UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Topics for STAT3819 Project in Statistics and Actuarial Science (6 credits)
(2012 – 2013)

1. Introduction to the Gerber-Shiu function in ruin theory

The Gerber-Shiu function introduced by Gerber and Shiu (1998) unifies and generalizes the study of various ruin-related quantities in ruin theory, including the time of ruin, surplus prior to ruin and the deficit at ruin of the insurance company. In this project, the student will learn the nice analytic properties of the Gerber-Shiu function as well as its applications. Recent advances and generalizations of Gerber-Shiu function will also be discussed. The student is assumed to be familiar with software packages such as Mathematica, Maple or Matlab.

Supervisor: **Dr. Eric C.K. Cheung**, Department of Statistics and Actuarial Science (MW502J)
(eckc@hku.hk)

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**, Department of Statistics and Actuarial Science (MW522)
(kccg@hkucc.hku.hk)

3. H shares and A shares

Many Chinese companies in China are dual-listed in Hong Kong and China (Shanghai or Shenzhen) by issuing H shares and A shares respectively, with price discrepancies having been found between them. The student who takes this project is expected to study the relationship between the movements of H shares and A shares, taking into account various economic factors.

Requirement: Knowledge of financial markets and SAS programming.

Supervisor: **Dr. K.S. Chong**, Department of Statistics and Actuarial Science (MW504)
(kschong@hku.hk)

4. Simulation of Brownian Motion

Brownian motion is a continuous-time stochastic process that plays an important role in the mathematical theory of finance. It is the basic building block of the underlying stock price model used by a large potion of modern finance theory. Monte Carlo simulation of the sample path of a Brownian motion is often needed in the evaluation of portfolios or analysis of option pricing. In this project, student will study the basic theories of Brownian motion and explore different algorithms on the simulation of the sample path, on a regular time grid and on a spatial grid defined by the first passage times.

Requirement: Knowledge in programming language like FORTRAN or C+ is a MUST.

Supervisor: **Dr. Y.K. Chung**, Department of Statistics and Actuarial Science (MW504)
(yukchung@hku.hk)

5. Analysis of DNA profiling data

DNA profiling has become a very powerful method for forensic human identification since its inception in 1985. It is regarded as one of the most important discoveries in forensic science since the introduction of fingerprinting. In this project, the student will learn some basic background in genetics and DNA profiling. The student will investigate some recent problems in DNA profiling.

Supervisor: **Prof. W.K. Fung**, Department of Statistics and Actuarial Science (MW523)
(wingfung@hku.hk)

6. Applications of influence curve

The identification of influential observations starts from the regression framework. One popular approach is the influence curve method. The influence curve is essentially the first derivative of an estimator, viewed as functional, at some distribution. It can be used to study several local robustness properties which are defined and intuitively interpreted. This project is a revision of the existing applications of influence curve for various statistical models.

Requirement: Knowledge of programming in SAS, FORTRAN or C++ is essential.

Supervisor: **Dr. C.W. Kwan**, Department of Statistics and Actuarial Science (MW508)
(cwkwan@hku.hk)

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**, Department of Statistics and Actuarial Science (MW519)
(hrntlkf@hku.hk)

8. Statistical inference for local estimating equations

A problem in nutritional epidemiology had motivated Carroll, Ruppert and Welsh (1998, Journal of the American Statistical Association) to develop a general class of nonparametric estimators obtained by a set of what they term local estimating equations. Generality of their approach is exemplified by the fact that it covers as special cases methods like Nadaraya-Watson kernel regression, local polynomial regression, local likelihood estimation and varying coefficient generalised linear models. This project explores different applications of local estimating equations and studies problems of statistical inference based on local estimating equations.

Supervisor: **Prof. Stephen M.S. Lee**, Department of Statistics and Actuarial Science (MW528)
(smslee@hku.hk)

9. Bootstrap approximation in time series modeling

The traditional time series modeling and further inference are based on the normality assumption or large enough sample size. In the real applications, the normality may be broken and the results may not be accurate for the moderate or small sample sizes. The bootstrap is a computer-intensive method, and the information in the real data is repeatedly used. Hence it may provide more accurate results. This project hopefully can train students for some bootstrap methods to dependent data, and some knowledge of computer languages such as FORTRAN or C is required since a little more computation will be involved.

Supervisor: **Dr. G. Li**, Department of Statistics and Actuarial Science (MW502H)
(gdli@hku.hk)

10. Nonlinear Modelling of Environmental Time Series

The changes in environmental factors, for example, air temperature and rainfall, affect our daily life in recent years. Due to rapid changing in the atmospheric conditions, the modelling and hence prediction of these highly fluctuated factors become a challenging task since the conventional linear approach may be inappropriate. In this study, the nonlinearity and the non-Gaussianity of some environmental times series of Hong Kong will be explored by modern statistical techniques.

Requirement: Basic knowledge of linear time series is preferred and the knowledge of R programming is required for this project.

Supervisor: **Dr. Gilbert C.S. Lui**, Department of Statistics and Actuarial Science (MW506)
(csglui@hku.hk)

11. Risk Analysis of Mandatory Provident Fund and Universal Retirement Scheme

Mandatory Provident Fund (MPF) has been in place in Hong Kong for more than ten years. During this period, the system has been affected by various factors such as reluctance of participation, downturn in investment market and high levels of administration fees. In recent years, several political groups have promulgated the establishment of Universal Retirement Scheme (URS). In this project, the student will analyze the risk factors affecting these two types of retirement security programs. The student is expected to collect information relating to MPF and URS in Hong Kong and other jurisdictions. The end result should be a comparison of the levels of retirement income and funding requirements under various scenarios.

Supervisor: **Dr. Louis F. K. Ng**, Department of Statistics and Actuarial Science (MW505)
(flouisng@hku.hk)

12. The Bayesian Paradigm of Inference and Prediction

One of the recent trends in statistics development is that the frequentist approach advocated by Sir R.A. Fisher has become old-fashioned as evidenced by its being called the "classical approach," while the Bayesian approach is viewed as the modern approach. This study will involve the admitted student (a) to go through the early history of both approaches, (b) to appreciate the differences in inference and prediction between the two approaches and, if any, the inherent problems, and (c) to explore the possible ways of verifying the validity of meaning for each approach. The student is expected to search literature and to understand to a certain extent so that he/she can explain in writing to people about the points listed above.

Supervisor: **Prof. K.W. Ng**, Department of Statistics and Actuarial Science, (MW502)
(kaing@hku.hk)

13. Sample surveys with sensitive questions by non-randomized response techniques

In epidemiological, medical, public health, psychological and sociological surveys, investigators may require to solicit information on sensitive or highly personal questions such as sex, abortion, illegitimate birth, AIDs, illegal betting, shoplifting, drug-taking, tax evasion, annual income, family violence, students' cheating behavior and so on. When sensitive questions are asked directly, some respondents may refuse to answer; even worse, they might provide wrong answers to maintain privacy. To overcome the aforementioned difficulty, Warner (1965, JASA) proposed a so-called randomized response (RR) technique that allows investigators to obtain sensitive information while protecting privacy of respondents. Since the introduction of the Warner model, voluminous work related to the RR technique has been developed in various directions over the past 45 years. Despite these advances, all RR models have the following limitations: (I) a lack of reproducibility (i.e., the same respondent is expected to give the same answer by the design if the survey is conducted again); (II) a lack of trust from the interviewees because randomization device (RD) is controlled by interviewer; and (III) a higher cost due to the use of RDs.

In the past several years, to address some of these issues, without using any RDs some researchers have developed several non-randomized response (NRR) techniques including the crosswise design, the triangular design, and the parallel design for a single sensitive question with a binary outcome. In this project, the admitted candidates are expected to (i) collect real data in Hong Kong for a specific sensitive topic using the three models, (ii) to compare their performances, and (iii) to analyze these data by using likelihood-based method and Bayesian method via R program.

Supervisor: **Dr. Gary G.L. Tian**, Department of Statistics and Actuarial Science (MW520)
(gltian@hku.hk)

14. Investigation of Semi-Static Hedging for Exotic Options

In the context of option pricing theory, two popular hedging methods are dynamic hedging and static hedging, in which both methods have their own merits and limitations. Kolkiewicz and Liu (2012) proposed a so-called semi-static hedging method for the Guaranteed Minimum Withdrawal Benefit (GMWB) in variable annuities. It is of interest to investigate the effectiveness of this hedging strategy for other exotic options. Students are expected to have some background knowledge in some path-dependent options. In this project, students will study the theoretical framework of relative issues and will conduct simulation studies to investigate the performance of the semi-static hedging method.

Requirement: Knowledge in programming language like Mathematica, MATLAB or R.

Supervisor: **Dr. K.P. Wat**, Department of Statistics & Actuarial Science (MW529)
(watkp@hku.hk)

15. 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**, Department of Statistics & Actuarial Science(MW511)
(rwong@hku.hk)

16. On an extension of Peaks-Over-Threshold model

This project extends the threshold model proposed by Wong and Li (2010) in any one of the directions including frequentist approach, Bayesian analysis, and testing method with applications to environmental or insurance data. Potential students are expected to be proficient in writing, to have knowledge in Fortran programming language, and to read the following references:

- Dupuis, D. J. (2012), Modeling waves of extreme temperature: the changing tails of four cities, *Journal of the American Statistical Association*, **497**, 24-39.

- Lee D. (2012), Statistical inference of a threshold model in extreme value analysis, *Mphil Thesis*, The University of Hong Kong, Hong Kong (to appear).

- Wadsworth J. L. and Tawn J. A. (2012), Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modeling, *Journal of the Royal Statistical Society*, Series B, **74**, 543–567.

Supervisor: **Dr. Tony, S.T. Wong**, Department of Statistics and Actuarial Science (MW506)
(wongtonyst@hku.hk)

17. Applications to risk theory with a class of Erlang mixures

It is known that a variety of distributions are of mixed Erlang type, in which case computational formula exists for many quantities of interests in risk theory. In this project, the student studies distributional properties of the class of Erlang mixtures as well as various risk theoretic applications including analysis of insurer's surplus process and discounted aggregate claims. In particular, the student is assumed to possess computational skills and statistical background to apply the methods of Lee and Lin (2010) who studied fitting of Erlang mixtures to data by using the EM algorithm and parameter reduction techniques.

Supervisor: **Dr. J.K. Woo**, Department of Statistics and Actuarial Science (MW530)
(jkwoo@hku.hk)

18. Ruin Probability Approximation by Moment Fitting in a Regime Switching Model

In the literature, there are two popular ruin probability approximation methods, one is the De Vylder approximation and the other one is the Beekman-Bowers approximation. In this project, we will investigate these two methods in a regime switching insurance risk model.

I will provide all the related literature and ideas to the student who takes this project. The student should have taken the risk theory course. The student needs to do some theoretical work and some numerical examples.

Supervisor: **Prof. H. Yang**, Department of Statistics & Actuarial Science (MW526)
(hlyang@hku.hk)

19. Analysis of large data sets: new tools from random matrix theory

Large data sets refer to data where the number of variables, or data dimension say p, is large compared to the sample size, say n. Modern statistical problems involve frequently such large data sets from various fields like genomic data analysis, financial portfolio optimization or design of wireless communication networks. For example in a genomic micro-array, p is several thousands and n several hundreds. Classical multivariate statistical tools dramatically fail to analyse these large data sets: either there are not applicable any more or lack efficiency.

New tools have emerged recently from the theory of random matrices. Most of them are based on the distribution of eigenvalues of sample covariance matrices which are computable from the data. Classical tools like Hotelling T^2 (tests on the mean), testing of generalized linear hypothesis (for regression or MANOVA) have been corrected or adapted to cope with large data sets. The theory behind is appealing and applications to large-dimensional data analysis are significant.

In this project, the student will i) learn some fundamental theorems from the theory of random matrices; ii) learn some new statistical tools developed recently; iii) start some own thinking about unsolved problems or perform some simulation experiments in order to get a deeper understanding of these results. I have included below an expository paper on the subject for a first introduction.

First reading:
Z.D. BAI (2005). High dimensional data analysis. *COSMOS*, Vol. 1, No. 1, 17–27.
(downloadable from: http://web.hku.hk/~jeffyao/Bai-cosmos-05.pdf)

Supervisor: **Dr. J.F. Yao**, Department of Statistics & Actuarial Science (MW502G)
(jeffyao@hku.hk)

20. Group Sequential Methods

Conventional clinical trial designs typically determine the total sample size in advance and only perform one final analysis after all the data are collected according to the planned sample size. These methods are rigid and lack flexibility. By contrast, group sequential methods are much more flexible, which regularly examine the efficacy data over administratively convenient intervals, and also monitor possible futility stopping along the course of the study (Jennison and Turnbull, 2000). Students will explore and compare different group sequential methods. Extensive R programming and simulations are needed.

Requirement: Programming in R.

Supervisor: **Dr. G. Yin**, Department of Statistics and Actuarial Science (MW502E)
(gyin@hku.hk)

21. Interval Estimation of Sharpe Ratio

The building blocks of the Sharpe ratio-expected returns and volatilities-are unknown quantities that must be estimated and are therefore subject to estimation error. How accurately is Sharpe ratio measured? This project reviews several methods of constructing a confidence interval for Sharpe ratio. Simulations will be conducted to compare their performance.

Supervisor: **Dr. Philip L.H. Yu**, Department of Statistics and Actuarial Science (MW521)
(plhyu@hku.hk)

22. Forecasting the Ranks of Asset Classes

In this project, we will study several statistical methods of modeling ranks and apply them to forecast the ranks of asset classes and hence determine the optimal asset class allocation.

Supervisor: **Dr. Philip L.H. Yu**, Department of Statistics and Actuarial Science (MW521)
(plhyu@hku.hk)

23. Resampling Tests for Comparing *k* Cumulative Incidence Functions in a Competing Risks Model

In the competing risks framework, an individual is exposed to *k* risks, but the actual cause of failure is due to only one of the risks. It is of great interest to examine whether there are possible differences in mortality from different causes of failure. In this project, resampling methods for comparing *k* risks in a competing risks model are studied.

Supervisor: **Prof. K.C. Yuen**, Department of Statistics and Actuarial Science (MW531)
(kcyuen@hku.hk)

24. Tests of extremal index based on inter-exceedances

The inter-exceedance times of a time series will be studied. Based on this, some test statistics of the extremal index are developed. Properties of the test statistics such as the distributions and powers are studied using theoretical and simulation methods. Real examples are also studied.

Requirement: Knowledge in programming language like FORTRAN or SAS is a MUST.

Supervisor: **Dr. Z. Zhang**, Department of Statistics and Actuarial Science (MW511)
(zhangz08@hku.hk)

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