

Bio³ 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.

Bio³ 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.

Bio³ 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.

Bio³ 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

Bio³ 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.

Bio³ 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.

Bio³ 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

Bio³ 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.