普通高等教育“十一五”规划教材

概率论与数理统计

主　编　徐雅静
副主编　段清堂　汪远征

科学出版社
www.sciencep.com

普通高等教育“十一五”规划教材

概率论与数理统计

主　编　徐雅静
副主编　段清堂　汪远征

科学出版社
北　京

内 容 简 介

本书由概率论与数理统计两部分组成. 概率论部分(第 1～5 章)包括随机变量及其分布、随机变量的数字特征、大数定理和中心极限定理等研究随机性问题的理论基础. 数理统计部分(第 6～10 章)包括参数估计、假设检验、回归分析和方差分析等基本统计分析方法.

本书增加了概率统计的发展历史简介、概率统计应用方面的例题及应用案例,以激发学生的学习兴趣;引进计算机数据处理与统计分析技术,便于学生对理论的理解和应用方法的掌握;增加实验内容,图文并茂地给出实验过程,便于学生自学掌握.

本书可供普通高等院校工科类、理科类(非数学专业)、经济管理类有关专业的大学生作为教材使用,也可供广大教师和相关人员参考.

图书在版编目(CIP)数据

概率论与数理统计/徐雅静主编. —北京:科学出版社,2009
(普通高等教育"十一五"规划教材)
ISBN 978-7-03-025178-7

Ⅰ.概… Ⅱ.徐… 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(2009)第 137558 号

责任编辑:李晓鹏 赵 靖 / 责任校对:朱光光
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

*

2009 年 8 月第 一 版 开本:B5(720×1000)
2009 年 8 月第一次印刷 印张:19 1/2
印数:1—5 000 字数:375 000

定价: 27.00 元

前　　言

“概率论与数理统计”是一门研究随机现象的应用性很强又颇具特色的数学学科，它在工程技术、科学研究、经济管理、企业管理、人文社科等众多领域都有广泛的应用. 它的理论与方法向各个学科渗透，是近代科学技术发展的特征之一，并由此产生了许多新的交叉学科，如生物统计、医学统计、统计物理、数学地质等. 它又是许多新兴的重要学科的基础，如信息论、控制论、可靠性理论、人工智能、信息编码理论和数据挖掘等. “概率论与数理统计”改变了原有单一学科发展的思路，对各门学科的发展具有极大的支撑作用，它的方法可以帮助各个领域的研究工作者更快地获得成功. 在理论联系实际方面，它是数学学科中最活跃、最有发展前途的分支之一.

“概率论与数理统计”是高校理、工、农、医、经济类、管理类等本科专业必修的一门重要的基础课，也是相关专业硕士研究生入学考试的一门必考科目，更是本科学生运用随机思维模式解决本专业相关问题的实用课程. 由于该课程的理论内容抽象深奥，应用方法灵活多样，应用中涉及大量复杂的公式和数值计算，偏重理论的传统教材和手工计算方法限制了它的应用，而现代计算机技术和众多统计软件早已为解决这一问题提供了条件. 为适应现代教学理念、方法和手段，改变学生认为“概率论与数理统计”难学、难记、不会用的现状，我们编写了这部《概率论与数理统计》教材.

本书由概率论与数理统计两部分组成. 概率论部分(第 1～5 章)包括随机变量及其分布、随机变量的数字特征、大数定理和中心极限定理等研究随机性问题的理论基础. 数理统计部分(第 6～10 章)包括参数估计、假设检验、回归分析和方差分析等基本统计分析方法.

本书在编写过程中认真贯彻落实教育部“高等教育面向 21 世纪教学内容和课程体系改革计划”的精神，并严格执行教育部数学与统计学教学指导委员会最新修订的工科及经济管理类“概率论与数理统计教学基本要求”，同时参考了近几年来国内外出版的有关教材，结合了编者多年来一线教学的实践经验. 其主要特色如下：

(1) 增加概率统计的发展历史简介、概率统计应用方面的例题及应用案例，以激发学生的学习兴趣.

在各章的首部增加了应用案例，通过案例教学，使抽象的概念、理论与实际应用紧密结合，便于学生理解与应用. 各章最后对应用案例的分析和解答旨在培养学

生综合运用有关理论和方法分析问题、解决问题的能力，把数学建模的思想方法融入到概率论与数理统计的课堂教学中.

选用和编制了一些能够反映当代概率论与数理统计应用方向的例题和习题，这些例题和习题不但具有一定的启发性，也能够使读者体会到概率论与数理统计在现实生活中应用的广泛性.

(2) 略去一些复杂的理论推导和烦琐的手工计算，引进计算机数据处理与统计分析技术，重点放在对理论的理解和应用方法的掌握上.

"概率论与数理统计"中有相当部分的理论推导，这虽然可以培养学生的理论分析能力，但对理解基本概念及解决实际问题帮助不大. 考虑到课程学时的限制，我们对理论推导部分做了弱化处理，在计算上尽量借助于计算机来完成，把重点放在对实际问题的分析和基本方法运用上. 在讲清基本理论的基础上，尽量做到与理工科、经济管理类专业的实际需要相联系，突出其应用性的特点.

(3) 增加概率统计实验内容，图文并茂地给出实验过程，便于学生自学掌握.

对于概率论与数理统计中较为复杂的计算，我们引入了应用非常广泛的数据处理软件——Excel 软件，并设计了大量实验. 在众多统计软件中，之所以选择 Excel 软件，是因为 Excel 软件的基本操作已为广大师生所熟悉，对于概率计算和统计数据的处理与分析，只需通过简单自学即可上机操作.

在 Excel 实验的设计中，我们采用了两种不同的处理方式. 实验方法一是实验型：直接利用 Excel 的函数和公式完成实验内容，帮助学生加深对所学知识的理解和记忆，还可以通过改变参数或个别数据观察动态的计算结果，便于教师课堂演示教学. 实验方法二是应用型：直接使用 Excel 的"数据分析"工具完成实验内容，简单、便捷，以应用为目的.

实验内容的学习可以根据教学时数，采用全讲、选讲或学生自学的方式进行.

(4) 根据各章内容特点设计了不同类型的习题，学生可根据需要有选择地进行练习.

本书大部分章节的习题分 A、B 两部分，A 中配备了填空题、选择题、解答题和应用题，希望通过这些基本题目的训练，学生能够较好地掌握基本概念、理论和基本方法；B 中选用了往年的部分考研题或较难的题目，旨在强化学生对理论内容的深入理解和掌握. 应用题大多需要借助于计算机完成，旨在训练学生利用计算机进行数据处理和统计分析解决实际问题的能力.

学习和使用本书需要读者具备"高等数学"与"线性代数"课程的基本知识. 本书知识系统、详略得当、举例丰富、讲解透彻、难度适宜，适合作为普通高等院校工科类、理科类(非数学专业)、经济管理类有关专业的"概率论与数理统计"课程的教材，也可作为相关专业人员和广大教师的参考资料.

本书由徐雅静任主编，段清堂、汪远征任副主编. 第 1 章由李鹏、杨林、贺勤编

写,第 2、3 章由段清堂、黄士国、荣民希编写,第 4、5 章由曲双红编写,第 6 章及附录由汪远征、卢金梅编写,第 7 章由焦万堂编写,第 8 章由徐雅静编写,第 9 章由徐雅静、成军祥编写,第 10 章由成军祥编写.全书由徐雅静统稿.

在本书的编写过程中,得到郑州轻工业学院、河南理工大学、河南工业大学等院校的大力支持,在此表示感谢!

限于编者水平,以及编写时间仓促,书中难免有错漏与不当之处,恳请读者批评指正.

编 者

2009 年 5 月

目　录

第1章 概率论基础

概率论是从数量化的角度来研究现实世界中一类不确定现象(随机现象)规律性的一门数学学科,20世纪以来,广泛应用于工业、国防、国民经济及工程技术等各个领域.本章介绍随机事件与概率、古典概型与几何概型、条件概率与乘法公式等概率论中最基本、最重要的概念和概率计算方法.

【概率论简史】

概率的概念形成于16世纪,与用投掷骰子的方法进行赌博有密切的关系.

1654年,一个名叫德梅尔(De Méré,法)的赌徒就"两个赌徒约定赌若干局,且谁先赢 c 局便算赢家,若在一赌徒胜 a 局($a<c$),另一赌徒胜 b 局($b<c$)时便终止赌博,问应如何分赌本"为题求教于数学家帕斯卡(Pascal,法,1623~1662).帕斯卡与费马(Fermat,法,1601~1665)通信讨论了这一问题,并用组合的方法给出了正确的解答.

1657年,惠更斯(Huyghens,荷,1629~1695)发表的《论赌博中的计算》是最早的概率论著作.论著中的第一批概率论概念(如数学期望)与定理(如概率加法、乘法定理)标志着概率论的诞生.

18世纪初,伯努利(Bernoulli,法,1700~1782)、棣莫弗(De Moivre,法,1667~1754)、蒲丰(Buffon,法,1707~1788)、拉普拉斯(Laplace,法,1749~1827)、高斯(Gauss,德,1777~1855)和泊松(Poisson,法,1781~1840)等一批数学家对概率论作出了奠基性的贡献.

1812年,拉普拉斯所著的《概率的分析理论》实现了从组合技巧向分析方法的过渡,开辟了概率论发展的新时期.

19世纪后期,极限理论的发展成为概率论研究的中心课题,是概率论的又一次飞跃,为后来数理统计的产生和应用奠定了基础.切比雪夫(Chebyshev,苏,1821~1894)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.切比雪夫的成果后来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.

1933年,柯尔莫哥洛夫(Kolmogorov,苏,1903~1987)在他的名著《概率论基础》一书中,提出了概率公理化定义,并得到数学家们的普遍承认.公理化体系给概率论提供了一个逻辑上的坚实基础,使概率论成为一门严格的演绎科学,取得了与其他数学学科同等的地位,并通过集合论与其他数学分支紧密联系起来.

在公理化的基础上，现代概率论不仅在理论上取得了一系列突破，在应用上也取得了巨大的成就，其应用几乎遍及所有的科学领域，如天气预报、地震预报、工程技术、自动控制、产品的抽样调查、经济研究、金融和管理等领域.

1.1　随机试验与样本空间

1.1.1　随机试验

客观世界中存在着两类现象：一类是在一定条件下必然出现的现象，称为**必然现象**；另一类是在一定条件下可能出现也可能不出现的现象，称为**随机现象**.

在标准大气压下，100℃的纯水必然沸腾；向上抛一枚石子必然下落；同性电荷必不相互吸引等都是必然现象的例子. 以往的各种数学学科的主要内容就是研究必然现象中的数量规律.

自然界和社会上发生的随机现象也是广泛存在的. 例如，抛掷一枚硬币，有可能正面朝上，也有可能反面朝上；一天内进入某超市的顾客数；某种型号电视机的寿命. 这些都是随机现象的例子.

研究随机现象是通过随机试验来进行的，概率论中把满足以下特点的试验称为**随机试验**：

(1) 可以在相同条件下重复进行；

(2) 每次试验的可能结果不止一种，并且能事先明确试验的所有可能结果；

(3) 进行一次试验之前不能确定哪一种结果会出现.

随机试验通常用大写字母 E 表示.

虽然一次随机试验的结果不能完全预言，但是，在相同条件下大量重复此试验时，则会呈现出一定的数量规律性，这种在大量重复试验中所呈现出的固有的规律性称为**统计规律性**. 例如，多次重复抛一枚硬币得到正面朝上大致有一半，同一门炮射击同一目标的弹着点按照一定规律分布，等等. 概率论就是研究随机现象中数量规律性的一门数学学科.

随机试验(以后简称试验)是一个广泛的术语，它既包括各种科学实验，也包括对客观事物进行的“调查”、“观察”、“测量”等.

1.1.2　样本空间

定义 1.1　随机试验的一切可能基本结果组成的集合称为**样本空间**，记为 $\Omega=\{\omega\}$，其中 ω 表示基本结果，又称为**样本点**.

研究随机现象首先要了解它的样本空间.

【例 1.1】　下面给出几个随机试验的样本空间.

“抛一枚硬币观察哪一面朝上”：$\Omega_1=\{$正面向上，反面向上$\}$.

“抛一颗骰子观察朝上一面的点数”:$\Omega_2=\{1,2,3,4,5,6\}$.

“某品牌电视机的寿命”:$\Omega_3=\{t\mid t\geqslant 0\}$.

“110 每天接到的报警次数”:$\Omega_4=\{0,1,2,\cdots\}$.

“圆心在原点的单位圆内任取一点”:$\Omega_5=\{(x,y)\mid x^2+y^2\leqslant 1\}$.

关于样本空间的几点说明:

(1) 样本空间中的元素可以是数也可以不是数;

(2) 样本空间中的样本点可以是有限多个的,也可以是无限多个的. 仅含两个样本点的样本空间是最简单的样本空间.

1.2 随机事件及其概率

1.2.1 随机事件

定义 1.2 随机试验的若干个基本结果组成的集合称为**随机事件**,简称**事件**,只含有一个基本结果的事件称为**基本事件**.

随机事件常用大写英文字母 $A,B,C,\cdots$表示.

关于随机事件概念的几点说明:

(1) 任一事件 A 是相应样本空间的一个子集,基本事件就是只含有一个样本点的事件.

(2) 当子集 A 中某个样本点出现了,就说事件 A 发生了,或者说事件 A 发生当且仅当 A 中某个样本点出现了.

(3) 样本空间 Ω 包含所有的样本点,作为自身的子集,在每次试验中它总是发生的,称为**必然事件**. 空集$\varnothing$不包含任何样本点,它作为样本空间的子集,在每次试验中都不发生,称为**不可能事件**.

【例 1.2】 掷一颗骰子的样本空间为:$\Omega=\{1,2,3,4,5,6\}$.

事件 A=“出现 5 点”,它由 Ω 的单个样本点“5”组成,它是一个基本事件,可记为 $A=\{5\}$;

事件 B=“出现奇数点”,它由 Ω 的三个样本点“1,3,5”组成,可记为 $B=\{1,3,5\}$;

事件 C=“出现的点数不大于 6”,它由 Ω 的全部样本点“1,2,3,4,5,6”组成,是必然事件,可记为 $C=\Omega$.

事件 D=“出现的点数大于 6”,Ω 中任一样本点都不在 D 中,所以 D 是不可能事件,可记为 $D=\varnothing$.

1.2.2 事件间的关系及运算

事件是一个集合,因而事件间的关系及运算实质上是集合间的关系及运算,当

然，我们关心的是其概率含义. 下面对事件的讨论总是假设在同一个样本空间 Ω 中进行.

1. 事件间的关系

1）子事件

如果属于事件 A 的样本点也属于事件 B，则称 A 为 B 的子事件，记为 $A\subset B$. 其概率含义是：A 发生 B 必发生.

2）事件相等

如果事件 A 与事件 B 满足 $A\subset B$ 且 $B\subset A$，则称 A 与 B 相等，记为 $A=B$. 其概率含义是：A，B 中有一个发生另一个也必发生.

3）互不相容

如果事件 A 和 B 没有相同的样本点，则称 A 与 B 互不相容(或互斥). 其概率含义是：A，B 不同时发生.

2. 事件运算

1）事件 A 与 B 的和

事件 A 与 B 的和事件定义为：由至少属于 A，B 之一的样本点全体组成的集合，记为 $A\cup B$. 其概率含义是：A，B 至少有一个发生.

2）事件 A 与 B 的积

事件 A 与 B 的积事件定义为：由既属于 A 又属于 B 的样本点组成的集合，记为 $A\cap B$ 或 AB. 其概率含义是：事件 A 与 B 同时发生.

事件 A 与 B 互不相容当且仅当其积事件为不可能事件，即 $AB=\varnothing$.

事件的和与积运算可推广到有限个或可列个事件，假设有事件 $A_1,A_2,\cdots,A_n,\cdots$，则 $\bigcup\limits_{i=1}^{n}A_i$ 称为**有限和**；$\bigcup\limits_{i=1}^{\infty}A_i$ 称为**可列和**；$\bigcap\limits_{i=1}^{n}A_i$ 称为**有限积**；$\bigcap\limits_{i=1}^{\infty}A_i$ 称为**可列积**.

3）事件 A 与 B 的差

属于事件 A 而不属于事件 B 的样本点的全体组成的集合称为 A 与 B 的差事件，记为 $A-B$. 其概率含义是：A 发生而 B 不发生.

4）**对立事件**

由在 Ω 中而不在 A 中的样本点组成的集合称为 A 的对立事件(逆事件)记为 $\overline{A}$. 其概率含义是：A 不发生. 显然，$\overline{A}=\Omega-A$，$\overline{\overline{A}}=A$.

3. 事件运算满足的定律

事件的运算性质和集合的运算性质相同，设 A，B，C 为事件，则有

交换律：$A\cup B=B\cup A$；$AB=BA$.

结合律：$(A\cup B)\cup C=A\cup(B\cup C)$，$(AB)C=A(BC)$.

分配律：$(A\cup B)C=(AC)\cup(BC)$，$(AB)\cup C=(A\cup C)(B\cup C)$.

对偶律：$\overline{A\cup B}=\overline{A}\,\overline{B}$；$\overline{AB}=\overline{A}\cup\overline{B}$.

【例 1.3】 向指定目标射三枪，观察射中目标的情况. 用 A_1、A_2、A_3 分别表示事件“第一、二、三枪击中目标”，试用 A_1、A_2、A_3 表示以下各事件：

(1) 只击中第一枪；

(2) 三枪都没击中；

(3) 至少击中一枪；

(4) 至多击中两枪.

解 (1) 事件“只击中第一枪”，意味着第二枪不中，第三枪也不中. 所以可表示成 $A_1\overline{A}_2\overline{A}_3$.

(2) 事件“三枪都没击中”，就是事件“第一、二、三枪都未击中”，所以可表示成 $\overline{A}_1\overline{A}_2\overline{A}_3$.

(3) 事件“至少击中一枪”，就是事件“第一、二、三枪至少有一次击中”，所以可表示成 $A_1\cup A_2\cup A_3$ 或 $A_1\overline{A}_2\overline{A}_3\cup\overline{A}_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2A_3\cup A_1A_2\overline{A}_3\cup A_1\overline{A}_2A_3\cup\overline{A}_1A_2A_3\cup A_1A_2A_3$.

(4) 事件“至多击中两枪”意味着三枪不能同时击中，所以可表示成 $\overline{A_1A_2A_3}$ 或 $\overline{A}_1\cup\overline{A}_2\cup\overline{A}_3$.

1.2.3 事件的概率及性质

随机事件有其偶然性的一面，即在一次试验中它可能发生也可能不发生，但是在大量重复试验中它又呈现出内在的规律性，即它发生的可能性大小是确定的，且是可以度量的，所谓随机事件的概率，概括地说就是用来描述随机事件发生的可能性大小的数量指标，它是概率论中最基本的概念之一.

1. 频率与概率的统计定义

定义 1.3 设 E 为任一随机试验，A 为其任一事件，在相同条件下，把 E 独立地重复做 n 次，n_A 表示事件 A 在这 n 次试验中发生的次数(称为**频数**). 比值 $f_n(A)=n_A/n$ 称为事件 A 在这 n 次试验中发生的**频率**.

易知频率有如下性质：

(1) 对于任一事件 A，有 $0\leqslant f_n(A)\leqslant 1$；

(2) 对于必然事件 Ω，有 $f_n(\Omega)=1$；

(3) 对于互不相容的事件 A,B，有

$$f_n(A \cup B) = f_n(A) + f_n(B).$$

人们在实践中发现:在相同条件下重复进行同一试验,当试验次数 n 很大时,某事件 A 发生的频率具有一定的“稳定性”,就是说其值在某确定的数值上下摆动.例如,大量重复地抛掷一枚硬币,正面出现的频率总在 0.5 附近波动,随着抛掷次数的逐渐增大,频率逐渐稳定于 0.5. 历史上许多统计学家用抛硬币的方法对频率的稳定性进行验证,他们的结论如表 1.1 所示,这些结论都说明了随着试验次数的增加,频率向 0.5 收敛.

表 1.1　投币试验

试验者	掷币次数 n	正面次数 n_A	频率 n_A/n
De Morgan	2048	1061	0.5180
Buffon	4040	2048	0.5069
Kerrich	7000	3516	0.5022
Kerrich	8000	4034	0.5042
Kerrich	9000	4538	0.5042
Kerrich	10000	5067	0.5067
Feller	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005
Romanovsky	80640	40173	0.4982

一般来说,试验次数 n 越大,事件 A 发生的频率就越接近一个确定的数值.这个数值说明了事件 A 发生的可能性的大小,据此给出概率的统计定义.

定义 1.4　设有随机试验 E,若当试验的次数 n 充分大时,事件 A 发生的频率 $f_n(A)$稳定在某数 p 附近波动,则称数 p 为事件 A 的**概率**,记为:$P(A)=p$.

值得注意的是,概率的统计定义是以试验为基础的,但这并不是说概率取决于试验.事件 A 出现的概率是事件 A 的一种属性,也就是说完全决定于事件 A 本身的结果,是先于试验客观存在的.概率的统计定义只是描述性的,一般不能用来计算事件的概率.通常只能在 n 充分大时,以事件出现的频率作为事件概率的近似值.根据频率和概率的这种关系和理论研究的需要,受频率性质的启发,苏联数学家柯尔莫哥洛夫于 1933 年给出了概率的公理化定义.

2. 概率的公理化定义与性质

定义 1.5　设 Ω 是一随机试验的样本空间,对于该随机试验的每一个事件 A 赋予一个实数,记为 $P(A)$,称为事件 A 的概率,如果集合函数 $P(\cdot)$满足下列公理.

(1) 非负性:对于每一个事件 A,有 $P(A)\geqslant 0$.