Science Atlantic Mathematics, Statistics, and Computer Science Conference 2021

Friday 22 October 2021 - Sunday 24 October 2021

St. Francis Xavier University

ST. FRANCIS XAVIER UNIVERSITY

Book of Abstracts

Introduction

The Mathematics, Statistics and Computer Science, Science Atlantic Conference is an annual regional event and will be hosted virtually by St. Francis Xavier University, October 22nd to 24th. This year Science Atlantic is celebrating its 60th anniversary as an organization and this annual conference, designed specifically to attract the interest of undergraduate students, dates back to 1978.

The conference provides a venue at which students and faculty can network, present research and share ideas and features two competitions (mathematics and computer science), undergraduate research talks and three named keynote lectures. The Blundon lecture (mathematics), the Fields lecture (statistics) and the Sedgwick lecture (computer science) are 50 minutes each and are designed to engage and educate the undergraduate audience on these fields and how they interact.

This year the competitions will be held in several remote locations on Saturday morning. Saturday afternoon and Sunday morning will be highlighted by undergraduate research talks and plenary speakers. Graduate research talks and the award ceremony will take place on Sunday morning.

Plenary Speakers and Abstracts

Blundon Lecture: Dr. Caroline Colijn

Abstract: COVID-19 modelling and applications in the pandemic

Mathematical modelling has been highly prominent during the pandemic. In this talk I'll describe what COVID-19 models are (at the population level), how they work, and I will introduce several models describing COVID-19 transmission. I will end with perspectives on the next steps for the virus and for our response.

Biography

Dr. Colijn's work is at the interface of mathematics and the epidemiology and evolution of pathogens. She holds a Canada 150 Research Chair in Mathematics at Simon Fraser University for Evolution, Infection and Public Health.

She leads the Mathematics, Genomics and Prediction in Infection and Evolution (MAGPIE) research group, which, alongside their own research, provides public health support with pandemic modelling. Their work involves case-forecasting, vaccination parameter estimation, genomic epidemiology and other topics in relation to COVID-19. Professor Colijn is the co-lead of the new Canadian Network for Modelling Infectious Disease (CANMOD). This collaboration between Public Health Agency of Canada and NSERC increases Canada's capacity for infectious disease modelling to directly support short, medium, and long-term public health decisions. It builds and coordinates national capacity by sharing research problems, models and estimates, data files and expertise. Throughout the pandemic, she has advised and collaborated with government and public health institutions about COVID-19. In 2020, Professor Colijn, was announced as a recipient of the Radio Canada Scientist of the Year prize for 2020 alongside three other mathematicians for her work to model the impact of physical distancing on the infection curve in British Columbia.

Professor Colijn develops mathematical tools connecting sequence data to the ecology and evolution of infections. She also has a long-standing interest on the dynamics of diverse interacting pathogens. For example, how does the interplay between co-infection, competition and selection drive the development of antimicrobial resistance? To answer these questions, her group is building new approaches to analyzing and comparing phylogenetic trees derived from sequence data, studying tree space and branching processes, and developing ecological and epidemiological models with diversity in mind. She is a founding member of Imperial College London's Centre for the Mathematics of Precision Healthcare. She received her PhD in mathematics from the University of Waterloo.

Sedgwick Lecture: Dr. Benjamin Haibe-Kains

Abstract: The Hard Path to Transparency and Reproducibility in Cancer Bioinformatics

One of the main challenges in precision oncology consists of developing predictors of drug response to select the most beneficial therapy for each individual patient. In this context, preclinical models are crucial to study the association between molecular features of tumor cells and response to chemical perturbations. However, only few predictors of drug response have been successfully translated to clinical settings. Such a low success rate is due not only to the complexity of the mechanisms underlying anticancer drug response, but also to multiple factors that can be controlled in the research settings. These factors include the inevitable noise in high-throughput biological experiments and the ever-increasing sophistication of the analytical pipelines used to develop predictors of drug response. In this presentation, I will present our attempts to characterize experimental noise, account for it in the predictive modeling and how new software platforms can be used to improve transparency and reproducibility in cancer Bioinformatics.

Biography

Trained as a computer scientist, Dr. Benjamin Haibe-Kains earned his PhD in Bioinformatics at the Université Libre de Bruxelles (Belgium). He was a postdoc in the Quackenbush group at the Dana-farber Cancer Institute and Harvard School of Public Health (USA). Dr. Haibe-Kains started his own laboratory at the Institut de Recherches Cliniques de Montréal (Canada) and he is now Principal Investigator at the Princess Margaret Cancer Centre. His research focuses on the integration of high-throughput data from various sources to simultaneously analyze multiple facets of diseases, with a particular emphasis on cancer. Dr. Haibe-Kains and his team are using publicly available genomic datasets and data generated through his collaborations to better understand the biology underlying carcinogenesis and to develop new predictive models in order to significantly improve disease management. Dr. Haibe-Kains' main scientific contributions include several prognostic gene signatures in breast cancer, subtype classification models for ovarian and breast cancers, as well as genomic predictors of drug response in cancer cell lines. Dr. Haibe-Kains has published more than 150 peer reviewed publications with a high citation impact of 23830 citations.

Fields Lecture: Dr. Luke Bornn

Abstract

In this talk I will explore how players perform, both individually and as a team, on a basketball court. By blending advanced spatio-temporal models with geography-inspired mapping tools, we are able to understand player skill far better than either individual tool allows. Using optical tracking data consisting of hundreds of millions of observations, I will demonstrate these ideas by characterizing defensive skill and decision making in NBA players.

Biography

Dr. Luke Bornn is recognized as a world leader in sports analytics, and is Co-Founder and Chief Scientist at Zelus Analytics, a world-leading sports analytics company providing sports intelligence to professional teams. Dr. Bornn was Vice President, Strategy and Analytics for the NBA Sacramento Kings and served as Head of Analytics for A.S. Roma of the Italian Serie A Football League, where he worked closely with managers, coaches and sports scientists to measure and evaluate athletes and performance. In addition to his work with soccer and basketball teams, the British Columbia native has previously held tenure-track professorships in Statistics at both Harvard University and Simon Fraser University. Bornn is a frequent contributor to the field of sports analytics, authoring research articles for the Journal for Quantitative Analysis, the Annals of Applied Statistics and the Journal of the American Statistical Association amongst others. His academic research is focused on developing statistics and machine learning methods for high dimensional spatio-temporal data, with a primary focus on extracting insights from player tracking data in sports. He was a finalist for the MIT SSAC research awards from 2014 through 2019, receiving the award in 2015 and 2019. He received his M.S. and Ph.D. in Statistics from the University of British Columbia.

Industry Lecture: Dr. Nithum Thain

Abstract: Machine Learning Fairness

Fairness is a fundamental consideration in the design and training of machine learning algorithms. We see time and again how even well-intentioned systems designers can inadvertently build bias into their ML models. In this talk we will introduce some of the concepts of the growing field of ML Fairness. We put ourselves in the shoes of someone building an ML algorithm to diagnose disease, and by stepping through the training process, we see how fairness issues might arise. We introduce some of the terms and techniques of fairness before stepping back and discussing the current state of the field.

Biography

Dr. Nithum (Nith) Thain is a Senior Research Engineer for Google and is recognized as an expert in Artificial Intelligence. Dr Thain holds numerous degrees in Mathematics from Queen's University (BSc) and McGill (MSc, PhD), along with an MBA from Oxford as a Rhodes Scholar. He completed a postdoctoral fellowship at Simon Fraser University, where he worked on cutting-edge algorithms to quickly classify tuberculosis strains using minimal genetic data. He subsequently joined Jigsaw, a unit within Google that explores threats to open societies, and builds technology that inspires scalable solutions. There, he did research on using natural language processing techniques in machine learning to help tackle online abuse. At Google Brain, he has worked on issues of AI Responsibility and Fairness. He has also taught Natural Language Processing with Deep Learning at the UC Berkeley School of Information and has advised numerous start-up companies on machine learning and data science approaches to assist with business decision-making and analytics. Since leaving Oxford, Nithum has done all of this while mainly residing in Newfoundland, the province where he grew up. Thain is a frequent contributor to the field of Machine Learning, Natural Language Processing, Artificial Intelligence, and Game Theory, authoring numerous articles, including proceedings in the International Conference on World Wide Web and the Association for Computing Machinery Conference on AI, Ethics, and Society.

Contents

CS Undergraduate / 3

Metaheuristic Exploration via Deep Learning Object Localization

Author: Mengyu Zang[1]

Co-author: Antonio Bolufe-Rohler [1]

[1] *University of Prince Edward Island*

Heuristic and metaheuristic optimization algorithms have enjoyed success as the method of choice for solving many real world problems due to their flexibility and speed. Optimization requires that a metaheuristic perform both exploration and exploitation. The role of exploration is to find the most promising region of a search space; recent publications have shown that exploration is the most critical and challenging part of the optimization process. In this research we present an entirely new approach to exploration based on the use of deep learning.

We have successfully represented an objective function as a 2-dimensional image; created two different datasets using the point where the global optimum is located as the target attribute to be predicted. We designed, trained and tested different deep neural network convolutional models for predicting the region where the optimum is located, and we have performed extensive experimentation to illustrate the effectiveness of this approach.

Math Undergraduate / 20

Distinguishing number of graphs generated by the symmetry groups Dn

Author: Meredith Cudmore-Keating[1]

[1] *St. Francis Xavier University*

The distinguishing number of a graph is the smallest number d such that a colouring (not necessarily proper) on the graph with d colours can only be preserved with the trivial automorphism. A group G is 2-generated if it contains two elements which can generate all of G. The 2-generated graph of G is the graph whose vertex set is those elements of G which can 2-generate it, and with two vertices adjacent if they 2-generate G. These graphs have a high level of symmetry, so the question of their distinguishing number is of interest. This talk explores work to date on the 2-generated graph of the symmetry group D_n and their associated distinguishing numbers.

CS Undergraduate / 6

Utilizing machine learning components to better optimize the Particle Swarm Optimization algorithm.

Author: Essayas Kassa[1]

Co-author: Antonio Bolufe-Rohler [1]

[1] *University of Prince Edward Island*

Optimization techniques are increasingly being used in research to improve algorithms. A good example of an optimization algorithm is part of the powerful family of a swarm-based algorithm called Particle Swarm Optimization (PSO). PSO was initially inspired from the collective behavior of social animals. PSO finds a solution to the optimization problem by using nodes that act like swarm particles which flow through the data driven by their own and neighbors' best performance. In the presentation I will be talking about how we utilized machine learning components to better optimize the Particle Swarm Optimization algorithm in making decisions.

Math Undergraduate / 13

Reconfiguration Graphs and Dominating Sets

Author: Amanda Porter[1]

Co-author: Margaret-Ellen Messinger [1]

[1] *Mount Allison University*

We study k-domination reconfiguration graphs based on the token/replacement model first proposed by Haas and Seyffarth. Every vertex represents a dominating set in a graph of size k or less and two vertices, X, Y are connected if and only if Y can be obtained by removing or adding a vertex to the dominating set X. We provide a complete characterization for Eulerian reconfiguration graphs when $k = |V(G)|$. We also consider what k values that allow Eulerian k-domination reconfiguration graphs for multipartite graphs.

Math Undergraduate / 18

Red, Blue, Green Poset Games

Author: Alexander Clow[1]

[1] *St. Francis Xavier University*

This talk examines Red, Blue, Green (partizan) poset games under normal play. Poset games are played on a partially ordered set where players take turns choosing an element of the partial order and removing every element greater than or equal to it in the ordering. The Left player can choose Blue elements (Right cannot) and the Right player can choose Red elements (while the Left cannot) and both players can choose Green elements. Red, Blue and Red, Blue, Green poset games have not seen much attention in the literature, do to most questions about Green poset games (such as CHOMP) remaining open. This talk will present generalizations of the results I presented at Science Atlantic last year about Green poset games (which I proved with Dr. Stephen Finbow, StFX) to Red, Blue, Green poset games as well as survey some distinct results that come out of my ongoing collaboration with Dr.Neil McKay (UNBSJ).

CS Undergraduate / 9

Multilingual Phishing Email Detection Using Federated Learning

Author: Dakota Staples[1]

Co-author: Saqib Hakak [1]

[1] *University of New Brunswick*

Phishing emails plague our modern-day society and with every year only continue to become more of a problem. There are many research studies showing how to detect phishing emails automatically and many tools already exist to do so. These models are very accurate at detecting phishing emails as well, with high precision rates and low recall and false positive rates. However, they contain issues with users' privacy and with multilingual detection. This research aims to define a new proof of concept model. The model this paper proposes not only detects phishing emails, but protects the user's privacy, and detects phishing emails in multiple languages. To achieve this, we propose a federated learning model. Different machine learning models will be trained and tested using existing phishing-based multi-lingual datasets by identifying unique features. In case there is no such dataset available, phishing emails written in different languages will be collected from large publicly available datasets such as PhishTank, and other online forums and websites. Finally, using the concept of federated learning, a machine learning model with best results will be selected and assigned the role of global model. Different edge devices (configured to detect phishing emails in French and English languages respectively) will download the global model and keep on improving the learning process of the global model by periodically sending the updates.

Math Undergraduate / 8

Infinite Fibonacci Trees and other Binary Trees Generated by Linear Operations.

Author: Jesse Preston[1]

Co-author: Karyn McLellan [1]

[1] *Mount Saint Vincent University*

Using two operations on an initial value can create an infinitely growing set of values. When representing the growth of these sets as a binary tree, it is seen that some operations cause an infinite Fibonacci tree to form.

We will prove which operations can be used to create a set of values including all integers, which of these operations will form an infinite Fibonacci tree, and why certain operations fail to accomplish either.

The structures and patterns behind these binary trees will also be analyzed, as there are interesting differences in the trees when changing the operations slightly.

CS Undergraduate / 11

The Practical Efficiency of Regular Expression Membership Algorithms

Author: Justin Gray[1]

Co-author: Stavros Konstantinidis [1]

[1] *Saint Mary's University*

Regular expressions encode text patterns and define languages of symbolic words. The membership problem decides if a given word is an element of the language described by a given regular expression. This problem has various well-studied algorithms, but current research only shows asymptotic complexity and performance with respect to samples of randomly generated regular expressions. Our research aims to answer how the algorithms perform when using practical regular expressions used in the real-world on a representative test set of words.

A set of compatible regular expressions have been collected from public GitHub repositories. Each compatible expression (i.e., no backreferences, lookaheads, or improper formatting) is then converted into an equivalent unambiguous mathematical representation. For each distinct expression, we have tested Thompson, Glushkov, position, follow, and partial derivative NFA constructions, as well as partial derivatives and exponential backtracking directly on the regular expression tree. These algorithms have been implemented into a modified version of the Python's FAdo package and include UNIX-inspired extensions such as character classes, the wild dot, and partial Unicode support. Preliminary results indicate that despite the backtracking algorithm being exponentially bound, its practical efficiency is faster than any other tested method.

Math Undergraduate / 14

Constructive Analysis in Agda

Author: Zachary Murray[1]

[1] *Dalhousie University*

Formal proofs are both impractical to write and difficult to verify by hand. The automation tools of proof assistants can make formal proofs convenient to write while verification is left to the computer. However, most mathematics, especially analysis, is unformalized, leaving a large gap between modern research and computerized proof. We will examine the practicality of formal mathematics in the Agda proof assistant and discuss my work on an Agda constructive analysis library.

CS Undergraduate / 12

A Novel Approach to Modelling Risk Factors for Mass Murder

Authors: Ethan C. Draper[1]; Ethan S. Heavey[1]

Co-authors: Margo C. Watt [1]; James A. Hughes [1]

[1] *St. Francis Xavier University*

Mass murders, while devastating, are relatively infrequent events. Their intermittence, however, poses a great challenge for researchers to develop accurate tools for risk assessment. While caution must always be exercised in active-shooter situations, the development of tools to predict on-scene outcome has the potential to inform law enforcement on appropriate actions and safety precautions. These risk assessment tools predict one of two potential outcomes: arrest or death (which may be by suicide or by law enforcement). Assessing the outcome of an act of extreme violence involves examining both static (e.g., age, sex, personal history) and dynamic (e.g., substance use, mental disorder symptoms) factors. Many static and dynamic risk factors have been examined independently, but their interplay has not yet been empirically studied. The aforementioned risk factors were analyzed with tools from the burgeoning field of Computational Social Science using a dataset of mass shooters from The Violence Project (TVP) (Version 3). TVP considered over 100 variables of interest from mass shootings occurring in America between 1966-2020, which was supplemented with Canadian cases from the same time frame. Forward Selection, Backward Selection, Recursive Feature Elimination, and Principal Feature Analysis methods were used to determine the risk factors with the greatest predictive validity of on-scene outcome. Using these risk factors, the researchers created a Shallow Neural Network model as a novel tool to help professionals predict on-scene outcome in active mass shooter cases. This model may also provide insight for future work in violence prediction and thus prevention.

Math Undergraduate / 5

Exceptional Gegenbauer Polynomials and Confluent Darboux Transformations

Author: James Munday[1]

[1] *Dalhousie University*

Classical orthogonal polynomials are families of orthogonal polynomials that arise as solutions to Sturm-Liouville eigenvalue problems, and are entirely classified by the classical families of Hermite, Laguerre, and Jacobi polynomials. Exceptional orthogonal polynomials also arise as solutions to Sturm-Liouville eigenvalue problems; however, we allow the polynomial sequences to miss a finite number of "exceptional" degrees, resulting in new orthogonal polynomials that are generalizations of the classical families. This talk will introduce a construction for multi-parameter exceptional Gegenbauer polynomials via the isospectral deformation of the classical Gegenbauer operator. We are able to obtain a fully explicit description of the operators and polynomials in question through the use of confluent Darboux transformations.

CS Undergraduate / 16

Security in the World of IoT

Author: Mohammadhossein Moghaddas Jafari[1]

[1] *Memorial University of Newfoundland*

The Internet of Things (IoT) has been a part of the disruptive technologies and has impacted many areas in our life. It has been described as 'the next major computing paradigm' because of its convenience and effectiveness in virtually every area of our daily lives. However, the IoT is also one of the most vulnerable technologies to cyberattacks, with many security researchers even predicting that IoT security will be the next 'bug' in computer systems. As sensitive and private information is exchanged between connected devices, privacy becomes a major concern. Among many important issues transparency and reliability are considered as new challenges that differentiate IoT from the conventional Internet and industries. This paper focuses on security threats on IOT technology and how to minimize them. We examine existing studies and suggested solutions to increasing privacy issues from a variety of perspectives in this study to identify dangers and mitigations. We assess the privacy difficulties and concerns that arise in IoT systems. We outline potential IoT solutions that address a wide range of privacy concerns.

CS Undergraduate / 19

Improving The TLS Client Puzzle Extension's Design Against DDoS Attacks.

Author: Emmanuel Blay[1]

[1] *University of Prince Edward Island*

The continuous enhancement of the various security components of the internet is what helps us navigate the quickly evolving threat landscape that presents several vulnerabilities and bad actors who are working hard to exploit the corresponding attack vectors of which a Distributed Denial of Service (DDoS) attack is a part of. The recent Transport Layer Security (version 3) Client Puzzle extension proposed by Thomas Nygren introduces the use of a client puzzle or proof-of-work concept, as opposed to the traditional rate-limiting techniques, to help stall massive influxes of client connections thereby preventing a denial of service on the part of the host server. This research aims at discussing how the initial design proposed by Nygren could be enhanced in terms of design, performance, and overall efficiency. In this paper, we discuss how we can utilize the design of Finite State Machines or Finite State Automaton to draw up a system for improving the storage of client puzzles, and by extension, the deployment of puzzle payloads and solution verification. We'll be utilizing a library called GRAIL that allows us to define and manipulate finite languages and their corresponding finite-state automata. This library allows us to concatenate a new finite-state automata to another finite-state automata, generate all the words in the regular language defined by a given finite-state automaton, and verify that a given expression belongs to the regular language defined by a finite-state automata. With these mechanisms in place, this research will be proposing a much more resource-efficient way of generating puzzles such that the host server does not get computationally burdened during an attack. We will be summarizing the research with a performance test on a couple areas of interest that would prove that this new scheme of serving client puzzles during DDoS attacks is indeed efficient.

Math Undergraduate / 21

Geophysical Inversion For Flood Risk Management

Authors: Eli Vandenberg[1]; Heidi Hebb[1]

Co-author: Peter Lelievre [1]

[1] *Mount Allison University*

Many regions across Atlantic Canada rely on dykes and other earthen flood barriers to support agricultural activities and to protect millions of dollars of commercial and residential infrastructure. Many dykes are in increasingly poor condition from repairs, compaction, erosion and other factors. Rising sea levels from climate change presents an ever increasing hazard. Hence, there is a growing concern for the health of these structures and their ability to resist breaching and flooding in the future. We are using geophysical inversion to image the internal structure and composition of these dykes. Inversion is a computationally intensive procedure, relying on accurate numerical solution of the differential equations that describe the physical phenomena involved, and development of numerical optimization routines tailored to the specific inverse problem at hand. Thorough procedures to use geophysical inversion to accurately image the internal composition of dykes have not been thoroughly explored and tested. We are investigating an integrated imaging approach, which combines the use of electrical resistivity tomography (ERT) and electromagnetic (EM) induction methods. Both produce data that is dependent on the conductivity of the earth being surveyed, which is an indicator of the salinity of the fluid in the soil, and therefore the data can provide information about the flow of saltwater through the dykes and about possible erosional features. The two types of data are inverted to produce candidate earth models that could have given rise to the measured data, and which can be interpreted to indicate information about dyke structure and composition. We are investigating whether or not this combination of survey methods, when combined with numerical inversion practices which incorporate sensible a-priori information, will produce a monitoring method that is accurate, time and cost effective, and sustainable.

Math Undergraduate / 24

An Analysis of Population Persistence in Y-Shaped River Networks

Author: Dylan Smith[1]

[1] *Memorial University of Newfoundland*

We study the population dynamics of aquatic organisms growing logistically in a Y-shaped river network. The habitat is viewed as a tree-like metric graph with the population density satisfying a reaction-diffusion-advection (RDA) equation on each segment along with the appropriate junction and boundary conditions. Steady states can be viewed as solutions of a system of two first order ODEs (subject to appropriate boundary conditions). Geometrically, they are represented by orbits in the phase plane, generated by the corresponding flow operator. Any positive steady state in a Y-shaped network is represented by a certain configuration of three orbits in the phase plane.

We are interested in determining the minimum size of a Y-shaped river network with a fixed geometry (specified by the ratios of its segments) for which the persistence of a population is possible. For a single river stretch (with hostile or outflow downstream boundary conditions), explicit formulas are known for such minimum length.

In this talk we will give typical examples of orbit configurations and discuss the dependence of minimal network length for persistence on the geometry of the network.

Stats Undergraduate / 17

Intensity-Duration-Frequency curves for dependent datasets

Author: Anas Boukili Makhoukhi[1]

Co-authors: Salah-Eddine Adlouni [1]; Wafaa El Hannoun [1]

[1] *Université de Moncton*

According to the latest report of the Intergovernmental Panel on Climate Change (IPCC), Atlantic Canada is part of the region in the world where extreme rainfall is expected to increase in intensity and frequency. Based on probabilistic approaches, the intensity-duration-frequency (IDF) curves are a major decision-making tool for the engineering design of various water resources infrastructure (urban and agricultural drainage systems, bridges, dikes, etc.). In unlikely hypotheses, the most IDF curves commonly used assume the Gumbel distribution for the annual maximum intensities of different durations and independence between all these durations.

In this study, the question of estimating IDF curves is treated according to two different approaches. A classic univariate analysis, under the assumption of independence between the different durations, but rejecting Gumbel's distribution in favor of the more general GEV one. The second part corresponds to a multivariate analysis using D-Vines copulas. This modeling makes it possible to consider the dependence between the sub-hourly measurements, of major interest in urban hydrology. In our case study located in Moncton, the quantiles of the sub-hourly durations estimated by the dependency model are higher than those estimated marginally by GEV or Gumbel distribution. Consequently, this result shows the significant effect of the choice of the marginal distribution and considering the dependence between the durations, on the IDF curves. The models offered are flexible and can be applied to different weather stations.

Stats Undergraduate / 15

Multiple Observers Ranked Set Samples for Shrinkage Estimators

Author: Andrew Pearce[1]

Co-author: Armin Hatefi [1]

[1] *Memorial University of Newfoundland*

Ranked set sampling (RSS), as a powerful data collection technique, is used for situations where measuring the study variable requires a costly and/or tedious process while the sampling units can be ranked easily (e.g., osteoporosis research). In this paper, we develop ridge and Liu-type shrinkage estimators under RSS data from multiple observers to handle the collinearity problem in estimating coefficients of linear regression, stochastic restricted regression and logistic regression. Through extensive numerical studies, we show that shrinkage methods with the multi-observer RSS result in more efficient coefficient estimates. The developed methods are finally applied to bone mineral data for analysis of bone disorder status of women aged 50 and older.

Keywords: Ranked set sampling, Multiple observer, collinearity, Ridge estimator, Stochastic restricted regression, Logistic regression.

Math Undergraduate / 22

A Piecewise Linear Spline Maximum Entropy Method for Frobenius-Perron Operators of Multi-dimensional Transformations

Author: Adam Smith[1]

Co-author: Shafiqul Islam [1]

[1] *University of Prince Edward Island*

In 1976, Mathematician Tien-Yien Li published a solution to a conjecture by Ulam concerning a finite approximation to the Frobenius-Perron operator. This accomplishment was a breakthrough for the numerical approximation of the invariant densities that describe the statistical behaviour of dynamical systems. Twenty years later, Jiu Ding and Aihui Zhou extended this method to multi-dimensional transformations. Since then, several different methods have been developed to approximate these invariant densities. Here, we present a piecewise linear spline maximum entropy method for the approximation of invariant densities corresponding to multi-dimensional transformations. Applications are considered and numerical results are explored.

Graduate and Contributed / 4

A Pursuit Evasion Game on the Plane

Author: Mehdi Salimi[1]

[1] *St. Francis Xavier University*

In this presentation, we review one of the most old and classical pursuit evasion problems. Then we study a pursuit game in the plane. Controls of players satisfy on the integral or geometric constraint. We introduce a winning strategy for the pursuer and show that using this strategy the pursuer catches the evader.

Graduate and Contributed / 10

A molecular vector embedding representation for enhanced toxicity prediction of Tox21

Author: Mohammed Kassab[1]

Co-author: Othman Soufan [2]

[1] *Department of Computer Science, St. Francis Xavier University, Nova Scotia, Canada*

[2] *St. Francis Xavier University*

Every day, we manufacture a myriad of new chemicals for pharmaceutical use. While it is possible to design a chemical to have a particular cellular effect, it is difficult to ensure the overall safety of the new substance to the living organism in comparison. To ensure the overall safety of the chemical, in-vivo and in-vitro chemical tests or assays are implemented, which among other considerations, are slow, expensive, and arguably unethical. The advancement of artificial intelligence provided a promising alternative in the form of in-silico tests that utilize large amounts of data to predict the toxicity of a substance over many assays instantly.

In this study, we process Tox21 data comprehensively and prepare toxicity prediction information for 9000 chemicals coming from 1500 assays. This makes for more than 13.5 million potential interactions. Traditionally, similarity-based approaches [SVM, kNN] have been successfully utilized to detect toxicity, especially in combination with various fingerprinting techniques. Recently, we showed that an ensemble of shallow models can achieve a macro F1 score of 0.8048. However, with the breakthroughs in deep learning and the growth of available data, a group of researchers won the Tox21 challenge using a deep neural network (DeepTox) with an extremely large number of fingerprint-based features. Deep learning was also successfully used to embed chemicals into vectors (mol2vec). In this study we aim to explore the extent that learnable molecular embeddings can be used for toxicity classification.

Graduate and Contributed / 7

Option price approximation by polynomial functions

Author: Joy Liu[1]

[1] *Dalhousie University*

In this presentation we will use using polynomial approximations to approach the option pricing problem in finance.

First we use polynomial functions to approximate the payoff function in one-dimension. The payoff for a call option, as a function of the stock price at expiration S is given by $f(S) = \max(S - K, 0)$, where K is a constant called strike price. We approximate this piece-wise defined function with polynomials obtained using a least-squares criteria. For our implementation we use the polyfit function in MATLAB.

For the asset prices we assume a Geometric Brownian motion (also known as the Black-Scholes model). Then we use the same methodology to approximate prices of options that depend on two assets, such as spread options. For the spread option price approximation, we work under two different models: (i) a Geometric Brownian motion and (ii) a Geometric Brownian motion with Poisson jumps. We show some numerical results for the two models, and compare them to the corresponding results from Monte Carlo simulation.

Graduate and Contributed / 23

Comparing the Surrounding and Containment Numbers of a Graph

Author: Caleb Jones[1]

[1] *Memorial University of Newfoundland*

We analyze the surrounding and containment variants of the game of cops and robber on graphs. In an attempt to resolve the conjecture that $s(G) \leq \xi(G)$ for all graphs G, we develop several bounds on $s(G)$ and $\xi(G)$ in terms of graph parameters. We use a simple induction process to show that one can assume $s(G-u) \leq \xi(G-u)$ for all non-cut $u \in V(G)$ when trying to prove the conjecture true.