Department of Statistics

STATISTICS COLLOQUIUM

ALEXANDRE TSYBAKOV

CREST-ENSAE
CNRS

Optimal and Adaptive Variable Selection

MONDAY, October 2, 2017 at 4:30 PM
Eckhart 133, 5734 S. University Avenue
Refreshments before the seminar at 4:00PM in Jones 111

ABSTRACT

We consider variable selection based on n observations from a high-dimensional linear regression model. The unknown parameter of the model is assumed to belong to the class S of all s-sparse vectors in R^p whose non-zero components are greater than $a > 0$.Variable selection in this context is an extensively studied problem and various methods of recovering sparsity pattern have been suggested. However, in the theory not much is known beyond the consistency of selection. For Gaussian design, which is of major importance in the context of compressed sensing, necessary and sufficient conditions of consistency for some configurations of n,p,s,a are available. They are known to be achieved by the exhaustive search decoder, which is not realizable in polynomial time and requires the knowledge of s. This talk will focus on the issue of optimality in variable selection based on the Hamming risk criterion. The benchmark behavior is characterized by the minimax risk on the class S. For Gaussian design, we propose an adaptive algorithm independent of s,a, and of the noise level that nearly attains the value of the minimax risk. This algorithm is the first method, which is both realizable in polynomial time and is consistent under almost the same (minimal) sufficient conditions as the exhaustive search decoder.

This talk is based on a joint work with C.Butucea, M.Ndaoud and N.Stepanova.
