THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Topics for STAT3799 Directed Studies in Statistics (6 credits)
(Offered in both 1st and 2nd semesters of 2019 – 2020 for STAT3799)

1. Mortality Projection and Longevity Risk

Life insurance companies face different types of risks. Within the life annuity business we may find what is called 'longevity risk', which refers to the possibility that annuitants live longer than expected according to the life tables used for pricing, determination of benefits and technical reserves.

This situation represents a threat to life annuity business, and therefore we need to rely on projected life tables that account for the improvement in mortality, a fact that has been observed since the second half of the 20th Century in most developed countries.

The student who takes this project is expected to study the most widespread models in the literature for mortality projection in order to mitigate this type of risk.

Requirement: STAT3901 and STAT3909.

Supervisor: **Dr. A. Benchimol**, benchi@hku.hk, Dept of Statistics and Actuarial Science

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

3. Random Neural Network in Supervised Learning

Following literature to devise a practical implementation of random neural networks in prediction as a supervised learning strategy. Adaptation of the algorithms in different types of data and identify any numerical instability. The project is suitable for students with basic programming skills and interest in statistics/data mining.

Supervisor: **Dr. Simon K.C. Cheung,** simonkc@hku.hk, Dept of Statistics and Actuarial Science

4. Familial Database Search on Two-Person DNA Mixtures Using Peak Area Information

For crime cases in which no suspect can be identified based on non-DNA evidences such as fingerprints or witness reports, the police force may search a database of DNA profiles from previously convicted criminals or unsolved crime cases. If no offender profile in the database perfectly matches the crime trace, an additional search can be performed, hoping that an individual in the database is a close relative of the perpetrator and can be identified through the search. The role of familial database search as a crime-solving tool has been increasingly recognized by forensic scientists. In this project, student will study the basic knowledge of DNA fingerprinting, DNA database search, methodologies on resolving two-person DNA mixtures based on peak area information; and explore the strategies of forensic investigations from familial database search results.

Requirement: Knowledge of course STAT3608 or equivalent. Strong knowledge in programming language like C++ or R and computational algorithms such as MCMC would be essential.

Supervisor: **Dr. Y.K. Chung**, yukchung@hku.hk, Dept of Statistics and Actuarial Science

5. Chronological Age Prediction Based on DNA Methylation

Over the years, the correlation between DNA methylation levels and chronological age has been discovered in different species. In this project, we are going to use the 450K Human Methylation Beadchip data for human age prediction. High dimensional variable selection methods and machine learning models will be attempted. The students are expected to write computer programs using software such as R.

Supervisor: **Prof. Tony W.K. Fung**, wingfung@hku.hk, Dept of Statistics and Actuarial Science

6. Latent Profile Analysis

Latent profile analysis is a method for analyzing the relationships among manifest data when some variables are unobserved. The unobserved variables are categorical, allowing the original dataset to be segmented into a number of exclusive and exhaustive subsets. This project is to explore some latent profile analysis methods. It includes the evaluation of some existing algorithm and models as well as the determination of the number of latent classes and variable set.

Requirement: Knowledge of some computer programming languages is essential.

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

7. 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 programming language like FORTRAN or C++.

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

8. 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

9. Inference about Ordered Binomial Probabilities

Consider the following hypothetical example. Two groups of candidates apply for graduate studies at a renowned university. The two groups are fairly homogeneous except that the first group have a professional qualification but the second group do not. We are interested in the success rates of the two groups in getting offers from the university. At first sight this seems like a rather simple two-sample problem concerning two binomial probabilities, one for each group. However, common sense tells us that the first group should have a success rate at least not smaller than that of the second group, so that the two binomial probabilities are actually “ordered” in a known manner. With such prior knowledge, standard statistical inference about binomial probabilities suddenly becomes not so standard. This project investigates the issues involved in the above problem setting, and studies different plausible solutions to the problem.

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

10. 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

11. 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

12. 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: **Dr. G.D. Li**, gdli@hku.hk, Dept of Statistics and Actuarial Science

13. Computational Methods for Statistical Inference of Dynamical Models

Dynamical models are commonly used in many fields, including population changes in ecology, transmission of disease in epidemiology and etc, and often defined by stochastic process or latent-variable structure. Many of these realistic models do not have closed-form likelihood function, and standard statistical inference like maximum likelihood estimator is difficult to implement. In this project, you will study various computer simulation-based inference, including approximate Bayesian computation and particle filter, and their applications in dynamical models. The project will focus on intensive computational algorithms, and hence experience on statistical computation programming are required.

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

14. Modeling of Social Media Data

In this project, the students will implement latent Dirichlet allocation models to analyse social media data to discover hidden semantic structure in the social media. The students need to know python programming languages and data crawling skills.

Supervisor: **Dr. Z.H. Liu**, zhhliu@hku.hk, Dept of Statistics and Actuarial Science

15. Cointegration in Financial Analysis (for 2nd Semester only)

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

16. Modelling Inline Warrants in Hong Kong

In July 2019, Hong Kong Exchanges and Clearing Limited (HKEX) launched a new type of financial derivative called inline warrant (界內證), which is essentially a product of financial engineering in a way to capture the potential price fluctuation of the underlying stock or index within a particular range. Like other financial derivatives or structured products, the inline warrants can serve the market speculators as well as the need for risk management for hedging purpose.

This project aims to study the features of inline warrants currently issued in the Hong Kong market and to model their price using option pricing formulae and/or simulation. Comparison with other synthetic products or real products in other markets can be performed and simulation study based on stock price modelling should be carried out. Applications of inline warrants in risk management are also to be investigated.

Students taking this project are expected to have fundamental knowledge in financial derivatives and option pricing (e.g., a pass in STAT3618, STAT3910 or FINA2322). Basic modelling skills such as programming and/or spreadsheet modelling are needed.

References:

- https://www.hkex.com.hk/Products/Securities/Inline-Warrants
- https://www.hkex.com.hk/-/media/HKEX-Market/Products/Securities/Naming-Conventions-of-Stock-Short-Name-by-Product-Types/English-Inline-Warrant-Factsheet-v2,-d-,0.pdf

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

17. A Statistical Study on Financial Market Anomalies

Market anomalies can refer to strange patterns in financial data which violate the efficient market hypothesis (EMH). Some famous market anomalies include weekend effect, January effect and size effect. Traders using technical analysis and trading strategies may earn abnormal profits from market inefficiency.

This project aims to study various market anomalies based on statistical analysis. Investigations should be made on the existence or significance of the effects of any market anomalies in various financial markets.

Students taking this project are expected to study the relevant literature and conduct statistical tests using real market data. Elementary programming skills may be required to process large amount of data.

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

18. A Study on Operational Risk Analysis

Operational risk is one of the three major financial risks (the other two being credit risk and market risk) defined and categorized in the context of risk management.

This project aims to study the major topics regarding operational risk and its modelling, measurement and management in the existing literature, including the Basel regulations on capital requirements for banks' operational risk charges.

Students taking this project are expected to study the relevant literature and official documents extensively and to summarize the readings in an organized and comprehensive way for a qualitative part. A basic modelling of operational risk charge using the advanced measurement approach (AMA), i.e., the loss distribution approach (LDA), is expected for a quantitative part in the study.

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

19. A Study on Liquidity Risk Modelling

Liquidity risk management is attracting more attention especially upon the implementation of Basel III. The industry and the regulators realized that maintaining a healthy liquidity level for a bank is as important as the traditional sound risk management in credit risk, market risk and operational risk. Yet the study of liquidity risk modelling is not as popular as that of the three major financial risks mentioned above.

This project aims to study different models of liquidity risk in a statistical context. The study should be a mix of qualitative content (e.g., Basel regulations on liquidity capital requirements) and quantitative component (e.g., simulation study).

Students taking this project are expected to study the relevant literature and official documents extensively and to construct and compare different models for liquidity risk.

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

20. Further Topics in Portfolio Theory

This project aims to study some further topics in portfolio theory which may be more advanced or less mentioned in a traditional study of the Markowitz's model and related topics.

A list of potential topics is given as follows:

- Black-Litterman model
- Kelly criterion
- Advanced topics in asset allocation, e.g., active asset management
- Beta estimation, e.g., Blume's adjustment method
- Tracking portfolio analysis

Students taking this project must first have fundamental knowledge in mean-variance analysis, or modern portfolio theory (MPT), as well as pricing models like the capital asset pricing model (CAPM). Basic modelling skills such as programming or spreadsheet modelling are needed.

Requirement: Pass in STAT3609 or FINA2320

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

21. Ruin Theory

Description: Ruin theory is motivated by the practical concern of solvency for insurer's surplus. To gain more insight into this issue, different risk matrices and surplus models are developed such that mathematical analysis is made viable. The student who takes this project is expected to study the tools and theories pertained to this subject area. All the related literature will be provided.

Supervisor: **Dr. Jeff Wong**, tywong1988@gmail.com, Dept of Statistics & Actuarial Science

22. Investigation of Non-normality in a Simple Errors-in-variables Model

In a classical linear regression model, it is usually assumed that the predictive variable is not subject to any kind of random error. However, it is not always true in many applications. In addition, it is also a common practice to assume that the error in the regression model is normally distributed. Unfortunately, we may often find that most real data sets do not really exhibit such nice properties. In this project, student will investigate the non-normality situation where the errors in a regression model exist. Computer programming skill is required.

Requirement: Strong knowledge in computer programming and statistical simulation technique is a must.

Supervisor: **Dr. Raymond W.L. Wong**, rwong@hku.hk, Dept of Statistics & Actuarial Science

23. Subgroup Analysis with Applications to Clinical Trials

In many scientific studies, one major challenge is to identify and confirm a subgroup of individuals for which a treatment or an intervention is effective. This project will conduct an extensive review of various methods on investigating a potential heterogeneity of covariate effects across subgroups. It includes exploratory and confirmatory subgroup analysis, frequentist, Bayesian and decision-theoretic approaches and, fixed-sample, group-sequential, and adaptive designs. Numerical studies and published case studies will be provided for illustration.

Requirement: R or Python

Supervisor: **Dr. J.F. Xu**, xujf@hku.hk, Dept of Statistics & Actuarial Science

24. Change Measure: Survey and Applications

Change measure is a useful and powerful tool in a number of areas, including mathematical finance, actuarial science and probability theory. In this course, we will study various change measure techniques, such as Girsanov theorem, Esscher transform, and applications in option pricing, premium calculation. The student needs to have some probability background in order to take this course. An advanced probability course (with some measure theory included) is preferred.

Supervisor: **Prof. H.L. Yang**, hlyang@hku.hk, Dept of Statistics & Actuarial Science

25. Wishart Matrix and the Marchenko-Pastur Law

Wishart matrix is a matrix model for sample covariance matrix from a multivariate normal distribution. It has a long history and many interesting results exist for its eigenvalues and eigenvectors. When the dimension increases to infinity, the empirical distribution of the eigenvalues converges to the celebrated Marchenko-Pastur law.

In this project, students will learn some basis theory on Wishart matrix and multivariate normal distributions. Some techniques from random matrix theory will be needed to derive the Marchenko-Pastur distribution. A good command of multivariate analysis and matrix algebra is required.

Supervisor: **Prof. Jeff J.F. Yao**, jeffyao@hku.hk, Dept of Statistics & Actuarial Science

26. Deep Learning with Application in Artificial Intelligence

This project will focus on extracting useful information from structured and unstructured data and formulating statistical models for inference and prediction. In particular, we will develop deep learning, including deep neural networks for imaging analysis and computer vision and natural language processing for text data analysis. Extensive computation will be needed and real data will be used for analysis and illustration.

Requirement: The student needs to have experience with Python and R programming.

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

27. Personalized Ranking and Selection with Applications to Personalized Medicine

This project considers the problem of ranking and selection with covariates which aims to identify a decision rule that can determine the best alternative as a function of observable covariates. A linear model is proposed to capture the relationship between the mean performance of an alternative and the covariates. In this project, we will study some ranking and selection procedures and demonstrate their usefulness via a case study of selecting the best patient-specific treatment.

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

28. Clustering High-Dimensional Time Series Data with Applications

Time series clustering problems arise when we observe a sample of time series and we want to group them into different categories or clusters. This is an important area of research for different disciplines such as stock clustering and bacterial clustering. In this project, we will consider a few existing clustering approaches and apply them to both simulated and real life data sets.

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

29. Insurance Risk Models with Dependent Risks

In classical risk theory, the assumption of independence in the study of the surplus process of an insurance company plays an important role. Since this assumption is rather restrictive and unrealistic, insurance risk models with dependent risks have been studied extensively in the past few decades. In this project, a number of these models will be discussed. In particular, for each of these models, numerical and simulation studies will be carried out to assess the impact of the dependence structure on some actuarial quantities related to ruin.

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

30. Explainable Artificial Intelligence Approach to International Large-scale Educational Assessment Data Mining

This project investigates the international large-scale assessment datasets such as PISA and TIMSS through the explainable artificial intelligence approach. We study the interpretable machine learning algorithms that balance between prediction accuracy and model explainability, then benchmark the model performance with traditional latent variable modeling. In particular, an explainable neural network framework will be considered for new model development. Proficient skills with Python/TensorFlow/PyTorch programming are required.

Supervisor: **Dr. A.J. Zhang**, ajzhang@hku.hk, Dept of Statistics and Actuarial Science

31. GWAS Summary Data Exploration

GWAS has been successful in finding the genetic associated variants with complex traits and diseases. Summary-association statistics (the per SNP allele effect-sizes or log-odds ratio) are often public available. The main focus of this area is to model the summary association statistics with accounting for linkage disequilibrium among variants from a reference panel such as 1000 Genomes. We shall work on modeling GWAS summary data to disentangle the relationship between functional annotations with SNP effect-size. Student taking this project are expected to have strong interest in statistical genetics and good knowledge of programming.

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

32. Improvements on Several R Functions for Time Series Analysis

This project aims to overcome some shortcomings in several R functions related to time series analysis. Students are assumed to know/learn some basic knowledge of time series analysis and the R language.

Requirement: Knowledge of course STAT4601 or STAT8003, and skills in statistical programming using R

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

33. Non-linear Time Series Analysis

Non-linear time series models have achieved a great success in real applications. This project aims to give a study on the modelling and statistical inference of many non-linear time series models such as threshold AR, GARCH, and their variants. Students are expected to use these methodologies to analyze real data sets.

Requirement: Statistics and Matlab.

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

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