Queen's University at Kingston
Department of Mathematics and Statistics

Summer 2020 Research Projects

Check back for updates

Last updated: 30 January 2020

Supervisor: Troy Day
Project Title: The Sampling Distribution of Multivariate Mutual Information
Project Description: Information-theoretic ideas are increasingly being applied in the life sciences. One interesting recent application is to evolutionary biology. It has been shown that multivariate mutual information (MMI) provides a natural way to quantify how the non-independence among organismal traits constrains a population's ability to evolve. In applying this measure to real biological populations, however, one needs to know what magnitude of MMI is expected simply because of sampling error (i.e., its sampling distribution). In this project we will attempt to derive the sampling distribution of MMI under various biologically relevant conditions.
Student's Role: To master the background material required to understand the problem thoroughly and then to try various approaches for deriving the sampling distribution. Time permitting we will also examine the robustness of the results using simulated data and well as transcriptomic data.
Prerequisites (req'd background/level of study): Some background in probability. Some knowledge of statistics and information theory would also be beneficial.

Supervisor: C Devon Lin
Project Title: Discovering Hidden Features Using Gaussian Process Classification with Random Projection Pursuit Features
Project Description: Link prediction is a fundamental scientific problem in an online social network such as Facebook and LinkedIn. It aims to predict the probability that a new link will be established. Machine learning methods such as Naïve Bayes, support vector machines and decision trees have been used. However, these methods build models using only the observed features, and do not consider the impact of unobserved factors, for which we do not have data and could influence link establishments in a social network. We propose a new predictive model that incorporates the hidden features in the linkage establishment.
Student's Role: Conduct literature reviews; Understand several machine learning methods; Application of the proposed methodology to the LinkedIn data; Conduct simulation studies; Produce the research report.
Prerequisites (req'd background/level of study): Have taken courses in first and second math/stat core courses, linear regression, familiar with the statistics programming language R, and preferably computational data analysis and statistical learning courses.

Supervisor: Felicia Magpantay

Project Title: Mathematical modeling with imperfect/incomplete covariate information

Project Description: Mathematical modeling is a powerful tool to for studying complex systems. It allows us to test the consistency of our hypotheses on underlying mechanisms and examine informative structures in noisy data. A common issue that arises is incomplete and imperfect covariate information. For example, when fitting disease models to data on childhood infections, we may have long-term data on disease incidence but incomplete or imperfect data on vaccine coverage. One way to work with this is to treat vaccine coverage as another type of data to be fitted (instead of as an exact covariate for the model). This allows us to model errors in the measurements of these covariates. In this project we will test the use of clamped B-splines and stochastic bridges to solve these problems.

Student's Role: Students working on this project will spend the first weeks studying mathematical modeling using partially observed Markov processes. They will then learn to use the code I have prepared for implementing the algorithms described above and test them on both simulated and real data. Advanced students may also study the theory behind these methods and examine its limitations.

Prerequisites (req'd background/level of study): Undergraduate students applying to work on this project should have a theoretical probability background (MTHE/STAT 351 or equivalent) and have some basic familiarity with R.

Supervisor: Giusy Mazzone

Project Title: Linearization Principles for Systems of Ordinary Differential Equations

Project Description: This project will focus on the study of stability and long-time behaviour of solutions to nonlinear systems of ordinary differential equations (ODEs). It is well-known that for systems of ODEs which are "locally linear", information about the stability can be obtained by studying the spectral properties of a (suitably constructed) linear system of ODEs. Whenever this linear approximation is possible, we obtain what is generally known as "linearization principle". The aim is to develop general stability principles for systems which are not "locally linear" in the classical sense. For such system, for example, the coefficients matrix is not invertible and equilibria are not isolated, instead they form a (nontrivial) nth-dimensional manifold. It becomes then fundamental to provide a qualitative behaviour of the trajectories approaching such manifold of equilibria. Applications span from solid mechanics to fluid dynamics.

Student's Role: The student is expected to learn the classical stability principles existing in the literature, have regular meetings with the supervisor to discuss the material, and apply the theory to real-world problems

Prerequisites (req'd background/level of study): MATH/MTHE 280, MTHE 237 or MATH 231

Supervisor: Catherine Pfaff
Project Title: Geodesics in Culler-Vogtmann Outer Space
Project Description: Culler-Vogtmann Outer Space is a single object encoding all possibilities for finite graphs with lengths on their edges. We would be building diagrams encoding distance minimizing paths through this object (efficient ways to deform the graphs).
Student's Role: Learning about Culler-Vogtmann Outer Space & building the encoding diagrams
Prerequisites (req'd background/level of study): Ability to code, group theory, real analysis

---

Supervisor: Yanglei Song
Project Title: Contextual multi-armed bandit problem
Project Description: This project focuses on the multi-armed bandit problem with context, which is illustrated using the following example. Consider multiple treatments for a disease, whose efficacy depends on patients' characteristics (context), such as genes. As a new patient arrives, based on his/her context and past knowledge, the doctor needs to select a treatment, of which the outcome is observed. The doctor may attempt to assign the "best" treatment based on the collected information so far, but also need to explore each treatment sufficiently to make informed decisions. This trade-off between exploitation and exploration lies at the heart of multi-armed bandit problems. The goal is to minimize the regret against an oracle, who knows how the outcome depends on context and treatment. We will in particular consider the case where the context may be high dimensional, and explore/improve existing methods.
Student's Role: The student will first conduct a literature review, and discuss the materials with the supervisor through regular meetings. Then the student will implement and compare several existing methods, and write a final report.
Prerequisites (req'd background/level of study): A strong background in probability/statistics (at the level of STAT 351/353); familiarity with at least one programming language.

Supervisor: Glen Takahara with Wesley Burr (Trent)
Project Title: Gap Filling Time Series
Project Description: Many time series in practice contain gaps while many statistical procedures for analysing the data require regularly sampled data with no gaps. Thus, filling in gaps is a basic problem in time series analysis. Ideally, we want the interpolated data to have the same statistical properties as the observed data, or at least the statistical properties that we are interested in. A naive but very common gap filling procedure is linear interpolation. The pros are that this procedure is very simple to implement and for short gaps works well enough that any bias introduced into statistical procedures is negligible. The cons are that the linearly interpolated data are obviously statistically different from the observed data and does not work well for longer gaps.
Being a basic problem in time series analysis, many gap filling procedures have been proposed, most of which are based on emulating the statistical and/or deterministic structure in the observed data in some way. One approach that is less well explored is to use multiple correlated time series to gap fill in one of them. In this project the student will investigate:

1) the use of a time series with no gaps to help gap fill a second time series with gaps that is highly correlated with the first time series, and/or the use of two highly correlated time series, both with gaps, to help gap fill one another;
2) a comparison of gap filling procedures via simulation;
3) contribution to a general purpose R package for gap filling time series.
Student's Role: The student will write the final report in LaTeX.
Prerequisites (req'd background/level of study): Familiarity with R for statistical computing.

Supervisor: David Wehlau with Eddy Campbell
Project Title: The invariant theory of reflection groups
Project Description: Eddy Campbell and David Wehlau have found a numerical invariant that provides a new understanding of the invariant theory of subgroups of reflection groups. The algebraic objects involved in this study can be programmed in the symbol manipulating languages, Sage, Gap and Magma and our understanding of the numerical invariant thereby advanced. Using these computer languages, we may study representations of such subgroups over number fields, their rings of integers, and localizations at various primes. The project provides a student the opportunity to develop a deeper understanding of these concepts, as well as an understanding of research in mathematics.
Student's Role: Learn to program in Sage and/or Gap and Magma to perform calculations associated with some beautiful mathematics: how does the geometry of the representation of a subgroup of a reflection group affect the algebraic structure of its ring of invariants, and vice versa? Such questions are at the core of the subject of algebraic geometry. There is an existing collection of programs written in Magma by Eddy and David that will help guide progress.
Prerequisites (req'd background/level of study): A course in group theory and ring/field theory.

Supervisor: Serdar Yuksel
Project Title: Interaction of Control, Information, and Probability
Project Description: The interaction between control, information, and probability is a very general phenomenon that arises in many control and information transmission problems. On the one hand, a controller requires information on the unknowns that affect the operation of an underlying system; on the other hand, the quality of the relevant information itself is typically affected by the choice of the control actions, and the transmission of information over communication channels and the process of shaping the source output and recovering the transmitted signal can themselves be viewed as control design problems.
The goal of this project is to study the optimal design of information structures in stochastic control systems. Depending on student interest, we will focus on one of the following three main themes
- (i) Decentralized stochastic control and the geometric properties of information structures
- (ii) Optimal control under information constraints and optimal design of coding and control policies in networked control
- (iii) Optimal partially observable stochastic control problems and filter stability under incorrect priors or system models.

Student's Role: The expectation is that the student(s) would study a wide variety of resources, take part in the research discussions through regular meetings, and prepare a technical document at the end of the summer to report the findings. Implementations of some of the findings through Matlab or another software will also be likely.
Prerequisites (req'd background/level of study): A strong background in probability/statistics, real analysis and systems theory would be desirable. 3rd or 4th year students would likely be more suitable for the project, though this is not necessary.

Supervisor: Imed Zagiua with Francois Rivest (CISC, Queen's)
Project Title: Multi-Satellite Scheduling problem
Project Description: Each day, thousands of tasks are requested and satellite schedules must be generated to maximize the number of tasks completed. One can turn the problem into a directed topographical graph of valid transitions from one task to the next where it remains to find the path collecting the largest number of tasks. This would be a simple problem if it wasn't for the fact that each task only has some probability of success (because of weather). Therefore, once a task is assigned, the value it has at its second occurrence is discounted to the remaining probability of success. This makes the problem NP-Complete.
Student's Role: Reviewing the literature on the existing heuristics and methodologies (some using a mixture of heuristics and dynamic programming for example). Testing a number of new heuristics and algorithms we are developing. Collaborating with graduate students working on a deep learning approach to solve the problem. Possibly proposing new algorithms, heuristics, or formulation of the problem for deep learning.
Prerequisites (req'd background/level of study): Experience in Python programming, Ability to program basic search algorithms on graphs (eg, depth-first search, A*, or Dijkstra). Course on Graph Theory an asset. Experience with Deep Learning an asset.

OTHER SUMMER POSITIONS

This non-research project is conditional on funding.

Supervisor: Peter Taylor
Project Title: Rabbitmath—a new grade 12 math curriculum
Project Description: In summer 2020 we will be constructing a new kind of school math curriculum at the grade 12 level. The grade 11 curriculum is already in draft form and is being tested in a few Ontario classrooms. See www.rabbitmath.ca. We are looking for students who are interested in being part of the Grade 12 phase.
Student's Role: Generation of ideas, construction of examples and problem sets. Construction of Python activities and animations.
Prerequisites (req'd background/level of study): Good mathematical, coding, and writing skills. Good Imagination. A concurrent education student specializing in math or CS at the high-school level would be ideal.

The following project is being offered by Dr. Mohan Chaudhry of the Department of Mathematics and Computer Science at the Royal Military College of Canada. Students interested in applying to work on this project should contact Dr. Chaudhry directly by email <chaudhry-ml@rmc.ca> or phone (613-541-6000/6460):

Project Title: Solving functional equations and inverting transforms that arise in the study of Markov models

Project description: In recent years, discrete-time queues have gained importance due to digitization of signal pro-cessing devices, communication and computer networks. Such queues could be useful for the analysis and design of digital systems. Much interest in discrete-time queues has been generated due to their applicability in the Broadband integrated services digital network which is considered as a key technology in providing a variety of multimedia services.

Many of the analytic solutions in queueing and other stochastic processes are given in various transforms or in functional equations. Distribution of the number of customers served during a busy period for the late arrival system with delayed access (LAS-DA) as well as for the early arrival system (EAS) Geom/G/1 has been discussed by many researchers. Most of the researches mentioned there either give only the functional equation or using that equation they give a few moments. Further, whereas some researchers solve that equation numerically, others just give analytic results in terms of convolutions. The explicit analytic results are given only for simple cases such as Geo/Geo/1 and Geo/D/1. Further, their procedure cannot be extended to give results for more general cases such as Geom/NB/1, Geom/B/1, Geom/U/1, where NB, B, and U indicate, respectively, negative binomial, binomial and uniform distributions. We have recently solved the model Geom/G/1 which gives explicit results for various service-time distributions. We now want to apply the method that we have used to solve the model Geom/G/1 to more complex models such as GeomX/G/1 which involves bulk arrivals.

All the previous students who came to work for me have gone for graduate work and after completing graduate work as Master's and/or Ph. D. students, they are now working at various places.

Student role: The student's role will be to solve functional equations for models such as GeomX/G/1 and/or invert transforms of other queueing models using mathematical tools such as MAPLE/MATLAB or MATHEMATICA and QROOT, a software developed by us. While working for this, he/she will also do mathematical typing by using Latex and/or MS word. This will be a great learning experience for a student who is wishing to go for graduate work