

Submittee: Vincenzo Coia
Date Submitted: 2015-03-10 12:05
Title: SFU/UBC Joint Statistics Seminar
Event Type: Lecture-Seminar-Series

Location:
SFU Harbour Center, Vancouver B.C.

Dates:
February 28, 2015

Topic:
Statistics

Methodology:

The seminar consisted of:

- Six presentations from graduate students in Statistics, each 25 minutes in length + 5 minutes of questions (three from UBC and three from SFU) //
 - Two presentations from faculty in Statistics, each 50 minutes + 10 minutes of questions (one from UBC and one from SFU) //
 - Lunch is also offered to encourage networking. //
-

Objectives Achieved:

- Participants shared and discussed ideas from the presentations. //
 - Networking (grad students from both SFU and UBC talked over lunch). //
-

Scientific Highlights:

N/A

Organizers:

Coia, Vincenzo, Department of Statistics, The University of British Columbia //
Dinsdale, Daniel, Department of Statistics, The University of British Columbia //
Di Gravio, Chiara, Department of Statistics, The University of British Columbia //
Grosskopf, Michael, Department of Statistics & Actuarial Science, Simon Fraser University //

Speakers:

Vincenzo Coia, Department of Statistics, The University of British Columbia //
Title: "Flood Warning: An Application of High-Quantile Regression" //

Abstract: "Alberta was largely unprepared for the big flood when it hit in 2013, perhaps due to a forecast of the expected river flow. In this talk, I propose a method to obtain a 'bigger picture' of the Bow River's flow by estimating upper quantiles of river flow, given some predictors like rainfall and snowmelt. Existing methods in Quantile Regression may not be sufficient to deal with this, so I use a newer technique that uses copulas, along with a new estimator." //

Huijing Wang, Department of Statistics & Actuarial Science, Simon Fraser University //

Title: "Analysis of Time to TB Infection data by a Mixture Cure Model with MNAR covariates" //

Abstract: "A cohort of Tuberculosis (TB) contacts were identified from TB patients and followed for about ten years at BCCDC. To analysis of time to TB infection of the cohort, a Mixture Cure Model was formulated and estimated by the likelihood based approaches via EM algorithm. The difficulty of missing not at random (MNAR) covariates was addressed at the same time by using additional info. The analysis results identified risk factors of TB development and provided useful info for cost-efficient TB screening strategies." //

Xiaoli Yu, Department of Statistics, The University of British Columbia //

Title: "Two-stage Sequential Designs for Binary Dose-Response Clinical Trials" //

Abstract: "Binary dose-response experiments are widely used in biomedical research. One major goal of such experiments is to estimate the dose-response relationship. Many existing designs for binary dose-response experiments have focused on accurately estimating a single quantile on the dose-response curve. However, in many applications, researchers are often interested in knowing the dose-response relationship over a range of dose levels. Then simultaneously estimating several quantiles on a dose-response curve is often helpful to better characterize this relationship. Motivated by this observation, we propose a two-stage sequential design which simultaneously estimates several quantiles on a dose-response curve. The proposed design consists of two stages. The first stage is a standard fix-point experiment used to provide a temporary and rough estimation of the parameters of the dose-response model. The second stage is a fully sequential experiment used to determine the most informative subsequent dose levels assigned to experimental subjects. Simulations indicate that the proposed two-stage design compares favorably with many existing designs under various scenarios in terms of root mean square error." //

Dr. Tom Loughin, Department of Statistics & Actuarial Science, Simon Fraser University //

Title: "Improving Communication with Collaborators by "Doctoring" the Environment" //

Abstract: "Practicing statistics in a team environment usually involves working with partners whose knowledge in their own areas is tremendous, but whose statistical expertise may be minimal. Furthermore, their interest in, and appreciation of what the statistician can offer may vary, as may their ability to communicate their ideas to someone who does not understand their discipline well. It is nonetheless necessary to partner with these collaborators and to form the best team possible. Communication is the key to a successful partnership. It is the Statistician's responsibility to translate a collaborator's ideas into a statistical problem that can be solved, and to translate the solution back into something that the collaborator can use.

In this talk I will discuss specific strategies that can facilitate this communication process and improve the Statistician's effectiveness and standing within a team. In particular, I draw the analogy between our job and that of a family physician, and use examples from both practices to highlight the strategies." //

Dr. Sara Mostafavi, Department of Statistics, The University of British Columbia //

Title: "Statistical integration of genomics data to derive gene functional networks and to predict gene function" //

Abstract: "In this talk I'll present two projects related to predicting gene functional relationships from heterogeneous types of genomics data. In the first part of the talk, I'll focus on the problem of gene function prediction, where gene co-functional networks are used to make predictions about the most likely function for a specific gene. The algorithms I'll introduce use a graph-based integration

framework, where graphs (aka networks) are used to facilitate a common representation. These algorithms result in state-of-the-art in accuracy and scalability. In the second part of the talk, I'll describe an approach, based on learning sparse Gaussian Graphical Models, for learning sparse graphs from gene expression data, while taking into account tissue-specificity of co-expression relationships." //

Yi Xiong, Department of Statistics & Actuarial Science, Simon Fraser University //

Title: "Analysis of Spatio-Temporal Data for Forest Fire Control" //

Abstract: "This project aims to establish the relationship of forest fire behavior with ecological /environmental factors, such as forest structure and weather. We analyze records of forest fires during the fire season (May to September) in 1992 from the Forest Fire Management Branch of Ontario Ministry of Natural Resource (OMNR). We start with a preliminary analysis of the data, which includes a descriptive summary and an ordinary linear regression analysis with fire duration as the response. The preliminary analysis indicates that the fire weather index (FWI) used by Natural Resource of Canada is the most relevant together with fire location and starting time. We apply semi-variogram and Moran's I, the conventional methods for exploring spatial patterns, and extend them to investigate spatio-temporal patterns with the fire data. Evaluations of the extended Moran's I statistic with the residuals of the ordinary linear regression analysis reveal a large departure from the independence and constant variance assumption on the random errors.

It motivates two sets of partially linear regression models to accommodate possible nonlinear spatial/temporal patterns of the forest fires. We integrate univariate and bivariate Kernel smoothing procedures with the least squares procedure for estimating the model parameters. Residual analysis indicates satisfactory fittings in both sets of regression analysis. The partially linear regression analyses find that the association of fire duration with FWI varies across different fire management zones, and depends on the fire starting time." //

Tingting (Crystal) Zhao, Department of Statistics, The University of British Columbia //

Title: "User-friendly Bayesian analysis of continuous time Markov chains" //

Abstract: Bayesian analysis of continuous time, discrete state space time series is an important and challenging problem, where partial observation and large parameter sets call for user-defined structured prior distributions. We proposed to use Generalized linear models to construct such prior distributions. An important obstacle with Bayesian GLM modelling of continuous time Markov chains (CTMCs) is that classical Markov chain Monte Carlo techniques are too ineffective to be practical. We address this by using an auxiliary variable construction combined with an adaptive Hamiltonian Monte Carlo (HMC) algorithm. The sampling algorithm and the model make it efficient to construct complex CTMC with prior known structure. We use both synthetic and real phylogenetic protein data to demonstrate the flexibility and scalability of our method, where a structured prior on amino acid based on physicochemical properties is exploited to obtain more accurate rate matrix estimates. //

Elena Szefer, Department of Statistics & Actuarial Science, Simon Fraser University //

Title: "Joint analysis of imaging and genomic data to identify associations related to cognitive impairment" //

Abstract: "Both genetic variants and brain region abnormalities are recognized to play a role in cognitive decline. In this project, we explore the relationship between genome-wide variation and region-specific rates of decline in brain structure, as measured by magnetic resonance imaging. The correspondence between rates of decline in brain regions and single nucleotide polymorphisms (SNPs) is investigated using data from the Alzheimer's Disease Neuroimaging Initiative 1 (ADNI-1), a study of Alzheimer's disease and mild cognitive impairment. In these data, the number of SNP and imaging biomarkers greatly exceeds the number of study subjects. To explore these data, we therefore look to modern multivariate statistical techniques that find sparse linear combinations of the two datasets having maximum correlation. These methods are particularly appealing because they greatly reduce the dimensions of the data, providing a low-dimensional representation of the

data to explore. Regularization of the correlation structure through a "sparse" singular value decomposition makes multivariate analysis on a large set of biomarkers possible. Using sparse linear combinations of the two datasets also incorporates variable selection into the analysis, providing insight into which genetic variants are associated with cognitive decline. Resampling techniques are used to examine the validity of the results by exploring their reproducibility in independent test sets, and by assessing the stability of the variable selection." //

Links:

<http://stat.ubc.ca/~vincen.coia/seminar.html>
