THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Topics for STAT4798 Statistics and Actuarial Science Project (6 credits)
(Offered in both 1st and 2nd semesters of 2023 - 2024 for STAT4798)

1. Natural hedging between longevity and mortality risk

A life insurer faces longevity risk, which arises from a systematic improvement in future life expectancies and the subsequent increase in the actuarial value of their annuity liabilities. On the other hand, the life insurer may face mortality risk, which arises from a systematic decrease in future life expectancies and the subsequent increase in the actuarial value of their death benefit (life insurance) liabilities. To manage those two risks, the life insurer may seek an optimal balance between annuities and death benefit insurances. This project seeks to find an optimal mix. To obtain this mix, an insurance pricing strategy can be analyzed, or risk-sharing arrangements can be sought with an appropriate counterparty. The data is available open source, and the literature will be provided. Students are expected to have good knowledge in programming languages such as R.

References:

- Cox, S. H., & Lin, Y. (2005). Natural hedging of life and annuity morality risks. Journal of Risk and Insurance , 11, 1-15.

Supervisor: **Dr. T.J. Boonen,** tjboonen@hku.hk, Dept of Statistics and Actuarial Science

2. Mitigating Biases in Machine Learning Applications

Machine learning methods have achieved tremendous successes in many applications. However, it has recently come to people's attention that many machine learning data sets contain significant biases, and machine learning models trained on such data sets may further amplify existing biases. For example, the activity "cooking" is over 33% more likely to involve females than males in a training set, and a trained model further amplifies the disparity to 68% at test time. This amplification of biases can cause serious social and ethical issues, and therefore studying and developing machine learning methods that mitigate biases is of ultra importance. In this project, students will explore and study existing methods along this line of research and develop new methods to better mitigate biases and enhance equity in machine learning applications.

The target students are senior undergraduate students with a strong background in deep learning and python (PyTorch/TensorFlow) programming.

Supervisor: **Dr. Y. Cao,** yuancao@hku.hk, Dept of Statistics and Actuarial Science

3. Dependence Structures in Multiple Life Insurances and Annuities

The price of a multiple-life insurance/annuity product depends not only on the marginal distributions of the underlying future lifetimes, but also on their dependence structure. In this project, the effect of dependence structure on the actuarial present values will be studied. In the course of the research, the student will learn some basic theory of dependence structures.

Supervisor: **Prof. K.C. Cheung**, kccg@hku.hk, Dept of Statistics and Actuarial Science

4. Stock market forecasting and stock investment in practice

The stock market is known for its inherent volatility and complexity, making it a challenging environment to predict with certainty. To become a successful stock investor, one needs to have macro level understanding of the stock and financial markets, their trends and movements as well as micro level understanding of individual stocks, related businesses and accounting measures. This project aims to provide students with hands-on virtual experiences of stock investment. Students are expected to build the following three types of models. First, based on the macro level information only, the macro model that predicts the overall market movements and produces clear buying and selling signals in light of the long term market movement cycle. Second, based on the principles of value investment and the development cycle of the industries, the portfolio selection models that design combinations of stocks suitable for the purposes of short-, medium- and long-term investments. Third, trading models that provide guidance for daily, weekly, monthly and quarterly buying and selling. In- and out-of-sample validations should be conducted to verify the quality of these models, and real-world virtual trading of no less than one month should be conducted to test the profitability of the daily and weekly trading strategies.

Software required: R or Python, Excel

Supervisor: **Mr. Harrison Y.Y. Cheung**, hcheung4@hku.hk, Dept of Statistics and Actuarial Science

5. Brain Imaging Analysis with Statistical Learning

Brain imaging analysis has played a central role in understanding the functions of human brain. From Electroencephalography (EEG) to magnetic resonance imaging (MRI) and from MRI to functional MRI (fMRI), the advancement of brain imaging technologies has benefitted tremendously to the diagnosis and treatment of brain disease. In this project, students will learn to develop statistical machine learning models to analyze brain imaging data and make predictions for brain diseases. As imaging data is usually represented as tensors (or multidimensional arrays), tensor decomposition and tensor regression methods would also be studied. Students are expected to have good knowledge in programming languages such as Python or R.

Supervisor: **Dr. L. Feng**, lfeng@hku.hk, Dept of Statistics and Actuarial Science

6. Approximate Inference for Bayesian Models

Markov chain Monte Carlo (MCMC) methods are considered the gold standard for inference in Bayesian models. However, in modern settings like machine learning, large datasets and high-dimensional models have become the norm. This presents a challenge to MCMC, as it is inherently serial and computational demanding. As a result, alternative scalable approximate methods for Bayesian inference are being developed. These include variational Bayes, expectation propagation, Laplace's approximation, the Bayesian bootstrap, and others. The aim of the project is to investigate the application of these methods in complex settings and evaluate their respective merits and weaknesses.

Requirement: Experience in Python or R; familiar with Bayesian inference

Supervisor: **Dr. Edwin C.H. Fong**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

7. Semiparametric regression analysis of interval-censored data

Interval-censored data arise frequently in medical, financial, and sociological research, where the event of interest is only known to occur within a time interval. For example, in large cohort studies of chronic diseases, participants can only be examined periodically (e.g., every year), such that the disease onset is only known to occur between two successive examinations. In this project, students will study semiparametric regression models for various interval-censored data, and ideally, implement one estimation approach to solve a real data problem. Students are expected to have basic knowledge in survival analysis, and those who are familiar with at least one programming language are preferred.

Supervisor: **Dr. Y. Gu**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

8. Joint analysis of multiple types of data in health studies

Health studies usually involve multiple types of data, such as longitudinal, survival, recurrent event, and competing risks data. For example, data collected in cancer research include repeated measures of risk factors, time to tumor recurrences, time to death from cancer, and time to death from other diseases. Modeling each data type separately without accounting for their dependence would lead to biased estimation. In this project, students will study joint analysis of multiple types of data via shared random effects models and apply this technique to a real dataset. Students are expected to have basic knowledge in survival analysis and longitudinal data analysis, and those who are familiar with at least one programming language are preferred.

Supervisor: **Dr. Y. Gu**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

9. Variable selection with censored outcomes and missing covariates in high-dimensions

Recent technological advances have facilitated the collection of various high-throughput data, such as genetic and imaging data, which are important covariates in cancer and Alzheimer's disease research. For economic or logistic reasons, however, some covariates may not be collected for all subjects, resulting in missing data. In this project, students will study variable selection for a censored outcome with high-dimensional, potentially missing covariates through latent-variable models and penalized likelihood approach (e.g., adaptive LASSO). Depending on students' capability and interest, more challenging problems could be investigated further, such as informative censoring and missing not at random mechanism. Students are expected to have basic knowledge in survival analysis and high-dimensional statistics, and those who are familiar with at least one programming language are preferred.

Supervisor: **Dr. Y. Gu**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

10. Deep/machine learning meets survival analysis: a new risk prediction framework

Deep/machine learning techniques have been used extensively for prediction of binary or continuous outcomes. However, prediction of an event subject to censoring is much less studied, mostly due to the difficulty of unknown event times among censored subjects. This project investigates deep/machine learning methods for risk prediction of censored outcomes. Students will explore various learning techniques, such as neural network, support vector machines, and decision trees, and apply them to the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Knowledge and hands-on experience in deep and machine learning is required.

Supervisor: **Dr. Y. Gu**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

11. Estimating time-varying treatment efficacy against infectious diseases

In phase-3 clinical trials or observational studies of infectious diseases (e.g., COVID-19, HIV/AIDS), subjects typically receive the treatment at different times and may experience crossover, such that the treatment indicator changes over time. It is important to estimate the time-varying treatment efficacy while adjusting for other time-varying confounders, such as community transmission and disease incidence. When reinfections are possible, it is also important to take past infection history into account, as it can affect the estimation of the treatment efficacy. This project investigates this problem and provides an interesting application to a COVID-19 study. Students with basic knowledge in survival analysis and excellent programming skills are preferred.

Supervisor: **Dr. Y. Gu**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

12. Open-world object discovery with deep learning

Deep learning has achieved remarkable success in many tasks, even surpassing humans, for example in image classification. However, the success comes at the cost of intensively labeled data, e.g., ImageNet which contains over 1.2 million manually annotated images. When a trained classification model meets an image from an unseen class, it often mistakenly predicts the image as one of the seen classes with high confidence. In other words, current learning models struggle to handle open-world problems where there are unseen or unfamiliar objects. In this project, the students will study the open-world object discovery problem with deep learning and develop solutions to enable the model to deal with unseen or unfamiliar objects.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

13. Content Generation with Diffusion Models

Diffusion models have shown promising results in visual content creation, driving the intriguing development of new generation image generation platforms such as Stable Diffusion and Midjourney. Though encouraging advancements have been achieved, the full potential of diffusion models is yet to be discovered. Despite generating visually appealing images, diffusion models can also be applied to generate other types of content, such as 3D models and videos.

In this project, students will study and explore diffusion models for different content generation tasks, showcasing their versatility and potential for different types of content creation.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

14. Efficient Adaptation of Pretrained Large-scale Language Models

However, the training of such models is immensely expensive. These models usually comprise billions of parameters and necessitate Internet-scale training data. Therefore, it is intriguing to investigate their capabilities without requiring retraining, but with some efficient adaptation that allows them to accomplish other tasks they were not trained on.

In this project, students will examine existing large-scale language models, showcase their applications, and explore efficient adaptation methods for these models. This exploration will not be limited to text data but will also incorporate data from other modalities, such as visual data.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

15. Implicit Neural Representations

With the advance of deep learning in computer vision, implicit neural representations appear to be a novel way to parameterize all kinds of signals. Unlike the conventional discrete signal representations (e.g., images are discrete grids of pixels, 3D shapes are often discrete grids of voxels/point clouds/meshes, and audio signals are discrete samples of amplitudes), implicit neural representations parameterize a signal as a continuous function that maps the domain of the signal (i.e., a coordinate, such as a pixel coordinate for an image) to whatever is at that coordinate (for an image, an R, G, B color). In this project, students will study and develop methods for using implicit neural representations to process visual information, such as images and 3D shapes, with potential applications including image super-resolution and 3D shape reconstruction.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

16. Effective self-supervised learning with large-scale unlabeled data

The success of modern machine learning techniques is driven by large-scale datasets with human annotations. However, it is not possible to annotate a large-scale dataset for all possible tasks. Some tasks may require domain-specific expertise and there is no large-scale data available, for example, medical images for a rare disease. Self-supervised learning, which requires no human annotations, appears to be an intriguing direction. It aims at learning useful representations in an unsupervised manner, which can be effectively used for various downstream tasks like object recognition, detection, and segmentation in visual data. In this project, the students will study various self-supervised deep representation learning techniques and develop solutions for effective self-supervised learning with large-scale real-world unlabeled data.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

17. Benford's law

Benford's Law describes the probability mass function of the kth significant digit in certain datasets. The goal of this project is to understand the underlying probabilistic details and to show how this result can be used for fraud detection.

Requirement: Knowledge in R

Supervisor: **Dr. M. Hofert**, mhofert@hku.hk, Dept of Statistics and Actuarial Science

18. Empirical beta copulas

Empirical beta copulas are smooth nonparametric copula estimators for any multivariate dataset with continuous margins. The goal of this project is get familiar with these copulas and investigate their properties.

Requirement: Knowledge in R

Supervisor: **Dr. M. Hofert**, mhofert@hku.hk, Dept of Statistics and Actuarial Science

19. The rearrangement algorithm for variance-reduction

The rearrangement algorithm is an algorithm to numerically determining optimal rearrangements of samples from marginal distributions such that the sum of the corresponding random variables has minimal variance. The goal of this project is to investigate the algorithm's performance as a variance-reduction method in classical Monte Carlo simulation applications.

Requirement: Knowledge in R

Supervisor: **Dr. M. Hofert**, mhofert@hku.hk, Dept of Statistics and Actuarial Science

20. Numerically determining value-at-risk subadditivity

As a high quantile of a distribution function, value-at-risk is a widely used risk measure. The problem is that it can sometimes violate the widely accepted notion of diversification also known as subadditivity. The goal of this project is to derive a condition on the Monte Carlo sample size that allows one to determine numerically, for a given confidence level, whether value-at-risk is subadditive.

Requirement: Knowledge in R

Supervisor: **Dr. M. Hofert**, mhofert@hku.hk, Dept of Statistics and Actuarial Science

21. Random matrices under dependence

Random matrix theory typically assumes the random entries of a high-dimensional, square matrix to be iid. Of interest is then the empirical distribution of all eigenvalues of this random matrix, with applications to estimated covariance matrices. The goal of this project is to present various ways of introducing dependencies in random matrices and to numerically investigate their influence on the distribution of eigenvalues (of the random matrices or their corresponding covariance matrices).

Requirement: Knowledge in R

Supervisor: **Dr. M. Hofert**, mhofert@hku.hk, Dept of Statistics and Actuarial Science

22. Applications of unsupervised learning

Unsupervised learning aims at representing structure in the input data, often by means of features. The resulting features can be used as input for classification tasks or as initialization for further supervised learning. Traditional methods, including principal component analysis, factor analysis and independent component analysis and deep learning methods, including autoencoders and variational autoencoders are considered.

The objective of the project is to explore and compare various unsupervised methods.

- Literature review of various unsupervised learning in the recent years.
- Apply to a real data set to identify any hidden features.
- Conduct simulation of various scenarios and evaluate the efficiency of various methods.

Requirement: Knowledge in Python.
Knowledge in multivariate statistics and ANN.

Supervisor: **Dr. C.W. Kwan**, cwkwan@hku.hk, Dept of Statistics and Actuarial Science

23. Analysis of correlated zero-inflated count data

In many medical and public health investigations, the count data encountered often exhibit an excess of zeros, and very frequently this type of data are collected on clusters of subjects or by repeated measurements on each subject. For example, in the analysis of medical expenditure, members in the same family may exhibit some correlation possibly due to housing locality, genetic predisposition, similar dietary and living habit. Ignoring such correlation may lead to misleading statistical inference. This project will survey the models and methods in the literature and apply them to a real data set.

Requirement: Knowledge in R or Python.

Supervisor: **Dr. Eddy K.F. Lam**, hrntlkf@hku.hk, Dept of Statistics and Actuarial Science

24. Applications of Extreme Value Models

Extreme value theory concerns the behaviour of maxima or minima, and has been used extensively in areas such as finance, hydrology, engineering and meteorology where the occurrence of extremes may have catastrophic consequences. In this project, the student will learn the basic modelling techniques for data of extremes and will apply such models to data sets of practical interest. The emphasis is on conceptual understanding of the underlying theory and interpretation of the fitted models.

Requirement: The student should be competent in computer programming. Knowledge in or willingness to learn the R programming language is essential.

Supervisor: **Dr. David Lee**, leedav@hku.hk, Dept of Statistics and Actuarial Science

25. Resampling Methods for Regression

Recent years have found increasing use of resampling methods in regression studies. Examples include the paired bootstrap, the residual bootstrap, the wild bootstrap, random perturbation, bagging, etc. In this project we explore their potential applications in contemporary regression settings where statistical inference remains prohibitively difficult.

Supervisor: **Prof. Stephen M.S. Lee**, smslee@hku.hk, Dept of Statistics and Actuarial Science

26. Applications of Secure Blockchain Solution

In this project we begin with a review of the basic architecture for blockchain in Python. This includes state transition rules, method for creating blocks, mechanisms for checking the validity of transactions, blocks, and the full chain. Next, we will create new blocks from data, validate the new blocks and add them to the existing blockchain.

Security is of the utmost importance in any blockchain architecture, in this project we will discuss 3 popular verification methods: public key cryptography, digital signature algorithm and trusted time-stamping. Finally, we will construct practical blockchain solutions to current fintech problems.

Supervisor: **Dr. Eric A.L. Li**, ericli11@hku.hk, Dept of Statistics and Actuarial Science

27. Introduction to Quantum Computing Algorithms

First we begin with a basic understanding of quantum computing (QC). Then we move on to some popular QC algorithms, written in Javascript and Python. In addition to constructing these QC codes, we will also provide the meanings, purposes and theoretical bases of these QC codes.

The QC algorithms we will cover include: Deutsch-Jozsa Algorithm, Simon's Algorithm, Super Dense Coding, Period Finding, and Shor's Factoring Algorithm. The last one is particularly important in modern cryptography: given an integer which is a product of two distinct prime numbers, this algorithm finds one of its prime factors.

Supervisor: **Dr. Eric A.L. Li**, ericli11@hku.hk, Dept of Statistics and Actuarial Science

28. Statistical Inference for Tensor Data

Tensors have been used in many fields and have provided powerful applications in various practical domains. They generalize vectors and matrices and have been studied from different viewpoints. The study of tensor methods has a long history in statistics. In the era of big data, tensor data appear frequently in the forms of video data, spatio-temporal expression data, relationship data in recommending and mining, and latent variable models, from a vast range of statistical applications. However, the extension of methods for dealing with matrices to tensors is much more difficult than those from vectors to matrices. This project targets to several tensor-based statistical methods.

Supervisor: **Prof. G. Li**, gdli@hku.hk, Dept of Statistics and Actuarial Science

29. Deep Learning Approach for Stochastic Control Problems

A stochastic optimal control problem deals with uncertainties when making decisions to maximize or minimize an objective function. It is widely used in deriving the optimal trading strategy in the financial field and the optimal insurance strategy in actuarial science. However, the ``curse of dimensionality" can quickly rise when solving a high dimensional stochastic control problem (e.g., a portfolio with a bunch of stocks, bonds, and insurances). Although no rigorous proof exists, some studies show that the deep learning approach can effectively reduce the ``curse of dimensionality" phenomenon.

How to use a neural network to compute the optimal trading and insurance strategies for the high dimensional stochastic control problem? This is a promising direction worth of study.

You may need the following theories and techniques to conduct this research:
1. The basic theories of Mathematical Finance to model a portfolio optimization problem (like Financial Economics I & II).
2. Monte Carlo approach to simulate stochastic market scenarios.
3. Neural network algorithm to maximize or minimize an objective function.

Supervisor: **Dr. W. Li**, wylsaas@hku.hk, Dept of Statistics and Actuarial Science

30. Privacy preservation for federated learning in healthcare

Artificial intelligence (AI) approaches have shown great promise for augmenting clinical workflows. However, access to large quantities of diverse training data is needed to develop robust models. Notably, sharing data across institutions is not always feasible due to security and privacy concerns. As such, Federated Learning (FL) approaches allow for multi-institutional training of deep learning models without the need to share data. However, FL comes with security and privacy concerns as well. Specifically, the data insights exchanged during FL training can leak information about institutional data. In addition, the collaborative nature of the FL workflow can introduce new issues when there is a lack of trust among the entities performing the distributed compute.

In this project, the students will study the current privacy threats and associated threat mitigations for FL workflows. Students are also encouraged to design new and robust privacy preserving models for FL in healthcare.

Requirement: Knowledge in machine learning/deep learning, proficient in python (PyTorch/TensorFlow) programming.

Supervisor: **Dr. L. Qu**, liangqqu@hku.hk, Dept of Statistics and Actuarial Science

31. Diffusion models for medical image restoration and synthesis

Medical imaging is an essential element for biomedical research and has demonstrated tremendous success in a wide range of areas, such as disease diagnosis, monitoring, or treatment. However, most existing medical imaging equipment is often cost-prohibitive and not always accessible in clinic. Thus, it is it is crucial to develop methods to reconstruct/synthesize high-quality medical images from low-cost, facilitating doctors with high diagnostic image quality for diagnostic decision. Recently, Denoising Diffusion Probabilistic Models have achieved remarkable success in various image generation tasks compared with Generative Adversarial Nets (GANs). In this project, the students will study and explore the diffusion models and apply it to medical image restoration and synthesis.

Requirement: Knowledge in machine learning/deep learning, proficient in python (PyTorch/TensorFlow) programming.

Supervisor: **Dr. L. Qu**, liangqqu@hku.hk, Dept of Statistics and Actuarial Science

32. Tackling data heterogeneity challenge in federated learning

Federated learning is an emerging research paradigm enabling collaborative training of machine learning models among different organizations while keeping data private at each institution. Despite recent progress, there remain fundamental challenges such as non-convergence and the risk of catastrophic forgetting, particularly when dealing with real-world heterogeneous devices and non-IID (independent and identically distributed) data. In this project, students will have the opportunity to study the impact of data heterogeneity on federated learning performance. They will explore various techniques and strategies to mitigate the negative effects of non-IID data on model convergence and learning. Moreover, students are encouraged to design new and robust federated learning algorithms that can effectively tackle the challenges posed by non-IID data distribution across participating organizations. By engaging in this project, students will gain valuable insights into the complexities of federated learning and develop critical skills in designing and implementing advanced machine learning solutions in real-world, heterogeneous environments.

Requirement: **Requirement**: Knowledge in machine learning/deep learning, proficient in python (PyTorch/TensorFlow) programming.

Supervisor: **Dr. L. Qu**, liangqqu@hku.hk, Dept of Statistics and Actuarial Science

33. Cointegration in Financial Analysis

The goal of this project is to test cointegration in financial time series. Students are required to have basic understanding of cointegration and some knowledge of computer programming.

Supervisor: **Dr. C. Wang**, stacw@hku.hk, Dept of Statistics and Actuarial Science

34. A Study on Capital Allocation Principles using Risk Measures

Risk measures and capital allocation principles are popular topics in the actuarial literature. In particular, risk measures based on convex order and VaR-based as well as CTE-based capital allocation rules were widely studied in recent years. This project aims to investigate different principles of capital allocation. Students taking this project are expected to study the relevant literature and give some numerical examples in order to compare various allocation principles.

All papers below can be accessed online in public domain or from HKUL by HKU students.

References:

- Balog, D., Bátyi, T. L., Csóka, P., and Pintér, M. (2017). Properties and Comparison of Risk Capital Allocation Methods. *European Journal of Operational Research*, 259(2), 614-625.
- Denault, M. (2001). Coherent Allocation of Risk Capital. *Journal of Risk*, 4(1), 1-34. http://neumann.hec.ca/pages/michel.denault/J%20of%20Risk%202001.pdf
- Dhaene, J., Goovaerts, M. J., and Kaas, R. (2003). Economic Capital Allocation Derived from Risk Measures. *North American Actuarial Journal*, 7(2), 44-56. Also a discussion by Eddy Van den Borre.
- Dhaene, J., Henrard, L., Landsman, Z., Vandendorpe, A., and Vanduffel, S. (2008). Some Results on the CTE-based Capital Allocation Rule. *Insurance: Mathematics and Economics*, 42(2), 855-863.
- Dhaene, J., Tsanakas, A., Valdez, E. A., and Vanduffel, S. (2012). Optimal Capital Allocation Principles. *The Journal of Risk and Insurance*, 79(1), 1-28.
- Laeven, R. J. A. and Goovaerts, M. J. (2004). An Optimization Approach to the Dynamic Allocation of Economic Capital. *Insurance: Mathematics and Economics*, 35(2), 299-319.
- Singh, M. K., (2002). Risk-Based Capital Allocation Using a Coherent Measure of Risk. *The Journal of Risk Finance*, 3(2), 34-45.
- Zheng, C. and Chen, Y. (2015). Allocation of Risk Capital Based on Iso-Entropic Coherent Risk Measure. *Journal of Industrial Engineering and Management*, 8(2), 530-553. http://www.jiem.org/index.php/jiem/article/view/1375/681
- Zhou, M., Dhaene, J., and Yao, J. (2018). An Approximation Method for Risk Aggregations and Capital Allocation Rules based on Additive Risk Factor Models. *Insurance: Mathematics and Economics*, 79, 92-100.

Supervisor: **Dr. K.P. Wat**, watkp@hku.hk, Dept of Statistics and Actuarial Science

35. Trustworthy AI with applications in healthcare

High-stakes decision-making in areas like healthcare, finance and governance requires accountability for decisions and for how data is used in making decisions. Many concerns have been raised about whether Artificial Intelligence (AI) models can meet these expectations. AI models are often complex black-boxes and thus have varying, unknown failure modes that are revealed only after deployment: models fail to achieve the reported high accuracies, lead to unfair decisions, and sometimes provide predictions that are plain unacceptable given basic domain knowledge.

This project will study and explore trustworthy AI technology with regard of model generalizability, stability, fairness and explainability. The application of such technology in healthcare domain (such as medical image analysis, health informatics) will be analysis and illustration.

Requirement: The student needs to have experience with Python programming and be familiar with basic machine learning/deep learning technique.

Supervisor: **Dr. L. Yu**, lqyu@hku.hk, Dept of Statistics and Actuarial Science

36. Multimodal AI with applications in healthcare

Most of the current applications of AI in medicine have addressed narrowly defined tasks using one data modality, such as a computed tomography (CT) scan or retinal photograph. In contrast, clinicians process data from multiple sources and modalities when diagnosing, making prognostic evaluations and deciding on treatment plans. The development of multimodal AI models that incorporate data across modalities such as medical images, EHRs, and genomic data can partially bridge this gap and enable broad applications in healthcare.

This project will study and explore multimodal AI models and demonstrate its applications in healthcare domain by analysing image, text, or even genomic data.

Requirement: The student needs to have experience with Python programming and be familiar with basic machine learning/deep learning.

Supervisor: **Dr. L. Yu**, lqyu@hku.hk, Dept of Statistics and Actuarial Science

37. Applications of Graph Neural Networks

Graphs are all around us; real world objects are often defined in terms of their connections to other things. A set of objects, and the connections between them, are naturally expressed as a graph. Graph Neural Networks (GNNs) are a class of deep learning methods designed to perform inference on data described by graphs. It can be directly applied to graphs and provide an easy way to do node-level, edge-level, and graph-level prediction tasks.

This project will study and explore GNN methods and demonstrate its applications in biomedical data analysis, drug discovery, or natural language processing.

Requirement: The student needs to have experience with Python programming and be familiar with basic machine learning/deep learning.

Supervisor: **Dr. L. Yu**, lqyu@hku.hk, Dept of Statistics and Actuarial Science

38. Optimality Studies with Dependent Risks

Due to the complexity of modern insurance and financial products, contemporary insurance risk models have taken many realistic features into consideration. In the actuarial literature, the incorporation of realistic features such as dividends, investment and reinsurance into the basic insurance risk process has generated a lot of interesting research on optimality in the past two decades. This project aims at studying optimal dividends, investment and/or reinsurance for an insurance risk models with dependent risks.

Supervisor: **Prof. K.C. Yuen**, kcyuen@hku.hk, Dept of Statistics and Actuarial Science

39. Bayesian Change Point Detection in Financial Time Series

Time series data are commonly observed in the real world, of which the patterns and trends are of great interest, especially in the financial industry. Fluctuations are frequently observed in financial time series data. Statistical approaches to locate abrupt variations driven by changes in policy, event, and market sentiment have raised great concerns. In this project, students will study various Bayesian change point detection algorithms and learn how to implement those techniques in real financial time series data.

Requirement: Knowledge in R or Python

Supervisor: **Dr. C. Zhang**, zhangcys@hku.hk, Dept of Statistics and Actuarial Science

40. Phase II Clinical Trial Design with Time-to-event Outcomes

Clinical trial design plays a crucial role in drug development, with the primary objective being to establish the effect of the investigated intervention. Following the assessment of safety and toxicity in Phase I trials, Phase II trial focuses on the effectiveness of the intervention for patients under specific conditions. In this project, students will learn and develop Phase II clinical trial designs with time-to-event outcomes. Students with fundamental knowledge in biostatistics are preferred.

Requirement: Knowledge in biostatistics and R programming

Supervisor: **Dr. C. Zhang**, zhangcys@hku.hk, Dept of Statistics and Actuarial Science

41. Statistical Modelling for Biological/Medical Data

(This project will be offered in Semester 1 only.)

In this project, the students will implement statistical methods to analyse real biological/medical data set to understand/interpret biology/disease etiology. Statical methods include Bayesian methods, variable selection, network analysis, etc.

Requirement: Students need to know at least one programming language (such as R, Python, etc) and basic data analysis skills.

Supervisor: **Dr. Dora Y. Zhang**, doraz@hku.hk, Dept of Statistics and Actuarial Science\

42. Multiple Output Online Non-stationary GPs

The goal of this project is to implement an online algorithm for multiple output Gaussian processes. The student will extend a Sequential Monte Carlo sampler for online Gaussian processes by writing a linear co-regionalization kernel to model multiple time series signals. Possible applications include medical settings or financial settings. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

43. Online Spectral Mixture Kernel

The goal of this project is to implement a method to estimate the parameters in the flexible "Spectral Mixture Kernel" in an online setting using a Sequential Monte Carlo algorithm. Applications of this method include medical or financial settings. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

44. Online Student-t Process Algorithm

The goal of this project is to implement an online inference algorithm to learn a heavy tailed Student-t process for time series analysis. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

45. Non-linear Network Embedding

The goal of this project is to model relational data as a non-linear decomposition of a lower dimensional representation of the relations between observations. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

46. A Bayesian Hypothesis Testing Approach for Generative Adversarial Networks

This project involves combining the popular Generative Adversarial Network with various forms of Bayesian hypothesis testing. If successful, the Bayesian hypothesis testing GAN could have stronger classification abilities and could possible reduce the risk of mode collapse. Prior knowledge of deep learning and strong programming ability in Python and deep learning packages like PyTorch, Tensorflow or Keras are required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

47. Forecasting Time Series: with Application to Stocks Trading

This project aims to forecast forward behavior of stock prices using neural networks. Simulated trading strategies based on the forecast results are also required.

Requirement: Knowledge of course STAT3612 or STAT8017, AI/machine learning/deep learning, and skills in statistical programming using either SAS, R, or C++.

Supervisor: **Dr. Z. Zhang**, zhangz08@hku.hk, Dept of Statistics and Actuarial Science

48. Financial data analysis

This project aims to analyze the financial data by using the time series models, causal semantics, or machine learning techniques. Students are expected to use these methodologies to analyze real data sets, and develop useful trading algorithms.

Requirement: At least one programming language and knowledge about financial time series analysis

Supervisor: **Dr. K. Zhu**, mazhuke@hku.hk, Dept of Statistics and Actuarial Science

49. Machine learning methods for analysing single-cell genomic data

In recent years, the rapid development of single-cell sequencing technologies brings unprecedented opportunities to disentangle the heterogeneity in cell populations, including different immune cells, differentiation trajectory or cancer mutations. However, it remains highly challenging to decipher how a biological system functions and the underlying patterns of the data, not only because of high technical noise but also the high dimensions of gene or other molecular feature space.

Therefore, this project aims to develop machine learning methods, likely in a form of probabilistic models or deep learning methods, to analyse single-cell genomic data. Experimental data is available for both model validation and biological exploration. The student is expected to have interests in biomedical data, but no previous experience is required.

Requirement: Experience with Python or R programming.

Supervisor: **Dr. Y. Huang**, yuanhua@hku.hk,
School of Biomedical Sciences & Department of Statistics and Actuarial Science

******** **END** ********