International Chinese Statistical Association

泛華統計協會

Canada Chapter

Advances and Innovations in Statistics and Data Science

The 4th ICSA – Canada Chapter Symposium

August 9-11, 2019
Queen's University, Kingston
Ontario, Canada

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques
Data • Discoveries • Decisions
Données • Découvertes • Décisions

Faculty of
Arts and Science

Queen's University
Department of Mathematics and Statistics

Welcome to ICSA – Canada Chapter 2019 Symposium

The Department of Mathematics and Statistics at Queen's University welcomes you to the 4th International Chinese Statistical Association (ICSA) Canada Chapter Symposium in Kingston! We are thrilled to be hosting this exciting academic event. The 2019 symposium will feature a wide range of scientific programs focusing on the latest advances and innovations in statistics and data science as well as two short courses on the cutting-edge statistical methodology. It will bring together over 130 statisticians and researchers from academia, industry and government in many different countries, especially from China, United States, and Canada. We hope to provide a platform to exchange ideas and explore opportunities for collaborations, and that junior statisticians (including the current students) are able to learn from and interact with senior members of the profession.

ICSA-Canada Chapter was founded in 2013. The first symposium was held in Toronto in 2013, the second and third symposium were held in Calgary and Vancouver in 2015 and 2017. The statistics community in Canada has been growing rapidly in past years, and the symposia serve as one of the best venues to bring together statisticians in Canada and around the world to present and discuss research and practices in different fields of statistics and their applications. The organizing committee of the 2019 symposium has been working tirelessly over the past twelve months to make sure that the symposium will be a successful one. More detailed information about the symposium, including a full version of the program book with abstracts of all talks, can be found at http://www.icsa-canada-chapter.org/symposium2019/

Kingston, known as the first capital of Canada, is rich in history and culture. It boasts vibrant arts and cultural communities, and internationally-recognized historical sites. For the list of the tourism places, please visit the symposium website.

Devon Lin and Wenyu Jiang
Local Organizing Committee
Department of Mathematics and Statistics
Queen's University

Sponsors

The 4th ICSA-Canada Chapter Symposium acknowledges the generous supports received from the following sponsors (in alphabetical order).

- **Canadian Statistical Sciences Institute (CANSSI)**
- **Department of Mathematics and Statistics, Queen's University**
- **Faculty of Arts and Science, Queen's University**

Committees and Chairs

ICSA – Canada Chapter Executive Committee

- Liqun Wang, Chair, University of Manitoba
- Changbao Wu, Past-Chair, University of Waterloo
- Yingwei Paul Peng, Chair-Elect, Queen's University
- Leilei Zeng, Secretary/Treasurer, University of Waterloo
- Jianan Peng, East Canada Representative, Acadia University
- Devon Lin, Central Canada Representative, Queen's University
- Longhai Li, West Canada Representative, University of Saskatchewan

Symposium Organizing Committee

- Liqun Wang (Chair, ICSA-Canada Chapter) University of Manitoba
- Wenqing He (Chair of the Program Committee) University of Western Ontario
- Devon Lin (the Local Committee) Queen's University
- Wenyu Jiang (the Local Committee) Queen's University
- Lang Wu (Chair of 2017 Symposium) University of British Columbia
- Xuewen Lu (Chair of 2015 Symposium) University of Calgary
- Leilei Zeng (Treasurer and Secretary) University of Waterloo
- Changbao Wu (Past Chair, ICSA-Canada Chapter) University of Waterloo

Symposium Website Manager

- Zhiyong Jin, University of Manitoba

Local Volunteers

- Kian Blanchette
- Xinyi Ge
- Jiani Heng
- Jordan Kokocinski
- Na Li
- Wanlu Li
- Wen Teng
- Feng Yang
- Qianhui Yu
- Jianwei Yue
- Zimo Zhu

General Information

Symposium Venue

Biosciences Complex (116 Barrie Street) and School of Kinesiology and Health Studies (28 Division Street) will be used for plenary talks and scientific sessions. The short courses, two plenary talks, and registration will take place at Biosciences Complex, while the scientific sessions will be hosted at both buildings.

Campus Wifi

SSID: QueensuSecure_WPA2 LOGIN: mathwifi PASSWORD: M@thWiFi2019

Registration

Registration will be available in the Atrium of Biosciences Complex on 2 pm – 6pm on Friday, Aug 9. The registration desk will stay open in the Atrium of Biosciences Complex from 8:30am to 5:30pm on Saturday, Aug 10 and from 8:30am to 10am on Sunday, Aug 11.

Lunch

The buffet lunches are available for short courses participants and symposium participants at Leonard Hall (150 Queen's Crescent).

Banquet

Banquet will take place at Isabel Bader Center at 390 King Street West on Aug 10th. The Isabel Bader Center is about 15-minute beautiful walk along the Lake Ontario from Biosciences Complex. The open bar is available at 6pm while the banquet starts at 7:15pm.

Annual General Meeting

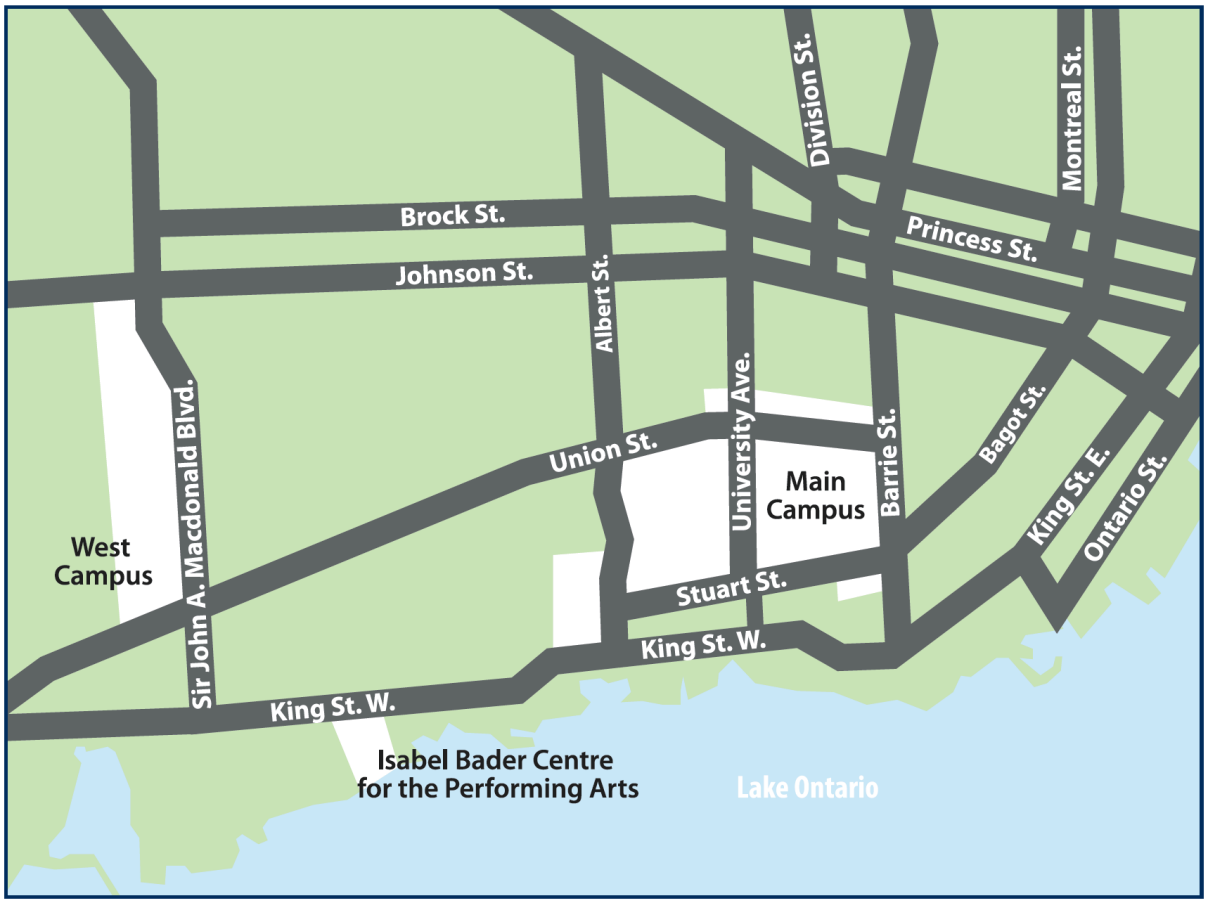
The annual general meeting (AGM) will take place at 5:45 pm – 6:30pm on Aug 10 at Room 1120 in Biosciences Complex.

One Thousand Island Cruise

The cruise takes place between 2:30 – 5pm on Aug 11, Sunday. It will set out in the town of Gananoque (35-minute drive east of Kingston). The bus will pick up all the participants at Leonard Hall at 1:10pm on Aug 11, and drop off on campus around 5:45pm.

Campus Security, Emergency and Hospitals

For campus emergency, contact 613-533-6111. There are two hospitals in downtown Kingston, Kingston General Hospital and Hotel Dieu Hospital.



Schedule for Friday, August 9

Room 1120, Biosciences Complex

- Short Course 1. *Statistical Analysis of High Dimensional Data.*
  Presenter: ***Wenqing He****, University of Western Ontario*

- Short Course 2. *Introduction of Reinforcement Learning.*
  Presenter: ***Linglong Kong****, University of Alberta*

9:00am – 10:20am	Short Course 1, Part I
10:20am – 10:40am	Coffee Break
10:40am – 12:00pm	Short Course 1, Part II
12:00pm – 2:00pm	Lunch Break (Leonard Hall)
2:00pm – 3:20pm	Short Course 2, Part I
3:20pm – 3:40pm	Coffee Break
3:40pm – 5:00pm	Short Course 2, Part II

Schedule for August 10 and 11

8:45am-9:00am, Opening Remarks

Plenary Talk I
9am-10am, Saturday, August 10th

Session 1: ***Keynote Speech 1***
Organizer and Chair: Liqun Wang, University of Manitoba
Room: B-1102, Time: 9:00AM-10:00AM

(1) Tony Cai, University of Pennsylvania
Title: *When Statistics Meets Computing*

Abstract: In the conventional statistical framework, the goal is developing optimal inference procedures, where optimality is understood with respect to the sample size and parameter space. When the dimensionality of the data becomes large as in many contemporary applications, the computational concerns associated with the statistical procedures come to the forefront. A fundamental question is: Is there a price to pay for statistical performance if one only considers computable (polynomial-time) procedures? After all, statistical methods are useful in practice only if they can be computed within a reasonable amount of time.

In this talk, we discuss the interplay between statistical accuracy and computational efficiency in two specific problems: submatrix localization and sparse matrix detection based on a noisy observation of a large matrix. The results show some interesting phenomena that are quite different from other high-dimensional problems studied in the literature.

Coffee Break, Biosciences Atrium

Parallel Sessions A
10:30am - 12:10pm, Saturday, August 10th

Session 2: ***New Development in the Analysis of Complex Structured Data***
Organizer: Wenqing He, University of Western Ontario
Chair: Dongsheng Tu
Room: B-1102, Time: 10:30AM-12:10PM

(1) Zhanfeng Wang, University of Science and Technology of China
Title: *Nonparametric Random Effects Functional Regression Model using Gaussian Process Priors*

Abstract: For functional regression model with functional responses, we propose a nonparametric random effects model using Gaussian process priors. It can model the heterogeneity nonlinearly and model the covariance structure

nonparametrically, enabling longitudinal study of functional data. The model also has a flexible form of mean structure. We develop a flexible and accurate procedure to estimate unknown parameters and to calculate random effects nonparametrically using penalized least squares combining with a MAP (maximum a posterior). This results in a more accurate prediction. The statistical theory, including information consistency, has been discussed. Simulation studies and two real data examples show that the proposed method performs well.

(2) Xingqiu Zhao, The Hong Kong Polytechnic University
Title: *Semiparametric Inference for the Functional Cox Model*

Abstract: This article studies penalized semiparametric maximum partial likelihood estimation and hypothesis testing for the functional Cox model in analyzing right-censored data with both functional and scalar predictors. Deriving the asymptotic joint distribution of finite-dimensional and infinite-dimensional estimators is a very challenging theoretical problem due to the complexity of semiparametric models. For the problem, we construct the Sobolev space equipped with a special inner product and discover a new joint Bahadur representation of estimators of the unknown slope function and coefficients. Using this key tool, we establish the asymptotic joint normality of the proposed estimators and the weak convergence of the estimated slope function, and then construct local and global confidence intervals for an unknown slope function. Furthermore, we study a penalized partial likelihood ratio test, show that the test statistic enjoys the Wilks phenomenon, and also verify the optimality of the test. The theoretical results are examined through simulation studies, and a right-censored data example from the Improving Care of Acute Lung Injury Patients study is provided for illustration.

(3) Xin Liu, Shanghai University of Finance and Economics
Title: *Ensembling Imbalanced-Spatial-Structured Support Vector Machine*

Abstract: The Support Vector Machine (SVM) and its extensions have been widely used in various areas due to its great prediction capability. However, research gaps still remain. In particular, these methods cannot effectively handle imbalanced data with spatial association which commonly arise from many studies such as cancer imaging studies. In this paper, we propose the ensembling imbalanced-spatial-structured support vector machine (EISS-SVM) method which is useful for both balanced and imbalanced data. Not only does the proposed method accommodate the association between the response and the covariates, but also it accounts for the spatial correlation existing in the data. Our EISS-SVM classifier offers a flexible classification tool that embraces the usual SVM as a special case. The proposed method outperforms the competing classifiers, which is demonstrated by simulation studies. The good performance of the proposed method is further confirmed by the application to the real imaging data arising from an ongoing prostate cancer research conducted at the University of Western Ontario, Canada.

Session 3: ***Statistical Inference in Mixture and Other Nonregular Models***

Organizer: Pengfei Li, University of Waterloo
Chair: Chunlin Wang
Room: B-1120, Time: 10:30AM-12:10PM

(1) Pengfei Li, University of Waterloo
Title: *Full Likelihood Inference for Abundance from Capture-Recapture Data*
Abstract: Capture-recapture experiments are widely used to collect data needed to estimate the abundance of a closed population. To account for heterogeneity in the capture probabilities, Huggins (1989) and Alho (1990) proposed a semiparametric model in which the capture probabilities are modelled parametrically and the distribution of individual characteristics is left unspecified. A conditional likelihood method was then proposed to obtain point estimates and Wald-type confidence intervals for the abundance. Empirical studies show that the small-sample distribution of the maximum conditional likelihood estimator is strongly skewed to the right, which may produce Wald-type confidence intervals with lower limits that are less than the number of captured individuals or even negative. In this talk, we present a full empirical likelihood approach based on this model. We show that the null distribution of the empirical likelihood ratio for the abundance is asymptotically chi-square with one degree of freedom, and the maximum empirical likelihood estimator achieves semiparametric efficiency. Simulation studies show that the empirical-likelihood-based method is superior to the conditional-likelihood-based method: its confidence interval has much better coverage, and the maximum empirical likelihood estimator has a smaller mean square error. We analyze three data sets to illustrate its advantages. This talk is based on the joint works with Yukun Liu, Jing Qin, and Yang Liu.

(2) Abbas Khalili, McGill University
Title: *Feature Selection and Estimation in Finite Mixture of Varying Coefficient Regression Models*
Abstract: There is a vast literature on the problems of feature selection and estimation in linear and generalized linear varying coefficient regression models. Motivated by a real data analysis in an attempt to identify risky SNPs in Osteoporosis disease, in this talk we introduce a new method for the aforementioned problems in the context of finite mixture of varying coefficient regression models. We discuss large sample properties of the proposed method and evaluate its finite sample performance via extensive simulations. We then present our real data analyze.

(3) Jiahua Chen, University of British Columbia
Title: *Learning Finite Mixture Models by Minimum Wasserstein Distance Estimator*
Abstract: When a population exhibits a level of heterogeneity, finite mixture models provide an easy interpretation: the population is made of several homogeneous subpopulations all from a parametric distribution family. As early as in 1894, Pearson used a two-component Gaussian mixture to fit a crab data set, suggesting the existence of two subspecies. Pearson used the method of moments

likely for the ease of numerical computation. Contemporary practice in statistics favours the learning by maximum likelihood for statistical efficiency and the convenient EM-algorithm. The maximum likelihood estimator (MLE) searches for a distribution in the assumed distribution family that attains the minimum Kullback-Leibler divergence from the empirical distribution. Such minimum distance principle can be applied to learn mixtures based on any distances between two distributions. In the machine learning community, the Wasserstein distance has drawn increased attention for its intuitive geometric interpretations and it is successfully employed in many new applications. We study the minimum Wasserstein distance estimator for learning finite Gaussian mixtures. We establish its statistical consistency and demonstrate its superior performances in some applications compared with a penalized version of MLE as the MLE is not well defined for finite Gaussian mixtures.

(4) Paul Marriott, University of Waterloo
Title: *The Geometry of Finite Mixture and Other Non-Regular Model*

Abstract: This talk looks at a range of geometric approaches that throw light on non-regular behaviour in mixture and other models.The classical Information Geometric (IG) approach was based on putting differential geometric structures on models which have smooth manifold structure. Many Important models, however, include boundaries, closures and singularities for which richer geometric tools need to be used

Session 4: ***Functional Data Analysis***
Organizer and Chair: Peijun Sang, University of Waterloo
Room: K-106, Time: 10:30AM-12:10PM

(1) Pang Du, Virginia Tech
Title: *A Two-Sample Test for Spectral Data*

Abstract: In many medical procedures, a monitoring of procedure status can be achieved through assessment of spectral data collected over time. In this talk, we propose a two-sample test for spectral data collected at two time points. Each spectrum is represented by wavelets with thresholding. The test statistic is adapted from the one in Horvath et al (2013) proposed for continuous data. We shall evaluate the test performance with simulations and show its use in the monitoring of dialysis with Raman spectra of waste dialysate samples.

(2) Gregory Rice, University of Waterloo
Title: *Goodness-of-Fit Tests for Functional GARCH Models*

Abstract: Building upon some recent advances on white noise tests for functional data, we construct asymptotically valid Goodness-of-Fit tests for functional time series models of conditional heteroscedasticity. These new tests are illustrated with applications to risk management with high-frequency intra-day return data.

(3) Bing Li, Pennsylvania State University
Title: *On Post Dimension Reduction Statistical Inference*

Abstract: The methodologies of sufficient dimension reduction have undergone extensive developments in the past three decades. However, there has been a lack of systematic and rigorous development of post dimension reduction inference, which has seriously hindered its applications. The current common practice is to treat the estimated sufficient predictors as the true predictors and use them as the starting point of the downstream statistical inference. However, this naive inference approach would grossly overestimate the confidence level of an interval, or the power of a test, leading to the distorted results. In this paper, we develop a general and comprehensive framework of post dimension reduction inference, which can accommodate any dimension reduction method and model building method, as long as their corresponding influence functions are available. Within this general framework, we derive the influence functions and present the explicit post reduction formulas for the combinations of numerous dimension reduction and model building methods. We then develop post reduction inference methods for both confidence interval and hypothesis testing. We investigate the finite-sample performance of our procedures by simulations and a real data analysis.

Session 5: ***Statistical Methods for Complex Data***
Organizer and Chair: Juxin Liu, University of Saskatchewan
Room: K-107, Time: 10:30AM-12:10PM

(1) Xuejing Meng, Hubei University of Economics
Title: *Hierarchical Bayesian Models and Its Application in Ecological Environment*

Abstract: In environmetrics, many researchers are interested in models and methods for the time evolution of certain variables. These variables (e.g. wind, temperature, moisture, population) are always observed with spatial and spatio-temporal features. In this research, we model the spread of species using advection-diffusion-reaction partial differential equation (PDE) instead of diffusion-reaction PDE within a hierarchical Bayesian model. The model takes into account the existence of possible trend and gives the diffusion process of species under this trend. We model the process as a Poisson response with the trend coefficients and spatially varying diffusion coefficients as well as a damping term using an advection-diffusion-reaction PDE that realistically mimics the species spread process. We demonstrate a simulation as an example to illustrate our results.

(2) Alexander de Leon, University of Calgary
Title: *Classification and Prediction with Mixed Measurements Subject to Misclassification*

Abstract: Identification of underlying latent classes (or subpopulations) to account for unobserved heterogeneity in the population is a challenging statistical problem, mainly because no explicit information about the latent classes is available (e.g., no gold standard exists for diagnosing a disease), or access to such information, even when available, may be costly and time consuming. Although

latent class analysis via finite mixture models is often used successfully in applications, it may fail with data that poorly separate the latent classes (e.g., a disease biomarker with poor diagnostic accuracy). Borrowing strength from readily accessible auxiliary information (e.g., expert opinions, questionnaires, screening tests) that can serve as surrogate classifiers, even if imperfect, may yield improved results in such settings. We develop in this paper a joint modelling approach that combines continuous and categorical data from multiple sources, including imperfect surrogate classifiers subject to misclassification, in order to better identify and separate the latent classes for more accurate classification and prediction. We outline maximum likelihood estimation for the joint model using the EM algorithm, and we show empirically via simulations that our methodology yields better estimates of the underlying latent class distributions than those obtained by ignoring the auxiliary information, while providing joint assessments of the imperfect surrogate classifiers. We use real diagnostic data on dry eye disease, for which no gold standard is available, to illustrate our methodology.

(3) Mary E. Thompson, University of Waterloo
Title: *Spatial Multilevel Modeling in the Galveston Bay Recovery Study Survey*
Abstract: The Galveston Bay Recovery Study survey was a longitudinal survey of residents of Galveston County and Chambers County in the aftermath of Hurricane Ike, which made landfall in the Galveston Bay, Texas area in September 13, 2008 and caused widespread damage. One of the primary objectives was to examine the extent of symptoms of Post-Traumatic Stress Disorder (PTSD) in the resident population over the following year and a half. Wave 1 of the survey was conducted between November 17, 2008 and March 24, 2009 after the hurricane; Waves 2 and 3 consisted respectively of two month and one year follow-ups. With the use of a stratified, 3-stage sampling design, data were collected from 658 residents. The first stage of sampling within strata was the selection of clusters, or "area segments. The objective of our analysis is to model the course of the repeated PTSD measures as a function of individual characteristics and area segment, and to examine the analytical and visual evidence for spatial correlation of the area segment effect. To incorporate design information, we describe a multilevel analysis that uses the composite likelihood approach of Rao, Verret and Hidiroglou (Survey Methodology, 2013) and Yi, Rao and Li (Statistica Sinica, 2016). We compare this with a Bayesian multilevel analysis.

(4) Lang Wu, University of British Columbia
Title: *Survival Models with Truncated Time-Varying Covariates*
Abstract: In the analysis of longitudinal data and survival data, joint models are useful since the longitudinal data and survival data are often strongly associated. In practice, the longitudinal data can be highly complicated, such as being truncated and mixed types of discrete and continuous. In this talk, I will discuss some recent work to address these data complications in joint models. Another challenge for joint models is computation, since the likelihoods of joint models often involve high-dimensional and intractable integrations. I will also

discuss a computationally efficient approximate likelihood method. The models and methods will be applied to the analysis of a recent HIV vaccine dataset.

Session 6: ***Recent Development in the Statistical Modeling of Complex Data***
Organizer: Dongsheng Tu, Queen's University
Chair: Wenqing He
Room: B-2109, Time: 10:30AM-12:10PM

(1) Xinyi Ge, Queen's University
Title: *A Threshold Linear Mixed Model for Identification of Treatment-Sensitive Subsets in a Clinical Trial Based on Longitudinal Outcomes and a Continuous Covariate*

Abstract: Identification of a subset of patients who may be sensitive to a specific treatment is an important problem in clinical trials. In this paper, we consider the case where the treatment effect is measured by longitudinal outcomes, such as quality of life scores assessed over the duration of the clinical trial, and the subset is determined by a continuous covariate, such as a biomarker. A single-index threshold linear mixed model is introduced, and a smoothing maximum likelihood method is proposed to obtain the estimation of the parameters in the model. Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is employed to maximize the proposed smoothing likelihood function. The proposed procedure is evaluated through simulation studies and application to the analysis of data from a randomized clinical trial on patients with advanced colorectal cancer.

(2) Yi Niu, Dalian University of Technology
Title: *Variable Selection for Marginal Mixture Cure Models via Penalized GEE*

Abstract: Correlated failure time data with long-term survivors are common in biomedical and clinical research. When there are a large number of predictors available, variable selection is always an important issue when modeling such data with a survival model. Existing variable selection methods for survival data with a cure fraction are limited and based on the penalized likelihood under the EM algorithm. In this paper, we consider a marginal proportional hazards mixture cure model and propose a penalized generalized estimating equation approach to select important variables and to estimate regression coefficients simultaneously in the marginal model. The proposed method explicitly models the correlation structure within clusters or correlated variables by using a prespecified working correlation matrix. We conduct an extensive simulation study to assess the performance of the proposed method. We illustrate the proposed approach on data from a smoking cessation study.

(3) Dongdong Li, Harvard Medical School
Title: *Comparison Between Breast Cancer Survivors and the General Population in Age at Cardiovascular Disease with Observations Subject to Informative Censoring*

Abstract: In an attempt to compare the age at cardiovascular disease between breast cancer survivors and the general population in a large cancer survivorship

study, we proposed an approach to deal with informative censoring caused by a terminating event, which is often encountered in observational studies. We model the event time jointly with the terminating event by an Archimedean copula function. This allows us to account for informative censoring, and it yields a consistent estimator of the marginal and conditional survivor function in the semicompeting risks data setting. We propose an easy-to-implement inference procedure using pseudolikelihood approach. The proposed approach was applied to the motivating breast cancer study, but the approach has much broader application in any semi-competing risk setting when the marginal or conditional survivor function is of interest.

Session 7: ***Advances in Analysis of Complex Data***
Organizer and Chair: Changbao Wu, University of Waterloo
Room: B-2111, Time: 10:30AM-12:10PM

(1) Song Cai, Carleton University
Title: *A Simple Variable Selection Method Under Two-Fold Subarea Models in Small Area Estimation*

Abstract: We aim to develop a simple and effective method for variable selection in a two-fold subarea-level model for small are estimation. The two-fold model under consideration consists of a sampling model and a linking model. Due to the complex structure of the two-fold model, existing variable selection methods for regression models are not directly applicable. We propose to first transform the linking model into regular regression model with IID errors. We then transform the sampling model accordingly and use the observed data and the transformed sampling model to approximate an information criterion such as AIC or BIC for the transformed linking model. Variable selection procedures then can be carried out using the approximated information criterion. Simulation studies show that the proposed method outperforms a few simple competitors especially when the variance of the area-level random effect in the linking model is substantial.

(2) Hon Yiu So, University of Waterloo
Title: *COrrelated and Misclassified Binary Observations in Complex Surveys (COMBOS)*

Abstract: Misclassification has long been a common problem in medical surveys and public health data. One way to handle misclassification in clustered or longitudinal data is to incorporate the misclassification model in a generalized estimating equations (GEE) approach. However, in the existing literature, most such methods are developed in a non-survey setting. To apply the GEE in a complex survey design appropriately, we propose a pseudo GEE method for the analysis of survey data. In this research, we have focused on misclassification in clustered or longitudinal outcomes and developed a marginal analysis method to handle binary responses which are subject to misclassification using pseudo GEE. The proposed methodology has various attractive features including simultaneous inference for both marginal means and association parameters, as well as robustness to model misspecification.

(3) Changbao Wu, University of Waterloo
Title: *Some Theoretical and Practical Aspects of Empirical Likelihood for Complex Survey Data*

Abstract: This paper provides an overview on two parallel approaches to design-based inference with complex survey data: the pseudo empirical likelihood methods and the sample empirical likelihood methods. The general framework covers parameters defined through smooth or non-differentiable estimating functions for analytic use of survey data as well as descriptive finite population parameters, and the theory focuses on point estimation, hypothesis tests and variable selection under an arbitrary sampling design. Major practical issues for the implementation of the methods, including computational algorithms, are briefly discussed. Results from simulation studies to compare the finite sample performances of the two approaches are presented.

Session 8: ***New Advances in Nonparametric Statistics for Big Data***
Organizer: Zhengwu Zhang, University of Rochester
Chair: Yanglei Song
Room: K-104, Time: 10:30AM-12:10PM

(1) Yafei Wang, Beijing University of Technology
Title: *Estimation of Functional Linear Model with Incomplete Observations*

Abstract: In this paper, the estimation of functional regression model is considered under two observation patterns for functional curves, and functional linear model as a representative is used for analysis. In both cases, we combine the classical functional principal component analysis method into partially observed functional data framework to get an estimator of the slope function, and then give the convergence rate of the proposed estimator. Finally, simulation studies and real data analysis are presented to illustrate the efficiency of proposed method. And we also use simulation studies to investigate finite sample behavior of proposed method. A further exploration is also conducted to analyze the superiority of imputation method in the view of predicting the missing part of functional data.

(2) HaiYing Wang, University of Connecticut
Title: *Optimal Subsampling: Sampling with Replacement vs Poisson Sampling*

Abstract: Faced with massive data, subsampling is a commonly used technique to improve computational efficiency, and using nonuniform subsampling probabilities is an effective approach to improve estimation efficiency. In the context of maximizing a general target function, this paper derives optimal subsampling probabilities for both subsampling with replacement and Poisson subsampling. The optimal subsampling probabilities minimize functions of the subsampling approximation variances in order to improve the estimation efficiency. Furthermore, they provide deep insights on the theoretical similarities and differences between subsampling with replacement and Poisson subsampling. Practically implementable algorithms are proposed based on the optimal structural results, which are evaluated by both theoretical and empirical analysis.

(3) Xing Qiu, University of Rochester
Title: *iSPREAD: A Personalized Inference Pipeline for Medical Images*
Abstract: Subject-specific longitudinal study is vital for investigation of pathological changes of lesions and disease evolution. In recent years, we developed the iSPREAD (improved Spatial Regression Analysis of Diffusion tensor imaging) medical image analysis pipeline for this purpose. iSPREAD is a non-parametric permutation-based statistical framework that combines spatial regression and resampling techniques to achieve effective detection of localized longitudinal diffusion changes within the whole brain at individual level without a priori hypotheses. It uses an anisotropic spatial smoothing technique based on the celebrated Perona-Malik diffusion equation that captures the anisotropic and inhomogeneous nature of human brain. It can be used with several voxel-level summary statistics to detect linear or nonlinear temporal trends. We also developed several global summary statistics based on functional data analysis to quantify abnormalities at the whole brain level. Our method achieves high sensitivity and specificity in extensive simulations and real data applications.

(4) Yize Zhao, Yale University
Title: *Bayesian Subtyping Analysis with Application on Multi-Scale Molecular Data*
Abstract: Investigating cancer genome based on multi-type omics data and how it advances personalized medicine is a global medical issue. Though some of the existing clustering methods are capable to character certain degree of concordant and heterogeneity across data types, none of them has incorporated biological network information within and across molecular modalities under cancer subtype discovery. Meanwhile, it is biologically important to identify the core set of biomarkers that are informative to the similarity among samples in each subtype. In this work, with the goal to achieve cancer subtype discovery, we construct a unified clustering model with an incorporation of biological network within and across different molecular data types and simultaneously identifying informative molecular biomarkers for each subtype. Different from existing parametric methods, we adopt a nonparametric approach based on Bayesian Dirichlet process mixture (DPM) models, which is more adaptable to different data types, robust to statistical assumptions and has no constrain on the number of clusters. The performance of the proposed model has been assessed by extensive simulation studies and GCTA.

Provided Lunch, Leonard Hall (150 Queens Crescent)

Parallel Sessions B
1:50pm - 3:30pm, Saturday, August 10th

Session 9: ***Innovative Computing Methods in Biostatistics***
Organizer and Chair: Yeying Zhu, University of Waterloo
Room: B-1102, Time: 1:50PM-3:30PM

(1) Wenqing He, University of Western Ontario
Title: *Dynamic Tilted Current Correlation for High Dimensional Variable Screening*

Abstract: High dimensionality brings distorted result and computational burden to statistical analysis of this type of data. In the ultra-high dimensional setting, the theory of Sure Independence Screening was introduced to significantly reduce the dimensionality of variables to a moderate scale below the sample size and to preserve the true model with probability tending to 1. The outstanding performance of SIS stimulates the researchers to investigate more and better methods on high dimensional variable screening. The performance of SIS depends on the marginal correlation which is unreliable when the dimension is high. In reality, the importance of the variables cannot be easily ranked by their marginal correlations when there are high correlations among predictor variables. Due to the dimensionality, important predictors may have small marginal correlations with the response, while unimportant predictors may be highly correlated with the response variable due to the associated or spurious correlation with the important predictors. To remove those unimportant predictors and keep real important predictors, we propose a new estimator for the correlation between the response and variables in high dimensional settings, and a new screening technique termed dynamic tilted current correlation screening (DTCCS) is employed to do the variable screening. The new method reduces high spurious correlation among predictor variables in a data-driven fashion. We show that DTCCS is able to discover all relevant predictor variables within a finite number of steps when the dimensional of the true model is finite. DTCCS's sure screening property, consistency property and computational complexity are justified theoretically and numerically. To confirm the effectiveness of the proposed methods, extensive simulation studies are conducted and a real data analysis is invoked to illustrate the screening and model selection procedure of DTCCS.

(2) Dehan Kong, University of Toronto
Title: *Causal Inference in Imaging Genetics*

Abstract: Understanding the workings of human brains and their connection with dementia behavior is a central goal in medical studies. In this talk, I will introduce a new method to identify the causal relationship between hippocampal atrophy and dementia behavior in Alzheimers disease. We consider a 2D hippocampal surface exposure and develop a causal inference procedure which can account for high dimensional potential genetic confounders. We examine the performance of our method using a large-scale imaging genetic dataset from the Alzheimers Disease Neuroimaging Initiative study.

(3) Liqun Diao, University of Waterloo
Title: *Regression Trees for Interval-Censored Data*

Abstract: Tree-based methods are useful tools to identify risk groups and conduct prediction by employing recursive partitioning to separate patients into different risk groups. Existing loss based" recursive partitioning procedures wouldn't be used in the presence of interval censoring. We propose a new regression

tree algorithm through censoring unbiased transformations of loss estimators and realize it by making censoring unbiased transformations for the response variable. Simulations and a data analysis demonstrate a strong performance of the proposed regression tree algorithm compared to previously used methods.

(4) Yingwei Peng, Queen's University
Title: *A Support Vector Machine Based Semiparametric Mixture Cure Model*
Abstract: The mixture cure model is an extension of standard survival models to analyze survival data with a cured fraction. Many developments in recent years focus on the latency part of the model to allow more flexible modeling strategies for the distribution of uncured subjects, and fewer studies focus on the incidence part to model the probability of being uncured/cured. We propose a new mixture cure model that employs the support vector machine (SVM) to model the covariate effects in the incidence part of the cure model. The new model inherits the features of the SVM to provide a flexible model to assess the effects of covariates on the incidence. Unlike the existing nonparametric approaches for the incidence part, the SVM method also allows for potentially high-dimensional covariates in the incidence part. Semiparametric models are also allowed in the latency part of the proposed model. We develop an estimation method to estimate the cure model and conduct a simulation study to show that the proposed model outperforms existing cure models, particularly in incidence estimation. An illustrative example using data from leukemia patients is given.

Session 10: ***Handling Complex Featured Data***
Organizer: Grace Y. Yi and Weixin Yao, University of Waterloo and University of California
Chair: Grace Yi
Room: B-1120, Time: 1:50PM-3:30PM

(1) Suojin Wang, Texas A&M University
Title: *Oracally Efficient Estimation and Simultaneous Inference in Partially Linear Single-Index Models for Longitudinal Data*
Abstract: In this presentation, we discuss oracally efficient estimation and asymptotically accurate simultaneous confidence band (SCB) for the nonparametric link function in the partially linear single-index models for longitudinal data. The proposed procedure works for possibly unbalanced longitudinal data under general conditions. The link function estimator is shown to be oracally efficient in the sense that it is asymptotically equivalent in the order of one over root n to that with all true values of the parameters being known oracally. Furthermore, the asymptotic distribution of the maximal deviation between the estimator and the true link function is provided, and hence an SCB for the link function is constructed. Finite sample simulation studies are carried out which support our asymptotic theory. The proposed SCB is applied to analyze a CD4 data set.

(2) Bing Han, RAND Corporation
Title: *Synthetic Estimation for Causal Inference*

Abstract: A major difficulty for practitioners of Rubin Causal Model (RCM) is to choose from the large number of available estimators. Numerical and empirical studies showed that the conclusions across methods can be highly variable and that many distinct approaches have been recommended by different authors. To address this challenge, we propose a synthetic estimator based on the classic linear model averaging theory. The synthetic estimator is a convex combination of multiple candidate estimators with the goal of achieving an optimal mean squared error. We discuss the properties and computational details of the proposed synthetic estimator. We demonstrate by numerical examples that the synthetic estimator has a robust performance across various data generating strategies, while any single candidate estimators performance is usually volatile.

(3) Yanglei Song, Queen's University
Title: *Distribution Approximation and Bootstrap for Suprema of Incomplete U-Processes*

Abstract: This paper studies the non-asymptotic inference for the supremum of an incomplete, non-degenerate U-process. The process is indexed by a function class of order r, whose complexity possibly increases with the sample size n. For each function, its corresponding U-statistic involves the average of $O(n^r)$ numbers, which is prohibitively demanding even for moderate r. Thus we study its incomplete version, where each subsample of size r is included in the average with a very small probability. We first approximate the supremum of such incomplete U-process by that of an appropriate Gaussian process in the Kolmogorov distance and then propose valid bootstrap methods to address the practical issue of unknown covariance function. Finally, we discuss its application in testing the qualitative features of nonparametric functions.

(4) Zhezhen Jin, Columbia University
Title: *Modeling Treatment Outcomes in Transplant*

Abstract: Transplant is a treatment option for diseases with damaged or destroyed bone marrow and for diseases with terminal and irreversible organ failures. Analysis of transplant data is to identify medical treatments and procedures for best clinical outcomes. The clinical outcomes are often characterized by events, such as graft failure, infections and death. In this talk, I will use pediatric allogeneic hematopoietic cell transplantation (alloHCT) and kidney transplantation to present commonly used statistical methods and to illustrate the challenges and issues in data analysis.

Session 11: ***Recent Developments in Statistical Methods for Skewed Longitudinal Data and Time Series Data***
Organizer and Chair: Ying Zhang, Acadia University
Room: K-106, Time: 1:50PM-3:30PM

(1) Toby Kenney, Dalhousie University
Title: *On Consistency of Approaches to Ranking Problems.*

Abstract: The ranking problem is to order a collection of units by some unobserved parameter, based on observations from the associated distribution. This problem arises naturally in a number of contexts, such as business, where we may want to rank potential projects by profitability; or science, where we may want to rank variables potentially associated with some trait by the strength of the association. There have been a number of methods for this problem in the literature. The majority are empirical Bayesian, and optimise the expectation of a chosen loss function for a given prior distribution. Previous research has focused exclusively on choice of loss function. In this talk we compare the consistency of these methods in high dimensions (i.e. when the number of units to be ranked exceeds the number of observations from each unit). We consider the effect of both choice of prior and loss function, including consideration of cases where these are misspecified. We show that there is little difference in the consistency of methods with only very light-tailed prior distributions being less consistent.

(2) Zhou Zhou, University of Toronto
Title: *Frequency Detection and Change Point Estimation in Complexly Oscillating Time Series*
Abstract: We consider the problem of detecting evolutionary oscillating frequencies of a signal when it is contaminated by non-stationary noises with complex time-varying data generating mechanism. A high-dimensional progressive periodogram test is used to accurately detect all oscillating frequencies. A further local changepoint detection algorithm is applied in the frequency domain to detect the locations where the oscillation pattern changes. A Gaussian approximation scheme to cumulative sums of high dimensional non-stationary time series is established, which is of independent interest.

(3) Renjun Ma, University of New Brunswick, Canada
Title: *Tweedie Mixed Models for Skewed Longitudinal Data*
Abstract: Generalized linear mixed models have played an important role in the analysis of longitudinal data; however, traditional approaches have limited flexibility in accommodating skewness and complex correlation structures. In addition, the existing estimation approaches generally rely heavily on the specifications of random effects distributions; therefore, the corresponding inferences are sometimes sensitive to the choice of random effects distributions under certain circumstance. In this project, we incorporate serially dependent distribution-free random effects into Tweedie generalized linear models to accommodate a wide range of skewness and covariance structures for discrete and continuous longitudinal data. An optimal estimation of our model has been developed using the orthodox best linear unbiased predictors of random effects. Our approach unifies population-averaged and subject-specific inferences. Our method is illustrated through the analyses of patient-controlled analgesia data, Framingham cholesterol data and the basic symptoms inventory data.

(4) Hao Yu, University of Western Ontario

Title: *Text Ranking Based on Time Series and Entropy*

Abstract: Text ranking is to find a relationship among a collection of documents in terms of some special rules or criteria. Ranking order can be represented from the highest quality document to the lowest quality one. Different ranking rules may lead to different ranking orders. In this talk, we will show how a text can be digitized into a time series based on a selection of keywords. Then we will use an improved DTW method as well as Entropy to rank texts. Their performance will be demonstrated through a specific set of Amazon question/answer data.

Session 12: ***Recent Developments in Empirical Likelihood Method***
Organizer and Chair: Pengfei Li, University of Waterloo
Room: K-107, Time: 1:50PM-3:30PM

(1) Chunlin Wang, Xiamen University
Title: *A Semiparametric Homogeneity Test for Comparing Quality of Life Outcomes in Cancer Clinical Trials*

Abstract: This paper is concerned about the test for the difference in the distributions of multi-group proportional data, which is motivated by the problem of comparing the distributions of quality of life (QoL) outcomes among difference treatment groups in clinical trials. The proportional data, such as QoL outcomes assessed by answers to questions on a questionnaire, are bounded in a closed interval such as $[0, 1]$ with continuous observations in $(0, 1)$ and, in additional, excess observations taking the boundary values 0 and/or 1. Common statistical procedures used in practice, such as t- and rank-based tests, may not be very powerful since they ignore the specific feature of the proportional data. In this paper, we propose a three-component mixture model for the proportional data and a density ratio model for the distributions of continuous observations in $(0, 1)$. A semiparametric test statistic for the homogeneity of distributions of multi-group proportional data is derived based on the empirical likelihood ratio principle and shown asymptotically distributed as a chi-squared random variable under null hypothesis. A nonparametric bootstrap procedure is proposed to further improve the performance of the semiparametric test. Simulation studies are performed to evaluate the empirical type I error and power of the proposed test procedure and compare it with likelihood ratio tests under parametric distribution assumptions, rank-based Kruskal-Wallis test, and Wald-type test. The proposed test procedure is also applied to the analysis of QoL outcomes from a clinical trial on colorectal cancer that motivated our study.

(2) Yilin Chen, University of Waterloo
Title: *Estimation of Proportions with Non-Probability Survey Samples Using Pseudo-Empirical Likelihood Method*

Abstract: In survey questionnaires, binary responses such as, yes/no, agree/disagree, satisfied/not satisfied, are one of the most commonly used format. Collected binary data are then used to estimate proportions of the population who has certain characteristics. In this paper, we propose to estimate population proportions with samples from non-probability based surveys. We propose a pseudo-empirical

likelihood (PEL) inferential procedure, and show that resulting point estimator for a population proportion has desirable doubly robust property. Two methods of constructing PEL ratio confidence intervals for the population proportion are proposed, one is based on the limiting distribution of adjusted PEL ratio statistic and the other uses the bootstrap calibrated PEL. A simulation study shows that, when sample size is small, PEL ratio based confidence intervals have better coverage rate and more balanced tail error rate than the commonly-used wald-type confidence intervals.

(3) Meng Yuan, University of Waterloo
Title: *Semiparametric Empirical Likelihood Inference with General Estimating Equations*

Abstract: Density ratio model has been proved to be a powerful statistical tool to link the distributions of multiple related populations. In this talk, we consider the statistical inference under the DRM with additional auxiliary information which is expressed in terms of estimating equations. We study the asymptotic normality of the unknown parameters in the DRM and/or estimation equations and further establish the chi-squared limiting distribution for the empirical likelihood ratio of these parameters. We also propose an empirical likelihood ratio test for testing the validity of the auxiliary information. Simulation studies show that utilizing the auxiliary information will result in more efficient estimation method and more powerful statistical tests.

(4) Puying Zhao, Yunnan University
Title: *Generalized Empirical Likelihood Inferences for Nonsmooth Moment Functions with Nonignorable Missing Values*

Abstract: The main purpose of this study is to develop parameter identifiability and statistical inferences for a class of possibly over-identified nonsmooth moment functions with nonignorable missing data. Assuming a parametric model on the respondent probability, we propose a propensity score-based nonparametric imputation approach that uses an instrumental variable to address model identifiability in the presence of nonignorable missing data. A set of augmented inverse probability weighting moment functions is constructed as a basis for inferences performed using the generalized empirical likelihood method. Under some mild regularity conditions, we establish the large-sample properties of the resultant two-step generalized empirical likelihood estimators and generalized empirical likelihood ratio statistics for the case in which the propensity score is estimated parametrically using a correctly specified model. A derivative-free optimization method based on the simulated annealing algorithm is developed to implement the proposed methods. The methods are illustrated using simulations and an application to a data set on the serum-cholesterol levels of heart-attack patients.

Session 13: ***Statistical Inference with Constraints and Complexities***
Organizer: Lang Wu, University of British Columbia
Chair: Hongbin Zhang, City University of New York
Room: B-2109, Time: 1:50PM-3:30PM

(1) Sanjoy Sinha, Carleton University
Title: *Constrained Inference in Mixed Models for Clustered Data*
Abstract: Mixed models are commonly used for analyzing clustered data, including longitudinal data and repeated measurements. Unrestricted full maximum likelihood (ML) methods have been extensively studied in the literature for analyzing generalized, linear, and mixed models. However, constraints or parameter orderings may occur in practice, and in such cases, we can improve the efficiency of a statistical method by incorporating parameter constraints into the ML estimation and hypothesis testing. In this talk, I will discuss constrained inference with generalized linear mixed models (GLMMs) under linear inequality constraints. Methods will be assessed using both Monte Carlo simulations and actual survey data from a health study.

(2) Abdus Sattar, Case Western Reserve University
Title: *Joint Modeling of Longitudinal and Survival Data with Covariates Subject to Limit of Detection*
Abstract: We develop and study an innovative method for jointly modeling longitudinal response and time-to-event data with a covariate subject to a limit of detection. The joint model assumes a latent process based on random effects to describe the association between longitudinal and time-to-event data. We study the role of the association parameter on the regression parameters estimators. We model the longitudinal and survival outcomes using linear mixed-effects and Weibull frailty models, respectively. Because of the limit of detection, missing covariate (explanatory variable, x) values may lead to the non-ignorable missing, resulting in biased parameter estimates with poor coverage probabilities of the confidence interval. We define and estimate the probability of missing due to the limit of detection. Then we develop a novel joint density and hence the likelihood function that incorporates the effect of left-censored covariate. Monte Carlo simulations show that the estimators of the proposed method are approximately unbiased and provide expected coverage probabilities for both longitudinal and survival submodels parameters. We also present an application of the proposed method using a large clinical dataset of pneumonia patients obtained from the Genetic and Inflammatory Markers of Sepsis study.

(3) Michelle Xia, Northern Illinois University
Title: *Simultaneous Adjustment of Covariate Misclassification and Missingness in Regression Models*
Abstract: In this talk, we discuss simultaneous adjustment of misclassification and missingness of categorical covariates in regression models. In simulation studies, we demonstrate that including observations with missingness and/or multiple surrogates of the covariate help alleviate the efficiency loss caused by the misclassification. In addition, we study the efficacy of misclassifcation adjustment when the number of categories increases for the covariate under concern. The method is applied to adjust for misclassification and missingness in the self-reported drug use status from the Longitudinal Studies of HIV-Associated Lung Infections and Complications.

(4) Jianan Peng, Acadia University
Title: *A Generalization to Dykstras Algorithm for Restricted Least Squares Regression*

Abstract: A common problem in statistics is to minimize a least squares regression subject to certain restrictions. Dykstra (1983) proposed an iterative algorithm for the case that the restriction is a finite intersection of some closed convex cones and showed that the procedure converges correctly. In this talk, we extend his algorithm to a more general case that the restriction is a finite intersection of affine transformations of some closed convex cones. We will show that the algorithm converges to the desired solution as long as the feasible set is not empty. Some examples are presented

Session 14: ***New Developments of Statistical Methods for High-Dimensional Data and Complex Sampling Designs***
Organizer: Xuewen Lu, University of Calgary
Chair: Amy Wu
Room: B-2111, Time: 1:50PM-3:30PM

(1) Zhonglei Wang, Wang Yanan Institute for Studies in Economics and School of Economics, Xiamen University
Title: *Bootstrap Inference for the Finite Population Total under Complex Sampling Designs*

Abstract: Bootstrap is a useful tool for making statistical inference, but it may provide erroneous results under complex survey sampling. Most studies about bootstrap-based inference are developed under simple random sampling and stratified random sampling. In this paper, we propose a unified bootstrap method applicable to some complex sampling designs, including Poisson sampling and probability-proportional-to-size sampling. Two main features of the proposed bootstrap method are that studentization is used to make inference, and the finite population is bootstrapped based on a multinomial distribution by incorporating the sampling information. We show that the proposed bootstrap method is second-order accurate using the Edgeworth expansion. Two simulation studies are conducted to compare the proposed bootstrap method with the Wald-type method, which is widely used in survey sampling. Results show that the proposed bootstrap method is better in terms of coverage rate especially when sample size is limited.

(2) Qihua Wang, Chinese Academy of Sciences
Title: *Bias-Corrected Kullback-Leibler Distance Criterion Based Model Selection with Covariables Missing at Random*

Abstract: Let Y be the response variable, and (X, Z) the covariable vector. We consider the model selection problem for $f_{Y|X,Z}(y|x, z)$ with X missing at random, where $f_{Y|X,Z}(y|x, z)$ is the conditional probability function of Y given (X, Z). Two novel model selection criteria are suggested. One is called bias-corrected Kullback-Leibler distance (BCKL) criterion and another one is called empirical-likelihood-based bias-corrected Kullback-Leibler distance (ELBCKL)

criterion. Both the criteria specify a parametric model, which do not need to be correct, for $f_{X|Y,Z}(x|y,z)$, the conditional probability function of the missing covariates given the observed variables. It is shown, however, that the model selection by both the proposed criteria is consistent and that the population parameter estimators, corresponding to the selected model, are also consistent and asymptotically normal even if the parametric model for $f_{X|Y,Z}(x|y,z)$ is misspecified. This is a remarkable superiority of our proposed criteria to some existing model selection strategies. Extensive simulation studies are conducted to investigate the finite-sample performances of the proposed two criteria and a thorough comparison is made with some related model selection methods. The simulation results show that our proposals perform competitively especially when the conditional distribution of the missing covariates given the observed variables is misspecified. Supplementary materials for this article are available online.

(3) Yuehua Wu, York University
Title: *Simultaneous Multiple Change-Point Estimation*

Abstract: Simultaneous multiple change-point estimation Multiple change-point problems can be found in many areas of science and engineering. To estimate all change-points in a data sequence is of great importance. A statistical analysis without considering their existence may lead to an incorrect or improper conclusion. We will present some examples to illustrate the multiple change-point problem, and show the importance to include multiple change-points in data modelling. Owing to the rapid development in model selection methods, a multiple change-point estimation method can be built upon by converting a multiple change-point estimation problem into a variable selection problem via proper segmentation of a data sequence. We will discuss recent developments in multiple change-point estimation using this methodology and their applications in real problems.

(4) Hao Ding, York University
Title: *General M-Estimation in High-Dimensional Data Analysis*

Abstract: Regularized M-estimation is widely used in high-dimensional regression models. In this talk, a general theory of high-dimensional multivariate M-estimation is studied. We consider a general multivariate linear regression model under the regime where the dimension of regression parameters (p) increases with the sample size (n) and p/n tends to a positive constant. Asymptotic properties of the ridge-regularized high-dimensional multivariate regression M-estimate and un-regularized M-estimate are both established. The numerical simulations are presented to verify our theoretical results.

Session 15: ***Recent Advances of Statistical Methods for Emerging High-Dimensional Biomedical Applications***
Organizer and Chair: Yize Zhao, Yale University
Room: K-104, Time: 1:50PM-3:30PM

(1) Gen Li, Columbia University
Title: *Zero-Inflated Poisson Factor Model with Application to Microbiome Ab-*

solute Abundance Data

Abstract: Dimension reduction of high-dimensional microbiome data facilitates subsequent analysis such as regression and clustering. Most existing reduction methods cannot fully accommodate the special features of the data such as count-valued and excessive zero reads. We propose a zero-inflated Poisson factor analysis (ZIPFA) model in this article. The model assumes that microbiome absolute abundance data follow zero-inflated Poisson distributions with library size as offset and Poisson rates negatively related to the inflated zero occurrences. The latent parameters of the model form a low-rank matrix consisting of interpretable loadings and low-dimensional scores which can be used for further analyses. We develop an efficient and robust expectation-maximization (EM) algorithm for parameter estimation. We demonstrate the efficacy of the proposed method using comprehensive simulation studies. The application to the Oral Infections, Glucose Intolerance and Insulin Resistance Study (ORIGINS) provides valuable insights into the relation between subgingival microbiome and periodontal diseases.

(2) Yuping Zhang, University of Connecticut
Title: *Integrative Structural Learning of Heterogeneous Exponential Markov Random Fields*

Abstract: Motivated by the discovery of conditional dependence relationships embedded in high-dimensional heterogeneous correlated data, we develop a new generalized data integration framework to jointly learn multiple heterogeneous exponential Markov random fields. We establish an approximate likelihood inference problem, and propose an efficient algorithm. The practical merits of the proposed integrative structural learning method are demonstrated through simulations and real applications to multiple heterogeneous multi-view genomic data.

(3) Linglong Kong, University of Alberta
Title: *High-Dimensional Spatial Quantile Function-on-Scalar Regression in Neuroimaging Analysis*

Abstract: With modern technology development, functional data are often observed in various scientific fields. Quantile regression has become an important statistical methodology. In this talk, we cdevelops a novel spatial quantile function-on-scalar regression model, which studies the conditional spatial distribution of a high-dimensional functional response given scalar predictors. With the strength of both quantile regression and copula modeling, we are able to explicitly characterize the conditional distribution of the functional or image response on the whole spatial domain. Our method provides a comprehensive understanding of the effect of scalar covariates at different quantile levels and also gives a practical way to generate new images for given covariate values. Theoretically, we establish the minimax rates of convergence for estimating coeffi cient functions under both fixed and random designs. We further developan effi cient primal-dual algorithm to handle high-dimensional image data. Simulations and real data analysis are conducted to examine the finite-sample performance.

Joint work with Zhengwu Zhang, Xiao Wang and Hongtu Zhu.

Coffee Break, Biosciences Atrium

Parallel Sessions C
4:00pm - 5:40pm, Saturday, August 10th

Session 16: ***Challenges in the Analysis of Survival Data***
Organizer and Chair: Lang Wu, University of British Columbia
Room: B-1102, Time: 4:00PM-5:40PM

(1) Grace Yi, University of Waterloo
Title: *Analysis of Error-Prone Survival Data under Additive Harzards Models*
Abstract: Covariate measurement error has attracted extensive interest in survival analysis. Since Prentice (1982), a large number of inference methods have been developed to handle error-contaminated data, and most methods are addressed to proportional hazards models. In contrast to proportional hazards models, additive hazards models offer a flexible alternative to delineate survival data. However, there is relatively less research on measurement error effects under such models, although some authors investigated this problem. In this talk, I will discuss several methods to correct for measurement error effects under additive hazards models. These methods will be justified both theoretically and empirically.

(2) Qi Cui, Simon Fraser University
Title: *A Copula Model-Based Method for Regression Analysis of Dependent Current Status Data*
Abstract: This talk discusses regression analysis of current status data, which arise when the occurrence of the failure event of interest is observed only once or the occurrence time is either left- or right-censored. Many authors have investigated the problem, however, most of the existing methods are parametric or apply only to limited situations such that the failure time and the observation time have to be independent. In particular, Ma et al. (2015) proposed a copula-based procedure for the situation where the failure time and the observation time are allowed to be dependent but their association needs to be known. To address this restriction, we present a two-step estimation procedure that allows one to estimate the association parameter in addition to estimation of other unknown parameters. The asymptotic properties of the resulting estimators are established and a simulation study is conducted and suggests that the proposed method performs well for practical situations. Also an illustrative example is provided.

(3) Tao Wang, School of Mathematics, Yunnan Normal University
Title: *A Mixed Effects Model for Analyzing Complex AIDS Clinical Data*
Abstract: Chinese government started providing highly active antiretroviral therapy (HAART) free of charge since 2002. In such HIV clinical data set three

kinds of correlative AIDS progression markers, i.e. unbalanced longitudinal CD4 , CD8, and viral RNA data are included. These data are typically intermittently missing and informatively left censored. Up to date, few model-based curative effect evaluation research on Chinese HAART based on such data set have been published, and almost all of them only focused on modeling CD4 dynamic process without taking CD8, viral RNA data and the missing, censored mechanism into account, thus great bias and false result may be yielded. In this talk, we propose a parsimonious generalized mixed effects model to jointly inference the dynamic progression of CD4, CD8 and viral RNA data for such HIV clinical data sets. We characterize the CD4 CD8 and viral RNA dynamic progress, by taking the correlation among such three AIDS progression markers into account .Simulation studies and real data analysis demonstrate that our model performs well and is appropriate for evaluating HAART in practice.

(4) Hongbin Zhang, City University of New York
Title: *Joint Model of Accelerated Failure Time and Mechanistic Nonlinear Model for Censored Covariates, with Application in HIV/AIDS*

Abstract: For a time-to-event outcome with censored time-varying covariates, a joint Cox model with a linear mixed effects model is the standard modeling approach. In some applications such as AIDS studies, *mechanistic* nonlinear models are available for some covariate process such as viral load during anti-HIV treatments, derived from the underlying data-generation mechanisms and disease progression. Such a mechanistic nonlinear covariate model may provide better-predicted values when the covariates are left censored or mismeasured. When the focus is on the impact of the time-varying covariate process on the survival outcome, an accelerated failure time (AFT) model provides an excellent alternative to the Cox proportional hazard model since an AFT model is formulated to allow the influence of the outcome by the entire covariate process. In this article, we consider a nonlinear mixed effects model for the censored covariates in an AFT model, implemented using a Monte Carlo EM algorithm, under the framework of a joint model for simultaneous inference. We apply the joint model to an HIV/AIDS data to gain insights for assessing the association between viral load and immunological restoration during antiretroviral therapy. Simulation is conducted to compare model performance when the covariate model and the survival model are mis-specified.

Session 17: ***Multivariate Data Exploratory and Modeling Methods***
Organizer and Chair: Renjun Ma, University of New Brunswick
Room: K-1120, Time: 4:00PM-5:40PM

(1) Connie Stewart, University of New Brunswick Saint John
Title: *A Folded Model for Compositional Data Analysis*

Abstract: Vectors of non-negative components carrying only relative information, and often normalized to sum to one, are referred to as compositional data and their sample space is the simplex. Compositional data arise in many applications across a variety of disciplines such as ecology, geology and economics

to name a few. Aitchisons log-ratio methods developed in the 1980s have, since this time, been a popular approach for analyzing compositional data and have motivated much of the recent research in the area. In this talk, a flexible class of models for data defined on the simplex by means of a transformation from the simplex to Euclidean space is presented. The transformation uses a folding procedure and is an extension of the alpha-transformation that transforms compositional data from the simplex to a subset of Euclidean space. Our research suggests that parameter estimation via the EM algorithm is efficient and consistent. Further, our proposed model has the potential to provide an improved fit over the traditional log-ratio based models, and can be advantageous over a similar model without folding. Simulation study results and examples will be used to illustrate our findings.

(2) Jianhua Hu, Shanghai University of Finance and Economics
Title: *Response Variable Selection in Multivariate Linear Models*

Abstract: Multivariate linear regression analysis is an important technique for modeling the predictive relationships of multiple related responses on a set of common predictors. Numerous studies have been conducted on situations where responses are given and only predictors are subject to variable selection. In practice, however, the number of responses that are truly depend on the predictors is not known prior to data analysis, and thus responses are needed to variable selection, a topic on which limited research has been done. In this talk, we address the recent developments of the response selection problem. (1) We introduce a response best subset (RBS) model, an efficient estimation procedure for performing response selection and regression coefficient estimation via using a penalty function to response variables. RBS model is suitable to large scale responses and fixed predictors. (2) When covariance can be efficiently estimated, a two-stage RBS estimation and its oracle properties are investigated. (3) A simultaneous response and predictor selection (SRPS) model is developed when responses and predictors both are needed to be selected in multivariate linear regression analysis, which simultaneously investigates response selection, predictor selection and estimation to regression coefficients in the standard multivariate linear regression, group adaptive lasso and the response best subset selection contexts. The SRPS model with a sreening method as the first step can solve the variable selection problems with large scale numbers of responses and predictors.

(3) Hong Gu, Dalhousie University
Title: *Poisson Measurement Error Corrected PCA, with Application to Microbiome Data*

Abstract: We study the problem of computing a Principal Component Analysis of data affected by Poisson noise. We assume samples are drawn from independent Poisson distributions. We want to estimate principle components of a fixed transformation of the latent Poisson means. Our motivating example is microbiome data, though the methods apply to many other situations. We develop a emiparametric approach to correct the bias of variance estimators, both for untransformed and transformed (with particular attention to log-transformation)

Poisson means. Furthermore, we incorporate methods for correcting different exposure or sequencing depth in the data. In addition to identifying the principal components, we also address the non-trivial problem of computing the principal scores in this semiparametric framework. Most previous approaches tend to take a more parametric line. For example the Poisson-log-normal (PLN) model, approach. We compare our method with the PLN approach and find that our method is better at identifying the main principal components of the latent log-transformed Poisson means, and as a further major advantage, takes far less time to compute. Comparing methods on real data, we see that our method also appears to be more robust to outliers than the parametric method.

(4) Wilson Lu, Acadia University
Title: *Exploring Properties of an Algorithm on Unequal Probability Sampling without Replacement (UPSWOR)*

Abstract: We use both theoretical argument and simulation study to explore the main features of an easy-to-implement UPSWOR algorithm, including the calculation of inclusion probabilities and variance estimation.

Session 18: ***Recent Advances in Statistical Methodology***
Organizer and Chair: Amy Wu, York University
Room: K-106, Time: 4:00PM-5:40PM

(1) Yichen Qin, University of Cincinnati
Title: *Model Confidence Bounds for Model Selection*

Abstract: In this article, we introduce the concept of model confidence bounds (MCB) for variable selection. Similarly to the endpoints in the familiar confidence interval for parameter estimation, the MCB identifies two nested models (upper and lower confidence bound models) containing the true model at a given level of confidence. Instead of trusting a single selected model obtained from a given model selection method, the MCB proposes a group of nested models as candidates and the MCB's width and composition enable the practitioner to assess the overall model selection uncertainty. A new graphical tool — the model uncertainty curve (MUC) — is introduced to visualize the variability of model selection and to compare different model selection procedures. The MCB methodology is implemented by a fast bootstrap algorithm that is shown to yield the correct asymptotic coverage under rather general conditions. Our Monte Carlo simulations and real data examples confirm the validity and illustrate the advantages of the proposed method.

(2) Baisuo Jin, University of Science and Technology of China
Title: *Learning Block Structure in U-Statistic Based Large Correlation Matrices*

Abstract: Learning the block structure efficiently in a large correlation matrix is not a trivial work, because of the unknown number of groups and arbitrary order of variables. In this paper, we provide a novel approach using large dimensional random matrix theory that targets directly identifying the number of groups and the group structures of a large number of variables. The proposed approaches are

conceptually simple, efficient and can be easily implemented. The asymptotical properties are also established under mild conditions. The solid performance of our approach is demonstrated in extensive simulations and two data examples.

(3) Xiaoping Shi, Thompson Rivers University
Title: *Non-Euclidean Graph-Based Change-Point Test*
Abstract: The change-point detection has been carried out in terms of the Euclidean minimum spanning tree (MST) and shortest Hamiltonian path (SHP), with successful applications in the determination of authorship of a classic novel, the detection of change in a network over time, the detection of cell divisions, etc. However, these Euclidean graph-based tests may fail if a dataset contains random interferences. To solve this problem, we present a powerful non-Euclidean SHP-based test, which is consistent and distribution-free. The simulation shows that the test is more powerful than both Euclidean MST- and SHP-based tests and the non-Euclidean MST-based test. Its applicability in detecting both landing and departure times in video data of bees flower visits is illustrated.

(4) Yingying Fan, University of Southern California
Title: *Large-Scale Inference for Networks*
Abstract: Characterizing the exact asymptotic distributions of high-dimensional eigenvectors for large structured random matrices poses important challenges yet can provide useful insights into a range of applications. To this end, in this paper we introduce a general framework of asymptotic theory of eigenvectors (ATE) for large structured symmetric random matrices with heterogeneous variances, and establish the asymptotic properties of the spiked eigenvectors and eigenvalues for the scenario of the generalized Wigner matrix noise, where the mean matrix is assumed to have the low-rank structure. Under some mild regularity conditions, we provide the asymptotic expansions for the spiked eigenvalues and show that they are asymptotically normal after some normalization. For the spiked eigenvectors, we establish novel asymptotic expansions for the general linear combination and further show that it is asymptotically normal after some normalization, where the weight vector can be arbitrary. We also provide a more general asymptotic theory for the spiked eigenvectors using the bilinear form. Simulation studies verify the validity of our new theoretical results. Our family of models encompasses many popularly used ones such as the stochastic block models with or without overlapping communities for network analysis and the topic models for text analysis, and our general theory can be exploited for statistical inference in these large-scale applications. This talk is based on joint works with Jianqing Fan, Xiao Han and Jinchi Lv.

Session 19: ***Causal Inference for Complex Data Challenges***
Organizer and Chair: Dehan Kong, University of Toronto
Room: K-107, Time: 4:00PM-5:40PM

(1) Michael Wallace, University of Waterloo
Title: *Measurement Error and Precision Medicine*

Abstract: Precision - or personalized - medicine aims to improve patient outcomes by basing treatment decisions on subject-level information. Treatment decision rules - known as dynamic treatment regimes or adaptive treatment strategies - formalize the process of precision medicine. Identifying the best treatment regimes is a topic that has received widespread attention in the biostatistical literature over the past decade, with a large variety of novel methodologies for optimal treatment regime estimation being developed.

Measurement error describes when measurements differ from those we wished to observe. Despite its ubiquity in longitudinal health data, measurement error remains virtually untouched across the precision medicine literature. In this setting measurement error not only presents as a familiar obstacle to valid estimation and inference, but also threatens the very principle behind precision medicine itself, with treatment decisions often dependent on error-prone observations.

We present foundational work regarding the problem of measurement error in precision medicine in general, and treatment regime estimation in particular. We will discuss the consequences of it being ignored alongside some theoretical results concerning estimation.

(2) Linbo Wang, University of Toronto
Title: *Causal Inference with Confounders Missing not at Random*

Abstract: It is important to draw causal inference from observational studies, which, however, becomes challenging if the confounders have missing values. Generally, causal effects are not identifiable if the confounders are missing not at random. We propose a novel framework to nonparametrically identify causal effects with confounders subject to an outcome-independent missingness, that is, the missing data mechanism is independent of the outcome, given the treatment and possibly missing confounders. We then propose a nonparametric two-stage least squares estimator and a parametric estimator for causal effects.

(3) Yeying Zhu, University of Waterloo
Title: *Covariate Balancing by Optimizing Kernel Distance*

Abstract: An important goal in many causal effect procedures is to achieve balance in the covariates. Compared to other balance measures, such as absolute standardized mean difference (ASMD) and Kolmogorov Smirnov (KS) statistic, kernel distance is one of the best bias indicators in estimating the causal effect. That is, the balance metric based on kernel distance is shown to have the strongest correlation with the absolute bias in estimating the casual effect. We recommend using kernel distance to measure balance across different treatment groups and propose a new propensity score estimator by setting the kernel distance to be zero. The kernel distance constraints are solved by generalized method of moments. Simulation studies are conducted across different scenarios varying in the degree of nonlinearity in both the propensity score model and the outcome model. An application to data from the International Tobacco Control (ITC) policy evaluation project is provided.

Session 20: ***Advances in Longitudinal and Survival Studies***
Organizer and Chair: Wei Liu, York University
Room: B-2109, Time: 4:00PM-5:40PM

(1) Guohua Yan, University of New Brunswick
Title: *Poisson Mixed Models for Longitudinal Binomial Data with Random Number of Trials*

Abstract: Binomial data frequently occur in almost all disciplines. The traditional logistic regression focuses on the number of successes while treating the number of trials as constant. In some studies, however, the number of trials is random. Taking the randomness of the number of trials into consideration provides additional insight on the probability of success. For example, a larger proportion of success might due to an increased number of successes, or a decreased number of failures, or both. In this study, we introduce a Poisson mixed model for longitudinal binomial data, which models the number of successes and the number of failures simultaneously. As a result, the number of trials is regarded as random, even zero. The model also allows for flexible temporal correlation structures through the introduction of random effects.

(2) Xiaoming Lu, Memorial University
Title: *A Joint Model of Longitudinal Quantiles and Multiple-Censored Survival Times*

Abstract: We propose a new joint modelling method to combine right- and interval-censored survival times and longitudinal quantiles to capture the underlying effects among them. Three models are linked by manifest variable and latent random variables to capture the dependence between and within those three models. We assume an asymmetric Laplace distribution (ALD) for the longitudinal responses and apply Cox models for the survival process in order to accommodate the data structure. A Monte Carlo expectation maximization strategy is applied for the estimation, which can be directly used under any distributional assumptions for longitudinal measurements and random effects. Our proposed joint model is illustrated through simulation studies and applied to a sample of data from the PAQUID cohort aiming to study the disease of dementia among the elderly.

(3) Wei Liu, York University
Title: *Joint Likelihood Inference for Semiparametric Nonlinear Mixed-Effects Models with Covariate Measurement Errors and Change Points*

Abstract: We propose a semiparametric nonlinear mixed-effects response model incorporating measurement errors and missing data in time-varying covariates and change points. The covariate measurement error models and models for the times of change points on response trajectories are introduced for joint likeliood inference. We propose two approaches to obtain approximate maximum likelihood estimates of the joint model parameters simultaneously. We illustrate the proposed approaches to analyze a real dataset. A simulation study is conducted to evaluate these proposed approaches.

Session 21: ***Advances in Statistical Prediction with Real World Applications***
Organizer and Chair: Yan Yuan, University of Alberta
Room: B-2111, Time: 4:00PM-5:40PM

(1) Anis Sharafoddini, University of Waterloo
Title: *A New Insight into Missing Data in Predicting Outcomes of Intensive Care Unit Patients*

Abstract: The data missing from electronic health records in Intensive Care Units (ICUs) are substantial and unavoidable. However, missing data are not always random and can be informative of patient health status. In this talk, I will present our results about the informativeness of missing data in mortality prediction in ICUs. In particular, our results demonstrated that the presence or absence of lab measurements can be considered as a potential predictor of mortality in ICUs and can improve the mortality prediction performance when being used with actual measurements. Our initial case study shows promise for more in-depth analysis of missing data and its informativeness in prediction applications in ICUs.

(2) Joel Dubin, University of Waterloo
Title: *Some Prediction Problems in Health Research*

Abstract: Prediction of health outcomes is an important component for determining how to make recommendations and treat individuals. Regarding treatment, the intensive care unit (ICU) is a place where many such decisions are made. A primary goal for ICU patients is treating them to achieve positive outcomes (e.g., hospital discharge alive, improvement from in-hospital ailments, extended survival). A major analytical issue is the preponderance of information available at ICU entry (e.g., age, sex, co-morbidities, prescriptions, vital signs), and longitudinally (e.g., vital sign changes, dynamic renal function, in-ICU treatment). I will present a few interesting analytic challenges In predictive modeling that my collaborators and I have encountered from a large ICU database, and discuss a few remedies that we have investigated; these emphasize implementation of a patient similarity step in an effort to improve predictive accuracy.

(3) Zhihui (Amy) Liu, Princess Margaret Cancer Centre, University Health Network
Title: *Deriving Dosimetric Predictors of Toxicity for Radiation Therapy Planning*

Abstract: The dose-volume histogram (DVH), which relates the radiation dose to tissue volume receiving it, is a commonly used measure for radiation treatment planning and associating radiation dose to outcomes, including adverse effects such as normal tissue toxicity. Previously, radiation dose has been correlated with toxicity using biological models and individual dosimetric parameters such as quantiles of the dose or volume. However, in principle DVH is a functional exposure and can be modeled using functional data analysis tools. On the other hand, the complement of the DVH can be interpreted as a probability distribution. In this talk we review the functional approach and as an alternative, building on the representation as a probability distribution, we propose produc-

ing features of radiation dose using sample statistics. For example, in addition to quantiles, we can characterize the distribution through moment-based measures, such as mean, variance, skewness and kurtosis. We demonstrate the methods in predicting toxicity in anal cancer patients treated between 2008-2013 at Princess Margaret Cancer Centre. These results may be used clinically to optimize radiation therapy planning to limit normal tissue toxicity.

Session 22: ***Statistical Learning Methods for High-Dimensional Data***
Organizer and Chair: Longhai Li, University of Saskatchewan
Room: K-104, Time: 4:00PM-5:40PM

(1) Xuewen Lu, University of Calgary
Title: *Bi-Level Variable Selection Method for Multivariate Failure Time Data*
Abstract: In this work, a bi-level variable selection method is proposed for multivariate failure time data via an adaptive hierarchical lasso. In the regression setting for multivariate failure time data, the coefficients corresponding to the same prediction variable is treated as a natural group, then the variable selection is considered at the group level and individual level simultaneously. The proposed bi-level variable selection method can select a prediction variable at two different levels: the first level is the group level, where the prediction variable is important to all failure types; the second level is the individual level, where the prediction variable is only important to some failure types. Simulation results show that the proposed method outperforms the classical L1-norm penalty methods in the sense of removing unimportant variables in different failure types. An algorithm based on cycle coordinate descent (CCD) is provided to carry out the method. The asymptotic oracle properties of the proposed variable selection method are obtained. A generalized cross-validation (GCV) method is constructed for the tuning parameter selection and model performance is assessed based on model errors. Finally, the proposed method is applied to analyze a real-life data set.

(2) Wei Xu, University of Toronto
Title: *Statistical Learning in High-Dimensional Microbiome and Genomic Data*
Abstract: Recent genomic sequencing technologies have enabled researchers to unveil the wide variability of bacteria presented within different locations of the body, i.e. the microbiome, and how it relates to human diseases. However, our understanding of how microbiomes affect diseases is still unclear. It is necessary to better understand both environmental and host genetic factors impact the composition of the microbiome to improve diagnosis and disease management. Efficient statistical and bioinformatics tools are needed to overcome these knowledge gaps. In this talk, I will introduce some general characteristics of the human microbiome sequencing data that can be clustered into operational taxonomic units (OTUs) at each taxonomic level by next-generation sequencing. Several analytic approaches will be introduced to summarize and assess the single or multiple OTUs and genetic risk factors using different methods. Specific features of the microbiome sequencing data will be explored such as non-negative, highly skewed sequence counts with excess zeros, and clustered taxonomic structure.

(3) Karen Kopciuk, AHS, University of Calgary
Title: *Modelling Ordinal Response Data with High Dimensional, Mixed Covariate Types.*

Abstract: Ordinal response regression models, such as the proportional odds (PO) or partial proportional odds (PPO) models, are flexible models that can handle mixed covariate types including continuous, binary and categorical covariates. However, identifying covariates associated with an ordinal response in high dimensional settings is challenging for several reasons. For example, the choice of model appropriate for specific assumptions, the challenge of modelling categorical covariates (which may also be ordered) and the issue of identifying and selecting covariates associated with the ordinal response from amongst a large number of potential candidate variables are few of them. We explored these issues in a real data set where cancer stage (I-III) was the ordered response with more than 150 measured mixed covariate types (continuous, binary and ordered categorical) for 492 women with breast cancer. We employed two common feature selection methods, elastic net and random forests, in several ordinal regression models as well as in logistic and multinomial regression models. We also explored a sequential approach where individual covariates were evaluated for their marginal association as well as meeting the proportionality assumption but treated the ordered categorical covariates as continuous. Obviously, this approach can lead to biased estimates due to the mismodelling of the covariates. We are currently comparing these methods and models in simulation studies that explore the data features that lead to optimal covariate and model selection. In particular, identification of covariates that do not have the same importance for each response level (non-proportional) is a key feature of our study. Future work will focus on developing methods to simultaneously select features of any type and the appropriate model for the high dimensional, mixed covariate type setting.

(4) Maxime Turgeon, University of Manitoba
Title: *A Tracy-Widom Empirical Estimator for Valid P-Values with High-Dimensional Datasets*

Abstract: Recent technological advances in many domains including both genomics and brain imaging have led to an abundance of high-dimensional and correlated data being routinely collected. Classical multivariate approaches like Multivariate Analysis of Variance (MANOVA) and Canonical Correlation Analysis (CCA) can be used to study relationships between such multivariate datasets. In this work, I explain how valid p-values can be derived for these multivariate methods even in high dimensional datasets. The main contribution is an empirical estimator for the largest root distribution of a singular double Wishart problem; this general inferential framework underlies many common multivariate analysis approaches. Through simulations and data analysis, I demonstrate the performance of this approach.

5:45-6:30 ICSA-Canada Chapter Annual General Meeting (AGM) Room 1120 in Biosciences Complex.

6:00-8:45 Banquet
Isabel Bader Center, 390 King Street West

Plenary Talk II
9am-10am, Sunday, August 11th

Session 23: ***Keynote Speech 2***
Organizer: Liqun Wang, University of Manitoba
Chair: Wenqing He
Room: B-1102, Time: 9:00AM-10:00AM

(1) Runze Li, Penn State University
Title: *Linear Hypothesis Testing for Generalized Linear Models*
Abstract: This paper is concerned with testing linear hypotheses in high-dimensional generalized linear models. To deal with linear hypotheses, we first propose constrained partial regularization method and study its statistical properties. We further introduce an algorithm for solving regularization problems with folded-concave penalty functions and linear constraints. To test linear hypotheses, we propose a partial penalized likelihood ratio test, a partial penalized score test and a partial penalized Wald test. We show that the limiting null distributions of these three test statistics are chi-square distribution with the same degrees of freedom, and under local alternatives, they asymptotically follow non-central chi-square distributions with the same degrees of freedom and noncentral parameter, provided the number of parameters involved in the test hypothesis grows to infinity at a certain rate. Simulation studies are conducted to examine the finite sample performance of the proposed tests. Empirical analysis of a real data example is used to illustrate the proposed testing procedures.

Coffee Break, Biosciences Atrium

Parallel Sessions D
10:30am -12:10pm, Sunday, August 11th

Session 24: ***Predictive Methods in Insurance and Finance***
Organizer and Chair: Chengguo Weng, University of Waterloo
Room: B-1102, Time: 10:30AM-12:10PM

(1) Guojun Gan, University of Connecticut
Title: *Nested Stochastic Valuation of Large Variable Annuity Portfolios: Monte Carlo Simulation and Synthetic Datasets*
Abstract: Dynamic hedging has been adopted by many insurance companies to mitigate the financial risks associated with variable annuity guarantees. In order to simulate the performance of dynamic hedging for variable annuity products, insurance companies rely on nested stochastic projections, which is highly computationally intensive and often prohibitive for large variable annuity portfolios. Metamodeling techniques have recently been proposed to address the computational issues. However, it is difficult for researchers to obtain real datasets

from insurance companies to test metamodeling techniques and publish the results in academic journals. In this paper, we create synthetic datasets that can be used for the purpose of addressing the computational issues associated with the nested stochastic valuation of large variable annuity portfolios. The runtime used to create these synthetic datasets would be about 3 years if a single CPU were used. These datasets are readily available to researchers and practitioners so that they can focus on testing metamodeling techniques.

(2) Chengguo Weng, University of Waterloo
Title: *DSA Algorithms for Mortality Forecasting*
Abstract: It has been well recognized that borrowing information from populations with similar structural mortality patterns and trajectories is helpful to the mortality forecasting of a target population. One crucial step to gain an enhanced forecasting accuracy lies in the selection of a proper group of populations. To the best of our knowledge, however, no structured method exists to select the group flexibly and effectively. In our paper, we consider the mortality forecasting for a general target population from the Human Mortality Database (HMD). We develop an effective procedure to select a group of populations from the HMD to enhance the mortality prediction of the target population. Instead of grouping populations according to geographical or socioeconomic information, we obtain the group from the mortality data themselves via some machine learning methods. We design a DSA (deletion-substitution-addition) algorithm to choose the best grouping, which has both reliable explanatory power for current mortality patterns and superior performance in terms of forecasting accuracy for each target country.

(3) Hong Li, University of Manitoba
Title: *Gompertz Law Revisited: Forecasting Mortality in a Multi-Factor*
Abstract: This paper provides a flexible multi-factor framework to address some ongoing challenging issues in mortality modeling, particularly focusing on the mortality curvature and old age mortality plateau. In particular, we extend the Gompertz law (Gompertz 1825) to include factors capturing the curvature of mortality increase over age, as well as the decelerating mortality increase for the very old ages. The proposed framework permits a convenient estimation and prediction algorithm. An extensive empirical analysis is conducted using the proposed framework and different existing Gompertz-based mortality models with a large number of populations. We find that allowing a more flexible age pattern of mortality decline leads to better fitness of historical data, as well as out-of-sample forecast performance for most populations. Moreover, the proposed factor model may produce more reasonable mortality forecasts in the long-term.

(4) Jiandong Ren, Western University
Title: *Analysis of a Multivariate Compound Loss Model*
Abstract: We introduce a compound multivariate distribution designed for modeling insurance losses arising from dierent risk sources in insurance companies. The distribution is based on a discrete-time Markov Chain and generalizes the multivariate compound negative binomial distribution, which has been widely

used for modeling insurance losses. We derive fundamental properties of the distribution and discuss computational aspects facilitating calculations of risk measures of the aggregate loss, as well as allocations of the aggregate loss to individual types of risk sources. Explicit formulas for the joint moment generating function and joint moments of dierent loss types have been derived, and recursive formulas for calculating the joint distribution oered. Several special cases of particular interest have been analyzed and an illustrative numerical example provided.

Session 25: ***Bayesian Methods in Biostatistics***
Organizer and Chair: Shirin Golchi, McGill University
Room: B-1120, Time: 10:30AM-12:10PM

(1) Eleanor Pullenayegum, Hospital for Sick Children
Title: *Bayesian Approaches to Health Utility Estimation for Economic Evaluations*

Abstract: Economic evaluations are used to inform decisions on which treatments should be publicly reimbursed. An important component to these evaluations is the quality adjusted life years gained under the new treatment, which is captured using health utilities. Measurement of health utilities is complex, and can employ both discrete choice experiments and time trade-off tasks. As a result, analysis must capture complex correlation structures and correctly characterize uncertainty. This talk will show how Bayesian methods can use the posterior predictive distribution to correctly account for uncertainty, and show that the current manner in which economic evaluations are done severely overestimates precision in the quality adjusted life years gained. While researchers had hoped that data from discrete choice and time trade-off tasks could be combined to yield more accurate inference, both joint Bayesian mixed models and joint Bayesian latent class models reveal that agreement between the two is poor. Finally, we will see how allowing for correlation in the data through spatially correlated random effects can substantially reduce the extent of uncertainty of utilities elicited using time trade-off tasks.

(2) Shirin Golchi, McGill University
Title: *Use of Historical Individual Patient Data in Analysis of Clinical Trials*
Abstract: Historical data from previous clinical trials, observational studies and health records may be utilized in analysis of clinical trial data to strengthen inference. Instances of cases where historical data are crucial are clinical trials where assigning patients to a control group is considered unethical, patient accrual rates are low or required sample sizes for detection of the actual treatment effects are unaffordable/infeasible. Use of any external data should be conditioned on the validity of the source. More specifically, external data should only be used if it can be considered as a representative sample of the target population. Since the current data is considered the most immediate proxy to the target population, historical studies are used on the basis of their similarity to the current study. Under the Bayesian framework incorporation of information obtained from any

other source than the current data is facilitated through construction of an informative prior. The existing methodology for defining an informative prior based on historical data relies on measuring similarity to the current data at the study level that can result in discarding individual patient data (IPD). This talk is focused on a family of priors that utilize IPD to empower statistical inference. The proposed IPD-based prior is shown to outperform existing methods in terms of accuracy and precision of parameter estimates in a simulation study and is applied to analysis of data available from a selection of trials in second line non-small cell lung cancer.

(3) Bingshu Chen, Queens University
Title: *Joint Modeling of Biomarker and Survival Outcome for Clustered Data*
Abstract: In clinical trials, it is often desirable to evaluate the effect of a prognostic factor such as a marker response on the survival outcome. However, the marker response and survival are usually associated with some potential unobservable factors. In this case, the conventional statistical methods that model these two outcomes separately may not be appropriate. In this paper, we propose a joint model for the marker response and survival outcomes with clustered data. It provides an efficient statistical inference by considering these two outcomes simultaneously. We focus on a special type of marker response: binary outcome, which is investigated together with survival data using a cluster-specific multivariate random effect variable. A multivariate penalized likelihood method is developed to make statistical inference for the joint model. The standard errors obtained from penalized likelihood method are usually underestimated. This issue is addressed using a Jackknife resampling method. We conduct extensive simulation studies to assess the finite sample performance of the proposed joint model and inference method in different scenarios. The simulation studies show that the proposed joint model have excellent finite sample properties compared to the separate models when there exists a strong underlying association between the marker response and survival data. Finally, we apply the proposed method to a Hodgkin's lymphoma study conducted by Canadian Cancer Trials Group to explore the prognostic effect of disease remission on the overall survival.

(4) Audrey Béliveau, University of Waterloo
Title: *BUGSnet: a New All-in-One R Package to Improve the Quality and Reporting of Bayesian Network Meta-Analyses*
Abstract: Network meta-analysis (NMA) is a set of statistical tools conventionally used to establish comparative efficacy/safety of more than two interventions using data extracted from a systematic literature review of randomized controlled trials. Several reviews have noted shortcomings regarding the quality and reporting of network meta-analyses. We suspect that this issue may be partially attributable to limitations in current NMA software which do not readily produce all of the output needed to satisfy current guidelines. To better facilitate the conduct and reporting of NMAs, we have created an R package called BUGSnet (Bayesian inference Using Gibbs Sampling to conduct a Network meta-analysis). BUGSnet contains a suite of functions that can be used to describe the evidence

network, fit Bayesian NMA models, assess the model fit and convergence, assess the presence of heterogeneity and inconsistency, and output the results in a variety of formats including league tables and surface under the cumulative rank curve (SUCRA) plots. We provide a demonstration of the functions contained within BUGSnet by recreating a Bayesian NMA found in the second technical support document composed by the National Institute for Health and Care Excellence Decision Support Unit (NICE-DSU). We hope that this software will help to improve the conduct, reporting and reproducibility of NMAs.

Session 26: ***Stochastic Models for Dynamic Biological Data***
Organizer: Ximing Xu, Nankai University
Chair: Xueli Xu
Room: K-106, Time: 10:30AM-12:10PM

(1) Yun Cai, Dalhousie University
Title: *Deconvoluiton Density Estimation with MLE Method*

Abstract: Data obtained for density estimation are often contaminated with measurement error or sometimes convergence error. The budget and time concern to collect noise free data limit the amount of high-fidelity data available. To solve the problem, there exist a lot methods targeting on estimating density with abundant contaminated data and knowledge of the error distribution. Most of the method are based on Fourier Transformation and more focused on symmetric error distribution. Here we develop a deconvolution method based on constrained MLE. Our method is able to accurately estimate the density with limited noisy data and samples of pure error, without any other knowledge of the error distribution, even when the signal noise ratio is very low. The comparison with other method conclude that our method give a better estimation with both symmetric and unsymmetric error. And our method shows ability to not only save resource for collecting data but also save time for some statistic analysis procedure.

(2) Libai Xu, Nankai University
Title: *Stochastic Generalized Lotka-Volterra Dierential Equation Model with an Application to Learning Microbial Community Structures*

Abstract: Inferring microbial community structure based on temporal metagenomics data is an important goal in microbiome studies. The deterministic generalized LotkaVolterra dierential equations were used to model the dynamics of the microbial data, however, these approaches fail to take random environmental uctuations into account, which may deteriorate the estimates severely. In this paper, we propose a new stochastic generalized Lotka-Volterra dierential equation model, where the random perturbations of Brownian motion in the model can naturally account for the external environmental eect of microbial community. We establish new conditions and show various mathematical properties of the solutions including general existence and uniqueness, stationary distribution, ergodic property and long time behavior. We further develop an approximate maximum likelihood estimator based on discrete observations and systematically

investigate the consistency and asymptotic normality of the proposed estimators. Our method is demonstrated using simulation studies and an application from the moving picture temporal microbial dataset.

(3) Shen Ling, Dalhousie University
Title: *Handling Block-Missing Data with Model Combination Methods*
Abstract: Block-wise missing is a common occurrence in regression and classification problem. In this paper, we consider one such situation, where one part of observations is fully observed while the other part is only partially observed with some block of data missing. This type of problem has received some attention in the literature. We present a new model combination approach to effectively handle the block missing problem. The method works by a linear combination of full and partial model. We propose several ways to approximate the combination parameter, which is shown to be consistent. We also show that the combination method works better than using the full or partial model alone. Simulation studies are conducted to evaluate our new method. The method is also applied to some real data.

(4) Xueli Xu, Nankai University
Title: *Pupil Center Shift in Young Adults: Large Directional Data Analysis*
Abstract: Purpose: To investigate the pupil center shift with age and correlated trend between the left and right eyes in young adults. Design: A cross-sectional study. Subjects: 8525 subjects (17050 eyes) aged 18-44 years from one center for directional data analysis. Another set of 8184 subjects (16368 healthy eyes) from another center were used for validation. Methods: The subjects were divided into five age groups: 18-20, 21-25, 26-30, 31-35, and 36-44 years. The Pentacam HR system was used to obtain the images of the pupil. Directional statistics was introduced to the model and test the change of pupils positions represented by polar coordinates (r,). Regression splines were used to study the trend of pupil size (d) with age. Main Outcome Measures: The changes of polar angle (), polar-radius (r) and size (d) of pupil. Results: Among 17050 eyes, 70mean polar angle of pupil centers in the right and left eyes were (107.21.4) and (108.81.8), respectively; the mean polar-radius was (0.1830.104) mm and (0.1770.103) mm, respectively; and the mean pupil-diameter was (3.460.678) mm and (3.420.662) mm, respectively. The polar-angles between right and left eyes were negatively correlated with a circular correlation coefficient -0.378 (P¡0.0001). For both eyes, the polar radius had an overall trend of increase with age (correlation coefficient=0.380, P¡0.0001), shifting to the temporal side; However, pupil size had a decrease trend (correlation coefficient=0.816, P¡0.0001). The trends of the pupil polar-angles and polar-radius changes with age were similar in both centers. Conclusions: Our data showed that the pupil center of left eye shifts counterclockwise, while that of the right eye moves clockwise as a function of age, and the distance between the center of the pupil and corneal apex becomes larger in young adults. Directional statistics proved to be a powerful analytical tool to visualize and understand the shift of pupil position.

Session 27: ***New Techniques for Modern Data Analysis***
Organizer: Linglong Kong and Jingjing Wu, University of Alberta and University of Calgary
Chair: Linglong Kong
Room: K-107, Time: 10:30AM-12:10PM

(1) Yan Yuan, University of Alberta
Title: *The Index Lift in Data Mining Has a Close Relationship with the Association Measure Relative Risk in Epidemiological Studies*

Abstract: Background Data mining tools have been increasingly used in health research, with the promise of accelerating discoveries. Lift is a standard association metric in the data mining community. However, health researchers struggle with the interpretation of lift. As a result, dissemination of data mining results can be met with hesitation. The relative risk and odds ratio are standard association measures in the health domain, due to their straightforward interpretation and comparability across populations. We aimed to investigate the lift-relative risk and the lift-odds ratio relationships, and provide tools to convert lift to the relative risk and odds ratio.

Methods We derived equations linking lift-relative risk and lift-odds ratio. We discussed how lift, relative risk, and odds ratio behave numerically with varying association strengths and exposure prevalence levels. The lift-relative risk relationship was further illustrated using a high-dimensional dataset which examines the association of exposure to airborne pollutants and adverse birth outcomes. We conducted spatial association rule mining using the Kingfisher algorithm, which identified association rules using its built-in lift metric. We directly estimated relative risks and odds ratios from 2 by 2 tables for each identified rule. These values were compared to the corresponding lift values, and relative risks and odds ratios were computed using the derived equations.

Results As the exposure-outcome association strengthens, the odds ratio and relative risk move away from 1 faster numerically than lift, i.e. —log(odds ratio)— —log(relative risk)— —log(lift)—. In addition, lift is bounded by the smaller of the inverse probability of outcome or exposure, i.e. lift min(1/P(O), 1/P(E)). Unlike the relative risk and odds ratio, lift depends on the exposure prevalence for fixed outcomes. For example, when an exposure A and a less prevalent exposure B have the same relative risk for an outcome, exposure A has a lower lift than B.

Conclusions Lift, relative risk, and odds ratio are positively correlated and share the same null value. However, lift depends on the exposure prevalence, and thus is not straightforward to interpret or to use to compare association strength. Tools are provided to obtain the relative risk and odds ratio from lift.

(2) Brian Franczak, MacEwan University
Title: *Applications using Mixtures of Asymmetric Distributions*

Abstract: Classification can be lucidly defined as the process of assigning group labels to sets of observations. When a finite mixture model is used for classifi-

cation in either an unsupervised, semi-supervised, or supervised setting, one can refer to this process as model-based learning. In this talk, we will discuss a mixture of shifted asymmetric Laplace (SAL) distributions and extensions thereof. Compared to the well-known mixtures of Gaussian distributions, the mixtures of SAL distributions can parameterize skewness in addition to both location and scale, making it well suited for the analysis of data with homogeneous subpopulations that are not symmetric. Some details regarding the development of the mixture of SAL distributions will be provided and an expectation-maximization based parameter estimation scheme will be outlined. The classification performance of the mixture of SAL distributions will be demonstrated using several real data sets.

(3) Ejaz Ahmed, Brock University
Title: *Big Data Big Bias Small Surprise!*

Abstract: Nowadays a large amount of data is available, and the need for novel statistical strategies to analyze such data sets is pressing. This talk focuses on the development of statistical and computational strategies for a sparse regression model in the presence of mixed signals. The existing estimation methods have often ignored contributions from weak signals. However, in reality, many predictors altogether provide useful information for prediction, although the amount of such useful information in a single predictor might be modest. The search for such signals, sometimes called networks or pathways, is for instance an important topic for those working on personalized medicine. We discuss a new post selection shrinkage estimation strategy that takes into account the joint impact of both strong and weak signals to improve the prediction accuracy, and opens pathways for further research in such scenarios.

(4) Mihye Ahn, University of Nevada, Reno
Title: *Spatially Weighted Reduced-Rank Framework for Group Analysis of Functional Neuroimaging Data*

Abstract: Recently, much attention has been received on the analysis of functional imaging data to delineate the intrinsic functional connectivity pattern among different brain regions within each subject. However, only few approaches for integrating functional connectivity pattern from multiple subjects have been proposed. The goal of this study is to develop a reduced-rank model framework for analyzing the whole-brain voxel-wise functional images across multiple subjects in the frequency domain. Considering the neighboring voxels with different weights, the frequency and spatial factors can be extracted. Imposing sparsity on the frequency factors enables us to identify the dominant frequencies. In addition, the spatial maps can be used for detecting group difference, when the comparison between different groups is of specific interest. Simulation study shows that the proposed method achieves less spatial variability and better estimates of frequency and spatial factors, compared to some existing methods. Finally, we apply the proposed method to ADNI data.

Session 28: ***Analysis of Complex Structured Data***

Organizer and Chair: Wenqing He, University of Western Ontario
Room: B-2109, Time: 10:30AM-12:10PM

(1) Liqun Wang, University of Manitoba
Title: *Variable Selection and Estimation in Generalized Linear Models with Measurement Error*

Abstract: We study the variable selection problem in linear and generalized linear models when some of the predictors are measured with error. We demonstrate through numerical examples how measurement error (ME) affects the selection results and propose a regularized instrumental variable (RIV) method to correct for the ME effects. We show that the proposed estimator has the oracle property in a linear model and we derive its asymptotic distribution under general conditions. We also investigate the performance of this method in generalized linear models. Our simulation studies show that the RIV estimator outperforms the naive estimator in both linear and some generalized linear models. Finally, the proposed method is applied to a real dataset.

(2) Ying Zhang, Acadia University
Title: *Some Nonparametric Rank-Based Statistics in Time Series Trend Analysis*

Abstract: Nonparametric sign- or rank-based statistics are popular in analyzing for trend/changes in time series data. In this talk, we will discuss several well-known nonparametric rank-based procedures in the literature and address the issues with the time series data applications that we recently conducted. Finally, we will present a new procedure for time series trend analysis by moving windows of the time series.

(3) Longhai Li, University of Saskatchewan
Title: *Estimating Cross-Validatory Predictive P-Values with Integrated Importance Sampling for Disease Mapping Models*

Abstract: An important statistical task in disease mapping problems is to identify out- lier/divergent regions with unusually high or low residual risk of disease. Leave-one-out cross-validatory (LOOCV) model assessment is a gold standard for computing predictive p-value that can flag such outliers. However, actual LOOCV is time-consuming because one needs to re-simulate a Markov chain for each posterior distribution in which an observation is held out as a test case. This paper introduces a new method, called iIS, for approximating LOOCV with only Markov chain samples simulated from a posterior based on a full data set. iIS is based on importance sampling (IS). iIS integrates the p-value and the likelihood of the test observation with respect to the distribution of the latent variable without reference to the actual observation. The predictive p-values computed with iIS can be proved to be equivalent to the LOOCV predictive p-values, following the general theory for IS. We com- pare iIS and other three existing methods in the literature with a lip cancer dataset collected in Scotland. Our empirical results show that iIS provides predictive p-values that are al- most identical to the actual LOOCV predictive p-values and outperforms the existing three methods, including the recently proposed ghosting method by Marshall and Spiegelhalter

(2007).

(4) Wenyu Jiang, Queen's University
Title: *Testing for Cluster Heterogeneity in Joint Modeling of Survival Time and a Marker Response in Clinical Trials*

Abstract: Large-scale clinical trials often recruit patients from multiple institutions to attain desired sample sizes. Heterogeneity among institutions can be described by cluster-level random effects. For a Hodgkins Lymphoma clinical trial HD.6 across 29 medical institutions, Wang and Chen (2019) proposed to model a remission response and survival time jointly, while accounting for cluster heterogeneity. For this complicated joint model, we propose a test to check if the cluster effects do exist in the first place. If not, then we can simply leave out the cluster random effects in the analysis. We develop the test by extending a score test for homogeneity (Liang, 1987) to the joint model scenario, and explore the variance estimation for the score statistic by both bootstrap and asymptotic derivation. We carry out simulation studies and find both approaches give correct test sizes while the bootstrap method achieves higher power. For the HD.6 data, we do not find enough evidence for cluster-level random effects. The analysis of the joint model of Wang and Chen (2019) can be much simplified, that is, it is not necessary to model cluster heterogeneity for patients from different institutions.

Session 29: ***Contributed Talks***
Organizer and Chair: Francois Marshall, Queen's University
Room: B-2111, Time: 10:30AM-12:10PM

(1) Kaili Jing, University of Ottawa
Title: *Distributed Hard Screening for Massive Data*

Abstract: Feature screening is a powerful tool for modeling high dimensional data. It aims at reducing the dimensionality by removing most irrelevant features before an elabo- rative analysis. When a dataset is massive in both sample size N and dimensionality p, classic screening methods become ineffcient or even infeasible due to the high computational burden. In this paper, we propose a distributed screening method for the large-N-large-p setup. The new method is built upon an ADMM updating procedure of l0-constrained consensus regression, where data are processed in m manageable segments by multiple local computers. In the procedure, the local computers improve screening results iteratively by communicating with each other via a global computer. The joint effects between features are also accounted naturally in the screening process. It thus provides a computationally viable and reliable route for screening features with big data. Under mild conditions, we show that the proposed updating procedure is convergent and leads to an accurate screening even when m = o(N). Moreover, with a proper starting value, the procedure enjoys the sure screening property within finite number of iterations. The promising performance of the method is supported by extensive numerical studies.

(2) Xingxiang Li, University of Ottawa
Title: *Distributed Feature Screening via Componentwise Debiasing*
Abstract: When the sample size N and the number of features p are both large, the implementation of classic screening methods can be numerically challenging. In this paper, we propose a distributed screening framework for big data setup. In the spirit of "divide-and-conquer", the proposed framework expresses a correlation measure as a function of several component parameters, each of which can be distributively estimated using a natural U-statistic from data segments. With the component estimates aggregated, we obtain a final correlation estimate that can be readily used for screening features. This framework enables distributed storage and parallel computing and thus is computationally attractive. Due to the unbiased distributive estimation of the component parameters, the final aggregated estimate achieves a high accuracy that is insensitive to the number of data segments m specified by the problem itself or to be chosen by users. Under mild conditions, we show that the aggregated correlation estimator is as efficient as the centralized estimator in terms of the probability convergence bound and the mean squared error rate; the corresponding screening procedure enjoys sure screening property for a wide range of correlation measures. The promising performances of the new method are supported by extensive numerical examples.

(3) Zhiyang Zhou, Simon Fraser University
Title: *Functional Continuum Regression*

Abstract: Functional principal component regression (PCR) can fail to provide good prediction if the response is highly correlated with some excluded functional principal component(s). This situation is common since the construction of functional principal components never involves the response. Aiming at this shortcoming, we develop functional continuum regression (CR). The framework of functional CR includes, as special cases, both functional PCR and functional partial least squares (PLS). Functional CR is expected to own a better accuracy than functional PCR and functional PLS both in estimation and prediction; evidence for this is provided through simulation and numerical case studies. Also, we demonstrate the consistency of estimators given by functional CR.

(4) Wen Teng, Queen's University
Title: *Simultaneous Confidence Bands for Treatment-Biomarker Interaction Effects in Clinical Trials Based on Local-Partial Likelihood*

Abstract: In clinical trials, treatment effects may vary with the levels of a biomarker. By the local-partial likelihood technique, the treatment-biomarker interaction effect can be modeled as a flexible function of the biomarker without pre-specified forms. The overall trend of biomarker-dependent effects can be estimated, along with confidence bands. Inference based on confidence bands over the range of thebiomarker is more informative than that based on pointwise confidence intervals. In our method, we approximate the asymptotic distribution of maximum absolute deviation of the estimator for treatment-biomarker interaction effect, then develop the simultaneous confidence band based on the limiting distribution. The performance of this approach is studied in simulation,

and we also illustrate the use of the proposed simultaneous confidence bands on a prostate cancer clinical trial.

(5) Dave Riegert, Queen's University
Title: *Function Estimation - A Possible Approach to Non-Stationarity*
Abstract: The relationship between geoelectric and geomagnetic fields can be modelled through the estimation of transfer functions. The estimated transfer functions can then be used to infer information about the Earth's subsurface structure. Assumptions used during the modelling process are informed by Maxwell's equations, including that the relationship is linear and stationary. However, in the 1700's it was observed that the geomagnetic field varies more in the early afternoon and less in the early evening (Graham, 1724) which implies that the geomagnetic field, as measured on Earth, is non-stationary. As a result, transfer function estimation could benefit from new approaches with the goal of obtaining a more accurate estimate of the relationship between electric and magnetic fields on Earth. I outline one such approach and provide examples

(6) Francois Marshall, Queen's University
Title: *A Monte Carlo Study of the Distributions in Spectral Analysis*
Abstract: Spectral analysis is a branch of signal processing in which time-series data are Fourier-transformed prior to conducting statistical analysis. The advantage of the Fourier-transform step is that the transform of the noise component of the stochastic process is often IID complex-normal. Based on a Monte Carlo analysis, the talk will provide new motivation for the notion of IID complex-normal Fourier transforms. Novelcentral-limit theorems will be presented which motivated the considered time-series models in the simulation. Correlation structure is enriched using linear filters, and it will be shown how the complex-normal behaviour of Fourier transforms the output processes remains preserved.

Provided Lunch, Leonard Hall (150 Queens Crescent)

2:30pm -5pm Cruise Trip

The cruise will set out in the town of Gananoque (35-minute drive east of Kingston). The bus will pick up all the participants at Leonard Hall at 1:10pm on Aug 11, and drop off on campus around 5:45pm.

List of Participants

Note: late registrants are not listed here.

Last Name	First Name	Organization	Email
Ahn	Mihye	University of Nevada, Reno	mahn@unr.edu
Bhatia	Himesh	Queen's University (Kingston)	himeshbhatia@live.com
Blanchette	Kian	Queen's University	kmlb@protonmail.com
Bliveau	Audrey	University of Waterloo	audrey.beliveau@uwaterloo.ca
Cai	Tony	The Wharton School	tcai at wharton dot upenn dot edu
Cai	Song	Carleton University	scai@math.carleton.ca
Cai	Yun	Dalhousie University	yn298893@dal.ca
Chen	Bingshu	Queens University	bingshu.chen@queensu.ca
Chen	Jiahua	UBC	jhchen@stat.ubc.ca
Chen	Yilin	University of Waterloo	1010cyl@gmail.com
Cui	Qi	Simon Fraser University	qi_cui@sfu.ca
de Leon	Alexander	University of Calgary	adeleon@ucalgary.ca
Diao	Liqun	University of Waterloo	l2diao@uwaterloo.ca
Ding	Hao	York University	dinghaostat@gmail.com
Du	Pang	Virginia Tech	pangdu@vt.edu
Dubin	Joel	University of Waterloo	jdubin@uwaterloo.ca
Fan	Yingying	University of Southern California	fanyingy@marshall.usc.edu
Franczak	Brian	MacEwan University	franczakb@macewan.ca
Gan	Guojun	University of Connecticut	guojun.gan@uconn.edu
Ge	Xinyi	Queen's University	16xg3@queensu.ca
Golchi	Shirin	McGill University	golchi.shirin@gmail.com
Gu	Hong	Dalhousie University	hgu@dal.ca
Han	Bing	RAND Corporation	bhan@rand.org
He	Wenqing	University of Western Ontario	whe@stats.uwo.ca
Heng	Jiani	Queen's Univeristy	18jh51@queensu.ca
Hu	Jianhua	Shanghai University of Finance and Economics	hu.jianhua@mail.shufe.edu.cn
Jiang	Wenyu	Queen's University	wenyu.jiang@queensu.ca
Jin	Baisuo	University of Science and Technology of China	jbs@ustc.edu.cn
Jin	Zhezhen	Columbia University	zj7@cumc.columbia.edu
Jing	Kaili	University of Ottawa	kjing073@uottawa.ca
Kenney	Toby	Dalhousie University	tkenney@mathstat.dal.ca
Khalili	Abbas	McGill University	abbas.khalili@mcgill.ca
Kokocinski	Jordan	Queen's University	11jdk13@queensu.ca
Kong	Dehan	University of Toronto	dehan.kong@utoronto.ca
Kong	Linglong	University of Alberta	lkong@ualberta.ca
Kopciuk	Karen	University of Calgary	kakopciu@ucalgary.ca
Li	Bing	Pennsylvania State University	bxl9@psu.edu
Li	Dongdong	Harvard Medical School	dli.stats@gmail.com
Li	Gen	Columbia University	gl2521@cumc.columbia.edu
Li	Hong	University of Manitoba	lihongfinance@yahoo.com
Li	Linke	Queen's University	15ll48@queensu.ca
Li	Longhai	University of Saskatchewan	longhai.li@gmail.com
Li	Na	Queen's university	18nl6@queensu.ca
Li	Pengfei	University of Waterloo	pengfei.li@uwaterloo.ca
Li	Runze	Pennsylvania State University	runzeli@psu.edu
LI	Wanlu	Queen's University	18wl2@queensu.ca
Li	Xingxiang	University of Ottawa	xli396@uottawa.ca
Lin	Chunfang	Queen's University	devon.lin@queensu.ca
Ling	Lisa	Dalhousie University	lisaling22@live.cn
Liu	Juxin	University of Saskatchewan	liu@math.usask.ca
Liu	Shikai	Queen's University	16sl33@queensu.ca
Liu	Wei	York University	liuwei@mathstat.yorku.ca
Liu	Xianhui	Jiangxi University of Finance and Economics	liuxh@mail.Ustc.edu.cn
Liu	Xin	Shanghai University of Finance and Economics	liu.xin@mail.shufe.edu.cn
Liu	Zhihui	Princess Margaret Cancer Centre	zhihuiamy.liu@uhnresearch.ca
Lou	Wendy	University of Toronto	wendy.lou@utoronto.ca
Lu	Wilson	Acadia University	wilson.lu@acadiau.ca
Lu	Xiaoming	Memorial University	xiaoming.lu@mun.ca
Lu	Xuewen	University of Calgary	xlu@ucalgary.ca
Lu	Zihang	University of Toronto	zihang.lu@mail.utoronto.ca
Luo	Liping	Hengyang Normal University	luolp3456034@163.com
Ma	Renjun	University of New Brunswick	renjun@unb.ca
Marriott	Paul	University of Waterloo	pmarriot@uwaterloo.ca

Marshall	Francois	Queen's University	francois.marshall@queensu.ca
Meng	Xuejing	Simon Fraser University	xuejing_meng@sfu.ca
Niu	Yi	Dalian University of Technology	niuyi@dlut.edu.cn
Peng	Jianan	Acadia University	jianan.peng@acadiau.ca
Peng	Yingwei	Queen's University	pengp@queensu.ca
Pullenayegum	Eleanor	Hospital for Sick Children	eleanor.pullenayegum@sickkids.ca
Qin	Shanshan	York University	qinsslzu@gmail.com
Qin	Yichen	University of Cincinnati	qinyn@ucmail.uc.edu
Qiu	Xing	University of Rochester	xing.qiu@gmail.com
Ren	Jiandong	University of Western Ontario	jren6@uwo.ca
Rice	Gregory	University of Waterloo	grice@uwaterloo.ca
Riegert	David	Queen's University	driegert@gmail.com
Sang	Peijun	University of Waterloo	psang@uwaterloo.ca
Sattar	Abdus	CWRU	sattar@case.edu
Sharafoddini	Anis	University of Waterloo	a.sharafoddini@gmail.com
Shi	Xiaoping	Thompson Rivers University	xshi@tru.ca
Singh	Somya	Queens university	17ss262@queensu.ca
Sinha	Sanjoy	Carleton University	sinha@math.carleton.ca
So	Hon Yiu	University of Waterloo	honyius@yahoo.com.hk
Song	Yanglei	Queen's University	yanglei.song@queensu.ca
Stewart	Connie	University of New Brunswick Saint John	cstewart@unb.ca
Sun	Zhaohan	University of Waterloo	z227sun@uwaterloo.ca
Teng	Wen	Queen's University	16wt@queensu.ca
Thompson	Mary E	University of Waterloo	methomps@uwaterloo.ca
Tu	Dongsheng	Queen's University	dtu@ctg.queensu.ca
Turgeon	Maxime	University of Manitoba	Max.Turgeon@umanitoba.ca
Wallace	Michael	University of Waterloo	michael.wallace@uwaterloo.ca
Wang	Chunlin	Xiamen University	wangc@xmu.edu.cn
Wang	HaiYing	University of Connecticut	haiying.wang@uconn.edu
Wang	Liming	Nanjing University of Finance & Economics Hongshan College	wanglimingnuist@163.com
Wang	Linbo	University of Toronto	linbo.wang@utoronto.ca
Wang	Liqun	University of Manitoba	Liqun.Wang@umanitoba.ca
Wang	Qihua	AMSS, Chinese Acacdemy of Sciences	qhwang@amss.ac.cn
Wang	Suojin	Texas A&M University	sjwang@stat.tamu.edu
Wang	Tao	Yunnan Normal University	wtaokm@263.net
Wang	Yafei	Beijing University of Technology	wyfzye@sina.com
Wang	Zhanfeng	University of Science and Technology of China	zfw@ustc.edu.cn
Wang	Zhonglei	Wang Yanan Institute for Studies in Economics	wangzl@xmu.edu.cn
Weng	Chengguo	University of Waterloo	chengguo.weng@uwaterloo.ca
Wu	Changbao	University of Waterloo	cbwu@uwaterloo.ca
Wu	Lang	University of British Columbia	lang@stat.ubc.ca
Wu	Yuehua	York University	wuyh@mathstat.yorku.ca
Xia	Chaoxiong	Northern Illinois University	cxia@niu.edu
Xu	Chen	Univeristy of Ottawa	cx3@uottawa.ca
Xu	Libai	Nankai University	libaixuyx@outlook.com
Xu	Wei	Princess Margaret Cancer Centre	wxu@uhnres.utoronto.ca
Xu	Xueli	Nankai University	15900388365@163.com
Yan	Guohua	University of New Brunswick	guohuayan@gmail.com
Yang	Fan	University of Waterloo	fan.yang@iuwaterloo.ca
Yang	Feng	Queen's University	18fy4@queensu.ca
Yeh	Chi-Kuang	University of Waterloo	chi-kuang.yeh@uwaterloo.ca
Yi	Grace	University of Waterloo	yyi@uwaterloo.ca
Yu	Hao	University of Western Ontario	hyu@stats.uwo.ca
Yu	Qianhui	Queen's University	tracyyu@outlook.com
Yuan	Meng	University of Waterloo	m33yuan@uwaterloo.ca
Yuan	Yan	University of Alberta	yyuan@ualberta.ca
Yue	Jianwei	Queen's University	17jy84@queensu.ca
Zeng	Leilei	University of Waterloo	lzeng@uwaterloo.ca
Zhang	Hongbin	Graduate School of Public Health and Health Policy	hongbin.zhang@sph.cuny.edu
Zhang	Min	University of South China	emin1217@sina.com
Zhang	Ying	Acadia University	ying.zhang@acadiau.ca
Zhang	Yuping	University of Connecticut	yuping.zhang@uconn.edu
Zhao	Puying	Yunnan University, Yunnan, China	pyzhao@live.cn
Zhao	Xingqiu	The Hong Kong Polytechnic University	xingqiu.zhao@polyu.edu.hk
Zhao	Yize	Yale University	zhaoyize1988@gmail.com
Zhou	Zhiyang	Simon Fraser University	zhiyang_zhou@sfu.ca
Zhou	Zhou	University of Toronto	zhouzhou.stat@gmail.com
Zhu	Yeying	University of Waterloo	yeying.zhu@uwaterloo.ca

Zhu	Zimo	Queen's University	zimoz1994@gmail.com
Zhuang	Weiwei	University of Science and Technology of China	weizh@ustc.edu.cn

International Chinese Statistical Association

泛華統計協會

Membership Application & Renewal Form

Name	(Last)	(Middle)	(First)
(English)			
(Chinese)			
Affiliation:			
Gender (optional):			
Address			
Office	Address:		
	City:		
	State:	Zip Code:	Country:
	Email:	Telephone:	FAX:
Home	Address:		
	City:		
	State:	Zip Code:	Country:
	Email:	Telephone:	FAX:
Education			
	Degree:	Year Graduated:	
	University:		

Professional Occupation & Title

Occupation: Title:

Membership Classification & Fees

Regular (US$80)
Student (FREE) without voting right nor hard copies of ICSA Bulletin
Permanent (US$1000)
Spouse (50%)
Donations US$
Total Amount Paid: US$

Statistical Area of Interest (circle all applicable):

A: Agriculture
B: Business / Economics
C: Computing / Graphics
D: Education
E: Engineering
F: Health Sciences
G: Probability
H: Social Sciences
I: Biostatistics
N: Theory & Methodology

Would you like to join the Biometrics Section (please choose your answer)?

Yes No

Please Make Check Payable to: I.C.S.A. Mail This Form & Fees to:

ICSA c/o Hongliang Shi, 18 Shagbark Road, Concord, MA, USA, 01742-2029.
Email: icsa.finance@gmail.com.