

Bio³ 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

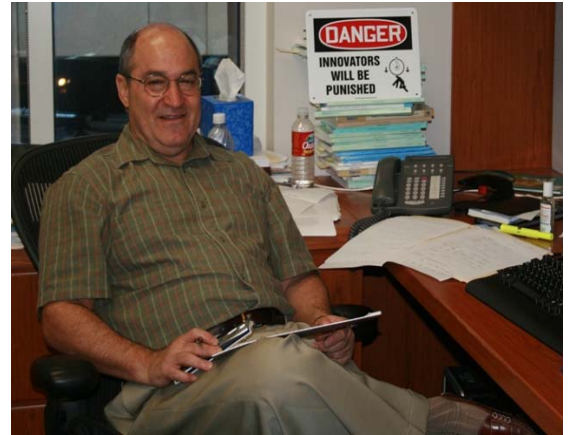
Bio³ Seminar Series

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.

Bio³ 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

Bio³ 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

Bio³ 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

Bio³ 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

Bio³ 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