalize P-values from traditional settings to high dimensional sparse regression models and establish their validity. We then propose scalable methods to estimating oracle P-values and develop a new host of tools based on oracle P-values for variable screening and variable ranking under the false discovery rate (FDR) control.