COLLOQUIUM

Department of Statistics and Probability
Michigan State University

Qing Wang
Pennsylvania State University

Topics in U-Statistics and Estimation

Thursday, February 9, 2012
A405 Wells Hall
10:20 a.m. - 11:10 a.m.
Refreshments: 10:00 a.m.

Abstract

The primary goal of my research has been to develop new methods, including theory and practical implementation, in the area of U-statistics. This area is quite old, with many important results first appearing in Hoeffding (1948). There have been many applications of U-statistics in nonparametric statistics. One area that is quite modern and active is cross-validation and risk estimation, although it has not traditionally been thought of as a U-statistic area.

As our motivating problem, we consider risk estimation in the context of nonparametric kernel density estimation. We consider U-statistic form estimators for the risk that arises from L2 and Kullback-Leibler loss functions. These risk estimators can then be used to select the bandwidth in the kernel density estimator. In order to overcome the large sampling variation of the cross-validation bandwidth selector, we propose a two-stage, "subsampling-extrapolation", bandwidth selection procedure that can help to reduce the variability of the traditional bandwidth selector dramatically. In this context, we also ask how one would assess the variance of the U-statistic that estimates the risk. This problem is quite general: we are interested in how we could estimate the variance of a general U-statistic when it is used as an unbiased estimator of the parameter of interest theta that has the form of theta=E[phi(X1,...,Xk)], where phi is a symmetric function of k arguments, and X1,...,Xk are i.i.d. from some distribution F. We devised an unbiased variance estimator for a general U-statistic which can be written in a quadratic form of the kernel function; it is valid as long as the kernel size k?n/2 (n is the sample size). In addition, this variance estimator is the best unbiased estimator and can be represented in a familiar ANOVA form as a contrast of between-class and within-class variation. As a further step to make our proposed methods more practical, we developed three subsampling schemes that can be used to realize the proposed estimators unbiasedly and more efficiently.

This is joint work with Dr. Bruce Lindsay.

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