概率论及统计应用

GAILULUN
JI
TONGJIYINGYONG

3

$P(AB)=P(A)P(B)$

汪忠志 主编

合肥工业大学出版社

概率论及统计应用

安徽工业大学"十五"规划教材

概率论及统计应用

主　编　汪忠志

编写组　范爱华　张敬和　侯为根

陈文波　汪忠志

合肥工业大学出版社

图书在版编目(CIP)数据

概率论及统计应用/汪忠志主编.—合肥:合肥工业大学出版社,2005.9
ISBN 7-81093-150-4

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.021

中国版本图书馆CIP数据核字(2005)第087051号

概率论及统计应用

主编 汪忠志　　　　责任编辑 朱移山

出 版	合肥工业大学出版社	版 次	2005年9月第1版
地 址	合肥市屯溪路193号	印 次	2006年1月第2次印刷
邮 编	230009	开 本	787×960 1/16
电 话	总编室:0551-2903038	印 张	23.25　　字 数 417千字
	发行部:0551-2903198	发 行	全国新华书店
网 址	www.hfutpress.com.cn	印 刷	安徽江淮印务有限责任公司
E-mail	press@hfutpress.com.cn	纸 张	山东光华纸业集团有限公司

ISBN 7-81093-150-4/O·21　　　　定价:25.00元

序

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，其应用得到更为长足的发展。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，故在编写过程中，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，尽力收集来自社会科学、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，从而使本书更具有时代气息。由此，学生不仅感受到概率统计的实用性，还会促使学生联想本专业的问题，激发学习和研究兴趣。

为了尽可能满足不同读者的需要，本书有一些小号字体或带 * 的内容和习题供读者选用，在学时有限时跳过这部分内容，并不影响本书的体系。本书编写了大量例题，有些事例是我们日常教学和在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，能有计算机和投影设备的配合，教学将会更为方便和有效。

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、数理学院和应用数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。由于作者才疏学浅，本书的缺陷和错误在所难免，

恳请广大教师和学生提出宝贵意见，我们将作进一步改进。

本书第一、二章由范爱华编写，第三、七章由张敬和编写，第四、六章由陈文波编写，第五、八章由汪忠志编写，侯为根编写每章附录并绘制了全部插图。

作 者

2005年9月

目　录

1　**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 …… 2

1.1.3　样本空间 …… 2

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 …… 15

1.4.5　全概率公式 …… 17

1.4.6　贝叶斯 Bayes 公式 …… 20

1.5　事件的独立性与相关性 …… 23

1.5.1　两个事件的独立性与相关性 …… 23

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　Bernoulli 概型 …… 28

综合例题 …… 34

2 一维随机变量及其分布 …… 49
2.1 随机变量及其分布 …… 49
2.1.1 随机变量的概念 …… 49
2.1.2 随机变量的分布函数 …… 51
2.2 离散型随机变量的分布函数 …… 52
2.3 离散型随机变量的概率函数(概率分布) …… 53
2.3.1 常见的离散型随机变量的概率分布 …… 54
2.3.2 缸的模型 …… 62
2.3.3 缸的模型应用实例 …… 62
2.4 连续型随机变量及其概率密度 …… 63
2.4.1 连续型随机变量及其概率密度函数 …… 63
2.4.2 常用的连续型随机变量 …… 66
2.5 随机变量的函数的分布 …… 77
2.5.1 离散型随机变量的函数的分布 …… 77
2.5.2 连续型随机变量的函数的分布 …… 78
综合例题 …… 82
3 多维随机变量及其分布 …… 92
3.1 二维随机变量及其分布 …… 92
3.1.1 二维随机变量 …… 92
3.1.2 二维随机变量的联合分布函数 …… 92
3.1.3 二维离散型随机变量的概率分布 …… 93
3.1.4 二维连续型随机变量及其联合概率分布 …… 94
3.1.5 n 维随机变量及分布函数 …… 96
3.1.6 几个常用的分布 …… 96
3.2 边缘分布 …… 99
3.2.1 离散型随机变量的边缘分布 …… 99
3.2.2 连续型随机变量的边缘分布 …… 101
3.3 二维随机变量条件分布 …… 103
3.3.1 二维离散型随机变量的条件分布律 …… 103
3.3.2 二维连续型随机变量的条件分布律 …… 105
3.4 随机变量的独立性 …… 107
3.4.1 二维随机变量的独立性 …… 107

3.4.2* 一般情形 …… 108
3.5 二维随机变量的函数分布 …… 113
3.5.1 和的分布 …… 113
3.5.2* 一般函数 $z = g(X, Y)$ 的分布 …… 118
3.5.3* 一般变换 …… 118
3.5.4 最大值、最小值的分布 …… 118
综合例题 …… 121

4 随机变量的数字特征 …… 133

4.1 数学期望 …… 133
4.1.1 数学期望的性质 …… 135
4.1.2 随机变量函数的数学期望 …… 138
4.1.3 数学期望的简单应用 …… 139
4.2 中位数、众数和分位数 …… 142
4.3 方差 …… 144
4.4 协方差及相关系数 …… 147
4.5 矩、协方差矩阵 …… 155
4.6 回归直线 …… 157
综合例题 …… 161

5 大数字律与中心极限定理 …… 178

5.1 大数定律 …… 178
5.1.1 问题的提出 …… 178
5.1.2 切比雪夫不等式与大数定律 …… 178
5.2 中心极限定理 …… 182
5.2.1 中心极限定理的提法 …… 182
5.2.2 中心极限定理 …… 186
5.2.3* 若干应用 …… 188
综合例题 …… 193

6 数理统计学的基本概念 …… 202

6.1 数理统计学的基本概念 …… 202
6.1.1 引例 …… 202
6.1.2 总体与样本 …… 204

6.1.3 统计量…… 206
6.1.4 经验分布函数…… 208
6.1.5 数理统计方法的特点…… 209
6.1.6 数理统计的基本思想…… 210
6.1.7 统计模型…… 210
6.2 正态样本统计量的抽样分布 …… 210
6.2.1 正态分布…… 211
6.2.2 χ^2(卡方)分布 …… 211
6.2.3 t 分布(学生分布)…… 214
6.2.4 F 分布 …… 216
综合例题 …… 221
7 **参数估计** …… 232
7.1 点估计概述 …… 232
7.1.1 频率替换法…… 232
7.1.2 矩估计法…… 233
7.1.3 极大似然法…… 234
7.1.4 极大似然估计的不变性原则…… 243
7.2 基于截尾样本的极大似然估计 …… 244
7.2.1 定时截尾…… 244
7.2.2 定数截尾…… 245
7.3 估计量优良性的评选标准 …… 246
7.3.1 无偏性…… 246
7.3.2 有效性…… 249
7.3.3 均方误差准则…… 249
7.3.4 一致性(相合性)…… 250
7.4 参数的区间估计 …… 252
7.4.1 枢轴量法…… 253
7.4.2 单个正态总体数学期望的区间估计…… 253
7.4.3 单个正态总体方差的区间估计…… 255
7.4.4 两个正态总体期望差的区间估计…… 257
7.4.5 两个正态总体方差比的区间估计…… 259
7.4.6 (0-1)分布参数的区间估计…… 260

7.4.7 单侧置信区间…… 262
综合例题 …… 265

8 假设检验 …… 278

8.1 假设检验的基本概念 …… 278
8.1.1 引例与问题的提法…… 278
8.1.2 假设检验的基本思想…… 279
8.1.3 假设检验的定义与步骤…… 280
8.2 正态总体参数的假设检验 …… 283
8.2.1 正态总体数学期望的假设检验…… 283
8.2.2 正态总体方差的假设检验…… 293
8.2.3 两正态总体期望差的假设检验…… 297
8.2.4 两正态总体方差比的假设检验…… 301
8.3* 似然比检验 …… 305
8.3.1 似然比检验的基本思想…… 306
8.3.2 似然比检验的一般步骤…… 307
8.4 两种类型的错误 …… 309
8.5 功效函数、用样本容量 n 来控制第二类错误…… 312
8.6 拟合优度检验 …… 315
8.6.1 非参数 χ^2 检验 …… 315
8.6.2 列联表独立性检验…… 321
综合例题 …… 328

习题答案…… 340

附　表…… 350

生活中最重要的问题其中占大多数实际上只是概率的问题。

——拉普拉斯

我又转念，见日光之下，快跑的人未必能赢，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

1 随机事件与概率

概率论是研究随机现象的规律性的数学分支，为了对随机现象的有关问题作出明确的数学描述，像其他数学学科一样，概率论具有自己的严格的体系和结构. 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

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，称之为**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，称之为**随机现象**.

在重力作用下，物体的位移随时间变化的函数 $x(t)$，是由二阶微分方程 $x''(t) = g$ 来描述，其中 g 为重力加速度，这是确定的、必然的.

随机现象也是广泛存在的. 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. 这些都是随机现象的例子. 许多影响事物发展的泅然因素的存在，是产生随机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**.一般地,一个随机试验要具有下列特点:

(1) 可重复性:试验原则上可在相同条件下重复进行;

(2) 可观察性:试验结果是可观察的,所有可能的结果是明确的;

(3) 随机性:每次试验将要出现的结果是不确定的,事先无法准确预知.

由于随机现象的结果事先无法预知,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.这一点被历史上许多人的试验所证明.表1.1列出Buffon等人连续抛掷均匀硬币所得的结果.从表中的数据可以看到,当抛掷次数很大时,正面出现的频率非常接近0.5,就是说,出现正面与出现反面的机会差不多各占一半.

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	4040	2048	0.5069
DeMorgan	4092	2048	0.5005
Feller	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005
Lomanovskii	80640	39699	0.4923

上面的试验结果表明,在相同条件下大量地重复某一随机试验时,各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.**概率论就是研究随机现象中数量规律的数学学科**.

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**,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**，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 $\Omega=\{$(正,正),(正,反),(反,正),(反,反)$\}$.这里，比如样本点 $\omega=$(正,反)表示第一枚硬币抛出正面而第二枚抛得反面.

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

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"表示一次射击未中，而以"1"表示命中.则样本空间 $\Omega=\{1,01,001,\cdots\}$.

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

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出现.如在例 1.1.2 中，若以每小时是否达到 5 次电话呼唤来区分这台电话机是否太繁忙，那么"不太繁忙"即不足 5 次的呼唤.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**，简称**事件**.通常用大写的字母 $A,B,\cdots$ 等表示.某事件发生，就是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 $\omega\in A$ 时发生.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**.相反，空集 Φ 不包含任何样本点，每次试验必定不发生，故称 Φ 为**不可能事件**.

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

如果事件 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 B **事件包含事件** A，或称**事件** A **包含于事件** B.记作 $B\supset A$ 或 $A\subset B$.

2. 事件相等

如果事件 A 包含事件 B，事件 B 也包含事件 A，则称**事件** A **与** B **相等**.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**，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**,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 的样本点所构成,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 $AB = \Phi$,则称 A 与 B 是**互不相容事件**.

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**,记作 $\bar{A}$. 易见,$\bar{A} = \Omega - A$.

8. 完备事件组

设 $A_1, A_2, \cdots, A_n, \cdots$ 是有限或可数个事件,如果其满足:

(1) $A_iA_j = \Phi, i \neq j, i, j = 1, 2, \cdots$

(2) $\bigcup\limits_i A_i = \Omega$

则称 $A_1, A_2, \cdots, A_n, \cdots$ **是一个完备事件组**. 显然,A 与 $\bar{A}$ 构成一个完备事件组.

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举如下:

表 1-2

符　号	集合论	概率论
Ω	全集	样本空间;必然事件
Φ	空集	不可能事件
$\omega \in \Omega$	Ω 中的点(或称元素)	样本点
$\{\omega\}$	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 中
$A = B$	集合 A 与 B 相等	事件 A 与事件 B 相等
$A \cup B$	集合 A 与 B 之并	事件 A 与事件 B 至少有一个发生
$A \cap B$	集合 A 与 B 之交	事件 A 与事件 B 同时发生
A	集合 A 之余集	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
$A \cap B = \Phi$	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

概率论研究的是随机现象的统计规律性.对于随机试验,如果仅知道可能出现哪些事件是不够的,更重要的是要知道各个事件发生可能性大小的量的描述(即数量化).这种量的大小我们称为事件的**概率**.

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 $m(0\leqslant m\leqslant n)$ 次,则量 $\frac{m}{n}$ 称为事件 A 在 n 次试验中发生的**频率**,记作 $f_n(A)$,即:$f_n(A)=\frac{m}{n}$.

它满足不等式:$0\leqslant\frac{m}{n}\leqslant 1$.

如果 A 是必然事件,有 $m=n$,则 $f_n(A)=\frac{m}{n}=1$;如果 A 是不可能事件,有 $m=0$,则 $f_n(A)=\frac{m}{n}=0$.就是说:必然事件的堑率为1,不可能事件的频率为0.

在掷硬币试验中,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 发生的频率围绕0.5这个数值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,而且试验次数越多,事件 A 的频率就越来越接近 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性,它是事件 A 在一次随机试验时发生可能性大小的度量.

定义 1.2.1 (概率的统计定义)在相同的条件下,进行大量重复试验,当试验次数充分大时,事件 A 的频率总围绕着某一确定值 p 作微小摆动,则 p 称为事件 A 的**概率**,记为 $P(A)$.

数字 p(即 $P(A)$)就是在一次试验中对事件 A 发生的可能性大小的数量描述.例如,用0.5来描述掷一枚均匀的硬币"正面"出现的可能性.在具体问题中,按统计定义来求出概率是不现实的.因此,在实际应用时,往往就简单地把频率当作概率来使用.

例 1.2.1 为了设计某路口向左转弯的汽车候车道,在每天交通最繁忙的时间(上午9点)观测候车数,共观测了60次(天),得数据如下:

表 1－3

等候车辆数	0	1	2	3	4	5	6	总和
出现的次(天)数	4	16	20	14	3	2	1	60
频　率	$\frac{4}{60}$	$\frac{16}{60}$	$\frac{20}{60}$	$\frac{14}{60}$	$\frac{3}{60}$	$\frac{2}{60}$	$\frac{1}{60}$	1

试求某天上午 9 时在该路口至少有 5 辆汽车在等候左转弯的概率.

解　设事件 A 表示"至少有 5 辆汽车在等候左转弯",在 60 次观测中,事件 A 发生的频率

$$f_{60}(A)=\frac{2+1}{60}=\frac{1}{20}=0.05$$

实际工作者往往就认为至少有 5 辆汽车在等候左转弯的概率为 0.05.

以频率取代概率在社会科学(例如经济科学)中已广泛地运用,即使试验次数 n 不大时也是如此.

1.2.2　组合记数

排列与组合是两类记数公式.它们的推导都基于如下两条记数原理:

(1) **乘法原理**　如果某事件需经 k 个步骤才能完成,做第一步有 m_1 种方法,做第二步有 m_2 种方法,…,做第 k 步有 m_k 种方法,那么完成这件事共有 $m_1\times m_2\times\cdots\times m_k$ 种方法.

(2) **加法原理**　如果某事件可由 k 类不同方法之一去完成,在第一类方法中又有 m_1 种方法完成,在第二类方法中又有 m_2 种方法完成,…,第 k 类方法中又有 m_k 种方法完成,那么完成这件事共有 $m_1+m_2+\cdots+m_k$ 种方法.

(3) **排列**　从 n 个不同元素中任取 $r(r\leqslant n)$ 个元素排成一列称为一个排列,按乘法原理,此种排列共有 $n\times(n-1)\times\cdots\times(n-r+1)$ 个,记为 P_n^r.若 $r=n$,称为全排列,全排列数共有 $n!$ 个,记为 P_n.

(4) **重复排列**　从 n 个不同元素中每次取出一个,放回后再取一个,如此连续取 r 次所得的排列称为重复排列,此种重复排列数共有 n^r 个.注意,这里的 r 允许大于 n.

(5) **组合**　从 n 个不同元素中任取 $r(r\leqslant n)$ 个元素组成一组(不考虑其间顺序)称为一个组合,按乘法原理,此种组合总数共有

$$\binom{n}{r}=\frac{P_n^r}{r!}=\frac{n(n-1)\cdots(n-r+1)}{r!}=\frac{n!}{r!(n-r)!}$$