International Workshop on Engineering Statistics

East China Normal University

Shanghai, China December 18, 2016

General Information

Organizing Committee Chairs:

Fugee Tsung, co-chair, Professor, Hong Kong University of Science and Technology

Xiaolong Pu, co-chair, Professor, East China Normal University

Organizing Committee Members:

Yukun Liu, Associate Professor, East China Normal University.

Dongdong Xiang, Assistant Professor, East China Normal University

Contact Information:

Yukun Liu: (86)13636472887, ykliu@sfs.ecnu.edu.cn

Dongdong Xiang: ddxiang@sfs.ecnu.edu.cn

Draft of Workshop Program

Saturday, December 17, 2016

9:00-18:00	**Registration** Location: Huhua International Hotel, 5858 South Hongmei Rd., Minhang District, Shanghai.
18:00	**Dinner** Location: Huhua International Hotel

Sunday, December 18, 2016

Location: Statistics Building, Room **105**, East China Normal University, 500 Dongchuan Rd., Minhang District, Shanghai.

Warm tip:
There will be a bus waiting for you outside the Huhua International Hotel at **7:50**.

8:20-8:30	**Opening Ceremony & Welcome Remark**
Session I	
8:30-8:55	Speaker 1: **Peihua Qiu**, University of Florida Title: Design SPC Charts Using p-Values and Process Monitoring Using a Dynamic Sampling Scheme
8:55-9:20	Speaker 2: **Amitava Mukherjee,** XLRI-Xavier School of Management Title: Condition Based Process Monitoring For Minimizing Fuzziness Surrounding Statistical Inference – A philosophical perspective with some examples and illustrations
9:20-9:45	Speaker 3: **Nan Chen,** National University of Singapore Title: Statistical Analysis of Simulation Output from Parallel Computing
9:45-10:10	Speaker 4: **Longcheen Huwang,** National Tsing Hua University Title: An EWMA chart for monitoring the covariance matrix of a multivariate process based on dissimilarity index
10:10-10:30	**Coffee/Tea Break**
Session II	
10:30-10:55	Speaker 5: **Min Xie,** City University of Hong Kong

Title: Modelling Bathtub Curve with Weibull Distribution for Lifetime Data Analysis

10:55-11:20 Speaker 6: **Xiaoming Huo,** Georgia Institute of Technology
Title:

11:20-11:45 Speaker 7: **Zhisheng Ye**, National University of Singapore
Title: Minimum Distance Estimation for the Generalized Pareto Distribution

11:45-12:10 Speaker 8: **Sheng-Tsaing Tseng,** National Tsing Hua University
Title: Nano-Sols Shelf-Life Prediction via Accelerated Degradation Model

12:10-14:10 **Lunch**

Session III

14:10-14:35 Speaker 9: **Jiahua Chen,** University of British Columbia
Title: Sequential Design for Binary Dose Response Experiments under Three Parameter Model

14:35-15:00 Speaker 10: **Lang Wu,** University of British Columbia
Title: Wood Property Relationships and Survival Models in Reliability

15:00-15:25 Speaker 11: **Suojin Wang,** Texas A&M University
Title: A new nonparametric test for checking the equality of the correlation structures of two time series

15:25-15:45 **Coffee/Tea Break**

Session IV

15:45-16:10 Speaker 12: **Yuejiao Fu,** York University
Title: Using differential variability to increase the power of the homogeneity test in a two-sample problem

16:10-16:35 Speaker 13: **Chengguo Weng**, University of Waterloo
Title: Regression Tree Credibility Model

16:35-17:00 Speaker 14: **Johannes Schmidt-Hieber,** University of Leiden
Title:

17:00-17:25 Speaker 15: **Teresa Oliveira,** University of Aberta
Title: p-Charts for Attribute Control considering life distributions and Using R

17:25-17:50 Speaker 16: **Amílcar Oliveira,** University of Aberta
Title: The role of np-Charts for Attribute Control based on Life Distributions

18:20 **Dinner**
Location: Huhua International Hotel

List of Abstract

Session I

Design SPC Charts Using p-Values and Process Monitoring Using a Dynamic Sampling Scheme

Peihua Qiu

Abstract: Conventional statistical process control (SPC) charts are designed using control limits; a chart gives a signal of process distributional shift when its charting statistic exceeds a properly chosen control limit. To do so, we only know whether a chart is out-of-control (OC) at a given time. It is therefore not informative enough about the likelihood of a potential distributional shift. In a recent paper, we suggested designing the SPC charts using p-values. By this approach, at each time point of process monitoring, the p-value of the observed charting statistic is computed, under the assumption that the process is in-control (IC). If the p-value is less than a pre-specified significance level, then a signal of distributional shift is delivered. This p-value approach has several benefits, compared to the conventional design using control limits. First, after a signal of distributional shift is delivered, we could know how strong the signal is. Second, even when the p-value at a given time point is larger than the significance level, it still provides us useful information about how stable the process performs at that time point. The second benefit is especially useful when we adopt a variable sampling scheme, by which the sampling time can be longer when we have more evidence that the process runs stably, supported by a larger p-value. A resulting dynamic sampling scheme will also be introduced in this talk. This is a joint research with Drs. Ansu Chatterjee, Zhonghua Li, and Zhaojun Wang.

Condition Based Process Monitoring For Minimizing Fuzziness Surrounding Statistical Inference – A philosophical perspective with some examples and illustrations

Amitava Mukherjee

Abstract: With significant development in the area of statistical process monitoring (SPM) over the last nine decades, more and more complexities creep both in the theory and methods related to the SPM procedures. While we have seen the development of a plethora of interesting tools and techniques for SPM, those, in turn,

often infuse many fuzzy areas within the domain of SPM and unfortunately, make a decision making process rather complicated. Starting from identification of appropriate of Phase-I distribution to choice of the best available SPM technique in Phase-II, there are certain degrees of fuzziness, in various areas. We provide some classic examples of illusions which lead to certain practical challenges both in the realms of parametric and nonparametric SPM. Certain conditions based techniques are discussed that can reduce the impact of fuzziness in some areas of SPM and ensure ease of decision making. Some philosophical perspectives are discussed rather than complex technicalities. Certain future research problems are identified.

Statistical Analysis of Simulation Output from Parallel Computing

Nan Chen

Abstract: This paper addresses the statistical output analysis of transient simulations in the parallel computing environment with fixed computing time. Using parallel computing, most unbiased estimators commonly used based on the output sequence compromise. To rectify this issue, this paper proposes an estimation procedure in the Bayesian framework. The proposed procedure is particularly useful when the computing time depends on the output value in each simulation replication. The effectiveness of our method is demonstrated through studies on queuing simulation and control chart simulation.

An EWMA chart for monitoring the covariance matrix of a multivariate process based on dissimilarity index

Longcheen Huwang

Abstract: In this talk, an EWMA chart for monitoring the covariance matrix of a multivariate process will be proposed based on the dissimilarity index of two matrices. Unlike the conventional EWMA charts for monitoring the covariance matrix which are either based on comparing the sum or the product or both of the eigenvalues of the estimated EWMA covariance matrix with those of the in-control (IC) covariance matrix which are assumed to be known. The proposed chart essentially monitors the covariance matrix by comparing the individual eigenvalues of the estimated EWMA covariance matrix with those of the estimated covariance matrix from the IC phase I data. The performance of the proposed chart will be compared with that of the best existing one when the process is represented by the multivariate normal distribution. The simulation results show that the proposed EMMA chart outperforms the best existing one for monitoring the covariance matrix in all situations considered. Further,

to prevent the control limit of the proposed chart developed using the limited IC phase I data from having too many false signals, a bootstrap re-sampling method is used to adjust the control limit to guarantee that the proposed chart has the actual IC ARL (average run length) not less than the nominal one with a certain high probability. Finally, an example is used to demonstrate the applicability and the implementation of the proposed chart.

Session II

Minimum Distance Estimation for the Generalized Pareto Distribution

Zhisheng Ye

Abstract: The generalized Pareto distribution (GPD) is widely used for extreme values over a threshold. Most existing methods for parameter estimation either perform unsatisfactorily when the shape parameter k is larger than 0.5, or they suffer from heavy computation as the sample size increases. In view of the fact that $k>0.5$ is not uncommon in numerous applications, including two illustrative examples used in this study, we remedy the deficiencies of existing methods by proposing two new estimators for the GPD parameters. The new estimators are inspired by the minimum distance estimation and the M-estimation in the linear regression. Large sample properties of the estimators are investigated. Through comprehensive simulation, the estimators are shown to perform well for all values of k under small and moderate sample sizes. They are comparable to the existing methods for $k<0.5$ while perform much better for $k>0.5$.

Nano-Sols Shelf-Life Prediction via Accelerated Degradation Model

Sheng-Tsaing Tseng

Abstract: In order to provide timely product's lifetime information to the customers, conventionally, manufacturers usually use temperature (or voltage) as an accelerating variable for shortening life testing time. Based on well-known life-stress relationship (such as Arrhenius reaction or inverse power model), it allows us to extrapolate the lifetime of highly-reliable products at a normal used condition. In this talk, however, we will present a real case study that the shelf-life prediction of nano-sol products can be successfully obtained by adopting pH value as an accelerating variable. An accelerated profile-degradation model is proposed to describe the time-evolution of the particle size distributions under three different pH values. Then, we can analytically construct the confidence interval for the shelf-life of nano-sol products under its normal use condition.

Session III

Sequential Design for Binary Dose Response Experiments under Three Parameter Model

Jiahua Chen

Abstract: In dose-response studies, experiments are often carried out according to an optimal design for the purpose of accurately locating a specific effective dose (ED) level. If the interest is in the dose-response relationship over a range of ED levels, classical optimality criteria are often mis-aligned. To cover this overlooked issue, a new criterion and a corresponding two- stage sequential ED-design have been proposed and found to have superior properties under the commonly used logistic and probit models. The use of logistic and probit models are mathematically convenient but they carry the risk of model mis-specification. To alleviate this risk, a three parameter logistic model may be used. We study the use of the two-stage sequential ED-design under the three parameter logistic model in this talk. We are pleased to find that the new design is as convenient to use and has anticipated efficient gains compared with other competing designs. This is a joint work with Xiaoli Yu, Department of Statistics, University of British Columbia.

Wood Property Relationships and Survival Models in Reliability

Lang Wu

Abstract: We study the relationship between lumber strength properties and their visual grading characteristics. This topic is central to the analysis of the reliability of lumber products in that it underlies the calculation of structural design values. The approaches described in the paper are adaptations of survival analysis methods commonly used in medical studies. Since each piece of lumber can only be tested to destruction with one method (i.e., each piece cannot be broken twice), modelling these strengths distributions simultaneously can be challenging. In the past, this kind of problem has been solved by subjectively matching pieces of lumber, but the quality of this approach is then an issue (however, see Evans, Johnson, and Green, 1984). The objective of our analysis is to build a predictive model that relates the strength properties to the recorded characteristics. The paper concludes that type of wood defect (knot), a lumber grade status (off~Vgrade: Yes/No) and a lumber~Rs module of elasticity (MOE) have statistically significant effects on wood strength. We find

that the Weibull accelerated failure time (AFT) model provides a better fit than the Cox proportional hazards (PH) model in our dataset.

A new nonparametric test for checking the equality of the correlation structures of two time series

Suojin Wang

Abstract: In this talk, we consider an order selection test to check the equality of two independent stationary time series in their correlation structures. The asymptotic distribution of the order selection test statistic under the null hypothesis is obtained. For many existing tests, consistency against general alternative hypotheses has not been established. On the other hand, we show that the proposed test is consistent not only under any fixed alternative hypothesis but also under a sequence of local alternative hypotheses. A simulation study is conducted to examine the finite sample performance of the test in comparison to some existing methods. We also apply the proposed test to an analysis of a biomedical data set.

Session IV

Using differential variability to increase the power of the homogeneity test in a two-sample problem

Yuejiao Fu

Abstract: We consider a particular two-sample homogeneity testing problem often encountered in case-control studies with contaminated controls, or in detecting a treatment effect when some subjects are not affected by the treatment in biological experiments. We propose an EM-test designed to simultaneously detect mean difference and differential variability in the two samples. We show that the EM-test statistic has a chi-squared null limiting distribution. The asymptotic properties of the EM-test under local alternatives are also investigated, and sample-size calculation is given. The main results are established for general location-scale family of distributions. Simulation results show that the EM-test outperforms the existing methods. Finally, two real data examples are used to illustrate the application of the proposed method.

Regression Tree Credibility Model

Chengguo Weng

Abstract: Credibility theory is viewed as cornerstone in actuarial science. This paper brings machine learning techniques into the area and proposes a novel credibility model and a credibility premium formula based on regression trees, which is called regression tree credibility (RTC) premium. The proposed RTC method first recursively binary partitions a collective of individual risks into exclusive sub-collectives using a new credibility regression tree algorithm based on credibility loss, and then applies the classical Buhlmann-Straub credibility formula to predict individual net premiums within each sub-collective. The proposed method effectively predicts individual net premiums by incorporating covariate information, and it is particularly appealing to capture various non-linear covariates effects and/or interaction effects because no specific regression form needs to be pre-specified in the method. Our proposed RTC method automatically selects influential covariate variables for premium prediction with no additional ex ante variable selection procedure required. The superiority in prediction accuracy of the proposed RTC model is demonstrated by extensive simulation studies.

p-Charts for Attribute Control considering life distributions and Using R

Teresa Oliveira

Abstract: Statistical Quality Control (SQC) is a crucial tool for quality improvement of products and services. In this talk we present charts for attribute control by means of the proportion (p) of defective items, named p-charts. Such charts were the first ones proposed by Walter A. Shewhart (1891-1967) and are often used in SQC to monitor the behavior of production processes by means of p over time, allowing the selection of unbalanced samples. When the sample size (n) is constant, charts for attribute control related to the number (np) of defective items, named np-charts, are a good alternative to the p-charts. We will review some topics on p-charts considering life distributions, see Leiva and Oliveira (2015). Some applications using R will be presented and the results will be illustrated and discussed.

The role of np-Charts for Attribute Control based on Life Distributions

Amílcar Oliveira

Abstract: In this talk we present charts for attribute control considering the number of defective items, called np-charts. The benefits of using np-charts instead of p-charts

are related to easier interpretation, once the sample size is constant and the fact that no calculation is required for each sample result.

An update to np-charts and some recent ideas on this topic are presented considering life distributions. We will show np-charts based on life distributions by an algorithm implemented in R language, and we will discuss examples using data of attributes and lifetimes.