Department of Statistics - Statistics Seminar – Fall 2014

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

9/8/2014

Tingting Zhang (University of Virginia)

"A Dynamic Directional Model for Effective Brain Connectivity using Electrocorticographic (ECoG) Time Series"

9/15/2014

Peder Olsen (IBM Research, NY)

9/22/2014

Thomas Lee (UC Davis)

"Fiber Direction Estimation in Diffusion MRI"

9/29/2014

Qiwei Yao (LSE, UK)

"Segmenting Multiple Time Series by Contemporaneous Linear Transformation"

We seek for a contemporaneous linear transformation for a p-variate time series such that the transformed series is segmented into several lower-dimensional subseries, and those subseries are uncorrelated with each other both contemporaneously and serially. The method boils down to an eigenanalysis and, if p is large, a permutation in terms of maximum cross-correlations or FDR based on multiple tests. The asymptotic theory is established for both fixed p and diverging p when the sample size n goes to infinity, reflecting the fact that the method also applies when the dimension p is large in relation to n. Numerical experiements with both simulated and real datasets indicate that the proposed method is an effective initial step in analysing multiple time series data, which leads to substantial dimension-reduction in modelling and forecasting high-dimensional linear dynamical structures.

10/6/2014

Ping Li (Rutgers)

"BigData: Hashing Algorithms for Large-Scale Search and Learning."

Abstract: The talk will begin with an interesting story about Cauchy distribution. Consider two data vectors, u and v, and a vector R of i.i.d. Cauchy variables. Pr{sgn(<u,R>) = sgn(<v,R>)} is essentially a monotonic function of the chi-square similarity (a nonlinear kernel) between u and v. This observation leads to useful bigdata (LINEAR) algorithms for building large-scale statistical models and searching for

near neighbors, in terms of the chi-square similarity (kernel). Chi-square similarity has been known for its superb performance in data generated from histograms (e.g., computer vision and NLP).

Modern applications of internet search and machine learning routinely encounter datasets with (hundreds of) billions of examples in billion or even billion square dimensions (e.g., documents represented by high-order n-grams). Developing novel algorithms for efficient search and machine learning has become an active area of research. Hashing can be very useful in many scenarios, for example,

(1) Some device only has limited computing/storage/power resources;

(2) To achieve higher accuracy, we may want to explicitly consider pairwise or 3-way interactions (or high-order n-grams) in linear models.

(3) We hope to reduce the complexity of learning models (e.g., deep nets) by hashing the inputs or hashing the outputs.

(4) Hashing is an effective way of indexing (and space partitioning), which allows efficient sub-linear time near neighbor search.

(5) Perhaps surprisingly, our newest research can show that, if designed carefully, hashing (which naturally leads to linear algorithms) can also model the nonlinear effect (e.g., nonlinear kernels). Examples of such kernels include resemblance, chi-square, and CoRE kernels.

This talk will cover a variety of hashing algorithms including sign Cauchy projections, b-bit minwise hashing, one permutation, and densified one permutation hashing, etc.

Bio-Sketch: Ping Li is Associate Professor at Rutgers University, in the Department of Statistics and the Department of Computer Science. He graduated from Stanford University with Ph.D. in Statistics (and Master's degrees in both CS and EE). Ping Li's research interests include probabilistic hashing algorithms for big data, machine learning, information retrieval, boosting, data streams, and compressed sensing. His research has been funded by the Department of Defense (DoD), Microsoft, Google, and the National Science Foundation (NSF). In particular, he was one of the PIs of the recent NSF-Bigdata program. Ping Li received the Young Instigator Award (YIP) from the Air Force Office of Scientific Research (AFOSR) and the YIP from the Office of Naval Research (ONR). He also won a prize in 2010 Yahoo! Learning to Rank Grand Challenge using "robust logitboost" (Li, UAI 2010).

10/13/2014

Alexander Tartakovsky (UConn)

"State-of-the Art and Recent Accomplishments in Quickest Changepoint Detection"

I will begin with an overview of optimal and nearly optimal quickest (sequential)changepoint detection methods, addressing a variety of scenarios and problem settings: simple and composite post-change hypotheses; conventionaliidmodels and general non-iid models; Bayesian, minimax and other optimality criteria. Classical results as well as certain recent achievements in more complicated quickestchangepoint detection problems with multiple populations/streams will be discussed. Certain Applicationswill also be addressed.

10/20/2014

Adrian Raftery (Univ. of Washington)

"Probabilistic population projections for all countries."

Abstract: Projections of countries' future populations, broken down by age and sex, are widely used for planning and research. They are mostly done deterministically, but there is a widespread need for probabilistic projections. I will describe a Bayesian statistical method for probabilistic population projections for all countries. These new methods have been used by the United Nations to produce their most recent population projections for all countries.

10/27/2014

Venkat Chandrasekaharn (Caltech)

Title: Latent Variable Graphical Model Selection via Convex Optimization

Suppose we have a Gaussian graphical model with sample observations ofonly a subset of the variables. Can we separate the extra correlations induced due to marginalization over the unobserved, hidden variables from the structure among the observed variables? In other words is it still possible to consistently perform model selection despite the unobserved, latent variables? As we shall see the key problem that arises is one of decomposing the concentration matrix of the observed variables into a sparse matrix (representing graphical model structure among the observed variables) and a low rank matrix (representing the effects of marginalization over the hidden variables). Such a decomposition can be accomplished by an estimator that is given by a tractable convex program. This estimator performs consistent model selection in the high-dimensional scaling regime in which the number of observed/hidden variables grows with the number of samples of the observed variables. The geometric aspects of our approach are highlighted, with the algebraic varieties of sparse matrices and of low rank matrices playing an important role.

Bio: Venkat Chandrasekaran is an Assistant Professor at Caltech in Computing and Mathematical Sciences and in Electrical Engineering. He received a Ph.D. in Electrical Engineering and Computer Science in June 2011 from MIT, and he received a B.A. in Mathematics as well as a B.S. in Electrical and Computer Engineering in May 2005 from Rice University. He was awarded the Jin-Au Kong Dissertation Prize for the best doctoral thesis in Electrical Engineering at MIT (2012), the Young Researcher Prize in Continuous Optimization at the Fourth International Conference on Continuous Optimization of the Mathematical Optimization Society (2013, awarded once every three years), an Okawa Research Grant in Information and Telecommunications (2013), and an NSF CAREER award (2014). His research interests lie in mathematical optimization and its application to the information sciences.

11/10/2014

Caroline Uhler (Institute of Science and Technology Austria)

“Parameter estimation for linear Gaussian covariance models”

Linear Gaussian covariance models are Gaussian models with linear constraints on the covariance matrix. Such models arise in many applications, such as stochastic processes from repeated time series data, Brownian motion tree models used for phylogenetic analyses and network tomography models used for analyzing the connections in the Internet. Maximum likelihood estimation in this model class leads to a non-convex optimization problem that typically has many local maxima. However, I will explain that the log-likelihood function is in fact concave over a large region of the positive definite cone. Using recent results on the asymptotic distribution of extreme eigenvalues of the Wishart distribution I will show that running any hill-climbing method in this region leads to the MLE with high probability.

11/17/2014

Constantine Frangakis (JHU)

“Deductive (Turing-machine) derivation of semiparametric efficient estimation”

Background: Researchers often seek robust inference for a parameter through semiparametric estimation. Efficient semiparametric estimation currently requires theoretical derivation of the efficient influence function (EIF), which can be a challenging and time-consuming task. If this task can be computerized, in the sense of a Turning machine, it can save dramatic human effort, which can be transferred, for example, to the design of new studies.

Problem: Although the EIF is, in principle, a derivative, simple numerical differentiation to calculate the EIF by a computer masks the EIF’s functional dependence on the parameter of interest. For this reason, the standard approach to obtaining the EIF relies on the theoretical construction of the space of scores under all possible parametric submodels. This process currently depends on the correctness of conjectures about these spaces, and the correct verification of such conjectures. The correct guessing of such conjectures, though successful in some problems, is a nondeductive process, i.e., is not guaranteed to succeed (e.g., is not computerizable), and the verification of conjectures is generally susceptible to mistakes.

Results: We propose a method that can deductively produce semiparametric locally efficient estimators. The proposed method is computerizable, meaning that it does not need either conjecturing, or otherwise theoretically deriving the functional form of the EIF, and is guaranteed to produce the desired estimates. The method is demonstared through two examples: one where the EIF is already known, and another where the EIF had not until now been found.

11/24/2014

Han Xiao (Rutgers)

Title: WALD TESTS OF SINGULAR HYPOTHESES

Abstract:

Motivated by the problem of testing tetrad constraints in factor analysis,we study the large-sample distribution of Wald statistics at parameterpoints at which the gradient of the tested constraint vanishes. When based on an asymptotically normal estimator, the Wald statistic converges to a rational function of a normal random vector. The rational function is determined by a homogeneous polynomial and a covariance matrix. For quadratic forms and bivariate monomials of arbitrary degree, we show unexpected relationships to chi-square distributions that explain conservative behavior of certain Wald tests. For general monomials, we offer a conjecture according to which the reciprocal of a certain quadratic form in the reciprocals of dependent normal random variables is chi-square distributed.

12/1/2014

Jason Fine (UNC Chapel Hill)

"Isotonic Proportional Hazards Models"

We consider estimation of the semiparametric proportional hazards model under unspecified baseline hazard function where the effect of a continuous covariate is monotone but otherwise unspecified. Previous work on full nonparametric likelihood estimation for such isotonic Cox proportional hazards regression with right censored data is computationally intensive, lacking theoretical justification, and may be prohibitive in large samples. We propose partial likelihood estimation. An iterative quadratic minorant (IQM) algorithm is proposed, with theoretically justified convergence properties. However, unlike with standard parametric isotonic regression models, the IQM derived from the partial likelihood cannot be implemented using standard pool adjacent violators techniques, increasing the computational burden, and evidences numerical instability. Iterative convex minorant (ICM) has been shown to be useful in such settings, but exhibits similar issues. An alternative pseudo iterative convex minorant algorithm (PICM) is presented which exploits standard PAVA techniques and is theoretically justified. Simulation studies demonstrate that PICM may yield orders of magnitude reductions in computing time versus IQM and, with improved empirical convergence and bias properties. Analysis of data from an HIV prevention study illustrates the practical utility of the isotonic methods in estimating nonlinear covariate effects.

12/8/2014

Fang Yao (U Toronto)

Title: Simultaneous white noise models and optimal recovery of functional data

We consider the white noise representation of functional data taken as i.i.d. realizations of a Gaussian process. We begin with establishing an asymptotic equivalence in Le Cam’s sense between an experiment which simultaneously describes these realizations and a collection of white noise models. In this context, we project onto an arbitrary basis and apply a novel variant of Stein-type estimation for optimal recovery of the realized trajectories. A key inequality is derived showing that the corresponding risks, conditioned on the underlying curves, are minimax optimal and can be made arbitrarily close to those that an oracle with knowledge of the process would attain. Empirical performance is illustrated through simulated and real data examples.