Biostatistics

Bioinformatics

Bio^3 Seminar Series

Contrasted Penalized Integrative Analysis

Shuangge Ma, PhD
Associate Professor of Public Health (Biostatistics)
Yale University

Abstract:
Single-dataset analysis of high-throughput omics data often leads to unsatisfactory results. The integrative analysis of heterogeneous raw data from multiple independent studies provides an effective way to increase sample size and improve marker selection results. In integrative analysis, the regression coefficient matrix has certain structures. In our study, we use group penalization for one- or two-dimensional marker selection and introduce contrast penalties to accommodate the subtle coefficient structures. Simulations show that the proposed methods have significantly improved marker selection properties. In the analysis of cancer genomic data, important markers missed by the existing methods are identified.

Friday, September 13, 2013 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.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Enriched Ensemble Methods for Classification of High-Dimensional Data

(Joint work with Javier Cabrera and others)

Dhammika Amaratunga, Ph.D.

Senior Director and Janssen Fellow in Nonclinical Biostatistics at Janssen Research & Development

Abstract:

A spate of technological advances has led to an explosion of high-dimensional data. One of the challenges of modern statistics is how to deal with this type of data. We will consider data from biomedical research that are characterized by the fact that they are comprised of a large number of variables measured on relatively few subjects, such as microarray or deep sequencing data. Classification and regression techniques are often used for analyzing this data, both for prediction as well as for identifying combinations of a few key variables associated with response. However, standard methods do not work well in this setting, due to the small sample size and surfeit of variables, a problem sometimes also exacerbated by the presence of non-specific signals. Enriched methods are a way of circumventing these difficulties. We will describe enriched methods, particularly enriched ensemble methods, that work well with this type of data. Real examples will be used to illustrate the methodology.

Friday, September 27, 2013 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.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Determining Change-points in Tumor Blood Flow using a Modified Information Criteria to Better Balance Complexity and Fit in a Semi-parametric Model.

Mary E. Putt, Ph.D., Sc.D.

Associate Professor of Biostatistics in Biostatistics and Epidemiology at the Hospital of the University of Pennsylvania, Perelman School of Medicine

Abstract:

Our work is motivated by a tumor biology study where blood flow naturally follows a non-linear pattern. An experimental cancer treatment disrupts blood flow; the duration and rate of decline appears to reflect treatment efficacy. With knowledge of the non-linear baseline blood flow, we used a smoothing spline model with unknown change-points to estimate the time of the change in flow, and blood flow at the change-points. We found that the choice of the smoothing parameter strongly influences the estimation of the change-point locations, and the function at the change-points. Choosing the smoothing parameter based on minimizing generalized cross validation, GCV, gave unsatisfactory estimates of the change-points. We propose a new method, aGCV, that re-weights the residual sum of squares and generalized degrees of freedom terms from GCV. The weight is chosen to maximize the decrease in the generalized degrees of freedom as a function of the weight, while simultaneously minimizing aGCV as a function of the smoothing parameter and the change-points. Compared to GCV, simulation studies suggest that the aGCV method yields substantially improved estimates of the change-points, as well as estimation of the function at the change-points. Remaining challenges involved in the development of valid and precise confidence intervals of the function at the change-points, as well as computational challenges will be discussed.

Friday, October 11, 2013 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

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Research at Census, including links to Biostatistics/Informatics

Thomas A. Louis, Ph.D.

Research & Methodology Directorate, U.S. Census Bureau
Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Abstract:

In order to meet the challenges of efficiently obtaining valid information and making it available to the public, research at the U.S. Census Bureau and survey research more generally burgeons. Many research goals and methods are similar to those addressed by and used in Biostatistics or Informatics. To set the scene, I briefly describe the Census Research & Methodology directorate, list major issues and approaches, then provide details on a small subset. Candidate topics include adaptive design (dynamic survey modes, R-factors in the National Survey of College Graduates, timing of mailing hard copy based on K-M curves, challenges of learning from experience), stopping rules, randomized experiments (the effect of interviewer training in the National Crime Victimization Survey), record matching, prediction (of response propensity, of occupancy, of the "fitness for use" of administrative records), imputation, Bayesian methods (design-consistent analyses, post-processed {confidence} intervals, benchmarking), small areea/spatio-temporal analysis (estimation of poverty rates, estimating omissions in the master address file), development and use of paradata (in the National Health Interview Survey), double-robustness, dynamic data posting ("OnTheMap" Local Origin-Destination Employment Statistics), disclosure avoidance/limitation, Big Data (opportunities and challenges), micro-simulation (benefits of research in designing the 2020 Census), and IT infrastructure (the Multi-mode Operational Control System). I close with a call for increased collaboration among statistical agencies and academe, building on the NSF-Census Bureau Research Network.

Friday, October 25, 2013 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

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Estimation of Mean Response Via Effective Balancing Score*

Zonghui Hu, Ph.D.

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases National Institutes of Health

Abstract:

We introduce effective balancing scores for estimation of the mean response under MAR (missing at random). Unlike conventional balancing scores, the effective balancing scores are constructed via dimension reduction free of model specification. Three types of effective balancing scores are introduced, carrying the covariate information about the missingness, the response, or both. They lead to consistent estimation with little or no loss in efficiency. Compared to existing estimators, the effective balancing score based estimator relieves the burden of model specification and is the most robust. It is a near-automatic procedure that is most appealing when high dimensional covariates are involved. We investigate both the asymptotic and the numerical properties, and demonstrate the proposed method in a study of HIV disease.

*Full list of authors for this paper: Hu Z., Follmann D.A., Wang N.

Friday, November 8, 2013 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.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

SHARE: Statistical and Synthetic Health Information Release with Differential Privacy

Li Xiong, Ph.D.

Associate Professor, Department of Biomedical Informatics, Emory University

Abstract:

Protecting privacy of human subjects while enabling large-scale analysis of clinical and public health data is a key challenge in health research. The current de-identification approach or microdata release is subject to various re-identification and disclosure risks and does not provide sufficient protection for our patients. I will present the SHARE (Statistical Health informAtion RElease) framework for sharing statistical or synthetic health data for secondary use with the state-of-art differential privacy guarantee. I will present our ongoing work on techniques for handing different types of data including relational, sequential, and time series data. I will also present case studies using real public health datasets and demonstrate the feasibility as well as challenges of applying the differential privacy framework on biomedical data.

Friday, November 22, 2013 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.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

The Role of Statistics in Regulatory Decision Making

Lisa LaVange, Ph.D.

Office of Biostatistics, Office of Translational Sciences

Center for Drug Evaluation and Research, US FDA

Abstract:

Statisticians at the FDA play an important role in the regulation of medical products. In the early stages of a drug development program, FDA statisticians provide advice on study design and analysis and review protocols for planned studies. Upon completion of a development program and submission of the dossier to the FDA for approval, statisticians are responsible for the statistical review of the sponsor's submission, including evaluating the quality of the data and the sponsors' analyses and reporting. If approval is granted, attention turns to monitoring the safety of the drug as it becomes more widely available, primarily through the evaluation of spontaneous adverse event reporting and additional post-marketing studies. Statisticians in the Office of Biostatistics review hundreds of new drug and biologic applications each year, and advise sponsors on protocols numbering into the thousands. In addition to remaining up-to-date on the newest statistical methodologies, they are often called upon for innovative approaches to difficult regulatory problems. I will give a broad overview of current office initiatives, including the development of guidance documents and a recent push for open and transparent collaboration with industry on methods development. The role of statistics in addressing critical unmet medical needs will be illustrated with a discussion of anti-infective drug development. Areas where innovative statistical solutions or greater clarity on existing approaches are still needed will also be discussed. I will also share my thoughts on the importance of leadership for our profession as a whole, regardless of employment sector.

Friday, December 6, 2013 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.

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Friday, January 10, 2013 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Peter Zhang, Ph.D.

Director, Otsuka Pharmaceutical Development & Commercialization, Inc.

Missing Data and Meta-Analysis: Application in Drug Development and Commercialization

Abstract:

Evidence based on medicine and personalized medicine is getting more attention in medical policy and practice. High development cost, low development success, cost-disciplined health-care policies, and intense competition demand an efficient drug development process. New regulatory trend is use meta-analysis not only as supportive analysis but also as primary evidence for regulatory approval. New compounds need to bring value to patients by being safe, efficacious, and cost-effective as compared with existing treatment options. Meta-analysis facilitates integration and utilization of summary-level efficacy and safety data, providing a quantitative framework for comparative efficacy and safety assessment. This Commentary discusses the application and limitations of MBMA in global drug development and commercialization. An example to gain CHMP positive opinion and market approval using evidence based meta-analysis is also given.

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

Biostatistics

Bioinformatics

Biomathematics

Bio^3 Seminar Series

Friday, January 24, 2014 10:00-11:00 am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

Joseph Ibrahim, Ph.D.

Alumni Distinguished Professor and Director of Graduate Studies
University of North Carolina Gillings School of Global Public Health

Bayesian Sequential Meta-analysis Design in Evaluating Cardiovascular Risk in a New Anti-diabetic Drug Development Program

Abstract:

Recently, the Center for Drug Evaluation and Research at the Food and Drug Administration (FDA) released a guidance that makes recommendations about how to demonstrate that a new anti-diabetic therapy to treat Type 2 diabetes is not associated with an unacceptable increase in cardiovascular risk. One of the recommendations from the guidance is that Phase II and III trials should be appropriately designed and conducted so that a meta-analysis can be performed. In addition, the guidance implies that a sequential meta-analysis strategy could be adopted. That is, the initial meta-analysis could aim at demonstrating the upper bound of a 95% confidence interval (CI) for the estimated risk ratio to be < 1.8 for the purpose of enabling a new drug application (NDA) or a biologics license application (BLA). Subsequently after the marketing authorization, a final meta-analysis would need to show the upper bound to be < 1.3. In this context, we develop a new Bayesian sequential meta-analysis approach using survival regression models to assess whether the size of a clinical development program is adequate to evaluate a particular safety endpoint. We propose a Bayesian sample size determination methodology for sequential meta-analysis clinical trial design with a focus on controlling the family-wise Type I error rate and power. The partial borrowing power prior is used to incorporate the historical survival meta-data into the Bayesian design. Various properties of the proposed methodology are examined and simulation-based computational algorithms are developed to generate predictive data at various interim analyses, sample from the posterior distributions, and compute various quantities such as the power and the Type I error in the Bayesian sequential meta-analysis trial design. The proposed methodology is applied to the design of a hypothetical anti-diabetic drug development program for evaluating cardiovascular risk.

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

Bio^3 Seminar Series

Friday, February 28, 2014 10:00-11:00 am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Joshua Sampson, Ph.D.
National Cancer Institute

Leveraging Local IBD Increases the Power of Case/Control GWAS with Related Individuals

Abstract:

Large case/control genome-wide association studies (GWAS) often include groups of related individuals with known relationships. When testing for associations at a given locus, current methods only incorporate the familial relationships between individuals. Here, we introduce the chromosome-based Quasi Likelihood Score (cQLS) statistic that incorporates local Identity By Descent (IBD) to increase the power to detect associations. In studies robust to population stratification, such as those with case/control sibling pairs, simulations show that the study power can be increased by over 50%. In our example, a GWAS examining late-onset Alzheimer's disease, the p-values among the most strongly associated SNPs in the APOE gene all decrease, with the smallest p-value decreasing from 1.23 x 10^-8 to 7.70 x 10^-9. Furthermore, as a part of our simulations, we reevaluate our expectations about the use of families in GWAS. We show that, although adding only half as many unique chromosomes, genotyping affected siblings is more efficient than genotyping randomly ascertained cases. We also show that the benefit of genotyping cases with a family history of disease is expected to decrease as GWAS search for SNPs with smaller effect sizes.

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

Bio^3 Seminar Series

Friday, March 28, 2014 10:00-11:00 am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Tudor Oprea, MD, Ph.D.

Professor, Department of Biochemistry & Molecular Biology, University of New Mexico Cancer Center

Drug Discovery and Repurposing from an Academic Perspective

Abstract:

This talk will highlight: 1) the discovery of the first selective agonist (G1) and antagonist (G15) for the G-protein estrogen receptor (GPER); 2) the discovery of potent small GTP-ase inhibitors for Cdc42 and Rac1, including Ketorolac; 3) the uncertainty of screening data from CEREP and DrugMatrix. These topics will be supported by our DRUGSDB project, a database with accurate chemical annotations for active pharmaceutical ingredients and controlled vocabularies for indications, contra-indications and off-label indications. DRUGSDB covers for all FDA-approved small molecule drugs. Additional topics include the use of ATC and INN codes for molecular similarity, as well as "Target Central", our new on-line service to categorize the druggable genome.

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

Bio^3 Seminar Series

Friday, April 11, 2014 10:00-11:00 am

Martin Marietta Conference Room, 1st Floor, Lombardi Cancer Center
Refreshments will be provided at 9:45am

Yun Zhou, PhD

The Russell H. Morgan Department of Radiology and Radiological Science
Johns Hopkins University School of Medicine

Biomathematical modeling and statistical analysis in quantitative positron emission tomography

Abstract:

Positron emission tomography (PET) is a main molecular imaging modality used to in vivo measure biochemical and physiological activities at molecular levels in human beings and laboratory animals. Quantitative PET is now considered as a standard imaging technique to measure physiological and biochemical parameters such as blood flow, glucose metabolism, receptor density, and drug occupancy in living subjects. The validated biomathematical modeling approach and statistical methods developed in dynamic PET studies are now widely adapted by dynamic MRI, CT, and SPECT. The talk will start with basic principles and main techniques developed in last 3 decades in quantitative PET. The challenges and recent developments on the integration of spatial-temporal analysis in multi-function and multi-modality PET/CT/MRI studies will be reviewed and discussed.

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

Bio^3 Seminar Series

Friday, April 25, 2014 10:00-11:00 am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Nandita Mitra, PhD

Associate Professor of Biostatistics
University of Pennsylvania

Assessing the sensitivity of cost-effectiveness measures to unmeasured confounding

Abstract:

The cost-effectiveness of medical treatments can be evaluated using observational data, but the results are susceptible to bias from unmeasured confounders. To assess the magnitude of this potential bias, investigators may conduct a sensitivity analysis where the results of the study are re-evaluated under a range of potential unmeasured confounders. Here, we propose a sensitivity analysis procedure for the Net Monetary Benefit (NMB), a popular method for combining cost and survival outcomes into a single result. Because costs and survival are often highly skewed and censored, we develop this procedure using a Gamma generalized linear model for cost and a Weibull accelerated failure time model for survival. If a specific unmeasured confounder is present, we show that there are closed-form expressions for the relationship between the naive expected values, found by adjusting for all measured confounders, and the true expected values which account for the unmeasured confounder. These relationships can be used to conduct a sensitivity analysis by adjusting the NMB for a wide range of hypothesized confounders. Our general formulas allow for any unmeasured confounder which can be characterized using a moment-generating function, and also allow for separate unmeasured confounders to influence cost and survival. Due to the potential correlation between costs and survival, we propose a bootstrap method for estimation of variance.

Using simulations, we evaluate the performance of the adjustment for expected cost differences, expected survival differences, and NMB. We demonstrate the use of this method in a comparison of two treatments for Stage II/III bladder cancer in a SEER-Medicare cohort.

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