MICHIGAN STATE UNIVERSITY
Department of Statistics and Probability

COLLOQUIUM

Seunggeun (Shawn) Lee
University of Michigan

Asymptotic properties of principal component analysis and shrinkage-bias adjustment in high-dimensional data

Tuesday, September 20, 2016
10:20 a.m. - 11:10 am
Refreshments 10:00 am
C405 Wells Hall

Abstract

With the development of high-throughput technologies, principal component analysis (PCA) in the high-dimensional regime is of great interest. Extensive efforts have been made to investigate the asymptotic behaviors of PCA in high-dimensional settings. Most of the existing theoretical and methodological results are based on the spiked population model in which all the population eigenvalues are equal except for a few large ones. Due to the presence of local correlation among features, however, this assumption may not be satisfied in many real-world datasets. To address this issue, we investigated the asymptotic behaviors of PCA under the generalized spiked population model. Based on the theoretical results, we proposed a series of methods for the consistent estimation of population eigenvalues, angles between the sample and population eigenvectors, correlation coefficients between the sample and population principal component (PC) scores, and the shrinkage bias adjustment for the predicted PC scores. Using numerical experiments and real data examples from the genetics literature, we showed that our methods could reduce bias and improve prediction accuracy.

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.