



The University of Chicago

Department of Statistics

Seminar Series

SAYAN MUKHERJEE

Department of Statistical Science
Institute for Genome Sciences & Policy
Department of Computer Science
Department of Biostatistics and Bioinformatics

**Predictive Models of Complex Traits:
Inference of Statistical Dependencies and Predictive Geometry**

MONDAY, October 13, 2008 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110.

ABSTRACT

Tumor progression is a complex (disease) trait. The challenge in modeling tumorigenesis is heterogeneity with respect to phenotype, stages of the disease, and genotype or gene expression variation. This is particularly challenging in the case of high-dimensional data.

We first develop an approach for modeling tumor progression in both the space of genes as well as a priori defined pathways. We infer both pathways relevant across stages of progression as well as localized to individual stages. We also infer pathway and gene networks relevant to progression between stages. This is applied to prostate and colon cancer. In the later case we generalize the Vogelstein model to pathways.

Our modeling tools are (Bayesian) hierarchical or multi-task regression models as well as a generalization inverse regression based on inference of gradients on manifolds. Statistical properties of this approach will be developed in the context of simultaneous dimension reduction and regression as well as graphical models. Consistency of dimension reduction as well as inference of the graphical model will be stated. An interesting observation is that the rate of convergence of both estimates depends on the dimension of the underlying manifold on which the marginal distribution is assumed to be concentrated, the result does not depend on the number of non-zero entries of the conditional independence matrix but on the rank of this matrix.

The second part of the talk describes a method to decompose pathways or gene networks into sub-networks and infer the relevance of these sub-components in explaining phenotypic variation. The approach which we call multiscale graphical models is strongly related to old ideas such as path analysis. Specifically, it is based on the idea of diffusion wavelets which in our application is a multiscale decomposition of a partial correlation matrix or the generator of a Markov chain. We describe results on yeast gene expression data to illustrate the method and then provide preliminary data on prostate cancer.

Inference will be using Bayesian models as well as regularized regularization algorithms.

Joint work with: Qiang Wu, Elena Edelman, Justin Guinney, Ding-Xuan Zhou, Mauro Maggioni, Kai Mao, Anil Potti

Please send email to Mathias Drton (drton@galton.uchicago.edu) for further information. Information about building access for persons with disabilities may be obtained in advance by calling Kathryn Kraynik or by email (kraynik@galton.uchicago.edu).