

物理学咬文嚼字之九十二 城邦与统计

曹则贤[†]

(中国科学院物理研究所 北京 100190)

2017-10-23 收到

[†] email: zxcao@iphy.ac.cn

DOI: 10.7693/wl20171111

你没到 States 去过罢!

——钱钟书《围城》

蒋介石：河南到底死了多少人？

大公报主编：政府统计 1062 人。

蒋介石：实际呢？

——电影《1942》¹⁾

摘要 Statistics 一词源自 stand (state, status), 是 science of states, 其发展得力于 state 管理事务的需求。统计源远流长、博大精深。统计物理让物理学有了新的面目和新的哲学。

1 引子

《武林外传》有一集，捕快燕小六见到陌生人就逼问：“姓嘛，叫嘛，从哪来，到哪去，家里几口人，人均几亩地，地里几头牛，说说说说！”²⁾ 请注意，1. 这里官家关切的是一些有价值的量(的分布)；2. 这是官家的急切行为——官家对百姓的人力、物力和隐私是非常严肃对待的，而那绝非仅仅出于好奇。官家，量的分布，这就引出了一个重要的科学概念：statistics。Statistics，汉译统计，译名相较于原文，丢失了很多关键的内容。Statistics，源自 stand (state, status)，西人从字面上就知道 statistics 是与 state (城邦、国家)有关的事务，

statistical physics，统计物理学，自然与体系的 state (状态) 有关。

2 State and stand

古希腊是由数百个城邦 (city-state) 组成的，其中著名的有雅典、斯巴达、柯林斯等，希腊人很为这个城邦历史而骄傲。City，城市，希腊语为 πόλις (polis)，由这个词而来的是 politics，城市事务，政治之谓也。State, status, 来自拉丁语 stare, to stand, 立，竖立，估计与柱子或者围墙有关，希腊语为 κατάσταση (katasiasi)。一个希腊城邦，要有市场和神庙(图 1)，可能还需要围墙。城邦这种结构应该是人类社会发展的一个阶段，小亚细亚有(图 2)，我们中国也有，相当于大

一点的圩子、寨子、关隘。一定规模的 state 里面就有组织和统治的问题，当代希腊语为 κράτος (kratos)，类似我国历史上周朝的诸侯国。诸侯国意义上的 state，德语词为 Staat。States 结成联盟，可以构成更大的国，美国是合众国，德国是联邦国，意思是一样的。

State，本意是 to stand，字面是这个意思的英文字还常见，比如前列腺，就是 prostate，pro (前)+state (立)。站立的相比跑动的，就是处于静态的，所以 static 有静态的意思。Hydrostatics，流体静力学，研究静止不动的流体的特征；electrostatics，静电学，研究静止不动电荷的势和相互之间的作用。静止流体还好理解，静止电荷可是不存在的事儿，但它反而被当成了电磁学的出发

1) 改编自刘震云的小说《温故一九四二》。

2) 到了如今，统计更加全面。客人住店，调查问题就劈头盖脸而来：“姓名？单位？老婆是不是二婚，孩子是不是你的？”——引自郭德刚相声。



图1 古希腊城邦之一柯林斯, Κόρινθος, 的遗址

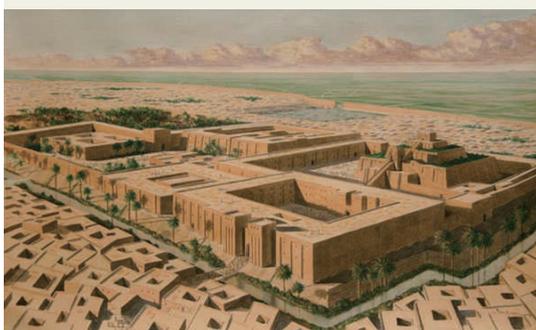


图2 苏美尔人的城

点——物理学抽象的 power, 就在这里。

State 还有状态的意思, 所谓物质的状态, 气、液、固和等离子体, 就是 states of matter。Status, 是 state 的拉丁语前身, 有状态 (condition) 的意思, 拉丁语的固体就是 status solidi。Status 有 (法律) 地位、身份等引申义, status quo 就是现状的意思。如果从中文学英语, 对 state 作为站立 (stand) 怎么会是状态的意思可能有迟疑, 但如果知道 state 对应的德语词是 Bestand, 应能消解部分怀疑。在距离 (distance, Abstand, 即分开站)、理解 (understand, verstehen, 即立于其下) 等词汇中, 都能看到 state 的身影。

3 城邦与统计

城邦要对其人民和人民的财富有明确的了解, 以利于管理³⁾和统治的延续。所谓 state, 就是一小部分人对他人之劳动成果甚至个人作为生殖资源可以剥夺、占有的势力范围 (马克思曾有差不多的表述)。亚历山大大帝在他的所有远征中都随军带着工程师、地理学家和测量师, 他们绘制被征服国家的地图, 记下这些国家的资源——可占有的资源才是 state 所关切的。由此, state 就衍生了一个重要的事务, sta-

tistics (统计)。在早期, statistics 的意义严格限制在关于城邦的信息上 (information about states), 其形容词形式 statistic, 希腊语, στατιστικός, 就是“国家的”意思。统计从来都是城邦层面上的事务。Statistics 一词来自拉丁语的 statisticum collegium (城邦委员会), 最后阶段应该是意大利语词汇 statista, 本意是官家的发言人或市民领袖 politician⁴⁾。德语的 Statistik, 由 Gottfried Achenwall (1749) 引入, 那时候已是分析国家数据的意思, 但强调那是忙公家事的科学 (science of state)。Statistics 经过 Statistik 传入英语是通过 1770 年的 *Bielfield's elementary universal education* 一书。该书说: 被称为统计的这门科

学告诉我们, 在已经探知的世界中, 所有现代国家的政治是如何组成的。一说是 1791 年, Sir John Sinclair 出版的 *Statistical Account of Scotland* 21 卷本的第一卷。19 世纪时, 统计的意思是数据的分析和分类, 后来被扩展到 “all collections of information of all types”。当然如今的意义既有数据搜集, 还有数据的分析以抽取信息 (statistical inference)。但是, 我们应该铭记 statistics 的本意, 按照 Webster 的 American dictionary 对统计的定义, 它是 “有关社会状态……以及国家的状态等事实的集合。”

统计一开始就是记账 (registration), 两千多年的时间里, 统计涉及的也就是可征用的人力和物力的清单。简单的记账后来发展成了统计这门学科, 赌博行为功不可没。赌博技术的发展导致了现代统计的开端 (The development of gambling techniques led to the beginning of modern statistics)^[1], 这已成定论。提起统计和概率论, 总让人想到骰子的形象。骰子的发明至今已有五千年的历史, 是罗马人将骰子从东方带到了西方。不过, 早在骰子这种高科技被发明之前, 人们就已经学会了用其它物件比如石子、小树枝来试运气了。

统计事务历来是城邦的大事。上世纪的国民政府下属两个著名的统计局, 分别为国民政府军事委员会调查统计局, 简称军统, 和国民党中央执行委员会调查统计局, 简称中统。中统和军统是凌驾于国民党党政军机关之上的特务组织。如今人们常在电视剧中见到抗战时期

3) 所谓管理, 是社会运行方程里的耗散项, 如果管理得太起劲以至于财富断了流, 就很尴尬了。中国历史上有 “竭泽而渔” 的说法, 可见这事不是虚构的。

4) Statesman 不仅是发言人, 他首先是公家的人。这样的人, 有时说点过头话 (overstatement) 也可以理解。

中统、军统的工作人员从事窃听、跟踪、调查、暗杀、爆破等活动，统计工作的重要性与复杂性由此可见一斑。一个 state，要对区域内的存在做出完整的 statistics，在没有监控系统、没有卫星通讯、没有计算机的年代几乎是不可能的事情。尤其是历史上的中国，幅员辽阔，地大物博，统计数据这样的混合物其可信度极低。一个典型的例子是战争人员统计。为什么先秦动辄都是几十万人的会战，明清时期人口多了却缩水成了几万人的规模？一个明显的原因是先秦官方有意无意地扩大统计口径，给后人留下了巨大的想象空间，难免会幻想出一幅气势磅礴、史诗般的战争画面。至于负面的内容，比如大饥荒造成的人口减少，则能遮掩就遮掩，哪里会认真统计。城邦处理数据时，很多时候又不能真按照统计学的规矩来，算是对这个词的讽刺。其实，这事也没啥好誉议的。历史总是被筛选和被遗忘的。谁是执掌筛选粗眼大筐的人呢？（刘震云《温故一九四二》）。统计从来都是大而化之的。有趣的是，统计物理也自然走上了粗粒化 (coarse-graining) 的途径，实在是由统计的本性决定的。

统计后来慢慢地变成了科学研究工具，有了点科学的气息。统计用于科学的一个看似浅显的例子是孟德尔 (Gregor Johann Mendel, 1822—1884) 的遗传规律。豌豆遗传的规律从数学的角度看是排列组合，这有别于黑色加白色等于灰色的简单混合 (图 3)。这是孟德尔遗传规律之价值所在。据信 Harold Hotelling (1895—1973) 的 statistics

的工作是导致纤维从理论的因素之一，其它的因素包括微分形式、拓扑、全局微分几何和联络理论^[2]，统计之内涵深厚由此可见一斑。按照薛定谔的说法，统计是我们时代非同寻常的发明 (……extraordinary invention of our time which goes by the name of statistics.)^[3]。

正确的统计不只对科学来说是好的，它还是最基本的。统计后来还进入了物理基础层面，但如何进行正确的统计却不是容易学得会的，坏的统计是科学的大麻烦^[4]。不幸的是，坏统计随处可见，甚至有人故意为之。统计的基本假设是大数目样本的存在，但是某些事件只有屈指可数的甚至单一的样本，也是没办法的事情。宇宙线研究依赖对天外粒子的偶然探测 (chance detection)，其结果很难复制，甚至原则上就没有机会复制，也就没有统计。然而一些研究者却硬是能编造出多宇宙的图景来作统计⁵⁾，或者算出方差或者置信度之类炫人眼目的数据来。统计对象的缺少样本、不可重复性以及采集的数据与研究对象的物理不相关让基于数据的所谓科学研究常常得出非常不科学的结论，基于数据自身开发的 p-value,

confidence (置信度) 计算其实并不能为别人带来 confidence (信心)⁵⁾——基于错误数据计算而来的置信度并不能赋予数据以科学性。由于无头绪的数据出发的统计，会产生一些源自这方法学

的所谓规律的发现，这些已经引起了严肃科学家的关注^[6]。薛定谔就直白地指出：“统计的方法因为缺少数学和逻辑的训练而名声不佳 (This method is discredited by the lack of mathematical and logical training).” 宇宙不是谁家的实验室，数据处理也不是有效的科学榨汁机。对于那些独家拥有少量几个数据的科学发现，保持一份谨慎是起码的科学精神。

4 统计物理

Robert Brown 是个植物学家、使用显微镜的专家。他的一个不懈努力追求的生活目标，是通过观察来发现生命力 (vis viva)⁶⁾ 的源泉。1827 年 6 月的某一天，他观察到花粉崩裂所释放的微粒在不停地运动。不幸的是，无机物颗粒也运动。这个颗粒在液体表面在不停地看似无规地运动着。统计解释了布朗运动，是统计物理进入的标志。

1867 年，麦克斯韦率先讨论了热力学第二定律的统计特性，1878 年提出了统计物理这个新词。麦克斯韦基于速度各项同性以及指数函数乘积的性质得出分子数关于动能

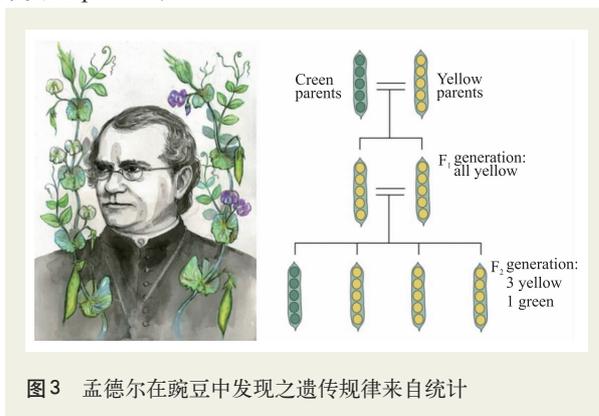


图3 孟德尔在豌豆中发现之遗传规律来自统计

5) 多宇宙, multi-verse, 的图景是想象出来以作统计的, 但与想象出来的有很多复制样本的 ensemble 不同, 后者原则上是可以观测的。

6) 作为物理概念的这个词指的是 mv^2 。

分布的公式是天才的杰作。1877年,玻尔兹曼假设分子能量有最小单元,分子能量只能是这个能量单元的整数倍(因此,笔者将1877年看作量子力学元年),配合平衡态即微观状态数最大的哲学,同样获得了麦克斯韦此前的结果,这就是著名的麦克斯韦—玻尔兹曼统计。统计力学甫一出手,即带来不俗的结果。由分布出发求和很容易,比如计算 $\sum_i n_i = N$, $\sum_i n_i^2 = N$;反过来,给定 N ,写成 $N = \sum_i n_i$ 或者 $N = \sum_i n_i^2$ (固体物理中会常用到此式),这是按照一定规则求 N 的partition,就有些难度了。Partition,分割、分拆,在几何、数论方面都会出现。Partition function,要义在于对总粒子数、总能量的拆分方式的研究,中文将之翻译成配分函数,几乎没见过有统计物理课本说清楚这一点。

统计物理的发展得力于试图用统计物理解热力学规律。系统的性质可由几个少数的参数,包括几何的、动力学的和热力学的参数加以描述。热力学的原理需要从构成宏观体系的微观部分构成以及支配微观部分之运动的动力学规律加以解释,这是统计力学的初衷。对微观动力学状态(state)作statistics,这统计物理,statistical physics,真是名副其实。平衡态物理量可以通过对微观动力学状态赋予一个同约束相融洽的几率分布计算出来,宏观值可以对具有不同几率的微观状态求平均而获得。统计物理,不止是几率概念和几率诠释扮演了重要的角色,而是概率进入了基础物理理论的层面,为因果律这种物理学的基石带来了讨论。统计是达成因

果律所决定的某些确定结果的捷径,还是说没有严格的因果关联⁷⁾?关于统计物理的一个误解是,统计物理似乎被当作是精确科学(exact sciences)的对立面,其关键词熵被当成无序的代名词。

作为量子力学基础的波函数几率诠释是1926年左右的事情,决定性方程得来的解的几率性诠释,这比较具有颠覆性。据Otto Stern说,外尔曾被人问到量子力学几率解释的重要性时,说:“in those years (1924 to 1927), everybody talked about probability.”相较于经典概率为[0, 1]之间的实数,量子力学其实是在用波函数 ψ 说话,而 ψ 是模为1的几率幅,是一个复函数。从经典概率到量子力学语境下的概率,这之间的conceptual gap是必须要面对的。任何替代几率诠释的诠释,而如果坚持把物理量当作自伴随算符,这之间如何协调就是个问题。统计是一个时常有悖于直觉的学科,量子统计肯定比量子力学带来更多的悖论。双胞胎联姻,外人眼里可能会错认出四对婚姻,但state发的结婚证书只有两套,理解了这些,可能有助于理解量子统计与经典统计之间的区别。

量子统计出现在量子力学的高级阶段,但必须牢记统计在量子之前,统计是量子的基础。实验的统计结果,很多时候是构造量子理论的出发点。一个典型的例子是统计和自旋的关联,实际上我们是根据统计行为把粒子分成了玻色子和费米子⁸⁾,这个统计行为和粒子的自旋有关。在基本粒子研究中,为了确立自散射实验得来的数据的统计,还会引入诸如色、味之类的量子数。

近代的大物理学家,几乎都精

于统计物理,而这之间尤以爱因斯坦为甚。在一般人眼里,统计不过是算平均和方差,甚至有些人对统计的关切只到平均值(期望值)的层面。比利时天文学家、统计学者Adolphe Quetelet是社会物理学概念的提出者,竟定义了平均人(l'homme moyen)这个概念,即各项指标是社会平均值(假设正则分布)的那么一位。平均值当然不足以再现具体的分布,更不能反映个体的酸甜苦辣——如果平均收入算是自家的收入的话,那笔者也该有富人的感觉。平均值是变量关于分布的一阶矩,下一个统计量,二阶矩,就是方差。爱因斯坦认识到方差,也即涨落,对理解物理现象的重要性。作为对统计物理之威力的理解就又进了一步。注意,高斯分布,或者正则分布、钟形分布,之所以特殊,是因为由平均值和方差可以完全确立高斯分布⁹⁾。统计物理博大精深,至于parastatistics, anionic statistics, braid statistics这些近期的研究前沿,光看看题目,哪个都不是好理解的。笔者不懂,也后悔未曾认真研习过一本统计物理的经典。

5 多余的话

统计的出现,是同城邦出现所带来的社会需求相联系的;其发展过程中,人类的嗜赌成性又起到了极大的促进作用。至于统计物理作为一门科学,应该理解为确实与自然事件的chance nature有关。

城邦对统计的需求,是旺盛的。它是统治的基础。有些人总担心自己健忘,其实大可不必,statistic一直在帮你做着statistics呢。如果

7) 高斯分布在量子力学中总出现是因为高斯分布函数的傅里叶变换还是高斯函数,可以编故事。

有必要，警察配合高科技能让你回忆起你三岁时在幼儿园偷吃韭菜鸡蛋饺子时到底是哪两颗牙齿中间塞了片韭菜叶子。二战期间德国的秘密国家警察(Geheime Staatspolizei, 其缩写 Gestapo 被译成盖世太保), 二战后东德的 Stasi (Ministerium fuer Staatssicherheit 中最后一个字国

家安全的缩写), 都是统计行业的翘楚。在电影《窃听风暴》(*Das Leben der Anderen*, 字面意思是别人的生活)中, 主角, Stasi 的一个工作人员, 整天的工作就是记录被统计对象读过哪几本书、买过几双鞋或者见过什么人。他整天盯着别人, 他就成了别人的生活的见证

者, 但却没有自己的生活。其实, 不只是专门统计机构的人员, 有些不是这行当的人也有这种监视别人家生活的强烈愿望, 这似乎是人性恶的顽固一面。

Statistics, the affair of states, the science of states for beings in the universe, do you understand?

参考文献

- [1] David F N. Games, Gods & Gambling: A History of Probability and Statistical Ideas. Dover Publications, 1998
- [2] Yang C N. The conceptual origins of Maxwell's equations and gauge theory, *Physics Today*, 2014, 67(11):45
- [3] Schrödinger E. Science and the human temperament. George Allen & Unwin Ltd., 1935
- [4] Nature editorials. Number Crunch. *Nature*, 2014, 506:131
- [5] Nuzzo R. Statistical error. *Nature*, 2014, 506:150
- [6] Lyons L. Discovery or fluke: statistics in particle physics. *Physics Today*, 2012, (7):45
- [7] Born M. *Natural Philosophy of Cause and chance*. Oxford: Clarendon Press, 1948
- [8] Uhlenbeck G E. Statistical mechanics and quantum mechanics. *Nature*, 1971, 232:449

读者和编者

订阅《物理》得好礼

——超值回馈《岁月留痕——<物理>四十年集萃》

部特推出优惠订阅活动：向编辑部连续订阅2年《物理》杂志，将获赠《岁月留痕——<物理>四十年集萃》一本(该书收录了从1972年到2012年在《物理》发表的40篇文章，476页精美印刷，定价68元，值得收藏)。

希望读者们爱上《物理》!

订阅方式(编辑部直接订阅优惠价180元/年)

(1) 邮局汇款

收款人地址：北京603信箱，100190

收款人姓名：《物理》编辑部

(2) 银行汇款

开户行：农行北京科院南路支行

为答谢广大读者长期以来的关爱和支持，《物理》编辑

户名：中国科学院物理研究所
帐号：112 501 010 400 056 99
(请注明《物理》编辑部)
咨询电话：010-82649266；82649277
Email: physics@iphy.ac.cn

