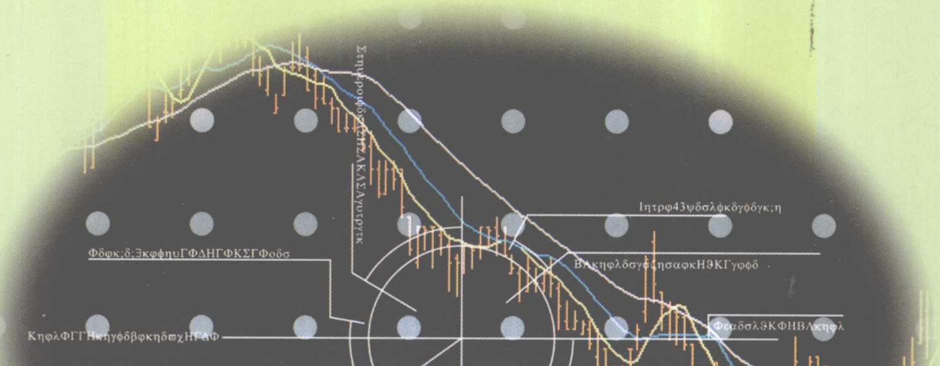
高等农林院校十五规划教材

全国高等农业院校教材

全国高等农业院校教学指导委员会审定

概率论与数理统计

程述汉　葛家麒　主编



中国农业出版社

8

全国高等农业院校教材
全国高等农业院校教学指导委员会审定

概率论与数理统计

程述汉　葛家麒　主编

中国农业出版社

图书在版编目（CIP）数据

概率论与数理统计/程述汉，葛家麒主编．—北京：中国农业出版社，2005.1
全国高等农业院校教材
ISBN 7-109-09557-6

Ⅰ．概…　Ⅱ．①程…②葛…　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135799号

中国农业出版社出版
（北京市朝阳区农展馆北路2号）
（邮政编码 100026）
责任编辑　朱　雷

北京通州皇家印刷厂印刷　　新华书店北京发行所发行
2005年2月第1版　　2006年12月北京第5次印刷

开本：787mm×960mm 1/16　　印张：17.75
字数：311千字
定价：23.80元

编写人员名单

主　编　程述汉　葛家麒

副主编　郑学高　周玉元

参　编　于　滨　刘素青

丁孝全　苏本堂

前　言

数学作为一门研究现实世界数量关系和空间关系的科学，在其产生和发展的历史长河中，一直和人们的生产活动及许多学科的发展密切相关.数学的特点不仅在于概念的抽象性、逻辑的严密性和结论的明确性，而且在于其应用的广泛性.随着计算机科学与技术的发展，使计算量不再成为数学方法应用的制约因素，促进了数学向各个领域的进一步渗透.

概率论与数理统计是数学的一个分支，也是高等农林院校本科数学教育中的一门主要课程，用于研究随机现象及统计规律性，是研究受随机性影响的数据的重要工具，其内容博大精深，其理论与方法对学习后继课程、研究其他学科有重要的借鉴作用.某些专业的硕士研究生入学考试、留学国外的资格考试试题中，也有一定比例的概率论与数理统计考题.

尽管概率论与数理统计是一门应用性很强的学科，但并不能把教材等同于应用手册.要想用好统计方法，除应熟悉与研究问题相关的专业知识外，对统计概念的直观解释及对统计方法的理论认识，都是非常重要的.因此，本书突出了如下特点.

一、保持体系完整.强调了数学的工具作用和培养学生逻辑思维能力的作用.既照顾到数学知识的系统性，又注重统计思想的传授.

二、追求简明实用.删略了一些繁琐的理论证明，详细叙述各种统计方法的计算过程，突出了数学的应用性.

三、增加数理统计实验.简明实用地综合介绍 EXCEL 和 MATLAB 数理统计方法，使读者容易使用这两大工具进行数据分析.这是本书作者为培养学生应用数学方法解决实际问题的能力所达成的共识.

本书由山东农业大学程述汉教授、东北农业大学葛家麒教授担任主编，西南农业大学郑学高副教授、湖南农业大学周玉元副教授担任

副主编.参加编写人员为程述汉(第六章§6.1、§6.2,第九章,第十章,第十一章§11.7),葛家麒(第二章,第三章),郑学高(第一章),周玉元(第四章,第五章),大连水产学院于滨(第七章),湛江海洋大学刘素青(第八章),山东农业大学丁孝全(第六章§6.3、第十一章§11.5、§11.6),山东农业大学苏本堂(第十一章§11.1、§11.2、§11.3、§11.4).

中国科学院数学与系统科学研究院张永光研究员仔细审阅了本书,并提出了许多宝贵意见,在此表示衷心感谢.

山东农业大学信息科学与工程学院丁世飞教授、张军本副教授给本书第十一章提供了有用的素材,并对本书的编写提出了宝贵意见.在此一并致谢.

书中疏漏和不妥之处,敬请读者批评指正.

编　者

2004年12月

目　　录

第一章　事件与概率

在自然界和人类社会中,人们观察到的现象大体可分为两类.一类是事前可预言的,即在一定条件下可以预言它一定会出现或一定不会出现.例如,“纯种紫花豌豆的后代开紫花”,“平面上三角形的内角和是 180°”,“在 101325Pa 大气压下,水加热到 100℃时沸腾”等现象必然发生;而“同性电荷相互吸引”,“早晨,太阳从西方升起”等现象必然不会发生.这类由试验条件完全决定其是否发生的现象称为**确定性现象**(或**决定性现象**).另一类现象是事前无法预言的,即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**.例如,抛掷一枚均匀硬币,可能出现正面,也可能出现反面,掷前无法确定会出现哪一面;幸运抽奖时,一张奖券可能中奖,也可能不中奖,事前无法预测,等等.

概率论与数理统计是研究和揭示随机现象统计规律性的一门数学学科,其理论与方法得到了广泛的应用.例如,使用概率统计方法进行气象预报、水文预报和地震预报,产品的抽样验收;在新产品研制时,为寻求最佳生产条件而进行试验设计和数据处理;在可靠性工程中,使用概率统计方法给出元件或系统的可靠性及平均寿命的估计;在自动控制中给出数学模型以便通过计算机控制工业生产等.其应用几乎遍及所有科学技术领域、工农业生产和国民经济各个部门中,而且还在不断向众多学科渗透并与之结合发展.这些特点不仅使概率论与数理统计成为十分活跃的数学分支,而且也是近代科学技术发展的特征之一.

§1.1　随机事件及其运算

1.1.1　样本空间与随机事件

为了叙述方便,我们把对自然现象、社会现象所进行的观察或科学实验,统称为**试验**.许多试验具有以下三个特点:

(1) 在相同的条件下可以重复进行;

(2) 每次试验的可能结果不止一个,且在试验之前已知试验的所有可能结果;

(3) 每次试验总是恰好出现这些可能结果中的一个,但在试验之前不能肯定会出现哪一个结果.

我们将具有上述三个特点的试验称为**随机试验**,简称**试验**,并记作 E.本书

中以后提到的试验均为随机试验.下面列举一些随机试验的例子.

E_1:抛掷一枚硬币,观察正面 H、反面 T 出现的情况;

E_2:从批量棉花种子中取 20 粒,观察发芽的种子数;

E_3:记录某公共汽车站某时刻的等车人数;

E_4:从三月龄的鸡群中,随机地抽取一只鸡,称其重量;

E_5:向平面上某目标射击,观察弹着点的位置;

E_6:从一批灯泡中任取一只,测试其寿命.

定义 1 随机试验 E 中可能出现的全部试验结果所组成的集合称为 E 的**样本空间**,记为 Ω.样本空间的元素称为**样本点**,记为 ω,即有 $\Omega=\{\omega\}$.

为研究方便起见,通常限定 Ω 中的试验结果,在每次试验中都有一个且仅有一个出现,即 **Ω 中的元素是最基本的、不能再分解的试验结果**.

记上述随机试验 E_k 的样本空间为 $\Omega_k(k=1,2,\cdots,6)$,则容易得到:

$\Omega_1=\{H,T\}$, $H=\{$出现正面$\}$, $T=\{$出现反面$\}$;

$\Omega_2=\{0,1,2,\cdots,20\}$, $\omega_i=i$ 表示$\{$有 i 粒种子发芽$\}$;

$\Omega_3=\{0,1,2,\cdots\}$, $\omega_i=i$ 表示$\{$等车人数为 $i\}$;

$\Omega_4=\{w \mid 0<w<\infty\}$, w 表示鸡的重量;

$\Omega_5=\{(x,y) \mid -\infty<x<+\infty, -\infty<y<+\infty\}$, (x,y)为弹着点的坐标;

$\Omega_6=\{t \mid 0\leqslant t<\infty\}$, t 为灯泡寿命.

建立样本空间事实上就是建立随机现象的数学模型,一个抽象的样本空间可以概括许多内容不相同的实际问题,例如 Ω_1 是只包含两个样本点的样本空间,但它既可以作为掷硬币出现正面或反面的模型,也可作为产品检验中产品合格与不合格的模型,还可用于公用事业排队现象中有人排队与无人排队的模型,以及作为气象预报中下雨与不下雨的模型等等.这说明尽管问题的实际内容不同,但有时却能归结为相同的概率模型.因此,我们常以抛掷硬币、摸球等这样一些既典型又形象且易于理解的例子阐明一些问题,以便使问题的阐述更明确,且使问题的本质更为突出.

称试验 E 的样本空间 Ω 的子集为 E 的**随机事件**①,简称**事件**,常用大写字母 $A,B,C,\cdots\cdots$ 表示.在每次试验中,当且仅当这一子集中的一个样本点出现时就称这一**事件发生**.

特别地,由一个样本点组成的单点集合,称为**基本事件**,例如,试验 E_1 有两

① 一般地,并不把 Ω 的一切子集都作为事件,因为这样将会对给定概率带来困难.关于这方面的讨论已超出本书的范围.

个基本事件$\{H\}$和$\{T\}$；试验E_2有21个基本事件$\{0\},\{1\},\cdots,\{20\}$.

样本空间Ω是自身的子集，从而是随机事件，它包含所有的样本点，在每次试验中必然发生，称为**必然事件**. 空集$\varnothing$是Ω的子集，从而是随机事件，但它不包含任何样本点，故在每次试验中都不发生，称为**不可能事件**. 例如，掷一枚均匀骰子的试验中，$\Omega=\{$点数不大于6$\}$是一个必然事件，因为在试验中不论哪一个基本事件发生，均导致“点数不大于6”这一结果出现；$\varnothing=\{$点数大于6$\}$不包含任何样本点，是不可能事件.

1.1.2　事件间的关系与运算

事件是一个集合，因此事件间的关系和运算就可按照集合论中集合之间的关系和运算来处理.

设Ω为试验E的样本空间，而A、B、$A_k(k=1,2,\cdots)$是Ω的子集.

1. 如果事件A发生必然导致事件B发生，则称 **A 是 B 的子事件**，或称**事件 B 包含事件 A**. 记作$A\subset B$或$B\supset A$，可用图1-1直观表示.

2. 若$A\subset B$且$B\subset A$，则称事件 **A 与 B 相等**或称 **A 与 B 等价**，记作$A=B$. 直观地说，$A=B$即A、B中含有相同的样本点.

3. “事件A与B中至少有一个发生”也是一事件，称为事件A与B的**和**或**并**，记作$A\cup B$，即$A\cup B=\{\omega\mid\omega\in A$或者$\omega\in B\}$，它的几何表示如图1-2.

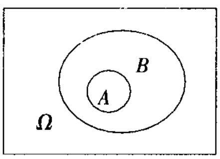


图1-1　$A\subset B$

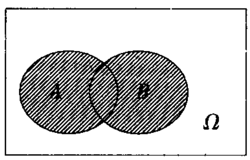


图1-2　$A\cup B$

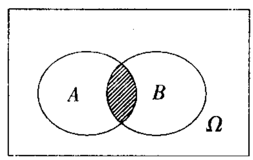


图1-3　$A\cap B$

类似地，“事件$A_1,A_2,\cdots,A_n$中至少有一个发生”称为$A_1,A_2,\cdots,A_n$的和事件，记为$\bigcup\limits_{k=1}^{n}A_k=A_1\cup A_2\cup\cdots\cup A_n$.

“可列个事件$A_1,A_2,\cdots,A_n,\cdots$中至少有一个发生”称为$A_1,A_2,\cdots,A_n,\cdots$的和事件，记作$\bigcup\limits_{n=1}^{\infty}A_n=A_1\cup A_2\cup\cdots\cup A_n\cup\cdots$.

4. “事件A与事件B同时发生”这一事件称为事件A与B的**积**或**交**，记作$A\cap B$或AB，即$A\cap B=\{\omega\mid\omega\in A$且$\omega\in B\}$，它的几何表示如图1-3.

类似地，“事件$A_1,\cdots,A_n$同时发生”称为$A_1,\cdots,A_n$的积事件，记作$\bigcap\limits_{k=1}^{n}A_k=A_1\cap A_2\cap\cdots\cap A_n$. 称$\bigcap\limits_{n=1}^{\infty}A_n=A_1\cap A_2\cap\cdots\cap A_n\cap\cdots$为可列个事件$A_1,\cdots,$

$A_n, \cdots$的积事件,即"可列个事件 $A_1, \cdots, A_n, \cdots$同时发生".

5. "事件 A 发生而 B 不发生"这一事件称为事件 A 与 B 的差,记作 $A-B$,即 $A-B=\{\omega \mid \omega \in A$ 且 $\omega \notin B\}$,如图 1-4.

6. 若事件 A 与 B 不可能同时发生,则称事件 A 与 B **互不相容或互斥**,也就是说 AB 是一个不可能事件,即 $AB=\varnothing$,如图 1-5. 对互不相容的事件 A、B,可以把和事件 $A \cup B$ 记作 $A+B$.

7. 若事件 A 与 B 满足条件 $A \cup B=\Omega$, $A \cap B=\varnothing$,则称事件 A 与 B 互为**对立事件**或互为**逆事件**. 通常将 A 的对立事件记作 $\overline{A}$,于是有 $A \cap \overline{A}=\varnothing$, $A \cup \overline{A}=\Omega$,如图 1-6 所示.

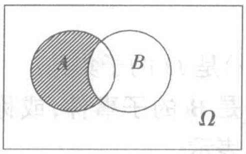


图 1-4 $A-B$

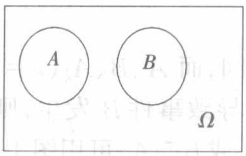


图 1-5 $AB=\varnothing$

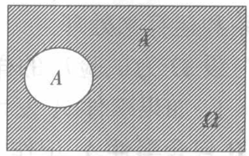


图 1-6 $\overline{A}$

由定义可知,A 与 B 互为逆事件就是指 A、B 不能同时发生,但在每次试验中必须发生其一且只能发生一个.

8. 事件的运算满足下列关系式:

(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:$(A \cup B) \cap C=(A \cap C) \cup(B \cap C)$,

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

(4) 对偶律(De. Morgan 公式):$\overline{A \cup B}=\overline{A} \cap \overline{B}$, $\overline{A \cap B}=\overline{A} \cup \overline{B}$.

例 1 甲、乙、丙三人对某目标射击,用 A、B、C 分别表示"甲击中"、"乙击中"和"丙击中",试用 A、B、C 表示下列事件:

(1) 甲、乙都击中而丙未击中;

(2) 只有甲击中;

(3) 目标被击中;

(4) 三人中最多两人击中;

(5) 三人中恰好一人击中.

解 (1) 事件"甲、乙都击中而丙未击中"表示 A、B 与 $\overline{C}$ 同时发生,即 $AB\overline{C}$.

(2) 事件"只有甲击中"表示 A 发生而 B、C 未发生,即 $A\overline{B}\,\overline{C}$.

(3) 事件“目标被击中”意味着甲、乙、丙三人至少有一人击中目标，表示为$A\cup B\cup C$.

(4) 事件“三人中最多两人击中”即“三人中至少有一人未击中”，可表示为$\overline{A}\cup\overline{B}\cup\overline{C}$.

(5) 事件“三人中恰好一人击中”即“三人中只有一人击中其余两人未击中”，可表示为$A\overline{B}\,\overline{C}+\overline{A}B\overline{C}+\overline{A}\,\overline{B}C$.

例 2　从某班学生中任取一名学生，A 表示选到的是男生，B 表示选到的是田径队员，说明下列关系式所表示的意义：

(1)$A\cap B=A$；(2)$A\cup B=A$.

解　(1) $A\cap B=A$ 等价于$B\supset A$，即若事件 A 发生，必导致事件 B 发生．所以，$A\cap B=A$ 表明该班的男生都是田径队员．

(2) $A\cup B=A$ 等价于$B\subset A$，即若事件 B 发生，必导致事件 A 发生．所以，此式表明该班的田径队员都是男生．

§1.2　概率的定义

一个事件在一次试验中可能发生，也可能不发生；不同的事件在同样的试验中发生的可能性有大有小．例如，在掷一枚均匀骰子时，事件 $A=\{$点数为 6$\}$发生的可能性比事件 $B=\{$点数为偶数$\}$发生的可能性小，因为 A 是B 的子事件．为了对随机试验有更深入的了解，人们希望对任一事件发生的可能性大小都能作出客观的描述，并用一个数值对它进行度量．

简单说来，随机事件 A 发生可能性大小的度量(数值)，称为 **A 发生的概率**，记作 $P(A)$．为研究概率的大小与性质，首先引入频率，用以描述事件发生的频繁程度．

1.2.1　频率

定义 2　如果随机事件 A 在 n 次重复试验中出现了 n_A次，则称 n_A是A 在这 n 次试验中发生的频数，称

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

为 A 在这 n 次试验中发生的**频率**．

不难证明，频率有如下性质：

(1)非负性：$0\leqslant f_n(A)\leqslant 1$；

(2)规范性：$f_n(\Omega)=1$；

(3) 有限可加性：对任意有限多个两两互不相容的事件 $A_1, A_2, \cdots, A_n$，有

$$f_n(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} f_n(A_i).$$

每 n 次试验，事件 A 的频率一般来说是不同的，也就是说频率具有随机性；但当 n 比较大时，就一般而言，能呈现某种规律性．为此，可先看下面的例子．

将一枚硬币抛 20 次，200 次，2 000次，各做 10 遍，得到表 1-1 中的试验数据(其中 n_H 表示正面出现的次数)．

表 1-1

试验序号	$N=20$		$N=200$		$N=2\,000$	
	n_H	$f_n(H)$	n_H	$f_n(H)$	n_H	$f_n(H)$
1	14	0.70	104	0.520	1 010	0.505 0
2	11	0.55	91	0.455	990	0.495 0
3	13	0.65	99	0.495	1 012	0.506 0
4	7	0.35	96	0.480	986	0.493 0
5	14	0.70	99	0.495	991	0.495 5
6	10	0.50	108	0.540	988	0.494 0
7	11	0.55	101	0.505	1 004	0.502 0
8	6	0.30	101	0.505	1 002	0.501 0
9	9	0.45	110	0.550	1 018	0.509 0
10	6	0.30	98	0.490	1 000	0.500 0

历史上，著名统计学家蒲丰(Comte de Buffon)和皮尔逊(Karl Pearson)曾进行过大量的抛掷硬币的试验，结果如表 1-2．

表 1-2

实验者	n	n_H	$f_n(H)$
蒲　丰	4 040	2 048	0.506 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

从上述数据可以看出，抛硬币次数 n 较小时，频率 $f_n(H)$在 0 与 1 之间随机波动，其幅度较大，但随着 n 增大，频率 $f_n(H)$呈现出一定的稳定性，即当 n 逐渐增大时，$f_n(H)$总是在 0.5 附近波动，而逐渐稳定于 0.5．

随机现象有其偶然的一面，也有其必然的一面．这种必然性表现为大量观察或试验中随机事件的频率的稳定性，即一个随机事件 A 发生的频率 $f_n(A)$通常在某个常数 $P(A)$附近摆动，而且当试验次数 n 很大时，频率 $f_n(A)$便会很接近常数 $P(A)$，这就是**频率稳定性**或随机现象的**统计规律性**．依照概率的统计定义，称上述 $P(A)$为**事件 A 的概率**．但我们必须指出，一般来说，n 越大 $f_n(A)$