哈工大数学研究院成立 5 周年

暨数学学科博士点设立 35 周年系列学术活动

统计科学专题研讨会

时间：2021 年 7 月 24-26 日

地点：哈尔滨工业大学一校区明德楼 B 区 201-1 学术报告厅

　　　腾讯会议　7 月 25 日会议号：644 950 482

与会专家:(按姓名首字母排序)

常晋源	西南财经大学
陈　敏	中国科学院数学与系统科学研究院
邓　柯	清华大学
郭建华	东北师范大学
郭水霞	湖南师范大学
刘小惠	江西财经大学
王启华	中国科学院数学与系统科学研究院
王兆军	南开大学
吴未迟	清华大学
姚　方	北京大学
张新雨	中国科学院数学与系统科学研究院
张月霞	多伦多大学
朱复康	吉林大学
朱利平	中国人民大学
邹长亮	南开大学

联系人：

崔　　嫣: ycui0311@163.com

端木昊随: duanmuhaosui@berkeley.edu

田 波 平: bopingt361147@hit.edu.cn

尹　　智: hustyinzhi@163.com

联系电话：0451-86413107

会议日程

7 月 25 日上午

时　间	开幕式
09:00-09:30	数学研究院院长许全华教授致辞, 东北师范大学副校长郭建华教授、中科院数学与系统科学研究院陈敏研究员致辞
09:30-09:45	合　照

时　间	报告人	题　目
09:45-10:30	王启华	大数据呼唤统计：统计的新进展
10:30-11:00		休　息
11:00-11:45	常晋源	Central limit theorems for high dimensional dependent data
12:00		午　餐

7 月 25 日下午

时　间	报告人	题　目
14:00-14:45	朱复康	Softplus INGARCH Models
14:45-15:30	端木昊随	On extended admissible decision procedures and their nonstandard Bayes risk
15:30-16:15	吴未迟	Nonparametric tests of long-range dependence for locally stationary process
16:15-16:45		休　息
16:45-17:30	张月霞	Inverse Probability Weighting-based Mediation Analysis for Microbiome Data
17:30-18:15	刘小惠 (线上)	Estimation of mean squared prediction error of observed best predictor under potential model misspecification
18:30		晚　餐

报告摘要

大数据呼唤统计：统计的新进展

王启华
中科院数学与系统科学研究院

大数据时代统计发展面临的主要挑战是数据的储存、计算问题、复杂的数据结构与数据异质性等问题。针对这些挑战，统计发展了一些新的推断方法，比如分布式推断、子抽样、在线推断及异质分析等方法。本讲座通过对这些方法的介绍，讲述了统计在大数据时代所取得的新进展。应该说这些进展还是初步的，只是统计在大数据时代朝大数据方向所迈出的第一步。大数据时代统计发展任重而道远!

Central limit theorems for high dimensional dependent data

常晋源
西南财经大学

Motivated by statistical inference problems in high-dimensional time series analysis, we derive non-asymptotic error bounds for Gaussian approximations of sums of high-dimensional dependent random vectors on hyper-rectangles, simple convex sets and sparsely convex sets. We investigate the quantitative effect of temporal dependence on the rates of convergence to normality over three different dependency frameworks (α-mixing, m-dependent, and physical dependence measure). In particular, we establish new error bounds under the α-mixing framework and derive faster rate over existing results under the physical dependence measure. To implement the proposed results in practical statistical inference problems, we also derive a data-driven parametric bootstrap procedure based on a kernel-type estimator for the long-run covariance matrices.

Softplus INGARCH Models

朱复康

吉林大学

During the last decades, a large variety of models have been proposed for count time series, where the integer-valued autoregressive moving average (ARMA) and integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) models are the most popular ones. However, while both models lead to an ARMA-like autocorrelation function (ACF), the attainable range of ACF values is much more restricted and negative ACF values are usually not possible. The existing log-linear INGARCH model allows for negative ACF values, but the linear conditional mean and the ARMA-like autocorrelation structure are lost. To resolve this dilemma, a novel family of INGARCH models is proposed, which uses the softplus function as a response function. The softplus function behaves approximately linear, but avoids the drawback of not being differentiable in zero. Stochastic properties of the novel model are derived. The proposed model indeed exhibits an approximately linear structure, which is confirmed by extensive simulations, and which makes its model parameters easier to interpret than those of a log-linear INGARCH model. The asymptotics of the maximum likelihood estimators for the parameters are established, and their finite-sample performance is analyzed via simulations. The usefulness of the proposed model is demonstrated by applying it to three real-data examples.

On extended admissible decision procedures and their nonstandard

Bayes risk

端木昊随

加州大学伯克利分校

Nonstandard analysis, a powerful machinery derived from mathematical logic, has had many applications in probability theory as well as stochastic processes. Nonstandard analysis allows construction of a single object——a hyperfinite probability space——which satisfies all the first order logical properties of a finite probability space, but which can be simultaneously viewed as a measure-theoretical probability space via the Loeb construction. As a consequence, the hyperfinite/measure duality has proven to be particularly in porting discrete results into their continuous settings.

The connection between frequentist and Bayesian optimality in statistical decision

theory is a longstanding open problem. For statistical decision problems with a finite parameter space, it is well known that a decision procedure is extended admissible (frequentist optimal) if and only if it is Bayes. Such connection becomes fragile for decision problems with an infinite parameter space and one must relax the notion of Bayes optimality to regain such equivalence between extended admissibility and Bayes optimality. Various attempts have been made in the literature but they are subject to technical conditions which often rule our semi-parametric and nonparametric problems. By using nonstandard analysis, we develop a novel notion of nonstandard Bayes optimality (Bayes with infinitesimal excess risk). We show that, without any technical condition, a decision procedure is extended admissible if and only if it is nonstandard Bayes. We conclude by showing that several existing standard results in the literature can be derived from our main result.

Nonparametric tests of long-range dependence for locally stationary process

吴未迟
清华大学

We study several KPSS-type tests for long memory in varying coefficient regression models. The responses and covariates considered are allowed to be locally stationary time series. We obtain the limiting distribution of the test statistics under null hypothesis, local alternatives as well as fix alternatives. We also provide a theoretical justified bootstrap approach for the implementation. The effectiveness of the tests are demonstrated by a simulation study and a real data analysis.

Inverse Probability Weighting-based Mediation Analysis for Microbiome Data

张月霞
多伦多大学

Mediation analysis is an important tool to study casual associations in biomedical and other scientific areas and has recently gained attention in microbiome studies. With a microbiome study of acute myeloid leukemia (AML) patients, we investigate whether the effect of induction chemotherapy intensity levels on the

infection status is mediated by the microbial taxa abundance. The unique characteristics of the microbial mediators---high-dimensionality, zero-inflation, and dependence---call for new methodological developments in mediation analysis. The presence of an exposure-induced mediator-outcome confounder, antibiotics usage, further requires a delicate treatment in the analysis. To address these unique challenges brought by our motivating microbiome study, we propose a novel nonparametric identification formula for the interventional indirect effect (IIE), a measure recently developed for studying mediation effects.

We develop the corresponding estimation algorithm and test the presence of mediation effects via constructing the nonparametric bias-corrected and accelerated bootstrap confidence intervals. Simulation studies show that the proposed method has good finite-sample performance in terms of the IIE estimation, and type-I error rate and power of the corresponding test. In the AML microbiome study, our findings suggest that the effect of induction chemotherapy intensity levels on infection is mainly mediated by patients' gut microbiome.

Estimation of mean squared prediction error of observed best predictor under potential model misspecification

刘小惠

江西财经大学

This paper is regarding robust small area estimation (SAE) in terms of measure of uncertainty. We consider estimation of the mean squared prediction error (MSPE) of the observed best predictor (OBP) in SAE under the Fay-Herriot model with potential model misspecification. Previously, it was thought that the traditional Prasad-Rao linearization method could not be used, because it is derived under the assumption that the underlying model is correctly specified. However, we show that, when it comes to estimating the unconditional MSPE, the Prasad-Rao (PR) estimator, derived for estimating the MSPE of OBP assuming that the underlying model is correct, remains first-order unbiased even when the underlying model is misspecified in its mean function. A second-order unbiased estimator of the MSPE is derived by modifying the PR MSPE estimator. The PR and modified PR estimators also have much smaller variation compared to the existing MSPE estimators for OBP. The theoretical findings are supported by empirical results including simulation studies and real-data applications.

数学学院简介

哈尔滨工业大学数学学院前身是创建于 1958 年的计算数学专业，1981 年开始培养基础数学和计算数学专业硕士，1986 年获得基础数学博士学位授予权（是国内最早的两所工科院校之一），1987 年成立数学系，2019 年成立数学学院。2001 年建立了数学学科博士后流动站，2005 年数学学科成为一级学科硕士学位授权点，2010 年数学学科成为一级学科博士授权点，2011 年统计学成为一级学科博士授权点。基础数学是省重点学科(2001 年)和国防科工委重点学科(2002 年)；应用数学是省重点学科（2001 年）。数学学科 2011 年成为省一级重点学科。2013 年基础数学和应用数学成为工信部重点学科。1997 年入选教育部首批七个"工科基础课程（数学）教学基地"之一；2020 年数学类专业入选教育部强基计划和基础学科拔尖学生培养计划 2.0 基地；2020 年获批成立黑龙江应用数学中心。

在教育部第四轮学科评估中，哈尔滨工业大学数学学科位列 A-，统计学位列 B。在 2020 年 10 月《美国新闻和世界报导》（US News）发布的世界大学数学专业排名中，我校数学学科排名全球第 80 位，在内地高校 45 个机构中位于第 14 位。在 2021 年发布的世界大学学科排名（QS World University Rankings）中，我校数学学科排名全球第 126 位，在内地高校 36 个机构中位于第 8 位；统计学排名全球第 101-150 位，在内地高校 17 个机构中并列第 7 位。在最新的 ARWU 排名中，数学学科位列全球第 76-100 位，在内地高校 93 个机构中并

列第 5 位。哈尔滨工业大学数学学科自 2013 年 5 月始终保持全球 ESI 前 1%行列。

学院现有专任教师 82 人，博士化率 91.5%；其中，国家杰出青年 1 人，中组部首届青年拔尖人才计划 1 人，教育部新世纪人才 1 人，龙江学者 1 人，中组部"万人青拔" 1 人，青年长江学者 1 人，黑龙江省杰出青年基金获得者 1 人，黑龙江省教学名师 4 人，龙江青年学者 1 人，宝钢优秀教师奖 7 人，黑龙江省优秀青年基金获得者 1 人；博士生导师 43 人，硕士生导师 63 人，教授 33 人。

学院现有本科专业三个：数学与应用数学（拔尖学生培养计划 2.0、强基计划（2020）、国家一流本科专业（2020））、信息与计算科学（拔尖学生培养计划 2.0、强基计划（2020）、国家一流本科专业（2019））、统计学（省一流本科专业（2020））。现有在读本科生 307 人，硕士研究生 144 人，博士研究生 195 人。现有：国家级精品资源共享课程 1 门，国家级精品课程 2 门，国家级精品在线开放课程 3 门，省级精品课程 4 门，省级优秀教学团队 1 个，省级优秀教材 2 部，省头雁团队 1 个（数学与人工智能交叉学科创新研究），省级重点实验室（计算与应用数学）1 个，省级领军人才梯队（计算数学）1 个。已培养本科生近 2000 人，硕士生近 1400 人，博士生近 400 人，其中涌现出一大批优秀学子：与境外高水平大学联合培养博士研究生 100 余人；长江学者、国家杰青等高层次人才 10 余人；8 位大学校长、副校长（如：哈工大副校长、电子科技大学副校长等）；国家百篇优博提名奖 3 人；教育部学术新人奖 3 人；20 余位省级学会

和国家二级学会理事长及副理事长；校优秀博士学位论文奖 16 人；世界华人数学家大会"新世界数学奖"博士金奖 1 人、本科生银奖 1 人。

数学学科依据国防和社会发展的需求及主流科研方向前沿发展趋势，形成了以传统优势方向为支撑，以新兴与交叉方向为主要生长点的学科格局。主要科研方向有：泛函分析及其应用、代数与数论、常微分方程与动力系统、科学与工程计算、偏微分方程与调和分析、数学物理反问题、运筹控制与优化、概率论与数理统计等。近年来承担国家重点研发计划等国家级课题 50 余项，科研经费千万余元。获黑龙江省科学技术奖一等奖、教育部高校科研优秀成果奖自然科学奖二等奖等多个科研奖高水平学术论文 100 余篇。在全国 SCI 高产机构的排名中，近几年一直在前 20 名，2012 年发表的 SCI 论文数量位居全国数学学科第 3 位。

网址：math.hit.edu.cn

数学研究院简介

哈尔滨工业大学数学研究院创建于 2016 年 7 月，首任院长由我校讲席教授许全华担任，研究院直接隶属于学校，是数学学院密不可分的合作伙伴。研究院以基础数学为基石，以从事国际一流的原始创新研究和培养杰出青年数学人才为第一要务，致力于推动数学科学的发展以及数学与物理、工程等领域的交叉研究。

研究院现有科研人员 18 人，其中高层次人才 7 人，分别为：菲尔兹奖得主吴宝珠；国家海外引才计划：许全华；国家海外引才计划、长江学者：吴黎明；国家海外引才计划（青年）：尹智、李科、熊枭、熊欢。研究院探索实行法国宽松管理模式，不片面追求论文数量或杂志级别，而是着力为科研人员提供利于事业发展的有效平台，积极打造一个愉快、舒适、和谐、向上的工作环境，让每名科研人员都能找到适合自己发展的方式和位置。

数学研究院重点打造现代分析、数论-代数-组合以及概率统计及其应用等优势基础学科方向。五年来，获批各类国家自然科学基金 15 项，博士后基金 7 项，2020 年获批国家自然科学基金重点项目 1 项，填补了我校数学学科在此项目中的空白；学院教师先后在《PNAS》，《Memoirs of the American Mathematical Society》，《Communications in Mathematical Physics》等国内外著名期刊发

表高水平论文 50 余篇。研究院组织举办了一系列具有国内外重要影

响力的学术会议，先后邀请中国科学院院士田刚、美国加州大学圣塔芭芭拉分校张益唐，以及哈佛大学、美国芝加哥大学、俄罗斯科学院等知名专家学者 300 余人到我校访问交流，并促使我校和法国弗朗什-孔泰大学签订双边合作协议。

我们相信，在学校的大力支持下，数学研究院将进一步加快发展步伐，不断开拓创新，促进学科间的交叉与融合，发展成在国内外具有重要影响的数学研究中心，助力学校"双一流"建设。