40th Annual Meeting of Alberta Statisticians

Book of Abstracts

MacEwan University
Edmonton, Alberta, Canada

September 29, 2018

Contents

Saturday September 29, 2018

Plenary Address

Douglas Wiens, University of Alberta

Robustness of Design: A Survey

When an experiment is conducted for purposes which include fitting a particular model to the data, then the ‘optimal’ experimental design is highly dependent upon the model assumptions - linearity of the response function, independence and homoscedasticity of the errors, etc. When these assumptions are violated the design can be far from optimal, and so a more robust approach is called for. We should seek a design which behaves reasonably well over a large class of plausible models. I will review the progress which has been made on such problems, in a variety of experimental and modelling scenarios - prediction, extrapolation, discrimination, survey sampling, dose-response, machine learning, etc...

Session on Statistical Genetics

Qing (Leah) Li, University of Calgary

Introduction to statistical genetics and precision medicine

One of the ultimate goals in the field of genetics is to develop genome-based phenotype predictors, in other words, ways to confidently predict biological phenomena from specific genetic variations. The successful development of such predictors will play an essential role in precision medicine. There are two recent technological advancements that have brought excitement to enact this goal. One is the availability of multi-scale data such as Electronic Medical Record (EMR) and other ’omics data, linked in a single database (e.g., a biobank) or distributed in multiple sites. The other is the success of machine learning algorithms, especially deep learning, in many fields that involve big-data. The integration of the high-throughput medical data and advanced big-data algorithms has been a dominant research topic for a while. However, many issues on how to integrate such data and quantify their performance are yet to be resolved. Inputs from professional statisticians tailoring to the setting of genetics and medicine are urgently needed.

In this talk, I will introduce the basic concepts and main branches of the fields of statistical genetics and precision medicine. These include (1) the history of the fields and existing models; (2) the current problems that are extensively studied by statistical geneticists; and (3) the repositories of publicly available biological and medical big-data and how to acquire the access.

Quan Long, University of Calgary

Analysis of Inter-tissue co-expression network using linear mixed models

Inter-tissue molecular interactions are critical to the function and behaviour of biological systems in multicellular organisms, but systematic studies of interactions between tissues are lacking. Also, existing studies of inter-tissue interactions are based on direct gene expression correlations, which can't distinguish correlations due to common genetic architectures versus biochemical or molecular signal exchange between tissues.

We developed a novel strategy to study inter-tissue interaction by removing effects of genetic regulation of gene expression (genetic decorrelation) using linear mixed models. We applied our method to the comprehensive atlas of gene expression across nine human tissues in the Genotype-Tissue Expression (GTEx) project to generate novel genetically decorrelated inter-tissue networks. From this we derived modules of genes important in inter-tissue interactions that are likely driven by biological signal exchange instead of their common genetic basis. Importantly we highlighted communication between tissues and elucidated gene activities in one tissue inducing gene expression changes in others. We reveal global unidirectional inter-tissue coordination of specific biological pathways such as protein synthesis. Using our data, we highlighted a clinically relevant example whereby heart expression of DPP4 was coordinated with a gene expression signature characteristic for whole blood proliferation, potentially impacting peripheral stem cell mobilization.

Thierry Chekouo Tekougang, University of Calgary

Bayesian Group Selection for Compositional Data: Application to Radiomic data in Glioblastoma disease

In this talk, we focus on the analysis of volumetric imaging features of brain cancer patients collected through The Cancer Genome Atlas (TCGA) project. Our main goal is to identify genes and their pathways whose are significantly associated with volumetric features. In particular, we focus on the glioblastoma multiform (GBM), as it is the most common malignant brain neoplasm, accounting for 23% of all primary brain tumours for which it still has an appalling prognosis. We propose a Bayesian hierarchical model for variable selection with a group structure in the context of correlated multivariate compositional response variables. More specifically, we model the volumetric data using a Dirichlet model by allowing for straightforward incorporation of available high-dimensional covariate information within a log-linear regression framework. We impose prior distributions that account for the overlapping structure between groups. Simulations and application to GBM disease show the importance of our approach.

Regression Applications and Methodology

Alex Mackie, MacEwan University

An Investigation of the Relationship Between Indigenous Suicide in Canada and Spatial Factors Using Categorical Analysis

Among Indigenous people in Canada, suicide has become a major cause of death over the past two decades (Tjepkema et al. 2009). In 2016, the suicide rate for the First Nations of Canada was twice the national average (Crawford 2016). While much research has been done, little can be found specifically investigating a spatial relationship between suicide and the Indigenous population in Canada. However, studies have been performed that show the potential for finding such a relationship. In Tennessee, an unexpected spatial relationship was found between suicide and church density suggesting that counties whose neighbours have a high church density are more susceptible to higher suicide rates (Difurio and Lewis 2017). Another paper examined many studies and found that rural communities tend to exhibit higher rates of suicide than urban ones (Jagodic, Agius and Pregelj 2012). Using the 2012 Aboriginal Peoples Survey Personal Use Microdata File (APS PUMF), which contains 326 categorical variables, we performed χ^2 tests for independence to assess the relationship between two variables associated with suicide and two variables associated with geographical location. Furthermore, we used logistic regression, treating suicidal thoughts as a response variable and requiring at least one of our geographic variables be a predictor. While our findings suggest a significant relationship between geographical factors and suicidal thoughts, logistic regression models chosen based on accuracy of their ROC curve and a low AIC showed little predictive ability of suicidal thoughts. Further investigation is required and is recommended to be done with the APS Analytical file rather than the PUMF.

Selvakkadunko Selvaratnam, University of Alberta

Asymptotics for maximum likelihood (ML) estimators of generalized linear models (GLM) for adaptive designs with applications.

Due to increasing discoveries of biomarkers and observed diversity among patients, there is growing interest in personalized medicine for the purpose of increasing the well-being of patients and extending human life. In fact, these biomarkers and observed heterogeneity among patients are useful covariates which can be used to achieve the ethical goals of clinical trials and improving the efficiency of statistical inference. Covariate-adjusted response-adaptive (CARA) design was developed to use information in such covariates in randomization to maximize the well-being of participating patients. In this paper, we establish conditions for consistency and asymptotic normality of maximum likelihood (ML) estimators of generalized linear models (GLM) for a general class of adaptive designs. We prove that the ML estimators are consistent and asymptotically follow a multivariate Gaussian distribution. The efficiency of the estimators and the performance of response-adaptive (RA), CARA, and completely randomized (CR) designs are examined based on the well-being of patients under a logit model with categorical covariates. We demonstrate that RA designs lead to ethically desirable outcomes as well as higher statistical efficiency compared to CARA designs if there is no treatment by covariate interaction in an ideal model. CARA designs were however more ethical than RA designs when there was significant interaction.

Adam Kashlak, University of Alberta

Sparse Covariance Regression

Linear regression in the high dimensional setting, $p >> n$, poses many challenges to researchers. Standard approaches such as LASSO assume sparsity in the parameter vector beta. We consider an alternative setting where the entries in the parameter vector beta are mostly non-zero under a different structural assumption that the covariance of the regressors is sparse. That is, for a design matrix $\mathbf{X}$, the matrix $\mathbf{X}^t\mathbf{X}$ is nearly diagonal. Under this structural assumption, we use a sparse matrix estimation technique to derive a novel estimator for beta and compare this new estimator to standard approaches such as ridge regression and LASSO. This project was joint work with Xi Fang, a visiting student from UIBE.

Statistical Smorgusbord

Peng Liu, University of Alberta

Efficient Low Rank Matrix Recovery by M-estimation

This paper considers the low rank matrix recovery from linear measurements. The loss function is based on M-estimation, and hence has a general form. Especially the Huber M-estimation is well known to have a performance robust to outliers, and we may expect that it will lead to an efficient approach for the matrix recovery problem, which actually is confirmed by our simulation studies. The alternating method is employed to minimize the objective function and, at each minimization step, the Nesterov's momentum algorithm is used accelerate the convergence. For the case with nonsmooth objective function, Nesterov's smoothing technique is first considered to obtain an optimal smooth approximation. The effectiveness of our method is validated by theoretical studies on convergence analysis, and by numerical studies on both synthetic and real data.

Wei Tu, University of Alberta

Quantile-optimal Treatment Regime based on Individual Gain

A treatment regime is a rule that assigns a treatment, among a set of possible treatments, to a patient based on individual characteristics. The problem of finding optimal treatment regime aims to identify the treatment regime that would lead to the best outcome on "average", provided the entire population of patients is followed. In the current literature, the optimal treatment regime is usually defined as the one that maximize the mean outcome. During this talk, we introduce quantile-optimal treatment regime. Quantile-optimal treatment regime provides a robust and more flexible alternative to the mean-optimal treatment regime, especially when the the outcome is skewed or censored. Specifically, the introduced framework aims to maximize the the individual gain of receiving the treatment against not. This is closely related to residual learning and concordance assisted learning proposed in the mean-optimal treatment regime literature. We conduct extensive numerical simulations and real data example to illustrate the performance of proposed framework. At last, we discuss the possible extensions to dynamic treatment regime.

Cristina Anton, MacEwan University

Statistical Analysis of a Model for Toxicity Assessment

We consider a stochastic model based on the logistic equation and linear kinetics to study the effect of toxicants with various initial concentrations on a cells' population. The EM algorithm and the unscented filter are used for parameter estimation and to predict the concentration of toxicants outside the cells. We also study the effect of parameter uncertainties.