Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Bayesian Methods for Proteomic Biomarker Discovery Using Functional Mixed Models

Jeffrey S. Morris, Ph.D.

Associate Professor

The University of Texas

MD Anderson Cancer Center

Abstract: Various proteomic assays yield spiky functional data, for example MALDI-TOF and SELDI-TOF yield one-dimensional spectra with many peaks, and 2D gel electrophoresis and LC-MS yield two-dimensional images with spots that correspond to peptides present in the sample. In this talk, I will discuss how to identify candidate biomarkers for various types of proteomic data using methods based on the Bayesian wavelet-based functional mixed models. This approach models the functions in their entirety, so avoid reliance on peak or spot detection methods. The flexibility of this framework in modeling nonparametric fixed and random effect functions enables it to model the effects of multiple factors simultaneously, allowing one to perform inference on multiple factors of interest using the same model fit, while adjusting for clinical for experimental covariates that may affect both the intensities and locations of the peaks and spots in the data. I will demonstrate how to identify regions of the functions that are differentially expressed across experimental conditions, in a way that takes both statistical and clinical significance into account and controls the Bayesian false discovery rate to a pre-specified level. Time allowing, I will also demonstrate how to use this framework as the basis for classifying future samples based on their proteomic profiles in a way that can also combine information across multiple sources of data, including proteomic, genomic, and clinical, and may also discuss improvements of the modeling framework that result in more robust inference. These methods will be applied to a series of proteomic data sets from cancer-related studies.

Friday, September 7, 2007 10:00-11:00 am

Martin Marietta Conference Room, Lombardi Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

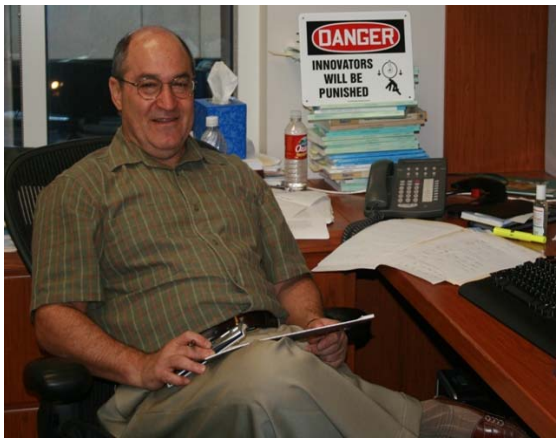
A Geometric Approach to Comparing Treatments for Rapidly Fatal Diseases

Peter Thall, Ph.D.

Professor

The University of Texas

MD Anderson Cancer Center



Abstract:

In therapy of rapidly fatal diseases, early treatment efficacy often is characterized by an event, "response," which is observed relatively quickly. Since the risk of death decreases at the time of response, it is desirable not only to achieve a response, but to do so as rapidly as possible. We propose a Bayesian method for comparing treatments in this setting based on a competing risks model for response and death without response. Treatment effect is characterized by a two-dimensional parameter consisting of the probability of response within a specified time and the mean time to response. Several target parameter pairs are elicited from the physician so that, for a reference covariate vector, all elicited pairs embody the same improvement in treatment efficacy compared to a fixed standard. A curve is fit to the elicited pairs and used to determine a two-dimensional parameter set in which a new treatment is considered superior to the standard. Posterior probabilities of this set are used to construct rules for the treatment comparison and safety monitoring. The method is illustrated by a randomized trial comparing two cord blood transplantation methods.

Friday, September 21, 2007 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

The Statistical Challenge of Studies with Errors-in-Covariates When Only the Means are Modelled

John Hanfelt, Ph.D.

Associate Professor
Emory University
Department of Biostatistics
Rollins School of Public Health

Abstract: Given the recent advances in convenient, flexible and powerful computer-intensive methods to analyze data, it is natural to wonder about the relevance of the `classical' theory of statistical inference. Here we discuss an application, namely studies with a covariate measured with error, that poses a severe statistical challenge when only the means of the observations are modelled. In this setting, standard methods of data analysis typically yield dramatically biased results -- even if computer-intensive methods are used. We draw upon the theory of bias reduction of profile estimating functions to arrive at inferences that are substantially less biased. We apply the proposed method to a study examining whether a biomarker measured with error (long-term alanine aminotransferase level) is related to length of hospital stay in patients treated for herpes zoster infections.

Friday, October 5, 2007 10:00-11:00 am

Martin Marietta Conference Room, Lombardi Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Probability of Detecting Disease-Associated SNPs in Case-Control Genome-Wide Association Studies

Mitchell H. Gail, Ph.D.

Chief of the Biostatistics Branch
Senior Investigator
Division of Cancer Epidemiology and Genetics
National Cancer Institute

Abstract: Some case-control genome-wide association studies (CCGWASs) select promising single nucleotide polymorphisms (SNPs) by ranking corresponding p-values, rather than by applying the same p-value threshold to each SNP. For such a study, we define the detection probability (DP) for a specific disease-associated SNP as the probability that the SNP will be "T-selected", namely have one of the top T largest chi-square values (or smallest p-values) for trend tests of association. The corresponding proportion positive (PP) is the fraction of selected SNPs that are true disease-associated SNPs. DP increases with genetic effect size and case-control sample size, and decreases with the number of non-disease-associated SNPs, mainly through the ratio of T to N, the total number of SNPs. We show that DP increases very slowly with T, and the increment in DP per unit increase in T declines rapidly with T. DP is also diminished if the number of true disease SNPs exceeds T. For a genetic odds ratio per minor disease allele of 1.2 or less, even a CCGWAS with 1000 cases and 1000 controls requires T to be impractically large to achieve an acceptable DP, leading to PP values so low as to make the study futile and misleading. We further calculate the sample size of the initial CCGWAS that is required to minimize the total cost of a research program that also includes follow-up studies to examine the T selected SNPs. A large initial CCGWAS is desirable if genetic effects are small or if the cost of a follow-up study is large.

Friday, October 19, 2007 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Multilevel Functional Principal Component Analysis

Ciprian Crainiceanu, Ph.D.

(Chip-ree-ann Cray-nee-cha-noo)

Assistant Professor

Johns Hopkins University

Department of Biostatistics

Abstract:

Modern research data have become increasingly complex, raising non-traditional modeling and inferential challenges. In particular, advancements in technology and computation have made recording and processing of functional data possible. Examples of functional data are time series of electroencephalographic (EEG) activity, anatomical shape, and functional MRI. The purpose of this talk is to describe statistical models for feature extraction from single-level (one or multiple functions per subject at one visit) and clustered or longitudinal (one or multiple functions per subject at multiple visits) functional data having a large number of subjects and large within- and between-subject heterogeneity. We introduce the framework and inferential tools for multilevel functional data (MFD) obtained by recording of functional characteristics at multiple visits. Though motivated by a novel experimental setting, the proposed methodology is general, with potential broad applicability to many high-throughput scientific studies. A prototypical example of MFD is the Sleep Heart Health Study (SHHS), which contains electroencephalographic (EEG) signals for each subject at two visits.

Friday, November 2, 2007 10:00-11:00 am

Martin Marietta Conference Room, Lombardi Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Ranges of Association Measures for Dependent Binary Variables

N. Rao Chaganty, Ph.D.
Department of Mathematics and Statistics
Old Dominion University

Abstract:

Analysis of longitudinal and clustered binary data is important in biomedical research. Numerous measures of association have been proposed in the literature for the study of dependence between the binary variables. These measures include correlations, odd ratios, kappa statistics and relative risks. In this talk I will discuss permissible ranges of these measures of association. Knowledge of these ranges is crucial for developing efficient estimation methods for real life data. I will show moment based methods such as generalized estimating equations, which ignore these ranges, could result in misleading p-values and incorrect conclusions.

Friday, November 16, 2007 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

DNA copy numbers and the Circular Binary Segmentation Algorithm

Venkatraman E. Seshan, PhD

Professor of Biostatistics, MSPH

Director Biostatistics Core HICCC

Columbia University

Abstract:

DNA sequence copy number is the number of copies of DNA at a region of a genome. The development of malignant tumors and their progression often involve alterations in DNA copy number. We will present the motivation for the Circular Binary Segmentation algorithm we developed (Olshen et al Biostatistics, 2004) to segment the genome into regions of equal copy number. We will also present refinements to the algorithm to handle the large arrays that are being used more commonly now (Venkatraman & Olshen Bioinformatics, 2007). We will present extensions to the problem such as parental copy numbers and the application to tumor data.

Friday, December 7, 2007 10:00-11:00 am

Martin Marietta Conference Room, Lombardi Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Mixed modeling with ambiguous cluster identifiers: An approach to haplotype-trait association studies

Andrea Foulkes, Sc.D.

Assistant Professor of Biostatistics
School of Public Health and Health Sciences
University of Massachusetts

Abstract:

Mixed effects modeling is a well-characterized method for the analysis of correlated data where correlation among observations can arise from repeated measures or clustering. In addition, it allows for characterizing gene-gene interactions while providing a flexible statistical framework to account for the confounding or mediating role of person specific covariates. Model fitting techniques generally assume the unit identifier (e.g. individual or cluster) is known; however, in the analysis of associations among genetic clusters and quantitative traits in unrelated individuals, this information is potentially unobservable. We describe a novel semi-parametric and fully likelihood-based approach to estimation when this identifier is not completely observable. The method is applied to data arising from a cohort of human immunodeficiency virus type-1 infected individuals at risk for therapy associated dyslipidemia.

Friday, January 18, 2008 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 1st and 3rd Friday of every month.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Measuring Signaling Activity in Cancer Cells using Bayesian Analysis of Microarray Data

Michael Ochs, Ph.D.

Associate Professor of Oncology

The Sidney Kimmel Comprehensive Cancer Center

Johns Hopkins University

Abstract:

Cellular signaling plays a critical role in carcinogenesis and is an extremely complex process that is still being elucidated. Nevertheless, understanding changes in signaling activity, especially as more therapeutics specifically target signaling proteins in cancer and other diseases, is critical to the development of personalized medicine and the development of novel therapies. In general, it is difficult to directly measure signaling protein states (e.g., phosphorylation) in vivo, however it has become routine to obtain global mRNA profiles with microarrays. We have developed techniques for isolating overlapping mRNA signatures using Bayesian Markov chain Monte Carlo, and we have extended these methods to direct estimation of transcription factor activity and the linking of this activity to changes in cell signaling in cancer cells during treatment. We demonstrate this approach using imatinib mesylate (Gleevec) treatment of a gastrointestinal tumor cell line.

Friday, February 1, 2008 10:00-11:00 am

Conference Room W302, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Data analysis for molecular fingerprinting of breast cancer

Dr. Reinhard Laubenbacher

Professor, Virginia Bioinformatics Institute

Professor, Department of Mathematics, Virginia Tech

Abstract:

Metabolomics is the study of cells by measuring profiles of all, or a large number, of their metabolites. This talk focuses on data analysis methods that are part of a metabolomics approach to study the progression of malignancy of breast epithelial cells. To identify robust molecular signatures that uniquely characterize early stages of malignant transformation, we employ a combination of detailed metabolic fingerprinting and data analysis using mathematical, statistical, and machine learning algorithms.

Friday, February 15, 2008 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics

This seminar series meets the 1st and 3rd Friday of every month.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Network Legos: Building Blocks of Cellular Wiring Diagrams

T.M. Murali, Ph.D.
Assistant Professor
Department of Computer Science
Virginia Tech University

Abstract:
Publicly-available data sets provide detailed and large-scale information on multiple types of molecular interaction networks in a number of model organisms. These multi-modal universal networks capture a static view of cellular state. An important challenge in systems biology is obtaining a dynamic perspective on these networks by integrating them with gene expression measurements taken under multiple conditions.

We present a top-down computational approach to identify building blocks of molecular interaction networks by
(i) integrating gene expression measurements for a particular disease state (e.g., leukaemia) or experimental condition (e.g., treatment with growth serum) with molecular interactions to reveal an active network, which is the network of interactions active in the cell in that disease state or condition and

(ii) systematically combining active networks computed for different experimental conditions using set-theoretic formulae to reveal network legos, which are modules of coherently interacting genes and gene products in the wiring diagram.

We analyse two human datasets using our method. A comparison of three leukaemias demonstrates how a biologist can use our system to identify specific differences between these diseases. A larger-scale analysis of 13 distinct stresses illustrates our ability to compute the building blocks of the interaction networks activated in response to these stresses and to use these building blocks to identify differences in the response of fibroblasts and HeLa cells to endoplasmic reticulum stress.

Friday, March 7, 2008 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 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Multi-block Matrix Factorization and Ordered Statistical Testing

S. Stanley Young, PhD

Assistant Director for Bioinformatics
National Institute of Statistical Sciences
in collaboration with Paul Fogel and George Luta

Abstract:

The -omic sciences, transcriptomics, proteomics and metabolomics, produce large data sets where the number of variables is massively larger than the number of observations. On the other hand, treatments often induce rather dramatic changes in a relatively limited number of underlying biological systems. There is a need to identify variables changed by treatments, increased or decreased. Our idea is to use matrix factorization with two blocks of data, one block is the starting matrix of data and the other block is the matrix of the reciprocals where each x_{ij} is replaced with $1/x_{ij}$. The elements of each of the factoring vectors are sorted from largest in absolute value to smallest and tested sequentially. We divide the total testing alpha among the vectors, but do not multiplicity adjust the sequential testing within the vectors. There is more statistical power as the testing within a vector is with no adjustment for multiple testing. We observe, as others have, that non-negative matrix factorization often groups variables into sets corresponding to separate mechanisms so the groups of significant genes are more interpretable. The method generalizes to multiple blocks of data. For background on non-negative matrix factorization see www.niss.org/irMF.

Friday, April 4, 2008 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 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Direct Regression Models for Survival Parameters Based on Pseudo-Values

John P. Klein, PhD
Professor and Head, Division of Biostatistics
Medical College of Wisconsin

Abstract:
Recently we have investigated the use of pseudo-values from a jackknife statistic constructed from a simple summary statistic as a way of direct regression modeling of survival probabilities. These pseudo-values, based on the difference between the complete sample and leave-one-out estimator, are used in a generalized estimating equation to obtain estimates of model parameters. The approach can be applied to direct regression modeling of the survival function over time, the cumulative incidence function for competing risk data, the restricted mean survival time, the mean quality of life, and the probabilities in a multi-state model. We illustrate many of these techniques using bone marrow transplant data from the Center for International Blood and Marrow Transplantation.

Friday, April 18, 2008 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Statistical issues in disease surveillance: A case study from ESSENCE

Cara Olsen, PhD

Biostatistics Consulting Center (CIV, USUHS)

Abstract:

Syndromic surveillance systems attempt to monitor the burden of disease in communities in real time, using health-related data and tools from statistics, epidemiology, informatics, and other disciplines. A potential benefit of such surveillance is early detection and tracking of infectious disease outbreaks.

The Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) is a syndromic surveillance system that monitors outpatient visits to military medical treatment facilities. This study examines whether ESSENCE can detect more infectious disease outbreaks, and detect them earlier, using joint monitoring of laboratory test orders and outpatient visit data rather than outpatient visit data alone. Statistical issues that arise from this question include which aberration detection algorithm is best suited to these data sources, how to quantify the tradeoffs among sensitivity, specificity and timeliness for detecting outbreaks, and how to monitor information from multiple data sources simultaneously.

Friday, May 2, 2008 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 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Bayesian Dose-finding Trial Designs for Drug Combinations

Guosheng Yin, Ph.D.

Assistant Professor
Department of Biostatistics
M. D. Anderson Cancer Center

Abstract:

Treating patients with a combination of agents is becoming commonplace in cancer clinical trials, with biochemical synergism often the primary focus. In a typical drug combination trial, the toxicity profile of each individual drug has already been thoroughly studied in the single-agent trials, which naturally offers rich prior information. We propose Bayesian adaptive designs to search for the maximum tolerated dose combination. We continuously update the posterior estimates for the toxicity probabilities of the combined doses. By reordering the dose toxicities in the two-dimensional probability space, we adaptively assign each new cohort of patients to the most appropriate dose. Dose escalation, de-escalation or staying the same is determined by comparing the posterior estimates of the toxicity probabilities of combined doses and the pre-specified toxicity target. We conduct extensive simulation studies to examine the operating characteristics of the design and illustrate the proposed method under various practical scenarios.

Friday, May 16, 2008 10:00-11:00 am

Conference Room E501, New Research Building

Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 1st and 3rd Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center