概率：了解不确定性

2016-02-09 方弦 [算法与数学之美](javascript:void(0);)

“概率”这两个字，除了课本以外，最常出现的地方也许就是天气预报中的“降水概率”，也就是未来几天下雨的可能性有多大。在数学中，概率论是专门研究“可能性”的一门分支。它涉及的问题非常广泛，内容远远超出了中学课本里那些刻板的习题。一切随机或者不确定的事件，都是概率论研究的范畴。上至气象下至金融，甚至连“磁铁的磁性怎么来的”这种物理问题，都可以用概率的方法来研究。

但这门学科的诞生却有些“不太光彩”。

**来自赌博的问题**

在1654年的一天早上，法国数学家布莱兹·帕斯卡收到了他的朋友贡博的一封来信。这位朋友自称“来自梅雷的骑士”，也算是一位业余数学家。他向帕斯卡提出了类似如下的问题：

两位贵族A与B正在进行一场赌局，赌注是每人500法郎，两人轮流掷硬币，得到正面则A得一分，反面则B得一分，每一局两人得分的机会相等，谁先得到6分谁就得到1000法郎。两人激战正酣，比分达到2比4之际，B突然有事需要终止赌局。赌注应该如何分配才最公平。

这一类问题被称为点数分配问题，早在16世纪就被研究过，但数学家当时的答案并不令人满意，在一些极端情况下会给出非常不合理的分配方案。也许这位“梅雷骑士”也见识过现实中这种赌局引起的矛盾，他希望帕斯卡能够解决这个问题。

帕斯卡对这个问题也很感兴趣。他向另一位业余数学家皮埃尔·德·费马发去一封信讨论这个问题。作为“业余数学家之王”，费马很快就给出了一个答案。他认为，不能单靠赌局停止时的比分或者各自获胜需要的分数来决定赌注的分配，而是应该考虑所有比赛的可能性中，双方获胜的比例。但列举所有的可能性的计算量非常大，帕斯卡继而提出了一个简化算法，完美地解决了点数分配问题。

实际上，他们的解答相当于计算两位玩家胜利概率的大小。在研究中，帕斯卡提出了“数学期望”的概念和著名的“帕斯卡三角形”（杨辉三角）。某个结果为实数的随机事件的数学期望，也就是所有结果按照发生概率加权之后的平均值。数学期望这个概念，掀开了概率论研究的序幕。

**什么是概率？**

很多概率问题有着特别的结构。对于某个非常简单的随机事件，比如说掷硬币，我们知道每种结果出现可能性的大小，这样的事件被称为“基本事件”。我们可以多次重复这些基本事件，假定它们发生的可能性不会改变，而且这些重复没有相互影响。如果我们将这些基本事件以合适的形式组合起来，就能得到一个更为复杂而有趣的系统。许多概率问题实际上就是对这些随机系统的各种性质的研究。比如说，在点数分配问题中，基本事件就是硬币的投掷，而系统则是赌局的具体规则，最后我们希望知道的则是每一方胜利的可能性大小。

在概率论发展的早期，数学家研究的问题大多比较简单，基本事件只有有限几种结果，组合的方式也相对简单。这样构成的随机系统又叫古典概型。随着数学的发展，数学家开始考虑更复杂的模型。18世纪的法国数学家布丰提出了这样一个问题：在数条间隔相等的平行线之间，随机投下长度与间距相等的一根针，它与这些平行线相交的概率是多少？在这里，因为角度与距离都是连续的值，基本事件有无数不同的结果，这样的随机系统被称为几何概型。

早在19世纪，概率论已经成为了一门枝繁叶茂的数学分支。有趣的是，“概率”这个概念的严格定义要等到20世纪才出现。对于古典概型，因为结果数量有限，概率的定义并没有含糊之处，但几何概型的情况更为复杂。考虑这样的一个问题：圆中的一条随机的弦，它的长度比圆内接正三角形的边长更长的概率是多少？这个问题又叫贝特朗悖论，它奇怪的地方在于，对于不同的选取“随机的弦”的方法，得到的概率也不相同，到底谁是谁非？要等到1933年，俄国数学家柯尔莫哥洛夫为概率论建立公理体系之后，这个问题的解答才变得昭然若揭。柯尔莫哥洛夫将概率模型建立在某一类所谓的“σ代数上的测度”上，这样的测度可以有很多种，不同的测度对应着不同的“随机”。而在贝特朗悖论中，选取随机弦的方法实际上对应着不同测度的选取，也就是不同的“随机”概念，那自然会得到不同的结果。

而到了现在，概率模型的种类越来越多也越来越复杂，系统可以包含无限个基本事件，而具体的组织方式也更复杂更有趣。随机图、渗流模型、自回避行走，这些概率模型早已不能用古典概型和几何概型来概括。也正因为有了这些复杂的模型，我们才能用概率论解决现实世界的种种难题。

**无处不在的分布**

如果让数学家评选概率论中最重要的定理，桂冠可能非中心极限定理莫属。它不仅是概率论中许多重要结果的基石，在别的学科中，尤其是计算机科学，它也有重要的应用，而在现实生活中，它是整整一个行业赖以生存的理论基础。

中心极限定理其实不止一个，可以说它是一连串定理的总称。它可以看作所谓“大数定理”的细化与推广。假设我们有一枚硬币，它掷出正反面的概率相等。那么，如果我们连续抛掷这枚硬币一万次，常识告诉我们其中大概有五千次是正面。这就是大数定理：对于某个基本事件独立地重复多次的话，某个可能性发生的次数占总数的比例会趋近于这个可能性发生的概率。

与大数定理不同的是，中心极限定理处理的是那些结果是实数的随机基本事件。它告诉我们，如果将许多相同而又独立的基本事件的结果取平均的话，这个平均值会趋向某个概率分布。根据大数定理，这个分布的数学期望就是基本事件的数学期望。而中心极限定理额外告诉我们的，就是这个概率分布必定是一个所谓的“正态分布”，而它的方差，也就是概率分布的“分散”程度，是基本事件的方差除以事件数目的平方根。也就是说，基本事件越多，平均值的不确定性就越小。将这个正态分布画成曲线的话，它就像一个大钟，中间高，但两头呈指数衰减，这也为它赢得了“钟形曲线”这个形象的名字。中心极限定理可以推广到取值范围是高维空间中一点的情况，“相同的基本事件”这个要求也可以被更弱的条件代替，只需要基本事件满足某些要求，而不需要完全相同。

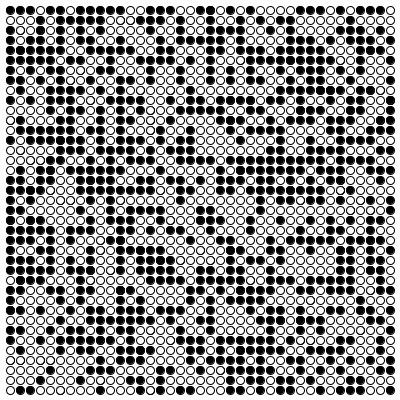
正态分布在自然界中随处可见，比如说人的身高和智力都服从正态分布。这是因为自然界中的很多现象都由各种因素千丝万缕的联系而决定，其中没有特别突出的因素。比如说人的身高，除了由许多不同的基因调控以外，后天的营养、环境、健康，甚至偶然的意外，都有着各自的影响。在这种情况下，如果将每个因素看成一个基本事件，并且假定这些因素各自的影响都差不多，将这些因素综合考虑，根据中心极限定理，得到的结果就非常接近正态分布。

中心极限定理也是保险这一整个行业的基础。每个人都会遇到各种各样的风险，比如事故、疾病等等，这些风险发生的概率都很低，但一旦发生，后果非常严重，并非每个人都能承受。而保险业，实际上就是通过保费与保险赔付的方式，将上千万人连结起来，每人付出相对小的代价，在万一不幸袭来时，就能获得一定的保障。由中心极限定理，这样由数量庞大的个案相加而成的保险业务，由于偶然因素导致无法赔付的概率非常小，而且参与的人数越多，风险就越小。为了确定保费与赔付，保险公司要做的就是根据大量统计数据精确地确定意外发生的概率，然后根据意外概率与收益确定保费与赔付的金额。这也是为什么现代的保险公司越来越重视概率与统计。

**理解复杂世界**

除了与不确定性相关的问题之外，概率论也与物理息息相关。法国物理学家皮埃尔·居里在攻读博士学位时，就发现了磁铁的一个有趣的性质：无论磁力多强的铁制磁铁，在加热到770摄氏度时，都会突然失去磁性。这个温度后来被称为铁的居里点。为什么磁铁会突然失去磁性？通过概率论与统计物理，我们现在明白，这种现象与冰雪消融、开水沸腾相似，都属于相变的范畴。

我们可以将磁铁里的铁原子想象成一个一个的小磁针。在磁铁还有磁性时，这些小磁针齐刷刷地指向同一个方向，但因为分子热运动的关系，每个小磁针都会时不时地动一下，但很快会被旁边的小磁针重新同化。物理学家将这个场景抽象成所谓的伊辛模型，通过对伊辛模型的研究，概率学家发现，当温度达到某个临界值时，整个体系就会由于热运动而不能保持统一的指向，也就是失去磁性。这个临界值就是居里点，而这样的对伊辛模型的研究也部分揭示了磁铁的一些微观结构的成因。

  
（图片很大，请耐心等待）

相变不仅仅局限于物理现象。流言的传播，传染病的爆发，还有微博的转发，都是一种相变过程，都存在某种临界值。比如说传染病，在适当的模型下，如果每个病人传染人数的平均值低于某个临界值，那么疾病就能被控制；如果高于临界值，就很可能导致疫病的全面爆发。对于疾病传播的研究，属于流行病学研究的范畴，而在概率论被引入流行病学研究之后，我们对如何防止与控制疫病爆发有了更深入的了解，这是能挽救成千上万人的知识。

概率论的应用远远不止这些，大至飞机失事搜救，小至垃圾邮件过滤，都能在其中找到概率论的身影。这个复杂的世界充满了不确定性，有些无伤大雅，有些却能致命。要驾驭这些不确定性，就要从了解它们开始。这就是概率论的意义。概率论不能为我们带来一个没有风险的世界，但它却能教会我们如何与风险和平共处。它带来的仅仅是关于不确定性的知识。

但知识，往往就是力量。

关于

本文已发表于《科学·24小时》2015年11月号，未经书面许可，禁止转载。