Department of Statistics - Statistics Seminar – Spring 2011

Statistics Seminars are on Mondays
Time: 12:00 - 1:00 PM
Location: Room 903, 1255 Amsterdam Avenue,
Tea and Coffee will be served before the seminar at 11:30 AM, Room 1025

_____

3/7/2011

Dr. Tony Cai, Department of Statistics, The Wharton School, University of Pennsylvania

Title: On Optimal Estimation of a Nonsmooth Functional

Abstract:
In this talk I will discuss some recent work on optimal estimation of nonsmooth functionals. These problems exhibit some interesting features that are significantly different from those that occur in estimating conventional smooth functionals. This is a setting where standard techniques fail. I will discuss a newly developed general minimax lower bound technique that is based on testing two fuzzy hypotheses and illustrate the ideas by focusing on the problem of optimal estimation of the l_1 norm of a high dimensional normal mean vector. An estimator is constructed using approximation theory and Hermite polynomials and is shown to be asymptotically sharp minimax.

This is joint work with Mark Low.

3/21/2011

Monday, March 21, 2011

Dr. Shuangge (Steven) MA, School of Public Health, Yale University

Integrative Analysis of Cancer Genomic Data

In cancer genomic studies, markers identified from the analysis of single datasets often suffer from a lack of reliability because of the small sample sizes. A cost-effective remedy is to pool data from multiple comparable studies and conduct integrative analysis. Integrative analysis of multiple datasets is challenging because of the high dimensionality of markers and, more importantly, because of the heterogeneity among studies. We consider penalized approaches for marker selection in the integrative analysis of multiple datasets. The proposed approaches can effectively identify markers with consistent effects across multiple studies and automatically accommodate the heterogeneity among studies. We establish the asymptotic consistency properties, conduct simulations, and analyze pancreatic and liver cancer studies.

3/28/2011

Dr. Yoram Singer, Google

Title: Entire Relaxation Path for Maximum Entropy Models

Abstract:
We describe a relaxed and generalized notion of maximum entropy problems for multinomial distributions. By introducing a simple re-parametrization we are able to derive an efficient homotopy tracking for the entire relaxation path. The end result is an algorithm that can provide optimal probabilistic estimates for any relaxation parameter using linear space and sub-lineartime. We also show that the Legendre dual of the relaxed maximum entropy problem is the task of finding the maximum-likelihood estimator with L1 regularization. Hence, our solution can be used for problems such aslanguage modeling with sparse parameter representation. We describe a simple large deviation bound that holds for any relaxation parameter and conclude with a demonstration and a discussion of potential applications.

4/4/2011

Dr. Jinchi Lv, USC

Title: Non-Concave Penalized Likelihood with NP-Dimensionality

Abstract:

Penalized likelihood methods are fundamental to ultra-high dimensional variable selection. How high dimensionality such methods can handle remains largely unknown. In this paper, we show that in the context of generalized linear models, such methods possess model selection consistency with oracle properties even for dimensionality of Non-Polynomial (NP) order of sample size, for a class of penalized likelihood approaches using folded-concave penalty functions, which were introduced to ameliorate the bias problems of convex penalty functions. This fills a long-standing gap in the literature where the dimensionality is allowed to grow slowly with the sample size. Our results are also applicable to penalized likelihood with the L_1-penalty, which is a convex function at the boundary of the class of folded-concave penalty functions under consideration. The coordinate optimization is implemented for finding the solution paths, whose performance is evaluated by a few simulation examples and the real data analysis. This is a joint work with Professor Jianqing Fan.

4/18/2011

Dr. Doug Nychka, National Center for Atmospheric Research
http://www.image.ucar.edu/~nychka/

"Surfaces for spatial data: a framework to understand
how Kriging and splines are just weighted averages."

Kriging is a non-parametric regression method used in geostatistics for estimating curves and surfaces for spatial data. Researchers in applied mathematics will recognize this method as related to radial basis function expansions for noisy data. The lack of statistical large sample theory for these very useful methods is a contrast to the well developed mathematical analysis of kernel smoothers. This talk outlines an approach to understand the mathematical properties of Kriging and along the way we also encounter splines. Some applications to climate temperature fields are given to illustrate the practical value of these methods.

4/25/2011

Dr. Naisyin Wang, University of Michigan

Title: Some Aspects of Analyzing Longitudinal or Functional Data Using Functional Data Analysis Methods

Abstract:

In this talk, we will focus on regression analysis that links functional or longitudinal covariate processes to a primary endpoint. The main regression methodologies we consider are originated from traditional functional data analysis. We will discuss various issues that could play a role in the outcomes of analysis. Some key focus would be on the choices of basis functions, global or local oriented, based on the goals of the study and other issues such as model determinations trailing from the decision. We will discuss asymptotic properties of the estimators we propose. Numerical outcomes from simulation and data analysis studies are used to illustrate our findings. The talk contains results from joint works with R.J.Carroll, Y.Li, N.Y. Wang and J.H.Zhou.

5/2/2011

Speaker: Deepak Agarwal

Affiliation: Yahoo! Research, Santa Clara, CA.

"Recommender Systems --- The Art and Science of Matching Items to Users."

Abstract

Algorithmically matching items to users in a given context is essential for the success and profitability of large scale recommender systems like content optimization, computational advertising, web search,

shopping, movie recommendation and so on. A key statistical problem that is essential to the success of such systems is to estimate response rates of some rare event (e.g. click-rates, buy rates, etc) when users interact with items. This is a very high dimensional estimation problem since data is obtained by interactions among several heavy-tailed categorical variables.

In this talk, I will discuss statistical techniques based on large scale multi-level hierarchical models, some of which have been deployed and are successfully recommending articles and ads to users on Yahoo! websites. The methods described are based reduced rank logistic regression, probabilistic matrix factorization, supervised Latent Dirichlet Allocation,
and multi-hierarchy smoothing.

About the Speaker

Deepak Agarwal is a statistician at Yahoo! who is interested in developing statistical and machine learning methods to enhance the performance of large
scale recommender systems. Deepak and his collaborators significantly improved article recommendation on several Yahoo! websites, most notably on
the Yahoo! front page. He also works closely with teams in computational advertising, yet another large scale recommender system. He serves as
associate editor for the Journal of American Statistical Association and has received four best paper awards in the past.