

Bio³ Seminar Series

Improving differential expression analysis with the consideration of genome-wide co-expression information

Yinglei Lai, Ph.D

Assistant Professor of Statistics
Department of Statistics
George Washington University, Washington D.C.

Abstract:

Microarrays have been widely used in biomedical studies. The differential expression analysis of microarray data is still an interesting topic. The control of false positives in differential expression analysis remains a major challenge although many statistical methods have been proposed for its improvement. Since genes interact with each other during cellular and molecular processes, an efficient incorporation of genome-wide co-expression information may significantly improve the detection of differential expression. We will address our recent research progress in this direction.

Friday, September 11, 2009 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Bio³ Seminar Series

Some lessons from our collaborative studies in esophageal cancer, prostate cancer, HIV, and breast cancer

George Bonney, Ph.D

Professor/Director

Statistical Genetics and Bioinformatics Unit,
National Human Genome Center at Howard University

Abstract:

The work of the Statistical Genetics and Bioinformatics Unit of the National Human Genome Center at Howard University involves the use of high level mathematical and statistical computing skills in biomedicine. Here I briefly discuss questions and results from some of our collaborative studies:

Esophageal cancer in Chinese families: Is alcohol really protective?

Multiple cancers in Texas families. Is the association with the p53 mutation causal?

Prostate Cancer in African American Men: Where are the genes?

HIV Prevalence and Incidence among Blacks in Washington DC:

Does it make sense to talk of estimation for the whole city using only the data from Howard University Hospital?

A Molecular Index for Breast Cancer Risk Assessment?

Can we really construct such an index for risk of invasive breast cancer?

Friday, September 25, 2009 10:00-11:00 am

New Research Building, E501

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Modeling and Inference of the Hazard Ratio Function

Song Yang, Ph.D

Office of Biostatistics Research

National Heart, Lung, and Blood Institute

National Institutes of Health

Abstract:

The hazard ratio provides a valuable tool for assessing a treatment effect with survival data, with the proportional hazards special case of the Cox model as a widely used example. In general, the hazard ratio is a function of time, and provides a visual display of the temporal pattern of the treatment effect. The proportional hazards assumption is often too restrictive, at least for the initial exploration of a treatment effect, while a nonparametric estimate of the hazard ratio function requires a bandwidth selection, and may result in increase in variance or bias. On the other hand, most semiparametric hazards models proposed so far imply certain restrictions on the hazard ratio that limit their utility. We investigate a model that allows monotone increasing or decreasing hazard ratio functions, including crossing hazards. This model provides a sufficient level of flexibility for many applications. The point estimates, point-wise confidence intervals, and simultaneous confidence intervals, or confidence bands, of the hazard ratio, are proposed under this model. We demonstrate the inference procedures in several examples, including the coronary heart disease data from the Women's Health Initiative estrogen plus progestin clinical trial. These examples, with a diverse range of time dependence of the hazard ratio from mild to severe, suggest that the hazard ratio under this class of models, its confidence intervals and confidence bands, provide very useful visual display tools for assessing the treatment effect with survival data.

Friday, October 9, 2009 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Impact of q-RT-PCR analytical methods on multi-center biomarker trials in colorectal cancer

Terry Hyslop, PhD

Associate Professor, Thomas Jefferson University
Director, Division of Biostatistics;
Director, Biostatistics Share Facility, Kimmel Cancer Center

Abstract:

Guanylyl cyclase 2C (GUYC2C), an emergent intestinal tumor suppressor, is the receptor for the paracrine hormones guanylin and uroguanylin, gene products frequently lost early in colon carcinogenesis [1,2]. Lymph nodes and tumor specimens were dissected from patients with AJCC stage I and II colon and rectal resections performed in the surgical departments of 7 academic medical centers and 2 community hospitals in the U.S. and Canada between January 2003 and June 2007. Follow-up, based on periodic evaluations, was confirmed for all patients through December 2007. GCC and beta-actin expression were measured based on standard curves formed from serial dilutions. Gene expression was also estimated by logistic regression analysis of amplification profiles from individual q-RT-PCR reactions, providing an efficiency-adjusted relative quantification based on parameter estimates from the fitted models [3]. We show that the measurement techniques developed impact the analysis and interpretation of this large multi-center prospective trial by reducing the measurement error in q-RT-PCR. In multivariable Cox models of n=257 early stage colorectal cancer patients [4], controlling for T stage, tumor location, lympho-vascular invasion, and tumor differentiation, GCC q-RT-PCR remains an independent predictor of recurrence (adjusted Hazard Ratio (AHR)=4.66, p=0.04, 95%CI=1.11, 19.57). Moreover, GCC q-RT-PCR is an independent predictor of disease free survival (AHR=3.27, p=0.03, 95% CI=1.15, 9.29). We also present findings in this population based on recursive partitioning analysis, where homogeneous risk sets are identified. Finally, preliminary analysis indicates that GUYC2C may also be used to identify subpopulations of early stage patients who may benefit from chemotherapy.

Friday, October 23, 2009 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Computer-Intensive Statistical Methodology with Applications to Translational Cancer Research

Kim-Anh Do, Ph.D

Professor

Department of Biostatistics

Division of Quantitative Sciences

The University of Texas MD Anderson Cancer Center



Abstract:

Early detection is critical in disease control and prevention. Biomarkers provide valuable information about the status of a cell at any given time point. Biomarker research has benefited from recent advances in technologies such as gene expression microarrays, and more recently, proteomics. The long term translational research goal is that if drugs can be targeted to specific tissues in the body, then dosage can be altered to achieve the desired effect while minimizing side effects such as toxicity. Motivated by specific problems involving such high throughput data, I have developed computer-intensive statistical methods based on nonparametric and semiparametric mixture model assumptions for real-time analysis in the context of biomarker discovery. Most biomarker-discovery projects aim at identifying features in the biomarker profiles (gene expression, phage, SAGE, mass spectrometry proteins) that distinguish cancers from normals, between different stages of disease development, or between experimental conditions (such as different treatment arms or different tissue types). Novel statistical methodology development will be highlighted with direct applications to cancer research challenges that address our long term translational goal.

Friday, November 13, 2009 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Bio³ Seminar Series

A unified approach to non-negative matrix factorization with application to large-scale biological data analysis and text mining

Karthik Devarajan, Ph.D

Assistant Professor
Biostatistics & Bioinformatics
Fox Chase Cancer Center

Abstract:

Non-negative matrix factorization (NMF) by the multiplicative updates algorithm is a powerful machine learning method for decomposing a high-dimensional nonnegative matrix V into two matrices, W and H , each with nonnegative entries, $V \sim WH$. NMF has been shown to have a unique parts-based, sparse representation of the data. The non-negativity constraints in NMF allow only additive combinations of the data which enables it to learn parts that have distinct physical representations in reality. Over the past decade, NMF has found successful applications in such diverse areas as natural language processing, information retrieval, image processing, speech recognition and computational biology for the analysis and interpretation of large-scale data. In this talk, we present a generalized approach to NMF based on Renyi's divergence between two non-negative matrices related to the Poisson likelihood. Our approach unifies various competing models and provides a unique framework for NMF. We demonstrate a link between NMF and some well-known statistical models. In addition, we describe an unsupervised clustering algorithm that utilizes this unified approach and discuss a parallel implementation of the algorithm using high-performance computing clusters. The applicability of our methods to molecular pattern discovery and text mining are illustrated using real-life and simulated data.

Friday, December 11, 2009 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Statistical issues in gene regulatory network reconstruction

Anthony Almudevar, PhD

Associate Professor, University of Rochester Medical Center
Department of Biostatistics and Computational Biology



Abstract:

The reconstruction of gene regulatory pathways and networks based on gene expression data is an important field in modern genetics. A number of graphical modeling techniques are widely used, including Bayesian networks and Boolean networks. However, statistical issues which are rigorously treated in classical biostatistical modeling tend not to be as widely considered in this newer field, but are no less important.

I will describe some recent research involving the use of experimental gene perturbation data to develop a Boolean network model of gene regulation in cancer tissue. The talk will focus on a number of new methodologies which apply various statistical principles to this problem. First, a method of testing for general network structure will be described, and I will show how it can be incorporated into a multiple hypothesis testing protocol, resulting in greater power than would be available with standard multiple testing adjustments. I will then show how implicit modeling techniques can be used to distinguish between the effects of persistent and transient perturbations, thereby more faithfully modeling the experimental data. Finally, computational Bayesian methodology will be used to assign confidence levels to specific network model features.

Friday, January 22, 2010 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Bio³ Seminar Series

Genetic Experimentation and Directed Bayes Modeling

Theodore T. Allen, PhD

Associate Professor, Ohio State University
Department of Integrated Systems Engineering
(with Hui Xiong, Françoise Seillier-Moiseiwitsch,
Nilgun Ferhatosmanoglu, Cenny Taslim, and David Woods)

Abstract:

This presentation begins by describing two relatively exotic types of genetics related experiments: two color array and genetic perturbation experiments. In both cases, it is possible and advisable to vary many controllable inputs simultaneously, generate samples, and test them. Our preliminary results for these areas are briefly reviewed along with a discussion of the benefits of experimental design in general.

Then, the presentation focuses on the more usual case in which there are no controllable physical factors or one factor, e.g., normal tissue or tumor tissue. The central issue explored then is the possibility for expert knowledge to be openly and explicitly incorporated into the analysis. There can even be multifactor experimentation on the experts to supplement the gene expression data.

The examples for this portion focus on clustering of genes hierarchically into groups. Yet, the concept of a “directed Bayesian” approach, in which all information from experts does not enter solely through the prior, is generally applicable. The motivations for directed Bayes methods include the fact that most statistical methods are “stupid” and do not know about physical laws or other considerations that may be obvious to biologists or others. Also, the team members using statistical models may not know, before seeing preliminary analysis, which physical laws are relevant for inclusion so that the results reflect reality as accurately as possible. In general, a simpler model form, mitigated using directed Bayes, might be more accurate than a more complicated form derived from a series of manual refinements and refitting.

Friday, February 26, 2010 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Empirical Likelihood and Its application in Survival Analysis

Mai Zhou, PhD

Professor, University of Kentucky
Department of Statistics

Abstract:

Empirical Likelihood is a nonparametric inference method with certain advantages. We first review the basics of the Empirical Likelihood ratio tests (Owen 2001). We then illustrate the advantages of Empirical Likelihood by examples related to the Kaplan-Meier estimator. Finally we discuss two commonly used Empirical Likelihood approaches with censored data estimating equations in the survival analysis and argue that one is better.



Friday, March 26, 2010 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Bio³ Seminar Series

AN ION MIXTURE MODEL FOR TIME-OF-FLIGHT DATA FROM PROTEIN MASS SPECTROMETRY

Mei-Ling Ting Lee, PhD

**Professor, Department of Epidemiology and Biostatistics
Director, Biostatistics and Risk Assessment Center
University of Maryland, College Park**

Abstract:

Protein mass spectrometry (MS) is playing a key role in protein and peptide identification and, consequently, has important research and clinical applications. Techniques for statistical modeling and analysis of MS time-of-flight (TOF) data are urgent needs. We present an ion mixture model and a first-hitting-time model for analyzing TOF mass spectra. We illustrate our inference methods using experimental data sets.

Friday, April 9, 2010 10:00-11:00 am
Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Fast Model-Based Analysis for Flow Cytometry

Cliburn Chan, Ph.D.

Assistant Professor, Department of Biostatistics and Bioinformatics
Duke University School of Medicine

Abstract:

Immune dysfunction and chronic inflammation play critical roles in essentially all major medical conditions, from infectious disease to cancer. Hence, the ability to monitor the immune response is essential for insights into disease mechanisms and effective immune-based therapies including vaccines, transplantation and other forms of immune modulation (e.g. monoclonal antibody therapy). As immune responses are typically controlled by extremely rare cell subsets (e.g. regulatory and antigen-specific polyfunctional lymphocytes), the information provided by aggregate assays such as microarrays and proteomics is limited. Flow cytometry is the dominant assay for immune monitoring and profiling, because it provides phenotypic and functional information at the single cell level. In this talk, I will introduce the basics of flow cytometry and recent technological developments in the field, and discuss the use of model-based analysis using Dirichlet process mixtures of Gaussians to automate flow cytometric analysis in a variety of different biomedical contexts. I will also describe recent innovations in GPU-based computing that provide orders of magnitude speed-ups for both EM and MCMC approaches to fitting statistical mixture models.

Friday, April 23, 2010 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.