

## **Department of Statistics - Statistics Seminar – Spring 2012**

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

---

### **1/23/2012**

Speaker: Daniela Yang, Department of Statistics, The Wharton School  
University of Pennsylvania

Title: Singular Value Decomposition for High-Dimensional Data

#### Abstract

Singular value decomposition is a widely used tool for dimension reduction in multivariate analysis. However, when used for statistical estimation in high-dimensional low rank matrix models, singular vectors of the noise-corrupted matrix are inconsistent for their counterparts of the true mean matrix. In this talk, we suppose the true singular vectors have sparse representations in a certain basis. We propose an iterative thresholding algorithm that can estimate the subspaces spanned by leading left and right singular vectors and also the true mean matrix optimally under Gaussian assumption. We further turn the algorithm into a practical methodology that is fast, data-driven and robust to heavy-tailed noises. Simulations and a real data example further show its competitive performance.

This is a joint work with Andreas Buja and Zongming Ma.

---

### **1/30/2012**

Speaker: Yingqi Zhao, University of North Carolina at Chapel Hill

Title: Estimating Individualized Treatment Rules Using Outcome Weighted Learning.

#### Abstract:

There is increasing interest in discovering individualized treatment rules for patients who have heterogeneous responses to treatment. In particular, one aims to find an optimal individualized treatment rule, which is a deterministic function of patient specific characteristics maximizing expected clinical outcome. In this paper, we first show that estimating such an optimal treatment rule is equivalent to a classification problem where each subject is weighted proportional to his or her clinical outcome. We then propose an outcome weighted learning approach based on the support vector machine framework. We show that the resulting estimator of the treatment rule is consistent. We further obtain a finite sample bound for the difference between the expected outcome using the estimated individualized treatment rule and that of the optimal treatment rule. The performance of the proposed approach is demonstrated via simulation studies and an analysis of chronic depression data.

---

**2/6/12**

Speaker: Rahul Mazumder, Department of Statistics, Stanford University

Title: Convex Regularization Algorithms for Learning Large Incomplete Matrices

Abstract:

In many applications measured data can be represented in a matrix  $X_{m,n}$ , for which only a relatively small number of entries are observed. The task is to "complete" the matrix based on the observed entries and has been dubbed the matrix completion problem. This is a fundamental problem arising for example in recommender systems (for eg the "Netix" prize) and collaborative filtering; internet and electronic-commerce industries; image processing and micro-array imputation, among others. In practice, the matrix dimensions frequently range from hundreds of thousands to even a million. Low-rank matrix modeling is often a model of choice for these problems but is computationally intractable. Using nuclear norm as the convex surrogate of the combinatorially hard rank constraint we develop a simple and efficient algorithm Soft-Impute which minimizes the reconstruction error subject to a regularization on the nuclear norm. With warm-starts this allows us to efficiently compute an entire regularization path of solutions on a grid of values of the regularization parameter. The computationally intensive part of our algorithm is in computing a low-rank SVD of a large dense matrix. Using problem structure, we show how to circumvent this computational bottleneck and demonstrate algorithmic scalability to problems of the Netix size, and even larger | all of which can be achieved within a very reasonable time. We address convergence properties and computational complexities of Soft-Impute and establish connections with other popular variants of matrix factorization. We show how our proposed framework can be adapted to more general regularized low-rank modeling frameworks, discussing specific applications in image processing, computer vision and collaborative filtering.

---

**2/13/2012**

Speaker Matt Hoffman, Princeton University

Title: Making Bayesian Inference Faster and Easier

Abstract: Analyzing data using hierarchical Bayesian models almost always requires approximate posterior inference techniques such as Markov chain Monte Carlo (MCMC) or variational Bayes (VB). These methods can be challenging to apply to complex models or large datasets. For example, popular MCMC methods such as Gibbs sampling can be very slow when applied to complex models with many parameters. VB is often much faster, but introduces bias and is less generally applicable. And any "batch" inference method will be unacceptably slow when applied to a sufficiently large dataset. In this talk, I will present two algorithms for efficient Bayesian inference: online VB and the no-U-turn sampler (NUTS). Online VB incrementally fits an approximation to the posterior, considering only a subset of the full dataset at each iteration. When applied to the latent Dirichlet allocation model, online VB is able to discover a set of topics from millions of Wikipedia documents in a fraction of the time needed by batch algorithms. NUTS is an MCMC algorithm that extends the Hamiltonian Monte Carlo (HMC) algorithm. HMC can be orders of magnitude faster than Gibbs sampling, but it requires careful

problem-specific tuning. NUTS both eliminates the need to hand-tune HMC and improves upon HMC's efficiency, making it easier to fit complex models quickly.

---

**2/20/2012**

Speaker: Dean Foster (U Penn)

Title: Linear methods for large data

Using random matrix theory, we now have some very easy to understand and fast to use methods of computing low rank representations of matrices. I have been using these methods as a hammer to improve several statistical methods. I'll discuss three of these in this talk. First, I'll show how these ideas can be used to speed up stepwise regression. Then I'll turn to using them to construct new linear features motivated by CCA's. Finally, I'll use these methods to get a fast way of estimating an HMM.

---

**2/27/2012**

Speaker: Hedibert Lopes (Chicago)

TITLE: Cholesky Stochastic Volatility Models for High-Dimensional Time Series

ABSTRACT:

Multivariate time-varying volatility has many important applications in finance, including asset allocation and risk management. Estimating multivariate volatility, however, is not straightforward because of two major difficulties. The first difficulty is the curse of dimensionality. For  $p$  time series, there are  $p(p+1)/2$  volatility and cross-correlation series. In addition, the commonly used volatility models often have many parameters, making them impractical for real application. The second difficulty is that the conditional covariance matrix must be positive definite for all time points. This is not easy to maintain when the dimension is high.

In order to simply maintain positive definiteness, we model the Cholesky root of the time varying  $p \times p$  covariance matrix. Our approach is Bayesian and we propose prior distributions that allow us to search for simplifying structure without placing hard restrictions on the parameter space. Our modeling approach is chosen to allow for parallel computation and we show how to optimally distribute the computations across processors. We illustrate our approach by a number of real and synthetic examples, including a real application with 94 time series ( $p=94$ ).

KEY WORDS: Bayesian modeling; Conditional Heteroscedasticity; Forward Filtering and Backward Sampling; Parallel Computing; Volatility Matrix.

---

**3/19/2012**

Sam Kou (Harvard)

Title: Multi-resolution inference of stochastic models from partially observed data

Stochastic models, diffusion models in particular, are widely used in science, engineering and economics. Inferring the parameter values from data is often complicated by the fact that the underlying stochastic processes are only partially observed. Examples include inference of discretely observed diffusion processes, stochastic volatility models, and double stochastic Poisson (Cox) processes. Likelihood based inference faces the difficulty that the likelihood is usually not available even numerically. Conventional approach discretizes the stochastic model to approximate the likelihood. In order to have desirable accuracy, one has to use highly dense discretization. However, dense discretization usually imposes unbearable computation burden. In this talk we will introduce the framework of Bayesian multi-resolution inference to address this difficulty. By working on different resolution (discretization) levels simultaneously and by letting the resolutions talk to each other, we substantially improve not only the computational efficiency, but also the estimation accuracy. We will illustrate the strength of the multi-resolution approach by examples.

---

**3/26/2012**

Ming Yuan (Georgia Tech)

Title : Adaptive Estimation of Large Covariance Matrices

Abstract :

Estimation of large covariance matrices has drawn considerable recent attention and the theoretical focus so far is mainly on developing a minimax theory over a fixed parameter space. In this talk, I shall discuss adaptive covariance matrix estimation where the goal is to construct a single procedure which is minimax rate optimal simultaneously over each parameter space in a large collection. The estimator is constructed by carefully dividing the sample covariance matrix into blocks and then simultaneously estimating the entries in a block by thresholding. I shall also illustrate the use of the technical tools developed in other matrix estimation problems.

---

**4/2/2012**

Prof. Arnaud Doucet (Oxford).

Title: Forward Smoothing in State-Space Models with Application to Maximum Likelihood Parameter Estimation

Abstract: Maximum likelihood parameter estimates in state-space models are generally computed using a gradient ascent procedure or an Expectation-Maximization (EM) procedure. Both approaches rely on a forward filtering-backward smoothing procedure which is computationally expensive for large data sets. Additionally the memory requirements increase linearly with the number of data as it is necessary to store the filtering distributions computed in the forward pass to carry out the backward pass. For nonlinear non-Gaussian state-space models, particle approximations of these algorithms can be easily derived but they obviously suffer from the same problems. We present a forward only version of the forward-backward procedure which bypasses entirely the backward pass, does not require storing the filtering distributions and allow us to implement online versions of gradient ascent and EM. We propose a non-standard particle implementation of the forward smoothing procedure which is provably numerically stable. This allows us to perform recursive maximum likelihood parameter estimation in

nonlinear non-Gaussian state-space models using particle algorithms which do not suffer from the particle path degeneracy problem.

This is joint work with Pierre Del Moral (INRIA Bordeaux) and Sumeet Singh (Cambridge).

---

**4/9/2012**

Vladimir I. Koltchinskii (Georgia Tech)

Complexity Penalties in Low Rank Matrix Estimation

Consider a problem of estimation of a large  $m \times m$  Hermitian matrix  $\Sigma$  based on i.i.d. measurements

$$Y_j = \text{tr}(\Sigma X_j) + \epsilon_j, j = 1, \dots, n,$$

where  $X_j$  are random  $m \times m$  Hermitian matrices and  $\{\epsilon_j\}$  is a zero mean random noise. The goal is to estimate  $\Sigma$  in the case when it has relatively small rank, or it can be well approximated by small rank matrices. There has been an extensive study of this problem in the recent years. Its important instances include matrix completion, where a random sample of entries of  $\Sigma$  is observed, and quantum state tomography, where  $\Sigma$  is a density matrix of a quantum system and it has to be estimated based on the measurements of  $n$  observables  $X_1, \dots, X_n$ . We will consider several approaches to such problems based on a penalized least squares method (and its modifications) with complexity penalties defined in terms of nuclear norm, von Neumann entropy and other functionals that “promote” small rank solutions and discuss oracle inequalities for the resulting estimators with explicit dependence of the error terms on the rank and other parameters of the problem. We will also discuss a version of these methods when the target matrix is a “smooth” low rank kernel defined on a graph.

---

**4/16/2012**

Dylan Small (Wharton School, UPenn)

Title: Case Definition and Design Sensitivity in Case Control Studies

Abstract:

A case-control study compares cases of some disease or disorder to some group of controls (non-cases), looking backwards in time to contrast the frequency of treatment among cases and controls. Cases are typically matched to controls on measured pretreatment covariates. However, in an observational study, there may be unmeasured pretreatment covariates that affect both treatment and outcomes. A sensitivity analysis asks: What magnitude of bias from unmeasured covariates would need to be present to materially alter the conclusions of a naïve analysis that presumes adjustments for measured covariates suffice to remove all bias?

The first step in designing a case-control study is to define a case of disease and a control. For example, the disease may have different severities and one needs to choose how severe a person's disease needs to be for the person to be a case. We examine the effects of this design decision on the sensitivity of conclusions to unmeasured biases. We develop an adaptive procedure for choosing the case definition based on the data to make the study as insensitive to unmeasured biases as possible asymptotically. This is joint work with Jing Cheng, Betz Halloran and Paul Rosenbaum.

---

**4/23/2012**

Chris Wiggins (Columbia University)

"Variational and hierarchical modeling for biological data"

Advances in biological technologies over the past two decades have dramatically increased the abundance of data available to biologists, and thereby changed the relationship between biology and statistics. While this is most famously celebrated in the subfield of genomics (both sequencing and functional genomics), there is increasing need in the subfield of molecular biology, particularly for methods based on generative models motivated by biologists' domain expertise. A natural set of tools is that provided by inference with latent variables. In this talk I'll introduce one application of a variational approach to inference; I then present current work on a closely-related hierarchical modeling approach, based on collaborations with the Gonzalez lab at Columbia, for understanding time-series data in single-molecule biophysics.

---

**4/30/2012**

Martin Wainwright (Berkeley)

TITLE: High-dimensional matrix decomposition: Applications and estimators

ABSTRACT: Consider a matrix that can be decomposed as the sum of two unknown matrices, one of which is approximately low-rank and the other having a complementary form of low-dimensional structure, such as bandedness, sparsity, or column-sparsity. Given noisy or partial observations, how to recover accurate estimates of the underlying decomposition?

Matrix decompositions of this type arise in many applications, among them robust forms of dimensionality reduction (PCA, canonical correlations etc.), collaborative filtering problems (e.g., Netflix and Amazon), and estimating the structure of Gaussian graphical models. Various researchers have studied conditions under which simple convex programs, based on the nuclear norm as a rank surrogate, can perform exact recovery based on noiseless observations. In practical settings, observations are likely to be noise-corrupted, and matrices only approximately low-rank. We describe a related convex relaxation for noisy observations, and sketch how recovery guarantees can be derived under milder conditions. These error bounds show that our method is information-theoretically optimal.

Based on joint work with Alekh Agarwal and Sahand Negahban.

Paper: <http://arxiv.org/abs/1102.4807>

**5/7/2012**

Douglas Simpson, Department of Statistics, University of Illinois at Urbana-Champaign

Title: Statistical Methods for Biomedical Research on Diagnostic  
Ultrasound

Abstract:

Diagnostic ultrasound is among the most widely used imaging techniques in biomedicine. Common uses include prenatal ultrasonic imaging of the fetus, echocardiogram images of the heart, ultrasound imaging of tumors in the breast and prostate. Current research aims to extend the range of applications and increase the diagnostic power of ultrasonic imaging through quantitative ultrasound technology. Statistical issues and results associated with these efforts will be presented including image segmentation, pattern recognition, tissue characterization and semiparametric functional data analysis.