

STATISTICS COLLOQUIUM

MONDAY, APRIL 21, 2008

TALK: 4:00 P.M. — Science Center, Rm. 309

RECEPTION: 5:15 PM – SCIENCE CENTER, 7TH floor

“A method of discovering influential variables”

Shaw-Hwa Lo
Department of Statistics
Columbia University

ABSTRACT

A trend in all scientific disciplines, based on advances in technology, is the increasing availability of high dimensional data in which are buried important information. A current urgent challenge to statisticians is to develop effective methods of finding the useful information from the vast amounts of messy and noisy data available, most of which are noninformative. We present a general computer intensive approach, built on our earlier work, of many potential explanatory variables, have an influence on a dependent variable Y . This approach is suited to detect influential variables, where causal effects depend on the confluence of values with other variables. It has the advantage of avoiding a difficult direct analysis involving possibly thousands of variables, by dealing with many randomly selected small subsets. The main objective is to discover the influential variables, rather than to measure their effects. Once they are detected, the problem of dealing with a much smaller group of influential variables should be vulnerable to appropriate analysis. In a sense, we are confining our attention to locating a few needles in a haystack. Examples will be provided.

This is a joint work with Herman Chernoff and Tian Zheng