概率论与数理统计

主编 陈文英 吴志丹 王艳芳 王 涛

科学出版社

概率论与数理统计

主　编　陈文英　吴志丹
　　　　王艳芳　王　涛
副主编　丁　巍　张洪阳
　　　　耿　莹　杨淑辉
　　　　王亚男

科学出版社
北　京

内 容 简 介

本书是普通高等院校非数学专业“概率论与数理统计”基础课教材. 全书共 9 章，主要内容包括：随机事件与概率、随机变量及其分布、多维随机变量及其分布、数字特征、大数定律与中心极限定理、数理统计的基本概念、参数估计、假设检验、概率统计在经济中的应用. 本书每节配有习题，每章末(第 9 章除外)均有本章小结并配有相应的自测题，书后附有参考答案.

本书可供高等院校工科、经济、管理、金融、旅游等专业的学生使用，也可作为工程技术人员、自然科学工作者和社会科学工作者的自学用书.

图书在版编目(CIP)数据

概率论与数理统计/陈文英等主编. —北京：科学出版社, 2012
ISBN 978-7-03-034265-2
Ⅰ. ①概… Ⅱ. ①陈… Ⅲ. ①概率论 ②数理统计 Ⅳ. O21
中国版本图书馆 CIP 数据核字(2012) 第 091783 号

责任编辑：张中兴 于俊杰／责任校对：刘小梅
责任印制：阎 磊／封面设计：迷底书装

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

北京市文林印务有限公司印刷
科学出版社发行 各地新华书店经销
*
2012 年 5 月第 一 版 开本：720 × 1000 (B5)
2012 年 5 月第一次印刷 印张：16 1/4
字数：305 000

定价：29.00 元

(如有印装质量问题，我社负责调换)

前　言

本书是由沈阳师范大学概率论与数理统计课程组的教师在多年教学实践的基础上编写的, 从编写体例、知识结构、重点难点, 到许多具体的细节处理, 都经过了课程组教师的深入讨论. 在编写的过程中, 也参考了国内外许多同类优秀教材.

概率论与数理统计课程是学生在大学里首次接触到以随机现象为研究对象的数学课程, 考虑到课程本身及教学对象的特殊性, 本书在编写时遵循了下列原则: ① 既理论体系严谨, 又通俗易读、详略得当. ② 弱化理论推导, 突出应用性特点, 贴近生活实际. ③ 结构合理, 脉络清晰, 概念准确, 重点突出. ④ 从实际问题出发引入概念、建立定理、选择例题、设置习题. ⑤ 归纳解题方法和技巧, 使读者 "有法可依". ⑥ 尽可能交代清楚知识产生的背景.

全书共 9 章, 主要内容包括: 随机事件与概率、随机变量及其分布、多维随机变量及其分布、数字特征、大数定律与中心极限定理、数理统计的基本概念、参数估计、假设检验、概率统计在经济中的应用. 每节都配有习题, 每章末 (除第 9 章) 都设有本章小结并配有自测题, 书后附有参考答案.

本书可作为普通高等院校非数学专业本科生学习概率论与数理统计的大学数学基础课教材, 也可作为工程技术人员、自然科学工作者和社会科学工作者的自学用书.

在本书编写过程中得到了沈阳师范大学计算中心的领导及科学出版社的领导和编辑的大力支持, 在此表示衷心的感谢.

由于编者水平有限, 书中难免有不妥之处, 热诚希望专家、同行和广大读者不吝赐教.

编　者

2012 年 1 月

目　　录

绪　论

概率论与数理统计是高等教育中一门重要的公共基础课, 所讨论和研究的问题与我们的现实生活有着密切的联系, 在近代物理、自动控制、地震预报和气象预报、工厂产品质量控制、农业试验和公用事业等方面都有广泛的应用. 课程的主要内容包括: 随机事件和概率、一维和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、参数估计、假设检验等内容.

概率论早在 15~16 世纪就已经开始萌芽. 意大利数学家帕乔利、塔尔塔利亚和卡尔丹的著作中都曾对 "赌金分配" 问题进行过讨论. 但率先试图对这类问题进行数学抽象和归纳的, 是法国数学家费马和帕斯卡. 1654 年, 法国的一位名为梅尔的骑士, 向帕斯卡提出了一个令他苦恼很久的问题, 史称 "赌金问题": 两个赌徒相约赌若干局, 先赢 s 局的就算赢了. 现在一个赌徒赢 a 局 $(a < s)$, 而另一个赢 b 局 $(b < s)$ 时赌博终止了. 问赌本应该如何分配才算合理? 帕斯卡就赌金问题通过书信与费马进行了深入交流. 正如对概率论作出卓越贡献的法国数学家泊松所说: "赌金问题乃是概率论的起源问题." 荷兰数学家惠更斯游学到巴黎, 也对赌金问题产生了浓厚的兴趣, 并于 1657 年发表了最早的概率论著作 ——《论赌博中的计算》.

上述数学家的著述中出现的第一批概率论概念, 标志着概率论的诞生. 而概率论作为一门独立的数学分支, 真正的奠基人是雅格布 • 伯努利. 他一生中最具创造力的著作 ——《猜度术》, 在其死后八年才出版, 该著作中首次提出了后来以 "伯努利定理" 著称的极限定理. "伯努利定理" 是 "大数定律" 的最早形式. 雅格布对大数定律的陈述与现代的标准概率著作十分一致. 大数定律第一次试图在单一的概率值与众多现象的统计度量之间建立演绎关系, 成为概率论通向广泛应用领域的桥梁. 伯努利之后, 法国数学家棣莫弗将概率论又向前推进了一步, 他提出了概率乘法法则, 正态分布和正态分布率的概念, 并给出了概率论的一些重要结果. 之后法国数学家蒲丰提出了著名的 "蒲丰问题", 引进了几何概率. 另外, 拉普拉斯、高斯和泊松等对概率论做出了进一步奠基性工作. 特别是拉普拉斯, 他是严密的、系统的科学概率论的最卓越的创建者, 在 1812 年出版的《概率的分析理论》中, 拉普拉斯以强有力的分析工具处理了概率论的基本内容, 实现了从组合技巧向分析方法的过渡, 使以往零散的结果系统化, 开辟了概率论发展的新时期. 泊松则推广了大数定理, 提出了著名的泊松分布. 这些科学家的努力工作, 使得概率论这门学科迅速成长起来, 但是还不能称其为一门严格的演绎科学.

最早对概率论严格化进行尝试的是俄国数学家伯恩斯坦和奥地利数学家冯 • 米西斯. 他们都提出了一些公理来作为概率论的前提, 但是他们的公理理论都是不完善的. 只有在测度论与实变函数理论的基础上才可能建立真正的公理化概率论. 19 世纪末期的俄国数学家及 20 世纪上半叶的苏联数学家为概率论走向公理化作出了不可磨

灭的贡献. 在这一时期, 极限理论的发展成为概率论研究的中心课题, 俄国数学家切比雪夫对此作出了重要贡献. 他建立了关于独立随机变量序列的大数定律, 推广了棣莫弗–拉普拉斯的极限定理. 切比雪夫的成果被其学生马尔可夫发扬光大, 影响了 20 世纪概率论发展的进程.

前苏联数学家柯尔莫哥洛夫的工作在概率论公理化进程中最为卓著. 他在 1926 年推导了弱大数定律成立的充分必要条件后, 又对法国数学家博雷尔提出的强大数定律问题给出了最一般的结果, 从而解决了概率论的中心课题之一 —— 大数定律, 成为以测度论为基础的概率论公理化的前奏. 1933 年, 柯尔莫哥洛夫出版了他的著作《概率论基础》, 这是概率论的一部经典性著作. 其中, 柯尔莫哥洛夫给出了公理化概率论的一系列基本概念, 提出了六条公理, 整个概率论大厦可以从这六条公理出发建筑起来. 柯尔莫哥洛夫的公理化体系逐渐得到数学家们的普遍认可. 由于公理化, 概率论成为一门严格的演绎科学, 并通过集合论与其他数学分支密切地联系着. 在公理化基础上, 现代概率论取得了一系列理论突破, 大大拓广了其应用范围.

数理统计学是一门应用性极其广泛的基础性学科, 它以概率论为基础, 伴随着概率论的发展而发展起来. 数理统计学通过建立数学模型, 收集整理数据, 进行统计推断、预测和决策. 数理统计方法在工农业生产、自然科学和技术科学以及社会经济领域中都有广泛的应用.

简单的统计思想古来有之, 现代意义的数理统计学在 19 世纪萌芽并发展, 在 20 世纪成熟.

19 世纪初期, 高斯由于计算行星轨道的需要建立了以 “最小二乘法” 为基础的对观测数据的误差分析, 并在对 “测地问题” 的研究过程中进一步完善了最小二乘法和对统计规律的研究. 高斯曾开设过 “最小二乘法及其在科学中的应用” 课程, 戴德金 (近代抽象数学的先驱) 就曾选修过高斯的最小二乘法课程. 高斯的工作促使统计学摆脱对观测数据的单纯描述, 向强调推断进行过渡.

英国统计学家 K. 皮尔逊对现代数理统计学的建立作出了重要贡献, 他提出了 “总体” 的概念, 把群体作为统计学的研究对象, 是 “大样本统计” 的前驱.K. 皮尔逊还给出了估计参数的一种方法 —— 矩法估计, 发展了德国测地学者赫尔默特发现的 χ^2 分布.K. 皮尔逊的学生戈塞特强调样本必须从总体中随机地抽取, 从而将统计学的研究对象从群体现象转变为随机现象.

现代数理统计学作为一门独立学科的奠基人是英国数学家费希尔, 他提出了许多重要的统计方法, 发展了正态总体下各种统计量的抽样分布, 建立了系统的相关分析与回归分析; 费希尔与叶茨共同创立了试验设计这一统计分支, 同时也是假设检验的先驱.

1946 年, 瑞典数学家克拉默发表论文, 用测度论系统总结数理统计的发展, 标志着现代数理统计学的成熟. 第二次世界大战之后, 美籍罗马尼亚数学家沃尔德提出了

序贯分析和统计决策理论, 引起了战后数理统计思想的革新.

随着现代数学的发展, 概率论与数理统计学将更广泛地与其他新科学新理论相结合, 不断产生新的边缘科学或新的统计分支.

中国概率统计领域内享有国际声誉的数学家有许宝騄和王梓坤.

许宝騄出身名门, 1936 年赴英国伦敦大学高尔顿实验室和统计系攻读学位. 1947 年开始在北京大学任教授, 直到 1970 年去世. 许宝騄的主要工作是在数理统计和概率论两个方面. 数理统计方面, 在 1938~1945 年期间, 他对多元统计分析中的精确分布和极限分布研究得到了重要的结果, 导出正态分布样本协方差阵特征根的联合分布和极限分布, 这些结果是多元分析中的基石. 以上这两方面的工作确立了他在数理统计中的国际地位. 晚年, 他致力于组合设计的构造, 也有重要的成果.

王梓坤, 中国科学院院士, 1929 年 4 月生, 江西吉安县人.1955 年, 王梓坤在南开大学任教期间, 经推荐考取了留苏研究生, 去莫斯科大学数学力学系攻读概率论. 王梓坤在苏联的导师是近代概率论的奠基人柯尔莫哥洛夫和杜布罗辛.1958 年, 王梓坤的博士论文《生灭过程的分类》在莫斯科大学的学术答辩会上一致通过, 获副博士学位回国.

回国以后, 王梓坤继续进行概率论的研究工作. 20 世纪 60 年代初, 他研究马尔可夫链的构造, 彻底解决了生灭过程的构造与泛函分布问题, 是开创这一领域研究的先驱; 70 年代, 他研究马尔可夫过程与位势论的关系, 求出了布朗运动与对称稳定过程末离球的时间与位置的分布, 并研究地震的统计预报问题, 著有《布朗运动与位势》、《概率与统计预报》等书; 80 年代, 他研究多指标马尔可夫过程, 在国际上率先引进多指标奥恩斯坦–乌伦贝克过程的定义, 并研究了它的性质; 90 年代初, 除继续上述工作外, 他还从事超过程的研究, 这是当前国际上最活跃的课题之一.

第 1 章　随机事件与概率

"概率论与数理统计" 是以数量化的方法来研究随机现象及其规律性的一门应用数学学科. 20 世纪以来, 概率论向各个领域的渗透已成为近代科学技术发展的重要特征之一, 并被广泛地应用到生产、生活的各个方面, 其理论和方法正在为时代发展和社会建设发挥着不可替代的独特作用.

本章介绍的随机事件与概率是概率论中最基本、最重要的概念之一. 其主要内容包括: 随机事件和样本空间, 事件间的关系与运算, 概率的定义、性质及计算方法, 条件概率与乘法公式, 全概率公式与贝叶斯公式, 事件的独立性与伯努利概型.

1.1　随机事件和样本空间

1.1.1　随机现象和随机试验

我们所指的试验是一个广义的概念, 它可指对某个过程的记录、对一个问题的调查、各种科学实验等.

1. 随机现象

什么是随机现象? 这可以用两个简单的试验来阐明:

试验 1　一袋装有 3 个外形完全相同的白球, 从中任取一球;

试验 2　一袋装有 3 个外形完全相同但颜色不同的球, 从中任取一球.

对于试验 1, 根据其条件, 我们就能断定取出的必是白球. 像这样在试验之前能断定结果的现象称为**确定性现象**. 确定性现象非常广泛. 例如, 同种电荷互相排斥; 标准大气压下, 水加热到 100°C 会沸腾; 边长为 a, b 的矩形, 其面积必为 $a \cdot b$. 诸如此类都是确定性现象.

对于试验 2, 根据其条件, 在球没有取出之前, 不能断定取出的是哪种颜色的球. 我们把在试验之前无法知道确切结果的现象称为**随机现象**. 随机现象更是广泛地存在于客观世界之中. 例如, 抛一枚硬币, 落地后可能出现正面, 也可能出现反面; 新生婴儿可能是男孩, 也可能是女孩; 将来某日某种股票的价格可能涨, 可能跌, 也可能价格不变. 诸如此类都是随机现象.

2. 随机试验

试验常用大写字母 $E, E_1, E_2, \cdots$ 表示. 下面看几个试验的例子:

E_1: 掷一枚骰子, 观察朝上出现的点数;

E_2: 先后抛两次硬币, 观察正面与反面出现的情况;

E_3: 记录一部热线电话在 2 分钟内接到电话的次数;

E_4: 按户调查农村居民年购买食品、家电的支出.

以上试验具有三个特点:

(1) 试验可以在相同条件下重复进行;

(2) 每次试验的可能结果不唯一, 但能事先明确试验的所有可能结果;

(3) 试验前不能确定哪一个结果会发生.

满足上述三个特点的试验称为**随机试验**. 以后我们所说的试验均指随机试验.

1.1.2 样本空间与随机事件

1. 样本空间

为了研究随机试验 E, 首先需要知道 E 的一切可能出现的结果. 我们把随机试验 E 的所有可能结果组成的集合称为 E 的**样本空间**, 记作 Ω. 其中每一个可能的结果称为**样本点**, 记作 ω.

例 1 写出 1.1.1 小节中随机试验 E_1, E_2, E_3, E_4 对应的样本空间 $\Omega_1, \Omega_2, \Omega_3, \Omega_4$.

解 $\Omega_1 = \{1,2,3,4,5,6\}$; $\Omega_2 =$\{(正, 正), (正, 反), (反, 正), (反, 反)\};

$$\Omega_3 = \{0,1,2,\cdots\}; \quad \Omega_4 = \{(x,y)|x \geqslant 0, y \geqslant 0\},$$

其中 Ω_1, Ω_2 中样本点的个数为有限个, 称为**有限样本空间**; Ω_3, Ω_4 中样本点的个数为无限个, 称为**无限样本空间**; 又因为 Ω_3 中的样本点可以按一定顺序排列, 又称为**可列样本空间**.

2. 随机事件

进行随机试验时, 人们常关心某一类结果是否发生. 如调查农村每户居民年购买食品、家电的支出是否分别大于 5000 元和 3000 元, 记 $A = \{(x,y)|x > 5000, y > 3000\}$, 显然 A 是 Ω_4 的子集. 一般地, 样本空间 Ω 的任意子集称为**随机事件**, 简称**事件**. 事件一般用大写的字母 A, B, C, A_1, A_2 等表示.

随机事件发生是常用的一个术语, 规定:

随机事件 A 发生 $\Leftrightarrow$ 随机试验时 A 中的一个样本点出现.

由一个样本点组成的单点集称为**基本事件**. 样本空间 Ω 有两个特殊的子集: 一个是空集 $\varnothing$, 它不包含任何样本点, 因此在每次试验中都不会发生, 称为**不可能事件**; 另一个是 Ω 本身, 由于它包含了试验所有可能的结果, 所以在每次试验中它总是会发生, 称为**必然事件**.

例 2 在公路上随机抽查 10 辆汽车, 考察其中公有车辆数, 写出样本空间并将下列事件用列举法表示为集合的形式:

$$
\begin{aligned}
A &= \{\text{没有公车}\},\\
B &= \{\text{有 1 辆或 2 辆公车}\},\\
C &= \{\text{公车不超过 3 辆}\},\\
D &= \{\text{公车多于 2 辆}\}.
\end{aligned}
$$

解　$\Omega = \{0,1,2,3,4,5,6,7,8,9,10\}$,

$$
A = \{0\}, \qquad B = \{1,2\},
$$
$$
C = \{0,1,2,3\}, \quad D = \{3,4,5,6,7,8,9,10\}.
$$

注　(1) 对于一个随机试验而言, 当试验的目的不同时, 样本空间往往是不同的. 如把篮球运动员投篮作为随机试验时, 若以考察是否命中为目的, 试验的样本空间 Ω={中, 不中}; 若以考察投篮的得分情况为目的, 试验的样本空间 Ω={0, 1, 2, 3}, 所以我们应从试验的目的来确定样本空间.

(2) 必然事件和不可能事件本来没有随机性可言, 但为了研究问题的需要, 常把它们看成随机事件的极端情况.

习　题　1.1

1. 判断下列试验是否为随机试验:

(1) 在一定条件下进行射击, 观察是否击中靶上红心;

(2) 在恒力作用下一质点做匀速运动;

(3) 在 4 个同样的球 (标号 1, 2, 3, 4) 中, 任取一只, 观察所取球的标号.

2. 有五件产品, 其中有一件次品 (记为 a), 四件正品 (记为 b_1, b_2, b_3, b_4). 从中一次取出两件, 观察结果. 写出样本空间.

3. 连续投三枚硬币, 观察正面与反面出现的情况. 写出样本空间及下列事件中的样本点:

$$
\begin{aligned}
A &= \{\text{第一枚出现正面, 第三枚出现反面}\};\\
B &= \{\text{第二枚出现反面}\};\\
C &= \{\text{至少出现一个正面}\}.
\end{aligned}
$$

4. 写出下列随机试验的样本空间:

(1) 将一枚骰子连投四次, 观察点数出现是 6 的次数.

(2) 丢甲、乙两枚骰子, 观察出现点数之和.

(3) 袋中有标号为 1, 2, 3 的三个球.

① 随机取两次, 一次一个, 取后不放回, 观察取到球的序号;

② 随机取两次, 一次一个, 取后放回, 观察取到球的序号;

③ 一次随机取两个, 观察取到球的序号.

5. 写出下列随机试验的样本空间:

(1) 记录一个小班 (30 人) 一次概率考试的平均分数 (以百分制记分);

(2) 生产某产品直到有 10 件正品为止, 记录生产产品的总件数.

6. 写出下列随机试验中的随机事件:

(1) 由 1,2,3 三个数组成没有重复数字的三位数;

(2) 10 个零件, 其中有 2 件次品, 随机的取 5 件.

A ={正品个数多于次品个数}, B={正品个数不多于次品个数}.

1.2 事件间的关系与运算

事件是一个集合, 事件间的关系和运算就是集合间的关系和运算, 只是在概率论中从事件的角度给出了新的术语.

1.2.1 事件间的关系与运算

1. 事件的包含与相等

若事件 A 发生必然导致事件 B 发生, 即 A 的每一个样本点都是 B 的样本点, 则称事件 A**包含于**事件 B, 或称事件 B**包含**事件 A, 记作 $A \subset B$ 或者 $B \supset A$. 图 1.1 是其文氏图.

若 $A \subset B$ 且 $B \subset A$, 即 A 与 B 含有相同的样本点, 则称事件 A 与事件 B**相等**, 记作 $A = B$.

2. 互斥事件 (互不相容事件)

若事件 A 与事件 B 不能同时发生, 即 A 与 B 没有公共的样本点, 则称事件 A 与事件 B 是**互斥事件**或**互不相容事件**. 图 1.2 是其文氏图.

互斥事件包含三种情形: ① A 发生 B 不发生; ② B 发生 A 不发生; ③ A 与 B 都不发生.

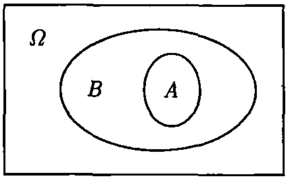

图 1.1

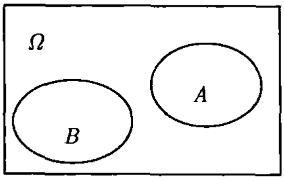

图 1.2

若事件 $A_1, A_2, \cdots, A_n$ 中的任意两个事件都互斥, 则称这些事件两两互斥. 同一样本空间中的基本事件是两两互斥的.

3. 对立事件

"事件 A 不发生" 这一事件称为事件 A 的**对立事件**, 记作 $\overline{A}$. 事件 A 的对立事件 $\overline{A}$ 就是 A 的补集. 图 1.3 是其文氏图.

一个事件与它的对立事件中有且只有一个发生. 对立事件一定是互不相容事件, 互不相容事件不一定是对立事件.

4. 事件的并 (或和)

事件 A 与 B 至少有一个发生, 称为事件 A 与 B 的**并** (或和), 记作 $A \cup B$(或 $A+B$). 事件 $A \cup B$ 发生意味着要么事件 A 发生, 要么事件 B 发生, 要么事件 A 与 B 都发生. 图 1.4 是其文氏图.

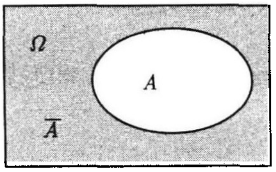

图 1.3

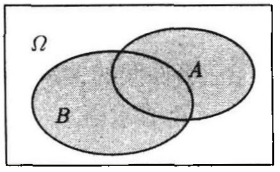

图 1.4

事件的和可以推广到多个事件的情形. n 个事件 $A_1, A_2, \cdots, A_n$ 的和事件记作 $\bigcup\limits_{i=1}^{n} A_i$, 它表示 $A_1, A_2, \cdots, A_n$ 中至少有一个发生.

5. 事件的交 (或积)

事件 A 与 B 同时发生, 称为事件 A 与 B 的**交**(或积)**事件**, 记作 $A \cap B$, 也简记作 AB. 图 1.5 是其文氏图.

类似于 n 个事件的和事件, n 个事件 $A_1, A_2, \cdots, A_n$ 的积记作 $\bigcap\limits_{i=1}^{n} A_i$, 它表示 $A_1, A_2, \cdots, A_n$ 同时发生.

6. 事件的差

事件 A 发生 B 不发生, 称为事件 A 与 B 的差, 记作 $A-B$. 图 1.6 是其文氏图.

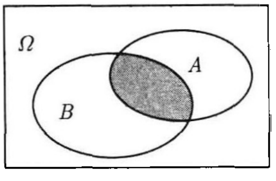

图 1.5

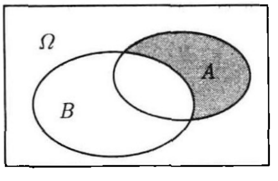

图 1.6

1.2.2　事件的运算性质

下面仅列出事件运算所满足的法则:

(1) **交换律**　$A \cup B = B \cup A$; $A \cap B = B \cap A$.

(2) **结合律**　$(A \cup B) \cup C = A \cup (B \cup C)$; $(A \cap B) \cap C = A \cap (B \cap C)$.

(3) **分配律**　$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

(4) **对偶律**　$\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$.