温州大学数理与电子信息工程学院现有教职工213人，其中专任教师170人，教授30人、副教授72人，高级职称比例为48%；博士110人，博士学位比例为52%。拥有“长江学者”特聘教授、“百千万人才工程”国家级人选、国家有突出贡献中青年专家、享受国务院政府特殊津贴专家等。学院现有数学、物理学、电子科学与技术等7个一级学科。其中，数学是浙江省重点学科。学院设有数学与应用数学、应用统计学、物理学、电子工程及其自动化、计算机科学与技术、信息与通信工程等10个本科专业。其中数学为国家特色专业；数学与应用数学为教育部首批专业。学院还设有学院计算中心、信息工程学院等教学平台。
会议概述

会议名称
统计与数据科学国际学术会议

会议主题
统计研究以及在大数据应用的未来趋势

会议时间
2017年12月22日-24日

会议地点
中国•浙江•温州

主办单位
温州大学    Florida Atlantic University

承办单位
温州大学数理与电子信息工程学院
<table>
<thead>
<tr>
<th>时间</th>
<th>内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00-8:30</td>
<td>宾馆乘校车至会场</td>
</tr>
<tr>
<td>8:30-8:40</td>
<td>校领导致辞 主持人：高利新（温州大学数理与电子信息工程学院常务副院长）</td>
</tr>
<tr>
<td>8:40-8:50</td>
<td>会议学术委员会主席致辞</td>
</tr>
<tr>
<td>8:50-9:10</td>
<td>合影</td>
</tr>
<tr>
<td>9:10-10:00</td>
<td>黎子良 Gradient Boosting for Modeling High-dimensional data: Mathematical theory and recent refinements 美国斯坦福大学</td>
</tr>
<tr>
<td>10:00-10:40</td>
<td>王友乾 Applied Statistics: Some Examples in Data Science 澳大利亚昆士兰科技大学</td>
</tr>
<tr>
<td>10:40-10:50</td>
<td>茶歇</td>
</tr>
<tr>
<td>10:50-11:30</td>
<td>张 浩 Big Data in Spatial Statistics 美国普渡大学</td>
</tr>
<tr>
<td>11:30-12:10</td>
<td>孙建国 Simultaneous Estimation and Variable Selection for Incomplete Event History Data 美国密苏里大学</td>
</tr>
<tr>
<td>12:10-13:40</td>
<td>午餐</td>
</tr>
<tr>
<td>13:50-14:30</td>
<td>殷向荣 A New Class of Measures for Testing Independence 美国肯塔基大学</td>
</tr>
<tr>
<td>14:30-15:10</td>
<td>林华珍 Efficient Estimation of the Nonparametric Mean and Covariance Functions for Longitudinal and Sparse Functional Data 西南财经大学</td>
</tr>
<tr>
<td>15:10-15:30</td>
<td>茶歇</td>
</tr>
<tr>
<td>15:30-16:10</td>
<td>吴尚武 On Security Properties of Privacy-preserving Data Collection Methods 美国佛罗里达大学</td>
</tr>
<tr>
<td>16:10-16:50</td>
<td>张立新 Adaptive Randomization: Efficiency, Selection Bias and Randomization Methods 浙江大学</td>
</tr>
<tr>
<td>17:00-18:00</td>
<td>会场乘校车至宾馆</td>
</tr>
<tr>
<td>18:00-20:00</td>
<td>晚餐</td>
</tr>
<tr>
<td>时间</td>
<td>内容</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>8:00-8:30</td>
<td>宾馆乘校车至会场</td>
</tr>
<tr>
<td>8:30-9:10</td>
<td>邵启满 Berry-Esseen Bounds of Normal and Non-normal Approximation for Unbounded Exchangeable Pairs 香港中文大学</td>
</tr>
<tr>
<td>9:10-9:50</td>
<td>苏中根 Asymptotic Normality for Linear Statistics in Hermitian Ensembles 浙江大学</td>
</tr>
<tr>
<td>9:50-10:00</td>
<td>茶歇</td>
</tr>
<tr>
<td>10:00-10:25</td>
<td>王纯杰 Bayesian Hierarchical Model of Clustered Current Status Data with Additive Hazards Model 长春工业大学</td>
</tr>
<tr>
<td>10:25-10:50</td>
<td>徐平峰 A Generalized ENS Algorithm for Model Selection with Incomplete Data 长春工业大学</td>
</tr>
<tr>
<td>10:50-11:15</td>
<td>何 勇 Adaptive Test for Mean Vector of High-dimensional Time Series Data with Factor Structure 山东财经大学</td>
</tr>
<tr>
<td>11:15-12:00</td>
<td>参观温州大学发绣馆</td>
</tr>
<tr>
<td>12:00-13:30</td>
<td>午餐</td>
</tr>
<tr>
<td>13:30-17:30</td>
<td>考察温州数学名人馆等</td>
</tr>
<tr>
<td>17:30-18:00</td>
<td>会场乘校车至宾馆</td>
</tr>
<tr>
<td>18:00-20:00</td>
<td>晚餐</td>
</tr>
</tbody>
</table>

**2017年12月25日 离会**

**会务组联系方式：**

蔡风景: 13695756075
黄玉飞: 13857725912
袁 野: 15258085606
薛漫芝: 13587686510
Gradient Boosting for Modeling High-dimensional Data: Mathematical Theory and Recent Refinements

Tze Leung Lai (黎子良)
Department of Statistics, Stanford University, Stanford, CA 94305
Email address: Lait@stanford.edu

Abstract: We begin with a review of the history of gradient boosting, dating back to the LMS algorithm of Widrow and Hoff in 1960 and culminating in Freund and Schapire’s AdaBoost and Friedman’s gradient boosting and stochastic gradient boosting algorithms in the period 1999-2002 that heralded the big data era. The role played by gradient boosting in big data analytics, particularly with respect to deep learning, is then discussed. We also present some recent work on the mathematical theory of gradient boosting, which has led to some refinements that greatly improves the convergence properties and prediction performance of the methodology.
Efficient Estimation of the Nonparametric Mean and Covariance Functions for Longitudinal and Sparse Functional Data
Huazhen Lin (林华珍)
Center of Statistical Research and School of Statistics,
Southwestern University of Finance and Economics, Chengdu, Sichuan, China
Email address: Linhz@swufe.edu.cn

Abstract: We consider the estimation of mean and covariance functions for longitudinal and sparse functional data by using the full quasi-likelihood coupling a modification of the local kernel smoothing method. The proposed estimators are shown to be consistent, asymptotically normal, and semiparametrically efficient in terms of their linear functionals. Their superiority to the competitors is further illustrated numerically through simulation studies. The method is applied to analyze AIDS study and atmospheric study.

Berry-Esseen Bounds of Normal and Non-normal Approximation for Unbounded Exchangeable Pairs
Qi-man Shao (邵启满)
Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong
Email address: qmshao@sta.cuhk.edu.hk

Abstract: An exchangeable pair approach is commonly taken in the normal and non-normal approximation using Stein's method. It has been successfully used to identify the limiting distribution and provide an error of approximation. However, when the difference of the exchangeable pair is not bounded by a small deterministic constant, the error bound is often not optimal. In this talk, using the exchangeable pair approach of Stein's method, a new Berry-Esseen bound for an arbitrary random variable is established without a bound on the difference of the exchangeable pair. An optimal convergence rate for normal and non-normal approximation is achieved when the result is applied to various examples including the quadratic forms, general Curie-Weiss model, mean field Heisenberg model, colored graph model. The talk is based on a join work with Zhuosong Zhang.
Asymptotic Normality for Linear Statistics in $\beta$-Hermitian Ensembles

Zhonggen Su (苏中根)
College of Data Science, Zhejiang University, Hangzhou, Zhejiang, China
Email address: suzhonggen@zju.edu.cn

Abstract: The $\beta$-Hermitian ensemble in $\mathbb{R}^n$ is a family of probability measures with joint probability density functions

$$p(x_1, x_2, \cdots, x_n) \propto \prod_{1 \leq i < j \leq n} |x_j - x_i|^{\beta} \prod_{i=1}^{n} e^{-x_i^2/2}, \quad x_i \in \mathbb{R}.$$  

The $\beta$-Hermitian ensemble is a well-studied object in both statistical physics and mathematical statistics. Here I will briefly report some asymptotic normality results for linear statistics with both smooth and discontinuous test functions.

Simultaneous Estimation and Variable Selection for Incomplete Event History Data

Jianguo Sun (孙建国)
Department of Statistics, University of Missouri, Columbia, MO 65211
Email address: sunj@missouri.edu

Abstract: This talk discusses regression analysis of incomplete event history data with the focus on simultaneous estimation and variable selection. Such data commonly occur in many areas such as medical studies and social sciences, and a great deal of literature has been established for their analysis except for the variable selection problem. To address this, we will present a new method, which will be referred to as a broken adaptive ridge regression approach, and establish its asymptotic properties including the oracle property and clustering effect. Numerical studies suggest that the proposed method performs well in practical situations and better than the existing methods. An application will be presented.
Applied Statistics: Some Examples in Data Science
You-Gan Wang (王友乾)
School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland 4001, Australia
Email address: you-gan.wang@qut.edu.au

Abstract: In this presentation, I will introduce estimating functions for dependent data and present some real data analysis. We will see new challenges in statistics and mathematics, which play an important role in understanding the “big” data problems.

On Security Properties of Privacy-preserving Data Collection Methods
Samuel S. Wu (吴尚武)
Professor and Associate Chair, Department of Biostatistics, University of Florida, 2004 Mowry Road, Gainesville, FL 32611
Email address: samwu@biostat.ufl.edu

Abstract: Recently our team developed new privacy-preserving data collection methods based on random matrix masking. The new technologies ensure that neither investigators nor analysts see the actual data, but standard statistical analysis can still be performed with the same results for masked data as for the original data. Specifically, multiple matrix masking procedures have been designed so that the raw data stay with research participants and only masked data with the same statistical utility are collected.
A New Class of Measures for Testing Independence
Xiangrong Yin (殷向荣)
Professor of Statistics, Department of Statistics, University of Kentucky,
725 Rose St., Lexington, KY 40536
Email address: yinxiangrong@uky.edu

Abstract: We introduce a new class of measures for testing independence between two random vectors, which uses expected difference of conditional and marginal characteristic functions. By choosing a particular weight function in the class, we propose a new index for measuring independence and study its property. To illustrate the use of such an index, two empirical versions are developed: slicing and kernel approaches. Their properties, asymptotics, connection with existing measures and applications are discussed. Implementation and Monte Carlo results are also presented.

Big Data Issues in Spatial Statistics
Hao Zhang (张浩)
Department Head, Department of Statistics, Purdue University,
250 N. University St., West Lafayette, IN 47907
Email address: zhanghao@purdue.edu

Abstract: One of areas where big data are collected is in environmental and climate studies. The Global Circulation Models or Regional Circulation Models can generate huge amount of data in space and time. Data collected through remote sensing or sensor networks are also huge. All these data are correlated spatially and temporally. One therefore has to deal with the huge covariance matrix in the traditional likelihood-based inferences or Bayesian inferences. When the dimension is very large, inversion of the matrix becomes infeasible numerically and also unstable due to the ill condition of the matrix. The issue becomes more complex when multivariate data are observed which can have different spatial scales. In this talk, I will discuss some recent developments in the theory and methods for dealing with the big spatial data, and in particular highlight some numerical algorithms that are applicable to large multivariate spatial data that are possibly of different scales.
Adaptive Randomization:
Efficiency, Selection Bias and Randomization Methods
Li-xin Zhang (张立新)
Associate Dean and Professor of Statistics, College of Data Science,
Zhejiang University, Hangzhou, Zhejiang, China
Email address: stazlx@zju.edu.cn

Abstract: Statistical inference of high-dimensional time series data is of increasing interest in various fields such as social sciences and biology. In this article, we consider the problem of testing the equality of high-dimensional mean vectors in the approximate factor model, which allows for time series dependence among distinct observations and more flexible dependence within observations. We propose a data-adaptive test based on the factor-adjusted data rather than on the directly observed data. By combining the tests with different norms, the proposed test adapts to various alternatives scenarios and thus overcomes the shortcomings of the tests based either on $\ell_1$-norm or $\ell_2$-norm. Multiplier bootstrap method is utilized to approximate the true underlying distribution of the proposed test statistics. Theoretical analysis shows that the proposed test enjoys desirable properties. Besides, we conduct thorough numerical study to compare the empirical performance of the proposed test with some state-of-the-art tests. A real stock market data set is analyzed to show the empirical usefulness of the proposed test.
Adaptive Test for Mean Vector of High-dimensional Time Series Data with Factor Structure

Yong He (何勇)
College of Statistics, Shandong University of Finance and Economics, Jinan, China
Email address: heyong@sdufe.edu.cn

Abstract: Statistical inference of high-dimensional time series data is of increasing interest in various fields such as social sciences and biology. In this article, we consider the problem of testing the equality of high-dimensional mean vectors in the approximate factor model, which allows for time series dependence among distinct observations and more flexible dependence within observations. We propose a data-adaptive test based on the factor-adjusted data rather than on the directly observed data. By combining the tests with different norms, the proposed test adapts to various alternatives scenarios and thus overcomes the shortcomings of the tests based either on $\ell_1$-norm or $\ell_2$-norm. Multiplier bootstrap method is utilized to approximate the true underlying distribution of the proposed test statistics. Theoretical analysis shows that the proposed test enjoys desirable properties. Besides, we conduct thorough numerical study to compare the empirical performance of the proposed test with some state-of-the-art tests. A real stock market data set is analyzed to show the empirical usefulness of the proposed test.
Bayesian Hierarchical Model of Clustered Current Status Data with an Additive Hazards Model

Chunjie Wang (王纯杰)
Changchun University of Technology
Email address: ciwang2014@126.com

Abstract: The additive hazard regression (AHR) model is known for its convenience in interpretation, as hazard is modeled as a linear function of covariates. In this paper, we investigate bayesian hierarchical model with additive hazard model under cluster current status data. An efficient computational scheme based on the Matropolis-Hastings algorithm is developed and has been implemented in R software. The proposed approach adopts the piecewise exponential model (PEM) and weibull distribution to model the log-baseline hazard function and allows to estimate the regression parameters and the baseline hazard function simultaneously. A simulation study is presented to show that the proposed method performs well with a finite sample and is easy to use in practice. The proposed methodology is further demonstrated by applying it to a lymphatic filariasis study.
A Generalized EMS Algorithm for Model Selection with Incomplete Data

Ping-Feng Xu (徐平峰)
Department of Statistics, Changchun University of Technology,
Changchun 130012, China
Email address: xupf_stat@126.com

Abstract: The EMS algorithm is a useful method for model selection with missing data. It performs E-step (Expectation step) and MS-step (Model Selection Step) alternately to find the minimum point of the observed generalized information criteria (GIC). However, sometimes it may not be numerically feasible to perform the MS-step, especially for high dimensional settings. In this paper, we seek only a decrease in the observed generalized information criteria in the MS-step. The resulting method is called a generalized EMS (GEMS) algorithm, which includes the EMS algorithm as a special case. We obtain several numerical convergence results of the GEMS algorithm. A useful special case is that all limit points of the EMS algorithm satisfy a necessary condition of the minimum points of the observed GIC under very weak conditions. We apply the GEMS algorithm for Gaussian graphical model selection and variable selection with missing data and compare with state of the art methods via numerical experiments.
温州大学南校区地图
路线一：导航温州大学岩松堂
路线二：步行至东屿村公交车站
乘坐131路公交至
高教园区站台下
路线三：步行至瓯江大厦公交站
乘坐52路或B111公交至
温州大学东大门公交站下
主办单位
温州大学  Florida Atlantic University
承办单位
温州大学数学与电子工程学院
统计与数据科学国际会议合影

2017.12.22-25 温州