

不确定性与 Bayes 统计

刘乐平 / 文

什么是确定？这是一个既简单而又复杂的问题。说其简单，因为不确定=不+确定，英文单词亦如此 Uncertainty=Un+certainty 确定之外就是不确定。论其复杂 因为如再问起什么是确定的 那就不容易回答了。

在我们生活的这个世界，不确定性无处不在，无时不有。且不论浩瀚的宇宙因何而生，无际的星云为何而去，也不谈神秘基因的排列组合，微小电子的舞动轨迹。就用发生在我们每个人身边的，我们在日常生活中肯定、必须、确定要遇见的不确定性来说，就比比皆是。

在一个人成为一个人之前，他（或她）的父母的结合肯定是不确定的，是两小无猜，还是媒妁之言，是一见钟情，还是随遇而安（如再往上一代考证，他的爷爷与奶奶，她的外公外婆，不确定性的程度就更大）。在这个人出生之前，是男是女，是高是矮、是胖是瘦、是俊是丑……不能确定；出生之后，是成龙成凤，还是无作无为；是功成名就，还是身败名裂；平平淡淡、风风火火……无法预知。这样看来，人的一生，用通俗的话讲是“非常可能不一定”。稍微严密些，可以从不确定性方面总结出一个公式，这就是“不确定性的n次方，其中，n的值对每个人来说都是不确定的。”

会计的不确定性、管理的不确定性、企业的不确定性、生活的不确定性、社会的不确定性、国家的不确定性……。是否可以这样推理，只要能想出一个名称，你就可以非常肯定的在其后加上三个字：“不确定”。

我们先来看一下“不确定”的定义，按照通常的做法，查词典或百科全书。根据

韦氏英语词典中的解释：“不确定性是从英文——Uncertainty一词而来，表示‘不确定的性质或状态’，不确定的包括：不肯定的、不可靠的、模糊的、不稳定的、无规则的、变化无常的、偶然的等等”。用上现代化手段，从“金山词霸2002”中寻找：“不肯定、不确定、无把握、未决定的状态或性质”；美国传统词典（双解）和现代英汉综合大词典有这样一些与不确定性相关的句子：“我对自己的决定越来越怀疑”，“导致接受假设或作出决定的犹豫的心理疑惑状态”……。

我们再来看学者的研究。我们生活在一个知识经济时代。知识不仅是构成人的认识能力的重要基础，而且也是知识经济的最主要生产要素。过去我们总是把知识看作是确定性的化身，但是由于20世纪的科学革命和计算机网络的迅猛发展，知识凸显出不确定性。吴国林在他的“论知识的不确定性”一文中从哲学的高度，论述了科学知识的不确定性。他认为，“确定”有两层意思：一是从事物或过程本身存在或状态来看，它们的存在是不变的或是真(truth)的(具有真实性、真理性)，人们对此没有怀疑；二是从人们对事物或过程的认识来看，事物的存在或真能够完全为人们所认识，能够运用其“真”或规律肯定地预见未来。因此，“确定”的涵义不仅是指事物或过程本身的客观性，而且还意味着主体对客体的知道、认识和理解程度。认识了“确定”的涵义，我们就容易定义“确定性”与“不确定性”。确定性就是指事物或过程具有“确定”的那样一种性质。不确定性是指事物或过程不具有“确定”那样的性质，或是确定性的缺乏。……确定性与不确定性可以从存在论与认识论角度加以区分。从客观事物本身界说确定性与不确定性，可

称之为存在论意义的确定性与存在论上的不确定性。如自然界本身具有不确定性。从事物或过程的客观性角度来看，不确定性大致可以用“随机性”、“不肯定性”、“不完备性”、“不稳定性”、“没有确定的真理性”来表达。

我们又生活在一个市场经济时代，在现在市场经济的浪潮中，市场的不确定性与经济的不确定性息息相关。在经济学领域，李拉亚对不确定性问题进行了比较系统的研究，他在《通货膨胀与不确定性》一书中，给出了不确定性的两种定义：一是与概率事件相联系，通常用随机变量的方差来定义；另一种定义与概率事件没有联系，不确定性是一种没有稳定概率的随机事件。在证券投资领域，汤光华在《证券投资的不确定性》中，从语言学的角度对不确定性给出了解释：“不确定性是事物属性状态的不稳定或不可确知的，人们一般认为“不确定性”是指未来，其实无论过去、现在和未来，都有‘不确定性’”。

知识的不确定性、经济的不确定性……就连科学都具有不确定性。汤光华在他的书中列出了历史上许多伟人的经典名言：数学，这颗科学王冠上的明珠，被看作是一种从不确定性中追求确定性的方法，在传统上被看成是描述物理世界的不可动摇的真理。伽利略有句名言：“自然这本书是用数学语言写成的”，就表达了这样一种信念。可美国著名数学史家克莱因在他的著作《数学：确定性的丧失》中明确指出：“数学设计并不是自然界所固有的。真理的丧失，数学和科学的不断增长的复杂性以及不确定性，是大多数数学家抛弃了科学，由于数学院内的灾害，他们退到特殊的数学领域，在那里，证

明的方法似乎是安全的。”爱因斯坦说过：“就涉及的一些数学命题而言，它们是不确定的……；就它们是确定的来说，它们却又不涉及到现实。”在第二次世界大战期间，幽默作家和数学家Stephen Leacock说过这样一段非常幽默的话：“人们将告诉你：数学、哲学和神学现今都凑到一块了。在某种意义上表明……它们就像一起参加葬礼的三个人那样凑到一块，这个葬礼就是确定性的死亡。”罗素在《人类的知识》一书的最后一页说过这样一句话：“全部人类知识都是不确定的、不精确的和全面的。”

面对这个“不确定性”的世界，人们确实是需要各种各样的对付不确定性的手段。这些手段可以分成两类：一类是定性的，比如建立法律、规章和制度等来规避不确定性所带来的风险。另一类是定量的，用精确的方法来度量不确定性，在不确定性中寻找确定的规律。在定量手段中，统计方法具有无可比拟的优势。从某种意义上来说，就是因为有了不确定性，才有统计学的产生，才有统计学的发展。用杰出的当代贝叶斯统计学家James O. Berger教授的话来说：“统计学就是对不确定性进行测度的科学。”

统计学的天空浩瀚无垠，这里只想简单介绍这天空的一半——贝叶斯统计学。

早在三百年以前，人们就开始严肃地考虑当存在不确定性时如何进行推理这个问题。Samuel Kotz和吴喜之教授猜测“James Bernoulli(1713)恐怕是第一个构造该问题的人”。Bernoulli意识到在可应用于机会游戏的演绎逻辑和每日生活中的归纳逻辑之间的区别，对于他来说，这个未回答的问题在于前者的机理如何能帮助处理后者的推断问题。

而托马斯·贝叶斯(Reverend Thomas Bayes, 1702-1761)是对归纳推理给出精确定量表达方式的第一人。他死后发表的论文应该作为科学史上最著名的回忆录之一。对贝叶斯方法的评价很多，直到今天还存在不少争议。但陈希孺院士对贝叶斯的评价应该是非常客观的：“贝叶斯在18世纪上半叶欧洲学术界不算一个起眼的人物。在他生前，没有片纸只字的科学论著发表。那时，传播和交流科学成果的一种方式，是学者间的私人通信。这些信件许多都得以保存下来并发表传世，例如惠更斯-巴斯噶通信。但在贝叶斯生前，除在1755年有一篇致约翰·康顿的信(其中讨

论了辛普森有关误差理论的工作)见诸约翰的文件外，历史上也没有记载下他与当时的人有何重要的学术交往。不过，他一定曾以某种方式表现出其学术造诣而为当时的学术界所承认，因为他在1742年当选为英国皇家学会会员。在当时直到今天，这个称号相当于英国科学院院士。这个生性孤僻，哲学气味重于数学气味的学术怪杰，以其一篇遗作的思想重大地影响了两个世纪以后的统计学术界，顶住了统计学的半边天。”

1812年，Laplace在他的概率论教科书第一版中首次将贝叶斯思想以贝叶斯定理的现代形式展示给世人。Laplace本人不仅重新发现了贝叶斯定理，阐述得远比贝叶斯更为清晰，而且还用它来解决天体力学、医学统计、甚至法学问题。

目前被承认的现代贝叶斯统计工具的使用，应归功于Jeffreys(1939)、Wald(1950)、Savage(1954)、Raiffa, Schlaifer(1961)、Lindly(1972)和DeFinetti(1974-1975)。在20世纪90年代，由于高维计算上的困难，贝叶斯方法的应用受到了很大的限制。但随着计算机技术的发展和贝叶斯方法的改进，特别是目前MCMC方法的发展和WinBUGS软件的应用，原来复杂异常的数值计算问题如今变得非常简单，参数后验分布的模拟也趋于方便，所以现代贝叶斯理论和应用得到了迅速的发展。现在已很难发现一个人类的研究领域不存在某种水平的贝叶斯分析工作。

我们仅以国际上关于贝叶斯统计分析的专著数的增长为例。Berger教授2000年在《美国统计学联合会期刊》(JASA)中撰文指出：从1769年到1969年，200年间大概有15本著作出版，从1970年到1989年20年间，贝叶斯统计学的书籍仅有30本，然而从1990年到1999年的最近10年中，贝叶斯分析的专著就有60本出版，这还不包括数十本关于贝叶斯会议的文集和参考书。

什么是贝叶斯统计？茆诗松和王静龙教授的解释非常简单明了：“贝叶斯推断的基本方法是将关于未知参数的先验信息与样本信息综合，再根据贝叶斯定理，得出后验信息，然后根据后验信息去推断未知参数”。

什么是先验信息？为了对未知参数作统计推断(或统计决策)，我们需要从总体抽

取样本，并且愈多愈好。因为样本含有未知参数的信息，并且是“最新鲜”的信息。这是经典统计推断的主要依据。可是我们周围还存在有一些非样本信息。这些非样本信息主要来源于经验和历史资料。由于这些经验和历史资料大多存在于(获得样本的)试验之前，故又称为先验信息。先验信息同样也可以用于统计推断和统计决策，因为当需要对未来的不确定性作出统计推断时，当前的状态固然重要，但历史的经验也同样是举足轻重的。

对贝叶斯统计批评最多的是它的“主观性”。能否在统计推断中利用主观经验，加入先验信息，我国著名科学家钱学森的一段话非常值得我们深思：“处理复杂行为系统的定量学方法学，是科学理论、经验和专家判断力的结合，这是定量方法学，是半经验半理论的。提出经验性假设(猜想或判断)，是建立复杂行为系统数学模型的出发点。这些经验性假设(猜想或判断)不能用严谨的科学方法证明，但需要经验性数据对其确实性进行检测。从经验性假设(猜想或判断)出发，通过定量方法途径获得的结论，仍然具有半经验、半理论的属性。当人们寻求用定量的方法处理复杂行为系统时，容易注重于数学模型和逻辑推理，而忽视数学模型微妙的经验含义或解释。要知道，这样的数学模型看起来‘理论性’很强，其实不免牵强附会，从而脱离真实。与其如此，反不如从建模一开始就老实承认理论的不足，而求助于经验的判断，让定性的方法和定量的方法结合起来，最后定量。”钱先生的话无疑给我们在面对不确定性时，怎样将定性和定量的方法进行有机地结合指明了一个明确的方向。

2002年是贝叶斯诞辰300周年，本文最初的设想是为了纪念这位顶住统计学“半边天”的“圣人”。但由于诸多不确定的原因拖至现在才得以完成。贝叶斯的家乡是否有纪念活动，其他国家是否也有人撰文纪念他，我不得而知，不敢确定！(还有更加不确定的是一本文是否能够发表？)不过，不管怎样，可以确定的是正是因为有了不确定性，我们的世界才会这样美妙，我们的人生才会如此精彩，我们的统计才会充满生机！

