面向21世纪课程教材

Textbook Series for 21st Century

应用概率统计

吴 坚 主编

高等教育出版社
HIGHER EDUCATION PRESS

面 向 2 1 世 纪 课 程 教 材
Textbook Series for 21st Century

应用概率统计

吴 坚 主编

高 等 教 育 出 版 社
HIGHER EDUCATION PRESS

内容提要

本书是教育部“高等教育面向21世纪教学内容和课程体系改革计划”的研究成果，是面向21世纪课程教材．本书突出随机数学思想，注重概率论与数理统计的应用背景和方法，讲授的内容分为上、下篇．上篇包括随机事件与概率、条件概率与独立性、随机变量及其分布、多维随机变量及其分布、随机变量的数字特征和极限定理；下篇包括数理统计的一些基本概念、参数估计、假设检验、方差分析和回归分析．

本书可作为高等院校农林类、水产类及工科各专业的本科生教材，也可作为农林科技人员的参考用书，部分内容可供有关专业研究生选用．

图书在版编目(CIP)数据

应用概率统计/吴坚主编．—北京：高等教育出版社，2002.1 (2003 重印)
高等院校农林水专业本科生教材
ISBN 7-04-010182-3

Ⅰ.应...　Ⅱ.吴...　Ⅲ.①概率论-高等学校-教材②数理统计-高等学校-教材　Ⅳ.021

中国版本图书馆 CIP 数据核字(2001)第071843号

应用概率统计
吴坚　主编

出版发行	高等教育出版社	购书热线	010-64054588
社　　址	北京市东城区沙滩后街55号	免费咨询	800-810-0598
邮政编码	100009	网　　址	http://www.hep.edu.cn
传　　真	010-64014048		http://www.hep.com.cn

经　　销　新华书店北京发行所
排　　版　高等教育出版社照排中心
印　　刷　北京中科印刷有限公司

开　　本	787×960　1/16	版　　次	2002年1月第1版
印　　张	23.75	印　　次	2003年2月第2次印刷
字　　数	440 000	定　　价	22.50元

总　　序

本系列教材是在原华东地区高等农林水院校数学系列课程教材《高等数学》、《线性代数》、《概率论与数理统计》试用后的基础上，按照教育部“高等农林教育面向21世纪教学内容和课程体系改革计划”中有关项目的要求重新编写的，是项目4-6的研究成果.该系列教材各册如下:《高等数学(Ⅰ)》、《高等数学(Ⅱ)》、《线性代数》、《应用概率统计》、《数学模型及其应用》和《数学实验》，适用于高等农林院校本科各专业.本系列教材编委会由杨崇瑞、王凯捷、吴坚、杨琪瑜、任明荣教授组成.

由南京农业大学牵头，东北林业大学、华中农业大学、西北农业大学合作主持，安徽农业大学、浙江农业大学、中国农业大学、河北农业大学、东北农业大学、黑龙江八一农垦大学、北京农学院、解放军农牧大学共14所院校参加的教育部教改项目“高等农林院校本科数学(含生物统计)系列课程教学内容和课程体系改革的研究与实践”，在各有关院校的重视和项目组成员的共同努力下，已通过验收，并获得了专家组的好评.这套系列教材就是该项目研究成果的一部分.此外，它也是在经过几次会议和有关专家讨论后，在高等教育出版社的大力支持下确定出版的.

该系列教材的选题主要遵从如下的课程体系设置:

- 大学农学类数学系列课程
  - 第一平台：基础型课程(面向100%学生)
    - 高等数学
    - 线性代数
    - 应用概率统计
    - 数学实验
  - 第二平台：扩展型课程(面向30%~50%学生)
    - 高等数学(续)
    - 微分方程(续)
    - 数值计算
    - 最优化方法
    - ⋮
  - 第三平台：研究型课程(面向10%~20%学生)
    - 数学模型及其应用
    - 多元统计分析
    - 模糊数学及其应用
    - 系统决策及其应用
    - 生态数学
    - ⋮

该系列教材的出版,首先要感谢参与编写的有关人员,感谢农业部数学课程教学指导委员会的关心和支持,特别致意的是这套系列教材的总设计、该项目组的总负责人杨崇瑞教授,他未能看到这套教材的出版就溘然长逝.现在,该系列教材的顺利出版,是对杨崇瑞先生的莫大慰藉.

编委会十分感谢中科院院士、复旦大学教授李大潜先生担任本系列教材的主审.

由于该系列教材还是一个教改尝试,不免存在一些问题和不足之处,诚恳期望本系列教材的使用者提出意见和建议,以利今后的进一步修改和完善.

编委会

2000 年 10 月 10 日

前　言

概率论与数理统计是高等农林院校本科数学教育中的一门主要课程,同时它的理论与方法也是学习和研究其他学科的重要基础,并且在农业、经济、管理、金融、工程技术以及社会科学诸多领域中有着广泛的应用.由于本书较为侧重概率统计理论与方法的应用背景和实用性,故取名为应用概率统计.虽然如此,它仍是一门数学基础课程.

本课程属于随机数学范畴,讲授研究随机现象规律性的概率论基础知识和以处理统计试验数据为主的数理统计基本理论和方法.它以微积分、线性代数等知识为基础,综合性和应用性强.在学习中要特别注意基本概念和基本方法,并结合书中的一些实际例子理解.该教材分为上、下两篇,上篇为概率论基础,下篇为数理统计,内容较以前丰富.本书适用于高等农林水院校各专业,课内学时为54～72学时的都可选用,课时更少的院校可适当选择内容讲授.本书还可用作农林科技人员的参考书,部分内容可供有关专业研究生选读.

由于这次面向21世纪高等农林院校本科数学系列课程教学内容和课程体系的改革内容中含生物统计,因此,作为一种教改尝试,本书的内容除了包含原概率论和数理统计的基本教学内容外,还适当包含了生物统计中的一些概率统计通用知识.按项目要求的课程体系设置,再加上授课学时数和篇幅这些限制,像统计推断中的协方差分析和多元统计分析,以及试验设计及分析等内容不在本书讨论范围内.本书在讲授内容的处理上,照顾到了数学知识的系统性,适当增加了一些近代概率统计知识,更加侧重于概率统计方法的应用,突出了随机数学思想,强调了数学的工具作用以及作为训练学生逻辑思维能力的重要载体作用.同时,还注意在一些内容上留有接口,引导学生的进一步学习和思考.有条件的院校可根据情况安排一些与本教材内容配套的概率统计实验课教学,进一步提高学生的应用能力.本书的例题和习题较为丰富,教学中可适当选择.

本书由安徽农业大学吴坚教授担任主编,南京农业大学徐凤君副教授、上海水产大学周亚虹担任副主编.参加编写人员为吴坚(上篇:概率论基础第一至六章,下篇:数理统计第七章,第八章§8.3,第九章§9.3,第十章§10.2、§10.3,第十一章§11.3、§11.4),徐凤君(第八章§8.1、§8.2,第九章§9.1、§9.2),南京林业大学童春发(第十章§10.1、§10.4)和周亚虹(第十一章§11.1、§11.2).全书习题及附表由吴坚整理,全书由吴坚修改统稿.

编者十分感谢中科院院士、复旦大学李大潜教授担任本书的主审工作.

由于编者水平所限,书中的缺点和谬误之处在所难免,恳请使用本书的教师和读者批评指正.

吴坚

2001年7月于合肥

目　录

上篇　概率论基础

下篇 数理统计

上篇　概率论基础

概率论是数学中一个有特色的分支，它是从数量侧面研究随机现象规律性的数学学科，并且与其他数学分支有着紧密的联系，是近代数学的重要组成部分．概率论的理论和方法已经广泛应用于工农业和科学技术的许多研究领域和实践之中，而且还在不断向众多学科渗透并与之结合发展．这些特点不仅使概率论成为数学中非常活跃的分支，而且也是近代科学技术发展的特征之一．

本篇主要介绍概率论中的一些基本理论和方法，并为以后的数理统计学习提供必要的预备知识．通过本篇的学习，读者应初步掌握处理随机现象的基本理论和解决有关实际问题的基本能力．

第一章　随机事件与概率

§1.1　概率论的现实背景

在自然界和各种科学研究中，有许多现象的规律是以下列形式表达的："只要条件 $\mathscr{C}$ 一经实现，则事件 A 必然发生(或必然不发生)．"例如："如果平面图形是三角形(条件 $\mathscr{C}$ 实现)，那么这个图形的内角和是 180°(事件 A 必然发生)"；又如"纯种紫花豌豆的后代开紫花"这一现象是一定会发生的，而"在 101 325 Pa 的大气压下，水加热至 100 ℃时不沸腾"和"同性电荷相互吸引"等种现象是必然不会发生的．以下就称在条件 $\mathscr{C}$ 下必然发生的事件 A 为**条件 $\mathscr{C}$ 下的必然事件**(或简称为**必然事件**)，称在条件 $\mathscr{C}$ 下必然不发生的事件 A 为**条件 $\mathscr{C}$ 下的不可能事件**(或简称**不可能事件**)．必然事件的反面是不可能事件，而不可能事件的反面就是必然事件．它们虽然形式相反，但两者的实质是相同的．所有这类现象我们称之为**确定性现象**(或**决定性现象**)，它广泛地存在于自然和社会现象中，许多数学分支研究的是确定性现象的数量规律．

与确定性现象不同，在自然和社会现象中还大量存在与之有着本质区别的

另一类现象——**随机现象：在条件$\mathscr{C}$下，事件 A 可能发生也可能不发生**. 我们以后就将在条件$\mathscr{C}$实现时，可能发生也可能不发生的事件称为**条件$\mathscr{C}$下的随机事件**（或简称**随机事件**）. 下面我们举一些例子.

例 1 从某工厂生产的某种产品中任意抽取的一件产品可能是合格品，也可能是不合格品. 在此例中，条件$\mathscr{C}$是"从某工厂生产的某种产品中任意抽取一件产品"，事件 A 是"抽取的产品是合格品"，显然 A 是条件$\mathscr{C}$下的随机事件.

许多实际问题与此类问题同类，例如：掷一枚均匀硬币观察所出现的面，"正面"或"反面"；种子发芽试验中一粒种子的"发芽"或"不发芽"；一次射击的"中"或"不中"等现象.

同样，将"从某工厂的某种产品中任意抽取几件产品"看作条件$\mathscr{C}$，则"其中恰有 $k(0\leqslant k\leqslant n)$件合格品"是在条件$\mathscr{C}$下的随机事件.

例 2 对于一个公用电话系统来说，电话局收到用户的呼叫是具有偶然性的. 因此"在时间间隔$(a,a+t]$内，电话局收到用户的 k 次呼叫"是"该公用电话系统在$(a,a+t]$内工作"的条件下的随机事件.

例 3 射击中，"弹着点距目标的偏差为 x"及"偏差在 x_1 与 x_2 之间"都是"进行一次射击"条件下的随机事件；如果在目标所在的平面上建立坐标系（不妨设目标位于坐标原点），则"弹着点(ξ,η)位于区域$\{(x,y)\mid x_1\leqslant x<x_2,y_2\leqslant y<y_2\}$内"也是"进行一次射击"条件下的随机事件.

例 4 植物病害流行研究中，以气传方式从菌源中心传播出的"一个病害孢子在时刻 t 的位置(ξ,η,ζ)位于区域$\{(x,y,z)\mid x_1\leqslant x<x_2,y_1\leqslant y<y_2,z_1\leqslant z<z_2\}$内"是随机事件.

例 5 种群动态研究中"某个野外动物种群进入某个划定区域的数量为 k"是随机事件.

总之，以上所列举现象的一个共同特点是：在基本条件不变的情况下，一系列的试验或观察会得到不同的结果. 换句话说，就个别的试验或观察而言，它会时而发生这种结果，时而发生那种结果，呈现出一种偶然性，这种现象就是随机现象. 在随机现象中通常关心的是在试验或观察中某个结果是否发生，这些结果就是随机事件. 以后我们用大写字母 $A,B,C,\cdots$ 表示随机事件.

另外，我们注意到随机事件的这种不确定性，不是事件本身不明确，而是发生的条件不充分，使得在条件与事件之间不能出现确定性的因果关系，从而在事件的发生与否上表现出不确定的性质，这种不确定性就是**随机性**. 在客观世界中，还存在另一些不确定现象，例如**模糊现象**，它所反映的事物在概念本身就是模糊的（即一个对象是否符合这个概念也难以确定，也就是由于外延模糊而带来的不确定性）. 例如天气预报中的"多云"，医疗诊断中的"食欲不振"、"头痛"，日

常生活中的“美与丑”、“冷与热”等.这种由客观事物差异的中间过渡的“不分明性”而产生的不确定性人们称之为**模糊性**.当然,客观世界中还有一些其他不确定现象,有些现象还可以是几种不确定性共存的情况,这些都是**非确定性现象**,而概率论研究和处理的对象是随机现象[①].

如何去研究随机事件?一种自然的想法就是像我们研究必然事件那样,去进一步寻求随机事件发生的条件.但是只要对上面举出的一些例子分析一下,就会发现这种做法几乎是不可能的,而且是不必要的.例如对于例2,要想事先找出“在$(a,a+t]$内,电话局收到用户的k次呼叫”这一随机事件发生的条件是不可能的.另外,电话局所关心的是大致有多少次呼叫,从而可以估计电话局的线路及接线员是否够用,是否有浪费现象,因而寻求事件确切发生的条件也是不必要的.这就产生了求这个随机事件出现的“可能性”的另一自然想法.

这种寻求随机事件“发生的可能性”的想法,在人们实践中是很自然的.例如,人们对于工厂生产好坏的判别标准之一是它的产品合格率,即产品中合格品个数与产品总数的比值;一种药物对某种疾病的疗效用治愈率,即使用过该药的患者中治愈人数与总人数之比来判断;对于射手的射击技术的判别标准之一是他的命中率,即射击命中次数与射击总数之比.合格率、治愈率及命中率都是人们所关心的随机事件发生的可能性的一个数量化描述.

从这些例子可以概括出频数与频率的概念如下:若在条件$\mathscr{C}$的n次实现下,事件A发生的次数是n_A,称它为A在$\mathscr{C}$的n次实现下的**频数**;而称$f_n(A)=\dfrac{n_A}{n}$为事件A在$\mathscr{C}$的n次实现之下的**频率**(简称频率),它反映了事件A在$\mathscr{C}$之下发生的可能性.当我们进一步考察A的频率$f_n(A)$时,发现$f_n(A)$对于条件$\mathscr{C}$的不同的n次实现一般来说是不同的.但是大量的实践表明,对于固定的随机事件A来说,当n(条件$\mathscr{C}$实现的次数)较大时,$f_n(A)$有经常接近于一个常数的趋势,并且当n越大时,接近的程度也就更为显著,接近的次数也越经常,即在大量的试验或观察下呈现出明显的规律性——**频率稳定性**.

例如,在掷一枚硬币时,既可能出现正面也可能出现反面,预先作出确定的判断是不可能的.但若硬币均匀,直观上出现正面与出现反面的机会应该相等,即“出现正面”的可能性应为1/2,下面让我们进行抛掷硬币的试验来验证这种频率稳定性.

将一枚均匀硬币掷50次,结果如表1.1所示.第一列是抛掷的次数,记作n,第二列是抛掷的结果,记作H(正面向上)和T(反面向上).到任何指定的一

---

① 近些年来,一些学者将概率方法适当改变以适应于人工智能、决策、不确定性推理诸方面,取得了不少成果.

次抛掷，出现 H 的总数列在第三列，并记作 n_H. 最后，出现 H 的频率，用 $f_n(H)=\frac{n_H}{n}$ 表示，列在第四列. $f_n(H)$ 画在图 1.1 之中.

表 1.1

n	H 或 T	n_H	$f_n(H)$	n	H 或 T	n_H	$f_n(H)$
1	H	1	1.0000	26	T	12	0.4615
2	H	2	1.0000	27	T	12	0.4444
3	H	3	1.0000	28	H	13	0.4643
4	T	3	0.7500	29	T	13	0.4483
5	H	4	0.8000	30	H	14	0.4667
6	H	5	0.8333	31	H	15	0.4839
7	H	6	0.8571	32	H	16	0.5000
8	T	6	0.7500	33	T	16	0.4848
9	T	6	0.6667	34	T	16	0.4706
10	H	7	0.7000	35	T	16	0.4571
11	H	8	0.7273	36	H	17	0.4722
12	H	9	0.7500	37	T	17	0.4595
13	T	9	0.6923	38	T	17	0.4474
14	H	10	0.7143	39	T	17	0.4359
15	T	10	0.6667	40	T	17	0.4250
16	T	10	0.6250	41	H	18	0.4390
17	T	10	0.5882	42	T	18	0.4286
18	T	10	0.5556	43	T	18	0.4186
19	T	10	0.5263	44	T	18	0.4091
20	T	10	0.5000	45	H	19	0.4222
21	H	11	0.5238	46	H	20	0.4348
22	T	11	0.5000	47	H	21	0.4468
23	T	11	0.4783	48	H	22	0.4583
24	H	12	0.5000	49	T	22	0.4490
25	T	12	0.4800	50	H	23	0.4600

此实验已由多人做过. 历史上，著名统计学家蒲丰(Comte de Buffon)和皮尔逊(Karl Pearson)曾进行过大量抛掷硬币的试验，所得结果如表 1.2 所示.

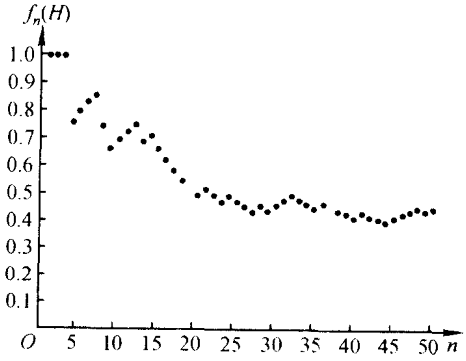

图 1.1

表 1.2

试验者	掷硬币次数	"出现正面"次数	频率 $f_n(H)$
蒲丰	4040	2048	0.5069
皮尔逊	12 000	6019	0.5016
皮尔逊	24 000	12 012	0.5005

从这些结果看,"出现正面"的频率总是在 1/2 附近波动,且抛掷次数越多,一般来说越接近于 1/2,这个常数 1/2 客观地反映了"出现正面"这一随机事件发生的可能性大小.读者可以思考,H 出现的频率是如何变得稳定并接近于 1/2.

同样,若多次测量同一个物体,其结果虽略有差异,但当测量次数增加时,就会越来越清楚地呈现一些规律性:测量值的平均值在某固定常数附近波动,诸测量值在此常数两边的分布呈现某种对称性.其他的随机现象,如灯泡寿命等,在进行多次观察或试验后,也都可以发现类似的规律性.

上述种种事实表明,随机现象有其偶然性的一面,也有其必然性的一面.这种必然性表现为大量观察或试验中随机事件发生的频率的稳定性,即一个随机事件发生的频率常在某个定值附近摆动,而且,试验次数越多,一般摆动越小.这种规律性我们称之为**统计规律性**.频率稳定的这种统计规律性说明随机事件发生的可能性大小是随机事件本身固有的,是不随人们意志而改变的一种客观属性,因此可以对它度量.

在寻求随机现象的规律时,常常采用模拟的方法.例如,代替掷一个均匀骰子,我们可以从一个恰好装有 6 个球(这 6 个球的大小、形状完全相同,分别标上 1,2,3,4,5,6 的号码)的袋中任意摸一个球.类似地,代替从一副纸牌中摸出一张,我们可以从一个装有 52 个球的袋子中任意摸出一个球(给球标上适当记号),以适当的方式用装球的袋子代替骰子、纸牌,只要

我们没有改变这些机会游戏中的机会，袋中摸球就能很好地模拟一些随机现象.如果你能编制计算机程序的话，也可以在计算机上模拟出你所要观察的某些随机现象.这样在短时间内就可以进行大量随机试验，更便于观察到随机现象的统计规律性.

对于一个随机事件 A，我们有理由认为在频率稳定性中 $f_n(A)$ 围绕其波动的那个与 A 有关的固定常数刻画了 A 的一个重要特性——事件 A 发生的可能性大小，记该数为 $P(A)$，我们就称 $P(A)$ 为**随机事件 A 的概率**.因此，概率度量了随机事件发生的可能性的大小.

实际上，上面提出的概率概念只是一个说明（或称为**概率的统计定义**），并不能作为严格的数学定义，这一严格定义可以在概率的公理化结构中得以解决.

对于随机现象，光讨论它可能出现什么结果，价值不大，而指出各种结果发生的可能性大小则具有很大意义.有了概率的概念就使我们能对随机现象进行定量研究，由此建立了一个新的数学分支——**概率论**.

人们早在 17 世纪中叶便开始了对随机现象的研究.在这些研究中建立了概率论的一些基本概念，如事件、概率、随机变量和数学期望等.当时研究的模型较为简单，就是现在通称的古典概型.

其后，随着生产实践的发展，特别是在射击理论、人寿保险、测量误差等工作中提出的一些概率问题，促使人们在概率论的极限理论方面进行深入研究.起初主要是对伯努利（Bernoulli）试验概型进行，之后则推广到更为一般的场合.极限理论的研究在 18 世纪和 19 世纪整整 200 年中成了概率论研究的中心课题.在 20 世纪初，由于新的更有力的数学方法的引入，这些问题得到了较好的解决.

虽然概率论历史悠久，但是它的严格的数学基础的建立以及理论研究和实际应用的极大发展却主要是 20 世纪的事情.在 19 世纪末以来，数学的各个分支广泛流行着一股公理化潮流，这个流派主张把最基本的假定公理化，其他结论则由此经过演绎导出.在这个背景下，1933 年，前苏联数学家科尔莫戈罗夫（А. Н. Колмогоров）提出了概率论公理化结构.这个结构综合了前人成果，明确定义了基本概念，使概率论成为严谨的数学分支，并对近几十年来概率论的迅速发展起了积极作用.

概率论现已大大地扩展了它的应用范围.特别在最近几十年，概率论的方法被引入工程技术、农业和社会经济学科.目前，概率论在无线电与自动控制、工厂产品的质量控制、农业试验、数量遗传、公用事业、金融保险和经济管理等方面都找到了重要应用.这些实际需要也有力地推动了概率论的新发展，有些还形成了边缘学科（如信息论、排队论等）.

在这个时期，由于生物学、农业试验和计算机等学科的推动，数理统计学也获得了很大的发展.它以概率论为理论基础又为概率论应用提供了有力的工具和新的课题，两者互相推动，迅速发展，成为相得益彰的姊妹学科.而概率论本身的研究则转入以随机过程为中心课题，取得了许多理论和应用上都有重要价值的研究结果.