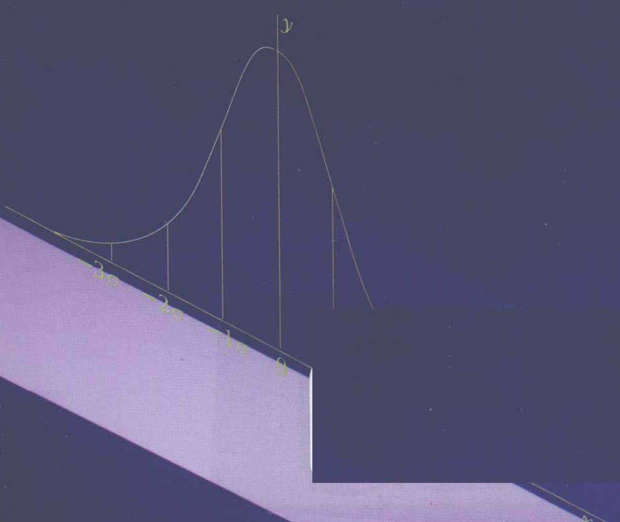
普通高等教育“十一五”规划教材

概率论与数理统计

主　编　张　薇
副主编　张　京　邵新慧　孙艳蕊



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　张　薇

副主编　张　京

邵新慧

孙艳蕊

科学出版社

北　京

内 容 简 介

本书是高等院校非数学专业的概率论与数理统计课程教材. 全书分上、下两篇. 上篇为概率论部分，内容包括随机事件与概率、随机变量及其分布、二维随机变量及其分布、随机变量的数字特征、大数定理与中心极限定理；下篇为数理统计部分，内容包括样本与抽样分布、参数估计、假设检验、线性回归分析、方差分析. 各章均配有适当、适量的分节习题和章末习题. 书末附有习题答案及 9 个附录，其中附录 9 介绍了如何利用 Excel 进行概率与统计计算.

本书可作为高等院校理工(非数学专业)、经管、农学等专业概率论与数理统计课程的教材，也可作为教学参考书和自学参考书.

图书在版编目(CIP)数据

概率论与数理统计/张薇主编. —北京：科学出版社，2010
普通高等教育“十一五”规划教材
ISBN 978-7-03-028280-4

Ⅰ. ①概… Ⅱ. ①张… Ⅲ. ①概率论-高等学校-教材②数理统计-高等学校-教材 Ⅳ. ①O21

中国版本图书馆 CIP 数据核字(2010)第 132819 号

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
铭浩彩色印装有限公司 印刷
科学出版社发行 各地新华书店经销
*
2010 年 7 月第 一 版 开本：B5(720×1000)
2010 年 7 月第一次印刷 印张：22 3/4
印数：1—5 500 字数：450 000

定价：35.00 元

(如有印装质量问题，我社负责调换)

前　言

概率论与数理统计是研究随机现象及其统计规律的一门核心数学学科，是被广泛应用的数学工具之一．目前，自然科学、经济科学和工程技术等领域的众多分支与随机理论的交融日益紧密，计算机技术的飞速发展和广泛应用又给这种交融插上了腾飞的翅膀，使概率统计更加广泛快速地渗透到科技研究的前沿．概率论与数理统计在培养学生的创新和应用能力方面占据了得天独厚的优势，它作为各类专业大学生数学基础课程的重要性勿庸置疑，相配套的教材在课程建设中的重要作用不言而喻．

根据教学需要，我们汇集东北大学理学院数学系任课教师20余年的教学实践经验，倾力编写了这本教材．教材正式出版前已在我校试用3个教学周期，几经修改与完善．

适用范围

本书既可作为高等学校理工（非数学专业）、经管、农学等专业概率论与数理统计课程的教材，也可作为教学参考书和自学参考书．

本书特点

• 易读、易懂．书中大部分章节均以常见的例子或易懂的知识引出基本概念，尽量简单论述基本理论，利于读者理解和掌握概率统计知识．

• 各章节例题和习题的选取紧密配合知识点，利于读者理解、掌握和运用所学知识．书中给出了许多实际例子，帮助读者从多个方面了解概率统计在众多领域中的应用．

• 计算机的使用．鉴于Excel软件的通用性，我们特别编写了附录9，详细介绍了如何利用Excel进行概率与统计计算．第6、7、8章的一部分例题和第9、10章的全部例题的结果数据及图像均由Excel软件给出．

• 在编排上注重知识的衔接与循序渐进．语言表达清晰明了，符号运用规范．

• 紧扣教育部本科教学大纲，与硕士研究生入学考试大纲接轨．

内容编排

全书分为上、下两篇．上篇由第1～5章组成，讲授概率论的基础知识，涉及随机事件与概率、随机变量及其分布、数字特征、大数定理与中心极限定理．下篇由第6～10章组成，讲授数理统计的基本概念、参数估计、假设检验、线性回归分析和方差分析．各章均配有适当、适量的分节习题和章末习题，书末附有习题答案和9个附录，其中附录1给出了常用的概率分布，附录2～7是常用分布数据表，附录8给

出了三个常用抽样分布的证明，附录 9 介绍了如何利用 Excel 进行概率与统计计算.

学时安排

概率论部分（前 5 章）教学约需 36 学时，概率统计部分（前 8 章）教学约需 56 学时，全程教学约需 68 学时.

编写人员

本书第 1、4、5 章由张京执笔，第 2、3 章由张薇执笔，第 6、7、8 章由邵新慧执笔，第 9、10 章由孙艳蕊执笔. 第 7、8 章末尾工具表及附录 1～7 由张薇执笔，附录 8 由邵新慧执笔，附录 9 由王洪曾执笔. 全书最后由张薇修改定稿.

致谢

本书获得了东北大学教材建设计划立项项目“概率论与数理统计教材建设”及东北大学教务处的支持，并得到科学出版社数理分社的大力支持，在此表示感谢. 在教材形成及试用过程中，任课教师提出了许多有益的建议，张国伟教授仔细阅读了大部分章节，提出了宝贵的修改意见，在此向他们致谢. 并向审稿人、科学出版社所有参与本书出版的工作人员、特别是编辑张中兴致谢.

本书不妥之处，敬请读者指正.

编 者

2010 年 6 月

目　　录

上篇　概　率　论

上篇　概　率　论

第1章 随机事件与概率

在没有外力的作用下，作匀速直线运动的物体必然持续作匀速直线运动；在标准大气压下，100℃的纯水必然沸腾. 像这类在一定条件下必然发生的现象我们称为**确定性现象**. 对于确定性现象，人们根据数学和科学的方法进行准确的计算，会得出确定的结果. 然而客观世界中还有其他现象. 例如，抛掷一枚硬币，落地时，是正面向上还是反面向上，这在抛掷前是不能确定的；下一个交易日股市指数可能上升，也可能下跌，而且升跌幅度的大小不能事先确定；记录一段时间内某电话机的呼叫次数，它可能是 0 次、1 次，也可能是 2 次，…. 这些现象的一个共同特点是：在基本条件不变的情况下，试验或观察到的结果呈现出偶然性，时而会出现这种结果，时而会出现那种结果. 像这类在一定条件下事先不能确定是否发生的现象我们称为**随机现象**. 在自然界和社会生活中存在着大量的随机现象. 如，刮风、下雨、地震、火山爆发、泥石流、山体滑坡、海啸的发生等，这些是自然现象. 如未来几年经济形势的好坏，股票债券的涨跌，彩票中奖与否，这些是社会生活现象. 在处理随机现象时，研究确定性现象的数学方法已不再完全适用，必须拓展及寻找新的数学方法，概率论与数理统计这门数学分支学科即是研究和揭示随机现象统计规律的一个重要工具.

§1.1 随机事件

1.1.1 样本空间

随着认识的深入和知识的积累，人们逐渐认识到，在随机现象偶然性的背后，隐藏着必然规律. 例如，(观察)抛掷一枚硬币，落地时可能是正面向上，也可能是反面向上，即一次的结果具有偶然性. 但是如果重复地抛掷 n 次，人们发现，随着 n 的增大，出现正面向上与出现反面向上的次数越来越接近，呈现出了规律性. 又如，(记录)一天 24 小时内某电话机的呼叫次数，前天可能是 6 次，昨天可能是 2 次，明天又可能是 10 次，…，每一天的结果都具有偶然性. 但是 1 个月、或 1 年、或 2 年…下来，人们发现，在这 1 个月、或 1 年、或 2 年…的时间内，呼叫次数往往呈现出某种规律性. 像这种"观察"、"记录"我们称之为**随机试验**，简称**试验**，记作 E. 它具有两个明显特征：试验可以在相同的条件下重复地进行，每次的结果不一定相同；试验可能出现的一切结果可事先预知，但不能事先预知每次试验会出现哪一个结果. 如上面两例中，抛掷一枚硬币只有"正面向上"和"反面向上"2 个可能结果；电话呼

叫次数必定是某个非负整数. 人们正是通过随机试验来探究随机现象及其规律的.

这里的"试验",它既可以是通常意义下的实验,也可以是对自然和社会现象的观察、记录等.

随机试验的所有可能结果组成的集合称为它的样本空间,记作 Ω. 试验的每个可能结果称为样本点,记作 ω.

例 1.1 抛掷一枚硬币,观察正面向上(用 H 表示)、反面向上(用 T 表示)出现的情况. 这是一个随机试验,其样本空间由 2 个样本点组成,即 $\Omega=\{H,T\}$.

例 1.2 将一枚硬币抛三次,观察正面向上(H)、反面向上(T)出现的情况. 这也是一个随机试验,其样本空间由 8 个样本点组成,即 $\Omega=\{HHH,HHT,HTH,HTT,THH,THT,TTH,TTT\}$.

例 1.3 记录某电话机在一小时内接收到的呼叫次数,其样本空间是非负整数的集合,$\Omega=\{0,1,2,\cdots\}$.

例 1.4 连续射击直到命中为止. 以"0"表示一次射击未中,"1"表示命中,则样本空间 $\Omega=\{l,01,001,\cdots\}$.

例 1.5 从某厂生产的相同型号的灯泡中抽取一只,测试它的寿命(即灯泡连续正常工作的小时数),其样本空间 $\Omega=[0,+\infty)$.

例 1.6 记录某地一昼夜的最低气温和最高气温. 若以 x,y 分别表示最低气温和最高气温,则样本空间 $\Omega=\{(x,y)\mid T_0\leqslant x\leqslant y\leqslant T_1\}$,这里 T_0,T_1 分别表示这一地区气温的下界与上界.

从以上 6 个例子中可以看出,样本空间 Ω 可以是数集,也可以是抽象元素的集合;可以是有限集,也可以是无穷可列集或无穷不可列集;可以是一维样本点,也可以是多维样本点的集合.

值得注意的是,由于试验目的的不同,即使是相同的随机试验,也可以得出不同的样本空间. 如例 1.2,如果关心的是出现正面向上的次数,则样本空间为 $\Omega=\{0,1,2,3\}$.

1.1.2 随机事件及其运算

在随机试验中,人们常常关心试验的某些结果是否出现. 例如,在例 1.3 中,若以一小时达到 20 次的呼叫来区分这台电话机是否繁忙,那么呼叫次数不足 20 次的试验结果为"不繁忙",它包含了样本空间中的前 20 个样本点 0,1,2,…,19. 再如,在例 1.5 中,若以灯泡寿命不足 1200 小时来区分灯泡质量的好坏,那么寿命小于 1200 小时的试验结果为"质量不好",它包含了样本空间中小于 1200 小时的样本点. 像这种由试验结果(样本点)组成的集合称为随机事件,简称事件[①]. 而其中

① 本书中未加特殊说明的事件都指随机事件.

仅由一个样本点组成的集合又称为基本事件；由所有的样本点组成的集合，即样本空间 Ω 称为必然事件，因为每次试验的结果必在 Ω 中，即 Ω 必然发生；不包含任何一个样本点的集合，即空集 $\varnothing$ 称为不可能事件，因为每次试验，$\varnothing$ 必定不发生. 上面的“不繁忙”与“质量不好”都是可能发生或可能不发生的随机事件；而“呼叫次数为25”的“繁忙”事件和“灯泡寿命为1200小时”的“质量好”事件则是基本事件. 在例1.1中，“出现正面向上或出现反面向上”的事件为必然事件，“既出现正面向上又出现反面向上”的事件为不可能事件.

常用大写字母 $A,B,C,\cdots$ 等标识事件. 上面的“不繁忙”事件可标识为 A，即 A 表示“一小时至多呼叫19次”，$A=\{$一小时至多呼叫19次$\}=\{0,1,2,\cdots,19\}$；“质量不好”事件可标识为 B，即 B 表示“灯泡寿命小于1200小时”，$B=\{$灯泡寿命小于1200小时$\}=\{x\mid 0\leqslant x<1200\}$；“呼叫次数为25”事件可标识为 C，即 C 表示“呼叫次数为25”，$C=\{$呼叫次数为25$\}=\{25\}$；“灯泡寿命为1200小时”事件可标识为 D，即 D 表示“灯泡寿命为1200小时”，$D=\{$灯泡寿命为1200小时$\}=\{1200\}$.

如果试验的结果(样本点)在事件 A 中，则称事件 A 发生. 在例1.3中，如果记录到的呼唤次数是10次，则“不繁忙”事件 A 就发生了，如果记录到的是28次，则“繁忙”事件就发生了.

由于事件可以用集合来表示，因此事件之间的关系与事件的运算可以按照集合论中集合之间的关系与集合的运算来规定. 以下给出这些关系和运算在概率论中的提法，并根据“事件发生”的含义，给出它们在概率论中的解释.

设 E 是一个随机试验，Ω 是 E 的样本空间，$A,B,A_i(i=1,2,\cdots)$ 都是 Ω 的子集合.

(1) $A\subset B$ 称为事件 B 包含事件 A. 其概率含义是：A 发生必然导致 B 发生.

$A=B$，即 $A\subset B$，且 $B\subset A$，称为事件 A 与事件 B 相等.

(2) $A\cup B=\{\omega\mid\omega\in A$ 或 $\omega\in B\}$ 称为事件 A 与事件 B 的和事件. 其概率含义是：当且仅当 A,B 至少有一个发生，事件 $A\cup B$ 才发生.

一般地，$\bigcup\limits_{i=1}^{n}A_i$ 是 n 个事件 $A_1,A_2,\cdots,A_n$ 的和事件，表示这 n 个事件至少有一个发生的事件；$\bigcup\limits_{i=1}^{+\infty}A_i$ 是可列无穷个事件 $A_1,A_2,\cdots$ 的和事件，表示这可列无穷个事件至少有一个发生的事件.

(3) $A\cap B=\{\omega\mid\omega\in A$ 且 $\omega\in B\}$ 称为事件 A 与事件 B 的积事件. 其概率含义是：当且仅当 A 与 B 同时发生，事件 $A\cap B$ 才发生. 积事件也写作 AB.

一般地，$\bigcap\limits_{i=1}^{n}A_i$ 是 n 个事件 $A_1,A_2,\cdots,A_n$ 的积事件，表示这 n 个事件同时发生的事件；$\bigcap\limits_{i=1}^{+\infty}A_i$ 是可列无穷个事件 $A_1,A_2,\cdots$ 的积事件，表示这可列无穷个事件同时发生的事件.

(4) $A-B=\{\omega|\omega\in A$ 且 $\omega\notin B\}$ 称为事件 A 与事件 B 的**差事件**. 其概率含义是:当且仅当 A 发生而 B 不发生,事件 $A-B$ 才发生.

(5) 如果 $A\cap B=\varnothing$,则称事件 A 与事件 B 是**互不相容的**或**互斥的**. 其概率含义是:A 与 B 不能同时发生.

如果一组事件中的任意两个事件都互不相容,则称该事件组为**两两互斥事件组**. 基本事件组是两两互斥事件组.

(6) 若 $A\cup B=\Omega, A\cap B=\varnothing$,则称 A 与 B 互为**对立事件**或**互逆事件**. 其概率含义是:每次试验,事件 A,B 有一个发生且仅有一个发生. A 的对立事件记为 $\overline{A}$,即 $\overline{A}=B$.

例如,某建筑工地,设 A 表示"缺少水泥",B 表示"缺少沙子",那么和事件$A\cup B$ 表示"缺少水泥或沙子",积事件 $A\cap B$ 表示"即缺少水泥,又缺少沙子",差事件 $A-B$ 表示"缺少水泥,但不缺少沙子",$\overline{A}$ 则表示"不缺少水泥",$\overline{B}$ 表示"不缺少沙子".

我们可以用图示的方式直观地表示以上事件之间的关系与事件的运算,如图 1.1~图 1.6 所示. 图中的矩形表示样本空间 Ω.

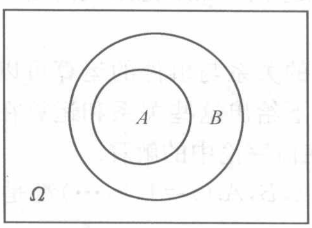


图 1.1 $A\subset B$

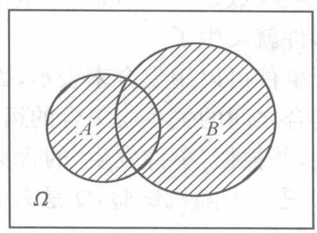


图 1.2 $A\cup B$

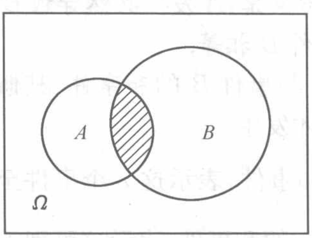


图 1.3 $A\cap B$

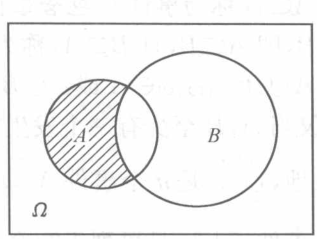


图 1.4 $A-B$

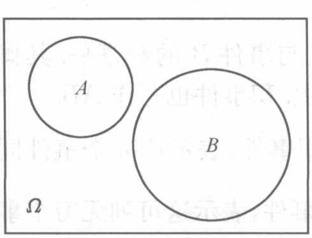


图 1.5 $A\cap B=\varnothing$

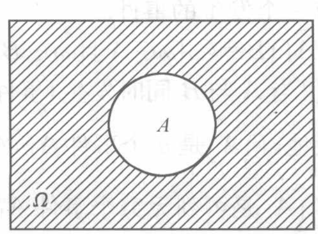


图 1.6 $\overline{A}$

与集合论中的集合运算一样,事件之间的运算也满足下述定律:

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

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

(4) 德·摩根(De Morgan)定律 $\overline{A\cup B}=\overline{A}\cap\overline{B}, \overline{A\cap B}=\overline{A}\cup\overline{B}$.

这些定律都可以推广到多个事件的情形. 例如,

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

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

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

(1)~(3)显然成立. 下面用概率论语言说明(4)中第一个等式. 如果事件$\overline{A\cup B}$发生,则 $A\cup B$ 不发生,即 A 与 B 都不发生,所以 $\overline{A},\overline{B}$ 都发生,$\overline{A}\cap\overline{B}$ 发生. 反之,如果$\overline{A}\cap\overline{B}$ 发生,则$\overline{A},\overline{B}$ 都发生,即 A,B 都不发生,所以 $A\cup B$ 不发生,即$\overline{A\cup B}$发生. 第一个等式成立. 第二个等式可类似地说明.

例 1.7 某工程队承包建造了 3 幢楼房,设 A_i 表示"第 i 幢楼房经验收合格",$i=1,2,3$. 试用 A_1,A_2,A_3 表示下列事件:

(1) 只有第 1 幢楼房经验收合格;

(2) 恰有 1 幢楼房经验收合格;

(3) 至少有 1 幢楼房经验收合格;

(4) 至多有 1 幢楼房经验收合格.

解 (1)"只有第 1 幢楼房经验收合格"包含"第 2,3 幢楼房经验收不合格"的意思,因此这个事件可以表示为 $A_1\overline{A}_2\overline{A}_3$.

(2)"恰有 1 幢楼房经验收合格"是指仅有一个合格,而另两个经验收不合格的意思,因此可以表示为

$$A_1\overline{A}_2\overline{A}_3\cup\overline{A}_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2A_3.$$

(3)"至少有 1 幢楼房经验收合格"是指 A_1,A_2,A_3 至少发生其一,因此可以表示为

$$A_1\cup A_2\cup A_3.$$

(4)"至多有 1 幢楼房经验收合格"是指(2)表示的事件与 3 幢均不合格事件的和事件,因此可以表示为

$$A_1\overline{A}_2\overline{A}_3\cup\overline{A}_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2A_3\cup\overline{A}_1\overline{A}_2\overline{A}_3.$$

习 题 1.1

1. 用集合的形式写出下列随机试验的样本空间 Ω 与随机事件 A:

(1) 抛一颗骰子,观察向上一面的点数. A 表示"出现奇数点".

(2) 对一个目标进行射击,一旦击中便停止射击,观察射击的次数. A 表示"射击不超过3次".

(3) 把单位长度的一根细棒折成三段,观察各段的长度. A 表示"三段细棒能构成一个三角形".

2. 把 $\bigcup_{i=1}^{n} A_i$ 表示成 n 个两两互不相容事件的和.

3. 在某班学生中任选一个同学,以 A 表示选到的是男同学,B 表示选到的人不喜欢唱歌,C 表示选到的人是运动员.

(1) 表述 $AB\overline{C}$ 及 $A\overline{B}C$;

(2) 什么条件下成立 $ABC=A$;

(3) 何时成立 $\overline{C}\subset B$;

(4) 何时同时成立 $A=B$ 与 $\overline{A}=C$.

4. 设 A,B,C 为三个随机事件,用 A,B,C 的运算及关系表示下列各事件:

(1) A 发生,B 与 C 不发生;

(2) A 与 B 都发生,而 C 不发生;

(3) A,B,C 中至少有一个发生;

(4) A,B,C 都发生;

(5) A,B,C 都不发生;

(6) A,B,C 中不多于一个发生;

(7) A,B,C 中不多于两个发生;

(8) A,B,C 中至少有两个发生.

§1.2 随机事件的概率

1.2.1 频率与概率

人们经过长期的实践发现,随机事件虽然在一次随机试验中可能发生,也可能不发生,呈现偶然性,且毫无规律而言,但在大量的试验中,结果则呈现出明显的规律性.下面就来研究这种规律性.

定义 1.1 在相同的条件下,进行 n 次相同的试验.在这 n 次试验中,事件 A 发生的次数 n_A 称为事件 A 发生的频数,比值 n_A/n 称为事件 A 发生的频率,记为 $f_n(A)$.

由定义,易见 $f_n(A)$ 具有如下基本性质:

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

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

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

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

例 1.8 考虑"抛硬币"试验.将一枚硬币抛 5 次、50 次、500 次,各做 10 遍,得到表 1.1 中所示数据(其中 H 表示正面向上,$f_n(H)$表示 H 发生的频率).

表 1.1

实验序号	$n=5$		$n=50$		$n=500$	
	n_H	$f_n(H)$	n_H	$f_n(H)$	n_H	$f_n(H)$
1	2	0.4	22	0.44	251	0.502
2	3	0.6	25	0.50	249	0.498
3	1	0.2	21	0.42	256	0.512
4	5	1.0	25	0.50	253	0.506
5	1	0.2	24	0.48	251	0.502
6	2	0.4	21	0.42	246	0.492
7	4	0.8	18	0.36	244	0.488
8	2	0.4	24	0.48	258	0.516
9	3	0.6	27	0.54	262	0.524
10	3	0.6	31	0.62	247	0.494

从表 1.1 中的数据可以看出,当抛掷次数为 5 或 50 时,频率 $f_n(H)$在 0 与 1 之间随机波动,且波动的幅度较大;当抛掷次数为 500 时,频率 $f_n(H)$随机波动的幅度明显变小,且 $f_n(H)$比较接近 0.5.

人们从直觉上感到,抛掷一枚均匀硬币,正面向上与反面向上出现的机会应该相等.这相当于说,在大量的试验中,$f_n(H)$应该接近于 0.5.历史上有不少人做过这个试验,部分试验结果现列入表 1.2.数据显示,随着试验次数 n 的增加,$f_n(H)$的波动呈现出稳定性,逐渐稳定于 0.5.

表 1.2

实验者	n	n_H	$f_n(H)$
德·摩根(De Morgan)	2048	1061	0.5181
蒲丰(Buffon)	4040	2048	0.5069
K. 皮尔逊(K. Pearson)	12000	6019	0.5016
K. 皮尔逊(K. Pearson)	24000	12012	0.5005

例 1.9 考察英语 26 个字母出现的频率.当观察字母的个数 n(试验的次数)较少时,频率有较大幅度的随机波动.但当 n 增大时,频率呈现出稳定性.G. Dewey 统计了 438023 个字母,得到一份英文字母频率统计表(表 1.3)[①].

① 引自 G. Dewey, Relative Frequency of English Spellings, New York: Teachers College Press, Columbia University, 1970.