

2015 Seminars

12/02/2015

Statistical Learning, Inference and Models for Big Data

Nancy Reid, PhD

Biography: Dr. Nancy Reid is University Professor and Canada Research Chair in Statistical Methodology at the University of Toronto. Her research interests are in statistical theory, likelihood inference, and design of studies. Along with her colleagues she has developed higher order asymptotic methods both for use in applications, and as a means to study theoretical aspects of the foundations of inference, including the interface between Bayesian and frequentist methods. She is the Director of the Canadian Statistical Sciences Institute. Dr. Reid received her PhD from Stanford University, under the supervision of Rupert Miller. She taught at the University of British Columbia before moving to the University of Toronto, and has held visiting positions at the Harvard School of Public Health, University of Texas at Austin, Ecole Polytechnique Federale de Lausanne, and University College London. She has been President of the Institute of Mathematical Statistics and the Statistical Society of Canada, and Vice-President of the International Statistical Institute. She is a Fellow of the American Association for the Advancement of Science, the Royal Society of Canada and the Royal Society of Edinburgh. In December 2014 she was appointed Officer of the Order of Canada.

Abstract: The Canadian Statistical Sciences Institute and the Fields Institute for Research in the Mathematical Sciences recently completed a six month thematic research program with this title. I will give an overview of the topics covered with emphasis on linkages between different areas, common problems, and common strategies. While the program was only able to cover a small fraction of the world of "Big Data", the breadth of the material covered by the large number of speakers was very stimulating.

11/04/2015

Identifying treatment effect heterogeneity using propensity score based quantile regression

Matthew Cefalu, PhD

Biography: Dr. Matthew Cefalu is an Associate Statistician at the RAND Corporation, where his research is primarily focused on the development and application of novel methods for causal inference. Examples of past and present research projects include the Health-Related Behaviors Survey of Military Personnel, an independent assessment of the VA healthcare system, and the CAHPS Hospital Survey. Dr. Cefalu received his PhD in Biostatistics from Harvard University in 2013.

Abstract: There is a vast literature on estimating causal effects from observational data, and the majority of these methods focus on estimating marginal treatment effects (i.e. treatment effects in the entire population). However, it is often of interest to identify subpopulations for whom the treatment is most effective. We will use locally weighted quantile regression, where locality is based on the propensity score, to identify if treatment effect heterogeneity is present. This method will be illustrated using data from a study assessing the efficacy of Motivational Enhancement Therapy-Cognitive Behavioral Therapy 5 in treating adolescents with cannabis-related disorders.

10/07/2015

REVEALER: Mapping Genomic Alterations to Functional Profiles of Pathway Activation, Gene Dependency and Drug Sensitivity

Pablo Tamayo, PhD

Biography: Dr. Pablo Tamayo is a Cancer Researcher at UC San Diego Moores Cancer Center and a Professor at the UCSD School of Medicine. Prior to UCSD he worked as a senior computational biologist at the Broad Institute of MIT and Harvard, as a consulting member of staff for the Advanced Analytics group at Oracle Corp., as senior researcher and chief scientist of Thinking Machines Corp., at the Theoretical Division (T-8) of the Los Alamos National Laboratory and as a research assistant at Boston University. He obtained a Ph.D. in Statistical Physics and a B.S. in Physics Engineering. During the last two decades he has worked on the study of cancer pathways, models of oncogene activation, models of pharmacological response, discovery of disease subtypes and integrated models to delineate and characterize cellular cancer states. He has been an original contributor to the development of many genomic data analysis methods including Gene Set Enrichment Analysis (GSEA), the Molecular Signatures Database (MSigDB) and the GenePattern Analysis Environment. His most recent work has focused on the development of experimental and computational models of oncogenic transformation, cancer vulnerabilities and catalogs of oncogenic states. He has also worked on an information-theoretic approach to find associations and co-analyze diverse types of cancer data with different statistical properties. He has published over 130 articles with over 35,000 citations. His publication list can be found in: <http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/40500040>

Abstract: Systematic efforts to sequence the cancer genome have identified many of the recurrent mutations and copy number alterations in tumors. However, in many cases the role(s) played by these alterations is not obvious and necessitates an effective functional characterization of the pathways and networks that these genomic alterations regulate. Here we introduce REVEALER (Repeated Evaluation of VARIABLEs conditionAL Entropy and Redundancy), an analysis method that enables the discovery of an ensemble of mutually exclusive genomic alterations correlated with “functional” phenotypes, e.g., the activation or dependency of oncogenic pathways. We use REVEALER to identify complementary genomic alterations that account for a large fraction of the “activated” or “dependent” samples with respect to four targets: the transcriptional activation of β -catenin and NRF2, MEK-inhibitor sensitivity, and KRAS dependency. REVEALER was able to “re-discover” several known features, as well as identify a number of novel associations, demonstrating the power of using information-theoretic association metrics to combine functional profiles with extensive characterization of alterations in cancer genomes.

09/02/2015

Exome Sequencing and Analysis of Phenotypic Extremes to Identify Genetic Modifiers of Iron Status in Hemochromatosis HFE C282Y Homozygotes

Christine McLaren, PhD

Biography: Dr. Christine McLaren is Professor and Vice Chair of the Department of Epidemiology in the School of Medicine at the University of California, Irvine. Dr. McLaren is also co-Leader of the Program in Cancer Prevention, Outcomes, and Survivorship and a member of the Biostatistics Shared Resource of the Chao Family Comprehensive Cancer Center, at UC Irvine. Dr. McLaren has focused on statistical modeling research and has concentrated on two important areas: (1) statistical modeling of biomedical data and (2) collaborative research in cancer. She is an elected Fellow of the American Statistical Association, in part for “innovative research in biology and medicine”. Dr. McLaren is Principal Investigator of the NIH/NDDK R24 grant, “Genetic Modifiers of Iron Status in Hemochromatosis HFE C282Y Homozygotes”.

Abstract: Approximately one million people in the United States are at risk for development of iron overload, attributable primarily to the genetic disorder known as hemochromatosis. In the NIH-funded Hemochromatosis and Iron Overload Screening (HEIRS) Study, 101,168 multi-ethnic participants in primary care were screened for iron overload and hemochromatosis. Dr. McLaren will describe her role as PI of a Field Center for the HEIRS Study and her contributions to study design and analyses. Her team enrolled over 20,000 primary-care patients in UC Irvine primary-care clinics and in community clinics throughout Orange County. She will also describe subsequent statistical studies designed to answer the question “What role do genetic modifiers play in determining iron accumulation in persons homozygous for the HFE C282Y genotype”.

07/01/2015

Case series analysis of infection-cardiovascular risk in patients on dialysis with exposure onset measurement error

Danh V. Nguyen, PhD

Biography: Danh Nguyen, PhD, is Professor in the Department of Medicine, Division of General Internal Medicine and Director of the Biostatistics, Epidemiology & Research Design (BERD) Unit in the Institute for Clinical and Translational Science, University of California Irvine. Prior to joining UC Irvine in 2013, he was Professor in the Division of Biostatistics, Department of Public Health Sciences, at UC Davis from 2003

Abstract: Cardiovascular disease and infection are major factors for morbidity and mortality in patients on dialysis. Hospitalization data from United States Renal Data System (USRDS) captures nearly all (> 95%) patients with end-stage renal disease in the U.S., the largest source of research data available for this population. Although the precise mechanisms by which infection may affect cardiovascular events are not fully known, infections may affect vascular endothelium, create a chronic sub-clinical inflammatory state that affects atherosclerosis, or may create a procoagulant state. Thus, we hypothesize that the time period following infection are associated with increased cardiovascular event risk. The self-controlled case series, or simply called case series, design/method and analysis of infection-cardiovascular risk in patients on dialysis using USRDS data presents several unique challenges, including (1) the timing of infection (or exposure) onset error since the time of infection is not known precisely, (2) misspecification of risk period, (3) and other inferential challenges, such as formal hypothesis testing. In this talk I will discuss current resolutions/developments for some of these challenges related to case series analysis and open topics in other areas of applications.

06/03/2015

Identifying Longitudinal Trends within EEG Experiments

Damla Senturk, PhD

Biography: Dr. Damla Senturk received her Ph.D. degree in Statistics from UC Davis in 2004 and joined the faculty in the Department of Statistics at Pennsylvania State University. She joined the faculty of the UCLA Department of Biostatistics in 2011 where she has been an Associate Professor in Residence since July 1st of 2013. Her main areas of statistical methodology research are longitudinal and functional data analysis, semiparametric adjustments in regression modeling and measurement error models. Her main collaborative research areas include psychiatry and nephrology.

Abstract: Differential brain response to sensory stimuli is very small (a few microvolts) compared to the overall magnitude of spontaneous electroencephalogram (EEG), yielding a low signal-to-noise ratio (SNR) in studies of event-related potentials (ERP). To cope with this phenomenon, stimuli are applied repeatedly and the ERP signals arising from the individual trials are averaged at the subject level. This results in loss of information about potentially important changes in the magnitude and form of ERP signals over the course of the experiment. In this paper, we develop a meta-preprocessing step utilizing a moving average of ERP across sliding trial windows, to capture such longitudinal trends. We embed this procedure in a weighted linear mixed effects model to describe longitudinal trends in features such as ERP peak amplitude and latency across trials while adjusting for the inherent heteroskedasticity created at the meta-preprocessing step. The proposed unified framework, including the meta-processing and the weighted linear mixed effects modeling steps, is referred to as MAP-ERP (Moving-Averaged-Processed ERP). We perform simulation studies to assess the performance of MAP-ERP in reconstructing existing longitudinal trends and apply MAP-ERP to data from young children with autism spectrum disorder (ASD) and their typically developing counterparts to examine differences in patterns of implicit learning, providing novel insights about the mechanisms underlying social and/or cognitive deficits in this disorder.

05/06/2015

Social Networks and Health: From Observation to Experimentation to Intervention

James Fowler, PhD

Biography: Dr. James Fowler earned a PhD from Harvard in 2003 and is currently a Professor at the University of California, San Diego. His work lies at the intersection of the natural and social sciences, with a focus on social networks, behavior, evolution, politics, genetics, and big data. Dr. Fowler was named a Fellow of the John Simon Guggenheim Foundation, one of Foreign Policy's Top 100 Global Thinkers, TechCrunch's Top 20 Most Innovative People, Politico's 50 Key Thinkers, Doers, and Dreamers, and Most Original Thinker of the year by The McLaughlin Group. He has also appeared on The Colbert Report. His research has been featured in numerous best-of lists including New York Times Magazine's Year in Ideas, Time's Year in Medicine, Discover Magazine's Year in Science, and Harvard Business Review's Breakthrough Business Ideas. Together with Nicholas Christakis, James wrote a book on social networks for a general audience called Connected. Winner of a Books for a Better Life Award, it has been translated into twenty languages, named an Editor's Choice by the New York Times Book Review, and featured in Wired, Oprah's Reading Guide, Business Week's Best Books of the Year, and a cover story in New York Times Magazine.

Abstract: From Framingham to Facebook, we have used a variety of social networks to measure, analyze, and change the effect of social networks on health. In this talk I will discuss a number of papers using different methods to better understand how networks function and what we can do to use them to make people healthier.

04/15/2015

Analysis of Longitudinal Data Under Biased Sampling

Yong Chen, PhD

Abstract: Over the past few decades, a dramatic increase in the incidence of obesity has become a worldwide health issue, contributing significantly as a risk factor of many diseases. Many individuals participate in web-based weight loss programs where their weights, physical activities and diets are self-reported. Such web-based program generated data poses new challenges to statistical modeling and inference, including subject-specific self-reporting times and outcome-dependent missingness. These challenges are known as biased sampling problem in statistical literature, and can lead to substantial bias in inference. In this talk, we propose a framework of novel statistical methods to efficiently detect and adjust for sampling bias, and to evaluate both the overall effectiveness of the weight loss program and the subject-specific effects of website usages on weight loss. The proposed methods provide elegant solutions for detecting and eliminating the impacts of biased sampling, and can achieve unbiased inference without fully specifying the complex data-generating mechanism. We apply the proposed methods to evaluate the effectiveness of a web-based program on weight loss, controlling the nonlinear trajectory of weights over time.

04/06/2015

Robust mixed-effects model for clustered failure time data: application to Huntington's disease event measures

Tanya P. Garcia, PhD

Biography: Tanya is an Assistant Professor in the Department of Epidemiology and Biostatistics at Texas A&M University, Health Science Center, School of Public Health. Previously, she worked in the Bioinformatics Training Program at Texas A&M University. She received a Ph.D. in Statistics from Texas A&M University in 2011 under the advisement of Prof. Yanyuan Ma. She earned a B.S. in Mathematics from the University of

California, Irvine in 2003, an M.S. in Industrial Engineering and Operations Research from the University of California, Berkeley in 2005, and an M.S. in Statistics from the University of Western Ontario in 2006. Her research interests include genetic mixture models, high-dimensional inference, measurement error, mixed models, neurodegenerative diseases, nonparametric models, semiparametric theory, measurement error, and survival analysis.

Abstract: An important goal in clinical and statistical research is estimating the distribution for clustered failure times, which have a natural intra-class dependency and are subject to censoring. We propose to handle these inherent challenges with a novel approach that does not impose restrictive modeling or distributional assumptions. Rather, using a logit transformation, we relate the distribution for clustered failure times to covariates and a random, subject specific effect such that the covariates are modeled with unknown functional forms, and the random effect is distribution-free and potentially correlated with the covariates. Over a range of time points, the model is shown to be reminiscent of an additive logistic mixed effect model. Such a structure allows us to handle censoring via pseudo-value regression and develop semiparametric techniques that completely factors out the unknown random effect. We show both theoretically and empirically that the resulting estimator is consistent for any choice of random effect distributions and for any dependency structure between the random effect and covariates. Lastly, we illustrate the method's utility in an application to the Cooperative Huntington's Observational Research Trial data, where our method provides new insights into differences between motor and cognitive impairment event times in genetically predisposed Huntington patients.

03/04/2015

Population Genetics, Biostatistics and Bioinformatics Issues in Individualized Medicine

Nicholas J. Schork, PhD

Biography: Nicholas J. Schork is a Professor and Director of Human Biology at the J. Craig Venter Institute (JCVI) and the Head of Integrated Genomics at Human Longevity, Inc. (HLI). He is also an adjunct Professor of Psychiatry and Family and Preventive Medicine (Division of Biostatistics) at the University of California, San Diego (UCSD). Prior to joining JCVI, Dr. Schork was, from 2007-2013, a Professor, Molecular and Experimental Medicine, at The Scripps Research Institute (TSRI), Director of Biostatistics and Bioinformatics at the Scripps Translational Science Institute (STSI), and Director of Research at Scripps Genomic Medicine, a division of Scripps Health. From 2001-2007 Dr. Schork was a Professor of Biostatistics and Psychiatry, and Co-Director of the Center for Human Genetics and Genomics, at UCSD. From 1994-2000, he was an Associate Professor of Epidemiology and Biostatistics at Case Western Reserve University in Cleveland, Ohio, and an Adjunct Associate Professor of Biostatistics at Harvard University. During 1999 and 2000, Dr. Schork took a sponsored leave of absence from CWRU to conduct research as the Vice President of Statistical Genomics at the French biotechnology company, Genset, where he helped guide efforts to construct the first high-density map of the human genome.

Dr. Schork's interests and expertise are in quantitative human genetics and integrated approaches to complex biological and medical problems, especially the design and implementation of methodologies to dissect the determinants of complex traits and diseases. He has published over 450 scientific articles and book chapters on the analysis of complex, multifactorial traits and diseases. A member of several scientific journal editorial boards, Dr. Schork is a frequent participant in U.S. National Institutes of Health-related steering committees and review boards, and has founded or served on the advisory boards of ten companies. In addition, he is currently director of the quantitative components of a number of national research consortia, including the NIA-sponsored Longevity Consortium and the NIMH-sponsored Bipolar Consortium. Dr. Schork earned the B.A. in Philosophy, M.A. in Philosophy, M.A. in Statistics, and Ph.D. in Epidemiology, all from the University of Michigan in Ann Arbor. .

Abstract: There is a great deal of attention surrounding 'individualized,' 'personalized,' and/or 'precision' medicine. Much of this attention has been motivated by technological advances in genetic and related molecular assays that have provided researchers with an unprecedented ability to identify and characterize the potentially unique determinants of an individual's disease susceptibility. However, as promising as these technologies are, their routine use in clinical settings will be hampered until they are appropriately vetted. In this talk, a number of studies are described that consider the use of genomic profiling to further efforts in individualized medicine. Focus is on the very thorny issues these studies have been designed to address, including dealing with patient genetic background heterogeneity, matching drugs to tumor genomic profiles in real-time clinical trial settings, exploring the utility of therapeutic interventions thought to be appropriate for an individual patient based on genomic profiling and monitoring genetically susceptible individuals. There is no doubt that individualized medicine will have a positive impact on health care, but only after some of the challenges it brings have been exposed and dealt with appropriately.

02/04/2015

Statistical Issues in the Analysis of Data from RNA-Seq Experiments

David Rocke, PhD

Abstract: RNA-Seq data are increasingly used for whole-genome differential mRNA expression analysis in lieu of gene expression arrays such as those from Affymetrix and Illumina. We review commonly used methods for this type of analysis, including DESeq, edgeR, and Cuffdiff2, by placing them within a common framework that allows comparisons of components of the methods as well as of the overall results. We also review a number of recent studies comparing these methods in terms of false positives and sensitivity, and add additional results of our own. We show that none of the existing methods is fully satisfactory, with most identifying large numbers of genes as differentially expressed even when there are none, but some will lead to better, more reliable results than others. This area is still early in its intellectual development and is changing rapidly, so there are substantial contributions that can be made.

1/30/2015

A multistate model for time to cancer recurrence and death incorporating a cured fraction

Jeremy Taylor, PhD

Biography: Jeremy M G Taylor PhD is the Pharmacia Professor of Biostatistics at the University of Michigan. He obtained a Bachelor's degree in Mathematics and a Diploma in Statistics from Cambridge University and a PhD in Statistics from University of California Berkeley. He was a faculty member in the Department of Biostatistics and the Department of Radiation Oncology at UCLA from 1983 to 1998. He is currently a faculty member in the Department of Biostatistics, the Department of Radiation Oncology and the Department of Computational Medicine and Bioinformatics and the Director of the Center for Cancer Biostatistics at the University of Michigan. He is the winner of the Michael Fry award from the Radiation Research Society and the Mortimer Spiegelman award from the American Public Health Association. He is a former Chair of the Biometrics section of the American Statistical Association and a Fellow of the ASA. He is the former chair of the Biostatistical Methods and Research Design grant review committee for the National Institutes of Health. He is currently one of the coordinating editors of Biometrics. He has 300 publications and research interests in longitudinal and survival data, cure models, methods for missing data, biomarkers, surrogate and auxiliary variables. He has worked extensively in AIDS research but currently mainly focuses on cancer research.

Abstract: Motivated by data from multiple randomized trials of colon cancer, we model time-to-cancer-recurrence and time-to-death using a multi-state model. We incorporate a latent cured state into the model to allow for subjects who will never recur. Parametric models that assume Weibull hazards and include baseline covariates are used. Information from the multiple trials are included using a hierarchical model. Bayesian

estimation methods are used. The model is used to assess whether there is improved efficiency in the analysis of the effect of treatment on time-to-death in each trial by using the information provided by earlier cancer recurrence. For subjects who are censored for death, multiple imputation is used to impute death times, where the imputation distribution is derived from the estimated model. Gains in efficiency are possible, although sometimes modest, using the extra information provided by the recurrence time.