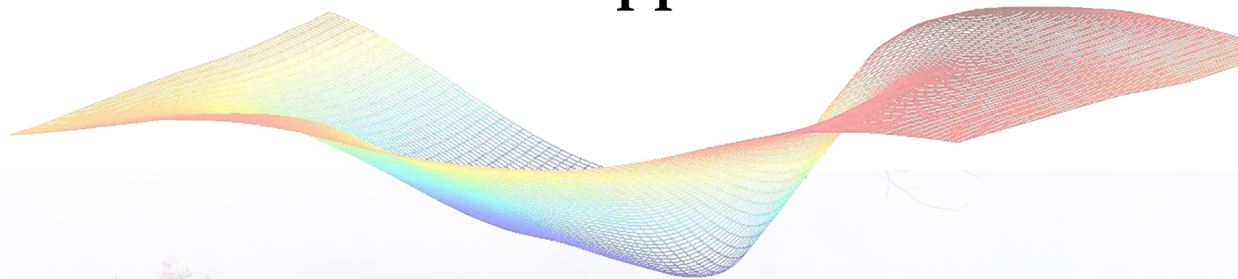


# Lilac International Conference on Application of Statistics



## LICAS 2019



June 24-25, 2019  
Harbin, China



### **Scientific Committee (members ordered by surname)**

Zhidong Bai	Honorary Chair	Northeast Normal University
T. Tony Cai	Co-Chair	University of Pennsylvania
Jun S. Liu	Co-Chair	Harvard University
Jianqing Fan		Princeton University
Jiti Gao		Monash University
Jianwei Ma		Harbin Institute of Technology
Boying Wu		Harbin Institute of Technology
Liming Wu		University of Clermont-Ferrand
Quanhua Xu		Harbin Institute of Technology
Xiaoping Xue		Harbin Institute of Technology
Lijian Yang		Tsinghua University
Qiang Ye		Harbin Institute of Technology
Guohua Zou		Chinese Academy of Sciences

### **Organizing Committee (members ordered by surname)**

Lijian Yang	Co-Chair	Tsinghua University
Ke Li	Co-Chair	Harbin Institute of Technology
Boping Tian	Co-Chair	Harbin Institute of Technology
Jiaqi Chen		Harbin Institute of Technology
Minghao Chen		Harbin Institute of Technology
Ru Fang		Harbin Institute of Technology
Yongqiang Fu		Harbin Institute of Technology
Zhong Guan		Indiana University South Bend
Heming Jiao		Harbin Institute of Technology
Shuilin Jin		Harbin Institute of Technology
Guanfeng Li		Harbin Institute of Technology
Longsuo Li		Harbin Institute of Technology
Yanpeng Li		Harbin Institute of Technology
Zhaoyan Li		Harbin Institute of Technology
Wei Liu		Harbin Institute of Technology
Zhenfeng Shi		Harbin Institute of Technology
Guangqiang Teng		Harbin Institute of Technology
Fenglin Tian		Harbin Institute of Technology
Li Wang		Harbin Institute of Technology
Haiyu Wen		Harbin Institute of Technology
Yu Xiao		Harbin Institute of Technology
Yuming Xing		Harbin Institute of Technology
Chiping Zhang		Harbin Institute of Technology
Dazhi Zhang		Harbin Institute of Technology
Fengmiao Zhang		Harbin Institute of Technology
Yangchun Zhang		Harbin Institute of Technology



# Conference Program

June 24th – 25th, 2019

Activity Centre 214, HIT, China

<b>June 23rd</b>	<b>Xiyuan Hotel, 西苑宾馆</b>
08:00-22:00	<b>Registration</b>
<b>June 24th</b>	<b>Activity Centre, 活动中心 214</b>
08:00-08:15	<b>Opening Ceremony</b>
08:15-08:40	<b>Group Photo</b>
08:40-10:25	<b>Session 1</b>
08:40-09:25	<b>Keynote Speech 1</b> Chair: T. Tony Cai (University of Pennsylvania) Generalized 4 moment Theorem with Application to Random Matrix Theory Speaker: Zhidong Bai (Northeast Normal University)
09:25-09:55	<b>Speech 2</b> Chair: Yi Zhao (Harbin Institute of Technology) Bivariate Continuous ARMA random fields Speaker: Yasumasa Matsuda (Tohoku University)
09:55-10:25	<b>Speech 3</b> Chair: Yi Zhao (Harbin Institute of Technology) Asymptotics in Quantum Hypothesis Testing Speaker: Ke Li (Harbin Institute of Technology)
10:25-10:35	<b>Tea Break</b>
10:35-12:20	<b>Session 2</b>
10:35-11:20	<b>Keynote Speech 4</b> Chair: Guohua Zou (Chinese Academy of Sciences) When Statistics Meets Computing Speaker: T. Tony Cai (University of Pennsylvania)
11:20-11:50	<b>Speech 5</b> Chair: Xinbing Kong (Nanjing Audit University) Maximum Approximate Bernstein Likelihood Estimation in Proportional Hazard Model for Interval-Censored Data Speaker: Zhong Guan (Indiana University South Bend)
11:50-12:20	<b>Speech 6</b> Chair: Xinbing Kong (Nanjing Audit University) Realized volatility matrix under ultra-high-frequency financial data Speaker: Zhi Liu (University of Macau)
12:20-14:00	<b>Lunch (Xiyuan Hotel, 西苑宾馆三楼大厅)</b>

14:00-16:15	<b>Session 3</b>
14:00-14:45	<b>Keynote Speech 7</b> Chair: Lijian Yang (Tsinghua University) Neural Networks Meet Statistics Speaker: Jun S. Liu (Harvard University)
14:45-15:15	<b>Speech 8</b> Chair: Li Wang ( Iowa State University) Progress on reciprocal characterization and equivalent transformation between complex networks and time series Speaker: Yi Zhao (Harbin Institute of Technology)
15:15-15:45	<b>Speech 9</b> Chair: Li Wang ( Iowa State University) Determining the number of communities in degree-corrected stochastic block models Speaker: Shujie Ma (University of California, Riverside)
15:45-16:15	<b>Speech 10</b> Chair: Li Wang ( Iowa State University) Identification of Trans-eQTLs Using Mediation Analysis with Multiple Mediators Speaker: Lin Hou (Tsinghua University)
16:15-16:25	<b>Tea Break</b>
16:25-18:10	<b>Session 4</b>
16:25-17:10	<b>Keynote Speech 11</b> Chair: Quanhua Xu (Harbin Institute of Technology) Maximum Entropy Estimation versus Classical Maximum Likelihood Estimation for Hidden Markov Chains Speaker: Liming Wu (University of Clermont-Ferrand)
17:10-17:40	<b>Speech 12</b> Chair: Shujie Ma (University of California, Riverside) Sufficient Dimension Reduction for Classification Speaker: Xin Chen (Southern University of Science & Technology)
17:40-18:10	<b>Speech 13</b> Chair: Shujie Ma (University of California, Riverside) Cream skimming: Theory and evidence from hospital transfers and capacity utilization Speaker: Ou Yang (University of Melbourne)
After 18:10	<b>Dinner</b> (Xiyuan Hotel, 西苑宾馆三楼大厅)

<b>June 25th</b>	<b>Activity Centre, 活动中心 214</b>
08:00-10:15	<b>Session 5</b>
08:00-08:45	<b>Keynote Speech 14</b> Chair: Jian Shi (Chinese Academy of Sciences) Model averaging prediction for time series models with a diverging number of parameters Speaker: Guohua Zou (Chinese Academy of Sciences)
08:45-09:15	<b>Speech 15</b> Chair: Zhong Guan (Indiana University South Bend) Multivariate Spline Estimation and Inference for Image-on-scalar Regression Speaker: Li Wang (Iowa State University)
09:15-09:45	<b>Speech 16</b> Chair: Zhong Guan (Indiana University South Bend) Estimation of a Mixed Bernstein Polynomial Model with Multidimensional Variable Density Speaker: Hongliang Liu (Harbin Normal University)
09:45-10:15	<b>Speech 17</b> Chair: Zhong Guan (Indiana University South Bend) Community Detection Based on the $L_\infty$ convergence of eigenvectors in DCBM Speaker: Yan Liu (Northeast Normal University)
10:15-10:25	<b>Tea Break</b>
10:25-12:10	<b>Session 6</b>
10:25-11:10	<b>Keynote Speech 18</b> Chair: Zhi Liu (University of Macau) Discrepancy between global and local principal component analysis on large-panel high-frequency data Speaker: Xinbing Kong (Nanjing Audit University)
11:10-11:40	<b>Speech 19</b> Chair: Xin Chen (Southern University of Science & Technology) Simultaneous Confidence Corridors for Mean Functions in Functional Data Analysis of Imaging Data Speaker: Guannan Wang (College of William & Mary)
11:40-12:10	<b>Speech 20</b> Chair: Xin Chen (Southern University of Science & Technology) Power of linear spectral statistics for testing the equality of two high dimensional covariance matrices and general Beta matrix Speaker: Xiaozhuo Zhang (Northeast Normal University)
12:10-14:00	<b>Lunch (Xiyuan Hotel, 西苑宾馆三楼大厅)</b>
14:00-15:30	<b>Session 7</b>
14:00-14:30	<b>Speech 21</b> Chair: Boping Tian (Harbin Institute of Technology)

	Modelling of Wear Rate with Oil Addition Disturbance Speaker: Jian Shi (Chinese Academy of Sciences)
14:30-15:00	<b>Speech 22</b> Chair: Heming Jiao (Harbin Institute of Technology) Optimal Subsampling Inference in the Big Data Era Speaker: Huiming Zhang (Peking University)
15:00-15:30	<b>Speech 23</b> Chair: Heming Jiao (Harbin Institute of Technology) A supplement on CLT for LSS under a large dimensional generalized spiked covariance model Speaker: Yangchun Zhang (Harbin Institute of Technology)
15:30-15:40	<b>Tea Break</b>
15:40-17:10	<b>Session 8</b>
15:40-16:10	<b>Speech 24</b> Chair: Yongchun Zhou (Harbin Institute of Technology) 金融风控模型应用介绍 Speaker: Kai Zhang (Ronghui Jinke)
16:10-16:40	<b>Speech 25</b> Chair: Haiyu Wen (Harbin Institute of Technology) 概率统计与人工智能之计算机视觉应用 Speaker: Quanhua Xue (ZTE Jetflow)
16:40-17:10	<b>Speech 26</b> Chair: Haiyu Wen (Harbin Institute of Technology) 结合集成学习的文本挖掘在信贷评分上的应用 Speaker: Yi Bo (Beijing Finup Group)
After 17:10	<b>Dinner</b> (Xiyuan Hotel, 西苑宾馆三楼大厅)

# ABSTRACTS

**Keynote Speech 1 (45 minutes)**

## **Generalized 4 moment Theorem with Application to Random Matrix Theory**

**Zhidong Bai (Northeast Normal University)**

**Abstract:** Many results in random matrix theory show universality, that is, the result is true for all distributions provided that their first 4 moments are matching with normal distribution. On the other hand, the universality provides an approach to investigate properties of RMT, that is, show universality first, then investigate the property under normality assumption. Recently, Tao and Vu proved a FMT for Wigner matrix. But they assumed much heavy assumptions, say the condition  $C_0$ . Now, we consider the CLT for normalized spiked eigenvalues, we shall show FMT first and then investigate the CLT of spiked eigenvalues. Because we only assume the match of first two moments and a proper of tail probability,  $x^4 P(|X_{ij}| > x) \rightarrow 0$ , which implies all moments less than 4 exist but the 4th moment may not exist. So, we call it GFMT.

**Personal Introduction:** Zhidong Bai has been an Elected Fellow of the Third World Academy of Sciences since 1989. Since 2002, he has been a Distinguished Professor of School of Mathematics and Statistics at Northeast Normal University, China. Since 1999, he has been a professor of Department of Statistics and Applied Probability at National University of Singapore. He received his doctorate from the Department of Mathematics in University of Science and Technology of China in May, 1982. And he was in the first batch of 18 PhDs in the history of the People's Republic of China. He is also a Fellow of International Statistical Institute, a Fellow of Institute of Mathematical Statistics, a Fellow of International Chinese Statistical Association, a Member of Chinese Society of Mathematics, executive director of Chinese Society of Probability and Statistics. He won the second prize in State Natural Science Award of China in 2012. He was selected as one of the Most Cited Chinese Researchers by Elsevier in 2016. He and Jack Silverstein published the classic monographs *Spectral analysis of large dimensional random matrices*. He is an

editor of *Random Matrices: Theory and Applications*, and was successively an editor of *Journal of Multivariate Analysis*, an associate editor of *Statistica Sinica*, *Journal of statistical planning and inference*.

**个人简介:** 白志东, 第三世界科学院院士, 2002 年至今任东北师范大学数学系特聘教授, 1999 年至今任新加坡国立大学概率与统计系教授。于 1982 年 5 月在中国科技大学数学系获得博士学位, 中国首批十八位博士之一。学术兼职包括国际统计协会 Fellow、国际数理统计 Fellow、国际泛华统计协会 Fellow、中国数学协会会员、中国概率统计协会常务理事。2012 年获国家自然科学奖二等奖。入选 2016 年中国高被引学者 (Most Cited Chinese Researchers) 榜单 (爱思唯尔, Elsevier)。2010 年与 Jack Silverstein 合作出版经典专著 *Spectral analysis of large dimensional random matrices*。现任 *Random Matrices: Theory and Applications* 主编, 曾担任 *Journal of Multivariate Analysis* 主编, *Statistica Sinica* 编委, *Journal of statistical planning and inference* 编委。

## Speech 2 (30 minutes)

### Bivariate Continuous ARMA random fields

Yasumasa Matsuda (Tohoku University)

**Abstract:** Brockwell and Matsuda (2017) extended CARMA models for time series to those for random fields, which we call as “CARMA random fields”. In this talk, we consider a bivariate extension of CARMA random fields to analyze spatially scattered bivariate observations. After defining bivariate CARMA random fields, we introduce Whittle likelihoods to estimate the parameters with applications to imputation for missing components of bivariate observations. There is a gap between discrete observations and continuous models to conduct the imputation. We employ Bayesian ways to fill the gap. We demonstrate them by applying to bivariate samples of precipitation and elevation over 7000 irregularly scattered observatory points in US continent.

**Personal Introduction:** Yasumasa Matsuda, professor of Economic Statistics, Graduate School of Economics and Management, Tohoku University. Research Topics include Fourier analysis of spatial data, analysis of financial time series by stochastic volatility models, graphical modeling for multivariate time



series. He awarded JSPS prize in 2010. He is associate editor for *Annals of the Institute of the Statistical Mathematics*.

**个人简介:** Yasumasa Matsuda, 日本东北大学经济统计系教授。主要研究空间统计的傅里叶分析、基于随机波动率模型的金融时间序列分析、多元时间序列的图建模等。2010 年获 JSPS 奖。现任《*Annals of the Institute of the Statistical Mathematics*》编委。

### Speech 3 (30 minutes)

#### Asymptotics in Quantum Hypothesis Testing

Ke Li (Harbin Institute of Technology)

**Abstract:** At first, I will review the mathematical foundations of quantum mechanics, and explain why we are interested in probability and statistics in the quantum world. Then, I will report two results in the asymptotics in quantum hypothesis testing: (1) quantum multiple Chernoff distance in the symmetric setting; (2) second-order asymptotics in the asymmetric setting.

**Personal Introduction:** Ke Li is a Professor at the Institute for Advanced Study in Mathematics, Harbin Institute of Technology (HIT). He has been selected into the Young overseas high-level talents introduction plan of China. Before joining HIT, he worked as a Postdoctoral Scholar at Caltech (Supervisor: Fernando Brandao), a Senior Research Associate jointly at IBM Watson Research Center and MIT (Supervisors: Charles Bennett, Aram Harrow, Graeme Smith, John Smolin), and a Research Fellow at NUS (Supervisor: Andreas Winter). He is mainly engaged in the study of quantum probability, statistics and information theory. His research is published in the leading journals of mathematics and physics, such as *Ann. Statist.*, *Comm. Math. Phys.*, *IEEE Tran. Inf. Theory*, *Phys. Rev. Lett.*, *Nature Phys.*

**个人简介:** 李科, 哈尔滨工业大学数学研究院教授, 入选中国青年千人计划。在加入 HIT 之前, 他曾在加州理工学院担任博士后学者(导师: Fernando Brandao), 他是 IBM 沃森研究中心和麻省理工学院的高级研究员(导师: Charles Bennett, Aram Harrow, Graeme Smith, John Smolin), 以及新加坡国立大学研究员(导师: Andreas Winter)。主要从事量子概率, 统计学和信息理论的研究。他的研究成果发表在如 *Ann. Statist.*, *Comm. Math. Phys.*,

*IEEE Tran. Inf. Theory, Phys. Rev. Lett., Nature Phys.* 等顶级数学和物理学期刊上。

**Keynote Speech 4 (45 minutes)**

**When Statistics Meets Computing**

**T. Tony Cai (University of Pennsylvania)**

**Abstract:** In the conventional statistical framework, the goal is developing optimal inference procedures, where optimality is understood with respect to the sample size and parameter space. When the dimensionality of the data becomes large as in many contemporary applications, the computational concerns associated with the statistical procedures come to the forefront. A fundamental question is: Is there a price to pay for statistical performance if one only considers computable (polynomial-time) procedures? After all, statistical methods are useful in practice only if they can be computed within a reasonable amount of time. In this talk, we discuss the interplay between statistical accuracy and computational efficiency in two specific problems: submatrix localization and sparse matrix detection based on a noisy observation of a large matrix. The results show some interesting phenomena that are quite different from other high-dimensional problems studied in the literature.

**Personal Introduction:** T. Tony Cai, professor and vice dean of the Wharton School, professor of Applied Mathematics & Computational Science Graduate Group, Associate Scholar of Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine in University of Pennsylvania, USA. He won the IMS Fellow in 2006, COPSS Presidents' Award (the highest prize in the field of statistics) in 2008, and IMS Medallion Lecturer in 2009. And he was served as president of the International Chinese Statistical Association in 2017. He has received the National Science Foundation Fund for eight consecutive times since 2000, the National Institutes of Health Foundation in 2012 and 2017, and the Wharton Global Initiative Fund in 2016. He was the editor of *Annals of Statistics*, and the associate editor of *Journal of the Royal Statistical Society: Series B*, *Journal of the American Statistical Association*, *Statistica Sinica*.

**个人简介:** 蔡天文, T. Tony Cai, 现任美国宾夕法尼亚大学沃顿商学院教授、副院长, 宾夕法尼亚大学应用数学及计算科学教授; 宾夕法尼亚大学医学院生物统计、流行病学及信息学系资深学者。2006 年荣获国际数理统计学会会士 (IMS Fellow), 2008 年荣获国际统计学最高奖考普斯“总统奖” (COPSS Presidents’ Award), 2009 年荣获国际数理统计协会 (IMS) 的 Medallion Lecturer, 2017 年任泛华统计学会主席。他自 2000 年以来, 连续 8 次获得美国国家科学基金会基金, 2012 年及 2017 年获得美国国家卫生研究院基金, 2016 年获得沃顿学院全球倡议基金。曾任《*Annals of Statistics*》主编、编委,《*Journal of the Royal Statistical Society: Series B*》编委,《*Journal of the American Statistical Association*》编委,《*Statistica Sinica*》编委。

**Speech 5 (30 minutes)**

**Maximum Approximate Bernstein Likelihood Estimation in  
Proportional Hazard Model for Interval-Censored Data**

**Zhong Guan (Indiana University South Bend)**

**Abstract:** Maximum approximate Bernstein likelihood estimates of density function and regression coefficients in the proportional hazard regression models based on interval-censored event time data are proposed and studied. A smooth estimates of the survival and density functions are then obtained. Simulation study is conducted to show the finite sample performance of the proposed method. The proposed method is illustrated by real data applications.

**Personal Introduction:** Zhong Guan is Full Professor with Tenure of Department of Mathematical Sciences in Indiana University South Bend, Doctor of Statistics, University of Toledo, Postdoctoral fellow in Division of Biostatistics, Yale University, a member of American Statistical Association, a member of American Mathematical Society. He was the Secretary General of the Probability and Statistics Society of Heilongjiang Province, a member of the board of Directors of the Heilongjiang Province Branch of the Applied Statistics Society of China.

**个人简介:** 关忠, 印第安纳大学南本德分校终身教授, 美国托莱多大学统计学博士, 耶鲁大学生物统计系博士后, 美国统计协会会员, 美国数学协会会员。曾任黑龙江省概率统计协会秘书长, 中国应用统计协会黑龙江省分会理事。

## Speech 6 (30 minutes)

### Realized volatility matrix under ultra-high-frequency financial data

Zhi Liu (University of Macau)

**Abstract:** The statistical inference of volatility matrix suffers from the effects of variate features of ultra-high-frequency data, such as market microstructure noise, asynchronous trading, multiple records, etc. In particular, the tick-by-tick transaction records are typically asynchronous among different assets and multiple records frequently appear. In this talk, I will review the existing feasible approaches. These methods, however, have to discard a large part of data, to keep the estimators consistent. Then I will introduce a data-efficient estimator of the realized volatility matrix which utilizes all of the data. The estimator is created by combining the congregation of the data within the synchronized time intervals and pre-averaging smoothing methods. We have established the asymptotic normality of proposed estimator. A studentized version of the central limit theorem is also proposed. I will also show improvement in estimation efficiency. Through a variety of synthetic data experiments, we assess the finite sample performance of proposed estimator and make comparison with existing methods. Finally, we implement the estimator with real NASDAQ high-frequency financial data sets.

**Personal Introduction:** Zhi Liu is an Associate Professor in Department of Mathematics, University of Macau. After graduating from the Hong Kong University of Science and Technology in 2011, he is mainly engaged in research on stochastic process statistics, financial statistics, bioinformatics and high-dimensional statistics, and has published many articles in international top journals in the fields of statistics, finance and bioinformatics.

**个人简介:** 刘志, 澳门大学数学系副教授。2011 年博士毕业于香港科技大学后, 主要从事随机过程统计、金融统计、生物信息学和高维统计等方面的研究, 并在统计学、金融和生物信息方面的国际顶级期刊上发表论文多篇。

**Keynote Speech 7 (45 minutes)**  
**Neural Networks Meet Statistics**  
**Jun S. Liu (Harvard University)**

**Abstract:** In recent years we have witnessed the tremendous advances of the neural-network-based machine learning methods and the excellent representation power of deep neural networks. We have recently been interested in investigating how we might take advantages of NNs for typical statistical problems. Here I will use a few examples to explain how ideas in DNN can inspire novel statistical methods, and how statistical ideas can help with NN constructions. One problem is related to uncertainty quantification and hierarchical Bayes analysis, for which we introduce the predictive-matching Generative Parameter Sampler (GPS). The GPS adopts a hierarchical modeling perspective such that each observation is modeled by an individual parameter and trains a NN to generate predictive distribution that marches the empirical distribution of the observed data. Our approach is illustrated for linear models, Poisson processes, and DNNs for classification. The results show that the GPS is successful in providing uncertainty quantification as well as additional flexibility beyond what is allowed by classical statistical procedures. Another problem is about neural network compression. We introduce a statistical principle to guide for pruning the trained network. Specifically, we keep the connections of a DNN that are significantly related to its outputs using a non-parametric penalized likelihood ratio test. Experimental results show that our approach achieves better lossless compression rates than existing techniques. The presentation is based on joint work with Minsuk Shin and Xing Xin.

**Personal Introduction:** Jun S. Liu, professor in Department of Statistics from 2000, professor of Biostatistics from 2001, Harvard University, USA, guest professor in Center for Statistical Science of Tsinghua University, China. He won the CAREER Award of National Science Foundation, 1995-1998 and Terman Fellow, Stanford University, 1995-1998. And he also won Mitchell Prize for the Best Bayesian Application Paper in 2000, COPSS Presidents' Award (the highest prize in the field of statistics) and IMS Medallion Lecturer in 2002. He was elected as IMS Fellow in 2004 and ASA Fellow in 2005.

And he won Morningside Gold Metal for Applied Mathematics in 2010, Distinguished Achievement Award of International Chinese Statistical Association in 2012, and Pao-Lu Hsu Award of International Chinese Statistical Association in 2016. He has been the associate editor of *Statistica Sinica* since 1999, and he was the Co-Editor and associate editor of *Journal of the American Statistical Association* and associate editor of *Biometrics*.

**个人简介:** 刘军, Jun S. Liu, 2000 年起任哈佛大学统计学系教授, 并兼任生物统计学系教授, 2014 年起任清华大学统计学研究中心客座教授。1995 年荣获美国国家科学基金会事业奖 (NSF CAREER Award) 和斯坦福大学特曼奖 (Stanford Terman Fellowship)。2000 年获得米切尔奖 (Mitchell Award)。2002 年荣获国际统计学最高奖考普斯“总统奖”(COPSS Presidents' Award), 以及国际数理统计协会 (IMS) 的 Medallion Lecturer。2004 年被选为国际数理统计学会会士 (IMS Fellow), 2005 年被美国统计协会选为会士 (ASA FELLOW)。2010 年荣获晨兴应用数学金奖 (Morningside Gold Medal)。2012 年荣获泛华统计协会颁发的 2012 年杰出成就奖 (Distinguished Achievement Award 2012)。2016 年荣获泛华统计协会颁发的“许宝騄奖”。1999 年至今任《*Statistica Sinica*》编委, 曾任《*Journal of the American Statistical Association*》主编、编委, 《*Biometrics*》编委。

### Speech 8 (30 minutes)

#### Progress on reciprocal characterization and equivalent transformation between complex networks and time series

Yi Zhao (Harbin Institute of Technology)

**Abstract:** The complex networks and time series are two classical paradigms to describe the real complex systems. But the complexity of systems determines that we merely can obtain the specific properties of the object given the single paradigm. To address this problem, recently transformation between complex networks and time series becomes promising research. People begin to pay much attention to their internal relationship. Inspired by the previous work, we emphasize the theoretical foundation of equivalent transformation between complex networks and time series so as to ensure the consistency of dynamics of complex system during transformation, thereby providing theoretical evidences for their reciprocal characterization. Based on that, we confirm



and quantify the corresponding relationship between both counterparts and provide characterization methods for their integrated application in practice so as to comprehensively understand complex systems from the dual perspectives of complex networks and time series.

**Personal Introduction:** Yi Zhao is a Professor of Shenzhen Graduate School, Harbin Institute of Technology and the Director of Applied Mathematics Research Center of Shenzhen Graduate School, Harbin Institute of Technology. In January 2007, he received his Ph.D. from the Hong Kong Polytechnic University. His research interests include nonlinear time series analysis, nonlinear dynamic systems, complex networks, and biomathematics. He has published 21 international journal articles about non-linear dynamic feature technology and biomedical applications, and 23 papers in international conferences, and obtained a national invention patent and a utility model patent authorization.

**个人简介:** 赵毅, 哈尔滨工业大学深圳研究生院教授, 哈尔滨工业大学深圳研究生院应用数学研究中心主任。2007 年 1 月获得香港理工大学博士学位, 研究方向包括非线性时间序列分析、非线性动力系统、复杂网络和生物数学。在非线形动力特征技术及在生物医学应用方面发表国际期刊论文二十一篇, 国际会议论文二十三篇, 获得国家发明专利和实用新型专利授权各一项。

### **Speech 9 (30 minutes)**

#### **Determining the number of communities in degree-corrected stochastic block models**

**Shujie Ma (University of California, Riverside)**

**Abstract:** We propose to estimate the number of communities in degree-corrected stochastic block models based on a pseudo likelihood ratio. For estimation, we consider a spectral clustering together with binary segmentation method. This approach guarantees an upper bound for the pseudo likelihood ratio statistic when the model is over-fitted. We also derive its limiting distribution when the model is under-fitted. Based on these properties, we establish the consistency of our estimator for the true number of communities. Developing these theoretical properties require a mild condition on the average degree – growing at a rate of  $\log(n)$ , where  $n$  is the number of nodes. Our proposed method is further illustrated by simulation studies and analysis of real-world

networks. The numerical results show that our approach has satisfactory performance when the network is sparse and/or has unbalanced communities.

**Personal Introduction:** Shujie Ma is an Associate Professor in Department of Statistics, University of California, Riverside. She is the associate editor of *Journal of Business & Economic Statistics*, *Computational Statistics and Data Analysis*, *The American Statistician*, *Statistica Sinica*, *Journal of Statistical Planning and Inference*. Her major fields are precision (personalized) medicine; factor models, large-scale data analysis; inference of high-dimensional data, functional data and nonlinear time series data; asymptotic theory, stochastic processes, extreme value theory; applications to gene-environment interaction, environmental risk assessment, medicine and financial data.

**个人简介:** 马舒洁, 加州大学河滨分校统计系副教授, *Journal of Business & Economic Statistics*, *Computational Statistics and Data Analysis*, *The American Statistician*, *Statistica Sinica*, *Journal of Statistical Planning and Inference* 的编委。主要研究领域为精准医学; 因子模型, 大数据分析; 高维数据、函数型数据和非线性时间序列数据分析; 渐近理论, 随机过程, 极值理论; 基因-环境互动、环境风险评估、医学和金融数据的应用。

**Speech 10 (30 minutes)**

**Identification of Trans-eQTLs Using Mediation Analysis with Multiple Mediators**

**Lin Hou (Tsinghua University)**

**Abstract:** Identification of expression quantitative trait loci (eQTLs) advances our understanding of genetics and regulatory mechanisms of gene expression in various organisms (Veyrieras et al., 2008). Previous studies suggest that trans-eQTLs may regulate expression of remote genes by altering the expression of nearby genes. Trans-association has been studied in the mediation analysis with a single mediator. However, prior applications with one mediator are prone to model misspecification due to correlations between genes. Motivated from the observation that trans-eQTLs are more likely to associate with more than one cis-gene than randomly selected SNPs in the GTEx dataset, we developed a computational method to identify trans-eQTLs that are mediated by multiple mediators. In simulation studies, multiple mediator analysis had

increased power to detect mediated trans-eQTLs, especially in large samples. In the HapMap3 data, we identified 11 mediated trans-eQTLs that were not detected by the single mediator analysis in the combined samples of African populations. Moreover, the mediated trans-eQTLs in the HapMap3 samples are more likely to be trait-associated SNPs. Our approach has improved the power of detecting mediated trans-eQTLs and advanced knowledge of gene regulation.

**Personal Introduction:** Lin Hou is an Associate Professor at the Center for Statistical Research in Tsinghua University. She got doctorate from Peking University in 2011 and worked as a postdoctoral researcher at the Yale School of Public Health from 2012 to 2015. After joining Tsinghua University in 2015, her main research areas are statistical genetics, bioinformatics and applied statistics.

**个人简介:** 侯琳, 清华大学统计学研究中心副教授。2011 年博士毕业于北京大学, 2012-2015 年间在耶鲁大学公共卫生学院做博士后研究。2015 年加入清华大学, 主要研究领域为统计遗传学、生物信息学和应用统计学。

### **Keynote Speech 11 (45 minutes)**

#### **Maximum Entropy Estimation versus Classical Maximum**

#### **Likelihood Estimation for Hidden Markov Chains**

**Liming Wu (University of Clermont-Ferrand)**

**Abstract:** Hidden Markov chains have many applications in economic, speech or gene recognition. However, its likelihood function can not be calculated when the sample is large, although there are related algorithms, but its computation is very large and slow. We first transform the high-dimensional problem into a two-dimensional problem, propose a maximum entropy estimator, prove its central limit theorem, and give its application in parameter estimation and hypothesis testing. The corresponding algorithm is also introduced, which is much faster than the known algorithm.

**Personal Introduction:** Liming Wu, professor in the Department of Mathematics, University of Clermont-Ferrand, France. His main research interests include large deviations theory and its applications in mathematical physics

and statistics, uniqueness and spectral analysis of operator semigroups, several probability analysis inequalities. He has independently solved Varadhan's guess, the unique conjecture of the quantum field ground state diffusion process, and cooperatively solved the Gross conjecture, introduced and established the concept and theory of consistent integrable operators, the unique concepts and theories of  $L_1$  and  $L_\infty$ . His research results published in top journals such as *Annals of Probability*, *Probability Theory and Related Fields*, *Journal of Functional Analysis*, *Annales de l'institut Henri Poincare (B) Probability and Statistics*.

**个人简介:** 吴黎明, 现任法国 Clermont-Ferrand 大学数学系教授。主要研究大偏差理论及其在数学物理、统计中的应用, 算子半群的唯一性、谱分析, 几种概率分析不等式。曾独立解决 Varadhan 猜测, 量子场基态扩散过程的唯一性猜想, 以及合作解决 Gross 猜想, 引入并建立一致可积算子概念及理论,  $L_1$  和  $L_\infty$  的唯一性概念及理论等。研究成果发表在 *Annals of Probability*, *Probability Theory and Related Fields*, *Journal of Functional Analysis*, *Annales de l'institut Henri Poincare (B) Probability and Statistics* 等顶级期刊上。

### Speech 12 (30 minutes)

#### Sufficient Dimension Reduction for Classification

Xin Chen (Southern University of Science & Technology)

**Abstract:** In this talk, we talk about a new sufficient dimension reduction approach designed deliberately for high-dimensional classification. This novel method is named maximal mean variance (MMV), inspired by the mean variance index first proposed by Cui, Li and Zhong (2015), which measures the dependence between a categorical random variable with multiple classes and a continuous random variable. Our method requires reasonably mild restrictions on the predicting variables and keeps the model-free advantage without the need to estimate the link function. Our method works pretty well when  $n < p$ . The surprising classification efficiency gain of the proposed method is demonstrated by simulation studies and real data analysis.

**Personal Introduction:** Xin Chen is an Associate Professor in Department of Mathematics, Southern University of Science & Technology, China. He got doctorate in University of Minnesota in 2010. Research interests: sufficient

dimension reduction, variable selection, high dimensional analysis, complex data analysis.

**个人简介:** 陈欣, 南方科技大学数学系副教授, 2010 年获得明尼苏达大学的博士学位。研究方向: 充分降维、变量选择、高维分析和复杂数据分析。

### **Speech 13 (30 minutes)**

#### **Cream skimming: Theory and evidence from hospital transfers and capacity utilization**

**Ou Yang (The University of Melbourne)**

**Abstract:** The paper examines cream skimming behaviour by studying hospital transfers in a mixed public-private hospital system. It proposes using capacity utilization to identify cream skimming. We develop a dynamic model with uncertain patient arrivals and hospital capacity constraints to clarify the conditions under which a profit maximizing hospital will engage in patient selection by transferring ‘hard’ patients—those with severe/complex conditions—to free up capacity to accommodate ‘easy’ patients who have less severe/complex conditions. Given finite capacity, public hospitals are strictly less likely to transfer patients than profit-motivated private hospitals at the same level of capacity. We test implications of the model using hospital administrative data from Victoria, Australia, and find empirical support to the theoretical predictions about cream skimming.

**Personal Introduction:** Ou Yang, Research Fellow at Melbourne Institute of Applied Economics and Social Research in The University of Melbourne, Senior Statistical Consultant, in Victoria Department of Health and Human Services, Australia. Research interests: applied economics, consumer behaviors. Society affiliation: Econometric Society; Australian Health Economics Society.

**个人简介:** 杨欧, 现任墨尔本大学应用经济与社会研究所研究员, 澳大利亚维多利亚卫生及公共服务部高级统计师。研究兴趣包括应用计量经济学、卫生经济学、消费者行为。社会兼职: 计量经济协会、澳大利亚卫生经济协会。

**Keynote Speech 14 (45 minutes)**

**Model averaging prediction for time series models with a diverging  
number of parameters**

**Guohua Zou (Chinese Academy of Sciences)**

**Abstract:** An important problem with model averaging approach is the choice of weights. In this paper, a generalized Mallows model averaging (GMMA) criterion for choosing weights is developed in the context of an infinite order autoregressive ( $AR(\infty)$ ) process. The GMMA method adapts to the circumstances in which the dimensions of candidate models can be large and increase with the sample size. The GMMA method is shown to be asymptotically optimal in the sense of obtaining the best out-of-sample mean-squared prediction error (MSPE) for both the independent-realization and the same-realization predictions, which, as a byproduct, solves a conjecture put forward by Hansen (2008) that the well-known Mallows model averaging (MMA) criterion from Hansen (2007) is asymptotically optimal for predicting the future of a times series. The rate of the GMMA based weight estimator tending to the optimal weight vector minimizing the independent-realization MSPE is derived as well. Both simulation experiment and real data analysis illustrate the merits of GMMA method in the prediction of  $AR(\infty)$  process.

**Personal Introduction:**Guohua Zou, researcher in Academy of Mathematics and Systems Science of Chinese Academy of Sciences, distinguished professor of Capital Normal University, China. He has got National Science Fund for Distinguished Young Scholars, and was elected as a member of New Century National Hundred, Thousand and Ten Thousand Talent Project, a member of Hundred-Talent Program in Chinese Academy of Sciences, and enjoys the special government allowance of the State Council of PRC. He is the fellow of International Statistical Institute, executive director of National Statistical Society of China, executive director of Chinese Association for Applied Statistics, associate editor of *Model Assisted Statistics and Applications*. He has presided over or participated in more than 20 national natural science fund projects and national practical projects, and the proposed forecasting methods have been adopted by the actual departments.



**个人简介：**邹国华，现任中国科学院数学与系统科学研究院研究员、首都师范大学特聘教授。国家杰出青年基金获得者、“新世纪百千万人才工程”国家级人选、中国科学院“百人计划”入选者、享受国务院政府特殊津贴。现任国际统计学会（International Statistical Institute）推举委员、中国统计学会常务理事、中国现场统计研究会常务理事、《*Model Assisted Statistics and Applications*》编委。主持或参加过二十多项国家自然科学基金项目以及全国性的实际课题，提出的预测方法被实际部门所采用。

**Speech 15 (30 minutes)**

**Multivariate Spline Estimation and Inference for Image-on-scalar  
Regression**

**Li Wang (Iowa State University)**

**Abstract:** Motivated by recent work of analyzing data in the biomedical imaging studies, we consider a class of image-on-scalar regression models for imaging responses and scalar predictors. We propose to use flexible multivariate splines over triangulations to handle the irregular domain of the objects of interest on the images and other characteristics of images. The proposed estimators of the coefficient functions are proved to be root-n consistent and asymptotically normal under some regularity conditions. We also provide a consistent and computationally efficient estimator of the covariance function. Asymptotic pointwise confidence intervals and data-driven simultaneous confidence corridors for the coefficient functions are constructed. Our method can simultaneously estimate and make inferences of the coefficient functions while incorporating the spatial heterogeneity and spatial correlation. Highly efficient and scalable estimation algorithm is developed. Monte Carlo simulation studies are conducted to examine the finite-sample performance of the proposed method. The proposed method is applied to the spatially normalized Positron Emission Tomography (PET) data of Alzheimer’s Disease Neuroimaging Initiative.

**Personal Introduction:** Li Wang is an Associate Professor in Department of Statistics, Iowa State University. Her primary interests lie in developing flexible, robust and computationally efficient statistical methodologies for solving problems arising from large-scale data with complex features, such as

high-dimensional, longitudinal, functional, heterogeneous or spatially and/or temporally correlated data. These statistical methodologies have important applications in many areas, such as engineering, genetics, geography, financial economics, meteorology, environmental, neuroimaging and biomedical science.

**个人简介:** 王丽, 爱荷华州立大学统计系副教授。主要兴趣在于开发灵活稳健且计算有效的统计方法, 以解决具有诸如高维数据、纵向数据、函数性数据以及异构或空间和/或时间相关的数等复杂特征的大数据问题。应用领域: 工程学, 遗传学, 地理学, 金融经济学, 气象学, 环境学, 神经影像学和生物医学科学等。

### **Speech 16 (30 minutes)**

#### **Estimation of a Mixed Bernstein Polynomial Model with Multidimensional Variable Density**

**Hongliang Liu (Harbin Normal University)**

**Abstract:** On the basis of density function estimation method, we mainly study the problem of estimating density functions by using multidimensional Bernstein polynomials defined on simplex. Different from previous studies on multidimensional density estimation, the multidimensional density estimation studied in this paper is based on the simplex region, and there are certain constraints between variables. Inspired by the Bernstein polynomial estimation model in onedimensional case, the mixed model of multidimensional Bernstein polynomials can be used to estimate the unknown multidimensional density, and the result is fine. This paper mainly focuses on two-dimensional situation. The EM algorithm is used to calculate the optimal two-dimensional Bernstein mixed model in the sense of maximum likelihood. Using the samples collected under the density function to fit, it can be seen that when the order  $m$  increases, the change of maximum likelihood function value of the hybrid model decreases gradually. At the same time, it can be seen that the mean square error between the estimated value and the real data decreases first and then increases. Then the optimal order of the two-dimensional Bernstein polynomial mixed model can be found by the method of change point. Combining with fitting simulation, we can see that the fitting effect of the model is good. Based on the two-dimensional Bernstein polynomial asymptotic expansion on

simplex, the asymptotic effect of the mixed model is proved. On the basis of the law of large numbers and the central limit theorem, the convergence of the mixed model under large samples is illustrated.

**Personal Introduction:** Hongliang Liu is an Associate Professor at Harbin Normal University. And he got doctorate in Center for Control Theory and Guidance Technology, Harbin Institute of Technology. His research areas are stability analysis and control of stochastic systems and time-delay systems.

**个人简介:** 刘宏亮, 哈尔滨师范大学副教授, 哈尔滨工业大学控制理论与制导技术中心博士。研究方向: 随机系统的稳定性分析与控制和时滞系统的稳定性分析与控制。

### Speech 17 (30 minutes)

#### Community Detection Based on the $L_\infty$ convergence of eigenvectors in DCBM

Yan Liu (Northeast Normal University)

**Abstract:** Spectral clustering is one of the most popular algorithms for community detection in network analysis. Based on this rationale, in this paper we give the convergence rate of eigenvectors for the adjacency matrix in the  $l_\infty$  norm, under the stochastic block model (BM) and degree corrected stochastic block model (DCBM), adding some mild and rational conditions. We also extend this result to a more general model, presented based on the DCBM such that the value of random variables in the adjacency matrix is not 0 or 1, but an arbitrary real number. During the process of proving the above conclusion, we obtain the relationship of the eigenvalues in the adjacency matrix and the corresponding ‘population’ matrix, which vary in dimension from the community-wise edge probability matrix. Using that result, we can give an estimate of the number of the communities in a known set of network data. To solve the problem that how to determine. Meanwhile we proved the consistency of the estimator. Furthermore, according to the derivation of proof for the convergence of eigenvectors, we propose a new approach to community detection – Spectral Clustering based on Difference of Ratios of Eigenvectors (SCDRE). Our simulation experiments demonstrate the superiority of our method in community detection.

**Personal Introduction:** Yan Liu, PhD of Statistics, study in School of Mathematics and Statistics, Northeast Normal University. Tutored by Zhi-dong Bai, the academician of The Third World Academy of Sciences. She studies in Northeast Normal University since 2014. The main research direction is the theory of large dimensional random matrix and related applications.  
**个人简介:** 刘研, 东北师范大学数学与统计学院, 统计学博士, 师从白志东院士, 2014 年至今就读于东北师范大学。主要研究方向是大维随机矩阵理论和相关应用。

**Keynote Speech 18 (45 minutes)**

**Discrepancy between global and local principal component analysis  
on large-panel high-frequency data  
Xinbing Kong (Nanjing Audit University)**

**Abstract:** The global principal component analysis (GPCA), PCA directly applied to the whole sample, is not reliable to reconstruct the common components of a large panel of high-frequency data when the factor loadings are time-varying, but it works when the factor loadings are constant. However, the local principal component analysis (LPCA) presented in Kong (2017)(2018) results in consistent estimates of the common components even if the factor loading processes follow Itô semimartingales. The LPCA is also suited for on-line computation in “big data” framework with restricted storage and memory. This motivates us to study the discrepancy between the GPCA and LPCA in recovering the common components of the large-panel high-frequency data. In this paper, we measure the discrepancy by the total sum of squared differences between common components reconstructed from GPCA and LPCA. We provide the asymptotic distribution of the discrepancy measure when the factor loadings are constant. Alternatively when some factor loadings are time-varying, the discrepancy measure explodes in a rate higher than  $\sqrt{pk_n}$  under some mild conditions on the time-variation magnitude of the factor loadings where  $k_n$  is the size of each subsample. We then apply the theory to testing the hypothesis that the factor loading matrix is a constant matrix. We show that the test performs well in controlling the type I error and detecting

time-varying loadings. Our real data analysis provides evidence that the factor loading matrices are always time-varying.

**Personal Introduction:** Xinbing Kong, professor in Institute of Statistical Science and Big Data, Nanjing Audit University from 2017, China. He was elected as a member of International Statistical Institute in 2016, a member of innovative and entrepreneurial talent plan in Jiangsu Province, a member of high-level talent shortage plan for universities and research institutes in Suzhou. He won Hong Kong Mathematical Society Best Doctoral Thesis Award and Youth New Star Award in School of Management, Fudan University. He is associate editor of *Random Matrices: Theory and Applications*.

**个人简介:** 孔新兵, 2017 年至今任南京审计大学统计科学与大数据研究院教授。他是国际统计学会 (International Statistical Institute, ISI) 的推选会员 (2016), 入选江苏省双创计划, 苏州市高等院校、科研院所高层次紧缺人才计划, 曾获得香港数学会最佳博士论文奖、复旦大学管理学院青年新星奖。现任《*Random Matrices: Theory and Applications*》编委。

### Speech 19 (30 minutes)

#### Simultaneous Confidence Corridors for Mean Functions in Functional Data Analysis of Imaging Data Guannan Wang (College of William & Mary)

**Abstract:** Motivated by recent work involving the analysis of biomedical imaging data, we present a novel procedure for constructing simultaneous confidence corridors for the mean of imaging data. We propose to use flexible bivariate splines over triangulations to handle irregular domain of the images that is common in brain imaging studies and in other biomedical imaging applications. The proposed spline estimators of the mean functions are shown to be consistent and asymptotically normal under some regularity conditions. We also provide a computationally efficient estimator of the covariance function and derive its uniform consistency. The procedure is also extended to the two-sample case in which we focus on comparing the mean functions from two populations of imaging data. Through Monte Carlo simulation studies we examine the finite-sample performance of the proposed method. Finally, the proposed method is applied to analyze brain Positron Emission Tomography

(PET) data in two different studies. One dataset used in preparation of this article was obtained from the Alzheimer's Disease Neuro-imaging Initiative (ADNI) database.

**Personal Introduction:** Guannan Wang is an Assistant Professor in Department of Mathematics, College of William & Mary. She graduated with a Ph.D. degree from Department of Statistics, University of Georgia in 2015. Research interests: statistical learning, large scale data analysis, non-/semi-parametric statistics, survey sampling, spatial data analysis.

**个人简介:** 王冠男, 美国威廉玛丽学院数学系助理教授。她于 2015 年毕业于乔治亚大学统计系, 获得博士学位。研究兴趣: 统计学习; 大数据分析; 非/半参数统计; 调查抽样; 空间数据分析。

### Speech 20 (30 minutes)

#### Power of linear spectral statistics for testing the equality of two high dimensional covariance matrices and general Beta matrix

Xiaozhuo Zhang (Northeast Normal University)

**Abstract:** Let general Beta matrix as  $B_n = S_n(S_n + \alpha_n(T_p^{-\frac{1}{2}})M_N(T_p^{-\frac{1}{2}})^*)^{-1}$ , where  $S_n, M_N$  are two independent sample covariance matrices with dimension  $p$  and sample size  $n$  and  $N$ , respectively,  $T_p$  is nonnegative Hermitian matrix whose inverse is bounded in spectral norm. We obtain some properties of general Beta matrix including its Limit Spectral Distribution (LSD) and the Central Limit Theorem (CLT) of its Linear Spectral Statistics (LSS) in this paper, moreover, we apply those results to test the equivalence of two high-dimensional covariance matrix. Specially, we can describe the power of some LSS with explicit mathematical expressions under any alternative hypothesis.

**Personal Introduction:** Xiaozhuo Zhang is a PhD student in the School of Mathematics and Statistics, Northeast Normal University. She has been tutored by Zhidong Bai since 2016. Her work focuses on the theory of large dimensional random matrix and related applications. She won the 2017 Master's National Scholarship.

**个人简介:** 张晓琢, 东北师范大学数学与统计学院, 统计学博士。2016 年至今就读于东北师范大学, 师从白志东院士。主要研究方向是大维随机矩阵理论和相关应用。曾获 2017 年度硕士研究生国家奖学金。



**Speech 21 (30 minutes)**

**Modelling of Wear Rate with Oil Addition Disturbance**

**Jian Shi (Chinese Academy of Sciences)**

**Abstract:** Wear has negative effect on the function of equipment, and therefore has attracted much attention in engineering. Conventionally, wear rate can be reflected by concentrations of some particular particles. In this talk, we will introduce a model for particle concentration in the circumstance of oil addition. Three kinds of data collection are considered. An efficient algorithm is developed when the oil addition information is completely missing. A real example is analyzed.

**Personal Introduction:** Dr. Jian Shi, graduated from Peking University, is a Professor at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. His research interests include statistical inference, biomedical statistics, industrial statistics and statistics in sports. He has held and participated in several projects of the National Natural Science Foundation of China as well as applied projects.

**个人简介:** 石坚, 博士, 毕业于北京大学, 现为中国科学院数学与系统科学研究院研究员, 博士生导师。主要研究方向有统计推断、生物医学统计、工业统计、体育统计。承担和参加过国家自然科学基金面上与重点项目, 以及其它应用课题的研究。

**Speech 22 (30 minutes)**

**Optimal Subsampling Inference in the Big Data Era**

**Huiming Zhang (Peking University)**

**Abstract:** To fast approximate the MLE with massive data, this paper study the optimal subsampling inference under the A-optimality criterion for generalized linear models (GLMs) with non-natural links as an importance example. The consistency and asymptotic normality of the estimator from the general subsampling estimators are established, and optimal subsampling probabilities under the A- and L-optimality criteria are derived. Furthermore, using Frobenius norm matrix concentration inequality, finite sample properties of the subsampling estimator based on optimal subsampling probabilities are also

derived. Since the optimal subsampling probabilities depend on the full data estimate, an adaptive two-step method is developed. Asymptotic normality and optimality of the sub-sampling estimator from this adaptive method are established. The proposed methods are illustrated and evaluated through numerical experiments on simulated and real datasets.

**Personal Introduction:** Huiming Zhang is a Ph.D. candidate in Statistics from Peking University. Research interests: high-dimensional statistics and probability, functional data, statistical models relating to count data and counting process, infinitely divisible and Lévy processes.

**个人简介:** 张慧铭, 北京大学统计学博士。研究兴趣: 高维统计和概率, 函数性数据, 与计数数据和计数过程相关的统计模型, 无穷可分与 Lévy 过程。

### Speech 23 (30 minutes)

#### A supplement on CLT for LSS under a large dimensional generalized spiked covariance model

Yangchun Zhang (Harbin Institute of Technology)

**Abstract:** Central limit theorem (CLT) for linear spectral statistics (LSSs) is widely used in large scale statistical inference when the sample size  $n$  and dimension  $p$  both tend to infinity. However, there always exists discrepancy between the sample mean and sample variance, and asymptotic mean and asymptotic variance when the CLT is applied for an LSS under spiked models. A major portion of the discrepancy is from the spiked eigenvalues, which depends on the dimensions  $(p, n)$  and the magnitudes of the spikes. In order to eliminate such discrepancy, we propose in this paper a supplement to the CLT defined as  $H_p$  CLT for a class of LSSs of sample covariance matrices. Simulation results demonstrate the success of the  $H_p$  CLT and exhibit its superiority to the original ones in various situations.

**Personal Introduction:** Yangchun Zhang is a Ph.D. candidate in Harbin Institute of Technology. Research interests: random matrix, extreme eigenvalues, etc. He visited the National University of Singapore via CSC from January 2018 to January 2019, and visited the Department of Mathematics, Hong Kong University of Science and Technology in June 2019.

**个人简介:** 张阳春, 哈尔滨工业大学在读博士, 研究兴趣: 随机矩阵, 极值特征值等。2018.01-2019.01 新加坡国立大学联合培养, 2019.06 香港科技大学数学系访学。

**Speech 24 (30 minutes)**

**金融风控模型应用介绍**

**Kai Zhang (Ronghui Jinke)**

**Abstract:** 业界应用比较广的主流统计模型算法介绍, 包括深度神经网络 (DNN)、梯度迭代决策树 (GBDT)、最近邻居法 (KNN)、逻辑回归 (LR), 和无监督评分卡等, 在精准营销、反欺诈、信用评估、额度授予等场景的应用, 涉及金融信贷、交易、理财等业界领域。

**个人简介:** 张凯, 融慧金科产品研究总监。本科毕业于哈尔滨工业大学; 博士毕业于美国密歇根州立大学。曾担任美国运通公司高级风控经理; 百度资深研发工程师。

**Speech 25 (30 minutes)**

**概率统计与人工智能之计算机视觉应用**

**Quanhua Xue (ZTE Jetflow)**

**Abstract:** 介绍概率统计在人工智能不同领域中的应用, 并结合公路、铁路等行业产品, 着重阐述计算机视觉使用到的概率统计学理论方法。

**个人简介:** 薛全华, 现任中兴飞流信息科技公司算法工程师。于 2012 年硕士毕业于哈尔滨工业大学数学系, 曾就职于中国电子科技集团公司第二十八研究所, 上海拓攻机器人有限公司, 主要从事区域网络化作战信息数据融合算法和人工智能算法。此后, 他加入中兴通讯先后从事视频智能分析, 深度学习平台构建及大数据分析等工作。

**Speech 26 (30 minutes)**

**结合集成学习的文本挖掘在信贷评分上的应用**

**Yi Bo (Beijing Finup Group)**

**Abstract:** 我们结合集成学习方法, 在金融借贷场景进行文本挖掘, 对比传统统计指标和朴素贝叶斯方法获得了较大的提升, 优化了业务和风险。对比用 wiki 预训练的 word2vec 向量建模效果相差不多, 且模型规模较小, 易于部署

和解释。我们先对每个材料进行分词, 筛选出一些高频词。然后同时抽样词与样本, 使用树深为 1 的 XGBoost 训练若干个模型, 并将他们分别拍平, 转化为类似 LR 模型的系数。接下来将这些 LR 模型的系数做平均, 即 Bagging 在一起。最后使用 Blending 的方法, 在留出样本上将每个材料的子模型组合在一起。应用此方法的模型很稳定, 上线半年几乎无衰减。且得到的子模型作为指标可以服务更多的模型, 甚至在做迁移学习时仍有较好的区分度。

**个人简介:** 伯毅, 北京凡普金科集团有限公司模型分析师。2018 年毕业于哈尔滨工业大学, 获得应用统计硕士学位。主要研究领域为 XGboost 算法, 金融建模, 深度学习, 文本分析等。





## Organizer

School of Mathematics, Harbin Institute of Technology

## Co-Organizers

Institute for Advanced Study in Mathematics, Harbin Institute of Technology

Center for Statistical Science, Tsinghua University

School of Mathematics and Statistics, Northeast Normal University

## Logistic Support

Yongchun Zhou (周永春) : 13654681396

Guangqiang Teng (滕广强) : 18845610601

Yangchun Zhang (张阳春) : 15765524268

Fenglin Tian (田枫林) : 18645081779

Yanpeng Li (李艳鹏) : 15045649975

