数　学　文　化

第 7 章
概率论与数理统计的思想方法与意义

概率论与数理统计是研究与揭示随机现象统计规律性的一门学科，它的起源与博弈现象有关。

在16世纪，意大利的一些学者开始研究赌博中的一些简单问题；

到了17世纪中叶，法国与荷兰的一些数学家基于排列组合方法，解决了一些较复杂的赌博问题。1812年，拉普拉斯在系统总结前人工作的基础上，写出了《概率的分析理论》，并在概率论中引入了更有力的分析工具，将概率论的发展推向一个新的阶段。

19世纪末，俄国数学家们用分析方法科学地建立了实际中遇到的许多随机变量近似地服从正态分布的理论，给出了概率的公理化定义，发展起了现代概率理论。数理统计虽然源于古代，但它的正式诞生应当是19世纪后期的事情。

概率论的建立为数理统计奠定了理论基础，而数理统计的发展又为概率论的应用提供了用武之地。两者互相推动，迅速发展。目前，概率论与数理统计已经广泛地应用于自然科学、技术科学、人文科学、社会科学等许多领域，它在经济、管理、工程、技术、教育、语言、生物、环保、国防等许多领域中的作用愈益显著。

章节目录

7.1 概率论与数理统计发展简史

7.1.1 概率论发展简史

首先提到的是文艺复兴时期的数学家、医学家J·卡当，他才华横溢，对数学贡献巨大，但却热衷于赌博。他不希望把时间花在不能获利的事情上，因此，他认真地研究牌技以及在一副牌中获得“A”的概率。他把自己的研究成果编成了一本手册，题为《赌博的游戏》。这是世界上第一部研究概率论的著作。他的研究除了赌博外还与当时的人口、保险业等有关，但由于卡当的思想未引起重视，概率概念的要旨也不明确，于是很快被人淡忘了。

大约100年以后，另一位赌徒梅累继续研究概率问题。可是他不具有像卡当那样的数学天分，所以不得不就这一问题去请教数学奇才帕斯卡。帕斯卡就梅累的问题与费马通了信，由此，帕斯卡和费马创立了概率论的一些基本结果。他们往来的信函中讨论了如下的“合理分配赌注问题”：甲、乙两人同掷一枚硬币。规定：正面朝上，甲得一点；若反面朝上，乙得一点。先积满3点者赢取全部赌注。假定在甲得2点、乙得1点时，赌局由于某种原因中止了，问应该怎样分配赌注才算公平合理？

当费马和帕斯卡通信讨论的问题被数学家惠更斯知晓后，他对这个问题进行了较为深入的研究。1657年，惠更斯的名著《论赌博中的计算》一书出版。此书是概率论的第一部成形的著作，书中提出了数学期望、概率的加法与乘法定理等基本概念。

1677年，法国数学家浦丰利用有名的浦丰投针问题给出了几何概率的概念。

使概率论成为一个独立数学分支的是瑞士数学家雅各布•伯努利。1713年出版了他的遗作《猜度术》，书中提出了现在称之为伯努利大数定律的概率论的第一个极限定律，起到了概率的理论奠基作用。

1812年，拉普拉斯的名著《概率的分析理论》出版，书中系统总结了前人关于概率的研究成果，使以前零星的概率知识系统化，而且明确给出了概率的古典定义，并引入分析方法，把概率论提高到一个新的阶段。

1733年，1809年棣莫佛与高斯分别独立地引进了正态分布的概念。1837年，法国数学家泊松发表著名论文《关于判断的概率之研究》，提出泊松分布。

1866年，俄国的切比雪夫建立了独立随机变量的大数定律，使伯努利与泊松的大数定律成为特例，并把棣莫佛与拉普拉斯的极限定理推广为一般的中心极限定理。

由于拉普拉斯的概率定义存在模糊的意义，1899年，法国科学家贝朗特提出了所谓的“贝朗特悖论”：在半径为r的圆内随机地选择弦，求弦长超过圆内接正三角形边长的概率。由于对“随机地选择”的不同理解，使得结果不唯一。概率论陷入危机之中。

为了克服古典概率的缺陷，数学家们开始创建概率的公理系统。俄国数学家伯恩斯坦、奥地利数学家冯•米西斯都提出了一些概率公理，但都不甚理想。1905年，法国数学家波莱尔用他创立的“测度论”语言来表述概率，为现代概率打开了大门。

1929年，前苏联数学家柯尔莫哥洛夫发表论文《概率论与测度论的一般理论》，首次给出了以测度论为基础的概率论公理结构；1930年，他的《概率论中的解析方法》开创了随机工程的一般理论（即马尔科夫过程）；1933年，他出版了名著《概率论基础》，建立了柯尔莫哥洛夫公理化概率论。

1934年，前苏联数学家辛钦提出“平稳理论”，建立了平稳随机过程理论。1942年。日本数学家伊藤清引进了随机微分方程，为随机分析理论奠定了基础。

1949年，柯尔莫哥洛夫与格涅坚科合作写出《独立随机变量与极限分布》，建立了弱极限理论。

7.1.2 数理统计发展简史

近代统计学是在概率论的基础上建立起来的。1662年，英国统计学家J. 格兰特组织调查伦敦的人口死亡率，并发表《从自然和政治方面观察死亡统计表》的专著，提出了“大数恒静定律”。

1763年，英国统计学家贝叶斯（T.Bayes）发表《论机会学说问题的求解》，提出“贝叶斯定理”，也就是从结果去对原因进行后验概率的计算方法。

19世纪中叶，比利时统计学家A.凯特勒把统计方法应用于天文、气象、物理、生物与社会学，并强调正态分布的用途，为统计方法的推广做了大量工作。同一时期，爱尔兰经济学家E埃奇沃斯引入了方差的概念。

1889年，英国生物学家高尔顿出版其著作《自然的遗传》，引入回归分析方法，给出了回归直线与相关系数等重要概念。高尔顿是生物统计学派的奠基人，他用统计方法研究遗传进化问题，第一次将概率统计原理应用于生物科学，明确提出“生物统计学”。

从19世纪末到二次世界大战结束，数理统计得到蓬勃发展并日臻成熟。这一时期，英国数学家皮尔逊发展了生物统计学与社会统计学的基本法则，发展了回归分析及相关理论，并于1900年提出了卡方统计量与卡方分布，建立了卡方检验法。1908年，皮尔逊的学生，英国科学家W. S. 戈塞特导出大统计量及其精确分布，建立了 t 检验法（也就是学生分布）。

现代数理统计的奠基人应该是英国数学家费歇尔（Fisher Ronald Aylmer，1890-1962）。1929年，他出版了《理论统计的数学基础》，对统计学中的相关系数、样本分布、多元分析以及统计方法在遗传与优生方面的应用都进行了研究，成为现代统计学的奠基性著作，在估计理论、假设检验、实验设计、方差分析等方面都作出了贡献。

1940年，瑞士数学家克拉默（H.Cramer）发表《统计学的数学方法》，运用测度论方法总结了数理统计的成果，使现代数理统计趣于成熟。

我国数学家许宝禄在数理统计和概率论这两个数学分支都有重要贡献。他的重要成绩有：1938年至1945年间，他在多元统计与统计推测方面发表了一系列论文，给出了样本协方差矩阵等概念，推进了矩阵论在数理统计学中的应用；他对高斯-马尔科夫模型中方差的最优预计的研究是其后关于方差分量和方差的最佳二次预计的众多研究的出发点；他推动了人们对全部相似检验进行研究等。

7.2 概率论与数理统计的基本思想

所谓概率，通俗地说：就是一件事情发生的可能性的大小。

在日常生活中，我们所使用的概率思想，主要是满足于估计一件事情发生的概率是大还是小，从而为我们的决策提供一种理性的支持。

我们来看看在古典概率中如何利用数学得到精确的概率值。例如，单独抛一枚骰子，出现“2”的概率是多少？解决这个问题的一种方法是，掷100000次骰子，然后计算出现“2”的次数。出现“2”的次数与100000的比就是所求的答案，或者差不多会接近真实的答案。但是，数学家们一般不会采用这种方法，而是静坐默思去找出解决这个问题的方法。

我们来看看帕斯卡和费马如何考虑这个问题的：一个骰子有6个面，由于在骰子的形状上或者在扔骰子的方式中，没有任何因素有利于某一面的出现，所以得到每一面正面朝上的可能性是相同的。六面出现的可能性相同，而仅仅只有一面也就是出现“2”的一面是有利情形，因为这就是我们所要求的那一面。因此出现“2”的概率就是1/6。如果我们对出现4或5这两面都感兴趣，我们则得到其概率为2/6，即6种可能性中的两种对我们有利；如果我们对出现4或5不感兴趣，那么将有4种有利的可能性，因此概率应该为4/6。

在古典概率中，一般地，计算概率值的定义是，如果有n种等可能性，而有利于一定事件发生的情形是m，那么这个事件发生的概率是m/n，而该事件不发生的概率是（n-m）/n。在这个概率的一般定义之下，如果没有有利的可能性发生，也就是说，如果事件是不可能的，则事件的概率为0；而如果n种可能性都是有利的，也就是说，如果事件是完全确定的，则概率为1。因此，概率值在从0到1的范围内变化，即从不可能性到确定性。

作为这个定义的一个例子，我们考虑从52张普通的一副扑克牌中，选取一张牌“A”的可能性。这里有52种等可能选择，其中有4种是有利的，因此，这个概率是4/52，即为1/13。

从52张一副的扑克牌中选取“A”的概率是1/13。围绕着这一命题的意义，经常会产生一些疑问。这个命题是否意味着，如果一人在这副扑克牌中取了13次(每一次都重复取牌，即将取过的牌又放回)，那么将一定会选中一张“A”呢？事实并不是这样，他可能取了30次或40次，也没有得到一张“A”。不过，他取的次数越多，则取得A的次数与取牌总次数之比将会趋近于1比13。

这是个合理的期望，因为选取的数目越大，每一张牌被取出的次数就会越相等。一个相关的错误想法是，假定如果一人取了一张“A”，比如说正好是在第一次取得的，那么下一次取出一张“A”的概率就必定小于1/13。实际上，概率依然是相同的，即为1/13，即使当3张“A”被连续抽中时也是如此。一副牌或一枚硬币，它们既没有记忆也没有意识，因此已经发生的事情不会影响未来。

注意，我们这里所讨论的问题要具有等可能性。例如，假定我们断说，一个人安全过街头的概率是1/2，因为只有两种可能性：安全通过或没有安全通过。如果这个命题成立，那我们就什么事情也别干了，只有坐在家里。这个命题的错误在于“安全通过或没有安全通过”这两种可能性不是等可能的。

例1 彩票中的数学问题

以某省福利彩票为例做说明。这种彩票玩法比较简单：2元一注，每注填写一张彩票；每张彩票由一个6位数和一个特别号码组成。每个数字均可填写0，1，……，9这十个数字中的任何一个；特别号码可以填写0，1，2，3，4这五个数字中的任何一个。每期开奖，开出一个6位数和一个特别号码作为中奖号码。设六个奖励等级：特等奖——奖券上写的6个数字与一个特别号码全部相同；一等奖——有6个连续数字相同；二等奖——有5个连续数字相同；三等奖——有4个连续数字相同；四等奖——有3个连续数字相同；五等奖——有2个相邻数字相同。每一期彩票以收入的50%作为奖金。三、四、五等奖奖金固定；一、二、特等奖的奖金浮动。

假如，中奖号码是123456，特别号码是0，那么，各个奖项的中奖号码和每注奖金如表7-1所示。

表7-1 某省福利彩票各个奖项的中奖号码和每注奖金

奖级	中奖号码	每注奖金
特等奖	123456+0	（奖金总额-固定奖金）×65%/注数 88万元（保底），500万元（封顶）
一等奖	123456	（奖金总额-固定奖金）×15%/注数
二等奖	12345,23456,...共2组20个	（奖金总额-固定奖金）×20%/注数
三等奖	1234,2345,3456, ...共3组300个	300元
四等奖	123,234,...共4组4000个	20元
五等奖	12,23,34,...共5组50000个	5元

(1) 中奖概率：以一注为单位，计算一注彩票的中奖率

特等奖 —— —张彩票上前6个号码有种可能选择，特别号码有5种选择，故一张奖券上的号码共有5×10^6种不同的填法。因此一注特等奖的中奖率为

$P_0=l/(5\times 10^6)=2\times 10^{-7}=0.0000002$；

一等奖中奖概率为： $P_1=1/10^6=0.000001$；

二等奖中奖概率为： $P_2=20/1000000=0.00002$；

三等奖中奖概率为： $P_3=300/1000000=0.0003$；

四等奖中奖概率为： $P_4=4000/1000000=0.004$；

五等奖中奖概率为：$P_5=50000/1000000=0.05$。

合起来，一注彩票的总的中奖率为上述之和：

$P=P_0+P_1+P_2+P_3+P_4+P_5=0.0543212\approx5.4\%$。

这就是说，每10000张彩票大约有540张得奖(包括从特等奖到五等奖)。

(2)彩票的期望值

因为彩票的返还率一般是50%，所以从总体上说，每注2元一张的彩票，其期望值应该是1元。下面来实际计算一下，看是否如此。决定彩票的期望值有两个因素，一是各个奖级的中奖率，二是各个奖励级别奖金多少。三、四、五等奖的奖金已经给出，中奖的概率也已知道，其他三个等级奖的奖金则可以计算出来。

根据规定，这三种奖级的奖金与三个因素有关：一是当期奖金总额，决定于销售的彩票总注数；二是上期“奖池”中的累积奖金；三是滞留到下期“奖池”的奖金。综合这几种因素，再结合对2001年2——4月发行的20期获奖情况统计的平均值，可以作如下假定：

第一，每一期售出100万注，奖金总额为100万；

第二，每期前三个奖级奖金取平均值；

第三，奖池的累积奖金以平均值计算。

结果如下表7-2：

表7-2 某省福利彩票各个奖项的中奖概率和奖金

奖级	概率	奖金（元）
特等奖	0.0000002	2000000
1等奖	0.000001	50000
2等奖	0.00002	5000
3等奖	0.0003	300
4等奖	0.004	20
5等奖	0.05	5

从而算得期望值

$E=0.0000002\times2000000+0.000001\times50000+0.00002\times5000+0.0003\times300+0.004\times20+0.05\times5$

$=0.4+0.05+0.1+0.09+0.08+0.25=0.97$(元)

即每一注彩票中奖的期望值约为0.97元，这与理论值（1元）非常接近。