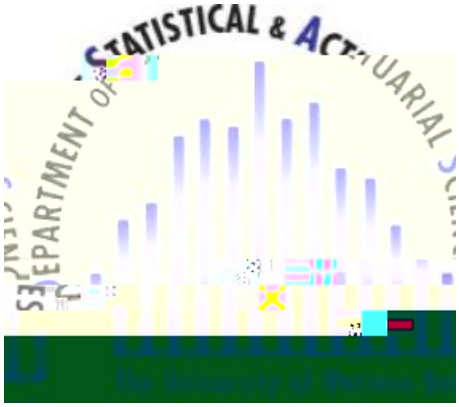
July 29, 2021



17th Annual Master's Day

.....3
.....4
.....4
.....4
.....4
.....5
.....6
.....6
.....6
.....7
.....7
.....8

9:00 a.m. Student and Faculty Zoom Log-in Now Open – *Master's Day – Open/Close*

9:15 a.m. Opening Remarks, Marcos Escobar-Anel (Graduate Chair)

9:30 a.m. Proceed to 2nd Zoom Link for Master's Day Presentations

Zoom Link
Yunxiao Dou
Haili Tian
Yijia Weng

Zoom Link
Neal Jackett
Mohsen Mollahajiaghari
Yaoxuan Shen
Xier Zeng

Zoom Link
Jiayue Tian
Zhaoqi Yang
Xinyi Zeng

Zoom Link
Yueqi Qiu
Pouya Sharifi
William Sirois

11:00 a.m. Completion of Master's Day Sessions

11:30 a.m. Rejoin *Master's Day – Open/Close* Zoom for closing remarks and Awards, Graduate Chair

: A Case Study of Understanding Risk Factors Associated with COVID-19 Mortality

Since the SARS-CoV-2 virus began its worldwide spread in March 2020, a significant fraction of health care resources has been devoted to dealing with confirmed and probably cases. It is crucial to efficiently allocate these scarce and at times constrained resources to patients in direst need, including those with a high risk of mortality from the infection. Individuals differ in demographics, geography, exposure history, disease severity indicators and outcomes, and presence of any underlying medical conditions and risk behaviors. Estimates of individual risks could help the way hospitals allocate resources to those with urgent need. We analyze a COVID-19 case surveillance public dataset from U.S. Centers for Disease Control and Prevention (CDC) to study the relationship between risk factors and death rates of COVID-19 to qualitatively identify key factors and quantitatively characterize the covariate effects. We employ three methods to deal with missing values in the dataset. Our analysis results help identify the risk factors that are associated with the COVID-19 death rates and provide assessment of mortality risks for different individuals.

Distribution of COVID-19 incubation time in Ontario, Canada

With the spread of COVID-19 epidemic, the research of central epidemiological parameters that determine the transmission of the disease, including the mean incubation time, becomes essential for taking effective health measures. This study used survival analysis to characterize the distribution of incubation time for confirmed COVID-19 patients in Ontario, Canada using the data from February 29, 2020, to May 22, 2021. Non-parametric and parametric methods are used to explore the distribution of incubation time, and the mean incubation time of COVID-19 is obtained. Stratification of factors and Accelerated Failure Time (AFT) regression model are also employed to examine the effects of the factors on the distributions of incubation times.

Meta-Analysis for the Average Incubation Time of COVID-19

Studying the incubation period of the coronavirus disease 2019 (COVID- 19), defined as the time interval from infection of the virus to the onset of clinical symptoms, is important to understand the infectious disease. Various studies have been available for estimating the average incubation time under different settings. To obtain a synthetic estimate, we search through online databases to collect the reports about average incubation time estimates and conduct meta-analyses on them. The pooled average incubation time is obtained through random effects models, and heterogeneity among different studies is quantified using Cochran's heterogeneity statistic Q and Higgin's & Thompson's I2 statistic. Subgroup analyses are conducted using mixed effects models and publication bias is assessed using the funnel plot and Egger's test. We conduct meta-analyses from different perspectives of viewing the data. Our study shows that using all those reported mean incubation estimates, the average incubation time is estimated to be 6.43 days with 95% confidence interval (CI) (5.90, 6.96), and that using all those reported mean incubation estimates together with those transformed median incubation estimates, the estimated average incubation time is 6.07 days with 95% CI (5.70,6.45).

COVID-19, Incubation Period, Meta-analysis

Examining the utility of tree-based regression techniques on the house price estimation

Classical regression techniques like simple linear regression and multiple linear regression may do a poor job in predicting results if the data being modelled is highly non-linear or if complex relationships exist between the explanatory and response variables. Furthermore, it may be difficult to specify a model or distribution for larger and complex datasets due to the sparsity of high dimensional data. We investigate the utility of tree-based methods and their ability to ameliorate some of these problems. We attempt to predict property values in the Ames, Iowa housing dataset using a gradient boosted decision tree model. Since we are interested in predicting valuations with low error, and not making inferences about the relationship between variables, we are satisfied with the tree-based approach.

Collective behaviour of financial markets

Financial markets are represented as complex systems. In this work, we focus on two markets, the cryptocurrency market and S&P/TSX 60. We analyze correlations between price changes of different cryptocurrencies and also between stocks in S&P/TSX 60. We review the previous works on the collective behavior of the cryptocurrency market with the help of graph theory. The minimum spanning tree of cryptocurrency correlations reveals distinct community structures. Then we work on S&P/TSX 60. We define a correlation graph for this market and with the help of spectral graph theory, we find distinct cluster structures.

Ocean freight rate modelling for the pricing and risk measurement of shipping risks

Abstract: The shipping industry has been thrown into a period of unprecedented volatility due to the whipsawing demand for retail. The sudden and dizzying heights of demand had to be catered by container ships hauling hundreds of thousands of tons at a time. Every idled ship in the world's fleet of merchant vessels was pretty much pulled into trade. Fierce competition for ocean freight capacity has become the new normal. With the new capacity slowly coming onstream, freight rates are expected to continue soaring to new highs and will most likely remain above their pre pandemic levels at least for the medium term horizon. Based on the salient features of the Capesize and Panamax indices data obtained from the Baltic Exchange, we develop a Markov switching framework to capture the indice dynamics. A multivariate Markov setting is proposed as well to accurately model the prices of the Forward Freight Agreements. Our modelling implementation also includes the valuation of an option and risk measure calculation that depends on the evolution of dry bulk indices.

Freight rates, Markov regime switching, geometric Brownian motion, risk management

Performance Evaluation of Clustering Methods for Brain Cells' Single-Cell RNA-Sequencing Data

Single-cell RNA-sequencing (scRNA-seq) technology has rapidly evolved for measuring the expression level of thousands of genes at from individual cells. A main task in scRNA-seq is to group cells by their gene expression profiles. Facing challenges because of the number of cells that can be much smaller than of genes, and zero inflation due to noise and dropout events, various clustering algorithms have been proposed. However, the brain system has greater complexity with many different cell types, presenting a challenge to statistical and computational analyses. Performance evaluation of clustering methods applied to brain cell's scRNA-seq data is required. In our work, we evaluate the performance of four clustering methods (SC3, Seurat, GiniClust, and BackSpin) based on three independent data sets using a computational pipeline and running all analyses on Compute Canada via Sharcnet. In terms of V-measure and visualization plots, BackSpin and Seurat perform best, but with many inferred groups. SEURAT owns highest AIR values and reasonable purity values as it estimates more appropriate number of groups compared with BackSpin and SC3. In terms of purity, GiniClust is preferred because it leads to more comparable total clusters to the true totals.

Securing PDF Estimates from Differentiated CDF Approximants

A novel approach to density estimation is proposed. First, the end points of the distribution are determined from an initial estimate that follows a four-parameter beta distribution. Next, a precise polynomial approximation to the distribution function is obtained from the adjusted empirical distribution function. Actually, moment-based and least-squares approximating polynomials are respectively utilized for the univariate and bivariate cases. Finally, these polynomials are differentiated and the resulting function is normalized in order to secure bona fide density estimates. What is more, any given quantile of the distributions can readily and conveniently be evaluated from the proposed polynomial representations of the distribution functions. An alternative density estimation methodology relying on bivariate Hermite orthogonal polynomials is also presented. As well, several illustrative examples are provided.

Exploring Demanding Wildland Fires and their Proximity to Human Interfaces

Wildfire has long been an important part of ecosystems, but large wildfires can be devastating. Thus, it is important to understand what conditions lead to "Demanding" fires. Our research is built on a conditional framework: given that fires have been reported, what is the probability that these fires would become "Demanding", which is defined as the fire exceeding size and lifetime thresholds. We use classification approaches such as Logistic Regression and Random Forest for our modelling. We explore the information on fires that were predicted by our modelling to be more difficult to control. We fit our model to training data for 10-year period from 2005-2014, and then use it to make predictions on more recently observed testing data. By combining the results of classifications and spatial analysis, we investigate which recently observed fires had a high chance of becoming a demanding fire, focusing on where they occurred relative to wildland urban, wildland industrial, and wildland infrastructure interfaces, which have dense property value and thus are of higher concern. Our data suggest that near-miss fire distributions do have spatial patterns, and that there are a large number of near-miss fires in WUI regions, notably in the Wildland-Infrastructure Interface, necessitating increased fire risk management concern.