● 高 等 学 校 教 材

概率论与数理统计

陈鸿建 赵永红 翁 洋

高等教育出版社
Higher Education Press

高等学校教材

概率论与数理统计

陈鸿建　赵永红　翁　洋

高等教育出版社

内容提要

本书是四川大学概率统计教研室在长期教材建设和试用的基础上编写而成。在内容上引入前沿知识，介绍自然指数分布族的基本理论和统计方法，并介绍了 SPSS 软件的应用。叙述流畅，推导严谨，注重方法的应用背景。习题丰富，内容新颖，涉及英语期末考试、网吧网管、机场误机人数、食堂座位安排、抽签结果等富有时代气息的问题。

全书内容包括：概率论基础知识、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、正态分布与自然指数分布族、极限定理、数理统计的基础知识、参数估计、假设检验、线性回归分析和方差分析、SPSS 软件简介等。

本书可作为高等学校理工科专业本科概率论与数理统计课程的教材，也可作为青年教师和青年科技工作者的参考书。

图书在版编目 (CIP) 数据

概率论与数理统计/陈鸿建，赵永红，翁洋. — 北京：高等教育出版社，2009.1

ISBN 978-7-04-024894-4

Ⅰ. 概⋯ Ⅱ. ① 陈⋯ ② 赵⋯ ③ 翁⋯ 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 (2008) 第 176392 号

策划编辑 张长虹 **责任编辑** 李华英 **封面设计** 于文燕
责任绘图 尹文军 **版式设计** 张 岚 **责任校对** 张 颖
责任印制 毛斯璐

出版发行	高等教育出版社	购书热线	010－58581118
社　　址	北京市西城区德外大街4号	免费咨询	800－810－0598
邮政编码	100120	网　　址	http://www.hep.edu.cn
总　　机	010－58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	北京嘉实印刷有限公司	畅想教育	http://www.widedu.com
开　　本	787×960　1/16	版　　次	2009 年 1 月第 1 版
印　　张	21.5	印　　次	2009 年 1 月第 1 次印刷
字　　数	400 000	定　　价	27.00 元

物料号　24894–00

前 言

随着近年来高校扩招, 四川大学与其他重点大学一样, 也面临着一个同样的、比以往更突出的问题, 既要培养符合一定要求的本科生, 也要培养大量高素质的硕士生、博士生。而在概率统计教学中, 同时要满足这两类学生的培养要求, 我们认为教材建设是一个重要环节。

概率统计是一门应用性较强的数学课程, 因此在教材中, 既需要强调基本概念和方法的实际应用背景, 也要尽可能严格地进行数学推导。

近二三十年来, 概率统计发展很快, 提出了许多能解决实际问题的分布、模型和方法。并且现代统计计算基本依赖于统计软件的应用。如何在概率统计教学中反映出现代知识的冰山一角, 也是我们考虑的问题。

基于以上三点想法, 我们编写了大学数学 (理工科非数学类专业) 概率论与数理统计教材。经过两年试用, 学生普遍反映较好。我们认为这本教材比以往教材更能符合四川大学这样的重点大学的实际需求。

本教材有以下特点:

一、叙述流畅自然, 结构合理, 推导严谨, 突出概率统计思想背景。

在引入随机变量的概念后, 我们立即定义分布函数。突出分布函数的作用是描述随机变量, 即描述随机试验的统计规律性。再由分布函数的两种基本类型引入离散型和连续型随机变量。分布函数在教材叙述中作用突出。许多概率论的基本问题都由分布函数展开。比如对二维随机变量 (X, Y), 有二维分布函数 $F(x, y)$。X 和 Y 又分别有一维分布函数 $F_X(x)$ 和 $F_Y(y)$。由它们之间的关系引入了边缘分布与随机变量独立性的概念。再比如连续型情形求随机变量函数 $Y = g(X)$ 的密度函数问题, 也是先求 Y 的分布函数。这样, 使得教材叙述自然流畅。

二、例题应用性强, 习题丰富, 每章附有复习分析题, 习题和例题中有许多新设计的题。

每章习题分为 A, B 两组。每组题又分为选择、填空及解答证明题。A 组题是基本题, 是为所有学生准备的, 是训练掌握基本概念和方法的。B 组题是为水平较好的学生准备的。通过 B 组题的训练, 学生能熟练掌握一般方法和概念的应用, 并掌握一些解概率统计习题的技巧。复习分析题是对较难的例题分析其解题步骤和方法, 训练学生的解题能力, 这部分是供水平较好的学生自学之用, 在

自学的基础上再做 B 组题效果会更好。可做完 B 组题的学生水平基本能达到硕士研究生入学考试水平。

三、尝试了在基础课教材中介绍现代知识。

书中介绍了近二三十年来对数理统计发展有着重要影响的自然指数分布族，与书中其他内容融合。自然指数分布族的知识在国内外同层次的教材中还尚未见到。另外还介绍了 SPSS 软件用于计算本教材涉及的统计问题。

本书是在四川大学概率统计教研室长期的教材建设经验积累上完成的，教研室对本书的构想进行了集体讨论。第一章至第八章由陈鸿建编写，第九章由翁洋编写，第十章和第十一章由赵永红编写，最后由陈鸿建统稿。马洪、刘晓石、林华珍教授及邹述超、何腊梅副教授仔细阅读了书稿并提供了宝贵意见，唐亚勇副教授对 SPSS 软件的介绍提供了帮助。赵永红为本教材配制了电子教案。如有需要请直接与高等教育出版社联系。

本书的出版得到了四川大学数学学院领导的大力支持。高等教育出版社的编辑张长虹为本书质量的提高付出了辛勤劳动，并提出了许多建议，在此一并表示感谢!

由于作者水平有限，难免有错误之处。恳请读者批评指正。

作　者

2008 年 6 月于成都

目　录

第一章 概率论基础知识

概率这个名词对大多数人来说并不陌生. 常听人们说买一注体育彩票中头等奖的概率很小; 巴西足球队与美国足球队相遇, 巴西队胜的概率很大等. 但要给概率一个明确的含义, 下一个准确的定义, 却不是几句话就能解释清楚的. 这一章, 我们就来解决这个问题, 并讨论有关概率论的基础知识.

首先, 在人类社会的生产实践和科学实验中, 我们可以观察到的客观现象形形色色. 但仔细观察, 这些客观现象可分为两类. 比如朝上掷一枚硬币, 由于地心引力的作用, 这枚硬币必然会落地. 又如一个标准大气压下, 水加热到 100 °C, 必定会沸腾. 这种一定条件下必定会出现唯一客观结果的现象叫确定性现象. 如果我们朝上掷一枚硬币, 考察落地后哪面朝上, 这时有两个可能客观结果, 即正面或反面朝上. 掷币之前我们就知道有这两个可能结果, 但不知道哪一个结果会出现, 掷币落地之后, 哪面朝上就知道了. 又比如一射手向一靶射击一次, 其成绩可能是 1 到 10 环中某一环, 也可能脱靶, 这时有 11 种可能的客观结果出现. 但射击之前不知道其成绩, 射击之后其成绩就确定了. 这种有两个或两个以上可能结果出现的客观现象叫随机性现象.

随机现象的可能结果有多个, 这是它的不确定性. 但这种不确定性中又蕴含着某种规律性. 如果我们重复抛掷一枚硬币许多次, 就会发现其正面朝上与反面朝上的次数大约各占一半. 这就是随机现象的统计规律性. 概率论与数理统计就是研究随机现象的统计规律性的一个数学分支.

由于随机现象普遍存在, 所以概率统计的应用十分广泛. 例如某车间有 200 台车床, 由于经常需要检修、测量、调换刀具、变换位置等诸多原因, 因此在生产时间, 各台车床也时常需要停车. 若每台车床有 60% 的时间在开动, 而每台车床开动时要耗电 1 kW, 那么要供给这个车间多少电力才能保证该车间正常生产呢?

显然, 若供给这个车间 200 kW 的电能则该车间能正常生产. 但因为每台车床开工率只有 60%, 平均起来同时工作的车床只有 120 台, 供给 200 kW 电力会造成浪费. 若供给 120 kW 电力又较少一些, 因为有时同时工作的车床会超过 120 台. 正确的答案是供给 141 kW 就够了. 这样, 因供电不足而影响生产的机会不到 0.1%, 即 8 h 工作中大约只有半分钟会碰到这种情况. 从而可以节约 59 kW 电力做其他用途. 这类问题的解决方法同学学了第六章之后便会得到.

§1.1　样本空间与随机事件

§1.1.1　随机试验

对随机现象的统计规律性进行研究, 要从随机试验着手.

一个科学实验, 或对一个自然现象和社会现象的观察, 我们都称为一个试验. 如果一个试验满足以下三个特点, 我们称为**随机试验**:

(1) 可在相同条件下重复进行;

(2) 试验的所有可能结果不止一个, 且试验前知道一切可能结果;

(3) 试验前不知哪一个可能结果出现, 试验后能客观确定出现的是哪一个结果.

随机试验以后简称为试验, 并用 E 表示试验.

不难验证, 抛掷一硬币观察落地后哪面朝上和记录一射手射击一次的成绩都是试验. 下面再看几个随机试验的例子.

E_1: 抛掷一硬币两次, 考察正面, 反面出现的结果.

E_2: 抛掷一硬币两次, 记录正面出现的次数.

E_3: 记录某网站一天中晚 8 点到 9 点的来访人数.

E_4: 记录某水文站一天早 7 点河流的水位.

E_5: 在单位圆内任取一点, 记录其座标.

§1.1.2　样本空间与随机事件

随机试验的所有可能结果在试验之前就知道, 而试验之后能确定试验中出现的结果是所有可能结果之一. 这样, 可以用集合与其元素描述它们.

Ω 表示一个试验的所有可能结果的集合, Ω 称为该试验的**样本空间**. 而这个试验的任何一个可能结果称为一个**样本点**, 记为 ω.

易见, 试验 E_1 到 E_5 的样本空间分别为

$$\Omega_1=\{HH,HT,TH,TT\},$$

其中 H 与 T 分别表示一次掷币中出现正面或反面.

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

注意 E_1 与 E_2 是不同的随机试验, 所以样本空间不同.

$$\Omega_3=\{0,1,2,\cdots\},$$

$$\Omega_4=\{x|a\leqslant x\leqslant b\},$$

其中 a,b 分别表示该水文站记录的历史最低, 最高水位.

$$\Omega_5=\{(x,y)|x^2+y^2\leqslant 1\}.$$

由上可见, 样本空间是由试验确定的. 它们可以是有限集, 也可以是无穷集: 可以是一维或多维的数集, 也可以是抽象的集合. 有时为了数学处理方便, 样本空间也可形式上扩大. 比如 Ω_4 也可写为 $[0,+\infty)$ 甚至 $(-\infty,+\infty)$.

知道了样本空间, 并不能满足人们实践上的需要. 比如掷一枚骰子, 考察出现的点, 则 $\Omega=\{1,2,\cdots,6\}$. 但对赌徒而言, 关心的是出现小点还是大点, 即出现 1,2,3 中某一点, 还是出现 4,5,6 中某一点. 又如 E_4 中人们关心的是水位是否超过警戒水位 C. 即人们关心样本空间的某子集中的样本点是否会在试验中出现.

一般地, 我们称样本空间的一个子集为一个**随机事件**, 简称为**事件**. 事件可用大写的 A,B,C 等表示.

我们称事件 A 在一次试验中发生, 当且仅当试验中出现的样本点 $\omega\in A$. 有几个特殊的事件需要强调. Ω 本身是 Ω 的子集, 它包含所有的样本点, 称为**必然事件**, 因为任何一次试验 Ω 都发生. 记 $\varnothing$ 为 Ω 的空子集, 它不包含任何样本点, 称为**不可能事件**, 因为任何一次试验 $\varnothing$ 都不发生. 比如 A 表示掷一次骰子中出现 8 点, 则 A 是不可能事件. 而仅含一个样本点 ω 的事件 $B=\{\omega\}$ 称为**基本事件**.

§1.1.3 事件的关系及运算

随机事件是样本空间的一个子集合, 因而可以根据集合论的知识来讨论事件间的关系与运算. 但是这些关系与运算有概率论的意义.

设试验的样本空间为 Ω, 而 $A,B,A_i(i=1,2,\cdots)$ 是 Ω 的子集, 是事件.

(1) 事件的包含与相等

若 $A\subset B$, 称事件 B 包含事件 A, 其概率意义为若事件 A 发生则事件 B 一定发生. 若 $A\subset B$ 且 $B\subset A$, 则称事件 A 与 B **相等**, 记为 $A=B$, 其概率意义为事件 A 与 B 同时发生与否.

(2) 事件的和 (并)

$A\bigcup B$ 称为事件 A 与 B 的**和 (并)** 事件, 它是由事件 A 与 B 产生的一个新事件, 表示事件 A 与 B 至少有一个发生的事件. 和事件可以推广到 $\bigcup\limits_{i=1}^{n}A_i$ 与 $\bigcup\limits_{i=1}^{\infty}A_i$, 它们分别表示有限个事件 $A_1,A_2,\cdots,A_n$ 中至少有一个发生或可数无穷个事件 $A_1,A_2,\cdots$ 中至少有一个发生的事件.

(3) 事件的积 (交)

$A\bigcap B$ 或 AB 称为事件 A 与 B 的**积 (交)** 事件, 它表示事件 A 与 B 都发生的事件. 同样, 事件的积可推广到 $\bigcap\limits_{i=1}^{n}A_i$ 与 $\bigcap\limits_{i=1}^{\infty}A_i$, 分别表示有限个或可数无穷个事件都同时发生的事件.

(4) 事件的差

事件 $A-B$ 称为事件 A 与 B 的差, 表示 A 发生而 B 不发生的事件.

(5) 互斥 (互不相容) 事件

若 $AB=\varnothing$, 即事件 A 与 B 不可能同时发生, 称事件 A 与 B 为**互斥**或**互不相容**的事件. 需要注意, 基本事件之间是两两互斥的.

(6) 互逆 (互为对立) 事件

若事件 A 与 B 有 $AB=\varnothing$, 且 $A\bigcup B=\Omega$, 则称 A 与 B 为**互逆** (**互为对立**) 事件. 因为此时, A 与 B 不可能同时发生, 但 A 与 B 必定有一个会发生, 所以称 B 为 A 的**逆** (**对立**) 事件, 记为 $\bar{A}$, 即 A 不发生. 这时有 $B=\bar{A}$ 和 $A=\bar{B}$.

易见, 对任意事件 A,B, 有 $A\bar{A}=\varnothing, A\bigcup\bar{A}=\Omega, A-B=A\bar{B}, \bar{\bar{A}}=A$.

(7) 完备事件组

若事件 $A_1,A_2,\cdots,A_n$ 两两互斥, 且 $A_1\bigcup A_2\bigcup\cdots\bigcup A_n=\Omega$, 称 n 个事件 $A_1,A_2,\cdots,A_n$ 为一个**完备事件组**, 或称为样本空间 Ω 的一个**有限划分**. 可见, 完备事件组是互为对立的事件的一个延伸.

例如, $\Omega=\{1,2,\cdots,6\}, A=\{1,2,4\}, B=\{1,3,5\}$, 则 A 与 B 互不包含; $A\bigcup B=\{1,2,3,4,5\}; AB=\{1\}; A-B=\{2,4\}; \bar{A}=\{3,5,6\}$; 六个基本事件 $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$ 两两互斥.

与集合的运算与关系一样, 直观上可用文氏图表示事件的关系与运算.

在图 1.1 中, 矩形表示样本空间 Ω, 矩形内每个点表示一个样本点. 前五个图两个圆代表事件 A 和 B, 第六个图小圆代表事件 A, 最后一个图代表由事件 $A_1,\cdots,A_6$ 构成的完备事件组.

与集合论中集合的运算一样, 事件间的运算有以下性质:

(1) **交换律**: $A\bigcup B=B\bigcup A, AB=BA$;

(2) **结合律**: $A\bigcup(B\bigcup C)=(A\bigcup B)\bigcup C$,

$A\bigcap(B\bigcap C)=(A\bigcap B)\bigcap C$;

(3) **分配律**: $A\bigcup(B\bigcap C)=(A\bigcup B)\bigcap(A\bigcup C)$,

$A\bigcap(B\bigcup C)=(A\bigcap B)\bigcup(A\bigcap C)$;

(4) **对偶律**: $\overline{A\bigcup B}=\bar{A}\bar{B}, \overline{AB}=\bar{A}\bigcup\bar{B}$.

对偶律一般形式为: $\overline{\bigcup_i A_i}=\bigcap_i \bar{A}_i, \overline{\bigcap_i A_i}=\bigcup_i \bar{A}_i$.

例 1.1 检查产品质量时, 从一批产品中任意抽取 5 件样品进行检查, 则可能的结果是: 未发现次品, 发现 1 个次品, $\cdots$, 发现 5 件次品. 设事件 A_i 表示"发现 i 件次品" $(i=0,1,\cdots,5)$, 试用 $A_i(i=0,1,\cdots,5)$ 表示:

(1) 完备事件组;

(2) 事件"发现 2 件或 3 件次品";

(3) 事件"最多发现 2 件次品";

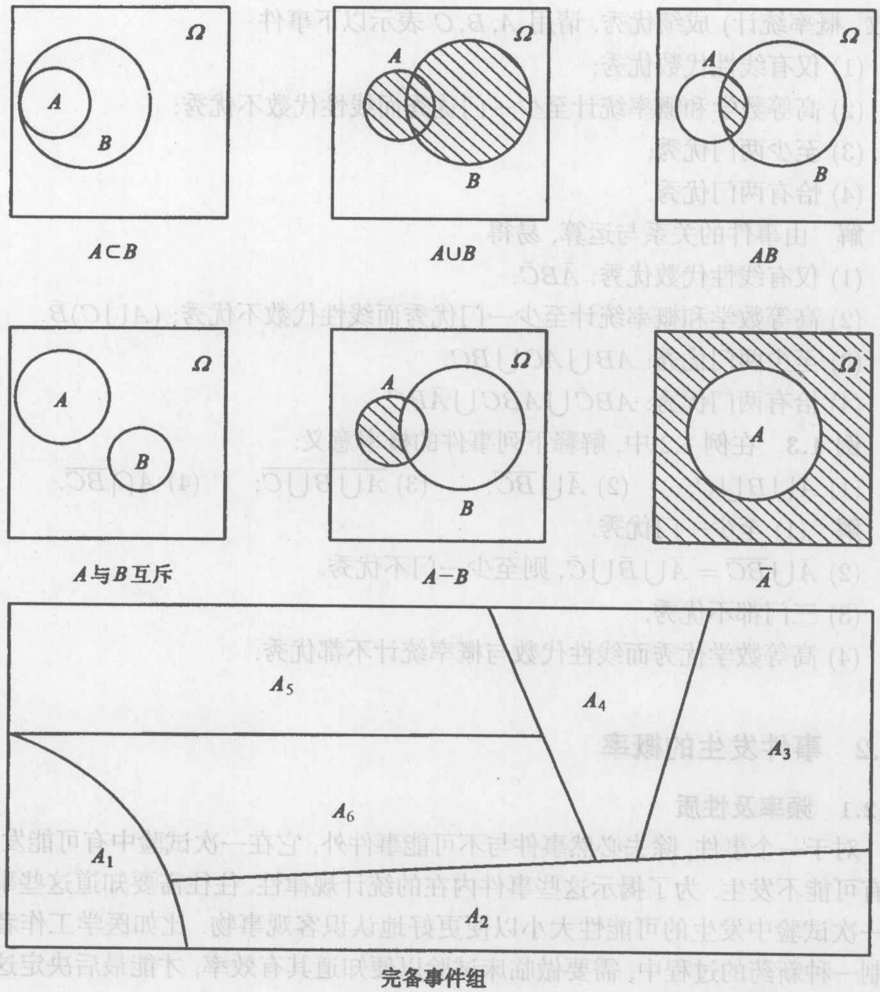

图 1.1 事件的关系与运算图

(4) 事件“至少发现一件次品”.

解 (1) 因为 $A_0, A_1, \cdots, A_5$ 两两互斥, 且其并是必然事件, 故 $A_0, A_1, \cdots, A_5$ 是一个完备事件组.

(2) 记 B 表示“发现 2 件或 3 件次品”, 则 $B = A_2 \bigcup A_3$.

(3) 记 C 表示“最多发现 2 件次品”, 则 $C = A_0 \bigcup A_1 \bigcup A_2$ 或 $C = \overline{A_3 \bigcup A_4 \bigcup A_5}$.

(4) 记 D 表示“至少发现一件次品”, 则 $D = \bar{A}_0$ 或 $D = A_1 \bigcup A_2 \bigcup \cdots \bigcup A_5$.

例 1.2 事件 A, B, C 分别表示一同学甲、乙、丙三门课程 (高等数学、线性

代数、概率统计) 成绩优秀. 请用 A,B,C 表示以下事件

(1) 仅有线性代数优秀;

(2) 高等数学和概率统计至少一门优秀而线性代数不优秀;

(3) 至少两门优秀;

(4) 恰有两门优秀.

解 由事件的关系与运算, 易得

(1) 仅有线性代数优秀: $\bar{A}B\bar{C}$.

(2) 高等数学和概率统计至少一门优秀而线性代数不优秀: $(A\bigcup C)\bar{B}$.

(3) 至少两门优秀: $AB\bigcup AC\bigcup BC$.

(4) 恰有两门优秀: $AB\bar{C}\bigcup A\bar{B}C\bigcup \bar{A}BC$.

例 1.3 在例 1.2 中, 解释下列事件的概率意义:

(1) $A\bigcup B\bigcup C$; (2) $\bar{A}\bigcup\overline{BC}$; (3) $\overline{A\bigcup B\bigcup C}$; (4) $A\bigcap\overline{BC}$.

解 (1) 至少一门优秀.

(2) $\bar{A}\bigcup\overline{BC}=\bar{A}\bigcup\bar{B}\bigcup\bar{C}$, 则至少一门不优秀.

(3) 三门都不优秀.

(4) 高等数学优秀而线性代数与概率统计不都优秀.

§1.2 事件发生的概率

§1.2.1 频率及性质

对于一个事件, 除去必然事件与不可能事件外, 它在一次试验中有可能发生, 也有可能不发生. 为了揭示这些事件内在的统计规律性, 往往需要知道这些事件在一次试验中发生的可能性大小以便更好地认识客观事物. 比如医学工作者在研制一种新药的过程中, 需要做临床试验以便知道其有效率, 才能最后决定这种药是否可以投入临床使用.

为了刻画事件在一次试验中发生的可能性, 我们首先引入频率的概念.

定义 1.1 在 n 次重复试验中, 若事件 A 发生了 k 次, 则称 k 为事件 A 发生的频数, 称 $\dfrac{k}{n}$ 为事件 A 发生的**频率**, 记为 $f_n(A)=\dfrac{k}{n}$.

由定义, 可知频率有如下性质:

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

(2) $f_n(\Omega)=1,\ f_n(\varnothing)=0$;

(3) 若 $A_1,A_2,\cdots,A_r$ 为 r 个两两互斥的事件, 则

$$f_n\left(\bigcup_{i=1}^{r}A_i\right)=\sum_{i=1}^{r}f_n(A_i).$$

性质 (1) 与 (2) 是显而易见的. 下面简单证明性质 (3).

考虑 $r=2$ 时, 因为 A_1 与 A_2 不可能同时发生, 故在 n 次重复试验中, $A_1\bigcup A_2$ 发生的频数 k 等于 A_1 发生的频数 k_1 与 A_2 发生的频数 k_2 之和. 即 $k=k_1+k_2$, 所以有

$$f_n(A_1\bigcup A_2)=\frac{k}{n}=\frac{k_1+k_2}{n}=f_n(A_1)+f_n(A_2).$$

再用归纳法, 就可以一般地证明性质 (3).

由于事件 A 在 n 次试验中发生的频率是 A 发生的频数与试验次数 n 之比, 频率大小表示了 A 发生的频繁程度. 频率越大, 事件 A 在 n 次试验中发生得越频繁, 就意味着 A 在一次试验中发生的可能性越大. 因此, 频率在一定程度上可以反映事件发生可能性的大小.

但是, 另一方面频率具有不客观性. 我们来看下面两组历史数据.

例 1.4 历史上, 许多著名的统计学家做过“抛硬币”试验, 得到如下数据 (见表 1.1):

表 1.1 抛硬币试验数据表

试验者	抛硬币次数 n	正面朝上次数 n_A	频率 $f_n(A)$
Buffon	4 040	2 048	0.506 9
Fisher	10 000	4 979	0.497 9
Pearson	12 000	6 019	0.501 6
Pearson	24 000	12 012	0.500 5

例 1.5 考虑某类种子发芽率试验. 从一大批种子中抽取 7 批种子做发芽试验, 其结果见表 1.2:

表 1.2 种子发芽率试验数据表

种子粒数	10	70	310	700	1 500	2 000	3 000
发芽粒数	9	60	282	639	1 339	1 806	2 715
发芽率	0.9	0.857	0.910	0.913	0.893	0.903	0.905

从这两组数据看, 频率具有波动性. 当 n 不同时, 频率不相同. (事实上, 即使 n 相同, 不同的试验者得到的频率也未必相同.)

再仔细观察这两组数据, 当 n 较小时, 频率波动较大; 而当 n 较大时, 频率波动越来越小, 且频率总稳定在一个客观数量附近. 掷币数据中这个稳定值是 0.5, 而种子发芽数据中这个稳定值是 0.9.

这个稳定值是一个客观数值, 它就能刻画事件 A 发生的可能性大小, 我们称为 A 发生的概率, 记为 $P(A)$. 而频率 $f_n(A)$ 有刻画 A 发生可能性客观的一面, 所以频率 $f_n(A)$ 叫做 A 的统计概率.