香港城市大學
City University of Hong Kong

Department of Systems Engineering and Engineering Management
City University of Hong Kong

International Statistical Institute Satellite Meeting

On

Statistics in Business, Industry and Risk Management

Sponsored by ISBIS and the ISI-CRA

23-24 August 2013

City University of Hong Kong

Hong Kong, China

International Statistical Institute Satellite Meeting
On
Statistics in Business, Industry and Risk Management

The 59th World Statistics Congress (WSC) organized by International Statistical Institute (ISI) will be held in Hong Kong during 25-30 August 2013. The 59th WSC provides a platform for the international statistical community to share and present the latest knowledge and innovation in statistics. The scientific programme encompasses a wide range of topics facilitating professional exchanges and sharing amongst experts and practitioners in various statistical spheres. Department of Systems Engineering and Engineering Management would like to this opportunity to organize the ISI Satellite Meeting prior to the WSC.

The ISI Meeting takes place in City University of Hong Kong during 23-24 August 2013. The theme of this Satellite Meeting will be on Statistics in Business, Industry and Risk Management.

Organized by:
Department of Systems and Engineering and Engineering Management
City University of Hong Kong

Sponsored by ISBIS and the ISI-CRA

23-24 August 2013, City University of Hong Kong

Programme Committee

Prof. David Banks	**Duke University, USA**
Prof. Vijay Nair	**University of Michigan, USA**
Prof. Vincenzo Esposito Vinzi	**ESSEC Business School, France**
Prof. Min Xie	**City University of Hong Kong**

Local Chair:

Dr. K. S. Chin	**City University of Hong Kong**

Local Co-Chair:

Prof. Kwok Leung Tsui	**City University of Hong Kong**

Program

Friday, August 23

Conference Venue: Lecture Theatre 7&9, 4/F, Academic Building 1, City University of Hong Kong

<table>
<tr><td>13:30-14:00</td><td colspan="2">Registration (in front of LT-9, coffee and tea will be served)</td></tr>
<tr><td>14:00-15:30</td><td colspan="2">Welcoming Remarks and Plenary Session (at LT-9)
Geoff McLachlan, University of Queensland, Australia
Non-Normal Mixture Models with Applications to Financial Data</td></tr>
<tr><td>15:30-16:00</td><td colspan="2">Coffee Break (in front of LT-9, light refreshments will be served)</td></tr>
<tr><td>16:00 -17:30</td><td>Technical Sessions

Quality Control (23/8, LT9)
Chair: Changsoon Park

Fugee Tsung, Hong Kong University, Hong Kong
Statistical Quality Control in a Service Environment
Jaeheon Lee, Chung-Ang University, South Korea
GLR Charts for Simultaneously Monitoring Sustained Shifts and Linear Drifts in the Process Mean
Nan Chen, National University of Singapore, Singapore
Monitoring Wafer Geometric Quality using Additive Gaussian Processes</td><td>Process Management (23/8, LT7)
Chair: Natalia M. Markovich

Ryan Ip, Hong Kong University, Hong Kong
Composite Likelihood Approach for Multivariable Spatio-Temporal Processes
Nozer Singpurwalla, City University of Hong Kong, Hong Kong
The Indicative and Irrealis Moods of Bayesians
Natalia M. Markovich, Russian Academy of Sciences, Russia
Risk Evaluation of Transmission Processes in Overlay Networks</td></tr>
<tr><td>18:00-19:30</td><td colspan="2">Evening Reception
(9/F, Amenities Building, CityU. Please gather in front of LT-9)</td></tr>
</table>

Saturday, August 24

Conference Venue: Lecture Theatre 7&9, 4/F, Academic Building 1, City University of Hong Kong

09:00-09:15	Registration (in front of LT-9, coffee and tea will be served)	
09:15-10:30	Technical Sessions Risk Analysis (24/8, LT9) Chair: K. K. Jose Kanichukattu **Tsuyoshi Nakamura**, Chuo University, Japan Modelling Intra-Individual Variations in Hair Minerals Measured by PIXE for Relative Risk Analysis **Marek Kimmel**, Rice University, USA Modelling the Detection of Lung Cancer in SEER Population **K. K. Jose Kanichukattu**, Palai Mahatma Gandhi University, India Generalized Brownian-Laplace Processes and Financial Modelling	Survival Analysis and Insurance(24/8, LT7) Chair: Yves-Laurent Grize **D. S. Hooda**, Jaypee University of Engineering and Technology, India On Generalized Measures of Past Entropies in Survival Analysis **S. C. Malik**, Maharshi Dayanand University, India Reliability Measures of a Standby System with Priority to Maintenance over Repair Subject to Random Shocks **Yves-Laurent Grize**, Baloise Insurance, Switzerland Non-life Insurance: Fertile Ground for Statistical Applications
10:30-11:00	Coffee Break (in front of LT-9, light refreshments will be served)	
11:00-12:15	Technical Sessions Sampling and Modelling (24/8, LT9) Chair: Lulu Kang **Rainer Göb**, University of Würzburg, Germany Recent Developments in Statistical Audit Sampling **Kristina Lurz**, University of Würzburg, Germany Exponential Smoothing with Covariates **Lulu Kang**, Illinois Institute of Technology, USA Density-Based Partitioning for K-fold Cross-Validation	Experimental Design (24/8, LT7) Chair: Martina Vandebroek **Mark Anderson**, Stat-Ease, USA Practical Aspects for Designing Statistically Optimal Experiments **Roger Hoerl**, Union College, USA Enhancing Big Data Projects Through Statistical Engineering Principles **Martina Vandebroek**, KU Leuven, Belgium Fast Algorithms to Generate Individualized Designs for the Mixed Logit Choice Model

Saturday, August 24

Conference Venue: Lecture Theatre 7&9, 4/F, Academic Building 1, City University of Hong Kong

12:15-13:30	Lunch (9/F, Amenities Building, CityU. Please gather in front of LT-9)	
13:30-14:45	Technical Sessions Reliability (24/8, LT9) Chair: William Meeker **William Meeker**, Iowa State University, USA Risk Assessment Based on Limited Field Data **Yili Hong**, Virginia Tech, USA Field Failure Prediction Based on Multi-Level Repair and System Usage Information **Sheng-Tsaing Tseng**, National Tsing-Hua University, Taiwan Joint Modelling of Lab and Field Data with Application to Warranty Prediction for Highly Reliable Products	Recent Advances in Process Control (24/8, LT7) Chair: Min Xie **Honghao Zhao,** City University of Hong Kong, Hong Kong A Window-Limited Generalized Likelihood Ratio Test for Monitoring Poisson Process with Linear Drifts **Jun Yang**, Beihang University, China Process capability indices and control charts based on the highest possibility interval/region **Zhonghua Li,** Nankai University, China Nonparametric Statistical Process Control of Univariate Process
14:45-15:15	Coffee Break (in front of LT-9, light refreshments will be served)	
15:15-16:30	Technical Sessions New Results in Structural, Stochastic and Statistical Reliability (24/8, LT9) Chair: Chen-ju Lin **Franciso J. Samaniego**, University of California at Davis, USA Estimating Component Reliability from System Failure-Time Data **Bo Henry Lindqvist**, Norwegian University of Science and Technology, Trondheim, Norway Some New Perspectives on the Signature of a Coherent System in Engineering Reliability **Chen-ju Lin**, Yuan Ze University, Taiwan Tool Wear Diagnosis and Tool Replacement Management	Studies on Reliability and Quality Analysis (24/8, LT7) Chair: Matthias H.Y. Tan **Zhisheng Ye**, Hong Kong Polytechnic University, Hong Kong The inverse Gaussian process as a degradation model **Mimi Zhang**, City University of Hong Kong, Hong Kong An Optimal Maintenance Policy for a Multi-state Deteriorating System **Matthias H.Y. Tan**, City University of Hong Kong, Hong Kong Minimax Space-Filling Designs for Finite Design Regions
16:30-17:30	Plenary Session and Closing Remarks(24/8, LT9) **Kwok-Leung Tsui**, City University of Hong Kong, Hong Kong Statistics Research in Prognostics and System Health Management	

Abstracts

Non-Normal Mixture Models with Applications to Financial Data

Geoff McLachlan, University of Queensland

Normal and t-mixture models have been widely applied to datasets from a variety of fields, including biostatistics, bioinformatics, finance, image analysis, and medical science, among many others. Finite mixtures of multivariate skew-symmetric distributions, in particular, the skew normal and skew t-mixture models, are emerging as a promising extension to the traditional normal and t-mixture modelling. Most of these parametric families of skew distributions are closely related. In this talk, we give a brief overview of various existing proposals for multivariate skew distributions for use in forming non-normal mixture models. We also discuss software implementation of EM-type algorithms for the fitting of these models by maximum likelihood. We compare the relative performance of some of these models in density estimation and clustering for two real datasets on some financial data. The first set concerns some bankruptcy data on American firms. The second set considers the estimation of the Value-at-Risk (VAR), which is frequently used in financial risk management as a measure of the risks of investment loss. The data for the latter example are taken from a portfolio of shares listed on the Australian Stock Exchange (ASX). (Joint work with Sharon Lee.)

Statistical Quality Control in a Service Environment

Fugee Tsung, Hong Kong University of Science and Technology

Driven by a new business environment including globalized economy, business automation, and business and technology innovations, the service sector keeps growing and now accounts for more than 50 percent of the labour force in the developed economies. It reaches as high as 80 percent in the United States and the United Kingdom. With the shift in economic focus from manufacturing to service, industrial and academic research facilities may need to apply more scientific rigor to the practices of service, such as discovering better methods to use statistics and mathematical optimization to increase quality, productivity, and efficiency to meet the challenges. This talk will focus on the development of statistical quality techniques, and discuss several technical challenges and recent extensions to the service engineering research area.

GLR Charts for Simultaneously Monitoring Sustained Shifts and Linear Drifts in the Process Mean

Milim Choi and Jaeheon Lee, Chung-Ang University

This paper considers the problem of monitoring the mean of a normally distributed process variable when the objective is to effectively detect both sustained shifts and linear drifts. The design and application of a generalized likelihood ratio (GLR) chart for simultaneously monitoring sustained shifts and linear drifts are evaluated. This GLR chart does not have any control chart parameters that need to be specified other than the control limit, and has the advantage that, at the time of the signal, estimates of both the change point and the shift and/or drift size are immediately available. The performance of the GLR chart is compared with that of other control charts, such as the standard cumulative sum (CUSUM) charts. We also compare the proposed GLR chart with the GLR charts designed for monitoring only sustained shifts and for monitoring only linear drifts. We expect that the proposed GLR chart has better performance for a wide range of shift sizes and drift rates relative to other control charts, when a special cause produces a sustained shift and/or a linear drift in the process mean.

Monitoring Wafer Geometric Quality using Additive Gaussian Processes

Nan Chen, National University of Singapore

Wafer geometric quality is an important quality characteristic in the semiconductor industry. However, it is difficult to monitor the geometric quality in the manufacturing process due to the challenges raised from the complexity of the data structure. In this article, we propose an additive Gaussian process (AGP) model to approximate the standard wafer geometric profile while quantifying the deviations from the standard when the process is normal. Based on the AGP model, two statistical tests are developed to determine whether a newly produced wafer is conforming or not. We have conducted extensive numerical simulations and real case studies, the result of which indicates that our proposed method is effective and has potentially wide applications.

Composite Likelihood Approach for Multivariable Spatio-Temporal Processes

Ryan H. L. Ip and W. K. Li, The University of Hong Kong

In recent years, modelling and predicting spatio-temporal processes has received numerous attentions in research. However, the high dimensional problem posed a great computational obstacle in practice. Traditional estimating approaches including likelihood and Bayesian methods become inefficient, if not infeasible, owing to inverting high dimensional matrices or computing high dimensional integrals. To rectify the problem, some authors suggested the composite likelihood approach, a simplified version of the full likelihood approach, which is found to be more computationally efficient while retaining nice properties of the full likelihood estimators including statistical consistency. Multivariable spatio-temporal processes, which include more than one variable, are usually seen in practice, especially for environmental quality monitoring networks. While most of the previous works on composite likelihood approaches have been focusing on univariate spatio-temporal processes, the objective of our work aims at developing a composite likelihood approach in modelling multivariable spatio-temporal processes. We propose to estimate the parameters of the cross-covariance functions based on a weighted pairwise joint composite estimating method. Its potential applications, including environmental quality monitoring and risk assessment concerning air pollution, water pollution, etc., will be discussed. Key Words: Environmental data, geostatistics, computational efficiency, weighted pairwise joint composite estimation.

The Indicative and Irrealis Moods of Bayesians

Nozer Singpurwalla, City University of Hong Kong

The title says it all.

Risk Evaluation of Transmission Processes in Overlay Networks

Natalia M. Markovich, Institute of Control Sciences of Russian Academy of Sciences

The paper contributes to the risk evaluation and quality-of-service assessment of transmission processes in overlay networks. Peer-to-peer systems like Skype and Internet IP TV provide examples of overlay networks. The latter can be considered as multi-agent systems with a random number of players (active customers of Internet) at each time moment. Such randomness causes problems in the observation and gathering of statistical information regarding the transmission processes and the required resources for a stable operation of the network. Thus, inferences regarding distributions of the networks characteristics based on available information are required. We consider the packet transmission processes that are induced by a peer-to-peer overlay disseminating real-time traffic as an available example of such a multi-agent system. It is technically very difficult to measure the length of the packet path and the end-to-end delay (i.e. the packet delivery time between sender and receiver peers). The problem is also that inter-arrival times between packets are dependent and heavy-tailed distributed. This generates clusters of exceedances of the transmission rate process over sufficiently high thresholds. The latter are interpreted as equivalent capacity of the channel. The clusters cause the loss of packets due to exceedances over the available channel capacity. Applying the theory of extreme events, we derive that the asymptotic distribution of the end-to-end packet delay is subexponential heavy-tailed if the inter-arrival times between packets are regularly varying heavy-tailed distributed. The packet length of the path is derived to be geometrically distributed. The probability to miss a packet is considered as a quality-of-service attribute showing the risk to lose the packet. The relation between the missing probability of transmitted packets of a variable bitrate video stream and the distributions of the length of the overlay path, the upload capacity of nodes and the playback delay at a destination node is presented. The exposition is illustrated by real data inferences from Skype and Internet IP TV measurements. Key Words: Overlay network, multi-agent system, risk evaluation, missing probability, heavy-tailed distribution, exceedances over threshold

Modelling Intra-Individual Variations in Hair Minerals Measured by PIXE for Relative Risk Analysis

Tomomi Yamada, Mie University; Todd Saunders, Nagasaki University; Tsuyoshi Nakamura; Chuo University; Koichiro Sera, Iwate Medical University; and Yoshiaki Nose, Kumamoto Health Sciences University

With the advent of the Proton Induced X-ray emissions (PIXE) technique, we can now obtain data concerning mineral amounts in a diversity of materials ranging from cloth and wine to prehistoric stone implements and aerosols. In this paper we are concerned specifically with the mineral amounts found in human hair. Hair mineral content is an ideal biomarker to measure elemental exposure. However, the associations between hair mineral amounts and individual health status are mostly unknown. In addition, the variability of hair mineral amounts in healthy subjects is generally so large that the study of them for medical practice has been limited. We undertook a cohort study to determine the association between hair mineral amounts and the onset of atopic dermatitis (AD) in infants. Eight hundred and thirty four mother-infant pairs, who donated hair samples during one and ten month health checkups, had their samples analysed by PIXE for 32 mineral concentrations. These mineral amount data together with individual AD family history were statistically examined for any associations between them. Results indicated that of the 32 minerals, only Selenium (Se) and Strontium (Sr) showed statistically significant associations for infants, while the same two elements were only marginally significant for mothers. To predict the probability of AD development, we performed logistic regression analysis, which provided a sensitivity of 65.9%, a specificity of 70.5% and a relative risk (RR) of 4.2 (J. Trace Elem. Med. Bio. 27, 126–131, 2013), all far better than any corresponding figures explicitly mentioned in previously published papers. In addition, we have observed extremely large intra- individual variations as well as peculiar nature of inter- individual variations in individual hair mineral amounts. It is well known that those variations, or measurement errors, degrade the association to the null. Thus, the objective of the present study is to determine the inter- individual distribution and the nature of the intra- individual variation for each mineral and to develop and describe statistical and simulation methods for sound and effective use of hair minerals for epidemiological relative risk assessment of various diseases.

Modelling the Natural History and Detection of Lung Cancer in SEER Population

Xing Chen, Zhejiang University; Millennia Foy, The University of Texas Health Science Centre; Marek Kimmel, Rice University; and Olga Y. Gorlova, The University of Texas MD Anderson Cancer Centre

Lung cancer is the most deadly cancer in men and women worldwide. The major reason for its high mortality rate is that it is less likely than many other cancers to be detected at an early stage. In this study, we developed a method for simulating, at an individual level, the characteristics of lung cancer in a population at the time of diagnosis, according to progression and detection of the disease. Lung cancer data from Surveillance, Epidemiology and End Results (SEER) database collected between 2004 and 2008 were retrieved and transformed to fit the lung cancer progression and detection model. The fitted model combined with a previously fitted carcinogenesis model was used to predict the distribution of age, gender, tumour size, and disease stage at diagnosis and the results were validated against independent data from the SEER database collected from 1988 to 1999. The model accurately predicted the gender distribution and median age of LC patients of diagnosis, as well as the tumour size and disease stage distribution. The results are discussed in the context of identification of the optimal high-risk group for mass screening programs.

Generalized Brownian - Laplace Processes and Financial Modelling

K.K. Jose Kanichukattu, Palai Mahatma Gandhi University

In this paper we review various stochastic processes including autoregressive processes developed recently for modelling data from financial contexts. In particular, we consider different Laplacian models and their generalizations. A general framework for Gaussian and non-Gaussian autoregressive models and their extensions is also developed and studied in detail with respect to Brownian-Laplace processes. Fractional extensions are also considered. An illustration is made with respect to a real data on exchange rates of Indian rupee and U.S. dollar. (Key words: Autoregressive processes, Brownian-Laplace processes, Financial Modelling, Fractional processes, Option pricing.)

On Generalized Measures of Past Entropies in Survival Analysis

D. S. Hooda, Jaypee University of Engineering and Technology

In the present paper we define two parametric generalized measures of past entropies in survival. Applying these measures we characterize life time distributions for continuous and discrete random variables. We also prove as corollaries that these entropies characterize uniform distribution for discrete and continuous random variables. In the end we study their applications in problems of survival Analysis.

Reliability Measures of a Standby System with Priority to Maintenance over Repair Subject to Random Shocks

S.C.Malik, Maharshi Dayanand University

The purpose of this paper is to evaluate reliability measures of a shock model developed for a system of two identical units- one is operative and other is kept as spare in cold standby. The operative unit suffers a random shock with some probability. A single server is provided to perform maintenance and repair of the unit. The unit undergoes for maintenance if it is affected by impact of shocks. However, repair of the unit is done when it fails due to some other reasons. Priority is given to the maintenance over repair of the unit. The shocked unit may work for the system. The unit works as new after maintenance and repair. Random shocks and failure times of the unit are exponentially distributed while maintenance and repair times are taken as arbitrary with different probability density functions. Several measures of system effectiveness are derived in steady state using semi-Markov process and regenerative point technique. The graphical behaviour of mean time to system failure (MTSF), availability and profit has also been observed giving particular values to various parameters and costs. (Keywords: Cold Standby System, Random Shocks, Maintenance, Repair, Priority and Reliability Measures.)

Non-life Insurance: Fertile Ground for Statistical Applications

Yves-Laurent Grize, Baloise Insurance, Switzerland

The business model of insurance companies, especially when active in non-life insurance, has seen dramatic changes over the last fifteen years, opening exciting new areas for statistical applications. Using practical examples from product development, pricing and marketing, I will describe how the activities of an insurance company can be viewed as an industrial process where data management and analysis play a key role, and therefore where statisticians can contribute and excel. Challenging applied research problems, especially involving large data sets will be indicated.

Recent Developments in Statistical Audit Sampling

Rainer Göb and Kristina Lurz, University of Würzburg

The evolution of audit sampling has been accompanied and reflected by national and international auditing standards like SAS 39 (1981), SAS 47 (1983), SAS 106 (2006), SAS 107 (2006), SAS 111 (2006). For long, the standards have been strengthening the position of judgmental selection in the auditing community. However, the recent international standards ISA 500 and 530 adopt a markedly different position: judgmental selection is not considered as audit sampling, the position of statistical sampling is considerably strengthened. The talk reviews the implications of the position of ISA 500 and 530 for stochastic modelling and for the design of statistical audit sampling. In particular, the use of recent results on two-sided confidence limits for proportions and for the mean of zero-inflated populations in audit sampling is demonstrated.

Exponential Smoothing with Covariates

Kristina Lurz and Rainer Göb, University of Würzburg

For long, exponential smoothing (ES) was considered rather a heuristic forecasting technique without a precise model foundation which guarantees optimality. In 1997, Ord et al. provided the basis for a solid model framework by formulating the single source of error (SSOE) state space scheme, which allowed demonstrating the optimality of the classical ES predictors. By introducing an additive term depending linearly on exogenous variables into the observation equation of the SSOE model, Wang (2006) developed the method of exponential smoothing with covariates (ES-Cov). The present study considers extensions and variants of ESCov in the following respects: i) the way of including covariates, ii) multiple seasonalities, iii) prediction intervals. The models are illustrated in industrial applications, in particular from the area of electrical load forecasting. (Keywords: Exponential smoothing with covariates; Time series; Exogenous factors; Prediction interval.)

Density-Based Partitioning for K-fold Cross-Validation

Lulu Kang, Illinois Institute of Technology

Cross-validation, sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. K-fold cross-validation is the most commonly used cross-validation method. Typically in K-fold cross-validation, the original sample is randomly partitioned into K sub-samples. In this research, we propose a new sampling method to partition the complete data set into K folds such that the probability distribution within each fold is similar to the probability distribution of the complete date set. The L2-norm of the difference between the empirical kernel density functions of the within-fold data and the complete data is used here as the criterion to search for the optimal partition of the K folds. An efficient optimization algorithm combining point-exchange and simulated annealing methods is developed to minimize the criterion and then obtain the optimal partitioning for the K-fold cross-validation. Numerical examples are shown to compare the proposed cross-validation methods and the conventional ones.

Practical Aspects for Designing Statistically Optimal Experiments

Mark J. Anderson and Patrick J. Whitcomb, Stat-Ease, Inc

Due to operational or physical considerations, standard (i.e. canned) factorial and response surface method (RSM) often prove to be unsuitable for actual experimentation. In such cases a computer-generated statistically-optimal design fills the breech. The talk explores vital mathematical properties for evaluating alternative designs. To assess goodness of design such evaluations must consider the model choice, specific optimality criteria (in particular D and IV), precision of estimation (based on the fraction of design space FDS), the number of runs (to achieve required precision), lack-of-fit testing, and so forth. All these issues are considered at practical level keeping the actual experimenters in mind. This brings to the forefront such considerations as subject matter knowledge (first principles), factor choice and the feasibility of the experiment design.

Enhancing Big Data Projects through Statistical Engineering Principles

Roger W. Hoerl, Union College

Massive data sets, or “Big Data” have become much more common recently, due to improved technology for data acquisition, storage, and processing of data. With the advent of Big Data, several disciplines, not only statistics, have developed new tools to analyze such data, including classification and regression trees (CART), neural nets, resampling or bootstrapping methods, and various clustering algorithms. These tools make high-powered statistical methods available to not only professional statisticians, but also to casual users. As with any tools, the results to be expected are proportional to the knowledge and skill of the user, as well as the quality of the data. Unfortunately, much of the data mining, machine learning, and Big Data literature may give casual users the impression that if one has a powerful enough algorithm and a lot of data, good models and good resultsare guaranteed at the push of a button.

Conversely, if one applies sound principles of statistical engineering (Hoerl and Snee 2012) to the Big Data problem, several potential pitfalls become obvious. I consider three important principles of statistical engineering that in my opinion have been either overlooked or underemphasized in the Big Data literature: The importance of using sequential approaches to scientific investigation, as opposed to the one-shot study so popular in the algorithms literature, The need for empirical modeling to be guided by subject-matter theory, including interpretation of data within the context of the processes and measurement systems that generated it, and The inaccuracy of the typical unstated assumption that all data are created equal, and therefore that data quantity is more important than data quality.

After introduction of the Big Data problem, and mention of the newer tools now available to analysts, I will discuss the problems that can arise when these statistical engineering fundamentals are ignored, even with well-designed and powerful analytic tools. Further, I will share my thoughts on how to improve data mining projects by incorporating these principles into the overall project. Combining the powerful analytic techniques now available with sound modelling principles and massive data that have a quality pedigree can produce groundbreaking results. However, Big Data projects must be built upon a sound statistical engineering foundation, and not upon the sandy foundation of hype and assumptions.

Fast Algorithms to Generate Individualized Designs for the Mixed Logit Choice Model

Deniz Akinc, Marjolein Crabbe, and Martina Vandebroek, KU Leuven

The mixed logit choice model has become the common standard to analyze transport behavior. Efficient design of the corresponding choice experiments is therefore indispensable to obtain precise knowledge of travelers' preferences. Accounting for the individual-specific coefficients in the model, this research advocates an individualized design approach. Individualized designs are sequentially generated for each person separately, using the answers from previous choice sets to select the next best set in a survey. In this way they are adapted to the specific preferences of an individual and therefore more efficient than an aggregate design approach. In order for individual sequential designs to be practicable, the speed of designing an additional choice set in an experiment is obviously a key issue. This paper introduces three design criteria used in optimal test design, based on Kullback-Leibler information, and compares them with the well-known D-efficiency criterion to obtain individually adapted choice designs for the mixed logit choice model. Being equally efficient to D-efficiency and at the same time much faster, the Kullback-Leibler criteria are well suited for the design of individualized choice experiments.

Risk Assessment Based on Limited Field Data

William Q. Meeker, Iowa State University

Nowadays, many consumer products are designed and manufactured so that the probability of failure during the technological life of the product is small. Most product units in the field retire before they fail. Even though the number of failures from such products is small, there is still the need to model and predict field failures in applications that involve safety and risk assessment. Challenges in modeling and predictions of failures arise because the retirement times are often unknown, and there are delays in field failure reporting. Motivated by an application to assess the risk of failure, we develop a statistical procedure to predict the field failures of products, which considers the impact of product retirements and reporting delays. Based on the developed method, we provide the point prediction for cumulative number of reported failures in a future time and the corresponding prediction interval to quantify uncertainty. We also conduct sensitivity analysis to show the effects of different assumptions on failure-time and retirement distributions and the values of their parameters. (This is joint work with Zhibing Xu and Yili Hong.)

Field Failure Prediction Based on Multi-Level Repair and System Usage Information

Yili Hong, Virginia Tech

Repairable systems in the field often receive repair actions at different levels. For example, a truck may have an engine replaced or have a component of the engine replaced. Thus we may consider three levels: system (truck), sub-system (engine), and component level. At system level, the system usage and environmental information can be available. At sub-system and component levels, the repair information can also be available through maintenance records. When the focus is on event process modeling and prediction of the component remaining life or replacements, multiple factors from different levels can affect the component failure process. In this paper, we develop a multi-level trend renewal process model to describe the event process for the component replacements. The proposed model can incorporate information from different levels and can explain the effects of factors at different levels on component failures. A field failure prediction procedure is also developed for the proposed model. (This is a joint work with Zhibing Xu and William Q. Meeker.)

Joint Modeling of Lab and Field Data with Application to Warranty Prediction for Highly Reliable Products

Sheng-Tsaing Tseng and Nan-Jung Hsu, National Tsing-Hua University

Motivated by a case study of smart phone, this talk addresses a prediction model for the field failure rates of multiple products which utilizes both the laboratory data and field failure data. The in-field laboratory test consists of go-no go attribute data, while the field data consist of early field failure warranty data (6 months in service). Due to the cost consideration, the test units in the laboratory test are very limited leading to very few failures or even no failures in general. For such sparse failure data, the conventional maximum likelihood inference will result in inaccurate estimation and predictions. To overcome this difficulty, we propose a prediction model of the products field failure rate (within a specific warranty period) by incorporating a hierarchical Bayesian procedure with the joint likelihood maximization. The methodology is applied to a real data set for illustration, in which the 83% of the variability in the mean failure rate can be successfully accounted by the proposed model.

A Window-Limited Generalized Likelihood Ratio Test for Monitoring Poisson Process with Linear Drifts

Honghao Zhao, Lianjie Shu, and Kwok-Leung Tsui, City University of Hong Kong

Besides monitoring Poisson process with step shifts, detecting Poisson mean shift subject to linear drifts attracts signicant attention in industry quality control and health care surveillance. A lot of methods have been proposed to monitor Poisson process mean with step shift. However their results do not guarantee good performance when they are implemented to detect the mean shift subject to linear drifts. In this paper, we formulate the generalized likelihood ratio (GLR) statistic to monitor Poisson rate under linear drifts. We discuss the design issue of the GLR method and compare the performance between the GLR method and several CUSUM-typed methods. Our results show that the GLR method presents generally better detection ability than the other alternative methods, especially when the true drift size is large. A simulated example is used to demonstrate the implementation of the proposed method.

Process Capability Indices and Control Charts That Are Based on the Highest Possibility Interval or Region

Jun Yang, Beihang University

For asymmetrical or multimodal distributions, there are often some out-of-control points are with higher possibility than some in-control points under the traditional natural tolerance or the probability control limits. This is unfair from the viewpoint of engineering. In order to solve this problem, the highest density interval is utilized to define the natural tolerance, which ensures that points in the natural tolerance are with higher possibility than those outside of the natural tolerance. Control limits are constructed by the highest possibility interval, which extends the existing Shewhart control limits. For a multivariate control chart, a control chart/region is similarly constructed from the highest possibility region, and it is argued that the traditional control chart is just a special case of this proposed multivariate control chart.

Nonparametric Statistical Process Control of Univariate Process

Zhonghua Li, Nankai University and Peihua Qiu, University of Florida

In this talk, some nonparametric statistical process control (SPC) control charts for univariate processes will be introduced and two newly proposed nonparametric SPC based on the idea of transformation and categorization, respectively, will be discussed in detail. Simulation comparisons and a real data example about daily exchange rates between the Korean Won and US Dollar show that our proposed control charts perform quite well in applications.

Estimating Component Reliability from System Failure-Time Data

Francisco J. Samaniego, University of California at Davis

The failure time data one collects from a sample of fielded systems provides indirect information about the performance of the systems components. Since it is often difficult to create or simulate field conditions in laboratory settings, the process of drawing inferences about component characteristics from data on system performance is of practical importance. Under the assumption that the system under study has components with i.i.d. lifetimes and a common distribution F, Bhattacharya and Samaniego (NRL, 2010) identified the asymptotic behavior of the nonparametric MLE of the underlying component reliability function $\bar{F}(t)$ and demonstrated the estimators good performance, even when sample sizes are moderate. Here, we extend this problem in two directions. First, we treat the estimation of $\bar{F}(t)$ under the same assumptions, but based on independent samples from $m \geq 2$ coherent systems. We compare the performance of two distinct approaches to estimating $\bar{F}(t)$: (1) convex combinations of the m individual NPMLEs of $\bar{F}(t)$ and (2) the estimator obtained from solving multiple point-wise maximum likelihood problems, i.e., one for each fixed time point t, and cobbling together these separate estimates to obtain an overall estimator of the function $\bar{F}(t)$. Both approaches lead to consistent, asymptotically normal estimators. The likelihood-based estimator is shown to be uniformly asymptotically superior. We then consider estimating $\bar{F}(t)$ when the system design is unknown. An unknown design can occur, for example, in military operations, where it is not uncommon to capture or gain control of a collection of like systems whose precise design is unknown. We assume that auxiliary data in the form of a variable K, the number of failed components at the time of system failure, is available from a subsequent autopsy. Given the data (T1,K1),..., (TN ,KN), one may estimate the unknown "signature" s of the system of interest (see Samaniego (2007)) and utilize the estimated reliability polynomial of the system to obtain a "Pseudo NPMLE" of $\bar{F}(t)$. This estimator is shown to be a consistent, asymptotically normal estimator of $\bar{F}(t)$. The performance of the latter estimator, which is based on the augmented data {(Ti,Ki)}, is compared to that of the estimator of $\bar{F}(t)$ based on the 10 data $\{T_i\}$ under the assumption that s is known. (Joint work with Peter Hall and Yin Jin.)

Some New Perspectives on the Signature of a Coherent System in Engineering Reliability

Bo H. Lindqvist, University of Trondheim, and Francisco J. Samaniego, University of California at Davis

The concept of signature of a coherent system with i.i.d. component lifetimes has been treated in several articles and in a recent monograph. A key result is the representation of the system survival distribution in terms of the signature vector (the index of the ordered component failure time that is fatal to the system). This leads to several results on stochastic comparison of system lifetimes, in particular for systems with different number of components. In this talk, we give a brief overview of the signature concept, aiming at complementing the theory in certain directions. One of these concerns the stochastic equivalence of systems of different sizes. While it is known that for a coherent system, the equivalent systems of larger size may be represented as stochastic mixtures of coherent systems, we show that they in fact can always be represented as single monotone systems with signature vectors that are easily calculated. Further, a necessary condition is obtained for the existence of a system with n - 1 components that stochastically dominates a given system of size n. The results are applied to optimization of the performance of systems under given cost constraints.

Tool Wear Diagnosis and Tool Replacement Management

Chen-ju Lin and Chun-hung Chien, Yuan Ze University

Determining the best time of tool replacement is critical to balancing production quality and tool utilization. A machining process would gradually produce defective workpieces as a tool wears out. To avoid additional production costs, replacing a tool before the yield drops below a minimum requirement is essential. On the other hand, frequent tool replacement would cause additional setup and tool costs. This paper proposes a system that diagnoses the state of tool wear and determines the time of tool replacement. The system constructs a hidden Markov model (HMM) to monitor the severity of tool wear by using the quality characteristics of products. The underlying number of states of tool wear is determined by using the Bayesian information criterion (BIC). Viterbi algorithm is then applied to estimate the tool wear of online data. A conditional yield of the next product is proposed to decide whether an operator should replace a tool or continue machining. The simulation analyses show that the accuracy of state estimation is above 90%. Furthermore, the decision process can make good use of tools whereas controlling yield. (Keywords: Hidden Markov Model, tool replacement, yield.)

The Inverse Gaussian Degradation Model

Zhisheng Ye, Hong Kong Polytechnic University, and Nan Chen, National University of Singapore

Degradation analysis has become a burgeoning research area in recent years to meet the requirement of complex systems. Previously, Wiener and Gamma process models have been extensively 12 studied for degradation modelling. This paper systematically introduces the third class, i.e., the inverse Gaussian (IG) process, as an effective degradation model. The IG process is shown to be a limiting compound Poisson process, which gives it a meaningful physical interpretation for modelling degradation of products deteriorating in random environments. Treated as the first passage process of a Wiener process, the IG process is flexible in incorporating random effects and covariates that account for heterogeneities commonly observed in degradations. This flexibility makes the class of IG process models much more attractive compared with the Gamma process, which has been thoroughly investigated in the literature of degradation modelling. The paper also discusses the statistical inference of the three random effects models and their model selections. It concludes with a real world example to demonstrate the applicability of the IG process in degradation modelling.

An Optimal Maintenance Policy for a Multi-state Deteriorating System

Mimi Zhang and M Xie, City University of Hong Kong, and Olivier Gaudoin, University of Grenoble, France

This paper proposes a sequential failure limit maintenance policy for a repairable system. The objective system is assumed to have k + 1 states, including one working state and k failure states, and the multiple failure states are classified potentially by features such as failure severity or failure cause. The system deteriorates over time and will be replaced upon the Nth failure. Corrective maintenance is performed immediately upon each of the first (N − 1) failures. To avoid the costly failure, preventive maintenance actions will be performed as soon as the systems reliability drops to a critical threshold R. Both preventive maintenance and corrective maintenance are assumed to be imperfect. Increasing and decreasing geometric processes are introduced to characterize the efficiency of preventive maintenance and corrective maintenance. The objective is to derive an optimal maintenance policy such that the long-run expected cost per unit time is minimized. The analytical expression of the cost rate function is derived, and the corresponding optimal maintenance policy can be determined numerically. A numerical example is given to illustrate the theoretical results and the maintenance procedure. The decision model shows its adaptability to different possible characteristics of the maintained system.

Minimax Space-Filling Designs for Finite Design Regions

Matthias H. Y. Tan, City University of Hong Kong

The problem of choosing a design that is representative of a finite candidate set is an important problem in computer experiments. The minimax criterion measures the degree of representativeness because it is the maximum distance of a candidate point to the design. A significant obstacle in the use of such designs is the lack of good construction algorithms. In this talk, we shall propose algorithms for finding minimax designs for finite design regions. We establish theoretical connections between minimax designs and the classical set covering location problem in operations research, 13 which is a binary linear program. These results are employed to design efficient procedures for finding globally optimal minimax and near-minimax designs.

Statistics Research in Prognostics and System Health

Kwok L. Tsui, City University of Hong Kong

Due to (i) concerns in product reliability, system safety, and failure prevention, and (ii) latest advancement in data collection technologies and modelling tools, there are tremendous opportunities in quantitative modelling research in prognostics and system health management (PHM) as well as system informatics (SI). In this talk we will present our view on research in data mining, surveillance, simulation, diagnostics and prognostics, and integration in system management. We will discuss important research problems in PHM and explain how they are connected to research in SPC, reliability, and industrial statistics.