

Courant Institute of Mathematical Sciences
Mathematics Colloquium
September 20, 2010

Speaker: **Jianqing Fan**, Princeton University

Title: A two-scale framework to variable selection in NP-dimensionality

Abstract: Ultrahigh-dimensionality characterizes many contemporary statistical problems from genomics and engineering to finance and economics. We outline a unified framework to ultrahigh dimensional variable selection problems: Iterative applications of vast-scale screening followed by moderate-scale variable selection.

The framework is widely applicable to many statistical contexts: from multiple regression, generalized linear models, survival analysis to machine learning and compress sensing. The fundamental building blocks are marginal variable screening and penalized likelihood methods. How high dimensionality can such methods handle?

How large can false positive and negative be with marginal screening methods?

What is the role of penalty functions? This talk will provide some fundamental insights into these problems. The focus will be on the sure screening property, false selection size, the model selection consistency and oracle properties. The advantages of using folded-concave over convex penalty will be clearly demonstrated. The methods will be convincingly illustrated by carefully designed simulation studies and the empirical studies on disease classifications using microarray data and forecast home price indexes at zip level.