

# STATISTICS COLLOQUIUM

TUESDAY, FEBRUARY 20, 2007

TALK: 4:00 PM — SCIENCE CENTER RM. 705

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

## “Sure Independence Screening for Ultra-High Dimensional Feature Space”

Jinchi Lv

Department of Mathematics  
Princeton University

### ABSTRACT

High dimensionality is a growing feature in many areas of contemporary statistics. Variable selection is fundamental to high-dimensional statistical modeling. For problems of large or huge scale  $p_n$ , computational cost and estimation accuracy are always two top concerns. In a seminal paper, Candès and Tao (2007) propose a minimum  $l_1$  estimator, the Dantzig selector, and show that it mimics the ideal risk within a logarithmic factor  $\log p_n$ . Their innovative procedure and remarkable result are challenged when the dimensionality is ultra high: the factor  $\log d_n$  can be large and their uniform uncertainty condition can fail.

Motivated by these concerns, in this paper we introduce the concept of sure screening and propose a fast and straightforward method via iteratively thresholded ridge regression, called Sure Independence Screening (SIS), to reduce high dimensionality to a relatively large scale  $p_n$ , say below sample size. An appealing special case of SIS is the componentwise regression. In a fairly general asymptotic framework, SIS is shown to possess the sure screening property for even exponentially growing dimensionality. With ultra-high dimensionality reduced accurately to below sample size, variable selection becomes much easier and can be accomplished by some refined lower-dimensional methods that have oracle properties. Depending on the scale of  $d_n$ , one can use, for example, the Dantzig selector or Lasso, the fine method of SCAD-penalized least squares in Fan and Li (2001), or the adaptive Lasso in Zou (2006).

This talk is based on collaborated work with Professor Jianqing Fan.