21 世纪高等院校教材 教育类

概率统计与微积分

主　编　田长生
副主编　徐庆和

科学出版社
www.sciencep.com

21世纪高等院校教材·教育类

概率统计与微积分

主　编　田长生
副主编　徐庆和

科学出版社
北　京

内 容 简 介

本书是根据教育部2003年颁布的《普通高中数学课程标准》编写的新教材．全书共10章，即：概率与统计概述、数据处理、随机世界的探秘、概率统计的实验设计、博弈论、数学前沿专题介绍、微积分概述、导数及其应用、积分及其应用、微积分的实验设计．主要分为两部分：课堂教学内容和教学内容的实验设计．本书对教学内容进行优化、整合、重组．在讲解重要概念和内容的同时，也介绍了学科的历史渊源和现代发展，以及数学前沿专题，并制作了教学需要的实验设计，同时也深层次地分析了教法，因而建立了新的课程体系．本书是教育数学教材，可以采用模块式教学法，是广大教师和专家丰富的教学经验的总结，是激发创新能力、提高运用数学知识建立教学新模式的重要步骤．

本书可供高等师范院校数学系教学使用，也可供各类教育院校和成人继续教育院校作为教师，包括中学教师培训、进修教材和教学参考书，同时也可作为大学本科生和研究生的学习指导书．

图书在版编目(CIP)数据

概率统计与微积分/田长生主编. —北京：科学出版社，2006
(21世纪高等院校教材・教育类)
ISBN 7-03-017106-3

Ⅰ．微… Ⅱ．田… Ⅲ．①微积分-师范大学-教材②概率论-师范大学-教材③数理统计-师范大学-教材 Ⅳ．①O172②O21

中国版本图书馆CIP数据核字(2006)第030045号

责任编辑：李鹏奇/责任校对：李奕萱
责任印制：张克忠/封面设计：陈 敬

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com

新蕾印刷厂 印刷
科学出版社发行 各地新华书店经销

*

2006年8月第 一 版 开本：B5(720×1000)
2006年8月第一次印刷 印张：22 1/2
印数：1—6 000 字数：432 000

定价：28.00元

前　　言

《概率统计与微积分》是根据教育部 2003 年颁布的《普通高中数学课程标准》编写的新教材，属于数学教育教材. 本教材可供高等师范院校数学系教学使用，也可供各类教育院校教师，包括中学教师作为培训和进修教材及教学参考书，同时也可作为大学本科生和研究生的学习指导书.

在 2002 年下半年，我们依据教育部新的数学课程标准初稿，就开始制定本书编写大纲和编写计划. 经过几年的努力、探讨、实践和反复修改与完善，终于完成了全书的编写工作. 本书的编写和出版，是贯彻国家在新世纪教育要现代化，教学内容和课程体制要改革的具体措施，是顺应当今计算机时代和数字化时代的具体体现，也是激发师生创新能力，建立教学新模式的重要步骤.

本书共 10 章：第 1 章概率与统计概述；第 2 章数据处理；第 3 章随机世界的探秘；第 4 章概率统计的实验设计；第 5 章博弈论；第 6 章数学前沿专题介绍；第 7 章微积分概述；第 8 章导数及其应用；第 9 章积分及其应用；第 10 章微积分的实验设计. 从内容上主要分为两大部分：1）课堂教学内容，包括第 1 章到第 3 章，以及第 5 章到第 9 章的内容. 2）教学内容的实验设计，包括第 4 章和第 10 章的内容.

本书力图体现如下特点：

1）结构新颖：书中考虑到教师教学的需要，既对教学中的重点和难点及抽象概念进行大量的实例分析，又用数形结合的方法，穿插算法、解法、证法的讲解，并用较大的篇幅，对教法进行深层分析. 与此同时，还介绍了重要概念和学科的历史渊源和现代发展，对教学内容进行优化、整合、重组，建立了新的课程体系，一方面给教师在教学时留有选择的余地，另一方面，也给读者留有自己动手、自己设计实验和自我学习的空间.

2）内容现代化：书中在第 6 章数学前沿专题介绍中，专门介绍了风险与决策、分形理论及其发展历程、混沌的特征与概念，用通俗和深入浅出的语言，展示了现代数学学科的主要内容和新发展，以开阔读者的视野. 此外，在各章中，都渗透了现代的教学内容和思想.

3）注重应用：本书不仅在引入新概念时注重概念的应用背景和应用分析，而且使用了很大的篇幅详细论述数学概念在日常生活、自然科学中的应用，并特别注重一些重要数学概念，例如导数、积分、概率、随机现象和随机变量的概念以及数据处理的方法在国民经济领域、企业管理、市场营销、证券股指、物价指数和其他现代生活中广泛而深入的应用，使读者体会到数学正在渗透到我们生活的方方面面和其

巨大的魅力. 与此同时,我们也注意到现代数学在初等数学中的渗透和应用. 在博弈论这一部分,用详实通俗的应用实例,论述了博弈论的深刻内涵以及它的新进展、新研究、新应用,给人以耳目一新的感觉.

4)体现现代数学技术:本书引用了在世界各地比较流行的 Mathematica 系统作为平台,有利于学生把系统学习和运用数学知识,及时地探索研究以及提高计算机技能和编程能力,这三者有机地结合起来.

5)强调创新:本书针对书中教学的内容,配以丰富详实的实验设计内容. 在实验设计中,不仅步步深入分层次地对书中最重要的概念和内容进行图文并茂的设计,以及进行动画设计,使一些抽象的数学概念变得易于理解,并且用了较大的篇幅,给出有利于读者进行探索和研究的实验设计,以有利于同学们在学习中培养创新意识.

总之,读者在本书的学习和实验设计中,通过观察、联想、类比和自己亲自练习及操作实验,可以提高五个能力,即:1)提高掌握数学知识、理解抽象概念的能力;2)独立思考和动手能力;3)运用数学知识的能力;4)计算机编程能力;5)创新和探索新知识的能力.

参加本书编写的作者有:田长生、徐庆和、魏凤荣、邬振明、丘文、龙永安、林智勇等老师.

在本书出版过程中,张景中院士表示了对本书出版的关注和支持. 科学出版社李鹏奇编辑花费了大量的精力,进行仔细地编辑和校对工作. 孙熙椿老师及周四清、田冶等同志对本书的出版做了很多工作. 在此,我们向所有对本书出版做出贡献和支持的单位和个人,谨表示诚挚的谢意!

此外,由于参考文献较多,受篇幅限制,不能一一列出,对参考文献未列出的作者,也在这里表示我们的谢意!

由于编者的学识水平有限,时间有限,书中不妥之处在所难免,衷心希望读者在使用中指正.

编　者

2006 年 7 月

目　　录

第 1 章　概率与统计概述

1.1　概率与统计溯源

概率论有着悠久的历史,它的起源与赌博问题有关.

据文献记载,最早提出这类问题的是意大利数学家帕西奥里(L. Pacioli). 他在 1494 年发表的数学巨著《算术、几何、比和比例摘要》中提出了这样一个问题:两个赌徒进行赌博,规定赢 6 次才算赢,而由于某种原因两个赌徒在一个赢 5 次另一个赢 2 次的情况下中断比赛,问这时应如何分配总的赌金? 帕西奥里的答案是按 5 : 2的比例给两个赌徒分配赌金,即按已赢的次数分配. 这看起来似乎是合理的,但是,如果考虑更一般的情况则会发现问题. 例如,若规定赢 16 次才算赢,现两个赌徒分别赢了 15 次与 12 次而比赛中断,那么按照帕西奥里的分成法,两人所分的赌金相差不多,但其中一个只要再赢一次就能得到全部赌金,而另一个则还得连赢 4 次. 这个分成法就显得不那么合理了. 到底应该怎样分成,帕西奥里没有正确地解决这个问题,但是他所提出的问题引起了人们的思索.

过了许多年之后,意大利的另一位数学家卡丹(J. Cardan)讨论了类似的赌博问题. 他去世后的 1663 年出版了他的著作《赌博之书》. 卡丹意识到了需要分析的不是已赌过的次数,而是剩下的次数. 在帕西奥里问题中,一个赌徒只需赢 1 次就可以得到全部赌金,而另一个需要连赢 4 次,因此,以后的赌博只有 5 种可能的结果,即第一个赌徒赢第 1 次或赢第 2 次或赢第 3 次或赢第 4 次,或者完全输掉. 可惜,卡丹思路对了,给出的计算方法却不正确. 另外,卡丹还对掷两颗骰子和掷三颗骰子时一切可能方法中有多少方法能得到某一点数的问题进行了研究. 与此同时,与卡丹在三次方程解法发明权上发生过争执的意大利数学家泰塔格利亚(N. Tartaglia)也曾讨论过类似的赌博问题. 尽管他们都没有正确地计算出合理分成的比例,但是,他们的工作却为概率论的方法做了早期的积累.

真正促使概率论产生的强大动力来自于社会实际问题的需要. 文艺复兴以后,随着生产的发展,社会的进步,尤其是商业发展的需要,航海事业也随之快速发展起来,意大利开始出现了海上保险业务. 到了 16 世纪末,欧洲许多国家又把海上保险业务扩展到了其他工商业. 保险问题普遍带有随机现象的色彩,为了既保证保险业主赢利,同时又能使人们愿意投保,就需要对保险问题中的大量随机现象的规律性进行分析研究,从而建立保险业的一般理论. 于是,概率论产生的时机来到了. 然而,上述实际问题所包含的随机现象,常常被太多的错综复杂的因素所干扰,使它

不是完全地呈现出“自然的随机状态”. 因此,像类似于赌博问题这样的随机性游戏模型成为简单的、便于研究的对象. 从这些简单问题中分析某些规律性的东西,例如“多次试验中的频率稳定性”等随机现象的规律性,成为当时人们所关注的问题.

17 世纪中叶,法国数学家帕斯卡(B. Pascal)、费马(P. de Fermat)和荷兰的数学家惠更斯(C. Huygens)基于排列组合方法,研究了一些较复杂的赌博问题,为概率论的产生做了奠基性的工作.

费马与帕斯卡都是法国“Mersenne 学院”的成员,俩人常常共同研究一些数学上的疑难问题,交流学术思想. 1654 年,帕斯卡的一个朋友梅累(C. Mere)向他提出一个类似“帕西奥里问题”的问题:

两个赌徒下的赌注相等,相约赌若干局,谁先赢 s 局谁就得到这笔赌金,由于某种原因赌博不能进行到底,赌博中止时一个赌徒还少赢一盘获胜,而另一个则还少赢两盘,问此时赌金应如何分成?

梅累的这个问题使帕斯卡思索很久而不得其解,于是他将问题提交给了费马. 此后两个人开始了频繁的通信讨论,并且就此开始了有关概率论和组合论的研究.

帕斯卡致费马的第一封信亡失了,现在无法得知帕斯卡在解梅累的这个问题中所用的方法. 但从他们留下的其他书信中仍然可看出他们解决这个问题的端倪. 虽然现在看来,这个问题已经不是十分复杂,但是,了解一下他们当时的讨论过程,仍然具有一定的意义.

费马一开始就想到利用组合来解概率问题,这是他的第一个特点,第二个特点是他十分重视区别各种概型,尤其是区别独立事件概型与条件事件概型. 他在 1654 年 7 月给帕斯卡的一封信中提出这样的问题:如果在赌博中某一赌徒轮到他掷第一次时自愿不掷,因为骰子是均匀的正六面体,要掷得某一规定的点数而获胜的可能性是 1/6,因此,若对方认可的话,这个赌徒应得到赌金的 1/6,接着他又不掷第二次,那么,他应得到经第一轮后剩下的赌金的 1/6 即原全部赌金的$(1/6)\times(5/6)=5/36$. 依次类推,如果他不掷第三次,他应得第二轮后剩下的赌金的 1/6,即原全部赌金的$(1/6)\times(5/6-5/36)=25/216$. 如果不掷第四次,他应将第三轮后剩下的赌金的 1/6,即原全部赌金的$(1/6)\times(25/36-25/216)=125/1296$. 在这封信中,费马还指出了帕斯卡在这个问题上所犯的错误. 起初,帕斯卡认为某一赌徒掷了前三次且都失利的情况下不掷第四次,也应得 125/1296 的总赌金. 费马指出,这种情况下,前几轮过后,整个赌金并没有发生变化,他第四次掷,仍然应得总赌金的 1/6. 而且,他还认为,每掷一次获胜的可能总是 1/6,这是独立的,不受前后是否获胜的影响.

1654 年 7 月 29 日,帕斯卡在得到费马来信后的第二天,给费马回了一封长信. 其中用纯算术方法讨论了梅累问题:两赌徒各出赌金 32,约定赢三局为胜. 帕斯卡详细地讨论了各种可能的情形:

(1) 假设甲赢了两局，乙赢了一局：那么甲再胜一局赢得全部赌金 64，也可能甲输掉这一局，形成 2∶2 的局面，在这种情况下中止赌博，他们应平分赌金. 于是，帕斯卡认为，在甲赢两局，乙赢一局的情况下中止，甲应该说：无论以上两类情形哪一个发生，自己首先应得一半的赌金，至于剩下的一半，两人获胜机会均等，应再平分这一半，即自己应得 32+16=48；

(2) 假设甲赢两局，乙未赢：那么甲再赢一局即赢得全部赌金 64；也可能甲输掉这一局，形成 2∶1 的局面，这是(1)所讨论的情形，帕斯卡认为，无论如何甲首先赢得 3/4 的赌金，剩下的 1/4，又因甲乙获胜机会均等，应再平分，故甲应得全部赌金的 7/8 即 56；

(3) 假设甲赢了一局，乙未赢：那么甲再赢一局，形成 2∶0 的局面，这是(2)所讨论的情形. 因此，如果乙胜了这一局，形成 1∶1 的局面，此时中止，甲乙两人应平分赌金. 所以，帕斯卡认为，无论如何甲首先应得 1/2 的赌金，乙应得 1/8 的赌金. 至于剩下的，他们获胜机会均等，应平分，于是甲得 32+12=44，乙得 8+12=20.

1654 年 8 月 24 日，帕斯卡又致信费马. 他首先对费马用组合来求解的方法表示钦佩. 接着采用了一种新的方法来讨论在甲差 2 局胜、乙差 3 局胜时如何将赌金分成的问题. 他的方法是：因为无论如何，再有 4 局就完全可以确定胜负，所以可以把玩 4 局转化为玩 1 局而同时掷 4 颗骰子的问题.

设对甲有利的骰面为 a，对乙有利的骰面为 b，那么会出现 16 种可能的排列：

a a a a	b a a b	b a a a	b b b b
a a a b	a a b b	b a b a	a b b b
a a b a	a b b a	a b a b	b a b b
a b a a	b b a a	b b b a	b b a b

显然，其中有 11 种情况有利于甲，5 种情况有利于乙，所以，赌金分成应按 11∶5 分配. 帕斯卡的这个方法比起他原来纯算术的方法有很大的进步，并且在某种程度上可以说是 n 重伯努利试验思想的雏形.

在此之后，帕斯卡还对三个赌徒的情形也做了讨论，即：三个赌徒中，若甲差一局胜，乙和丙都差两局胜，那么如何分配赌金. 他也用上述方法，得出 16∶5.5∶5.5 的结论，这与利用算术方法计算出来的比例 17∶5∶5 不同. 费马看到这个结果，立即回信给帕斯卡，指出分成比例应为 17∶5∶5. 1654 年 9 月 25 日，费马致帕斯卡一封极为重要的信，深入地论述了这个问题.

费马将这类问题推广到四个赌徒的情形，即甲差一局胜，乙、丙、丁差两局胜. 那么就会出现 81 种可能的排列，赌金应按 51∶15∶15∶15 来分成，即 17∶5∶5∶5 的比例. 可以证明，无论有多少赌徒，只要甲差一局胜而其他人差两局胜，则总是应按 17∶5∶5∶5 …∶5 的比例来分成. 之所以会如此，以三个赌徒为例，此时设对各赌徒有利的骰面分别为 a、b、c，则排列 acc 仅有利于甲，这是由于一旦 a 出现，甲

就赢了.同理,cca 也仅有利于丙.这里有顺序问题.即在三个赌徒参加的问题中,甲可能在第一局、或在第二局、或在第三局上赢,也可能不赢.如果用一个三面的骰子,可以简化问题的说明,这就是说,第一局他仅需要三面中的一面,所以他取胜的可能性是 1/3.如果第二局要进行,那么有 $3\times3=9$ 种不同的排列出现,但在第一局失利的情况下仅有两种排列对他有利,他分别在乙、丙赢第一局后他赢第二局,因此他获胜的可能性是 2/9.如果第三局要进行,那么有 $3\times3\times3=27$ 种不同的排列出现,在这种情况下,也只有两种排列对他有利,即赢的顺序是乙-丙-甲或丙-乙-甲,因此,他获胜的可能性为 2/27,由此推得,甲赢得赌金的机会应为 $1/3+2/9+2/27$.乙、丙获胜的可能性也同理可算.这里,费马开始更多地应用组合方法来分析概率问题.

在帕斯卡与费马的讨论中,虽然没有确切定义概率的概念,但是,他们在估计赌徒获胜的可能性时,考虑到了利用有利情形数与所有可能情形数之比,这实质上就是概率的概念,而在计算赌徒获得赌金的问题上所用的基本原则,本质上就是数学期望的概念.

荷兰数学家惠更斯也独立地研究了这一问题,并且于 1657 年发表了概率论最早的论著《论赌博中的计算》.由于帕斯卡、费马与惠更斯三人的工作,概率、数学期望这样的基本概念已经初步建立,并得到了相应的一些基本性质和计算方法,可以认为,概率论已经开始初步形成.

17、18 世纪之交,有许多数学家参与研究了概率问题,也有不少有关概率论方面的著作问世.其中影响最大的是瑞士数学家伯努利(J. I. Bernoulli)的名著《推想的艺术》.这部著作于 1713 年发表,它被誉为概率论方面的第一部重要著作.在这部著作中,伯努利推广了组合理论,用组合公式证明了帕斯卡曾提出过的、n 为正整数时的二项式定理.他得到了现在仍被称为“伯努利定理”的重要结论:若 p 是出现某一事件的概率,q 是不出现该事件的概率,则在 n 次独立试验中该事件至少出现 m 次的概率,等于二项式 $(p+q)^n$ 的展开式中从 p^n 项到包括 p^mq^{n-m} 项为止的各项之和.这是概率论最重要的定律之一“伯努利大数定律”的最早形式,伯努利大数定律是概率论中的第一个基本极限定理,由于它的重要地位,1913 年 12 月彼得格勒科学院举行庆祝会,纪念“大数定律”诞生 200 周年.从此,建立在经验分析基础之上的对于频率稳定性的估计开始理论化,概率论从对特殊问题的求解,发展到了理论概括的阶段.

伯努利之后,法国出生的数学家棣莫弗(A. de Moivre)对概率论的发展又做了巨大的推进,于 1718 年发表了概率论历史上的另一部重要著作《机会的学说》.书中提出了概率乘法法则,以及“正态分布”和“正态分布率”的概念,得到了概率论第二个基本极限定理即“中心极限定理”的原始形式,它是后来称之为“棣莫弗-拉普拉斯定理”的一个特例.

英国数学家辛普森(T. Simpson)于 1740 年出版了《论机会的性质与规律》,这也是一部重要的概率论著作. 20 年后法国数学家蒲丰(C. de Buffon)的《偶然性的算术试验》则把概率和几何结合起来,开始了几何概率的研究,引入了著名的"蒲丰投针问题":将一根长 $2l$ 的针,任意投在画有许多平行线的平面上,平行线间相距 $2a(a>l)$. 则可以证明针与任一直线相交的概率为 $p=\frac{2l}{a\pi}$. 通过用实验方法确定 p,即可以求得 π 的近似值. "蒲丰投针问题"给人的另外一个重要启示是,对人们感兴趣的某个量(例如 π),设计适当的随机试验,然后根据试验的结果去确定这个量. 现在随着电子计算机的发展,已经按照这个思路建立起一类新的计算方法,称为蒙特卡罗(Monte-Carlo)方法.

概率论方法在 18 世纪得到很大的发展,在应用方面也发挥了广泛的重要作用. 例如:丹尼尔·伯努利(D. Bernoulli)根据大量的统计资料,作出了种牛痘能延长人类平均寿命 3 年的结论,从而消除了一些人对其副作用的恐惧和怀疑;欧拉在《关于死亡率和人类增长问题的研究》、《关于孤儿保险》等文章中将概率论应用于人口统计和保险;泊松在《打靶概率论研究报告》中将概率论应用于射击问题的研究,等等.

19 世纪初,概率论开始朝着理论系统化的方向发展.

法国数学家拉普拉斯(P. S. de Laplace)可以说是概率论系统理论最卓越的创建者,他出版了好几部概率论著作,其中 1812 年出版的《分析概率论》是古典概率论系统理论的经典之作. 他全面总结了当时概率论的研究结果,并予以严密而系统的表述,他成功地证明了中心极限定理,建立了观察误差理论和最小二乘法,有效地发展了概率论在实际中的一系列重大应用.

拉普拉斯的观察误差理论和最小二乘法,经高斯(G. F. Gauss)的努力,又极大地向前推进了一步,高斯正式地奠定了最小二乘法和误差理论的基础. 泊松(Poisson)引入了概率论中重要的"泊松分布",推广了大数定律. 俄国数学家切比雪夫引入了著名的"切比雪夫不等式",并以这个不等式证明了概率论中两个最基本的结果"大数定律"和"中心极限定理".

概率论形成系统理论之后,它的广泛应用进一步迅速展开. 在 18、19 世纪概率成为风行的热门学科,几乎在所有的科学领域,不仅自然科学领域,而且包括神学在内的社会科学领域都企图借助于概率论去解决问题. 尤其是人们在对随机现象研究中试图得到可靠信息的想法,促成了一门新的学科"数理统计学"的建立. 但是,应用的广泛性又使得人们不得不重新对概率的基本理论进行检查,与此同时,对拉普拉斯的先验概率的定义,在应用中也提出了种种争论. 因此,概率论需要有一个牢固的逻辑基础,成为了当时必须解决的一个问题.

1917 年苏联的数学家伯恩斯坦(С. Н. Бернштейн)首先提出概率论的公理体

系. 1933 年苏联的另一位数学家柯尔莫哥洛夫(A. H. Колмогоров)以更完善的勒贝格测度论为基础,给出了概率论的公理体系. 这种体系不仅使现代意义的概率论的理论臻于严密完备,而且对论述无限随机试验序列或一般的随机过程给出了足够的逻辑基础. 因此,现代几乎所有的概率论的结论都是用测度论的观点来加以论述.

概率论传入我国,最早是 1896 年我国晚清数学家华蘅芳(1833～1902)在英国传教士付兰雅(J. Fryer)的协助下译出了第一本书,名为《决疑数学》. 后来,“Probability”又被译为“可遇率”、“或是率”、“公算”、“适遇”、“可能率”、“机率”、“盖然率”、“结率”等等. 1935 年《数学名词》定为“几率”或“概率”. 1956 年《数学名词》仍然把“概率”、“机率”并用,1964 年《数学名词补编》开始确定用“概率”,到 1974 年《英汉数学词汇》正式将“Probability”定为“概率”. 我国的数学工作者对概率较完整系统的学习和研究是从 20 世纪初才开始的,并且取得了一系列的成果.

相对于概率论,关于数理统计学的早期发展情况,学者们很少论述,可资参考的文献很少.

数理统计学是一门较年轻的学科,它主要的发展是从 20 世纪初开始,是随着概率论的发展而发展起来的. 在早期发展中,起领导作用的是以 R. A. Fisher 和皮尔逊(K. Pearson)为首的英国数理统计学派. 特别是 Fisher,在数理统计学的发展中起了独特的作用. 目前许多常用的统计方法以及教科书中的内容,都与他的名字有关. 其他一些著名的学者,如 W. S. Gosset、J. Neyman、E. S. Pearson、A. Wald 以及我国的许宝禄教授等,都做出了根本性的贡献. 他们的工作奠定了许多统计学分支的基础,提出了一系列有重要应用价值的统计方法以及一系列基本概念和重要理论问题. 有一种意见认为,瑞典统计学家克拉默(H. Cramer)在 1946 年发表的著作 *Mathematical Methods of Statistics*(《统计学的数学方法》)可以作为数理统计学进入成熟阶段的标志. 这样说也许并不过分,这是因为,虽然在此之前已出现了一些重要的统计学著作,特别是高斯和勒让德(A. M. Legendre)关于观测数据的误差分析与最小二乘法理论. 但第一次用严谨的且比较系统的数学方法总结到那时为止数理统计学的主要成就的,还是要推克拉默的这一部著作.

收集和记录数据的活动,在人类历史上自古就有. 翻开历史典籍,无论中外,都可以看到有很多关于人口、粮食、牲畜以及地震、洪水等自然灾害的记录. 实际上,在英语中,Statistics(统计学)一词源出于 State(国家),意指国家收集的国情资料. 另一方面,在历史上,也有许多人为研究某些问题而进行特定的观察试验,收集分析资料. 但这些情况,终究还不能认为是数理统计这门学科已经建立的标志,它不是现代意义下的数理统计学. 因为当时的许多工作,只停留在收集数据或对其进行一些简单的加工整理. 即使有时也作出了某种超出已有数据范围之外的推断,也只是基于一种朴素的直观想法,而未能把问题模型化而使之带有普遍意义,更谈不上

建立必要的基本概念和一般理论了. 这种情况延续了许多年，这是因为，没有一定的数学工具特别是概率论的发展，无法建立现代意义下的数理统计学. 也因为应用方面的要求还没有达到那么迫切，这就是说，当时的社会需求还不足以构成一股强大的推动力. 直到 19 世纪后半期至 20 世纪初，情况才起了较大的变化.

关于数理统计学正式诞生的时间，学者们提出过一些意见，但尚无定论. 有的认为这个时间"不早于 1850 年"，有的认为 20 世纪初 K. Pearson 关于统计量的极限分布的论文可以作为一个标志，也有人认为，直到 1922 年 Fisher 关于统计学的数学基础的一篇著名的论文发表，数理统计学才正式诞生. 这个时间可能失之过晚，不过，Fisher 的这篇论文首次概括了统计理论的现状和存在的问题，并提出了数理统计学的三个任务. 文中的观点的主要部分到现在仍基本有效. 因之，Fisher 这项工作无疑是数理统计学建立过程中的一个里程碑.

无论如何，至迟到 20 世纪 20 年代，数理统计学的基本理论和方法已建立起来. 20 世纪前 40 年有了迅速而全面的发展，到 40 年代时，已基本形成为一个成熟的数学分支.

20 世纪前 40 年是数理统计学辉煌发展的时期. 但最近的几十年，数理统计学的发展也十分显著. 许多此前已形成的统计分支，得到纵深的发展. 这不仅是由于二次大战后工农业生产和科学技术等方面迅速发展的需要，也由于电子计算机这一有力工具的出现. 许多统计方法的实施都涉及到大量的计算，在电子计算机得到广泛应用以前，这些方法的威力就难于发挥. 而在今天，许多已得到充分发展的统计方法，借助于电子计算机这一强有力的工具才真正在应用上发挥了巨大的作用.

1.2　概率与统计的作用与地位

目前，概率论与数理统计的理论与方法已经广泛地应用于自然科学、技术科学、社会科学以及人文科学的各个领域. 近年来，随着科学技术和社会生产力的迅速发展，它在经济、管理、工程、技术、生物、地质、气象等领域中发挥的作用越来越显著. 甚至可以说，几乎在人类活动的一切领域中都能程度不同地找到它的应用. 尤其是电子计算机的迅猛发展与普及，概率论与数理统计已成为处理信息、制定决策的一种重要理论和方法.

概率论与数理统计的理论与方法的应用之所以如此广泛，是因为实验是科学研究的根本方法，而随机性因素对实验结果的影响是无所不在的. 也正是由于这个原因，概率论与数理统计所起的作用是其他学科所不可替代的.

例如，在工农业生产中，一个常见的问题是：有一个(或多个)我们感兴趣的指标，如工业产品的某项质量指标，农业中的单位面积产量等. 客观上，一些因素对这个指标可能会产生影响，例如，工业生产中所用设备、原材料、配方和温度、压力及

反应时间等工艺因素，农业生产中所用种子品种、肥料类型和施放数量以及耕作方法等方面的因素. 为了得到最大的经济效益，需要了解这些因素对所感兴趣的指标起影响的具体情况：哪些因素是主要的，其影响有多大，因素与指标之间是否可建立某种数量上的联系等. 弄清楚了这些问题，就可以确定一组较好的生产条件. 这些都要通过做试验，就是把有关因素固定在某些水平上做试验，去观察感兴趣的指标值. 试验要经过精心的设计，而所得试验结果必然会受到大量随机因素的影响，这就必须用概率统计方法进行分析. 因此，随着近几十年来工农业生产的规模愈来愈大，数理统计学在这方面的应用也与日俱增.

概率统计方法在工业生产中所起到的另一个重要作用是：现代工业生产大多具有大批量生产及要求很高的可靠性的特点. 需要在连续生产的过程中进行质量控制，成批的产品在交付使用前要进行验收，由于检验可能带有“破坏性”或者有“经济性”、“时效性”等条件限制，这种验收一般不能是全面检验，而只能是抽样验收. 抽样具有随机性，这就需要根据数理统计学的原理，去制定适合各种满足要求的抽样验收方案. 另外，一个大型设备往往包含成千上万个元件. 由于元件数目很多，它们的寿命可视为随机的并且服从一定的概率分布规律. 整个设备的可靠性，与设备的结构及这种分布规律有关，因而可以用统计方法去估计. 为解决上面的这些问题，发展了一系列的统计方法，目前常提到的“统计质量管理”，就是由这些方法构成的.

概率统计方法在医药卫生领域中有着广泛的应用. 例如，治疗一种疾病的某种药物和治疗方法的效果，同样的药物和治疗方法，对不同的人产生的效果可能不尽相同，这里有随机性因素的影响. 要说明一种药物治疗某种疾病是有效（或无效）的，需要引用统计资料来说明. 这种资料的可信性，依赖于其数据的取得方法与使用的统计方法. 其他的方面，例如分析某种疾病的发生是否与特定因素有关（例如吸烟与患肺癌的关系），如果有关，其关系的密切程度如何？ 又如，在环境问题中，污染大气的许多有害成分对人体有何种程度的影响？ 也都带有明显的随机性，这类问题都需要利用数理统计的方法去研究，并且已经取得了许多好的成果.

数理统计方法在气象预报、地震和地质探矿等方面的应用也十分广泛. 在这个领域中，人们对事物的规律性认识还很不充分，表现在根据事物的现象去判断结果时，往往会出现判断失误. 使用统计分析方法可能有助于获得一些对潜在的规律性的认识，用以指导人们的行动. 不过，在人们对事物的规律性认识很不充分的情况下，一些起较大作用的系统性因素，尽管并不是随机的，但现在的人们还认识不到，只好当作随机性因素来处理. 这样，统计分析的精度或可靠性就往往还不尽人意.

在自然科学的研究过程中，概率统计的理论与方法也起到相当的作用. 自然科学的任务是揭示自然界的规律性. 一般是先根据若干观察或实验资料提出某种初步理论或假说，然后再通过各种途径通过实验去验证之. 一个好的统计方法有助于

提取观察和实验数据中带根本性的东西，即能够尽可能全面的获取实验数据所提供的事物的本质信息，因而有助于提出较正确的理论或假说. 在建立了一定的理论或假说后，数理统计方法又可以指导研究者如何去安排进一步的观察或实验，以使所得数据更有助于判定理论或假说是否正确. 数理统计学本身也提供了一些理论上完善的方法，以判断观察或实验数据与理论或假说的符合程度如何. 历史上一个著名的例子是遗传学中的 Mendal 定律. 这个根据观察资料提出的定律，经历了严格的统计检验. 数量遗传学的基本定律——Hardy-Weinberg 平衡定律，也是属于这种情况.

数理统计方法在社会、经济领域中也有很多应用. 在某些国家中，统计方法在这些方面的应用，比其在自然科学和技术领域中的应用更为显著. 统计方法在社会领域中的一项重要应用是抽样调查. 经验证明，经过精心设计和组织的抽样调查，其效果可以达到以至超过全面调查的水平. 在有些问题上，概率统计的作用甚至是其他方法所不可替代的. 例如，如果在一所学校中提出这样的问题：谁在考试中作过弊？可能不会得到真实准确的回答. 因为恐怕没有哪一个作过弊的人能够公开承认自己曾经作弊，显然，调查这种问题不应当采取直接询问的方式，即便采用不记名问卷调查，并且作出保密的承诺，但是问题太敏感了，被调查者仍然会存在种种顾虑. 在社会调查中，经常会遇到类似于这样的一些敏感问题或涉及个人隐私的问题. 例如：你认为你的班主任老师对工作负责任吗？你在初中阶段有过早恋行为吗？你信任你的老板吗？出于种种原因，对这一类问题，被调查者往往不愿意透露自己的某些真实思想及行为，因而不能对这种调查给予配合 . 对这一类问题，为了能够得到较为客观、真实的调查结果，通常采用的调查方法是所谓的“随机化问答技术”. 这一方法最早由 Stanley L. Warner 在 1965 年提出，较为有效地解决了这一类问题.

另外，对社会现象的研究有向定量化发展的趋势，例如人口学，确定一个合适的人口发展动态模型，需要掌握大量的观察资料，并使用包括统计方法在内的一些科学分析方法. 在经济科学中，定量化的趋势比其他社会科学部门更早更深. 早在 20 世纪二三十年代，时间序列的统计分析方法就用于市场预测. 现在，一系列的概率统计方法，都在经济科学中具有广泛的应用.

近年来，随着电子计算机技术的迅速发展和普及，为概率论与数理统计方法的广泛应用提供了强有力的支持，概率论与数理统计在向各个领域渗透，并产生了许多新的分支和边缘科学，例如生物统计、教育统计、心理统计、统计物理等. 不仅如此，概率论与数理统计还是许多新兴学科的理论基础，例如信息论、排队论、控制论、预测论、人工智能以及可靠性理论等.

概率论与数理统计作为理论严谨、应用广泛的数学分支正越来越受到人们的重视并发挥着重要的作用.

1.3 概率统计的思想方法与微积分的区别

作为一门研究随机现象的数学分支,概率统计与经典数学(例如微积分学)相比较,在思想方法与分析方法上有许多独特之处.微积分学主要研究的内容之一是函数理论,由函数定义可知,对自变量 x 的任意一个值,因变量 y 有确定的值与其对应,本质上,这就是对确定性现象的一种数学方法的描述,即在给定"自变量 x 取某值"的条件下,按照客观规律,因变量 y 有确定的结果与之对应.

但在实际问题中,这种"在一定条件下必然会产生某一确定结果"的情况往往只是相对而言,而其结果具有不确定性则更为常见,例如,人们熟知的重力加速度的测定,人们常取 9.8(单位:m/s^2)作为重力加速度的结果,有许多人也认为这是一定条件下必然产生的确定性结果,但在实际做测定试验时,只要要求一定的精度,多次试验的结果可能是不尽相同的,即试验的结果实际上具有不确定性.这种不确定性正是由于许多不为人们所知的随机性因素的影响,或是虽然人们知道是某种因素的影响,但以人们目前的科学技术水平上无法控制其对结果的影响所造成的.

对诸如此类的问题,概率统计有较好的解决方法,即若设 n 次观测结果为 X_1,X_2,$\cdots$,X_n,则取样本的均值

$$\overline{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n)=\frac{1}{n}\sum_{i=1}^{n}X_i$$

为最终的测定值.而大数定律则表明,当 n 充分大时,$\overline{X}$ 与所研究的理论值有大偏差的概率可以任意小.由此可见,从一定意义上说,在分析方法与处理方法上,可以认为概率统计对问题的研究更加细微.对于上面所提的问题,也希望读者在概率统计的学习中能够更深刻地理解样本均值的意义.

极限理论是微积分学中的另一个重要内容.以数列极限为例,$\lim\limits_{n\to\infty}a_n=A$ 的意义是大家熟知的,其描述性的说法是,当 $n\to\infty$时,a_n 无限向常数 A 趋近.基于此,在实际应用中,有近似式 $a_n\approx A$,只要 n 足够大,近似程度就可以足够的好.在概率论中,也常会碰到类似的问题,例如,当 n 足够大时,可以将 n 次独立试验中事件 B 发生的频率 $f_n(B)=\frac{m_B}{n}$作为事件 B 发生的概率 $P(B)=p$ 的近似,即有

$$\frac{m_B}{n}\approx p$$

但这两个近似式的意义是完全不同的,其分析方法与本质意义均有很大的差别,我们知道,前面的近似式:$a_n\approx A$,意味着只要 n 充分大,a_n 与常数 A 的偏差就可以任