THE UNIVERSITY OF HONG KONG
SCHOOL OF COMPUTING AND DATA SCIENCE
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Topics for STAT3799 Directed Studies in Statistics (6 credits)
(Offered in both 1st and 2nd semesters of 2026 - 2027 for STAT3799)

1. Statistical modelling of longevity risk

The aim of this project is to compare some stochastic mortality models. As a baseline model, the Lee-Carter model was developed in 1992. This model contains single latent factor that represents the developments of mortality over time. This model can be estimated using a Singular Value Decomposition (SVD). In the literature, there are many contributions that aimed to extend this mortality model. For instance, by adding an extra latent or observable factor, or by considering a multi-population model that predicts mortality using also the developments in related countries. In this project, the performance of the Lee-Carter is compared with such related models, both in-sample as well as the forecasting performance. 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:

- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. Journal of the American statistical association, 87(419), 659-671.

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

2. Privacy Protection in Machine Learning

In recent years, machine learning has become an important tool for data analysis in various domains, including healthcare, finance, and marketing. However, as ML models rely on large amounts of data, there is a growing concern about the privacy risks associated with the collection, storage, and processing of sensitive data. In this project, students will explore different techniques and tools for privacy-preserving machine learning, which aim to enable data analysis while protecting the privacy of individuals and organizations.

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

Supervisor: **Prof. Y. Cao**, yuancao@hku.hk, School of Computing and Data Science

3. Copulas in Risk Management

Copulas are functions that join multivariate distribution functions to their one-dimensional marginal distribution functions. The student who takes this project is expected to study the basic theory of copula and some of its applications in risk management. All the related literature will be provided.

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 the macro level understanding of the stock and financial markets, their trends and movements as well as the 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, and monthly 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 (or weekly) and monthly trading strategies.

Pre-requisite knowledge: STAT3600, STAT4601 and R programming

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

5. Queuing Theory

The study of waiting lines, called queuing theory, is one of the oldest and most widely used quantitative analysis techniques. The analytical models of waiting lines can help management evaluate the cost and effectiveness of service systems.

In this project, students will conduct a case study and explore the queuing systems with various statistical models, and apply simulation models to enhance the efficiency of the queuing system.

Supervisor: **Dr. Olivia Choi**, ochoi@hku.hk, Dept of Statistics and Actuarial Science

6. Project Management Analysis

Most realistic projects that corporates/organizations like Microsoft, General Motors undertake are large and complex. Project management is a process that allows project managers to plan, execute, track and complete projects effectively.

In this project, students will conduct a case study and evaluate the efficiency of the selected project by identifying the crucial activities, developing the Work Breakdown Structure and critical path, and hence provide recommendation for the management through the analysis of the Program evaluation and review technique (PERT).

Supervisor: **Dr. Olivia Choi**, ochoi@hku.hk, Dept of Statistics and Actuarial Science

7. Financial markets co-integration analysis

Financial market microstructure theory, a new research branch in financial market analysis, attracted attention from researchers (Busato and Handorf, 2001; Nagel 2016).

Over the years, international cross-listing has arisen a great deal of academic focus particularly on the segmented markets theory and the failure of the law of one price in multiple markets. The existence of long term co-movements of cross-listing company shares performance has importance implication on the Efficient Market Hypothesis as well as portfolio diversification (Taylor & Tonks, 1989; Kasa 1992).

Given the close economic and rapid growth relationships between many major financial markets, students are expected to look into market co-integration and shares performance of these dually listed companies.

Supervisor: **Dr. Olivia Choi**, ochoi@hku.hk, Dept of Statistics and Actuarial Science

8. Decentralized Insurance

Decentralized insurance is an emerging approach to risk sharing in which blockchain technology, smart contracts, data oracles, and decentralized governance are used to create transparent and automated insurance systems that reduce reliance on traditional centralized insurers. Participants may pool funds, set rules for coverage and claims, and use external data sources or oracles to trigger payouts automatically. This model aims to improve transparency, reduce administrative costs, and expand access to insurance, while also raising new challenges in regulation, security, incentives, governance, and risk management.

The student taking this project is expected to study the basic theory and applications of decentralized insurance. Topics may include risk pooling, premium calculation, parametric insurance, solvency, smart-contract-based insurance, oracle risk, and case studies of existing decentralized insurance protocols. All related literature will be provided.

Requirements: Calculus, linear algebra, probability, and statistics. Basic knowledge of actuarial science, optimization, and game theory would be helpful, but can be introduced as needed.

Supervisor: **Prof. Alfred W.F. Chong**, chongwf@hku.hk, Dept of Statistics and Actuarial Science

9. Generative AI for Virtual Cell and Genomic Response Prediction

Recent advances in single-cell genomics have generated large-scale datasets measuring how cells respond to genetic, chemical, and environmental perturbations. Generative AI provides a promising framework for building virtual-cell models that predict genomic responses under new experimental conditions. In this project, students will learn the fundamentals of single-cell perturbation modelling, survey recent developments in virtual-cell research, and study two to three state-of-the-art approaches based on models such as variational autoencoders, transformers, diffusion models, or flow matching. Students will reproduce and compare selected methods on public datasets, focusing on predictive accuracy, biological relevance, and generalization to unseen perturbations, cell types, dosages, or time points.

Requirements: Experience with Python programming and basic knowledge of machine learning or deep learning. Familiarity with PyTorch, statistics, or genomics is preferred.

Supervisor: **Prof. J. Du**, jinhongd@hku.hk, Dept of Statistics and Actuarial Science

10. Evaluation of Large Language Models for Scientific and Biomedical Applications

Large language models are increasingly used for scientific reasoning, biomedical question answering, literature analysis, and clinical decision support. However, their reliability, factual accuracy, and ability to generalize across specialized domains remain important concerns. In this project, students will learn about scientific and biomedical language models, survey recent evaluation methods and benchmarks, and study two to three state-of-the-art models. Students will evaluate selected models on public datasets, focusing on aspects such as accuracy, reasoning ability, factuality, robustness, uncertainty, and hallucination.

Requirements: Experience with Python programming and basic knowledge of machine learning or natural language processing. Familiarity with large language models, statistics, or biomedical applications is preferred.

Supervisor: **Prof. J. Du**, jinhongd@hku.hk, Dept of Statistics and Actuarial Science

11. Matrix Completion with Application to Recommender System

Developing efficient recommender system to track users' preference and provide tailored recommendations are becoming increasingly important in modern society. This project focuses on using matrix completion techniques to construct a recommender system, with the MovieLens dataset as the primary data source. Students are expected to have good knowledge in programming languages such as Python or R.

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

12. 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.

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

Supervisor: **Prof. Edwin C.H. Fong**, chefong@hku.hk, Dept of Statistics and Actuarial Science

13. Prediction-Powered Survival Analysis

Survival studies often have limited sample sizes and substantial censoring, while external AI models may provide useful but biased survival predictions. This project aims to explore and implement prediction-powered methods that integrate observed survival data with AI-based survival predictions to improve survival analysis. Students will compare conventional survival estimators with AI-assisted methods through numerical studies. Students should have a basic understanding of survival analysis. Familiarity with R or Python programming is expected.

Supervisor: **Prof. Y. Gu**, yugu@hku.hk, Dept of Statistics and Actuarial Science

14. Spatial AI: Open-world Multi-modal Scene Understanding with LLMs

Understanding a scene means more than labelling the objects in it. A capable system should recover the three-dimensional structure of the environment, reason about layout, occlusion, relative position and metric scale, and describe all of this in language that a person or another system can act on. Two limitations hold current methods back. First, most perception models operate over a closed set of categories and fail silently when they meet something unfamiliar, whereas real environments are open-world. Second, although large multi-modal models bring broad semantic knowledge, they remain surprisingly weak at basic spatial reasoning, often unable to judge which of two objects is nearer or whether one can be reached without moving another.

In this project, students will study the intersection of 3D scene representation and large multi-modal models, covering open-vocabulary detection and segmentation, structured 3D scene representations such as scene graphs, and the grounding of language in geometry. Students will develop and evaluate methods for open-world multi-modal scene understanding, for example by building language-queryable 3D scene representations, by discovering and naming previously unseen object categories, or by measurably improving the spatial reasoning ability of a multi-modal model.

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

References:

[1] Radford et al., Learning Transferable Visual Models From Natural Language Supervision, ICML 2021.
[2] Kirillov et al., Segment Anything, ICCV 2023.
[3] Kerbl et al., 3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023.

Supervisor: **Prof. K. Han**, kaihanx@hku.hk, School of Computing and Data Science

15. Agentic AI: Security Screening with Multi-modal Foundation Models

Security screening — of baggage, cargo and parcels — is a demanding perception problem. Threat items are rare, heavily occluded, deliberately concealed and viewed through cluttered X-ray or millimetre-wave imagery whose appearance shifts between scanner models and sites. A detector trained on one distribution degrades quickly on another, and a purely one-shot detector cannot express uncertainty or ask for more information. Meanwhile, multi-modal foundation models have brought broad visual and semantic priors, and agentic frameworks have shown that a model can plan over several steps, invoke tools, gather additional evidence and decide when to defer to a human.

In this project, students will study agentic systems built on multi-modal foundation models and apply them to security screening. Rather than producing a single classification, the agent is expected to reason over an image, consult auxiliary information such as a declared manifest, decide whether to request another view or a re-scan, and produce an auditable justification for its recommendation. Students will implement and evaluate such a pipeline on public screening datasets, examining detection performance on rare and occluded items, robustness under domain shift, and the quality of the explanations produced.

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

Supervisor: **Prof. K. Han**, kaihanx@hku.hk, School of Computing and Data Science

16. Physical AI: Physical Understanding and Simulation for Spatial Intelligence

Spatial intelligence is incomplete without physics. Knowing where objects are is only useful if a system can also anticipate what happens next: whether a stack will topple, whether a surface will bear a load, how an object will move when pushed. Humans do this from a young age with intuitive physical models that are approximate but robust. Machine learning systems, by contrast, often produce visually plausible predictions that are physically impossible, and their errors compound quickly over long horizons. A related development is that large video generative models are increasingly used as learned simulators of the visual world, which raises the question of how much genuine physical understanding such models actually acquire.

In this project, students will study physical understanding and simulation in the service of spatial intelligence: learned and differentiable simulators, intuitive physics benchmarks, and video generative models used as world simulators. Students will implement and evaluate a representative approach, probe where its predictions become physically implausible or drift over long rollouts, and develop improvements — for example by injecting physical constraints, by learning from simulation and transferring to real observations, or by designing better evaluations of physical plausibility.

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

References:

[1] Ha and Schmidhuber, World Models, NeurIPS 2018.
[2] Bruce et al., Genie: Generative Interactive Environments, ICML 2024.
[3] Yang et al., Learning Interactive Real-World Simulators, ICLR 2024.

Supervisor: **Prof. K. Han**, kaihanx@hku.hk, School of Computing and Data Science

17. Understanding stochastic dependence modelling

In various applications in statistics and actuarial science, copulas have been successfully applied to construct more adequate joint models. The goal of this project is to mathematically and stochastically understand how dependence can be modelled with copulas, how more flexible dependence models can be constructed parametrically or nonparametrically, what statistical properties they have, how random variates can be obtained from these dependence models, how they can be efficiently implemented and applied, etc.

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

18. 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

19. Post-model-selection Inference

When a correct statistical model is not known a priori, as is common in modern statistical analysis, one often needs to select a data-driven model before proceeding to statistical inference. Classical statistical theory is, however, developed under the assumption of a correct model and may no longer hold if the model is data-driven and is therefore random. This extra level of uncertainty induced by model selection should be taken into serious consideration for subsequent inferences, which requires almost a complete revision of existing statistical practice. This project investigates this problem and its potential solutions.

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

20. 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

21. 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

22. 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

23. 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.

Requirement: pass STAT3618: Derivatives and risk management or MATH3906: Financial calculus or pass STAT3910: Financial Economics I or STAT3911: Financial Economics II

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

24. 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: **Prof. L. Qu**, liangqqu@hku.hk, School of Computing and Data Science

25. 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: **Prof. L. Qu**, liangqqu@hku.hk, School of Computing and Data Science

26. Multimodal large language model for medical image analysis

Medical image analysis is a critical component in healthcare, contributing substantially to accurate diagnoses and effective treatment decisions. Despite its importance, the analysis of medical images presents considerable challenges due to their inherent complexity and variability. Large language models, such as ChatGPT and Qwen, have made significant strides in natural language processing tasks by mastering the contextual representation of words and sentences. However, their application in the domain of medical image analysis is still largely unexplored, with existing multimodal medical image analysis methods predominantly based on traditional general AI strategies.

This project offers students an opportunity to delve into the development and application of a multimodal large language model specifically for medical image analysis. This includes tasks like medical image segmentation, report generation, visual question answering, among others. Students will be exposed to the latest techniques in natural language processing and medical image analysis, fostering a deeper understanding of how to integrate these cutting-edge language processing techniques to enhance medical image analysis.

This project is particularly well-suited for students interested in exploring the intersection of AI and healthcare, providing them with practical experience in applying AI techniques in a critical and rapidly evolving field.

References:

Tong S, Brown E, Wu P, et al. Cambrian-1: A fully open, vision-centric exploration of multimodal llms. arXiv 2024

Zhang S, Xu Y, Usuyama N, et al. BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv 2024

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

Supervisor: **Prof. L. Qu,** liangqqu@hku.hk, School of Computing and Data Science

27. <u>Tackling data heterogeneity challenge in federated learning</u>

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

Supervisor: **Prof. L. Qu,** liangqqu@hku.hk, School of Computing and Data Science

28. <u>Cointegration in Financial Analysis</u>

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

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

29. <u>A/B testing</u>

A/B testing is an online randomized experiment to evaluate which strategy or web layout is better in online platform industry, such as Amazon, Google, Meta, Taobao, Wechat, Baidu, etc. In contrast to randomized controlled clinical trials, there are many new challenges in online experiments due to fast recruitment, large sample size, multiple tests as well as multi-arm bandit problems. This project will explore Bayesian approaches to enhancing efficiency of A/B tests while controlling the type I error rate (false positive).

Requirement: students need to know basic Bayesian inference and computation and familiar with R programming.

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

30. LLM/Agentic AI for Healthcare

This project explores how large language models and AI agents can support healthcare applications, including answering medical questions and automatically analysing medical images. Students may investigate how an AI agent can break down a medical imaging task, select suitable analysis tools, design an analysis framework, and interpret the results. The project will focus on practical problem-solving rather than advanced clinical knowledge. Students should be familiar with programming and large language models, while prior experience in medicine or medical imaging is not required.

Supervisor: **Prof. L. Yu**, lqyu@hku.hk, School of Computing and Data Science

31. LLM/Agentic AI for Scientific Discovery

This project explores how large language models and AI agents can assist scientific discovery, particularly in biomedical research and bioinformatics. Possible topics include summarising scientific literature, analysing biological or clinical datasets, identifying meaningful patterns, generating research hypotheses, and supporting the design of computational experiments. Students will learn how statistical reasoning, data analysis, and AI methods can be combined to address real-world biomedical questions. Students should be familiar with programming and large language models, but specialised biology or bioinformatics experience is not required.

Supervisor: **Prof. L. Yu**, lqyu@hku.hk, School of Computing and Data Science

32. 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

33. 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

34. Statistical Modelling for Biological/Medical Data

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

35. Bayesian Federated Learning

In this project, we will investigate new techniques for performing approximate posterior inference in the federated learning scenario. Here, we will have data distributed across different processors, thus we will need to think of algorithms that can accurately approximate the posterior distribution without leaking the data to other computers. Strong coding skills are a requirement.

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

36. 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

37. <u>Financial data analysis</u>

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

38. <u>Robust optimal risk sharing with applications to insurance</u>

Supervisor: **Prof. Michael B.Y. Zhu**, mbzhu@hku.hk, Dept of Statistics and Actuarial Science

39. <u>Posterior Contraction Rates for Bayesian nonparamtric methods</u>

Supervisor: **Prof. Y. Zhu**, yczhu@hku.hk, Dept of Statistics and Actuarial Science

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