陈松蹊老师采访



陈松蹊，国家特聘专家， 北京大学讲席教授，商务统计与经济计量系联合系主任、北京大学统计科学中心联席主任 国家首批“千人计划”入选者，加盟北大后主要致力于商务统计与经济计量学学科建设及北大统计学研究队伍的建设工作。

他是数理统计学会(Institute of Mathematical Statistics) 资深会员（fellow)，美国统计学会会士（fellow），国际统计学会 (International Statistics Institute) 当选会员 （elected member），国际数理统计学会 (IMS) 理事会常务理事( Council member)。他现在是The Annals of Statistics(统计年鉴) 副主编 (自2010年）；Journal of Business and Economic Statistics 副主编 （自2013年）；曾任Statistics and Its Interface 的联席主编 （2010-2013）

健桥：老师您大学是在北师大修读数学，您当时是怎么出于什么样的考虑选择了数学呢？

这个实际上具有时代的特点，我不是自己主动学数学的，我是被分配的。我79年考大学，当时最热的是生物，我的报名志愿都是我父亲和叔叔帮我报的，他们对我影响很大。他们说，你应该当一个医生，往后不管世界发生什么动乱，医生肯定是吃香的。而且同时我叔叔也不叫我离开北京，所以我的目标是读北京第二医学院，现在叫首都医科大学；中央直属的学校，不让我申请，所以我主要报的是北京市的学校，这样就不会被分配到外地，因为我叔叔当年就是被分配到外地去了，他们不想我重蹈覆辙。但由于我的分数已经挺高了，过了一本线，所以我需要填几个其它一本学校志愿，其中就包括了北师大的生物。后来这个档案到了北师大，北师大生物系老师跟数学系老师说，你看这个孩子，数学考得很高，物理也很好，应该学数学。就这样数学系把我录取了。我接到通知书一看是数学系，我也一愣，后来转念一想，学数学也行。整个过程非常的偶然。后来发现学数学实际上很适合我。不过如果我到医学院，接受另外一个训练，可能也会为好的医生，谁也说不准。

健桥：那您后来为什么选择统计了呢？

这个说来话长，因为我大学毕业是83年，中国已经开始进行改革，83年的时候我还来过北大，听厉以宁先生讲经济学的课，他讲一门课叫“英国病”，讲公共福利制度对经济效率的影响。当时是大学最后一学期，我每个礼拜都骑车到北大上课，非常享受。所以就想学经济，大学毕业后分配到北京经济学院做讲师，就是现在首都经贸大学。因为想研究经济，所以参加了一个讨论班，有北京市一些挺牛的人，学经济的、学文科的都在那里讨论。我是作为学数学的跟他们一块来学习。当时大家都是胸怀大志，想来改变中国，参与经济改革。当时自己也读了经济方面的书，《资本论》、萨缪尔森《经济学》、马克思的和西方的东西都读了一些。后来考研究生我准备到南开大学读计量经济，因为当时南开的计量经济是最好的。等到报名的时候，经济学院说不行，你不能去学经济，因为你是教数学的，你的研究生一定要到数学系念。你知道吗，30年来中国发生了很大的变化，现在人们有更多的选择，更多的对个人选择的尊重，那个时候好多都是安排的。所以没办法，我就回母校北师大去学数理统计。后来想数理统计也行，因为统计跟计量经济联系非常紧密，所以最后就学了统计学。

我是85年9月份（研究生）入学，同年10月份小平同志开放私人留学。之前的留学大都是公派的，85年10月开始，你自己联系到国外的学校，只要那边有人接收，你就可以走，留学热从那时开始。我85年开始念的研究生，没一个月就开放私人留学了。因为我那时是经济学院的代培研究生，必须把研究生读完。88年读完以后我马上就去国外读书了。正好那时有一个新西兰的教授David Vere-Jones（愿意接收我）。当时我北师大导师郑小谷老师以及到北师大访问的Vere-Jones，他俩帮我联系了去新西兰读书。到新西兰以后有一天在校园里看到一个澳洲国立大学招统计博士生的广告，联系人就是Peter Hall。由于它要求申请者有一篇小论文才行，我就把北师大硕士论文给翻译了，寄过去几个月就被接收了。

所以我是先去的新西兰，到了那里自然想去更好的学校，后来就去了ANU（澳洲国立大学），这个过程很随机，也非常偶然。

雪宁：去了澳洲读博，当时就确定了以后要做老师吗？

我一直愿意做老师，我喜欢成为学者，我的性格比较适合于这种，我对研究也很感兴趣。读博过程非常享受，我这两年学到很多东西，当然也有之前硕士阶段在国内的训练积累，所以我的博士只用了26个月，做了四篇论文。我跟Peter确实收获了很多东西，包括他的为人、他的学者风范以及他的治学精神，对我影响非常大。

健桥：对于统计系的博士生，您有什么建议吗？

我认为这是非常充实非常好的一个经历，这段时间对你的一生非常重要，能够跟你的导师一块找一个好的问题，把它解决了，同时又训练了自己的基本功，这真的是一段很棒的经历。

健桥：从您的个人发表还有讲授课程来看，您的理论气息很浓，您也有您的Mathematics Genealogy，您觉得统计理论和应用是什么关系？或者更尖锐一些，您认为您算是数学家还是统计学家？

我觉得我是统计学家。理论和应用，这俩都是统计。理论统计是利用理论来对统计学方法提供一个深入了解，用数学来验证它的一些性质，最终都是为了更好的应用。应用是统计学的“前锋”，一定要先有应用，你把这个应用做出来，之后慢慢升华出理论；弄出方法，再对其理论进行研究和评价。 如果理论能过得去，那它就是一个可以的方法，它过不去我们就去发明更好的方法，所以我想没有什么矛盾。

健桥：所以您也认为应该是先有个实际问题，从实际问题出发再做理论。

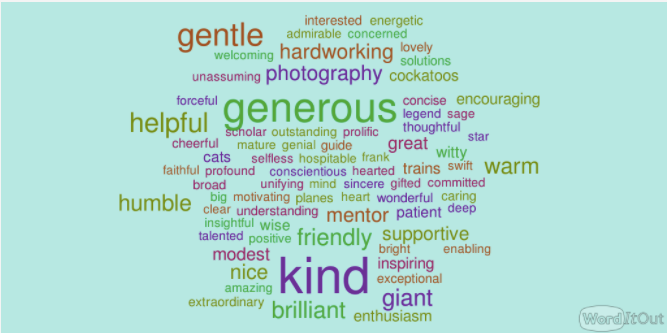
对，从宏观上，从我们学科发展来看是这样。但是对每一个人来讲，受他的环境或者他的老师的影响，每个人道路会不同。因为我的老师Peter非常理论，我受到他的影响做研究也是从理论角度，我看到什么问题我先想看能不能从理论上研究，如果理论上给出一个比较好的结果的话，那就很好了。我感觉理论上的洞察还是很重要。关于应用，我这几年做空气污染研究也接触了统计应用。应用不容易，应用有时候比理论还难。大家都是统计，实际上我觉得统计这个学科现在非常好，有一批非常优秀的人做应用来扩大统计的影响，把统计的底蕴做得很广；同时也有做理论的人把这一学科做得很深，很扎实。我想统计学科为什么有今天，主要是这两路人一起把这个大楼盖起来，你光有一方面不行。

统计整个就是一盘棋， 你不能说“车“是最厉害的，“炮“就不厉害，或者“象”就不厉害。而且我认为千万不要说我是做理论的，你应用怎么样；或者说我做应用的，你理论怎么样。大家都要尊重彼此，都不容易。我现在做一些应用，我就非常欣赏那些做应用的，非常困难，因为你要花大量时间处理数据。做应用，没有好的数据，质量不好，是做不出来的。相对来说理论可能更单纯一点，但你要有理论工具，有数学上的准备，这也需要很多功夫。总之统计是一个非常包容的，能够介于两个方面的学科，大家应该一块来促进它的发展。

健桥：到目前为止，您觉得对您学术生涯影响最大的人是谁，能不能分享一些你们之间的趣事呢？

那就是Professor Hall 了，他是一个非常友善的人。我在去年10月份的Annals(编者注：Annals of Statistics，权威统计学期刊)上写了一篇关于他的学术贡献的回顾文章，其中也提及了我和他之间的一个故事。

1990年我从新西兰坐飞机到堪培拉读博士。我下飞机已经晚上十点多了，他开车去接，之前他已经拿到了我们住的房子钥匙。因为当时大家都很穷，出国行李里面锅碗瓢盆都会带着，东西很多、很重。Peter身体非常强壮， 一手一个拿起行李来，直接给弄到车上。然后开到了住的地方，他问几楼，“三楼”，然后噌噌噌搬着行李就上去了。我都非常不好意思，我想这哪像老师，这是一个搬运工啊。你能感觉到他特别没有架子，非常非常和善。我们开研讨会，有时候找不到投影仪，他会去隔壁房间或者楼上给你拎一个特别沉的投影仪来。当然他也会给你很大压力，他学术上非常严格。但是只要不是学术的东西，他都特别generous and kind（慷慨友好）。在他去世以后，我们北大统计中心给他做了一个纪念主页，Terry Speed（编者注：前加州大学伯克利分校统计学教授）用所有的悼词留言制作了一个词云，都是kind（友好）， generous（慷慨）这些词，而不是说mathematics statistics（数理统计）或bootstrap（编者注：一种统计上重抽样方法，Peter Hall 教授对其理论发展做出了巨大贡献）。这些东西也有，但是最大的字还是他的人格，像kind ，generous。所以跟他学到很多，他非常大气。



（Peter Hall 的词云图，来源http://www.stat-center.pku.edu.cn/Peter\_Gavin\_Hall/）

健桥： 给我们读者推荐一本书吧，非统计领域的。

应该说我自从念博士以后读的（非专业书）就显著的少了。年轻的时候，爱看海明威的小说，但是我觉得他的小说太消沉了，结尾都是非常负能量。

如果推荐一本书，还是我自己80年代读的一本，我前些日子看到一个年轻人在读，叫《光荣与梦想》。它是美国近代历史的一个简单的缩写，那本书实际上写的蛮有激情的。

如果说中文的话，我最近看的比较喜欢的应该是《明朝那些事儿》。

健桥：您现在同时在爱荷华州立大学和北大教书，您觉得国内国外两所学校有什么区别或共同点吗？

我觉得Iowa State（爱荷华州立大学）统计系有着十分优良的统计传统，理论跟应用结合得非常好，你能看到它的老师在各个系兼职，这有点像我们北大统计中心招的老师是跟光华、数学院、生物进行兼职。但Iowa State（爱荷华州立大学）一直都这样做，他的教职人员有40个，差不多有20个人都是joint department（联合教职，即在不同的系兼职），（雪宁：40个人是一个非常大统计系；陈老师：是，非常大），所以他是一个良性发展的很好的例子，统计学为其他学科做出了贡献，人家得到实惠，也来反哺你，统计有这个特点，所以Iowa State有这样一个传统。

北大的话，我觉得北大也是历史非常悠久的学校，同时中国的发展很令人激动，很多变革正在发生。中国的发展，很有动力，很有能量，也有很多令人激动的挑战，我觉得都挺好的。

健桥：我们之前有一篇翻译的Leo Brieman的文章，畅想了20年后的统计系，你觉得未来20统计系会怎么发展呢？

**统计学目前发展得很好，但是我们要看到潜在的危机，尤其现在大数据来了，我们应该思考如何应对，怎么使我们的东西能应用到科学里面去，能融入到数据科学里面。**现在统计在各个方面做得已经很好了，下面怎么进一步地保持这个优良传统，我感觉还是充满挑战的。并不是我现在已经这样了，20年以后我一定还是这样。现在的情形充满了不确定，各种可能都有。现在发展太快了，我当然不想看到我们比现在差。但是如果我们不努力的话，比现在差是有可能的(雪宁：居安思危)对，**我觉得实际上是挺危的，而且我认为我们在应用方面一定要做出好的东西来**。在美国来讲的话，统计很有底蕴，应用很好；而在中国，我们的统计在理论上有一定的基础，但在应用方面还有很大的发展空间，我们需要回答统计在实际中到底多有用，有多大贡献这个问题。我想如果我们既有应用方面又有理论，那会非常强劲；但如果我们只是关起门来做理论的话，我不敢讲20年以后会是什么样。因为现在情况是非常动态的，好多人在有了数据以后，他们完全能够绕过你，自己产出结果，你能做的分析人家都能做。（哈哈）尽管实际上并不是这样，可表面上看是这样。像统计推断，一般人做不了，但是人家可能不做推断而只通过描述统计就能产出满意的结果，于是不想纠结那些东西。在这种情形下，我们可能就发不出声音来，我们就可能被遮蔽了。所以当前还是充满挑战的。任何一个时候一个学科都有挑战，但是我觉得现在可能挑战更大，就看你怎么处理这个事。

健桥：我觉得这些挑战也是一个学科富有希望和活力的标志。其实这就关系到后面的问题了，您觉得计算机、数据科学和统计之间应该是什么关系呢？

依我来看统计还是不同于计算机的，因为现在有些最顶尖的CS（编者注：computer science，即计算机科学）人在学统计，学数理统计，也往Annals上投文章、发文章。他们还是非常青睐统计的杂志。但是对统计方面来说，如果人家学会了我们统计推断，他们又天生计算能力很强，他们就很有优势。因为数据就是要计算，来做任何事情你都要能计算。如果你计算不出来，只是证明定理的话，我觉得还不到位。刚才我们说重要科学问题我们要介入进去，在里边有个声音，要发挥我们的影响，所以最后不落实到“算”不行。我觉得统计在计算方面，在实现这方面，应该好好设计一下，包括我们的课程和对学生的训练。

雪宁：咱们现在的课程设计好像计算方面不是很多的

健桥：一个感觉是说本科先不学统计，先学计算机，然后研究生学统计可能会更好，您觉得是吗？

我觉得没有一个一定的路径，只要是你能够把这方面给考虑到就可以。我感觉大家都不差，只要你意识到这个问题，就能够提高。现在的当务之急是要保持你的声音，统计推断是我们的看家本领，对吧？还有我们的抽样，我们众多的统计方法如Bootstrap，变量选择，降维技术等，如果我们再把计算，大型数据的分析给加强，我觉得统计还是很好的。但这需要我们付出辛苦努力，发展数据科学大家是互相补充的，大家一起来促进数据科学发展。统计是从一个方面来做贡献，CS从一方面做贡献，数学也从另外一方面做贡献。我觉得这是一个大的领域，大家都有机会，但是我们要有意识，要有意地对数据科学做贡献。

健桥：您觉得既然统计是给数据科学做出贡献，所以您认为数据科学也不会取代统计是么？

统计也不至于那么容易就被取代。数据科学，如果没有统计学也不行，我认为也都是互补的。

健桥：如果选择研究生，您会更注重他什么能力呢？

这个都在变化。因为这几年在做应用了，现在觉得要有非常扎实的数学能力 再加上有计算的能力。

健桥：可能之前比较注重数学，但是现在就越来越看重计算？

对，我现在感觉计算很重要，好多东西实际上还是看计算的。但相对来说，我可能还是认为数学更重要，因为计算能力相对来说学起来容易一些。**同时我还想强调一下中文写作应该提高，**我们做宣传写文稿，中文写作也挺重要的，当然现在写中文的机会越来越少，沟通都是微信了。中文写作宣传方面，统计之都就做得很好。对一般的同学，如果我要建议的话，我觉得要提高数学能力，计算能力还有中文写作能力。

编者注：2015年3月和2016年3月，陈松蹊带领的北大光华和北大统计科学中心团队先后发布了两份空气质量评估报告，《空气质量评估报告：北京城区2010-2014年PM2.5污染状况研究》和《空气质量评估报告（二）：中国五城市空气污染状况之统计学分析》。第二份报告一经发出，便引起了国内外媒体的关注，美国《纽约时报》网站和国内《参考消息》先后对其进行了报道。2016年3月17日《空气质量评估报告（三）：北京地区2013-2016年区域污染状况评估》在北京大学正式发布。

健桥：接下来想跟您讨论一下关于PM2.5项目的问题。虽然说统计能带我们进入各个学科的后花园，但就现在污染问题来说，大气物理，大气化学，统计都在做，我个人有时候觉得他们从大气机理出发，很科学，例如现在北京雾霾的预报都是用的人家模型，即使上午天气晴朗，也能准确预测下午雾霾；统计比较起来反而没那么“科学”，您怎么看呢？

目前有各种大气模式，各个国家都有发展，我们国内也有发展研发自己的模式，这些模型确实也很有效。

但是大气的变化，例如大气化学反应，充满着不确定性。因此这些模式的误差，实际上挺大的，尤其是在重污染的情况下。我前一段时间参加一个大气污染的研讨会，环保部有一个空气质量模型的达标的标准，它规定相对误差正负70%就达标了。从这你能看出大气模式还是有很大不确定性，这时候统计学就有很大的用武之地，例如最明显的是对空气质量的评估。我们团队现在做的是在去除气象因素影响后给出一个公平的空气质量水平，而不是完全受气象条件的干扰，这对空气质量管理非常重要。做大气模式不涉及这个，这却恰恰属于统计中的observation study（观测研究）研究内容。再比如现在最关心的是重污染的原因，里面实际上有很多的问题不是很清楚，在物理和化学层面上尚不明晰。往往就是这些尚未清楚的地方是造成误差最大的原因，所以仍有改进的空间，我觉得大气物理、大气化学跟统计一块能够对这个非常具有挑战性的问题提出有效的解决方案。

健桥：您之前说的一句话令我印象特别深刻，您说它们都是基于模式的，它们的模式不一定准，但我们是基于实证的，实证加模式会比使用模式要好。

就是对误差比较大的地方，需要通过统计的实证，得到一些wisdom（智慧）来促进它，比如给这个模式提供一个好的初始场，使它预报得更加准确。

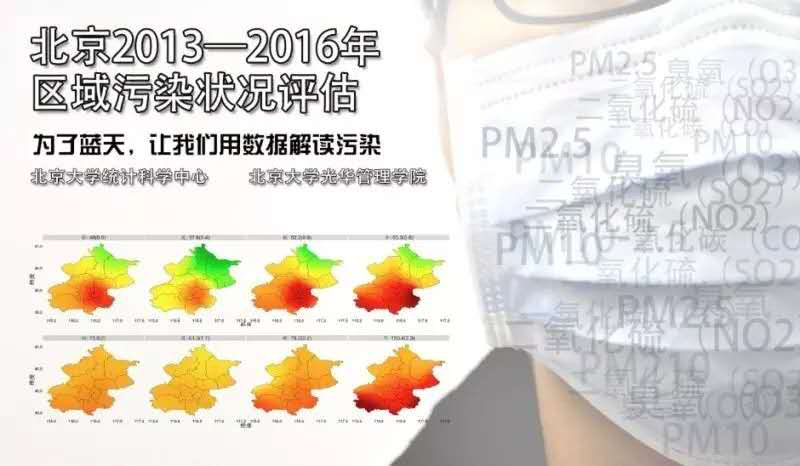
健桥：我很好奇，您这个项目的动机最初的动机是什么？

（这个项目开始的）原因最初是因为我邀请Tibshirani（编者注：斯坦福大学统计学教授）来访问北大，但他老犹豫，从2013年秋天开始谈，几个月都不能敲定这件事。后来我说你可以春天或夏天来，因为那时候北京的空气质量好，他说我非常高兴，你终于把这问题提出来了。他给我寄来一个（公布空气质量的）网站，说你看看这个。我以前也不太在意空气污染，但自从看了这个网站就开始想研究空气质量的课题，为中国的大气污染防治做点什么。所以14年2月底回到北京，恰逢两会期间，污染很严重。2014年3月3号憋了好几天以后我说Let’s do something（我们来做些什么吧）

这个项目另外一个间接原因是大数据对我的触动。从13年开始，随着大数据概念热起来，我参加了国内的一些活动，也给过报告，后来觉得还应该自己真正做一个实际应用的问题。

雪宁：那段时间大数据比较热，现在也热，但那个时候更热，

当时我觉得只谈大数据的概念有点虚，自己应该做一个实实在在的（研究），之前也在从事高维统计研究，但都在比较理论的层面，说同应用也有联系，但仍是两回事。所以正好有Tibishirani这样一个契机，也是为了自己的健康，也为了老百姓的健康，我是打算做一个应用的研究。当时我考虑对这个环保问题从统计的角度来给出一方面的意见和看法。我们能不能把这事儿分析地更全面一点，帮助把这问题彻底解决了。

（第三份空气质量评估报告封面）

健桥：您最近发了第三份报告，这个事您打算再做几年？

我们之前做的两份报告都没有资金支持。去年我们拿到了国家重点研发专项“大气污染成因防治损害评估研究”的一个项目。我们一个统计团队拿了环境方面的项目，主要是前面两个报告闯出牌子来了。我们发了两篇文章，一个在《英国皇家协会会刊》，一个在《地球物理研究杂志》，在环境圈子里得到他们的认可。大家也看到监测数据也需要研究分析，所以有这方面的需求，加上我们团队这方面基础也很好，所以项目就给我们了。我们这个项目现在快一年了，四年的项目，还有三年多一点。

健桥：我觉得Tibshirani这个事也可能是一个缩影，代表了一批国外科学家对是否来中国教书的这种顾虑，很多人其实都担忧这个空气问题吧。

对，当时我意识到这是一个很重要的问题，其实早前几年并不是很有这个意识。

健桥：谈到统计学在我国的情况，您也了解过统计之都这样一个统计学专业组织，您对于统计之都有着什么样的期待？

一开始了解统计之都是通过谢益辉，益辉当时是Iowa State的学生，他跟我介绍过这方面的事。我看着你们一点一点的成长起来，非常成功，而且我觉得你们为中国统计做了很多，像普及统计的基本知识以及统计应用，非常了不起。你们全是志愿者，从中你能看到中国统计的希望。我觉得你们很有热情，你们现在办的一些活动，论坛，微信，还有你们的R语言会议，非常了不起。我现在出去跟环保部门的人聊，发现他们都用R。前年在北大的R语言会议，汉生给我打电话说松蹊你得帮我跟学校沟通，邱德拔体育馆要得到校领导批准才能够使用，因为参加的人数太多了。你能看出这个会议的影响力，他不是说一个教室就能给搞定的，你需要一个体育馆，你需要安全部门的批准，动静很大，影响力很广。

雪宁：那年有4000来个人，人非常非常多。

健桥：我觉得这个会很体现统计和其他学科的交叉，我参加R会认识了来自很多其他学科的人。

对，我觉得这做得非常好。现在都说大数据，我觉得还是要做实事。我认为R会议真是把统计在大数据，在数据科学这方面的发展的领导力体现了出来。我觉得你们年轻人做了很多。我非常欣赏你们这个，我们需要做这种超前的事情。统计之都靠的是一拨人员干几年，之后下一拨人接着传承几年。我觉得这是为统计服务，也是为社会服务，统计之都做的很大气。

谢谢陈老师，还希望陈老师以后多多指导支持！