

## Statisticians at Work: Inspiration, Aspiration, Ambition

### 统计学者的工作及风范: 灵感、抱负、雄心 \*

C.F. Jeff Wu (吴建福)

(Georgia Institute of Technology)

作者简介: 吴建福教授生于台湾新竹, 1971年毕业于台湾大学数学系, 1976年在美国加州大学伯克利(Berkeley)分校获得统计学博士; 1977-1988年任威斯康辛大学麦迪森(Madison)分校教授, 1988-1993年任加拿大滑铁卢大学统计与精算学系的通用汽车及加拿大理工基金会讲座教授, 1993-2003年任密歇根大学H.C. Carver统计学讲座教授; 2003年起, 任佐治亚理工学院工业与系统工程系Coca Cola讲座教授。

吴建福教授于1987年获得COPSS奖(国际统计学四十一岁以下学者的最高成就奖), 2000年被选为台湾中研院院士, 2004年当选为美国国家工程院院士, 是第一位统计学家获此殊荣。此外, 他还获得过多项国际顶级奖励。学术界对他的评价是: “他的贡献始终是专业严格性与实际重要性的理想结合”; “吴和George Box是两位领袖, 他们为设计、分析和优化实验奠定了严谨的基础, 并发展了一套完整的方法”; 吴建福“对参数设计和质量改进作出了原创性贡献, 创建了一套现代实验设计体系, 培养了一代质量科学专业工作者和教授”。

吴建福教授是最早与我国统计界建立联系的海外学者之一, 为推动我国统计学的发展作出了杰出的贡献。改革开放以来, 应邀访问吴建福教授的我国学者达40余人, 他也先后数十次来华访问、讲学。特别是近几年, 在他的倡议和指导下, 中国科学院数学与系统科学研究院创立了我国第一个“质量科学研究中心”, 并与国家纳米科学中心成立了“纳米技术与统计科学交叉研究联合实验室”。吴教授亲自担任这两个“中心”的学术委员会海外主任, 为两个“中心”的建设和发展做了大量工作。

1985年5月我第一次到中国来, 距离这次我被邀请已经隔了25年。这次我该讲什么呢? 我是这么想的, 假如20年前的中国像个穷人, 现在的中国就像个富人, 或者快变成富人。那么对富人讲话应该跟对穷人不一样, 所以今天我想讲的题目是怎么看待学术成就和学术风范。为什么呢? 各位从历史上看, 一个国家要变成大国, 变成真正受人尊敬的大国, 或者一个科学界, 要衡量它的成熟以及它本身的品质, 其中一个很重要的衡量标准是它怎么对

\*第九届全国概率统计会议大会特邀报告, 2010年10月22日。

本文2010年10月22日收到。

待学术成就和学术风范. 这个问题我们可以从三种境界来看, 这三种境界用英文来说就是: Inspiration, Aspiration, Ambition. 自从我接受邀请后, 一直在想这个问题, 在我脑袋里大概几个月了. 其中我马上想到的是这三个英文单词, 它们的英文的押韵也很好, 但中文就很难翻译, 这里我翻译成: 灵感、抱负和雄心.

下面先解释一下这三种境界, 然后以统计学一百年来一些重要大人物的故事为例加以说明, 包括他们的重要创作, 以及他们之间的互相争执. 最后镜头再转向中国, 看一下中国过去的问题和成就, 以及对未来的建议.

Inspiration: 艺术、科学里称之为灵感, 宗教里则称之为启示.

Aspiration: 可以说是抱负、志向. 这个词可以比较实际一点, 但也有精神层面的含义.

Ambition: 比较难翻译, 好的含义是雄心、热望, 不好的含义就是野心、企图. 这个词假如有太多的目标性或功利性就不好.

第一个, 先讲Karl Pearson (1857–1936). 他是Biometrika (1901)的创始人, 在十九世纪末到二十世纪初, 在Fisher出现以前, 他是最重要的统计学家, 是统计应用于农业与生物学的奠基者. Pearson在1900提出了现在常用的Pearson chi-square, 拟合优度的 $\chi^2$ 检验. 但有意思的在于, 他认为: 假如用于一个 $r$ 列 $c$ 行的列联表作 $\chi^2$ 检验时, 它的自由度是 $rc - 1$ , 因为用了平均数, 导致用掉一个自由度.

下一个人物是R.A. Fisher (1890–1962). Fisher在1922年就已经很出名了. Fisher提出了最严格的自由度, 对自由度给出了一个几何的说法. 对于 $r \times c$ 的列联表, Fisher认为由于要估计 $r$ 列每一列的平均值, 所以要用到 $r - 1$ 个自由度, 同样 $c$ 行要用 $c - 1$ 个自由度, 这样它的自由度是 $(r - 1) \times (c - 1)$ , 这是正确的. 这时候Pearson就暴怒了, 他说“Fisher这个观点是完全错误的, 这个作者所做的事对统计科学一点好处都没有. 我相信我的批评者会原谅我把他跟唐吉歌德做比较, Fisher就像唐吉歌德拿着长矛要去打风车. 他这样会把自己毁掉, 或者他的整个理论会被毁掉” (1922年). 这是一个不好的ambition的例子.

Fisher也不是个好惹的人物. Fisher在1950年的论文集里写到Pearson的时候, 他说“一个人对自由的评论很容易愤怒是老昏的征象. Pearson在很年轻的时代就已经开始老昏了”. 英国并不是和东方一样学霸很厉害, 英国是一个自由讨论的地方, 教授的力量很大, 但是你可以反抗他, 反抗他你也有地方去. 从这几个例子就看出来, 其他的例子我没有时间讲. 英国是一个很奇怪的组合体, 它有一定的封建, 但却可以完全地自由讨论, 甚至所用的语言相当尖锐. 假如英国人用尖锐的语言批评你, 不要生气, 因为这是他们文化的一部分.

Fisher是最伟大的统计学家, 其他人跟他相比差得远. Fisher 1919年从剑桥毕业后到了Rothamsted农业试验站. 这是他的第一份工作, 当然Fisher得到了很好的推荐信. Fisher本科读数学, 也读生物, 但当时他对农业没有任何接触. 在那里, 有很多农业育种栽培的实验, 产生了很多的数据. Fisher在几年内就发明了两个大成果——方差分析和试验设计. 试验设计是跟分析、建模结合在一起的, 试验设计不只是布点选择. Fisher的试验设计有三个基本原则: 重复、分区组以及随机化, 其中最具有统计新意的是随机化. 我记得在我出

国前,曾在台湾当了两年兵,从1971年毕业到1973年在台湾军队里边,自己读点统计,以前没有读过.我当时印象很深的是,随机化的观念不是一般的数学观念,它是很独特的、有统计思想的观念.另外一个例子,也是Fisher的发明,所谓的似然函数.概率和统计最大的一个差别就是演绎跟推论.演绎是由假设推出一个结果;统计是看到现象,要去推断不知道的东西.所以同样一个分布密度函数在统计推断中是似然函数.统计上还有很多例子具有很独特的统计思想.因此统计学这一百年来变得这么壮大,有着自己一套很完整的知识遗产.所以要能做到大家,第一个要有灵感.但是你要把一个领域做得很大,还一定要有志向,一定要有能力把它做到底,有体力,能说能写,有门生,能够推广,能够应用,这不是一两把刷子.这是对对应到我们前面讲的Aspiration的例子.

Pearson还有个故事,同样是针对列联表的研究,关于如何用统计来估计相关性.由于列联表是离散的,他假设列联表的下方隐含一个连续的两维正态分布,然后再作截断,进行离散化来作计算. Pearson称此为四项相关(tetrachoric correlation),大概在1970年代人们会读到,但现在已经很少有人用了.很快,一个叫George U. Yule (1871–1951)的,发现这个做法不对,他发明了列联表领域很有名的比值比(odds ratio).就是说,我们看到的是离散的,比如下面 $2 \times 2$ 的列联表

|          |          |
|----------|----------|
| $n_{11}$ | $n_{12}$ |
| $n_{21}$ | $n_{22}$ |

这时就取 $\hat{\theta} = n_{11}n_{22}/(n_{12}n_{21})$ ,这是大家今天都在用的. Yule还算客气,他说“Pearson用正态分布给我们的只是一个假设性的相关系数,因为它根据的是你看不到的变数.用这个不必要而且没法验证的假设来做相关性的计算,看起来不是很好”.这本书在当时很有名,出了14版,是一个还不错的研究,是一个Aspiration的例子.

Pearson发脾气了,他说(1913年)“这个一点都没有思考过的赞美,会让很多统计的学生走入迷途”,“如果Yule先生的观点被接受的话,对近代统计学的发展会产生无法弥补的伤害.在我指导下的任何工作里, Yule先生的这个方法过去没用过,未来也永远不会用.假如Yule先生想在统计界有点名望的话,就一定要把他的观点撤回”.在1913年, Pearson还很厉害,当时Fisher还没起来的时候, Pearson是一个大学霸.可这个大学霸很快就被Fisher推翻了, Fisher比他厉害多了.许宝騄先生去英国读书时没有几个教授, Pearson是Professor of Statistics, Fisher是Professor of Genetics, 一个在二楼, 一个在三楼, 两边不讲话.

下一个是Jerzy Neyman (1894–1981),他是波兰人,也是很重要的人物.他到法国念的书,跟Émile Borel学过概率论, 1925年去伦敦大学的大学院,跟Karl Pearson学习. Neyman当时很年轻,我觉得Karl Pearson一开始并不看重他.从历史上看,他跟Karl Pearson的儿子Egon Pearson联手发明了现在很有名的Neyman-Pearson理论.这是假设检验的经典之作,我们一直在用这个东西.假设检验是Fisher提出来的, Karl Pearson这方面没做什么.但需要注意的是Fisher的假设检验没有最优理论.应该说Neyman是第一个把数

理统计的严格概念带进统计学的。不过有很多思想,甚至一些不算严格的数理统计的东西是Fisher做出来的,包括很多大样本理论。Neyman最早的一个观念是假设检验要提一个零假设和对立假设,两个互相对立以后,由相对比就可以得到Neyman-Pearson引理。这是数理统计中最重要的结果之一,因为它有最优性。这个结果证明很简单,几行话就证出来了。这是Neyman做的大事,把最优理论带进来了。后来Neyman自己一个人,很自然地假设检验去做区间估计、置信区间等等。Neyman-Pearson里面Neyman比Pearson更厉害。Egon Pearson是英国人, Neyman是移民,移民总是比较拼一点。而且Egon Pearson年轻时身体不好,他每年暑假要到苏格兰度假; Neyman就在伦敦继续努力写文章,就像我们中国人在美国很用功读书一样! 不过总的来讲, Neyman高明多了。后来Neyman和Fisher产生了争执,但没有全面的爆发。你知道点估计主要是Fisher做的,但是怎么做区间估计, Fisher提出了fiducial推断。这东西现在大家不太常用。Neyman要在二楼和三楼之间保持平衡,因为K. Pearson跟Fisher这两个都是不能惹的大人物, Neyman有办法保持平衡,但是久了也不行。所以1938年Neyman就搬到Berkley数学系去了,因为他在伦敦最多做到Reader或Senior Lecturer。那个时候教授就两个位置,除非一个死掉或者被暗杀,否则就没有教授的位置。一九五几年Berkley成立了统计系。所以Neyman跟Fisher一样都有和Inspiration、Aspiration对应的例子。

我是Berkeley毕业的,运气好,当学生时还看到Neyman两年多。他的生活习惯和一般人不一样,他又抽烟,又吃香肠,可是寿命很长,那个时代活到80多岁,我想他基因比较好。他每天中午都在lunch room吃饭,我当时发现这个伟人天天在那吃饭,平常也没有机会听他讲话,就每天带着我的东西去,还不敢带中国味道太重的东西,边吃饭边听他讲话。Le Cam也都会去。所以现在回想起那个时代很令人怀念,看到这么多大人物,以及他们的学术风范。

下一个是John W. Tukey (1915–2000)。我们先讲一下Jackknife。原来第一个提出来Jackknife的是英国人,叫M. Quenouille,他不是最有名的。早期做时间序列的会知道,他做multiple time series。我稍微提一下他的思想。对于一个参数的估计量 $\hat{\theta}_n = \theta(x_1, \dots, x_n)$ 。只要把数据中一个点 $x_i$ 拿掉:  $x_{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ , 就可以重新做估计 $\hat{\theta}_{(i)} = \hat{\theta}(x_{(i)})$ 。现在对每个数据点都这么做,得到的估计取平均值,记为 $\tilde{\theta} = (1/n) \sum \hat{\theta}_{(i)}$ 。一般情况下,  $\hat{\theta}_n$ 偏差的阶是 $E\hat{\theta}_n - \theta = O(1/n)$ 。因为 $\tilde{\theta}$ 和 $\hat{\theta}_n$ 偏差展开的第一项是一样的,记 $b_J = (n-1)(\tilde{\theta} - \hat{\theta}_n)$ , 就是 $\hat{\theta}_n$ 偏差的估计,它的尾项就是下一个阶,即 $E(b_J) = \text{bias}(\hat{\theta}_n) + O(1/n^2)$ 。Quenouille认为假如能够这样做,就把 $b_J$ 从原来的估计量中减掉,  $\hat{\theta}_J = \hat{\theta}_n - b_J = n\tilde{\theta}_n - (n-1)\tilde{\theta}$ , 叫做Quenouille点估计。这个东西不受重视。从历史来评断这个东西,他当初做这个问题是为了做时间序列。事后看,对Jackknife来讲,时间序列不是最好的例子。用于时间序列你必须拿掉一段,不能中间减掉一个点,要拿就拿前面一段,或后面一段。所以从统计上讲,它并不是一个很适当的setting。另外用这个方法时,把偏差减掉,往往方差会增加,因此这个方法不受重视。但是,精彩的是Tukey的眼光。



Tukey是我认为所有的统计学者中间的两位天才之一. Tukey和Fisher是做到常人难以想象的地步. Fisher是在遗传学与数理统计学领域; Tukey的博士论文是做拓扑学的, 但是他对统计学以及其他领域都有很大贡献, 比如信号处理、电机、计算机科学, 他都有基本贡献. 简单讲, Tukey看到Jackknife的文章后, 就把一阶变二阶. 定义 $\tilde{\theta}_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_{(i)}$ , 这里叫做Jackknife pseudo value. 假如是线性的估计量, pseudo value就是去掉第 $i$ 个数据, 假如是非线性的可以做投影. 记 $v_J = \{1/[n(n-1)]\} \sum_1^n (\tilde{\theta}_{n,i} - av.)^2$ , 其中 $av. = (1/n) \sum_1^n \tilde{\theta}_{n,i}$ . 显然 $v_J = [(n-1)/n] \sum_1^n (\hat{\theta}_{(i)} - \tilde{\theta})^2$ ,  $v_J$ 叫做Jackknife方差估计. Tukey做了两个猜测, 在数学上很含糊, 但这不重要, 因为他把原来一个没有人注意的东西变成了新的方向. 第一个猜测是 $\{\tilde{\theta}_{n,i}\}$ 是i.i.d.. 这个是错的, 但是没有人证明. 第二个是对的: 用pseudo value进行方差估计之后, 可以用来估计原来估计量的方差. 这是很重要的突破, 因为这个估计量的方差估计没有分布的假设, 而且还是在上世纪50年代发现的. 这是一个Inspiration的例子.

Tukey太有名了, 无所谓, 他的这篇文章从来没有发表, 就在1958年的Ann. Math. Stat. 29, 614页上发表个摘要, 他叫Preliminary Report, 后来也没有再发表了. 其实Tukey的数学远远比这个要好, 他不在乎. 最主要的是他在1958年就看到了可以用Jackknife做区间估计, 而且不要分布的假设. 如果了解Tukey, 就会知道他所做的事情远远超越他的时代, 这就是为什么他是天才. 当时那个时代的人还听不懂. Tukey上面的研究为什么这么重要? 因为第一, 他是第一个看到, 并提出用Jackknife作为再取样的推断方法; 第二个是我半开玩笑的, 他这篇文章只有一页, 影响力页数比(impact pages ratio)很大. 当然, 最重要的是他激励了Bradley Efron的突破, 就是bootstrap.

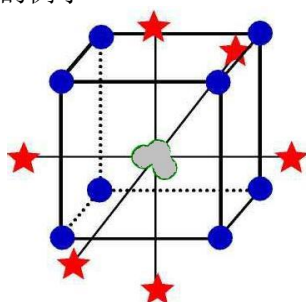
Efron在1979年发表的文章题目叫做“Bootstrap methods: another look at the jackknife”. 这个时代是我亲身经历的, 听到了Efron的报告. Efron的想法我了解, 我也曾问过他, 他是想从Jackknife中发现一个比Jackknife更好的方法. 在第一篇文章里他比较谦虚, 各位注意这篇虽然在Annal上发表, 可是没有太多严格的证明. 往下Efron的文章再提Jackknife, 因为bootstrap比较好, 事实也如此. 什么是bootstrap? 把我们的数据 $x_1, \dots, x_n$ 想成一个母体, 从这个数据中进行可放回的i.i.d.取样, 每取出一个bootstrap样本, 就重新去算估计量, 多次后计算 $v_{\text{boot}} = \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}^*)^2 / B$ , 就是Bootstrap方差估计. 最主要的在于: 它是基于数据的, 但是数据的分布是一个未知的 $F$ . 因为 $F$ 未知, 所以就用经验分布, 是来自于数据的, 记为 $\hat{F}$ . 大家想,  $x_1, \dots, x_n$ 是从 $F$ 中取样的, 我们为了猜测或是逼近它, 就用经验分布 $\hat{F}$ 当作母体对它取样, 这个取样就叫做bootstrap. 这样看来, bootstrap很简单自然, 但之前却没人想到. 这就是大家, 他能想到别人想不到的. Efron在2005年得到美国总统的科学奖章, 这个奖章在统计界特别难拿, 我看到某些领域拿到的人并不如Efron. 对Efron的表彰就提两件事情, 一个是bootstrap, 一个是他用几何的观念来看统计估计量的性质, 这是两大深刻的东西.

回到Tukey, 他有个贡献, 统计上可能不太知道. 60年代作信号处理, 他做出了电机工

程里的大问题、大贡献,就是所谓的Fast Fourier Transform (FFT, 快速傅氏变换). 在各种科学领域里面,不管是工程还是自然、社会科学都要做傅氏变换. 应用傅氏变换要离散化去做计算,而其中最大的一个问题就是怎么样算得快. 他在1965年与Cooley合作的文章“An algorithm for the machine calculation of complex Fourier Series”中提出一种算法,简单的讲就是,这个计算可以想成一个 $N \times N$ 矩阵乘以一个 $N$ 维向量. Tukey注意到,同样这个问题在试验设计领域中早已有研究,第一个做这个的就是Frank Yates. Yates原先是Fisher的助手,是大师级的人物,30年代的析因(factorial)试验是他发明的. 他提了一个Yates算法,早期教试验设计要同步计算主作用和各阶交互作用. Yates发明这个算法是在1930年代,那个时候只有手摇的机器,所以计算方法一定要最精简. Yates算法从今天的计算机的算法复杂度讲是最优,就是次数会更少. 做法是,比如数据是16个,怎么拆? 从小到大. 简单的讲,先作2再做4、做8再做16. Tukey注意到Yates的算法. Yates是针对二阶的析因试验,后来Box把它推到三阶,最后是I.J. Good推到素数幂. 信号处理跟试验设计的差别在于,它的理论受到应用环境的影响,就可以作出新东西. 在这个例子里,传统的试验设计,做试验很少能超过64次. 但是到了上世纪60年代作信号处理,傅利叶序列的 $n$ 是几千个,这时候最优的算法变得非常重要,否则根本没有办法计算. 也就是说,没有Tukey的Fast Fourier Transform这东西基本上是不能算的. 所以在信号处理,这是个基本突破. 有意思的是,这个基本突破用的工具是试验设计三十年前Yates提出来的,这个历史是很精彩的. 所以这是一个inspiration的例子.

再说George E.P. Box (1919–). 我Berkeley毕业后就去了威斯康辛,跟他共事. 伟大的人物也不是好惹的,他们有棱有角. Box有很多贡献,我们这里讲一件事. 他原来是读化学跟应用数学,二战时在毒气实验室做事,他根本不懂统计学. 他的上级跟他说,你懂数学,你要变成统计学家,今天就给我做东西. 他说我不懂怎么办? 他上级说很简单, Fisher有一大堆的书跟文章,你就拿去用. 所以Box跟我说,他就每天晚上看Fisher的东西,然后第二天现炒现卖,做实际工作. 我认为这是学统计最棒的方式. 而我们现在花很多时间一页一页地读,看到的数据都是别人的数据,为了在文献发表文章抄个数据. Box在不懂的时候同时做,很有意思. 战后他到帝国化学工业(Imperial Chemical Industries Ltd, ICI)工作. 有一个插曲,二十几年前Box是Fisher的女婿. Box到了帝国化学工业注意到一件事,厉害的人能看人所未看,见人所未见. 同样的东西在你眼前经过你就看不到,只能用传统的方法;只有有洞察力的人才会发现,这就是差别. 包括我带学生,有的学生有洞察力,有的学生只会套,没办法,这是上天给的. 在他发明反应曲面方法(response surface methodology)以前,他发现传统的都是用Yates和Finney的析因设计做最优选择. 析因设计是用于农业中的育种栽培,是离散的;他现在做的是化工过程的过程优化. 所以他注意到在过程的选择优化中,很多变量是定量的,回归模型比较适合;而相对地,在农业实验里很多的变量是定性,用的是方差分析. 这样就引出了他的突破和新典范. 其中一个就是他提出的中心复合设计,它不是一般的析因设计. 我就用这个很简单的例子说明,当然Box做了很多其他贡献,包括时间序列. 下

面是一个最简单的中心复合设计的例子



它有15到17个点. 你看, 角落深色的有8个点, 可以估计一次模型; 三个轴上各加两个☆点, 一共6个点, 可以估计二次的模型; 中心要放几个点, 可以估计误差方差. 这是很简单的想法, 可在上世纪50年代这是很重要的. 相比之下, 可以看到, 假如是做 $3^3$ 的析因设计, 则需要27个点. 当然你可以估计比较多的东西, 比如说一些三阶的交互作用, 但是没有用. Box的反应曲面理论实际上有三部分, 第一部分是怎么选设计, 这个东西后来对Jack Carl Kiefer (1924–1981)影响很大. 这也是个大人物. 他58岁就去世了, 是在Berkeley游泳, 出来洗热水澡, 心脏病发作就去世了. 他曾是我的老师, 在康奈尔任教. 我研究生三年级上半年的时候他在Berkeley呆了十个礼拜. 受他的影响, 我才选了最优设计的题目做. 我偷偷做完了, 写成论文才敢给老师Peter J. Bickel看. 老师看了吓了一跳, 说让你毕业, 然后他替我改英文就让我毕业. 所以我对Kiefer很熟, 他的学术风范很伟大. Kiefer的最优设计实际上受到Box一定的影响. 但Box常批评Kiefer, 这个可以说是一个ambition的例子. 故事很多, 时间限制就不说了. 但我要强调的是, 试验设计方法一般是个三部曲. 第一是如何找设计点, 第二是如何建模、分析. Box的建模用回归, 这个倒不难. 说个题外话, 在国内有个术语用得很好, 国内把这一类的设计, 包括中心复合设计, 叫做回归设计, 这个词用得很好. 英文没有regression design这个叫法, 它就一个一个设计单独取名. 但我觉得regression design这个叫法特别好, 它把这一类连续型的设计都包含在内. 第三是优化. 所以我说试验设计不只是布点. 再说个插曲, 虽然这是Box最伟大的贡献之一, 但这并不是Box在伦敦大学学院的博士论文, 因为他发现这个东西在当时没有数学; 当然后来发展得很完整. 他原来的工作不是很完整, 数学不够, 要拿博士是不行的, 所以他写了另外一篇文章发表在Biometrika上, 是有关likelihood ratio test的, 有点小名气. 这个历史典故也很有意思.

接下来, 讲我自己的工作. 试验设计这一百年来, 可以用下面的图来综合描述, 这是我做的分类, 共有四个分枝:

- (1) Regular orthogonal designs (Fisher, Yates, Finney, ...): 如 $2^{n-k}$ ,  $3^{n-k}$ 设计, 使用最小低阶混杂准则(minimum aberration criterion)
- (2) Nonregular orthogonal designs (Plackett-Burman, Rao, Bose): 如Plackett-Burman设计, 正交表
- (3) Response surface designs (Box): 拟合参数响应曲面

#### (4) Optimal designs (Kiefer): 由特定的模型或准则驱动的最优性

最早的就是二水平、三水平之类的设计,我称之为规则性的正交设计.在最近一、二十年,试验设计的最优性,叫做minimum aberration criterion.我跟我的学生以及合作者在这方面做得很多.另外是非规则性的正交设计,在构造方面有很多工作,第一个是Plackett-Burman. C.R. Rao定义了最广的一类所谓的正交表. C.R. Rao定义正交表时只有28岁.它的构造中很多很难的数学是R.C. Bose做的. Bose后来也搬到美国,他跟Rao都是美国的院士. Bose死得比较早,他在编码理论上有很大贡献,是离散设计、组合设计跟编码理论的大家;不过对这些成果在试验设计中怎么使用没有发展.目前主要是我跟我的学生在近十五年来做的一些工作.

简单地讲,我对这种设计有兴趣是这么来的:1986年我们在威斯康辛,在Box领导之下,跟贝尔实验室从国家科学基金会(NSF)得到一个很大的项目.当时美国的电子业和汽车业都被日本打败了.当时的日本跟今天的中国一样,外汇储备最多,美国逼它升值,日本听话就升值了,后来就不行了.日本当时的品质管理做得非常好,我们听说日本有个田口玄一,以及其他品质管理大师,所以我们就去访问日本了,大概一共有7个人.我注意到一件事,有一天下午,在日本名古屋的中部日本质量学会(Central Japan Quality Association)听报告,他们日本人给我们很多的案例研究,我注意到他们用的很多是国内叫做 $L_{12}$ 、 $L_{18}$ 和 $L_{36}$ 的正交表.试验次数都不是2的幂或3的幂,不是我说的规则性的正交设计,是非规则性的正交设计.我注意到他们画出来的数据,日本人是很有机密的,数据不拿出来,但是我记忆力好,我看他们做的东西,注意到很多都是很复杂的曲面.传统上这一类正交表只能估计主作用,可是他们最后验证都成功,也就是说,它是一个很复杂的响应曲面.我当时想,有个思想很直观,但没有严格证明,就是说这个非规则性设计后面肯定有深意,否则不可能这么成功.因为它的试验次数很少,传统的典范(paradigm)不能用.简单地讲,我们提出的新典范认为,这种设计有很复杂的别名(alias),尽管传统上只能用来筛选主作用,但应该可以估计交互作用.这只是我当时的猜想,后来我跟M. Hamada在实际数据的分析中得到了验证.再讲一个偶然的发展,有一个原则是我们两个提出来的,叫做效应遗传原则,这个原则是我们数据分析成功的原因. Hamada是我的学生的时候,他第一次做这个问题,发现电脑可以找到的模型很多,可分析出来的结果却互相不一致,当时就很混乱.我们都搬到加拿大以后,有一天我跟他说,我们是不是应该加个限制.有了这条新原则,很多原来的模型经过这个原则就被排除掉了.第二天早上我看到他笑咪咪的,我知道有名堂了,他告诉我“好结果出来了”.比如说,有一组工业上的实际数据,原来只能分析两个主作用,  $R^2 = 0.36$ ;但是用新方法分析出来找到了完全不一样的因子,包括它的交互作用,  $R^2$ 变成了0.9,只有三项.所以这是个比较重要的成果,这当然引起了后面很多重要的工作,这里就不讲了.

我花这么多时间准备演讲,是有目的的,希望看到中国做得越来越好.中国的统计在过去的四十年里的发展算是比较有限的.究其原因,我想第一个是传统比较弱,因为早期概率统计是和数学在一起,很多数学的老先生是大家,许宝騄先生回国后身体不好,又去世的



比较早, 所以吃了很多亏. 这个老生常谈我就不多说. 另一个是统计一直是二级学科, 在数学里边, 很多统计系都在数学学院里边, 所以很多数学的判断准则用到统计的职升、评奖上, 这个有时候会有负面影响. 理论的工作比较容易受重视. 当然, 怎么评价好的应用工作是很困难的, 大家没有共识. 但我要强调的是, 这是个非常大的国家, 这个国家跟欧洲一样大, 有多少好的问题, 但是都被我们给浪费掉了, 我对此很感遗憾. 我们没有在这几十年把握这么多的机会. 大家都知道, 过去二十年来中国有很大的进展, 千人、长江, 外界的交流, 而且海外的华人做得非常好, 已经成为一股势力, 也常回国帮忙, 这是非常好的事. 有很多指标上来看都是比较好的, 比如SCI的文章越来越多了, 顶级杂志的文章数量也增加了. 上世纪80年代我来的时候比较少, 不过这应该是很自然的事情, 因为年轻一代的英文比较好, 但也可能不是好事. 我还听说, 好像在用SCI来做评审指标. 这个可以用, 但不要用得太多, 否则就容易走火入魔. 而且整个来讲, 包括国内和海外的华人, 我总觉得是ambition比较多, inspiration比较少, 这是一种感触. 统计基本上可以分成三块: 理论、方法、应用. 理论呢, 我的分法不太一样, 理论里面可以分成两类: 技术性的和非技术性的. 一般所谓的数理统计是技术性的, 但有的理论统计没有用很多的数学推导, 却是关键性的东西. 比如早期的点估计, 一个很重要是C.R. Rao的结果, 几行就出来了. 所以重要的是你要先想到, 不是想到证明, 而是想到那个定理, 想到结构. 毕竟统计学不太一样, 统计学需要很多思维性的东西. 界面交叉的东西是最精彩的, 在中国这个方面没有充分利用起来, 而且在目前这个系统、制度之下可能不是最被鼓励的. 我听说中国把统计学术刊物排名, 有个叫四大天王, 这不是杭州灵隐寺那四大天王, 是Ann. Stat., Biometrika, J. Am. Stat. Assoc, J. Roy. Stat. Soc. B. 当然这些是很好的, 但是还有很多好的. 我举个例子, 假如是在特别的领域里面, 比如Technometrics, 工业统计里面大部分的突破在这上面发表, minimum aberration, resolution, fractional factorial, latin hypercube都是在Technometrics上发表的. 在Ann. Stat.上的都是已经看到突破后再做一些构造, 所以请大家不要这么表面化. 而且现在越来越多好的领域, 比如计算及生物里也有很多好的杂志. 所以我认为这样做有负面作用. 这个做法比较表面, 其中最大的原因我猜是这样, 假如你自己没有能力判断的时候, 用一个指标最容易. 但科学大国不会有这种事情, 它只看文章的内容好坏. 我认为要鼓励中国统计发展的话, 要像毛主席说的, 矫枉必须过正, 对于应用工作应该给予更多照顾. 应用统计学者在中国的实际工作中有很多贡献, 当然也有做方法的, 可是很少是这两个结合在一起, 成为一个体系. 可能过去没有这样一个传统, 没有这样一个模式让大家去模仿. 另外, 我猜想是做应用统计的人没有用心去写出文章来发表. 不过这里牵扯到一个问题, 就是应用统计的英文文章比较难写. 你研究的内容可能不错, 但英文不好, 第一关就被退稿了, 他看不懂就把你宰掉了. 由于这个语言的困难度, 它的门槛比较高一点.

我再来讲数理统计在中国. 我们今天在场的大部分是本科做数学出身的, 所以进入统计学领域做数理统计当然比较自然, 而且语言障碍少. 中国的数理统计从建国开始就有好成果, 虽然动乱时代成果少一些, 但一直有好成果, 现在也是, 这是非常好的. 但是

我注意到, 中国的数理统计, 尤其改革开放以后, 基本上是大样本理论占最大优势. 大样本理论作为优势应该保留. 但是我想解释一下, 大样本有好的, 也有不是特别好的; 有的有意思, 有的就不是很有意思. 接下来的内容对我来说很难讲, 但是这一次能有这个机会, 心里想什么就说什么. 我们先讲好的大样本, 我把它分成四类. 第一类是结果大概猜得到, 但是证明非常困难, 就是连大家都做不出来. 最好的例子大概就是C. Stone 1975年做的fully efficient adaptive estimation. 研究一个位置参数估计的最优性时, 假如总体分布的形式已知, 由Fisher information确定了估计量渐近方差的下限, 达到下限就最优; 但是假如总体分布完全未知, 必须用数据同时去估计分布和参数, 两步一起做, 是不是最后还得到full efficiency, 这是个非常难的问题. 我记得Peter J. Bickel给我的第一个问题是要把1975年Stone的文章推到回归, 但我推不出来. Stone在1980年代就当选为美国科学院院士, 就是凭借这个东西. 第二类是数学不是太难, 但是结果非常惊人. 关于这方面早期或者说历史上第一个最惊人的是Stein, 他的一个大发现, 叫James-Stein估计量. 传统上认为正态分布的样本均值不管对几维都是可容许的, 因为它是minimax估计. 但是Stein和James这两个人(说句题外话, James是Stein的学生, 没有名气, 他们俩合作, Stein按照拼音排序把他放在前面, 这又是一个学术风范的好例子), 他们最精彩的在于发现三维以上样本均值是不可容许的, 而且有一个比它更好的估计量, 就是Stein估计量. 所用的数学也很有意思. Stein估计量原来的证明也很巧, Stein的数学一向非常好. 过了几十年后, 他想到一个非常精彩的证明, 他发现了一个叫integration-by-parts的引理, 利用这个引理, 原来Stein的那个结果几行就做出来了. 而且这引理可以用在很多地方, 比如Poisson估计, 甚至用到估计之外的领域. 第一个用到点估计以外是李克昭(Li, Ker-Chau)做的. 他当时在普渡大学, 是助理教授, 他把这个引理用到了cross-validation, 是一个很精彩的例子. 第三类是很clever也很elegant, 比如Le Cam's Contiguity Lemma, 用来算一个检验的功效. 最后一类是技术上非常困难, 实际上又有贡献. 比如这几年研究红火的高维统计量, 有很多大师进入这个领域, 比如Bickel, Donoho, Johnstone; 国内出去的有很多也做得非常好.

但是也有一些大样本的研究我觉得不是很精彩. 第一类就是J. Friedman取名的Asymptopia, 他是把asymptotics和utopia这两个单词合起来, 所以我自己大胆地翻成“大样本的乌托邦”. 我准备演讲的时候问了Friedman, 因为我是在15年前他给的演讲里听说的, 最近很少听到这个词了. 他当时是指一个大样本的理论结果跟它的有限样本性质没有关系. 当然, 他说这话带点玩笑, 因为他本身是非常应用的. Friedman是分类树的发明者, 还有很多应用方法是他发明的. 但我觉得这个东西不是完全负面的. 比如说, 假如大样本不相合的话, 往往会让你对这个方法有一些新的看法. 一个很有名的例子是AIC (赤池准则). 当时一个日本人叫柴田, 他证明了这个准则的大样本在一个框架下不相合; 但是这个准则在小样本下往往还是不错的. 这种不一致性引出了很多人在模型选择理论上的其它方法的研究, 所以这是很正面的. 另外一类我取名为Gloried Taylor Series Expansions, 就是说, 这个东西假如不严格去推导, 做泰勒级数展开, 把尾项丢掉, 就可以得到我们想要的结

果. 比如说Biometrika就是这个style. 所谓的gloried是指很严格的把所有的过程补上去, 比如Annals上有很多的这样文章. 但有一点请各位注意, 去年的Annals of Statistics有4312页, 有多少页是重要的? 这是值得思考的. Berkeley的Terry Speed在IMS Bulletin, December, 2009上的Terence's Stuff "You want a proof?"中发表了一些比较轻松的言论. 我大概翻译几句, 他说“假如你摆一大堆假设的话, 证明就很容易. 但假如要检验你的假设比直接检验你的结论更难的话, 那你就参加了一个‘伟大’的传统”. 他这是半开玩笑的. 他的下一段比较认真, “假如你没办法证明的话, 你可以用模拟去检验它的结论, 这就是用模拟来证明”. 第三段是说, “统计学里我们需要的是推导而不是证明, 也就是说, 从某个假设到公式有个推导过程. 这个过程给我们一些洞察力, 这个过程可能是用数学方法(但不一定是使用 $(\epsilon, \delta)$ ), 可能是用模拟, 也可能只是描述性的, 我之为统计学家的证明”.

下面我有一个建议, 三个并重. 建议是“多点灵感、抱负, 少点野心、企图”. 几个月前, 严加安院士寄给我他的一篇文章, 我想起他有一句话说得非常有条理, 所谓的“大道至简, 大美天成”, 跟这个意思是很相关的. 三个并重是“成就与风范并重”, 当然我们要看重成就; “理论与应用并重”; “国内(中文)与国外(英文)并重”. 我注意到在国内举办的国际会议里, 尤其是应用统计, 英文的报告国内听众不一定听得懂, 尤其是讲得快的时候. 因此是不是有些资源应该放到国内的中文会议及杂志上?

最后是我对年轻研究者的三个建议: 第一个, “要质, 不要量”. 严院士说的一句古文我觉得非常好, “厚积薄发”. 第二个, “要本质, 不要标签”. 我知道这几年国内年轻人喜欢看标签, 访问学者的排名谁高谁低, 头衔谁多谁少, 你需要看本质. 文章不光看是哪个杂志, 而要看文章的内容是什么. 要做科学大国就要看本质. 第三个比较困难, “有所为, 有所不为”, 浅显的意思就是不好的东西不做, 这才能进到一个科学大国的最高境界. 好, 我讲到这里, 谢谢!

(中国科学院数学与系统科学研究院姜宁宁、李国英根据录音整理, 经吴建福教授核准)