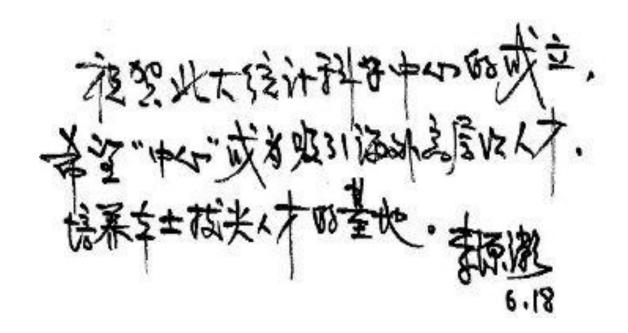
北京大学统计科学中心 2012 年年度报告

Tel: 0086-010-62760736 Fax: 0086-010-62760736

Mail: stat-center1@pku.edu.cn Http: www.stat-center.pku.edu.cn

Add: 北京大学燕东园 32 号小楼,100871



北京大学统计科学中心为跨学院的交叉学科研究机构,其目标是协调全校统计研究的力量,促进统计学与其他学科的交叉与融合,建设世界一流的统计研究机构。

目的与使命

整合全校统计研究力量,培养综合型的统计人才,推动统计学研究与教育走向世界,成为世界一流的统计研究机构。扩展统计学在其它学科的应用。

目录

3	综述:	一、
5	人员:	二、
6	中心要闻:	三、
12	论文发表:	四、
16	研究报告:	五、
5介绍:17	中心科研进	六、
23	学术讲座:	七、
29	开设课程:	八、
30	交流访问:	九、
36	统计咨询:	十、

一、 综述:

2012年北京大学统计科学中心继续在统计学人才引进、推动北大统计学的整体研究实力的增强、同其它院系的交叉合作等方面展开工作,工作取得了不少的成果。本年度中心利用学校所给予的特殊机制聘任了席瑞斌博士和涂云东博士,他们是中心分别同数学学院概率统计系和光华管理学院商务统计与经济计量系联合聘任。为了充分利用在海外的华人统计学家资源,促进、带动北京大学统计学科研水平的提高,中心在2012年先后聘任了美国卡耐基梅隆大学的金加顺教授和罗格斯大学的陈嵘教授为中心的特聘教授。他们二位均是国际上顶尖的统计学家,担任多个顶尖统计期刊的副主编、主编,他们已经开始开设暑假高级统计学短期课程,指导年轻教员。二位特聘教授正在积极筹划2013北大统计学短期课程,这一课程将面向全国的统计学研究生开放,将提高北大在全国统计学科的主导地位。

2012年中心继续积极开展同国际、国内统计学界的学术交流、合作。三月统计学大师Peter Hall 来北大教授了 16 小时的课程,来自国内各高校的一百多位师生参加了课程。2012年中心针对全校与统计学相关的研究生开设了 4 门短期课程,授课老师六人分别来自美国、澳大利亚及香港的著名高校。中心在 2012年资助了 4 个学术短期访问计划,支持北大同统计学相关的三个系的老师与国际同行的研究、合作,访问时间从一周到一个半月不等,共资助了四位国外的访问学者来北大进行研究。中心为了增进统计学在其它学科的应用水平,推动北大其它学科的科学水平,继续提供面向校内老师、同学的无偿的统计学咨询活动。

2012年中心举办学术报告共 46 场,其中来自海外的学者 35 场,国内的学者 11 场。中心联合光华管理学院商务统计和经济计量系及数学科学学院金融数学系举办了"金融建

模中的量化方法"学术论坛,来自北大、清华、中科院、人大、中财、南开等相关院系的师生及证券、银行、风投等行业的专业人士等上百人参加了此次论坛。

教育部在 2012 年对统计学作为一级学科第一次进行了评估。由中心牵头,联合校内各统计学相关院系组织队伍,集中成果积极参加了学科评估。我们在期待评估结果的同时,十分惴惴不安。因为这几年其他学校在统计学成为一级学科后,纷纷成立了统计学院,大量引进教员并大量招收学生,体量远远超过北大。(评估结果已于 2013 年一月公布,在87 所参评学校中,人大以 90 分排名第一,北大获得 88 分,以 2 分之差排在第二)。在这一日益竞争的环境中如何保持并提高北大的统计学,将是统计中心及北京大学应该思考并为之有所布局的迫切问题。

二、人员:

主任: 陈松蹊、耿直 副主任: 陈大岳

执行委员会

陈大岳 数学科学学院概率统计系

陈松蹊 光华管理学院商务统计与经济计量系

耿直 数学科学学院概率统计系

王汉生 光华管理学院商务统计与经济计量系

吴岚 数学科学学院金融数学系

全职教员

贾金柱 数学科学学院概率统计系

涂云东 光华管理学院商务统计与经济计量系

席瑞斌 数学科学学院概率统计系

特聘教授

陈嵘 金加顺 格罗斯大学

卡耐基梅隆大学

学术讲座召集人

贾金柱 数学科学学院概率统计系

统计咨询召集人

邓明华 数学科学学院概率统计系

张俊妮 光华管理学院商务统计与经济计量系

政府统计工作室召集人

房祥忠 数学科学学院概率统计系

行政人员

孟蕾、张卫香

三、 中心要闻:

1、商务统计与经济计量系博士生常晋源近期在国际统计学最顶级的学术期刊 The Annals of Statistics 上发表了题为 On the approximate maximum likelihood estimation for diffusion processes 的论文(第一作者)。

The Annals of Statistics 是由国际数理统计协会主编的刊物,旨在反映当代统计的许多方面的最高质量的研究论文发表。常晋源于 2009 年进入光华管理学院商务统计与经济计量系学习,为硕博连读研究生,该文章是在其导师陈松蹊教授的指导下完成的。

2、国际统计学大师 Peter Hall 教授访问北大统计中心并开办短期课程。

2012 年 3 月 16 日由北大统计科学中心主办,北京国际数学研究中心协办的 "Methodology and Theory for The Bootstrap"统计学短期前沿课程开课,主讲人是国际统计学大师 Peter Hall 教授,他以高超的学术造诣,饱满的热情教授了精心准备的共计 16 小时的课程。参与课程学习的除了本校师生,还有来自全国各地十多所高校的博士研究生及青年教师,总数近一百二十人,Peter Hall 教授在课程中系统介绍了 Bootstrap 方法的应用及理论性质,以及 Edgeworth 展开在 Bootstrap 理论研究中的应用,通过与其它统计方法的比较说明了 Bootstrap 方法的优点、存在的一些不足,及一些未解决的问题。课堂气氛活跃,课间互动积极,参加课程的师生们利用一切时间与 Peter Hall 教授就课程相关内容及统计学各领域的问题展开热烈探讨,Peter Hall 教授的大师风范和精彩的授课内容令大家收获良多。

周其凤校长于 3 月 19 日接见了 Peter Hall 教授,双方就当今国际统计学最新的发展及在各学科中的交叉应用展开讨论,并对中国统计学界在未来的发展方向和趋势交换了意见。相关资料:

Peter G. Hall 博士现任墨尔本大学和加州大学戴维斯分校的统计学教授,英国皇家学会院士,澳大利亚科学院院士,他是世界上最高产的数学家之一,也是统计中心的国际顾问。他在非参数统计,特别是估计曲线和重采样,鞅中心极限定理,平滑,密度估计,带宽选择等方面做出了重要的贡献,他的工作也围绕着众多的应用方向,横跨经济学、工程学、物理学和生物学等各个领域。

3、Carnegie Mellon University 的金加顺教授受聘为统计中心特聘教授。

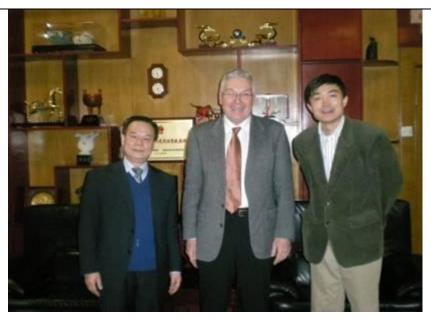
来自 Carnegie Mellon University 的金加顺教授于 2012 年 5 月 1 日起受聘为北大统计科学中心的特聘教授,任期两年。



Peter G. Hall 教授(前中)、田刚院士(前右四)、陈松蹊教授(前左五)、 耿直教授(前右三)、陈大岳教授(前左四)与师生们合影留念



授课现场



周其凤校长、Peter Hall 教授、陈松蹊联席主任合影留念

4、博士生常晋源同学获得国际数理统计协会 Laha Award.

光华管理学院商务统计与经济计量系 09 级硕博连读生常晋源同学,在继其论文发表在统计学顶级 刊物 The Annals of Statistics 上后,近日又获得国际数理统计协会(Institute of Mathematical Statistics) 颁发的 2012 年 Laha Award. Laha Award 是国际数理统计协会用于资助全球在读博士和博士毕业两年内的青年统计学者参加国际数理统计协会年会并宣读论文的奖项。该奖项每年颁发一次,每次获奖人数在 10 人左右。常晋源将于今年 7 月前往土耳其伊斯坦布尔参加 2012 年国际数理统计协会年会暨第八届全球概率统计会议,并报告文章 High dimensional empirical likelihood for generalized estimating equations with dependent data.

5、国际著名统计学家吴建福教授访问北大统计中心。

国际著名统计学家、北京大学统计科学中心咨询委员会委员吴建福教授于 5 月 7 日下午莅临位于北京大学燕东园 32 号小楼的统计科学中心,听取了中心联席主任陈松蹊教授的工作汇报,并会晤了刁锦寰先生和中心部分成员。8 日上午,由统计中心主办的统计学科发展座谈会在英杰交流中心第二会议室举行,来自北京大学、中科院、北京工业大学、北京师范大学、首都师范大学、北京理工大学、清华大学、美国弗吉尼亚大学等单位约 50 位老师同学参加了这次座谈会,座谈会由统计中心联席主任耿直教授主持,吴建福教授首先发表了题为"对中国统计发展的一些浅见"的

主题演讲,李国英、杨振海、何书元、陈敏、于丹和研究生代表等先后发言。座谈会录像可在北京大学讲座网上看到(http://lecture.pku.edu.cn/video1/).

吴建福(C.F. Jeff Wu)先生于 1971 年毕业于台湾大学数学系, 1976 年在加州大学伯克利分校获得统计学博士学位。先后执教于威斯康星大学、滑铁卢大学和密执安大学,现任乔治亚理工学院工程统计讲座教授。 吴建福教授的研究领域为实验设计和工程统计,在统计研究的筹划、分析和阐释方面做出了基础性贡献,对统计的实践具有深刻影响,特别是在工程领域。例如再抽样方法、试验设计理论、以及在工业统计方面的开拓性工作,丰富了用于优化产品和改进工艺流程的统计方法。迄今为止他发表论文 140 多篇,出版专著两部,指导博士生 37 名。他是美国工程院院士,中央研究院院士,1987 年获 COPSS 奖,2011 年应邀做 COPSS Fisher 演讲,并将在 2012 年美国统计联合年会(JSM)上做 Deming 演讲。早在上个世纪八十年代他就和国内同行交流合作,近年来多次访问中科院,积极提携年轻学者,我校的艾明要博士曾在他指导下工作一年半。

6、Rutgers University 的陈嵘教授受聘为统计中心特聘教授。

来自 Rutgers University 的陈嵘教授于 2012 年 6 月 1 日起受聘为北大统计科学中心的特聘教授,任期两年。

7、国际知名统计学家王永雄教授于2012年6月6日访问了北大统计中心。

国际知名统计学家、斯坦福大学统计系主任、北京大学统计科学中心国际顾问委员会主席王永雄教授(Professor Wing Hung Wong)于2012年6月6日访问了北大统计中心,并于当日下午2:00-3:00在理科一号楼1114教室作了题为"Multivariate density estimation"的学术报告。报告结束后,王永雄教授参加了由中心主任陈松蹊教授主持召开的执行委员会会议,并听取了执委会成员的工作汇报。王永雄教授于第二天拜会了周其凤校长,就统计学在北大的发展问题进行了讨论。

王永雄教授(Professor Wing Hung Wong)是美国斯坦福大学统计系教授,现任系主任,2009年当选美国国家科学院院士。他的实验室主要开发用于分析各种高通量基因组实验数据的方法和软件。具体研究内容包括数据分析,多变量分析,信息理论,蒙特卡罗采样,图论,线性及非线性方程,以及它们在计算分子生物学和系统生物学中的应用。近期,他的实验室利用生物实验和计算分析相结合的方式对生物发育过程中的顺式调控机制展开了研究。

8、郁彬教授当选 IMS 候任会长。

郁彬教授(美国人文与科学学院院士,加州大学伯克利分校讲座教授),近日当选为数理统计学会候任会长(President-Elect of the Institute of Mathematical Statistics),一年之后她将升任会长(President, 2013-2014),卸任会长之后还将担任为期一年的前任会长(Past President, 2014-2015)。

郁彬教授于 1980 年到 1984 年在北京大学数学科学学院学习。她现在是加州大学伯克利分校统计系和电子工程及计算机系教授。于 2009 年到 2012 年担任伯克利统计系主任。

自从 2005 年她每年会在北大访问 6 到 8 周,指导并与中心的年轻教员老师合作。她还联合指导北大数学科学学院的统计与信息技术系的微软联合实验室。她最近研究方向为统计机器学习理论及遥感、神经科学、文献检索等多样化跨学科领域。最近更与北京大学签订了"海外高层次人才引进计划"的三年聘用合同。

9、国际知名统计学家蔡天文教授于2012年6月21日访问了北大统计中心。

国际知名统计学家、宾州大学沃顿商学院统计学教授、Dorothy Silberberg 讲席教授蔡天文(Professor T. Tony Cai)于2012年6月21日访问了北大统计中心,并于当日下午2:00-3:00在理科一号楼1114教室作题为"High-dimensional statistical inference: From vectors to matrices"的学术报告。

蔡天文教授(Professor T. Tony Cai)于 1996 年毕业于康奈尔大学并获得博士学位。现任美国宾州大学沃顿商学院统计学教授、Dorothy Silberberg 讲席教授。蔡天文教授的研究方向为高维统计推断、大范围假设检验、非参数函数估计、函数数据分析、小波方法和应用、统计决策论。迄今为止他发表论文 80 多篇。历任泛华统计学会会刊副主编、美国统计学会会刊副主编、统计年刊副主编,现任统计年刊主编。2008 年获 COPSS 奖并于 2009 年作国际数理统计学会勋章讲座。

10、国际著名统计学家 Donald B. Rubin 教授于 2012 年 7 月 10 日访问了北大统计中心。

国际著名统计学家、哈佛大学统计系教授、John L. Loeb 讲席教授 Donald B. Rubin 于 2012 年 7 月 10 日访问了北大统计中心,并于当日下午 2:00-3:30 在北京大学光华管理学院新楼 110 教室作题为 "Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data" 的学术报告。

Donald B. Rubin 教授于 1970 年毕业于哈佛大学并获得博士学位。先后执教、任职于芝加哥大学、哈佛大学和美国国家经济研究局,现任哈佛大学统计系教授、John L. Loeb 讲席教授、Datametrics Research, Inc.主席及 National Opinion Research Center 副研究员。Donald B. Rubin 教授的研究方向为实验与观测研究中的因果推断、无应答的抽样调查及丢失数据情况下的推断、贝叶斯和经验贝叶斯技术的应用、开发和应用各学科的数据的统计模型。迄今为止他已发表论文 350 篇。他是美国科学院院士,曾获得过四次统计学界最著名的奖项,分别为美国统计学会的 Samuel S. Wilks 奖,Parzen Prize for Statistical Innovation,R. A. Fisher Lectureship 和 George W. Snedecor Award of COPSS.

11、统计学家 Minghui Chen 教授于 2012 年 7 月 11 日访问了北大统计中心。

著名统计学家、康涅狄格大学统计系教授 Minghui Chen 于 2012 年 7 月 11 日访问了北大统计中心,

并于当日上午 9:00-11:00 和下午 2:00-4:00 在北京大学光华管理学院新楼 217 房间教授 "Monte Carlo Methods in Bayesian Computation"的短期课程。

Minghui Chen 教授于 1993 年毕业于普渡大学并获得博士学位。曾任伍斯特理工学院 Harold J. Gay 讲席教授、SBSS 评委会主席等,现任康涅狄格大学统计系教授、International Chinese Statistical Association 主席、《Bayesian Analysis》主编等。Minghui Chen 教授的研究方向为贝叶斯统计方法,贝叶斯计算,分类数据分析,DNA 微阵列数据分析,缺少数据的分析(EM, MCEM, and Bayesian),蒙特卡罗方法,Prior Elicitation,前列腺癌数据的统计方法,统计建模,生存数据分析,变量的选择等。迄今为止他已发表论文 225 篇。

12、2012年7月17日举办"金融建模中的量化方法"学术论坛。

北京大学统计科学中心联合光华管理学院商务统计和经济计量系及数学科学学院金融数学系于2012年7月17日在光华新楼110会议室,举办了名为"金融建模中的量化方法"的学术论坛(A Workshop on Quantitative Methods in Finance), 主题为"金融量化方法及其在中国的实践"。特别邀请了美国普林斯顿大学的 Yacine Aït-Sahalia、中国人民银行的梁世栋先生、北京航空航天大学韩立岩教授、富国基金的李笑薇女士、南开大学王永进教授等作了关于我国金融界在量化方法研究和实践中的经验的系列报告。来自北大、清华、中科院、人大、中财、南开等相关院系的师生及证券、银行、风投等行业的专业人士等上百人参加了此次论坛。

13、耶鲁大学的赵宏宇被聘为我校讲座教授。

根据最近公布的2011年度国家级重大人才计划(教育部)学者名单,赵宏宇被聘为我校讲座教授。

赵宏宇 1986-1990 年就读于北京大学概率统计系,1995 年在美国加州大学伯克利分校取得统计学博士学位,现任耶鲁大学公共卫生学院生物统计系主任,Ira V. Hiscock 讲座教授。赵宏宇教授于2012 年 12 月 14 日访问了北大,并在数学学院作了题为"Biological network inference from genomics data" 的学术报告。

- 14、陈松蹊、王汉生教授就任 Journal of Business and Economic Statistics 副主编,任期 2012-2015年。
- 15、虞吉海担任SSCI国际学术期刊Spatial Economic Analysis的联合编辑,任期2012-2015年。
- 16、耿直担任第八届全国博士后管委会专家组成员。
- 17、艾明要自2012年1月起担任Journal of Statistical Planning and Inference 副主编。
- 18、艾明要自 2012 年起升任教授。



Yacine Aït-Sahalia 教授(前中),陈松蹊教授(前左四),吴岚教授(前右四), 韩立岩教授(前右三),王永进教授(前右二)等与参会者合影留念



主讲人 Yacine Aït-Sahalia 教授在会议现场

四、论文发表:

- [1] **Ai, M. Y.**, He, Y. Z. and Liu, S. M. (2012). Some new classes of orthogonal latin hypercube designs, *Journal of Statistical Planning and Inference*, 142(10), 2809-2818.
- [2] **Chen, D. Y.** and Ingram Olkin (2012). The grandparent of probability and statistics in china, *Statistical Science*, 27(3), 434-445.
- [3] Chu, W. J., **Ren, Y. X.** and Li, V. W. B. (2012). Small value probabilities for continuous state branching processes with immigration, *Science China Mathematics*, 55(11), 2259-2271.
- [4] Deng, K., Li, Y., Zhu, W. P., Geng, Z. and Liu, J. (2012). On delay tomography: fast algorithms and spatially dependent models, *IEEE Transactions on Signal Processing*, 60, 5685-5697.
- [5] He, P., Deng, K., Liu, D. L., Liu, J. and **Geng, Z.** (2012). Discovering herbal functional groups of traditional Chinese medicine, *Statist. Medicine*, 31, 636-642.
- [6] He, Y. Z. and Ai, M. Y. (2012). Complementary design theory for sliced equidistance designs, *Statistics and Probability Letters*, 82(1), 542-547.
- [7] Hou, L., Wang, L., Arthur Berg, Qian, M. P., Zhu, Y. P., Li, F. T. and **Deng, M. H.** (2012). Comparison and evaluation of network clustering algorithms applied to genetic interaction networks, *Front Biosci.* (*Elite Ed*), 4, 2150-2161.
- [8] Jiang, B., Song, K., Ren, J., **Deng, M. H.**, Sun, F. Z. and Zhang, X. G. (2012). Comparison of meta genomic samples using sequence signatures, *BMC Genomics*, 13, 730.
- [9] Kang Y., Deng R., Wang C., Deng T., **Deng, M. H.**, et al. (2012). Etiologic Diagnosis of Lower Respiratory Tract Bacterial Infections Using Sputum Samples and Quantitative Loop-Mediated Isothermal Amplification, *PLoS ONE*, 7(6), 38743.
- [10] A. E. Kyprianou, Liu, R. L. and **Ren, Y. X.** (2012). A. Murillo-Salas: Supercritical super-Brownian motion with a general branching mechanism and travelling waves, *Annales de l'Institut Henri Poincaré*, 48(3), 661-687.
- [11] A.E. Kyprianou and **Ren, Y. X.** (2012). Backbone decomposition for continuous-state branching processes with immigration, *Statistics & Probability Letters*, 82, 139-144.
- [12] Lan, W., **Wang, H. S.** and C.-L. Tsai (2012). A Bayesian information criterion for portfolio selection, *Computational Statistics & Data Analysis*, 56, 88-99.
- [13] Lee, L. F. and **Yu, J. H.** (2012). QML estimation of spatial dynamic panel data models with time varying spatial weights matrices, *Spatial Economic Analysis*, 7, 31-74.

- [14] Lee, L. F. and **Yu, J. H.** (2012). Spatial Panels: random components vs. fixed effects, *International Economic Review*, 53, 1369-1412.
- [15] Lee, L. F. and **Yu, J. H.** (2012). The c(a)-type gradient test for spatial autoregressive models, *Letters in Spatial and Resource Science*, 5, 119-135.
- [16] Li, J. and Chen, S. X. (2012). Two sample tests for high dimensional covariance matrices, *The Annals of Statistics*, 40, 908-940.
- [17] Li, X. J., Yang, D. N., **Xu, M. Y.,** Lei, M. (2012). China's labor-intensive manufacturing productivity dilemma-the perspective of corporate social responsibility, *Nankai Business Review*, 3, 122-130.
- [18] Liang, H., Wang, H. S., C.-L.Tsai (2012). Profiled forward regression for ultra high dimensional variable screening in semiparametric partially linear models, *Statistics Sinica*, 22, 531-554.
- [19] Liu, S. Q., Shen, G., Li, W. R., Geng, Z., Xu, T. M. (2012). China Orthodontic Specialists wrong deformity severity Consistency of subjective judgment, *Journal of Peking University* (*Health Sciences*), 44, 98-102.
- [20] **Liu, Y.** and Yang, F. X. (2012). Some ergodic theorems for a parabolic anderson model, *Acta Mathematica Sinica*, 28(12), 2443-2462.
- [21] **Liu, Y.** and Zhai, J. L. (2012). A note on time regularity of generalized Ornstein-Uhlenbeck processes with cylindrical stable noise, *C. R. Acad. Sci. Paris*, Ser. I, 350, 97-100.
- [22] Peng, L., Qi, Y. C., Wang, R. D. and **Yang, J. P.** (2012). Jackknife empirical likelihood method for some risk measures and related quantities, *Insurance Mathematics & Economics*, 51(1), 142-150.
- [23] Qiu, Y. M. and **Chen, S. X.** (2012). Test for bandedness of high dimensional covariance with bandwidth estimation, *Annals of Statistics*, 40, 1285-1314.
- [24] Shan, Z. C. and **Chen, D. Y.** (2012). Voter model in a random environment in Z^d, *Front. Math. China*, 7(5), 895-905.
- [25] Song, G. Y., Li, W. R., **Geng, Z.,** Xu, T. M. (2012). Subjective evaluation of orthodontic treatment satisfaction Discussion consistency, *Journal of Peking University (Health Sciences)*, 44, 103-107.
- [26] Song, K., Ren, J., Zhai, Z. Y., Liu, X. M., **Deng, M. H.** and Sun, F. Z. (2012). Alignment-Free sequence comparison based on next generation sequencing reads, *Proceedings of Research in Computational Molecular Biology 16th Annual International Conference*, April 21-24, 272-286.

- [27] Tan, X., Hu, L., L. L. J., Gao, G., Liu, Y., Qu, H., **Xi, R. B.**, Lu, Z. J., Park, P., Elledge, S. J.
- (2012). Systematic identification of synergistic drugpairs targeting HIV, *Nature Biotechnology*, 30, 1125-1130.
- [28] Tao, J. and **Yu, J. H.** (2012). The spatial time lag term in panel data models, *Economics Letters*, 117, 544-547.
- [29] The Cancer Genome Atlas Research Network (Including **Xi, R. B.**) (2012). Comprehensive genomic characterization of squamous cell lung cancers, *Nature*, 489, 519-525.
- [30] The Cancer Genome Atlas Research Network (Including **Xi, R. B.**) (2012). Comprehensive molecular characterization of human colon and rectaltumors, *Nature*, 487, 330-337.
- [31] Wang, H. S. (2012). Factor profiled sure independence screening, *Biometrika*, 99, 15-28.
- [32] Wang, L., Hou, L., Qian, M. P. and **Deng, M. H.** (2012). Integrating phosphorylation network with transcriptional network reveals novel functional relationships, *PLoS ONE*, 7(3), 33160.
- [33] Wang, Q., Peng, P. C., Qian, M. P., Wan, L. and **Deng, M. H.** (2012). Hybridization and amplification rate correction for Affymetrix SNP arrays, *BMC Medical Genomics*, 5, 24.
- [34] Wu, L., Qi, Y. C. and **Yang, J. P.** (2012). Asymptotics for dependent bernoulli random variables, *Statistics & Probability Letters*, 82(3), 455-463.
- [35] Wang, X. L., Chen, H., **Geng, Z.** and Zhou, X. H. (2012). Using auxiliary data for binomial parameter estimation with nonignorable nonresponse, *Communications in Statistics Theory and Methods*, 41, 3468-3478.
- [36] **Xi, R. B.**, Lee, S. and Park, P. (2012). A survey of copy number variation detection tools based on high-throughput sequencing data, *Current Protocols in Human Genetics*, 7, 7.19.
- [37] Yan, W., Hu, Y. Q. and **Geng, Z.** (2012). Identifiability of causal effects for binary variables with baseline data missing due to death, *Biometrics*, 68, 121-128.
- [38] **Yu, J. H.** and L.F. Lee (2012). Convergence: A spatial dynamic panel data approach, *Global Journal of Economics*, 1, 125006.
- [39] **Yu, J. H.**, R. deJong and L.F. Lee (2012). Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration, *Journal of Econometrics*, 167, 16-37.
- [40] **Zhang, F. X.** and Xia, A. H. (2012). On the asymptotics of locally dependent point processes, *Stochastic Processes and Their Applications*, 122, 3033-3065.

五、 研究报告:

- [1] Liu, H. Z. and **Jia**, **J. Z.**, On L2 error bounds of the elastic net when p>>n. 2012-01.
- [2] Cui, W., Yang, J. P. and Wu, L., Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles matrices. 2012-02.
- [3] Su, N. F., Qian, M. P. and **Deng, M. H.**, Integrative approaches for microRNA target prediction: combining sequence information and the paired mRNA and miRNA expression profiles. 2012-03.
- [4] Xu, Q. Q, Huang, Q. M., Jiang, T. T., Yan, B. W., Lin, W. S. and **Yao, Y.**, HodgeRank on random graphs for subjective video quality assessment. 2012-04.
- [5] Weinan E, Lu, J. F. and Yao, Y., The Landscape of complex networks. 2012-05.
- [6] Jiang, X. Y., Yao, Y., Liu, H. and Leonidas Guibas, Compressive network analysis. 2012-06.
- [7] Lee, L. F. and Yu, J. H., Spatial panels: Random components vs. Fixed effects. 2012-07.
- [8] Wang, R. D., Peng, L. and **Yang, J. P.**, Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. 2012-08.
- [9] Michael A. Kouritzin and **Ren, Y. X.**, A strong law of large numbers for super-stable processes. 2012-09.
- [10] Xu, Q. Q., Huang, Q. M. and Yao, Y., Online crowdsourcing subjective image quality assessment. 2012-10.
- [11] Jiang, B. C. and Ai, M. Y., Complementary design theory for uniform designs. 2012-11.
- [12] **Ren, Y. X.** and Yang, T., Multitype branching brownian motion and traveling waves. 2012-12.
- [13] Chen, Y. and **Liu, Y.**, On the eigenfunctions of the complex Ornstein-Uhlenbeck operators. 2012-13.
- [14] **He, Y. B.**, **Jia, J. Z.** and Yu, B., Reversible MCMC on markov equivalence classes of sparse directed acyclic graphs. 2012-14.
- [15] Lee, L. F., and **Yu, J. H.**, QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. 2012-15.
- [16] **Jia, J. Z.** and Karl Rohe, Preconditioning to comply with the irrepresentable condition. 2012-16.
- [17] Jiang, Q., Wang, H. S., Xia, Y. C. and Jiang, G. H., On a principal varying coefficient model. 2012-17.

- [18] **Chen, S. X.** and Ingrid Van Keilegom, Estimation in semiparametric models with missing data. 2012-18.
- [19] Li, C. X., Bessel Processes, Stochastic volatility, and timer options. 2012-19.
- [20] **Yu, J. H.**, R. de Jong and Lee, L.F., Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. 2012-20.
- [21] **Ai, M. Y.**, Li, K., Liu, S. M. and Dennis K. J. Lin, Balanced incomplete latin square designs. 2012-21.
- [22] Tao, J. and Yu, J. H., The spatial time lag term in panel data models. 2012-22.
- [23] Qian, J. Y. and **Jia, J. Z.**, Onpattern recovery of the fused lasso. 2012-23.
- [24] János Engländer, **Ren, Y. X.** and Song, R. M., Weak extinction versus global exponential growth of total mass for superdi usions corresponding to the operator Lu + β u ku2. 2012-24.
- [25] **Xi, R. B.**, Semin Lee and Peter Park, A survey of copy number variation detection tools based on high-throughput sequencing data. 2012-25.
- [26] An, B., Wang, H. S. and Guo, J., Testing the statistical significance of an ultra-high dimensional na webayes classifier. 2012-26.
- [27] **Chen, S. X.**, Peng, L. and Yu, C. L., Parameter estimation and model testing for markov processes via conditional characteristic functions. 2012-27.
- [28] **Tu, Y. D.**, Su, L. and A. Ullah, Testing additive separabilty of error termin nonparametric. 2012-28.

六、 中心科研进展介绍:

中心成立以来同数学学院、光华管理学院展开联合招聘,共有三名获得美国著名高校博士学位的年轻教员加盟统计中心,在中心的支持下在各自研究领域开展了一系列的科研活动,取得了一定的成果。

一)、贾金柱,副研究员,联合受聘于北京大学统计科学中心和北京大学数学学院概率统计系。2009年获北京大学博士学位,攻读博士期间赴美国加州大学伯克利分校交流访问1年,获学位后在加州大学伯克利分校做博士后。研究领域为高维统计推断、统计机器学习、因果推断等。在变量选择方法的理论研究、高维数据统计学习的应用、因果推断等领域发表论文多篇。



主持或参加的基金和项目:

- 1. NSFC 11101005, Sparse functional ANOVA and Sparse high-dimensional Bayesian networks. 2012/01-2014/12, 220K RMB, hosted.
- NSFC 61121002, Theory and methods in high fidelity vision and audio system 2012/01- 2015/12, 150K RMB, Joined

2012年,贾金柱在高维统计推断、高维统计方法的应用、因果推断领域做了深入的分析。以下是对各个成果的简介。

(1) Jinzhu Jia, Karl Rohe and Bin Yu (2013). The Lasso under Heteroscaddasticity. Statistical Sinica2013, pp 99 – 118.

人们对于 Lasso 这一方法的理解基于数据来自于齐次线性模型,即

$$Y = X \beta + \epsilon$$

其中C是一个和X独立的且每一个分量独立同分布的误差向量。但是实际问题中,如医学成像问题,误差往往随着观测量Y的增大而增大。这个时候是否仍然使用Lasso就成为一个亟待回答的问题。如果答案是肯定的,就说明了Lasso这一方法具有稳健性。

本文证明在误差和观测值成正比的时候, Lasso 仍然可以很好地选择相关的变量, 具有稳健性。 我们给出了 Lasso 在解决非奇性方差情形下的大样本性质, 为实际工作中使用 Lasso 这一方法奠定 了理论基础。我们指出即使误差向量e的协方差矩阵是已知的时候, 也不需要通过线性变换把非其 次线性问题转化为其次线性问题。这种变换对于 Lasso 的模型选择相合性有时候具有破坏作用。

(2) Jinzhu Jia and Karl Rohe. Preconditioning for sparse inference. (submitted)

Lasso 这一方法的大样本性质(相合性) 需要条件。它的符号相合(模型选择相合性)需要irrepresentable condition,它的 L2 相合性需要限制特征值条件 (restricted eigenvalue condition)。本文展示和提出如何规避稀疏线性回归分析需要的条件。给定 X, Y 满足如下线性回归方程:

$$Y = X \beta + \epsilon$$

本文提出如果设计矩阵 X 不满足 irrepresentable condition, 通过选择一个合适的矩阵 F, 对方程进行如下的预处理: 回归方程两边都乘以 F

$$FY = F X \beta + F \epsilon$$

通过选择合适的 F, 解如下的 Lasso 问题

$$\min_{\beta} || FY - FX \beta ||_2^2 + \lambda ||\beta||_1$$

可以去掉Lasso 所需要的 irrepresentable condition 或者 restricted eigenvalue condition. 我们提出的 F 如下构成:记 X 的 SVD 分解为

$$X = UDV^{T}$$
,

则 F 定义为 $F = UD^{-1}U^{T}$.

由于F的定义取决于X的SVD分解,我们称这种预处理方法为基于SVD分解的与处理方法。 我们的预处理使得变换后的设计矩阵(FX)的特征值只有0或者1。当变量的个数(n)小于观测数

据(p)的个数时,变换后的矩阵是正交矩阵,这个时候 irrepresentable condition 和 restricted eigenvalue condition 都成立。当变量的个数大于观测数据个数的时候,变换后的设计矩阵行正交,此时我们也证明了 irrepresentable condition 和 restricted eigenvalue condition 都以很大的概率成立。我们证明了在高维回归中,预处理可是规避 irrepresentable condition, 并且明显改进 Lasso 的性质。随机模拟也进一步证明了我们的理论。

(3) Junyang Qian and Jinzhu Jia. On pattern recovery of Fused Lasso.(submitted)

Fused Lasso 常常用于信号处理,特别是当临近的信号具有群组特征的时候,一个特殊的情形就是一维分段常数的信号。Fused Lasso 定义如下:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i} (y_i - \beta_i)^2 + \lambda \sum_{i} |\beta_i - \beta_{i-1}|,$$

其中, y 是带有噪声的信号, β 是待估计的去噪声的分段常数信号。事实上, 通过线性变换, fused Lasso 可以转化为 Lasso 问题。通过对等价的 Lasso 问题的分析, 我们指出, fused Lasso 几乎不可能是相合的。只有极少数情况下才是相合的。通过对等价的 Lasso 问题进行如同前面的预处理, 可以使得 Irrepresentbale condition 成立。这时,进行预处理后的估计就是一个相合的估计了。该工作进一步说明了预处理的重要性。图 1 是预处理前后信号估计的对比图。图中黑色的点是含有噪声的信号,红色是真实的不含噪声的信号。左图黑色是用 fused lasso 对信号的恢复,右图黑色是在预处理之后恢复的效果。对比发现,预处理之后,同一组的信号仍然是在同一组。预处理后的信号值与真实信号值有些差距,是由于 Lasso 的估计是有偏的估计造成的。这不是一个大问题,可以通过进一步的最小二乘解决,即对估计为同一组的信号进行平均即可。

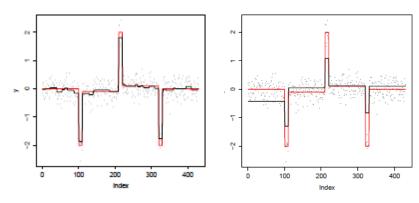


图 1. 左图: Fused Lasso 对信号的恢复; 右图: 预处理后的结果。红色是不含噪声的信号。黑色点是观测数据。右图很好地恢复了信号的分组模式。

(4) Luke Miratrix, Jinzhu Jia, Bin Yu, Brian Gawalt, Laurent El Ghaoui, Luke Barnesmoore and Sophie Clavier. Concise comparative summaries (ccs) of large text corpora with a human experiment (submitted). 为了能够理解公众如何看待世界,我们要能够从描述世界的文本中抽取具体的详细的总结性词语。本文发展了一套自动方法,该方法用于抽取关于人们感兴趣话题的总结性词语。这些总结词语理想地显示了这些话题和其他的概念之间的联系。我们使用了美国纽约时报2009到2011年的新闻以及华尔街日报2008到2011年的新闻标题作为例子。事实上我们的方法还可以用到更到的文章中。因为这些文章很多,单凭人们手工无法完成如此巨大的工作量。自动提取的方法就变得更加有用,它能够提供文献之间的比较,从而帮助人们进一步的阅读。

我们提出了一个一般的框架用于给定话题的总结性词语的提取,并且使用两个例子来说明我们方法的用途。我们的方法是基于稀疏分类方法的。与一些复杂的方法如LDA相比,该方法更简单、运行速度更快。

为了验证我们的工具,我们设计了一套人工问卷调查以比较不同的总结的优势。我们有三个主要的结论:首先Lasso在所有的模型选择方法中是表现最好的,其次,对于很长的文章,TF-IDF是一个很好的预处理方法;最后,对于很短的文章,L2方法是很好的预处理方法。

(5) Yangbo He, Jinzhu Jia and Bin Yu. Reversible MCMC on Markov equivalence classes of sparse directed acyclic graphs. (submitted).

图模型是一个很有用的表示因果关系的工具。但是通常从数据中不能惟一地确定因果关系。等价的因果关系称为等价类。描述等价类空间的性质对于认识等价类以及学习等价类都是非重要的。然而,已知的文献中,仅仅对20个点(变量)以内的等价类做过深入的分析。本文中,我们设计一个可逆非可约的马氏链。该马氏链的每一个状态为一个等价类。这个马氏链的平稳分布可以很容易的计算出来。有了这个马氏链以及平稳分布,我们就可以通过计算得到等价类的性质。我们的方

法可以处理数千个变量。对于稀疏的因果图模型,实验发现 (1) 大多数的边都是有向边, (2) 大多数的无向子图都只有少数几个点组成 (3) 无向子图的数量随着变量个数的增多线性增长。

二)、席瑞斌、研究员、联合受聘于北京大学统计科学中心和北京大学数学学院概率统计系。2002年在内蒙古大学数学系获得学士学位,2005年及2007年分别在北京大学数学学院和华盛顿大学圣路易斯分校数学系获得硕士学位,2009年在华盛顿大学圣路易斯分校完成博士论文,并于当年至2012年在哈佛医学院生物医学信息中心担任副研究员。其主要研究方向是Statistical theory and algorithm development和 Computational genomics and bioinformatics.



2012 年度,席瑞斌研究员一篇文章于 Current Protocols in Human Genetics (CPHG)发表,另一篇文章被 Genome Research (GR)接受。发表于 CPHG 的文章为一篇综述文章。此文章系统的回顾了最近几年发表的关于探测结构变异及拷贝数变异的算法,分析了这些方法的优缺点以及未来发展的方向。图 1 显示了目前基于高通量测序数据检测结构变异

及拷贝数变异的四种基本策略,即测序深度法(read depth),两端映射法(paired-end mapping), 短列分割法(split read)及从新拼装法(de novo assembly)。在发表于 GR 的文章中,我们综合的 分析了今年来发表于十几篇文章中的 8000 多个癌基因组中拷贝数变异的情况(即 meta analysis)。 此文章中,我们的分析表明,一些常见的与癌症的发生及发展密切相关的基因(如 TP53, EGFR, ERBB2 等)常常会在癌基因组中有拷贝数增加或减少,同时我们还发现一些潜在的可以指导寻找癌症治疗靶点的相关变异。

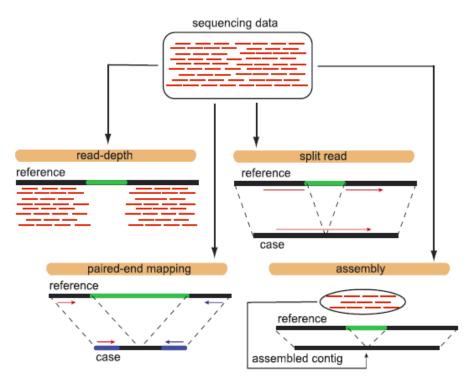


图 1 探测结构变异的四种基本策略

三)、涂云东,助理教授,联合受聘于北京大学统计科学中心和北京大学商务统计与经济计量系。2004年获得武汉大学理学学士学位,于 2006年获得武汉大学经济学硕士学位。2007年前往美国加州大学河滨分校攻读经济学博士学位,2012年加盟北大。



其主要研究领域为计量经济学理论和应用计量经济学。研究兴趣涵盖非参数建模,高维数据建模,信息计量经济学,金融计量学,经济预测,生产效率建模,宏观计量经济学等。他的两篇论文已经分别被世界一流期刊 Journal of Econometrics 和 Econometric Reviews 接受。他的研究致力于如何利用已有的信息或数据进行更加有效

的推断和经济预测,让数据自身来揭示背后的经济规律,减少经济研究中作出的不合理 假设。其当前研究主要有

- (1) 非参数模型和不可分模型试图让数据来确定模型的函数形式(不一定为线性或已知非线性)以及扰动项对因变量的扰动形式(不一定为可分),建立模型的估计和检验方法,并将其应用在经济学等应用问题中。
- (2)信息计量方法利用信息学里的信息度量熵以及最大熵原理对经济个体在经济运作中的贡献进行评估,在作出最少模型假设的情况下实现模型估计和检验。同时,利用熵作为不确定性的度量解决模型不确定性下的模型平均建模和统计推断。
- (3)高维数据建模着力改善当前高维数据建模中所作的模型选择和模型平均方法,提出新的模型选择和平均方法。
- (4)区间数据建模处理区间数据(例如股票市场的开盘收盘价格区间)的建模和统计推断。把传统的点数据建模拓展到符号数据,提出新的回归模型和分析方法。
- (5) 宏观计量对宏观经济变量(通货膨胀,经济增长等)建模,识别其动态过程和决定因素。

七、 学术讲座:

中心 2012 年共组织学术报告 46 场,报告人来自美国,澳大利亚,新加坡,台湾,香港地区及国内的高校、研究机构和国际知名企业。

[1]Model Averaging PartiaL Effect (MAPLE) Estimation with Large Dimensional Data 报告人: Yundong Tu, University of California

时 间: 2012年2月13日下午10:00-11:30; 地 点: 光华管理学院新楼217室

[2]Inference in Semiparametric Conditional Moment Models with Partial Identification

报告人: Shengjie Hong, University of Wisconsin-Madison

时间: 2012年2月16日下午2:00-3:00; 地点: 光华管理学院新楼217室

[3] Tests for Overidentifying Restrictions in Factor-Augmented VAR Models

报告人: Xu Han, North Carolina State University

时 间: 2012年2月20日上午10:00-11:30; 地 点: 光华管理学院新楼217室

[4]Covaraite-adaptive Designs For Personalized Medicine

报告人: Prof. Feifang Hu, Department of Statistics, University of Virginia And School

of Statistics, Renmin University of China

时 间: 2012年3月8日下午2:00-3:00; 地 点: 光华管理学院新楼217室

[5] Methodology and Theory for Partial Least Squares Applied to Functional Data

报告人: Prof. Peter G. Hall Professor and ARC Federation Fellow Department of Mathematics and Statistics University of Melbourne, Australia

时 间: 2012 年 3 月 15 日 (周四) 下午 4:00-5:00; 地 点: 中科院数学与系统科学研究院思源楼 712

[6]Multi-Stage Designs for Evaluation of Multidimensional Mixed Clinical Outcomes

报告人: Ming T Tan, Professor and Head, Division of Biostatistics and Bioinformatics, University of Maryland Greenebaum Cancer Center & Department of Epidemiology and Public Health, University of Maryland School of Medicine

时 间: 2012年3月16日下午4:10-5:00; 地 点: 理科一号楼 1493教室

[7]Structure Estimation in Semiparametric Regression

报告人: Prof. Jian Huang, Department of Statistics and Actuarial Science, College of Liberal Arts and Sciences epartment of Biostatistics School of Public Health, University of Iowa

时 间: 2012年3月29日下午2:00-3:00; 地 点: 光华管理学院新楼217室

[8] Solving the bond-pricing problem for non-affine interest rate diffusions

报告人: Prof. Robert Kimmel, EDHEC Business School

时间: 2012年4月12日下午2:00-3:00; 地点: 光华管理学院新楼217室[9]高维数据的结构降维

报告人: 郭建华教授, 东北师范大学

时间: 2012年4月17日下午2:00-3:00; 地点: 理科一号楼1556教室

[10]On the Wellposedness of Forward-Backward SDEs— A Unified Approach

报告人: Prof. Zhen Wu, 山东大学

时 间: 2012年4月19日2:00-3:00; 地 点: 光华管理学院新楼217室

[11]Comparing the Expected Misclassification Cost for Two Classifiers based on Estimates From the Same Sample

报告人: Prof. Kai Fun Yu, Tsinghua University

时间: 2012年4月26日下午4:00-5:00; 地点: 理科一号楼 1556 教室

[12]Post-Fisherian Experimentation: from Physical to Virtual

报告人: 吴建福教授, 美国乔治亚理工学院

时间: 2012年5月8日下午2:00-3:00; 地点: 北京国际数学研究中心报告厅

[13]当前中国宏观经济形势分析与预测

报告人: 李雪松教授, 中国社会科学院

时 间: 2012年5月11日下午3:00-4:30; 地 点: 光华管理学院新楼216室

[14]On estimation of the population spectrum from large dimensional covariance matrices

报告人: Prof. Jianfeng Yao, The University of Hong Kong

时 间: 2012年5月17日下午2:00-3:00; 地 点: 光华管理学院新楼217室

[15] An Introduction to Bayesian Analysis of Hierarchical Models

报告人: Prof. George C. Tiao, The University of Chicago、北京大学名誉教授

时 间: 2012年5月17日下午3:30-4:30; 地 点: 光华管理学院新楼217室

[16]A Sampler of Applied Statistics

报告人: Prof. Jianhua Huang, Texas A&M University

时间: 2012年6月1日上午10:30-11:30; 地点: 理科一号楼1556教室

[17]Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation

报告人: Prof. Yehua Li, Iowa State University

时 间: 2012年6月5日下午2:00-3:00; 地 点: 光华管理学院新楼217室

[18]Multivariate density estimation

报告人: Prof. Wing Hung Wong, Stanford University

时间: 2012年6月6日下午2:00-3:00; 地点: 理科一号楼1114教室

[19]Robust stepwise regression methods for high-dimensional variable selection

报告人: Prof. Jingshiang Hwang, Academia Sinica

时 间: 2012年6月7日下午2:00-3:00; 地 点: 光华管理学院新楼217室

[20]Design and Analysis Considerations in Clinical Trials With a Sensitive Subpopulation

报告人: Prof. Yan Daniel Zhao, University of Oklahoma Health Sciences Center at Tulsa

时 间: 2012年6月7日下午3:15-4:15; 地 点: 光华管理学院新楼217室

[21] High-Dimensional Statistical Inference: From Vectors to Matrices

报告人: Prof. Tony Cai, University of Pennsylvania

时 间: 2012年6月21日下午2:00-3:00; 地 点: 理科一号楼1114教室

[22]Semiparametric Double Balancing Score Estimation for Incomplete Data with Ignorable Missingness

报告人: Prof. Jing Qin, The National Institute of Allergy and Infectious Diseases

时间: 2012年6月21日下午3:15-4:15; 地点: 理科一号楼1114教室

[23]On the maximal size of Large-Average and ANOVA-fit Submatrices of a Gaussian Random Matrix

报告人: Prof. Andrew B. Nobel, The University of North Carolina at Chapel Hill

时间: 2012年6月26日下午2:00-3:00; 地点: 理科一号楼1556教室

[24] A new approach of functional estimation for high-dimensional inputs

报告人: Prof. Xiaoming Huo, The Georgia Institute of Technology

时 间: 2012年6月28日下午2:00-3:00; 地 点: 光华管理学院新楼217室

[25]Independent Component Analysis Involving Autocorrelated Sources with Application to Functional Magnetic Resonance Imaging

报告人: Prof. Haipeng Shen, The University of North Carolina at Chapel Hill

时 间: 2012年6月28日下午3:15-4:15; 地 点: 光华管理学院新楼217室

[26] Integrating ideas from statistics and engineering into behavioral intervention science

报告人: Linda M. Collins, The Pennsylvania State University

时 间: 2012年6月29日上午10:00-11:00; 地 点: 光华管理学院新楼216室

[27]Clinic and ambulatory BP measurement: A valid and cost-effective approach using the two-method measurement design

报告人: John W. Graham, The Pennsylvania State University

时 间: 2012年6月29日上午11:00-12:00; 地 点: 光华管理学院新楼216室

[28] Smoothed jackknife empirical likelihood method for ROC curves with missing data

报告人: Prof. Yichuan Zhao, Georgia State University

时间: 2012年7月5日下午2:00-3:00; 地点: 理科一号楼1303教室

[29]On the robustness of penalized variable selection to model misspecification

报告人: Prof. Jason Fine, The University of North Carolina at Chapel Hill

时 间: 2012年7月6日下午2:00-3:00; 地 点: 光华管理学院新楼216室

[30]Large Scale Multinomial Inference and Statistical Methods for Confounding Effects

in Genome Wide Association Studies

报告人: Prof. Jun Xie, Purdue University

时间: 2012年7月10日上午9:00-10:00; 地点: 理科一号楼 1560 教室

[31] Evaluating the Effect of Training on Wages in the Presence of Noncompliance,

Nonemployment, and Missing Outcome Data

报告人: Prof. Donald B. Rubin, Harvard University

时 间: 2012 年 7 月 10 日下午 2:00-3:00; 地 点: 光华管理学院新楼 110 室

[32]Multilabel Classification in Disease Diagnosis Using Public Gene Expression Data

报告人: Prof. Haiyan Huang, Associate Professor, Department of Statistics, University of California, Berkeley

时 间: 2012年7月24日上午10:00-11:00; 地 点: 理科一号楼1479教室

[33] Composite Joint Estimating Functions in Spatio-Temporal Models

报告人: Prof. Peter Song, University of Michigan

时 间: 2012年8月23日下午4:00-5:00; 地 点: 理科一号楼1479教室

[34]Interaction Detection for Index Models in High Dimension

报告人: Prof. Jun Liu, Department of Statistics, Harvard University

时 间: 2012 年 8 月 24 日下午 4:00-5:00; 地 点: 光华管理学院新楼 217 室

[35] How good is the first order asymptotic normality of MLE

报告人: Prof. Kai Ng, Patrick S C Poon Professor in Statistics and Actuarial Science, The University of Hong Kong

时 间: 2012年9月13日下午4:00-5:00; 地 点: 理科一号楼1418教室

[36]ARFIMA--GARCH 模型的自加权指数极大似然估计

报告人:陈敏,中科院数学与系统科学研究院研究员

时 间: 2012年10月18日下午3:00-4:00; 地 点: 理科一号楼1114教室

[37]统计学习和个性化推荐系统

报告人:郭翔,优推网 utuii.com 联合创始人兼首席数据官

时 间: 2012年10月18日下午4:00-5:00; 地 点: 理科一号1114教室

[38]缺失数据模型若干问题研究

报告人: 唐年胜教授,云南大学

时间: 2012年11月12日下午2:00-3:00; 地点: 理科一号楼1114教室

[39] Statistical Inference Using Maximum Correlation Portfolios

报告人: Prof. Robert L. Kimmel, National University of Singapore and EDHEC Business School

时间: 2012年11月13日上午10:00-11:30; 地点: 光华管理学院新楼216室 [40]Change-point Methodology and Its Applications

报告人: Prof. Tze Leung Lai, Department of Statistics, Stanford University

时 间: 2012年11月20日晚上7:30-9:00; 地 点: 光华管理学院新楼109室

[41]Optimal Investment Strategies for Consistent Performance

报告人: Prof. Weidong Tian, University of North Carolina at Charlotte

时间: 2012年11月27日下午2:00-4:00; 地点: 光华新楼 K02会议室

[42]A New Semiparametric Quantile Panel Data Model with Estimating the Growth Effect of FDI

报告人: Prof. Zongwu Cai, University of North Carolina at Charlotte and Xiamen University

时 间: 2012年12月14日上午10:30-12:00; 地 点: 光华新楼 K02会议室

[43] Robust Sparse Estimation of Multi-Response Regression

报告人: Prof. Xinwei Deng, Department of Statistics, Virginia Tech

时间: 2012年12月14日下午3:30-4:30; 地点: 理科1号楼1418教室

[44]Dependence notions through arrangement increasing survival and density functions with applications in optimal allocation problems

报告人: Prof. Jun Cai, Department of Statistics and Actuarial Science, University of Waterloo

时间: 2012年12月20日下午2:00-3:00; 地点: 理科1号楼1418教室

[45] Spectral Clustering by Higher Criticism Thresholding

报告人: Dr. Wanjie Wang, Carnegie Mellon University

时间: 2012年12月24日上午10:30-11:30; 地点: 理科1号楼1569室

[46]A generalized beta copula with applications in modeling multivariate long-tailed data

报告人: Prof. Zhengjun Zhang, University of Wisconsin-Madison

时间: 2012年12月24日下午2:00-3:00; 地点: 理科1号楼1114室

八、 开设课程:

本年度中心继续联合国内外多个国家地区杰出的统计学专家学者,为本校、本市及全国各地高校的学生和教师开办了四门短期课程,除了本校数学学院、光华管理学院等院系的学生学者外,来自中科院、清华、人大、北理工、北科大、北工大、中央财经、首经贸、华东理工、东北师大、华东师大、华中师范、华中理工、天津大学等十几所高校相关院系的师生也积极参加中心的各类课程讲座。

短期课程:

1) Methodology And Theory for The Bootstrap

授课教师: Prof. Peter G. Hall

Professor and ARC Federation Fellow

Department of Mathematics and Statistics, University of Melbourne

课程时间: 2012年3月16日(周五)、3月18日(周日)、

3月20日(周二)、3月22日(周四)

课时数: 16课时

课程地点:北京国际数学研究中心报告厅

2) Monte Carlo Methods in Bayesian Computation

授课教师: Prof. Ming-Hui Chen, Department of Statistics, University of Connecticut

课程时间: 2012年7月11日(周三)

课时数: 4课时

课程地点:北京大学光华管理学院新楼 217 教室

3) Variable Selection in Very High Dimensions

授课教师: Prof. Jiashun Jin, Statistics Department, Carnegie Mellon University
Prof. Yingying Fan, Information and Operations Management Department,
University of Southern California

课程时间: 2012年7月25日(周三)、7月26日(周四)、7月27日(周五)

课时数: 6课时

课程地点:北京大学光华管理学院新楼 217 教室

4) Random Matrix Theory

授课教师: Prof. Tiefeng Jiang, Department of Statistics, University of Minnesota Prof. Xinghua Zheng, Information Systems, Business Statistics and Operations Management at Hong Kong University of Science and Technology

课程时间: 8月6日(周一)、8月7日(周二)、8月8日(周三)

课时数: 5课时

课程地点:北京大学光华管理学院新楼 217 教室

九、 交流访问:

1) 中心 2012 年共接待海内外访问学者 57 人次,参与学术报告、短期课程、合作项目 等各类学术活动。

2012 visitor list

December

Visitor	Zhengjun Zhang	Visitor	Wanjie Wang
From	The University of Wisconsin	From	Carnegie Mellon University
Date	Dec. 24 - Dec.26	Date	Dec. 24

Host Prof.Jingping Yang Host Prof.Songxi Chen

VisitorJun CaiVisitorXinwei DengFromUniversity of WaterlooFromVirginia TechDateDec. 20- Dec. 24DateDec. 14

Host Prof. Yanxia Ren Host Prof. Jinzhu Jia

Visitor Zongwu Cai Visitor Yanping Yi

From North Carolina at Charlotte From Shanghai University

Date Dec.14- Dec.17 **Date** Dec.04

Host Prof. Yundong Tu Host Prof. Yundong Tu

November

Visitor	Weidong Tian	Visitor	Tze Leung Lai
From	North Carolina at Charlotte	From	Stanford University
Date	Nov.27 - Nov.29	Date	Nov.20- Nov.22
Host	Prof.Chenxu Li	Host	Prof.Chenxu Li

VisitorRobert L. KimmelVisitor唐年胜FromNational University of SingaporeFrom云南大学DateNov.13- Nov.15DateNov.12- Nov.13HostProf.Chenxu LiHostProf.Jinzhu Jia

October

Visitor	郭翔	Visitor	陈敏
From	优推网 utuii.com	From	中科院
Date	Oct.20	Date	Oct.18

Host Prof.Jinzhu Jia Host Prof.Jinzhu Jia

September

Visitor Kai Ng

From University of Hong Kong

Date Sept.13 - 15Host Prof.Jinzhu Jia

August

Visitor Jun Liu Visitor Peter Song

From Harvard University From University of Michigan

Date Aug.23 **Date** Aug.24 - 31

Host Prof.Songxi Cheng Host Prof.Songxi Chen

July

Visitor Haiyan Huang Visitor Minghui Chen

From University of California, Berkeley From University of Connecticut

Date Jul.24 Date Jul.11-13

Host Prof.Minghua Deng Host Prof.Zhi Geng

Visitor Donald B. Rubin Visitor Jun Xie

From Harvard University From North Carolina at Chapel Hill

Date Jul.10 – 12 **Date** Jul.6

Host Prof.Junni Zhang **Host** Prof.Hansheng Wang

Visitor Jason Fine Visitor Yichuan Zhao

From Georgia State University From Carnegie Mellon University

Date Jul.5-6 Date Jul.5

Host Prof.Hansheng Wang **Host** Prof.Junshan Shen

June

Visitor John W. Graham Visitor Linda M. Collins

From Pennsylvania State University From Pennsylvania State University

Date One week **Date** One week

Host Prof. Hansheng Wang **Host** Prof. Hansheng Wang

Visitor HaipengShen Visitor Xiaoming Huo

From North Carolina at Chapel Hill From Georgia Technology

Date Jun.28-30 Date Jun.28

Host Prof.Dayue Chen **Host** Prof.Hansheng Wang

VisitorAndrew B. NobelVisitorJing QinFromNorth Carolina at Chapel HillFromNIAID

Date Jun.26-29 **Date** Jun.20-22

Host Prof.Songxi Chen Host Prof.Songxi Chen

Visitor Tony Cai Visitor Yan Daniel Zhao

From University of Pennsylvania From OSU Center for Health

Date Jun.20-22 Date Jun.07

Host Prof.Dayue Chen **Host** Prof.Mingyao Ai

VisitorJingshiang HwangVisitorWing Hung Wong

From Sciences Academia Sinica From Stanford University

Date Jun.06- 09 **Date** Jun.05-08

Host Prof.Songxi Chen Host Prof.Songxi Chen

Visitor Yehua Li Visitor Jianhua Huang

From Iowa State From University Texas A&M

Date Jun.04-08 Date Jun.01

Host Prof.Songxi Chen Host Prof.Mingyao Ai

Visitor Cheng Li

From University HSPH & DFCI

Date Jun.01

Host Prof.Songxi Chen

May

Visitor George C. Tiao Visitor Jianfeng Yao

From The University of Chicago From University of Hong Kong

Date 2 weeks **Date** May.17

Host Prof.Songxi Chen Host Prof.Hansheng Wang

Visitor 李雪松 Visitor 吴建福

From 中国社会科学院 From Georgia Technology

Date May.11 Date May.07-10

Host Prof.Jihai Yu Host Prof.Dayue Chen

April

VisitorKaifun YuVisitorZhen WuFromTsinghua UniversityFrom 山东大学DateApr.26Date Apr.19

Host Prof.Mingyao Ai Host Prof.JingpingYang

Visitor 郭建华 Visitor Robert Kimmel

From 东北师范大学 From EDHEC Business School

DateApr.17DateApr.11-14

Host Prof.MingyaoAi Host Prof.Chenxu Li

March April

Visitor Jian Huang Visitor Ming T. Tan

From University of Iowa From University of Maryland

Date Mar.28-30 Date Mar.16

Host Prof.Songxi Chen Host Prof.Mingyao Ai

Visitor Peter G. Hall Visitor Feifang Hu

From University of Melbourne From University of Virginia

Date Mar.15-23 Date Mar.08

Host Prof.Songxi Chen Host Prof.Songxi Chen

February

Visitor Xu Han Visitor Shengjie Hong

From North Carolina State University From University of Wisconsin-Madison

Date Feb.20 **Date** Feb.16

Host Prof.Songxi Chen Host Prof.Songxi Chen

Visitor Yundong Tu

From University of California

Date Feb.13

Host Prof.Songxi Chen

- 2) 短期学术交流:
- ① 邀请宾夕法尼亚州立大学暴乐博士合作项目:生物统计方向应用(2012.07.10-08.26)
- 项目申请人: 邓明华
- 项目参加人: 邓明华、暴乐
- 暴乐: 毕业于University of Washington统计系,目前是Penn State University统计系的 Assistant Professor。
- 项目的研究内容:
 - (1)通过network clustering和Bayesian方法分析基因的功能相关性。这个题目是我现在进行的主要研究项目之一,希望通过这次合作解决network中成对基因的相互作用的非独立性问题。
 - (2)通过clustering和Bayesian方法分析人民医院提供的肺结核病数据,确定1535位结核病人的具体致病菌。
 - (3)暴乐教授在流行病数据分析方法有一定的积累,他的若干统计方法都成为联合国推荐的艾滋病分析软件EPP的重要组成部分: incremental mixture importance

sampling, stochastic infection rate model, mean-shift infection rate model, hierarchical model。我们可以考虑将类似的方法应用到正在进行的结核病研究。

- ② 邀请密歇根州立大学 Pingshou Zhong 教授合作项目(2012.06.27-07.26)
- 项目申请人:徐敏亚
- 项目参加人:徐敏亚、Pingshou Zhong、陈松蹊
- Pingshou Zhong: Michigan State University.
- ③ 邀请首尔大学的 Panki Kim 教授合作项目:研究带边界的分支 Levy 过程的性质 (2012.08.01-08.10)
- 项目申请人: 任艳霞
- 项目参加人: 任艳霞、Panki Kim
- Panki Kim: 于2004年在美国University of Washington获博士学位, 2004-2006在美国 University of Illinois at Urbana-Champaign做博士后两年, 之后一直在Seoul National University任教, 2008年晋升为副教授.

项目的研究内容: Panki Kim副教授是Levy过程、概率位势理论等方面的青年专家,研究成果丰富,国际交流非常活跃。任艳霞教授的研究方向主要是分支过程与超过程,而 Levy 过程与分支过程、分支 Hunt 过程、超过程等具有紧密的联系。双方试图研究带边界的分支 Levy 过程的性质,并希望探讨新的合作可能。

④ 邀请东京大学的舟木直久教授来校教授短期课程:Hydrodynamic limit and stochastic interacting systems (2012.09.19-09.28)

授课人: 舟木直久, 东京大学,舟木教授是国际概率论界在 Hydrodynamic limit and stochastic interacting systems 方面的权威之一。他曾在 2007 年获得过日本数学会的 秋季奖(该奖被认为是日本数学会的最高奖),以及在 2002 年获得第一次解析学会奖。

- 课程内容:
 - 1) 9月21日下午作了题为 Hydrodynamic limit for a multi-species system producing a self-organized aggregation 的学术报告.
 - 2) 9月20日下午,24日下午开设了题为 Some topics in stochastic partial differential equations 4 小时的短期课程.

3) 中心成员参加学术交流活动:

- [1] 3 月李辰旭出访日本东京,参加 Young Researchers Workshop in Finance;
- [2] 5 月李辰旭出访到新加坡国立大学访问交流;
- [3] 5 月中旬至 6 月中旬王明进到韩国首尔大学进行讲学;
- [4] 6月 15-19 日陈松蹊参加在希腊由"非参数统计国际协会"举办的会议,

并做"Testfor bandedness of high dimensional covariance matrices and bandwidth estimation"的报告;

[5] 6 月 20-24 日徐敏亚参加在香港举行的中国管理研究国际学会(IACMR),

并做 "When Emotionally Exhausted Employees Speak Out?" 的演讲和 "Safety Comes first: A Multilevel Examination of Employee Unsafe Behaviors and Safety Outcomes"的 论文圆桌讨论;

- [6] 6月11日-7月8日虞吉海到 Orleans University 访问交流, 并参加 18th International Conference on Panel data, Bank of France;
- [7] 2012 年 7 月 1 至 4 日, 艾明要在日本筑波大学举行的 IMS 亚太区国际统计会议 上作 30 分钟邀请报告 "Construction of orthogonal Latin hypercube designs";
- [8] 耿直于 2012 年 7 月 1 至 4 日,在日本筑波大学举行的 The Second Institute of Mathematical Statistics Asia Pacific Rim Meeting (IMS-APRM 2012)国际会议上做 50 分钟 Distinguished Lecture (特别讲座): Causal Inference and Learning Along Paths; [9] 7 月 5-6 日陈松蹊、李辰旭参加在湖南大学举办的"2012 金融工程与风险管理国际会议";
- [10] 7月 22 日在黄山举办的"2012年统计与管理科学国际学术会议"上陈松蹊担任大会程序委员会副主席。
- [11] 2012 年 8 月 1 日-12 月 31 日王汉生到美国 Pennsylvania State Univ.进行为期半年的合作研究;
- [12] 2012年8月7日到11日,席瑞斌参加由哈尔滨医科大学主办的第五届全国生物信息学与系统生物学学术大会,并作特邀报告。
- [13] 陈松蹊于 8 月 12-22 日参加新加坡国立大学数学研究所主办的 Meeting the Challenges of High Dimension,并做报告;

- [14] 2012年8月15至18日,艾明要在黑龙江大学举办的第八届海峡两岸统计与概率研讨会上作20分钟交流报告;
- [15] 李辰旭于9月出访日本京都,并参加 Workshop on Mathematical Finance and Related Issues;
- [16] 2012年11月9日,席瑞斌参加在中科院举办的2012年全国统计workshop,并作特邀报告;
- [17] 虞吉海于 12 月参加在北京举办的 International Conference on Policy Analysis Using Modern Econometric Methods.

十、 统计咨询:

中心为促进统计学与其他学科及相关单位的交叉合作,提高相关学科的数据分析水平,解决相关领域的科研及应用工作中遇到的统计相关问题,特面向校内外各相关院系及校外相关领域的单位提供了统计学咨询服务。本年度取得了一定的成绩,为相关院系及企事业单位解决了一系列问题,也提高了中心研究人员在实践应用中分析、研究水平:

1)、杨静平教授主持课题:中央国债登记结算有限责任公司的《中国国债发行策略的随机模拟模型研究》Ⅰ期和Ⅱ期项目。

国债发行成本与市场风险管理课题涉及国债发行管理方面的理论研究,具体而言,就是在成本与风险分析框架下帮助确定最优发债计划。该项研究将有助于财政部落实"十二五"规划中提出的完善预算管理和深化国债管理制度改革的要求,并进一步提升国债管理水平,同时也有助于中央结算公司进一步提升为发行人服务的水平。据财政部国库司国债管理处处长韦士歌介绍,从上世纪80年代开始,国际上对国债管理有关的理论探讨逐渐增多。目前的主流观点认为,国债管理的基本原则是在最小成本与最小风险两个目标之间进行权衡。基于这个原则设计的 CaR 模型已成为目前主流的国债管理工具之一,国际上许多国家的债务管理部门都在使用这一工具辅助国债发行管理工作。具体有以下几方面成果:

一是确定了中国的利率与宏观经济模型。课题组参考国外模型并根据中国实际情况加以 改造,同时从计量经济学方面进行了优化处理,使得模型的估计和预测效果更加稳定。 二是初步确定了融资需求模型及季度分配方案。课题组参考国外模型的做法确定了比较 可行的融资需求模型方案,同时还积极探索了更为复杂的融资需求模型。另外,根据中国的实际情况,课题组还建立了季度均衡融资方案。

三是摸索出具体的发行计划组合模拟路径实现方法,综合考虑多种风险计量方法。国外模型的相关文献中对国债发行的组合模拟方法涉及不多,课题组摸索出了具体的组合模拟路径,并且综合考虑了多个成本与风险指标。

- 2)、艾明要教授作为外协项目负责人,主持工业和信息化部技术基础项目"战斗部传爆序列的可靠性优化设计"的外协项目"计算机试验设计方法"。
- 3)、邓明华教授与北大生科院白书农教授就水稻雄蕊发育的全基因组表达数据分析进行合作。这项合作已经进行了两年多的时间,最近我们关于数据校正所提出的数学模型刚刚发表在 BMC Genomics 上 (BMC Genomics 01/2013 14(1):31.)。
- 4)、邓明华教授对中科院动物所韩春生研究员的蛋白质表达和基因表达的关系的统计分析进行咨询,文章发表在

http://www.mcponline.org/content/early/2013/01/16/mcp.M112.020123.full.pdf+html, 在文章中对中心参与咨询的教员及学生加以致谢。

5)、邓明华教授与中日友好医院李平研究员开展合作,对其糖尿病肾病中医方剂"糖肾方"的临床数据进行分析,并参与进一步的临床实验设计的讨论。

委员会

国际顾问委员会

主席: 王永雄 斯坦福大学

成 员: Peter J. Bickel 加州大学伯克利分校

 范剑青
 普林斯顿大学

 Peter G. Hall
 墨尔本大学

 黎子良
 斯坦福大学

马志明 中国科学院

吴建福 佐治亚理工学院

科学委员会

成 员: 陈大岳 北京大学数学科学学院

陈嵘 罗格斯大学,北京大学光华管理学院

陈松蹊 北京大学光华管理学院,国家级重大人才计划学者

刁锦寰 北京大学光华管理学院,芝加哥大学

耿直 北京大学数学科学学院

胡永华 北京大学医学部公共卫生学院

姚琦伟 伦敦政治经济学院,北京大学光华管理学院

(排序按首字母英文字母表)