

Workshop on Statistical Methodology and Applications to Biomedicine and Finance

In Honor of Professor Tze Leung Lai

Organizing Committee

Jie Chen, Novartis Pharmaceuticals

Yunxiao Chen, Columbia University

Qingfu Liu, Fudan University

Zhiliang Ying, Columbia University and Shanghai Center for Mathematical Sciences

Naiqing Zhao, Fudan University

Sponsor

Shanghai Center for Mathematical Sciences

Shanghai, China

December 4-5, 2015

This workshop provides a forum to review and discuss recent advances in statistical methodology and their biomedical and financial applications. It also celebrates the 70th birthday of Professor Tze Leung Lai, who has made seminal contributions to these areas.

SCHEDULE

Friday, Dec 4 Room 2001, Guanghua East Building, Fudan University

Morning Session I

| | |
|---------------|---|
| 8:20 - 8:30 | Opening Remark |
| 8:30 - 8:55 | Kani Chen, Hong Kong University of Science and Technology |
| 9:00 - 9:25 | Hong Chen, Shanghai Jiaotong University |
| 9:30 - 9:55 | Dongmei Han, Shanghai University of Finance and Economics |
| 10:00 - 10:30 | Coffee Break |

Morning Session II

| | |
|---------------|-----------------------------------|
| 10:30 - 10:55 | Xiaoqun Wang, Tsinghua University |
| 11:00 - 11:25 | Qingfu Liu, Fudan University |
| Lunch | |

Afternoon Session I

| | |
|-------------|--------------------------------------|
| 2:00 - 2:25 | Yuan Ji, University of Chicago |
| 2:30 - 2:55 | Jie Chen, Novartis Pharmaceuticals |
| 3:00 - 3:25 | Dejun Tang, Novartis Pharmaceuticals |
| 3:30 - 4:00 | Coffee Break |

Afternoon Session II

| | |
|-------------|---------------------------------------|
| 4:00 - 4:25 | Yuan-chin Ivan Chang, Academia Sinica |
| 4:30 - 4:55 | Zhongshan Yuan, Shandong University |
| Dinner | |

Saturday, Dec 5 Room 2001, Guanghua East Building, Fudan University

Morning Session I

| | |
|---------------|---|
| 8:30 - 8:55 | Zhengyan Lin, Zhejiang University |
| 9:00 - 9:25 | Dong Han, Shanghai Jiaotong University |
| 9:30 - 9:55 | Wei Jiang, Shanghai Jiaotong University |
| 10:00 - 10:30 | Coffee Break |

Morning Session II

| | |
|---------------|----------------------------------|
| 10:30 - 10:55 | Xinsheng Zhang, Fudan University |
| 11:00 - 11:25 | Zhongyi Zhu, Fudan University |
| Lunch | |

Afternoon Session I

| | |
|-------------|--------------------------------|
| 2:00 - 2:25 | Naiqing Zhao, Fudan University |
| 2:30 - 2:55 | Tony Guo, MSD China |
| 3:00 - 3:30 | Coffee Break |

Afternoon Session II

| | |
|-------------|--|
| 3:30 - 3:55 | Lingbo Yu, Roche Product Development in Asia Pacific |
| 4:00 - 4:25 | Chunquan Ou, Southern Medical University |

Abstracts

Yuan-chin Ivan Chang, Academia Sinica

Assessing the Predictive Power of Newly Added Biomarkers

As medical research and technology advance, there are always new biomarkers found and predictive models proposed for improving the diagnostic performance of diseases. Therefore, in addition to the existing biomarkers and predictive models, how to assess new biomarkers becomes an important research problem. Many classification performance measures, which are usually based on the performance on the whole cut-off values, were applied directly to this type of problems. However, in a medical diagnosis, some cut-off points are more important, such as those points within the range of high specificity. Thus, as the partial area under the ROC curve to the area under ROC curve, we study the partial integrated discriminant improvement (pIDI) for evaluating the predictive ability of a newly added marker at a pre-specified range of cut-offs. Theoretical property of estimate of the proposed measure is reported. The performance of this new measure is then compared with that of the partial area under an ROC curve. The numerical results use synthesized are presented, and a liver cancer dataset is used for demonstration purposes. (This is a co-work with Professor Zhanfeng Wang at University of Science Technology of China)

Hong Chen, Shanghai Jiaotong University

Default and Credit Risk Evaluation and Pricing - Big Data Approach

In this talk, I will present the motivation, initiative and preliminary results on the credit risk research I have been working on with my colleagues at SAIF.

Jie Chen, Novartis Pharmaceuticals

Sequential Testing of Disproportionality of Two Poisson Processes with Applications to Drug Safety?

Two sequences of independent, possibly non-homogeneous, Poisson counting processes are investigated sequentially to detect an increasing or decreasing trend of proportion for one process relative to another over a period of time. The problem is formulated as sequential testing of a null hypothesis of some types of odds ratios (OR). Conditional on both margins (subtotals of both processes in the subintervals and the sequential cumulative totals for each process), the random variables of events in the subintervals for one Poisson process follow a non-central multivariate hypergeometric distributions. Then the sequential testing of the null hypothesis is performed using sequential generalized likelihood ratio (sequential GLR) tests where explicit formulas are further developed and computational implementation is provided. The proposed methods are illustrated with an example in vaccine safety surveillance where an increasing reporting proportion of a rare adverse event is of interest.

Kani Chen, Hong Kong University of Science and Technology

Follow the Leader, or Follow the Loser?

It is always a mystery for portfolio managers whether following the leader or following the loser would achieve outstanding performance. Motivated by the universal portfolio and the aggregating algorithm, we propose the method of binary switch portfolio. The new method is based on a carefully designed classification rule and adaptively chooses the learning rate, a key tuning parameter in the aggregating algorithm, in an optimal fashion. We provide theoretical justification by giving a lower bound of the wealth, which is larger than that of the naive $1/N$ portfolio. Applications to real data show that the overall performance of the proposed binary switch portfolio is the best among several popular competing methods over varying time horizons and stock pools.

Tony Guo, MSD China

Consistent Evaluation for Treatment Effect in Multi-Regional Clinical Trials

In recent years, MRCTs were widely used in drug development to optimize valuable resource and further reduce the “drug lag” issue. However, how to evaluate the consistency of treatment effect among regions in a MRCT remains an unresolved issue and it may cause many review issues when a MRCT is used for new drug application in some regions. Recently, Harvard MRCT center and PKU have formed a joint working group to work on the consistency evaluation. In the presentation, the progress of the working group along with a new study design for consistency evaluation called nested group sequential SGDDP will be reported.

Dong Han, Shanghai Jiaotong University

Change-point Monitoring in Financial Sequences

In this talk we will mainly discuss on two kinds of control charts: reference-free cuscore chart and multi-chart, and use them to monitor the changes in IBM's stock returns and a financial network sequence.

Dongmei Han, Shanghai University of Finance and Economics

Effects of Shanghai Free Trade Zone on Economy: From a General Equilibrium View

We first construct Shanghai SAM(2010),then by constructing computable general equilibrium model (CGE), we investigate the economic impacts of Establishing Shanghai Free Trade Zone from a general equilibrium view. The main findings are: the increase of investment has a positive effect on the growth of GDP, the import and export, and the income of residents in Shanghai, and the promotion of the service industry is the most obvious. Logistics cost reduction has a significant promotion to the import and export of Shanghai. The effects on the GDP, the income of the residents are not obvious, while the effects on the second industry are obvious positive. The raise of TFP caused by investment, trade, finance and administration reform has positive effect on the economy of Shanghai.

Yuan Ji, University of Chicago

Bayesian Graphical Models with Application to Integrate Cancer Genomics Data

The Cancer Genomes Atlas (TCGA) data are unique in that multimodal measurements across genomics features, such as copy number, DNA methylation, and gene expression, are obtained on matched tumor samples. The multimodality provides an unprecedented opportunity to investigate the interplay of these features. Graphical models are powerful tools for this task that address the interaction of any two features in the presence of others, while traditional correlation- or regression-based models cannot. We introduce Zodiac, an online resource consisting of a large database containing nearly 200 million interaction networks of multiple genomics features produced by applying novel Bayesian graphical models on TCGA data through massively parallel computation. Setting a new way of integrating TCGA data, Zodiac, publically available at <http://www.compgenome.org/ZODIAC>, is expected to facilitate the generation of new knowledge and hypotheses by the community.

Wei Jiang, Shanghai Jiaotong University

Projection-based Process Monitoring based on Empirical Divergence

With the help of distributed sensing and high-speed wireless communication technologies, real-time data collection in complex system is becoming more and more common. The state of complex system are usually assessed through monitoring massive system quality characteristics. In order to efficiently monitor high dimensional data in complex system, projection methods are commonly used to reduce the dimensionality of process quality characteristics. Conventional projection methods such as projection pursuit (PP), have been used to estimate the process shift direction. However, implementing such methods highly rely on the normality assumption of a process. In this paper, we propose a general framework of projection-based methods in real-time monitoring by contrasting the newly observed data with the reference data. Because the degree of process deviating from in-control state can be measured as the divergence between the in-control distribution and the

distribution based on the newly observed data, the optimal projection is the direction in which the divergence between projected reference data and new observations is maximized. We further outline how the proposed method can be used in the on-line process monitoring. The performance of proposed monitoring procedure is compared with one class approaches such as SVDD-based method.

Zhengyan Lin, Zhejiang University

On Weak Convergence of Stochastic Processes to Stochastic Integrals

Weak convergence of various general functionals of partial sums of dependent random variables (statistics) to stochastic integrals now plays an important role in the modern probability theory and statistical theory. In this talk, we introduce the weak convergence of various general functionals of partial sums of causal processes to stochastic integrals driven by both the Brownian motion and Levy α -stable process.

Qingfu Liu, Fudan University

Arbitrage Activity and Price Discovery across CSI300 Spot, Futures and ETF Markets

We examine how does the introduction of exchange-traded fund (ETF) affect the arbitrage and price discovery mechanism between the China Securities Index (CSI) 300 spot and futures markets. We utilize a bivariate Smooth-Transition VECM (ST-VECM) to accommodate a two-speed error-correction mechanism. This allows us to differentiate price discovery contribution between no-arbitrage versus arbitrage states. Our analysis yields three main findings: i) Post-ETF trading, we see a substantial reduction in observed pricing errors. This is expected given the narrowing of arbitrage bounds due to lower transaction cost in trading ETFs; ii) The futures market still contributes more price discovery than its spot index and ETF counterparts; iii) Arbitrage activity migrated from the CSI300 spot predominately to the ETF traded in Shanghai, seemingly ignoring the ETF traded in Shenzhen. When arbitrage activity is present, the Gonzalo and Granger (GG 1995) price discovery contribution is a noisy measure since the VECM averages the error-correction mechanism between no-arbitrage and arbitrage states. A modified GG measure from the ST-VECM addresses this issue. We explain why the price discovery bound that corresponds to no-arbitrage state, provides a clearer indication of cross-market price discovery contribution.

Chunquan Ou, Southern Medical University

Assessing the Short-term Effects of Meteorological Factors and Air Pollution on Mortality

There are increasing concerns about urban air pollution and climate change. The assessment of their public health impacts is necessary for the development of interventions and measures targeted at the management and control of air pollution and extreme climate events. We obtained daily data of mortality, meteorological measures and concentrations of ambient air pollutants in Guangzhou, China, during 2003-2011. Quasi-Poisson regression models combined with distributed lag non-linear models were developed to assess the effects of air pollution and climatic factors on mortality. We found distributed lag effects of extremely low and high temperatures on mortality, and the magnitude and time course of the effects were modified by the levels of particulate matter. The short-term lag effects of other climatic factors and air pollution were also observed. Various stratified analyses were performed to identify susceptible subpopulations.

Dejun Tang, Novartis Pharmaceuticals

Some Applications for Quantitative Trial Design

There are many decision making points in drug development, which have significant impact to the quality and duration of clinical trials. To improve the possibility of making the right decision, the quantitative trial design is increasingly applied when designing clinical trials. It utilizes the innovative statistical methodology including modeling and simulation to quantify the decision making criteria, to assess the operational characteristics of the design, and to predict the outcomes of trial successes. This presentation is to provide some real case examples to demonstrate the principles and applications of the quantitative trial design.

Zhongshang Yuan, Shandong University

Chi-square Test for Group Difference of Pathways in Systems Epidemiology

Traditional epidemiology often pays more attention to the identification of a single factor than the pathway related to a disease, therefore with difficulty to deeply explore disease mechanism. Systems epidemiology aiming to integrate putative lifestyle exposures and biomarkers extracted from multiple omics platforms to offer new insights into the pathway mechanisms underlying disease on the human population level. One key but inadequately addressed question is to develop powerful statistic to identify whether one target pathway is associated with disease. Bearing in mind that pathway difference can result from not only changes of nodes but also changes of edges, we proposed a novel statistic for detecting the group difference of pathways, it can in principle capture the nodes change and edge change, as well as account for the pathway structure simultaneously. It follows chi-square distribution theoretically. Simulations showed it had better performance than the others

under various scenarios including only nodes change, only edges change and both nodes and edges change. Integrating genome-wide DNA methylation data, we analyzed one real data from Bogalusa cohort study and significantly identified a potential pathway from smoking to abdominal obesity through insulin signaling pathway.

Xiaoqun Wang, Tsinghua University

Handling Discontinuities in Financial Engineering: Good Path Simulation and Smoothing

Quasi-Monte Carlo (QMC) methods are deterministic versions of Monte Carlo methods and play an increasingly important role in the problems of pricing and hedging of complex financial derivatives. These problems are usually of high dimensionality and low smoothness (e.g., discontinuity), both of which may significantly affect the efficiency of QMC methods. This paper studies what defines a good path simulation method (PSM) for the underlying processes in problems with discontinuities and how to handle the challenging problems of high dimensionality and discontinuity concurrently. We develop a two-step procedure that perfectly combines dimension reduction and smoothing. Dimension reduction can be achieved by suitably generating the underlying processes and variable transformations, and discontinuities can be removed completely, significantly improving the smoothness of the functions. Numerical experiments demonstrate the extreme high performance of the proposed method for pricing options and estimating Greeks (or sensitivities).

Lingbo Yu, Roche Product Development in Asia Pacific

Statistical Considerations on Biomarker Analyses

Using biomarkers to identify patients who benefit most from a new drug is integral to achieving personalized health care, and biomarkers have the potential to revolutionise therapy. Thus clinical trial designs that are driven by predictive biomarkers to enable both evaluation of new treatments and identification of the patient subgroups might be more efficient. However it is a quite complex process of developing companion diagnosis, and there are many associated issues of Biomarker analyses. This talk will discuss basic concept of biomarkers, common pitfalls in biomarker analyses, and trial design considerations.

Naiqing Zhao, Fudan University

The Design and Statistical Consideration for Clinical Trials with Concomitant Medications as Strong Time-dependent Confounders

Zhongyi Zhu, Fudan University

Functional Linear Models with Fixed Effects

In this paper, we introduce a functional linear model with fixed effects for functional data where the predictor and response are random processes. The proposed model can be viewed as a generalization of the classical functional linear model, and characterizes individual specific source of variability. We implement the regularity procedure through a projection on the eigenfunction basis of the response process, leading to a special version of linear mixed effects model for panel data. In order to deal with the difficulty caused by a large number of individual effects, we use the penalty method to shrink individual effects, and propose a class of penalized least squares estimators. In a theoretical investigation, we establish asymptotic normality for the deviation between estimated and true regression function coefficients, and derive some asymptotic consistent properties for the predictions obtained from the fitted functional linear models with fixed effects. Some simulation studies and an application of intra-day volatility patterns of the *S&P 500* index are conducted to illustrate the finite sample performance of the proposed modeling framework and estimation methods.

Xinsheng Zhang, Fudan University

Testing the Equality of Large U-Statistic Based Correlation Matrices

In this talk, we provide a framework for testing the equality of two large U-statistic based correlation matrices, which include the rank-based correlation matrices as special cases. Our approach exploits extreme value statistics and the Jackknife estimator for uncertainty assessment and is valid under a fully nonparametric model. Theoretically, we develop a theory for testing the equality of U-statistic based correlation matrices. We then apply this theory to study the problem of testing large Kendalls tau correlation matrices and demonstrate its optimality. For proving this optimality, a novel construction of least favourable distributions is developed for the correlation matrix comparison. (This is joint work with Cheng Zhou, Fang Han, and Han Liu)
