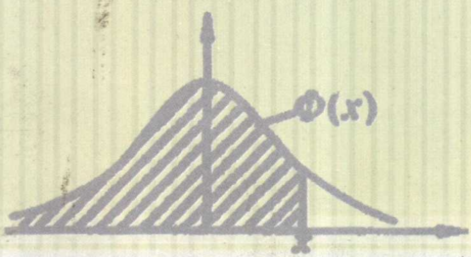
概率论与数理统计

马军英 编著



山东大学出版社
Shandong University Press

概率论与数理统计

马军英　编著

山东大学出版社

图书在版编目(CIP)数据

概率论与数理统计/马军英编著.—济南:山东大学出版社,2004.9

ISBN 7-5607-2858-8

Ⅰ.概…

Ⅱ.马…

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(2004)第 086370 号

山东大学出版社出版发行

(山东省济南市山大南路 27 号　邮政编码:250100)

山 东 省 新 华 书 店 经 销

山东旅科印务有限公司印刷

850×1168 毫米　1/32　12 印张　312 千字

2004 年 9 月第 1 版　2004 年 9 月第 1 次印刷

印数:1－3000 册

定价:19.00 元

前　言

概率论与数理统计是随机数学的两个分支,随机现象的非常广泛性自然就决定了它们在现代科学与应用中的重要性与广泛性。概率统计课已被列为高等学校理工科及管理、经济类等绝大多数专业的必修或选修课。从1997年以来,它被列入工学、管理学、经济学门类中的若干学科及专业硕士研究生入学统一考试的必考科目,尤其是从2003年开始,入学初试科目调整,数学科目作为专业基础课(一)进行考试,其权重在原来的基础上增加了50%,总分为150分,而概率统计内容占总分的25%~30%,这进一步说明了该课程作为专业基础课的重要地位,同时也极大地促进了该课程的教学。本书是在多年的教学实践、教学经验的基础上编写的,可作为高等学校理工科(非数学专业)及部分应用文科专业概率论与数理统计课程的教材,也可以作为自修本课程读者的读物。

本书在选材上力求贯彻"少而精,广而易懂"的原则,内容紧扣硕士研究生入学考试大纲,并以此规范文中的述语与记号。在内容安排上分概率与数理统计两大部分。概率论(1~5章)内容包括:随机事件及概率,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,极限定理。数理统计(6~9章)内容包括:数理统计的基本概念,参数估计,假设检验,方差分析与回归分析初步,其中第9章可根据专业的需要选用。本书在讲述上,对基本概念的引入尽可能通过人们所熟悉的事实直观描述,叙述力求简

洁、清楚，对推理论证在高等数学内容范围内尽可能严密，而对繁难的证明适当弱化，但对于相关知识解决什么问题都作了较为明确的解释。本书例题丰富，注意安排概率统计在工程、经济、教育及管理、生物等各专业的应用实例。本书各章后都配有相当数量的习题，其中一部分有一定的综合性与难度，可作为硕士研究生入学应试的练习题，题型包含填空、选择、计算、证明等，书末附有各章的习题答案，以供读者参考。

本书在编写过程中得到了张立琴、王爱云教授的大力支持和热情帮助。王秀丽、张玉芬、张燕老师也提出了许多宝贵意见，在此一并表示衷心的谢意。

由于编者水平所限，本书难免存在一些不足或不当之处，恳请读者批评指正。

编　者

2004 年 5 月

目 录

第 1 章　随机事件及其概率

概率论与数理统计是研究随机现象统计规律性的科学. 本章将主要介绍随机事件,随机事件的概率,概率的基本性质,条件概率和三个重要的概率公式——乘法公式、全概率公式和贝叶斯公式以及随机事件的独立性.

1.1　随机事件

在自然界与人类社会中,人们所能观察到的现象是各种各样的,但归纳起来,大体可以分为两类:一类是**确定性现象**,另一类是**随机现象**. 所谓确定性现象,即在一定条件下必然发生的现象. 例如,一枚硬币向上抛起后必然会落地;没有水分,种子不会发芽;直角三角形,斜边边长的平方是另两直角边边长的平方之和. 所谓随机现象,即在一定条件实现后,可能发生也可能不发生的现象. 例如,上抛一枚硬币,落地时是币值面(称为正面)向上还是徽花面(反面)向上,这在上抛前是不可确定的;一穴播两粒种子,可能两粒都出苗,也可能只有一粒出苗,还可能两粒都不出苗;一批新产品投入市场,可能畅销,也可能滞销;从事某项科学研究工作,其结果可能成功,也可能失败;炮手用同一门火炮向同一目标射击,无论怎样控制射击条件不变,各次弹着点总不尽相同. 尽管随机现象在一次或少数几次观察中其结果呈现出不确定性,但在作出大量重复观察或试验时,又会呈现出一定的规律性. 如,人们从长期实

践中知道，多次重复上抛同一枚均匀硬币，出现正面向上的次数占一半左右；同一门炮射击同一目标的弹着点按一定规律分布. 我们把这种在相同条件下进行大量重复试验，随机现象所呈现的规律性称为随机现象的**统计规律性**. 正是由于这种统计规律性，才诱发了人们叩开概率论与数理统计这扇科学大门的欲望和激情，使概率论与数理统计成为研究和揭示随机现象统计规律性的一个数学分支.

1.1.1 随机试验

试验，通常是指对现象的观察.

如果试验满足：

(1) 可在相同条件下重复进行；

(2) 每次试验可出现多种可能结果；

(3) 每次试验前能明确试验的所有可能结果，但不能确定试验后会出现哪一个结果.

则称该试验为**随机试验**，简称**试验**，用 E 表示.

下面给出一些随机试验的例子.

例 1.1 掷一枚均匀对称的骰子，观察着地时的点数.

例 1.2 记录一段时间内，某城市 110 报警次数.

例 1.3 从含有三件正品 a_1, a_2, a_3 和二件次品 b_1, b_2 的五件产品中任取二件，观察出现正品、次品的情况.

例 1.4 从一批电脑中，任取一台观察无故障运行的时间.

1.1.2 随机事件

随机试验的每一个可能的结果称为一个**随机事件**，简称**事件**. 随机事件通常用大写字母 $A, B, C, \cdots$ 表示. 例如，例 1.1 中随机试验可能的结果{出现 1 点}，{出现 2 点}，…，{出现 6 点}，这些都是随机事件. 再如例 1.2 中随机试验的结果{出现 0 次报警}，{出现 1 次报警}，…，{出现多于 10 次报警}，…等也是随机事件. 事件

是概率论中最基本的概念,正确地表示随机试验的各事件是学好概率论的基础.

事件又分基本事件和复合事件.**基本事件**是指不能再分解的事件.如例1.1中{出现1点},…,{出现6点},例1.2中{出现0次报警},…,{出现 n 次报警}等都是基本事件.**复合事件**是指由若干基本事件组成的事件.如例1.1中{出现奇数点},{出现偶数点},例1.2中{出现多于10次报警}等都是复合事件.

事件中有两个极端情况值得注意:一个是每次试验都发生的事件,称为**必然事件**,记为 Ω;另一个是每次试验中都不发生的事件,称为**不可能事件**,记为 $\varnothing$.显然,必然事件、不可能事件都是确定性事件,为了今后讨论问题的方便,也可将它们看作是两个特殊的随机事件.

1.1.3　样本空间

为了用数学方法描述随机现象,需要引入样本空间的概念.

随机试验 E 产生的所有基本事件构成的集合称为**样本空间**,记为 Ω.称其中的每个元素(基本事件)为一个**样本点**,记为 ω.即 $\Omega=\{\omega\}$.

例1.1(续)　在例1.1中,样本空间 Ω_1 有6个基本事件,即出现1点,2点,…,6点.故

$$\Omega_1=\{1,2,3,4,5,6\}.$$

例1.2(续)　在例1.2中,样本空间为

$$\Omega_2=\{0,1,2,\cdots\}.$$

例1.3(续)　在例1.3中样本空间为

$$\Omega_3=\{(a_1,a_2),(a_1,a_3),(a_1,b_1),(a_1,b_2),(a_2,a_3),(a_2,b_1),(a_2,b_2),(a_3,b_1),(a_3,b_2),(b_1,b_2)\}.$$

例1.4(续)　在例1.4中样本空间为

$$\Omega_4=\{t\mid t\geqslant 0,t\text{ 为无故障运行的时间,单位:小时}\}.$$

从以上例子中我们看到，样本空间可以是数集，也可以不是数集；可以是有限集，也可以是无限集.

有了随机事件和样本空间的概念后，对于随机试验 E 中的任何一个事件 A，都可以理解为该试验的样本空间 Ω 的子集. 现在，从集合论的观点再对上述的一些事件作如下说明：在随机试验 E 中，随机事件是以 E 中样本点为元素的集合；基本事件是只包含一个样本点的单元素集合；必然事件是由 E 中全体样本点组成集合，即 E 的样本空间 Ω；不可能事件是不包含 E 中任何一个样本点的集合，即空集 $\varnothing$. 而且，所谓**事件 A 发生**，是指当且仅当 A 中所包含的某个样本点在试验 E 中出现.

1.1.4　事件间的关系及运算

由于一个随机试验可产生许多随机事件，这些事件中有的简单，有的复杂. 概率论的重要任务之一就是研究随机事件的规律性，特别需要通过对简单事件的研究去把握复杂事件的规律性，这就导致了对事件间的关系及运算的研究.

1.1.4.1　事件的包含与相等

若事件 A 发生必然导致事件 B 发生，则称事件 B **包含**事件 A，记作 $A \subset B$ 或 $B \supset A$. 事件 B 包含事件 A，即 A 中的每个样本点必然在 B 中.

若事件 A,B 满足：$A \subset B$ 且 $B \subset A$，则称事件 A 与事件 B **相等**(或称**等价**). 记作 $A = B$.

例 1.5　在某条公路上随机抽查 8 辆汽车，考察违章车辆数. 其样本空间 $\Omega = \{0,1,2,\cdots,8\}$. 设 A 表示事件{违章车不超过 3 辆}，即 $A = \{0,1,2,3\}$；B 表示事件{有 2 辆或 3 辆违章}，即 $B = \{2,3\}$；C 表示事件{有 2 至 5 辆违章}，即 $C = \{2,3,4,5\}$；D 表示事件{有 4 至 8 辆违章}，即 $D = \{4,5,6,7,8\}$；E 表示事件{违章车辆数不少于 2 辆且不多于 5 辆}，即 $E = \{2,3,4,5\}$；F 表示事件{违章车辆数多于 4 辆}，即 $F = \{5,6,7,8\}$.

由事件的包含与相等定义易知，$B \subset A$，$C = E$.

1.1.4.2　事件的和(并)

事件 A 与事件 B 至少有一个发生所构成的事件，称为事件 A 与 B 的**和**(或**并**)，记作 $A \cup B$.

如例 1.5 中，$A \cup C$ 表示事件{违章车不超过 5 辆}，即 $A \cup C = \{0,1,2,3,4,5\}$. 由此可知，和事件 $A \cup B$ 的样本点由 A 和 B 的样本点组成，但公共的样本点只能取一次.

一般地，n 个事件 $A_1, A_2, \cdots, A_n$ 中至少有一个发生的事件称为 $A_1, A_2, \cdots, A_n$ 的和(并)，记作 $A_1 \cup A_2 \cup \cdots \cup A_n$，或 $\bigcup\limits_{i=1}^{n} A_i$. 同样，无限可列个事件 $A_1, A_2, \cdots, A_n, \cdots$ 至少有一个发生的事件称为这可列个事件的和(并)，并记作 $A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots$，或 $\bigcup\limits_{i=1}^{\infty} A_i$.

1.1.4.3　事件的积(交)

事件 A 与 B 同时发生所构成的事件，称为事件 A 与 B 的**积**(**交**)，记作 $A \cap B$，简记为 AB. 积事件 $A \cap B$ 的样本点由既属于 A 又属于 B 的公共样本点组成.

如例 1.5 中，A 和 C 的积事件 $A \cap C = \{2,3\}$.

类似地，事件 $A_1, A_2, \cdots, A_n$ 同时发生所构成的事件称为 $A_1, A_2, \cdots, A_n$ 的积(交)，记作 $A_1 \cap A_2 \cap \cdots \cap A_n$ 或 $A_1 A_2 \cdots A_n$. 对无限可列的情形，用 $\bigcap\limits_{i=1}^{\infty} A_i$ 表示事件 $A_1, A_2, \cdots$ 同时发生所构成的事件.

1.1.4.4　事件的差

事件 A 发生且事件 B 不发生所构成的事件，称为事件 A 和 B 的**差**，记作 $A - B$. 差事件 $A - B$ 的样本点由属于 A 而不属于 B 的样本点组成.

如例 1.5 中，A 和 C 的差事件为 $A - C = \{0,1\}$.

1.1.4.5 互不相容事件(互斥事件)

事件 A 与事件 B 不能同时发生,即 $AB=\varnothing$,则称 A 和 B 是**互不相容(或互斥)事件**.两事件互斥时,它们就没有公共的样本点.

如例 1.5 中,A 和 F,A 和 D 都是互不相容事件.

若 n 个事件 $A_1,A_2,\cdots,A_n$ 中的任意两个事件都互不相容,即 $A_iA_j=\varnothing(i\neq j,i,j=1,2,\cdots,n)$,则称这 n 个事件**两两互不相容**.基本事件是两两互不相容的.

1.1.4.6 互逆事件(对立事件)

若事件 A 与事件 B 必发生其一,但又不能同时发生,即 $A\cup B=\Omega$. $AB=\varnothing$,则称事件 A 与 B **互为逆事件(对立事件)**,也可称作 A 是 B 的**逆事件**,或 B 是 A 的逆事件. A 的逆事件记作 $\overline{A}$. 显然 $A\cup\overline{A}=\Omega,A\overline{A}=\varnothing,\overline{\overline{A}}=A,\overline{A}=\Omega-A$.

如例 1.5 中,事件 A 和 D 互为逆事件.

请读者注意,互逆事件必为互不相容事件,而互不相容事件却未必是互逆事件.

事件间的关系与运算可以用图形直观表示.如图 1.1 所示,平面上的矩形域表示样本空间 Ω,矩形域内不同图形表示事件 A,B,阴影部分表示事件 A 和 B 的各种关系和运算结果.

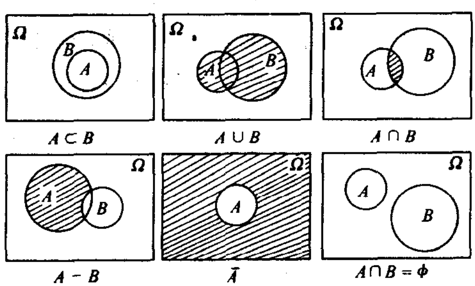


图 1.1

1.1.4.7　事件的运算性质

事件运算具有下列性质.

(1) 交换律：$AB=BA, A\cup B=B\cup A$；

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

(3) 分配律：$A(B\cup C)=AB\cup AC$,

$$A\cup(BC)=(A\cup B)(A\cup C);$$

(4) 德·摩根(De Morgan)对偶法则：

$$\overline{A\cup B}=\overline{A}\cap\overline{B}, \overline{A\cap B}=\overline{A}\cup\overline{B}.$$

此外，容易证明下列等式的正确性：

$A\cup A=A, A\cup\Omega=\Omega, A\cup\varnothing=A$,

$AA=A, A\Omega=A, A\varnothing=\varnothing$,

$A-B=A-AB=A\overline{B}, A\cup B=A\cup B\overline{A}$.

例 1.6　设 A,B,C 表示三个事件，用 A,B,C 表示下列事件：

(1) A 发生且 B 与 C 至少有一个发生；

(2) A 与 B 发生而 C 不发生；

(3) A,B,C 中恰有一个发生；

(4) A,B,C 中至少有两个发生；

(5) A,B,C 中至多有两个发生；

(6) A,B,C 中不多于一个发生.

解　(1) $B\cup C$ 表示 B 与 C 至少有一个发生，故所求事件可表示为 $A\cap(B\cup C)$.

(2) A 与 B 发生即 A 和 B 同时发生，就是 AB 发生，C 不发生即 $\overline{C}$ 发生，故所求事件可表示为 $AB\overline{C}$.

(3) A,B,C 中恰有一个发生没有指明究竟哪一个发生，因此这个事件可表示为 $A\overline{B}\,\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\,\overline{B}C$.

(4) 所求事件可表示为 $AB\cup BC\cup AC$.

(5) A,B,C 中至多有两个发生的逆事件为 ABC 发生，故这

个事件为$\overline{ABC}$.

(6) A,B,C,至多有一个发生意味着没有两个或两个以上的事件同时发生,因此这个事件可表示为$\overline{AB\cup BC\cup AC}$.

1.2 频率与概率

对于随机现象,我们不但要讨论它可能出现什么结果,更希望获得各种结果(随机事件)发生可能性大小的度量,这种度量应能反映随机现象所呈现出的统计规律性.本节,先引入频率,它描述了事件发生的频繁程度,进而引出刻画随机事件发生可能性大小的度量——概率.

1.2.1 频率与频率的稳定性

定义 1.1 如果在 n 次重复试验中,事件 A 发生了 μ_n 次,则称 μ_n 为事件 A 发生的**频数**,比值$\frac{\mu_n}{n}$称为事件 A 在 n 次试验中发生的**频率**,记为 $f_n(A)$,即

$$f_n(A)=\frac{\mu_n}{n}. \tag{1.1}$$

由于事件 A 发生的频率是它发生的次数与试验次数之比,所以,其大小表示 A 发生的频繁程度,频率大,事件 A 在一次试验中发生的可能性大;频率小,事件 A 在一次试验中发生的可能性小.

从频率的定义可知频率有下列性质:

(1) $0\leqslant f_n(A)\leqslant 1$;

(2) $f_n(\Omega)=1$;

(3) 设 $A_1,A_2,\cdots,A_k$ 是两两互不相容的事件,则

$$f_n(A_1\cup A_2\cup\cdots\cup A_k)=f_n(A_1)+f_n(A_2)+\cdots+f_n(A_k).$$

尽管随机事件 A 在一次试验中发生与否是偶然的,但在大量

的试验中,事件 A 发生的频率却随着试验次数的增大总在某一确定的数值(称为频率的稳定值)附近摆动,这种规律性称为**频率的稳定性**.

例如,在抛掷一枚均匀硬币时,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.但是,由于硬币是均匀的,直观上看来出现正面与出现反面的机会应该相等,亦即在大量的试验中,出现正面的频率应在 0.5 左右,历史上曾有不少人做过试验,其结果见表 1.1.

表 1.1

试验者	试验次数 n	发生正面的次数 μ_n	频率 $\frac{\mu_n}{n}$
德·摩根	2048	1039	0.5073
布丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

又如,曾有人统计过某个国家每年因没有写清地址或其他原因无法投递的信件数,这从常识上看,似乎没有什么规律性,但是经过统计之后惊人地发现,一年中这类信件数在全体信件中所占比例许多年几乎保持不变.见表 1.2.

表 1.2

年　份	信件总数 n	无法投递的信件数 μ_n	频率 $\frac{\mu_n}{n}$
1906	$983\cdot10^6$	54861	$56\cdot10^{-6}$
1907	$1076\cdot10^6$	53500	$50\cdot10^{-6}$
1908	$1214\cdot10^6$	59627	$49\cdot10^{-6}$
1909	$1357\cdot10^6$	62088	$46\cdot10^{-6}$
1910	$1507\cdot10^6$	76614	$51\cdot10^{-6}$