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 2020 - 2021 for STAT4798)

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

3. 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

4. Application of Latent Profile Analysis

Latent profile analysis (LPA) is a technique that aims to recover hidden groups from observed data. They are similar to clustering techniques but more flexible because they are based on an explicit model of the data, and allow you to account for the fact that the recovered groups are uncertain.

Objectives of the project:
- Explore and compare various unsupervised classification methods with LPA.
- Apply to a real longitudinal data set.

Requirement: Knowledge in a programming language (e.g. R, Python) and multivariate statistics.

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

5. 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

6. Big Data Analytics in Securities Market

This project aims to discover new knowledge of market volatility of US securities and Hong Kong securities. Students will learn some skills of financial data analysis, and use big data to analyze the securities markets. Some statistical techniques, social media analysis, and/or artificial intelligence methods will be explored in this project. Students with some basic knowledge in statistics, data mining, text mining, and/or programming will take an advantage. Students are expected to be willing to learn new methods and skills.

Supervisor: **Dr. Adela Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

7. Building a Predictive Model to Determine Epidemic Disease Outbreak Risk

This project aims to develop a predictive model of epidemic disease outbreak risk. Students will research the environmental factors of epidemic disease outbreak, and explore to use a new AI method for building a predictive model. Students will learn the methods of contact analysis, network analysis, and artificial intelligence in this project. Students are expected to be interested in research, have some basic knowledge in programming, and use statistics and data mining software.

Supervisor: **Dr. Adela Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

STAT4798

8. Data Visualization in Global Market Analysis

This project aims to discover new business opportunities between countries by analyzing the import and export trading data. Some analysis, such as network analysis, demand and supply analysis, etc., will be done in this project. Students will develop a knowledge map to visualize the import and export trading patterns, and new discovery. Students will learn how to apply statistical methods and data analysis theories in business analysis. Students who have basic knowledge in using statistics and data mining software are preferred.

Supervisor: **Dr. Adela Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

9. 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

10. 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

11. 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

STAT4798

12. 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

13. 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

14. 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

15. 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

16. 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

17.   Ruin Theory

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 T.Y. Wong**, jefftywong@hku.hk, Dept of Statistics & Actuarial Science

18.   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

19.   Statistical Learning of Recurrent Events Data

Recurrent events data is an important type of survival data, which is frequently encountered in practice. There has been a vast literature on recurrent events data analysis. For example, well-known methods include modeling the intensity process of recurrent events and modeling the marginal hazard of each recurrent event or the gap time between recurrent events. In addition, recurrent events problems can also be fit into the paradigm of multi-state models, for which transition probability or transition intensity between states can be estimated to characterize event progression. Another popular approach for recurrent events data is to specify covariate effects on mean or rate functions of recurrent events. This type of approach is attractive because mean or rate functions are more intuitive to interpret than intensity or hazard functions. This project will assess and compare the above-mentioned approaches. In addition, we will also employ various machine learning/deep learning techniques and develop powerful predictive models for recurrent events data.

Requirement: R or Python

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

20.   Reserve Calculation under Stochastic Interest Rate Models

Insurance companies need to have enough reserve all the time to pay the claims. Net Premium Valuation is the preliminary approach to calibrate actuarial liabilities, but it is increasingly replaced by Gross Premium Valuation due to some major disadvantages. Also the reserve depends on the mortality risk and the investment risk. Due to the long term nature of insurance products, the mortality and the interest rate are changing when the economy and/or other environments change. In the life contingencies courses, you always assume the interest is a constant. In this project, we consider the stochastic interest effect. The students should pass the life contingencies courses when they take this project.

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

21.   Distribution of Zeros of a Random Polynomial

Consider a sequence of i.i.d. random variables $a_0, a_1, ..., a_k, ...$ such that $P(a_0=0)=0$. For $n>=1$, consider the random polynomial

$$P_n(z) = a_0 + a_1 z + ... + a_n z^n.$$

Let m_n be the normalized counting measure of the n (complex-valued) zeros of P_n, that is, for any Borel set B in the complex plan,

$$m_n(B) = \ \#\{ \text{ zeros of } P_n \text{ in } B \} / n .$$

A famous theorem by Arnold, Ibragimov and Zaporozhets establishes that when n tends to infinity, m_n weakly converges to the uniform measure on the unit circle if and only if $E \setminus log [1 + |a_0|]$ is finite.

In this project, students will study the proof of this theorem.   A good command of mathematical analysis and first-level probability theory is required.

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

22.   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

23.   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

STAT4798

24. 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. Statistical 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

25. 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.Q. Zhang**, zhangz08@hku.hk, Dept of Statistics and Actuarial Science

26. 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 ********