

2014 Seminars

12/3/2014

Spatial Identification of Epilepsy Region

Giovanni Motta, PhD

Abstract: Epilepsy patients who are not able to adequately control their seizures with medications are sometimes treated with a neurosurgical procedure. The goal of this procedure is to remove the abnormal “epileptic” tissue causing seizures, and spare the normal tissue that is critical for brain function. However, current brain mapping technology has limited accuracy for mapping epileptic and normal brain tissue. This is especially problematic in the treatment of patients whose seizures arise from neocortex. To address these problems, we have been developing an experimental optical brain imaging technique for spatially mapping epileptic and normal cortical tissue. Better methods for the statistical analysis of the spatiotemporal optical imaging data are necessary for further development of this technique into a practical and reliable clinical tool.

In this paper we introduce a novel flexible tool, based on spatiotemporal statistical modeling of Optical Imaging, that allows for source localization of the epilepsy regions. The final goal is clustering (dimension reduction) of the pixels in regions, in order to localize the epilepsy regions for the craniectomy. We identify the spatial clusters of the pixels according to the temporal non-stationarity of the observed time series – rather than using spatial information. In a second step, we use non-parametric bootstrap and non-parametric density estimation to obtain the probabilities that a given pixel belongs to each of the clustered regions on the neocortex.

The advantage of our approach compared with previous approaches is twofold. Firstly, we use a non-parametric approach, rather than the (more restrictive) parametric or polynomial-based specification. Secondly, we provide a statistical method that is able to identify the clusters in a data-driven way, rather than the (sometimes arbitrary) ad-hoc currently used approaches.

To demonstrate how our method might be used for intra-operative neurosurgical mapping, we provide an application of the technique to optical data acquired from a single human subject during direct electrical stimulation of the cortex.

11/5/2014

A statistical approach to detecting patterns in behavioral event sequences

Hal S. Stern, PhD

Biography: Hal Stern is professor of statistics and dean of the Donald Bren School of Information and Computer Sciences at the University of California, Irvine. Stern came to UC Irvine in 2002 as the founding chair of the Department of Statistics. The Department now has 9 faculty and more than 40 graduate students in its MS/PhD programs. In 2010 he was named Ted and Janice Smith Family Foundation Dean of the Bren School. Prior to coming to UC Irvine he had faculty appointments at Iowa State and Harvard.

Within statistics he is known for his research work in Bayesian statistical methodology and model assessment techniques. He is a co-author of the highly-regarded graduate level statistics text Bayesian Data Analysis. Current areas of interest include applications of statistical methods in psychiatry and human behavior, atmospheric sciences, and forensic science. He is a Fellow of the American Statistical Association and the Institute for Mathematical Statistics and has served on several expert committees for the National Academies. Stern received his B.S. degree in Mathematics from the Massachusetts Institute of Technology in 1981 and the M.S. and Ph.D. degrees in Statistics from Stanford University in 1985 and 1987, respectively

Abstract: The identification of recurring patterns within a sequence of events is an important task in behavioral research. We consider a general probabilistic framework for identifying patterns by distinguishing between events that belong to a pattern and events that occur as part of background processes. Using this framework

we develop an inference procedure to detect sequences present in observed data and estimate the parameters governing these sequences. The model is applied to data from a study of the impact of fragmented and unpredictable maternal behavior on cognitive development of adolescents.

10/1/2014

Tree derived Survival risk groups in differentiating care for glioma patients

Annette Molinaro, MA, PhD

Biography: Dr. Molinaro is an Associate Professor in Residence at the Department of Epidemiology and Biostatistics, Department of Neurological Surgery, and Hellen Diller Family Comprehensive Cancer Center, at the University of California, San Francisco.

Abstract: We recently developed partDSA, a multivariate method that, similarly to CART, utilizes loss functions to select and partition predictor variables to build a tree-like regression model for a given outcome. However, unlike CART, partDSA permits both 'and' and 'or' conjunctions of predictors, elucidating interactions between variables as well as their independent contributions. partDSA thus permits tremendous flexibility in the construction of predictive models and has been shown to supersede CART in both prediction accuracy and stability. As the resulting models continue to take the form of a decision tree, partDSA also provides an ideal foundation for developing a clinician-friendly tool for accurate risk prediction and stratification.

With right-censored outcomes, partDSA currently builds estimators via either the Inverse Probability Censoring Weighted (IPCW) or Brier Score weighting schemes; see Lostritto, Strawderman and Molinaro (2012), where it is shown in numerous simulations that both proposed adaptations for partDSA perform as well, and often considerably better, than two competing tree-based methods. In this talk, various useful extensions of partDSA for right-censored outcomes are described and we show the power of the partDSA algorithm in deriving survival risk groups for glioma patient based on genomic markers.

9/2/2014

The effect of regional deprivation on mortality avoiding compositional bias: A natural experiment

Ursula Berger, PhD

Abstract: We assess the effect of regional deprivation on individual mortality by making use of a natural experiment: We followed up ethnic German resettlers from Former Soviet Union countries, who were quasi-randomly distributed across the socioeconomically heterogeneous counties of Germany's federal state North Rhine-Westphalia (NRW). This allows us to disentangle the contextual effect from compositional effects. We use data from the retrospective cohort study 'AMOR' on the mortality of resettlers in NRW (n=34 393). Based on the postcode of the last known residence we could link study participants to the municipalities of NRW. After a mean follow-up of 10 years, 2580 resettlers were deceased. When analyzing regional deprivation using in additive survival models, we explore the gain of more precise data on deprivation and of smaller regional entities? Our findings indicate that in terms of mortality, regional deprivation does matter.

<http://www.ibe.med.uni-muenchen.de/>

6/4/2014

TBA

Hernando Ombao, Ph.D.

Biography: Dr Ombao's research interests include: Time Series Analysis, Spatio-temporal modelling, Statistical Learning, Applications to Brain Science (fMRI, EEG, MEG, EROS)

<http://hombao.ics.uci.edu/hernando.html>

5/23/2014

A Hierarchical Model for Simultaneous Detection and Estimation in Multi-subject fMRI Studies

David Degras, Ph.D.

Assistant Professor, Statistics Department of Mathematical Sciences DePaul University College of Science and Health

Abstract: In this paper we introduce a new hierarchical model for the simultaneous detection of brain activation and estimation of the shape of the hemodynamic response in multi-subject fMRI studies. The proposed approach circumvents a major stumbling block in standard multi-subject fMRI data analysis, in that it both allows the shape of the hemodynamic response function to vary across region and subjects, while still providing a straightforward way to estimate population-level activation. An efficient estimation algorithm is presented, as is an inferential framework that not only allows for tests of activation, but also for tests for deviations from some canonical shape. The model is validated through simulations and application to a multi-subject fMRI study of thermal pain. The model is validated through simulations and application to a multi-subject fMRI study of thermal pain.

<http://depaul.academia.edu/DavidDegras>

5/7/2014

TBA

Babak Shahbaba, Ph.D.

Assistant Professor, Department of Statistics and Department of Computer Science, University of California, Irvine

Biography: Dr Shahbaba's research interest is related to developing new Bayesian methods and applying them to real-world problems. He is currently focusing on the following areas: Scalable Bayesian inference (fast MCMC methods that can be applied to large datasets); Developing new models that are sufficiently flexible and provide interpretable results; Incorporating appropriate priors into statistical models in order to improve their performance; Applying novel statistical methods to answer research questions in genetics, neuroscience, and cancer studies

<https://www.ics.uci.edu/~babaks/>

5/2/2014

Successive normalization/standardization of rectangular arrays

Richard Olshen, Ph.D.

Biography: Dr. Olshen is a Professor and Division Chief Division at the Biostatistics Department of Health Research and Policy at Stanford University School of Medicine

Abstract: When each subject in a study provides a vector of numbers/features for analysis, and one wants to standardize, then for each coordinate of the resulting rectangular array one may subtract the mean by subject

and divide by the standard deviation by subject. Each feature then has mean 0 and standard deviation 1. Data from expression arrays and protein arrays often come as such rectangular arrays, where typically column denotes "subject" and the other some measure of "gene." When analyzing these data one may ask that subjects and genes "be on the same footing." Thus, there may be a need to standardize across rows and columns of the matrix. We investigate the convergence of a successive approach to standardization, which we learned from colleague Bradley Efron. Limit matrices exist on a Borel set of full measure; these limits have row and column means 0, row and column standard deviations 1. We study implementation on simulated data and data that arose in cardiology. The procedure can be shown not to work with simultaneous standardization. Results make contact with previous work on large deviations of Lipschitz functions of Gaussian vectors and with von Neumann's algorithm for the distance between two closed, convex subsets of a Hilbert space. New insights regarding inference are enabled. Efforts are joint with colleague Bala Rajaratnam and have been helped by conversations with many others.

<http://statweb.stanford.edu/~olshen/>

4/2/2014

Standardized statistical framework for comparison of biomarkers: techniques from the Alzheimer's Disease Neuroimaging Initiative

Danielle Harvey, Ph.D.

Biography: Dr Harvey received her BA cum laude in mathematics from Pomona College and her PhD in statistics from University of Chicago. Her methodological interests span survival analysis, correlated event times, informative censoring, repeated measures, computational methods, and high-dimensional data as in MRI or PET scans. Collaborative research interests include work on Alzheimer's, cancer, end-of-life care, dosing errors, and health services and public health issues.

Abstract: Alzheimer's disease (AD) is widespread in the elderly population and clinical trials are ongoing, focused on elderly individuals with AD or at apparent risk for AD, to identify drugs that will help with this disease. Well-chosen biomarkers have the potential to increase the efficiency of clinical trials and drug discovery and should show good precision as well as clinical validity. We propose measures that operationalize the criteria of interest and describe a general family of statistical techniques that can be used for inference-based comparisons of marker performance. The methods are applied to regional volumetric and cortical thickness measures quantified from repeat structural magnetic resonance imaging (MRI) over time of individuals with mild dementia and mild cognitive impairment enrolled in the Alzheimer's Disease Neuroimaging Initiative. The methodology presented provides a standardized framework for comparison of biomarkers and will help in the search for the most promising biomarkers.

<http://biostats.ucdavis.edu/FacultyProfile.php?id=156>

3/5/2014

Local False Discovery Rate and Effect Size Estimation for Highly Polygenic Complex Traits

Wesley K. Thompson, Ph.D.

Biography: Dr. Thompson earned his Ph.D. in Statistics from Rutgers University in 2003, and his dissertation studies focused on the development of a Bayesian model for sparse functional data. He was appointed Assistant Professor of Statistics and Psychiatry at the University of Pittsburgh in 2005, and he collaborated with several senior investigators on clinical research studies on depression, sleep and sleep disorders, and physical illness across the lifespan. Dr. Thompson joined the UCSD Department in 2008 and he serves as the Director of Biostatistics at the Stein Institute for Research on Aging. Dr. Thompson's research interests center on the adaptation and application of statistical models of a dynamic covariation of multiple functional processes

in order to identify potentially causal relationships between brain function, depression, and physical health. This work is supported by a NIH Career Development Award that Dr. Thompson received in 2006. He is also interested in developing statistical models that may explain the underlying mechanisms of healthy cognitive aging.

Complex traits and disorders such as schizophrenia are multifactorial and associated with the effects of multiple genes in combination with environmental factors. These disorders often cluster in families, have no clear-cut pattern of inheritance, and have a high fraction of phenotypic variance attributable to genetic variance (high heritability). It is becoming increasingly clear that many genes influence most complex traits and disorders. In such a scenario with a very high number of risk genes ('polygenic'), each gene has a tiny effect. This makes it difficult to determine an individual's risk, and to identify disease mechanisms that can be used for development of new effective treatments.

Genome-wide association studies (GWAS) have identified many trait-associated single nucleotide polymorphisms (SNPs), but so far these explain only small portions of the heritability of complex disorders. This "missing heritability" has been attributed to a number of potential causes, including lack of typing of rare variants. However, it has been shown that a large proportion of the missing heritability is available within GWAS data when associations of SNPs are examined in aggregate. This implies the existence of numerous common variants with small genetic ('polygenic') effects. These effects cannot be reliably detected with traditional GWAS statistical methods given current sample sizes. Thus, there is a need for innovative statistical approaches to identify polygenic effects and reduce the proportion of 'missing heritability'.

In this talk I describe novel statistical tools that enhance gene discovery, improve replication rates of discovered risk gene variants, and improve estimation of polygenic risk scores. The basic framework relies on extensions of a Bayesian two-group mixture model (Efron, 2010) that assumes a large proportion of loci are either null (unassociated with the phenotype of interest) or have very small effects, but that a small proportion have larger (though still small) effect sizes. These models can incorporate a priori information regarding functional roles of SNPs or pleiotropic effects with multiple phenotypes. We demonstrate these methods on GWAS data from large Crohn's disease and Schizophrenia meta-analyses.

<http://psychiatry.ucsd.edu/faculty/wthompson.html>

2/21/2014

TBA

Donald B. Rubin, Ph.D.

Biography: Donald B. Rubin is John L. Loeb Professor of Statistics, Harvard University, where he has been professor since 1983, and Department Chair for 13 of those years. He has been elected to be a Fellow/Member/Honorary Member/Research Fellow of: the Woodrow Wilson Society, John Simon Guggenheim Memorial Foundation, IZA, IAB, Alexander von Humboldt Foundation, American Statistical Association, Institute of Mathematical Statistics, International Statistical Institute, American Association for the Advancement of Science, American Academy of Arts and Sciences, European Association of Methodology, British Academy, and the U.S. National Academy of Sciences. He has authored/coauthored nearly 400 publications (including ten books), has four joint patents, and he has made important contributions to statistical theory and methodology, particularly in causal inference, design and analysis of experiments and sample surveys, treatment of missing data, and Bayesian data analysis. Among his other awards and honors, Professor Rubin has received the Samuel S. Wilks Medal from the American Statistical Association, the Parzen Prize for Statistical Innovation, the Fisher Lectureship and the George W. Snedecor Award of the Committee of Presidents of Statistical Societies. He was named Statistician of the Year, American Statistical Association, Boston and Chicago Chapters. He has served on the editorial boards of many journals, including: Journal of Educational Statistics, Journal of American Statistical Association, Biometrika, Survey Methodology, and Statistica Sinica. Professor Rubin has been, for many years, one of the most highly cited authors in

mathematics in the world (ISI Science Watch), as well as in economics (Highly Cited Economists), with approximately 140,000 citations, with nearly 30,000 so far in 2012 and 2013 (according to Google Scholar). For many decades he has given keynote lectures and short courses in the Americas, Europe, Australia and Asia. He has also received honorary doctorate degrees from Otto Friedrich University, Bamberg, Germany and the University of Ljubljana, Ljubljana, Slovenia, and held the Honorary Belle van Zuylen Chair in the Department of Methodology and Statistics at the University of Utrecht, the Netherlands in 2012 -2013.

<http://statistics.fas.harvard.edu/people/donald-b-rubin>

2/5/2014

Alternative Tumor Measurement-based Phase II Clinical Trial Endpoints for Predicting Overall Survival (OS), using the RECIST 1.1 data warehouse

Ming-Wen An, PhD

Assistant Professor, Department of Mathematics, Vassar College, Poughkeepsie, NY

Biography: Ming-Wen An received her B.A. in mathematics from Carleton College and her Ph.D. in biostatistics from the Johns Hopkins Bloomberg School of Public Health. One of her research interests is in issues of study design for addressing missing data due to "loss to follow-up" (with applications to evaluating HIV treatment programs in Africa). She is also interested in cancer clinical trial methodology, specifically designs for validating biomarkers used in targeted therapy and identification of alternative endpoints for Phase II trials.

1/29/2014

Gaussian Oracle Inequalities for Structured Selection in Non-Parametric Cox Model

Jelena Bradic, PhD

Biography: Dr. Bradic received her Ph.D. in Operations Research and Financial Engineering from Princeton in Spring 2011 with a specialization in Statistics and Applied Probability under the direction of Jianqing Fan. Her research is in high dimensional statistics, stochastic optimization, asymptotic theory, robust statistics, functional genomics and biostatistics.

Abstract: In this paper, we study sparse structured estimation in the context of the high-dimensional non-parametric Cox proportional hazard's model with a very general family of group penalties. We study the finite sample oracle risk bounds of such regularized estimator and develop new techniques to do so. Unlike the existing literature, we exemplify differences between bounded and possibly unbounded non-parametric covariate effects. In particular, we show that unbounded effects can lead to larger prediction bounds, compared to simple linear models, in situations where the true parameter is not necessarily sparse. Moreover, we propose a sequence of sparse non-convex group regularizations. Interestingly, we identify a specific regime of the proposed non-convex estimation that allows the group SCAD penalty and the group Lasso penalty to have equivalent prediction errors. Oracle prediction bounds are also discussed for the group ℓ_0 penalty. Theoretical results for hierarchical and smoothed estimation in the non-parametric Cox model are also discussed as two examples of the proposed general framework.