



高等理工院校数学基础教材

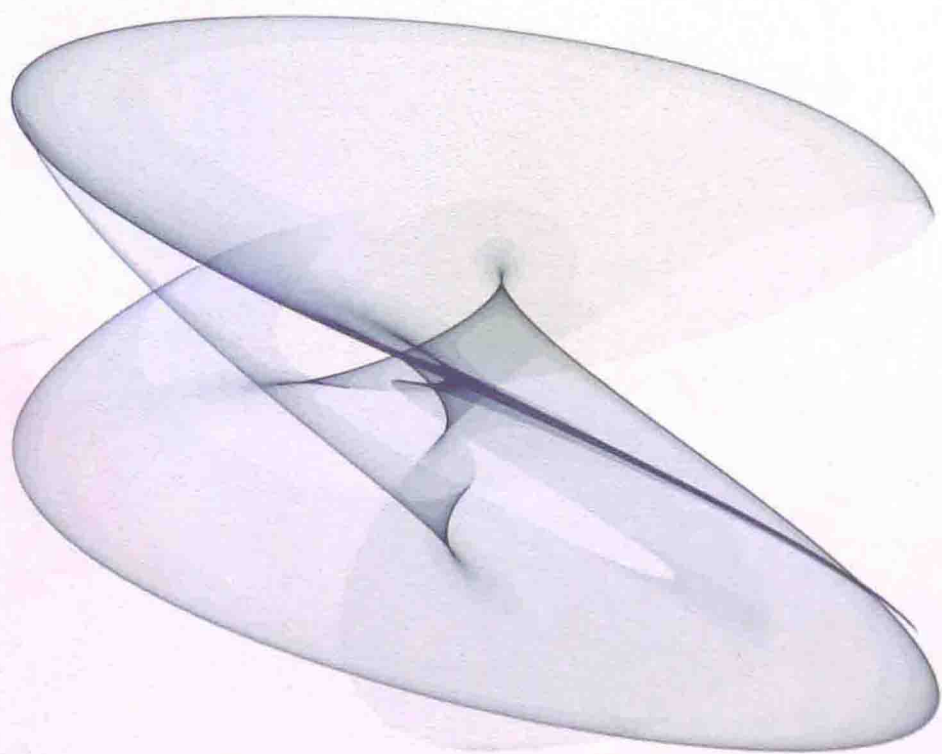
GAODENG LIGONG YUANXIAO
SHUXUE JICHU JIAOCAI

概率论与数理统计教程

GAILÜLUN YU SHULI TONGJI JIAOCHENG

第2版

汪忠志 主编



中国科学技术大学出版社



高等理工院校数学基础教材

概率论与数理统计教程

第2版

主 编 汪忠志

编写组 范爱华 张敬和 张永进

胡 萍 胡姿岚 吴绍凤

中国科学技术大学出版社

内 容 简 介

本书在系统介绍概率统计的基本理论和方法的同时,尽量收集来自社会科学、工程技术、医学、经济与金融和自然科学等领域的实际问题,进而讲解概率统计学是如何解决这些问题的.全书共分10章,内容包括:随机事件与概率,一维随机变量及其分布,多维随机变量及其分布,随机变量的数字特征,大数定律与中心极限定理,数理统计学简介,参数估计,假设检验,方差分析和回归分析简介,R语言简介等.

本书可作为高等院校理工科非数学专业的教材,也可供有关人员和教师参考.

图书在版编目(CIP)数据

概率论与数理统计教程/汪忠志主编.—2版.—合肥:中国科学技术大学出版社,2014.8

ISBN 978-7-312-03537-1

I. 概… II. 汪… III. ① 概率论—高等学校—教材 ② 数理统计—高等学校—教材 IV. O21

中国版本图书馆 CIP 数据核字(2014)第 152749 号

出版 中国科学技术大学出版社

安徽省合肥市金寨路 96 号,230026

<http://press.ustc.edu.cn>

印刷 合肥现代印务有限公司

发行 中国科学技术大学出版社

经销 全国新华书店

开本 710 mm×960 mm 1/16

印张 21

字数 399 千

版次 2010 年 8 月第 1 版 2014 年 8 月第 2 版

印次 2014 年 8 月第 3 次印刷

定价 33.00 元

第 2 版前言

本书第 1 版自 2010 年 8 月出版以来,已经历了 4 年的教学实践.这次我们根据在教学过程中积累的第一手材料,并且听取多次使用本教材的同事们的宝贵意见,对第 1 版的一些内容进行了必要的修正、删减或增补.另外,作为一门紧密联系实际课程,我们鼓励学生利用软件来解题.为此,根据当前国内外该课程教学的发展动向,由吴绍凤老师撰写了“R 语言简介”这一章,这样就使得本书更加完整.

参加第 2 版编写修订工作的有安徽工业大学工程数学教研室范爱华、张敬和、胡萍、张永进、胡姿岚、吴绍凤等.信息与计算科学教研室侯为根老师绘制了全部插图.全书最后由汪忠志统稿.

由于我们水平有限,书中一定还有不少缺点和错误,欢迎读者和专家批评指正.

最后我们要感谢中国科学技术大学出版社对本书出版的支持.

作 者
2014 年 7 月

前 言

概率论与数理统计作为现代数学的重要分支,在自然科学、社会科学和工程技术的各个领域有着极其广泛的应用,特别是近 20 多年来,随着计算机的迅速普及和数学软件的大量出现,其应用得到长足的发展.本书是按教育部颁布的高等学校理工科类专业核心课程“概率论与数理统计”教学大纲并参考近些年硕士研究生入学统一考试数学考试大纲编写而成的.

本书的大部分内容在安徽工业大学理工类专业课上讲授过多次.我们写作的指导思想是:考虑到初学者往往对本课程中的一些重要概念的实质领会感到困难,在编写过程中,尽量注意贯彻由浅入深、循序渐进和融会贯通的原则,力求既注重基本概念、基本理论和基本方法的阐述,又注重学生基本运算能力的训练和分析问题、解决问题能力的培养.

概率论与数理统计和现实生活非常贴近.本书在系统介绍概率统计的基本理论和方法的同时,尽量收集来自社会科学、工程技术、医学、经济与金融、自然科学等领域的实际问题,进而讲解概率统计学是如何解决这些问题的,使本书更具有时代气息.由此,学生不仅可以感受到概率统计的实用性,还会促使他们联想本专业的问题,激发学习和研究的兴趣.

概率论与数理统计是大学随机数学最重要的基础课,在学生数学能力培养中占有十分重要的地位,但由于它与学生比较习惯的连续量数学(如微积分)有不同特点,学生在学习时往往感到困难.本书针对学生学习的难点所在,从直观分析入手逐步过渡到严格的数学叙述,并通过大量的应用性问题,启发学生由浅入深,由表及里地掌握重要概念和方法,最终学会正确应用,从而有效地化解学生学习中的难点,提高教学效果.

为了尽可能满足不同读者的需要,本书有一些楷体字内容或带 * 号的内容和习题供读者选用,在学时有限时可跳过这部分内容,并不影响本书的体系.本书编写了大量例题,有些例题是我们在日常教学和考研辅导中收集起来的,有部分例题为了体现其典型性引自他人著作.在此,我们谨致谢意.

本书讲授 64 课时较为合适,能有多媒体设备的配合,教学将会更为方便、有效.

本书为安徽工业大学“十五”规划教材及安徽省教育厅重点建设学科的子项目之一.在写作过程中,得到了安徽工业大学教务处、数理学院和应用教学系领导的大力支持.陈松林、徐龙封两位教授在百忙之中认真审阅了教材初稿,提出了不少中肯的意见.在此基础上我们对教材做了认真修改,若现在书中仍有不妥之处,责任当由笔者自负.尽管作者倾尽全力,但囿于水平,书中的错漏与不妥之处在所难免,因此,作者殷切希望广大教师和同学不吝赐教,以期改正.

本书第 1~2 章由范爱华编写,第 3 章、第 7 章由张敬和编写,第 4 章、第 6 章由陈文波编写,第 5 章、第 8 章由汪忠志编写,第 9 章由张永进和胡姿岚编写,最后由汪忠志统稿.

我们感谢多年来担任本课程的授课老师,在我们致力于不断改进本书的过程中他们提出了许多有益的建议,他们是胡萍、花春、罗冬梅、李霞、李孜、宋静、王军秀、王二红、吴绍凤、许武玲、徐伟亮、闫志莲、张世涛、张杰、智丽萍等.

作 者

2010 年 8 月

目 次

第 2 版前言	(i)
前言	(iii)
第 1 章 随机事件与概率	(1)
1.1 基本概念	(1)
1.1.1 随机现象	(1)
1.1.2 随机现象的统计规律性	(1)
1.1.3 样本空间	(3)
1.1.4 随机事件及其运算	(3)
1.2 随机事件的概率	(5)
1.2.1 概率和频率	(5)
1.2.2 组合记数	(6)
1.2.3 古典概率	(7)
1.2.4 几何概率	(9)
1.2.5 主观概率	(10)
1.3 概率的定义与性质	(11)
1.3.1 概率的公理化定义	(11)
1.3.2 概率的基本性质	(11)
1.4 条件概率	(13)
1.4.1 引例	(13)
1.4.2 条件概率的定义	(14)
1.4.3 条件概率的性质	(15)
1.4.4 乘法公式	(16)
1.4.5 全概率公式	(16)
1.4.6 贝叶斯公式	(19)
1.5 事件的独立性与相关性	(22)

1.5.1 两个事件的独立性与相关性	(22)
1.5.2 有限个事件的独立性	(25)
1.5.3 相互独立事件的性质	(26)
1.5.4 伯努利概型	(28)
习题1	(29)
第2章 一维随机变量及其分布	(36)
2.1 随机变量及其分布	(36)
2.1.1 随机变量的概念	(37)
2.1.2 随机变量的分布函数	(38)
2.2 离散型随机变量的概率函数及分布函数	(39)
2.2.1 常见的离散型随机变量的概率分布	(40)
2.2.2* 缸的模型	(47)
2.2.3* 缸模型的应用实例	(47)
2.3 连续型随机变量及其概率密度	(48)
2.3.1 连续型随机变量及其概率密度函数	(48)
2.3.2 常用的连续型随机变量	(50)
2.4 随机变量的函数分布	(58)
2.4.1 离散型随机变量的函数分布	(59)
2.4.2 连续型随机变量的函数分布	(59)
习题2	(62)
第3章 多维随机变量及其分布	(66)
3.1 二维随机变量及其分布	(66)
3.1.1 二维随机变量	(66)
3.1.2 二维随机变量的联合分布函数	(67)
3.1.3 二维离散型随机变量的概率分布	(67)
3.1.4 二维连续型随机变量及其联合概率分布	(68)
3.1.5 几个常用的分布	(69)
3.2 边缘分布	(72)
3.2.1 离散情形	(73)
3.2.2 连续情形	(74)
3.3 二维随机变量的条件分布	(76)
3.3.1 离散情形	(76)

3.3.2 连续情形	(78)
3.4 二维随机变量的独立性	(79)
3.5 二维随机变量的函数分布	(82)
3.5.1 和的分布	(82)
3.5.2 一般函数 $Z = g(X, Y)$ 的分布	(87)
3.5.3 一般变换	(87)
3.5.4 极值分布	(87)
习题 3	(89)
第 4 章 随机变量的数字特征	(95)
4.1 数学期望	(95)
4.1.1 数学期望的性质	(96)
4.1.2 随机变量函数的数学期望	(99)
4.1.3* 数学期望的简单应用	(101)
4.2 中位数、众数和 p 分位点	(104)
4.3 方差	(106)
4.4 协方差及相关系数	(109)
4.5 矩、协方差矩阵	(115)
习题 4	(117)
第 5 章 大数定律与中心极限定理	(124)
5.1 大数定律	(124)
5.1.1 问题的提出	(124)
5.1.2 切比雪夫不等式与大数定律	(125)
5.2 中心极限定理	(129)
5.2.1 中心极限定理的提法	(129)
5.2.2 中心极限定理	(132)
5.2.3* 若干应用	(134)
习题 5	(136)
第 6 章 数理统计学简介	(139)
6.1 数理统计学的基本概念	(140)
6.1.1 引例	(140)
6.1.2 总体与样本	(141)

6.1.3	统计量	(144)
6.1.4	经验分布函数	(145)
6.1.5	数理统计方法的特点	(147)
6.1.6	数理统计的基本思想	(147)
6.1.7	统计模型	(148)
6.2	正态样本统计量的抽样分布	(148)
6.2.1	正态分布	(149)
6.2.2	χ^2 (卡方)分布	(149)
6.2.3	t 分布(学生分布)	(152)
6.2.4	F 分布	(153)
	习题 6	(157)
第 7 章	参数估计	(163)
7.1	点估计概述	(163)
7.1.1	频率替换法	(163)
7.1.2	矩估计法	(164)
7.1.3	极大似然法	(166)
7.1.4	极大似然估计的不变性原则	(173)
7.2	估计量优良性的评选标准	(174)
7.2.1	无偏性	(175)
7.2.2	有效性	(176)
7.2.3	均方误差准则	(177)
7.2.4	一致性(相合性)	(178)
7.3	参数的区间估计	(180)
7.3.1	枢轴量法	(181)
7.3.2	单个正态总体数学期望的区间估计	(181)
7.3.3	单个正态总体方差的区间估计	(183)
7.3.4	两个正态总体期望差的区间估计	(184)
7.3.5	两个正态总体方差比的区间估计	(186)
7.3.6	大样本区间估计	(188)
7.3.7	单侧置信区间	(189)
	习题 7	(191)

第 8 章 假设检验	(196)
8.1 假设检验的基本概念	(196)
8.1.1 问题的提法	(196)
8.1.2 假设检验的基本思想	(197)
8.1.3 假设检验的定义与步骤	(198)
8.2 正态总体参数的假设检验	(201)
8.2.1 正态总体数学期望的假设检验	(201)
8.2.2 正态总体方差的假设检验	(211)
8.2.3 两正态总体期望差的假设检验	(215)
8.2.4 两正态总体方差比的 F 检验	(218)
8.3* 似然比检验	(222)
8.3.1 似然比检验的基本思想	(222)
8.3.2 似然比检验的一般步骤	(223)
8.4 两种类型的错误	(225)
8.5 拟合优度检验	(229)
8.5.1 非参数 χ^2 检验	(229)
8.5.2 列联表独立性检验	(234)
8.6 连续型分布的柯尔莫戈洛夫检验	(238)
8.6.1 样本的经验函数	(238)
8.6.2 柯尔莫戈洛夫检验	(238)
8.7 双总体的秩和检验	(241)
8.7.1 两个总体分布的假设检验	(241)
8.7.2 秩和检验法	(242)
习题 8	(244)
第 9 章 方差分析和回归分析简介	(250)
9.1 单因素方差分析	(250)
9.2 一元线性回归	(257)
9.2.1 因变量 y 与自变量 x 之间的关系	(257)
9.2.2 一元线性回归模型及其参数估计	(258)
9.2.3 最小二乘估计量 a, b 的基本性质	(260)
9.2.4 σ^2 的无偏估计	(262)
9.2.5 线性相关显著性检验	(263)

9.2.6 预测与控制	(265)
9.2.7 可化为一元线性回归的曲线回归模型举例	(267)
习题 9	(269)
第 10 章 R 语言简介	(272)
10.1 R 简介	(272)
10.2 R 语言的应用	(275)
10.2.1 描述性统计	(275)
10.2.2 数据的分布	(275)
10.2.3 正态性检验	(279)
10.2.4 多元数据的数字特征及相关系数	(279)
10.2.5 参数估计	(280)
10.2.6 假设检验	(282)
10.2.7 方差分析	(288)
10.2.8 回归分析	(293)
习题答案	(297)
参考文献	(307)
附表 1 泊松分布函数表	(308)
附表 2 标准正态分布函数表	(310)
附表 3 χ^2 分布分位数 $\chi^2_{\alpha}(n)$ 表	(312)
附表 4 t 分布分位数 $t_{\alpha}(n)$ 表	(314)
附表 5 F 分布表	(316)

我又转念,见日光之下,快跑的未必能赢,力战的未必得胜,智慧的未必得粮食,明哲的未必得资财,灵巧的未必得喜悦,所临到众人的,是在乎当时的机会。

——《传道书》

第 1 章 随机事件与概率

概率论是研究随机现象规律性的数学分支,为了对随机现象的有关问题做出明确的数学描述,像其他数学学科一样,概率论具有自己严格的体系和结构.本章重点介绍概率论的两个基本概念——随机事件和概率.

1.1 基 本 概 念

1.1.1 随机现象

日常生活中,有些问题是很难给予准确无误的回答的.例如,抛掷一枚硬币,它可能是正面朝上,也可能是反面朝上,就是说,“正面朝上”这个结果可能出现也可能不出现;下一个交易日股市的指数可能上升也可能下跌,而且升跌幅度大小也不可能事先确定.这些问题的结果都是不确定的、偶然的,很难准确给出答案.许多影响事物发展的偶然因素的存在,是产生这类现象中不确定性的原因.

1.1.2 随机现象的统计规律性

虽然随机现象中出现什么样的结果不能事先预言,但是可以假定全部可能结果是已知的.在上述例子中,抛掷一枚硬币只会有“正面”与“反面”这两种可能结

果,而股指的升跌幅度大小充其量假定它可能是任意的实数.可见,“全部可能的结果的集合是已知的”这个假定是合理的,而且它会给我们的学习研究带来许多方便.

进行一次试验,如果其所得结果不能完全预知,但其全体可能结果是已知的,则称此试验为**随机试验**(random experiment).

由于随机现象的结果是事先无法预知的,初看起来,随机现象毫无规律可言.然而人们发现同一随机现象在大量重复出现时,其每种可能的结果出现的频率却具有稳定性,从而表明随机现象也有其固有的规律性.历史上许多科学家都对抛硬币会出现正反面的现象感兴趣,并做了许多试验.下面是一些著名的记录(表 1.1).

表 1.1 抛硬币试验

试验者	抛硬币次数	出现正面次数	出现正面频率
Buffon	4 040	2 048	0.506 9
De Morgan	4 092	2 048	0.500 5
Kerrich	7 000	3 516	0.502 2
Kerrich	8 000	4 034	0.504 2
Kerrich	9 000	4 538	0.504 2
Feller	10 000	4 979	0.497 9
Pearson	12 000	6 019	0.501 6
Pearson	24 000	12 012	0.500 5
Lomanovskii	80 640	39 699	0.492 3

这里的 De Morgan(1806~1871)是印度人,在逻辑学方面有重要的创新,在数学和数学史方面也有重要贡献;Buffon(1701~1788)是法国博物学家;Karl Pearson(1895~1936)是英国统计学家;Kerrich 是南非数学家,他的抛币试验是在德国人的集中营里进行的.二战爆发时,Kerrich 访问哥本哈根,德国入侵丹麦时他被关进 Jutland 集中营,在那里度过了漫长的岁月.为了消磨时光,他一次次地抛硬币,并记录了上面的结果.

从上面的数据可以看出,当抛掷次数很大时,正面出现的频率愈来愈接近于 0.5,就是说,出现正面与出现反面的机会差不多各占一半.上述试验的结果表明,在相同条件下大量地重复某一随机试验时,各种可能结果出现的频率稳定在某个确定的数值附近.称这种性质为**频率的稳定性**.频率的稳定性的存在,标志着随机现象也有它的数量规律性.

注 一般地,随机试验原则上是可以重复的,对不可重复的情况,我们不去研究(至少概率论上是这样认为的),因为我们无法确定其发生的频率.

1.1.3 样本空间

随机试验的每一个可能的结果称为一个**样本点**(sample point),因而一个随机试验的所有样本点也是明确的,它们的全体称为**样本空间**(sample space),习惯上分别用 ω 与 Ω 表示样本点与样本空间.

例 1.1.1 抛两枚硬币观察其正面与反面出现的情况.其样本空间由四个样本点组成,即 $\Omega = \{(\text{正}, \text{正}), (\text{正}, \text{反}), (\text{反}, \text{正}), (\text{反}, \text{反})\}$. 这里,比如样本点 $\omega = (\text{正}, \text{反})$ 表示第一枚硬币抛出正面而第二枚抛出反面.

例 1.1.2 观察某电话交换台在一天内收到的呼叫次数,其样本点有可数无穷多个: i ($i = 0, 1, 2, \dots$)次,样本空间为: $\Omega = \{0 \text{ 次}, 1 \text{ 次}, 2 \text{ 次}, \dots\}$.

例 1.1.3 接连射击直到命中为止.为了简洁地写出其样本空间,我们约定以“0”表示某次射击未中,而以“1”表示命中,则样本空间 $\Omega = \{1, 01, 001, \dots\}$.

例 1.1.4 观察一个新灯泡的寿命,其样本点也有无穷多个(t 小时, $0 \leq t < +\infty$),样本空间为: $\Omega = \{t | 0 \leq t < +\infty\}$.

1.1.4 随机事件及其运算

我们时常会关心试验的某一部分可能结果是否会出现.如在例 1.1.2 中,若以每小时是否达到 5 次电话呼叫来区分这台电话机是否太繁忙,那么“不太繁忙”即不足 5 次的呼叫,它由样本空间中前 5 个样本点 0, 1, 2, 3, 4 组成.由于它是由 Ω 中的一部分样本点组成的子集合,故在未来的一次试验中可能发生也可能不发生.称这种由部分样本点组成的试验结果为**随机事件**(random event),简称事件.通常用大写的字母 $A, B, \dots$ 表示.某事件发生,表示属于该集合的某一样本点在试验中出现.记 ω 为试验中出现的样本点,那么事件 A 发生当且仅当 $\omega \in A$ 发生.只包含一个样本点 ω 的单点集合 $\{\omega\}$,称为试验的一个**基本事件**.由于样本空间 Ω 包含了全部可能结果,因此在每次试验中 Ω 都会发生,故称 Ω 为**必然事件**(certain event).相反,空集 $\emptyset$ 不包含任何样本点,每次试验必定不发生,故称 $\emptyset$ 为**不可能事件**(impossible event).

在一个随机试验中,一般有很多随机事件,为了通过对简单事件的研究来掌握复杂事件,我们需要研究事件之间的关系和事件之间的一些运算.

1. 事件的包含

若事件 A 发生必然导致 B 发生,即属于 A 的每一个样本点一定也属于 B ,则

称事件 B 包含 (contain) 事件 A , 或者称事件 A 包含于事件 B , 记作 $B \supset A$ 或 $A \subset B$.

2. 事件的相等

若事件 A 包含事件 B , 事件 B 也包含事件 A , 则称事件 A 与 B 相等 (equivalent), 记作 $A = B$.

3. 事件的并

“事件 A 与 B 至少有一个发生”这一事件称作事件 A 与 B 的并 (union), 记作 $A \cup B$. 显然 $A \cup B$ 是由 A 与 B 的样本点共同构成的事件, 这与集合的并的含义是一致的.

4. 事件的交

“事件 A 与 B 都发生”这一事件称作事件 A 与 B 的交 (intersection), 记作 $A \cap B$ (或 AB). 显然 $A \cap B$ 是由 A 与 B 的公共样本点所构成的事件, 这与集合的交的含义是一致的.

5. 事件的差

“事件 A 发生而 B 不发生”这一事件称作事件 A 与 B 的差, 记作 $A - B$. 显然 $A - B$ 是由所有属于 A 但不属于 B 的样本点构成的事件, 这与集合的差的含义也是一致的.

6. 互不相容事件

若事件 A 与 B 不能同时发生, 也就是说 AB 是不可能事件, 即 $AB = \emptyset$, 则称 A 与 B 是互斥事件 (mutually exclusive event) 或互不相容事件 (incompatible event).

7. 对立事件

“事件 A 不发生”这一事件称作事件 A 的对立事件 (opposite event) 或逆事件 (complementary event), 记作 $\bar{A}$. 易见, $\bar{A} = \Omega - A$.

为了帮助大家理解上述概念, 现把集合论的有关结论与事件的关系和运算的对应情况列于表 1.2.

表 1.2

符号	集合论	概率论
Ω	全集	样本空间; 必然事件
$\emptyset$	空集	不可能事件
$\omega \in \Omega$	Ω 中的点 (或元素)	样本点

续表

符号	集合论	概率论
$\{\omega\}$	单点集	基本事件
$A \subset \Omega$	Ω 的子集 A	事件 A
$A \subset B$	集合 A 包含在集合 B 中	事件 A 包含于事件 B
$A = B$	集合 A 与 B 相等	事件 A 与 B 相等
$A \cup B$	集合 A 与 B 之并	事件 A 与 B 至少有一个发生
$A \cap B$	集合 A 与 B 之交	事件 A 与 B 同时发生
$\bar{A}$	集合 A 的余集	事件 A 的对立事件
$A - B$	集合 A 与 B 之差	事件 A 发生而 B 不发生
$A \cap B = \emptyset$	集合 A 与 B 没有公共元素	事件 A 与 B 互不相容(互斥)

1.2 随机事件的概率

1.2.1 概率和频率

对于随机事件,知道它发生的可能性大小是非常重要的.用概率度量随机事件的可能性大小能为我们的决策提供关键性的依据.随机事件在一次试验中是否发生带有偶然性,但大量试验中,它的发生具有统计规律性,人们可以确定随机事件发生的可能性大小.若随机事件 A 在 n 次试验中发生了 m ($0 \leq m \leq n$) 次,则量 m/n 称为事件 A 在 n 次试验中发生的频率(frequency),记作 $f_n(A)$,即 $f_n(A) = m/n$.它满足不等式: $0 \leq m/n \leq 1$.

如果 A 是必然事件,则有 $m = n$, $f_n(A) = 1$;如果 A 是不可能事件,则有 $m = 0$, $f_n(A) = 0$.也就是说,必然事件的频率为 1,不可能事件的频率为 0.

在抛硬币试验中,我们用 A 表示事件出现正面,由表 1.1 可以看出,随着试验次数 n 的增加, A 发生的频率在 0.5 这个数值附近摆动的幅度越来越小,即随机事件 A 发生的频率具有稳定性.一般地,在大量重复试验中,随机事件发生的频率,总是在某个确定值 p 附近徘徊,而且随着试验次数增多,事件 A 的频率就越接近 p ,数 p 称为频率的**稳定中心**.频率的稳定性揭示了随机现象的客观规律