JIAOCHENG

高等院校应用型人才培养规划教材——数学类
GAODENG YUANXIAO YINGYONGXING RENCAI PEIYANG GUIHUA JIAOCAI SHUXUE LEI

概率统计简明教程

GAILü TONGJI JIANMING JIAOCHENG

王三福 魏艳华 王丙参 张艺馨 编

SWJTUP 西南交通大学出版社

高等院校应用型人才培养规划教材——数学类

概率统计简明教程

王三福　魏艳华　王丙参　张艺馨　编

西南交通大学出版社

·成　都·

内容简介

本书针对高等院校非数学专业教学大纲与《全国硕士研究生入学统一考试数学考试大纲》对概率统计的要求，系统全面地介绍了概率统计的理论及其应用，读者只需具备高等数学基础即可读懂. 全书共6章，主要讲解事件与概率，随机变量（一维与多维）及其分布，随机变量的数字特征，大数定律与中心极限定理，统计量及其分布，参数估计，假设检验. 我们采用独具特色的方式处理教材，以较短的篇幅讲解必备的知识，并给出了习题的详细解答，方便教学，方便自学. 为适应信息化社会，我们在附录中给出了 MATLAB 与概率统计，随机模拟实验供读者参考，以加深对正文的理解.

本书可作为理工科、经济、管理各专业的本科生教材，也可作为相关专业的参考用书.

图书在版编目（CIP）数据

概率统计简明教程 / 王三福等编. —成都：西南交通大学出版社，2015.10
高等院校应用型人才培养规划教材. 数学类
ISBN 978-7-5643-4339-2

Ⅰ. ①概… Ⅱ. ①王… 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（2015）第 243919 号

高等院校应用型人才培养规划教材——数学类

概率统计简明教程

王三福 魏艳华 王丙参 张艺馨 编

*

责任编辑 张宝华
特邀编辑 刘文佳
封面设计 何东琳设计工作室

西南交通大学出版社出版发行
成都市金牛区交大路 146 号 邮政编码：610031 发行部电话：028-87600564
http: //press.swjtu.edu.cn
四川森林印务有限责任公司印刷

*

成品尺寸：185 mm × 260 mm 印张：12.75
字数：320 千字
2015 年 10 月第 1 版 2015 年 10 月第 1 次印刷
ISBN 978-7-5643-4339-2
定价：28.00 元

前　言

大学数学是自然科学的基本语言，是应用模式探索现实世界物质运动机理的主要手段. 目前，很多社会科学领域也引入了越来越多的数学，比如经济学、管理学等，甚至文科专业也开设文科高等数学，目的是提升大学生的必备数学素质. 大学数学是一门科学语言，然而现在已经变为高校各专业的通识教育. 对于非数学专业的大学生而言，大学数学的教育意义不仅仅是学习一种专业的工具，而是培育人的一种理性思维品格和思辨能力，是智慧的启迪，是创造力的开发，其价值远非专业技术教育所能相提并论的.

随着高校的扩招，我国高等教育快速实现了从精英教育向大众教育的转变，而教育规模的极速扩张，也给我国高等教育带来了一系列的变化、问题和挑战. 对于一般本科院校而言，学生基础薄弱，而传统大学数学教材重理论，轻实践，学生难以接受，结果导致学生厌学，后继专业课难以开展. 另外，为了培养应用型人才，课程设置开始淡化理论推导，突出实践，大学数学课的课时越来越少，比如线性代数、概率统计都压缩为 36 课时，而高等数学则变为一个学期，只开设 72 课时. 另外，随着社会的发展，计算机越来越普及，我们只有借助数学软件的强大功能才能站得更高，走得更远. 针对这种现象，我们尝试组织编写一系列大学数学公共课教材：

高等数学简明教程（72 ~ 90 课时）；

线性代数简明教程（36 ~ 54 课时）；

概率统计简明教程（36 ~ 54 课时）.

为了突出特色，本套教材将以最少的课时讲解专业必修的数学知识，并以理工科通用的数学软件 MATLAB 为基础增加计算机实现，以辅助教学. 同时也保证了理论上的完整，逻辑性强，主要适用于普通高等院校课时较少的大学数学公共课. 值得一提的是，我们在教材中插入了历史上对数学有杰出贡献的数学家简介，从他们身上既可以领略数学家坚韧不拔地追求真理的人格魅力和科学精神，也可以体会形形色色的人生，从而给自己以启迪.

概率论与数理统计作为高等学校本科数学教学中一门重要的基础课程，是进一步学习相关专业课的必备基础，非常重要. 因此，我们在编写《概率统计简明教程》分册时，博采百家之长，注重基本理论、概念、方法的叙述，坚持抽象概念形象化的原则，关注应用能力、解题能力的培养，读者只需具备高等数学基础即可读懂本书. 由于在每年考研中，数 1, 3 的概率统计内容占到了 20%以上，所以我们也参考了最新颁布的《全国硕士研究生入学统一考试数学考试大纲》的要求，力求使教材的体系、内容既符合本科生的特点，又兼顾报考研究生的学生需求. 书中很多例题直接采用了历年考研真题，我们尽量在 36 课时内，使必讲内容达到数 1 对概率统计的要求. 例题可以加深读者对理论的理解，为此我们配备了大量例题和习题，难度各异，以满足不同学生的需求. 本书采用了一些经典的例子和段落，在这里对这些材料的作者表示感谢.

全书共 6 章. 第 1~4 章为概率部分，包括随机事件与概率，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，大数定律与中心极限定理，并探讨了与概率论有关

的决策理论；第 5 ~ 6 章为统计部分，重点讲解统计量及其分布，参数估计，假设检验. 附录给出了概率统计简介，MATLAB 与概率统计，随机模拟实验，标准正态分布表，常见分布表，以供读者查阅并加深对正文的理解. 由于现在计算机软件非常普及，我们不再重点学习查表，而是利用软件直接给出结果. 为了和软件及思维习惯保持一致，书中采用下侧分位数. 另外，为方便学生自学和练习，本书也给出了比较详细的参考答案，供读者参考.

本书可作为理工类、经济、管理各专业的本科生教材，也可作为相关专业的参考用书. 概率论与数理统计，一般每学期总学时为 36 ~ 54 学时，包括习题课. 建议按第一到第六章的顺序分配学时如下：

（1）6 + 8 + 4 + 8 + 4 + 6 = 36 学时；

（2）8 + 10 + 8 + 8 + 6 + 10 = 50 学时，随机模拟实验 4 学时.

以上建议仅供参考，任课老师可根据实际需要合适安排各章学时并选择教学重点. 如果课时非常少，书中带*号的内容可供读者自学.

本书由天水师范学院数学与统计学院王三福、魏艳华、王丙参、张艺馨共同编写. 我们经常在一起讨论，切磋写法，经过反复讨论和修改后最终定稿. 在本书的编写过程中得到了学院领导的大力支持，统计教研室的同事认真审阅了书稿，提出了宝贵的修改意见，也得到了西南交通大学出版社有关各方和同仁的大力支持，特在此一并致以诚挚的谢意！

虽然我们希望编写出一本质量较高、适合当前教学实际需要的教材，但由于编者水平有限，书中难免存在错误和不妥之处，恳切希望读者批评、指正.

编　者

2015 年 2 月

目　录

1 概率论基本概念

未来是随机的，所以世界才五彩缤纷. 而概率论是研究随机现象的科学，**概率论才是生活真正的领路人**. 著名苏联数学家 Kolmogorov（柯尔莫哥洛夫）在《概率论基础》一书中提出了概率论的公理化体系. 概率论的发展史表明它是现代概率论的基础，具有里程碑式的意义. 很多文科教材编写者认为，概率公理化体系太难，所以基本不做介绍. 其实，对学过微积分的大学生而言，无论是理科还是文科，这并不难理解，反而能使他们认清概率的本质，能提升一个高度，并与微积分统一起来. 本章用大量通俗实例介绍这一公理体系，先由随机试验引出样本空间，进而给出概率的公理化定义及其基本性质，并给出概率的计算方法，最后介绍条件概率、全概率公式、贝叶斯公式和独立性.

1.1 随机事件

1.1.1 随机现象

在自然界与人类社会生活中，存在着两类截然不同的现象.

（1）**确定性现象**. 例如，每天早晨太阳必然从东方升起；在标准大气压下，水加热到 100 摄氏度必然沸腾；异性电荷相互吸引，同性电荷相互排斥，等等. 这类现象的特点是：在试验之前就能断定它有一个确定的结果，即在一定条件下，重复进行试验，其结果必然出现且唯一.

（2）**随机现象**，也称为**偶然现象**，是概率论与数理统计的研究对象. 例如，某地区的年降雨量；打靶射击时，弹着点离靶心的距离；某种型号电视机的寿命. 这些例子表明，在可控制条件相对稳定的情况下，由于影响这类现象的是大量的、时隐时现的、瞬息万变的、无法完全控制和预测的偶然因素在起作用，所以最终使得现象具有随机性. 注意，既然随机性是由大量无法完全控制的偶然因素引起的，那么随着科学的不断发展、技术手段的不断完善，人们可以将越来越多的因素控制起来，从而减少随机性的影响，不过完全消除随机性是不可能的.

在一定条件下并不总是出现相同结果的现象称为**随机现象**. 这类现象有两个特点：

① 结果不止一个；

② 哪一个结果出现，人们事先不能确定.

扩展阅读*：世界上为什么会有随机现象呢？比如掷骰子，按说不过是一种简单的物理现象，我们为什么不能预测呢？其实，骰子运动是一种“混沌”运动. 掷骰子，骰子离手的速度和转动速度等是由许许多多初始条件决定的. 骰子掉在桌面上滴溜溜滚动虽然是一种非常复杂的运动，但是，假如知道了全部初始条件，那也是可以用计算机计算出来的，因而也可预测最后出现的点数. 如果能制造出一个掷骰子的机器人，若它总是能够以完全相同的初始条件掷骰子，则总能得到完全相同的点数. 也就是说，如果知道了速度和转动等所有信息，

骰子出现什么点数就是可以预测的. 不过，绝大多数人都无法控制掷骰子的动作，而且也无法获得有关的信息，因此，对我们来说，在这些现象中出现的结果都是随机现象.

例 1.1.1 随机现象到处可见，比如：

（1）抛一枚硬币，可能正面朝上，也可能反面朝上；

（2）一天内某高速公路的交通事故次数；

（3）明天某时刻甘肃天水的温度；

（4）测量某物理量（长度、直径等）的误差.

由于随机现象的结果事先不能预知，初看起来似乎毫无规律，然而，人们发现：在相同条件下，虽然个别试验结果在某次试验或观察中可以出现也可以不出现，但在大量试验中却呈现出某种规律性，这种规律性称为**统计规律性**. 例如，在投掷一枚硬币时，既可能出现正面，也可能出现反面，预先做出确定的判断是不可能的，但是假如硬币均匀，直观上出现正面与反面的机会应该相等，即在大量的试验中出现正面的频率应接近 50%.

试验是对现象的观测（观察或测量），而**实验**是根据科学研究的目的，尽可能地排除外界的影响，突出主要因素并利用一些专门的仪器设备，人为地变革、控制或模拟研究对象，使某一些事物（或过程）发生或再现，从而去认识自然现象、自然性质和自然规律.

一个试验，如果满足：

（1）可以在相同的条件下重复进行；

（2）其结果具有多种可能性；

（3）在每次试验前，不能预言将出现哪一个结果，但知道其所有可能出现的结果，

则称这样的试验为**随机试验**.

简言之，在相同的条件下可以重复的随机现象称为**随机试验**. **随机试验是对随机现象的一次观测或试验**，通常用大写字母 E 表示，简称**试验**. 例 1.1.1 中（1）（4）是随机试验，（2）（3）由于不能重复进行（历史不可重演），故它们虽是随机现象，而不是随机试验.

1.1.2 样本空间

随机试验的一切可能基本结果组成的集合称为样本空间，用 Ω 表示；其每个元素称为**样本点**，又称为**基本结果**，用 ω 表示. 样本点是今后抽样的基本单元，认识随机现象首先要列出它的样本空间.

例 1.1.2 给出下面随机现象的样本空间.

（1）抛一枚均匀硬币，观察出现正反面情况，记 z 为正面，f 为反面，样本空间 $\Omega_1=\{z,f\}$；

（2）电话总机在单位时间内接到的呼唤次数，样本空间 $\Omega_2=\{0,1,2,\cdots\}$；

（3）测量误差的样本空间 $\Omega_3=\mathbf{R}$；

（4）掷两枚骰子，观察出现的点数，则样本空间 $\Omega_4=\{(i,j),i=1,2,\cdots,6,j=1,2,\cdots,6\}$.

对样本空间要注意以下几点*：

（1）样本空间是一个集合，元素可以是数，也可以不是数；表示方法有：列举法、描述法，如例 1.1.2 中（2）的表示方法就是列举法，用描述法表示为{呼叫次数为自然数}；

（2）样本点可以是一维的，也可以是多维的，如例 1.1.2 中（4）；可以有限个，如例 1.1.2 中（1）（4），也可以无限个，如例 1.1.2 中（2）（3）.

（3）对于一个随机试验而言，样本空间并不唯一，它由试验目的而定，但通常只有一个能提供最多信息的样本空间. 例如，在运动员投篮的试验中，若试验目的是考察命中情况，则样本空间 $\Omega=\{$中, 不中$\}$，若试验目的是考察得分情况，则样本空间 $\Omega=\{$0 分, 1 分, 2 分, 3 分$\}$.

今后在数学处理上，我们往往将据样本点个数为有限个或可列个的情况归为一类，称为**离散的样本空间**，而将样本点为不可列无限多的情况归为一类，称为**连续的样本空间**. 由于这两类样本空间有着本质差异，故分别称呼之.

扩展阅读*：众所周知，点是没有长度的，而线段是有长度的. 那么，由没有长度的点构成有长度的线段是怎么回事呢？其实，可列个点是没有长度的，而不可列个点就具有了长度，可见，由可列到不可列发生了质的飞跃. 那么，何为可列，何为不可列呢？给定集合 A,B，若存在 A 到 B 上的一一映射，则称 A 与 B 对等，记作 $A\sim B$. 如果两个集合对等，称它们具有相同的势. 若 $A\sim \mathbf{N}$，其中 $\mathbf{N}$ 为自然数集，则称 A 为可数集（可列集）. 不是可数集的无限集称为不可数集（不可列集）.

1.1.3 随机事件

在样本空间 Ω 中，具有某种性质的样本点构成的子集称为**随机事件**，简称**事件**，常用大写字母 A,B,C 等表示. 用集合论语言，随机事件就是样本空间 Ω 的子集，它包括基本事件和复合事件.

（1）由一个样本点构成的集合称为**基本事件**；

（2）由多个样本点构成的集合称为**复合事件**.

某个事件 A 发生当且仅当 A 所包含的一个样本点 ω 出现，记为 $\omega\in A$.

任何样本空间 Ω 都有两个特殊子集，即空集 $\varnothing$ 和 Ω 自身，其中，空集 $\varnothing$ 称为**不可能事件**，指每次试验一定不会发生的事件；Ω 自身称为**必然事件**，指在每次试验中都必然发生的事件. 严格来讲，必然事件与不可能事件反映了确定性现象，也可以说它们并不是随机事件，但为了研究问题的方便，常把它们作为特殊随机事件进行处理，即**退化的随机事件**.

经常会遇到这样的情况，我们感兴趣的是一个较为复杂的事件，但通过种种方法，可使之与一些较简单的事件联系起来，进而设法利用这种联系再通过简单事件去研究这些较为复杂的事件.

1）随机事件间的关系

（1）**事件的包含**：事件 A 发生必然导致事件 B 发生，则称 A 包含于 B 或 B 包含 A，记为 $A\subset B$ 或 $B\supset A$，即

$$A\subset B\Leftrightarrow\{若\omega\in A,则\omega\in B\}.$$

（2）**事件的相等**：若事件 $A\subset B$ 且 $B\subset A$，则称 A 与 B 相等，记为 $A=B$.

（3）**事件的互斥**：若事件 A 与 B 不能同时发生，则称 A 与 B **互斥**，也称为**互不相容**.

显然有：基本事件是互斥的；$\varnothing$ 与任意事件互斥.

（4）**事件的对立**：称事件 $B=\{A不发生\}$ 为 A 的**对立事件**或**逆事件**，常记为 $\overline{A}$.

作为样本空间的子集，逆事件 $\overline{A}$ 是 A 相对于样本空间 Ω 的**补集**.

对立事件一定互斥，但互斥事件不一定是对立事件. 请读者举出例子.

注：很多教材对$\subset$与$\subseteq$不加区分，认为两者等价，即不区分子集与真子集.

2）随机事件运算

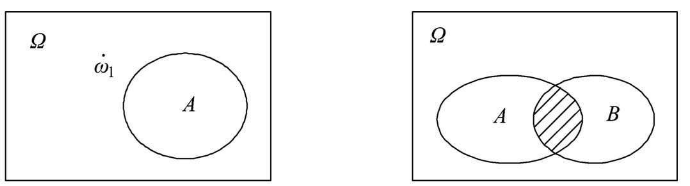
（1）**事件的并**：两个事件A,B中至少有一个发生的事件，称为事件A与B的**并**（**和**），记为$A\bigcup B$（或$A+B$），即

$$A\bigcup B=\{\omega\mid\omega\in A\text{或}\omega\in B\}.$$

（2）**事件的交**：两个事件A与B同时发生的事件，称为事件A与B的**交**（**积**），记为$A\bigcap B$（或AB），即

$$A\bigcap B=\{\omega\mid\omega\in A\text{且}\omega\in B\}.$$

为直观表示事件及其关系，在概率论中常用长方形表示样本空间Ω，用一个圆或其他几何图形表示事件A，点表示样本点ω_1，见图 1.1.1，这类图形称为**维恩**（Venn）**图**. 在考察事件关系或事件计算中，维恩图可以起到事半功倍的效果.



$\omega_1\notin A$　　　　$A\bigcap B$

图 1.1.1　事件的维恩图

（3）**事件的差**：事件A发生而事件B不发生的事件，称为事件A与事件B的**差**，记为$A-B$，或$A\setminus B$，即

$$A-B\Leftrightarrow\{\omega\in A\text{而}\omega\notin B\}.$$

（4）**事件的逆**：若事件A与B满足$A\bigcup B=\Omega$且$AB=\varnothing$，则称B为A的**逆**，记为$B=\overline{A}$，即

$$\overline{A}=\{\omega\mid\omega\notin A,\omega\in\Omega\}.$$

$A,\overline{A}$称为**互逆事件**或**对立事件**.

由前面可知，事件之间的关系与集合之间的关系建立了一定的对应法则，因而事件之间的运算法则与集合的运算法则相同.

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

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

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

（4）**德莫根（对偶）定律**：

$$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A_i}\text{（和的逆 = 逆的积），}\quad\overline{\bigcap_{i=1}^{n}A_i}=\bigcup_{i=1}^{n}\overline{A_i}\text{（积的逆 = 逆的和）}.$$

（5）**差积转换律**：$A-B=A\overline{B}$.

例 1.1.3 设 A,B,C 为任意三个事件，试用 A,B,C 的运算关系表示下列事件：

（1）三个事件中至少一个发生：$A\cup B\cup C$；

（2）没有一个事件发生：$\overline{A}\cap\overline{B}\cap\overline{C}=\overline{A\cup B\cup C}$（由对偶律）;

（3）恰有一个事件发生：$A\overline{B}\,\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\,\overline{B}C$；

（4）至少有两个事件发生：$AB\overline{C}\cup A\overline{B}C\cup\overline{A}BC\cup ABC=AB\cup BC\cup CA$.

1.2 概率的定义与计算

随机事件在一次试验中可能发生，也可能不发生，具有偶然性，但人们从实践中认识到，在相同的条件下进行大量重复试验，试验的结果具有某种内在规律性，即随机事件发生可能性的大小是可以比较的，可以用一个数字进行度量. 例如，在掷一枚均匀骰子的试验中，事件 A 表示“掷出偶数点”，B 表示“掷出 2 点”，显然事件 A 比事件 B 发生的可能性要大. **概率就是随机事件发生可能性大小的度量，介于 0 与 1 之间**. 事件发生的可能性越大，概率就越大，但事件的概率是如何定义的呢？在概率论的发展历史上曾经有过概率的古典定义、几何定义、频率定义和主观定义，这些定义各适合一类随机现象，那么如何才能给出适合一切随机现象的最一般定义呢？

一个随机试验可由一个概率空间 $(\Omega,\mathcal{F},P)$ 所描述，其具体定义由数学家 Kolmogorov 在 1933 年提出，称为 Kolmogorov **公理化体系**，它以最少几条本质特性出发刻画概率的概念，既概括了历史上几种概率定义的共同特性，又避免了各自的局限性和含混之处. 这一公理体系一经提出，就迅速获得举世的公认.

1.2.1 概率的定义

为了给随机试验提供一个数学模型，我们已经建立了样本空间 Ω，并把 Ω 的一些子集称为事件，介绍了事件的运算. 为了能自由地对有限个或可列个事件进行各种运算，并且运算的结果仍然是事件，这就需要给出事件域的定义.

定义 1.2.1 如果 $\mathcal{F}$ 是样本空间 Ω 中的某些子集的集合，它满足：

（1）$\Omega\in\mathcal{F}$；

（2）若 $A\in\mathcal{F}$，则 $\overline{A}\in\mathcal{F}$；

（3）若 $A_i\in\mathcal{F},i=1,2,\cdots$，则 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathcal{F}$，

则称 $\mathcal{F}$ 为 σ **域**，或 σ **代数**、**事件域**.

最简单的 σ 域为 $\{\Omega,\varnothing\}$，称为**平凡 σ 域**.

可见，在事件域 $\mathcal{F}$ 中至多涉及可列个事件的运算，且对差、有限交、有限并、可列交、可列并等运算是封闭的，即在 σ 域 $\mathcal{F}$ 中可以自由地进行有限个或可列个事件的各种运算，这为定义概率和全面研究随机现象奠定了基础. 今后，总是给出样本空间 Ω 后，立刻给出事件域 $\mathcal{F}$，并把 $\mathcal{F}$ 中的元素称为事件，而不在 $\mathcal{F}$ 中的 Ω 的其他子集皆不是事件，它们不在我们的研究范围之内. 如何确定事件域 $\mathcal{F}$，要根据实验类型和需要确定，在一般理论中，样本空间 Ω 和事件域 $\mathcal{F}$ 都假定事先给出，而在实际问题中，$\mathcal{F}$ 常理解为从随机事件中得到的全部

信息.

在概率论中，(Ω,F) 又称为**可测空间**，这里可测指的是 F 中的元素都具有概率，即都是可度量的.

扩展阅读*：既然有可测，那么有没有不可测的东西呢？1967 年，曼德布罗在国际权威的美国《科学》杂志上发表了一篇划时代论文，标题是《英国的海岸线有多长？统计自相似性与分数维数》，而他的答案却让我们大吃一惊：他认为，无论你做得多么认真细致，你都不可能得到准确答案，因为根本就不会有准确答案. 英国的海岸线长度是不确定的！它依赖于测量时所用的尺度.

定义 1.2.2 设 Ω 是一个样本空间，F 是由 Ω 的某些子集组成的一个事件域，如果对 $\forall A\in\mathrm{F}$，定义在 F 上的一个集合函数 $P(A)$ 与之对应，它满足：

（1）**非负性公理**：$0\leqslant P(A)\leqslant 1$；

（2）**正则性公理**：$P(\Omega)=1$；

（3）**可列可加性公理**：设 $A_i\in\mathrm{F}, i=1,2,\cdots$，且 $A_i\cap A_j=\varnothing, i\neq j$，有

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

则称 $P(A)$ 为事件 A 的概率，P 称为**概率测度**，简称为**概率**，三元总体 (Ω,F,P) 称为**概率空间**.

概率的公理化定义刻画了概率的本质，即**概率是集合函数且满足上述三条公理**. 事件域的引进使我们的模型有了更大的灵活性，在实际问题中我们根据问题的性质选择合适的 F，一般选 Ω 的一切子集为 F. 事件域可以保证随机事件经过各种运算后仍为随机事件. 注意：事件的概率是事件本身固有的属性，它是一个确定的数，不因在一次具体试验中事件是否发生而改变. 例如，掷硬币时正面出现的概率是 0.5，这是由硬币的形状对称、密度均匀等客观条件决定的. 如果一枚硬币正面是铜，反面是铝，则正面出现的概率就小于 0.5 了.

初次接触概率论的人，对"可能"这个词也许会不甚理解. 比如，我们说，坐火车安全，其实，它的真正意思是"坐火车很可能安全"，但这并不排除"坐火车会发生事故". 事实上，许多读者已经看到过火车事故的新闻了. 注意，这并不是说"坐火车要时刻提放事故，忧心忡忡"，而是让大家理解"**未来是随机的，一切皆有可能**"，其实，坐火车发生事故是小概率事件，在你的一生中是不可能发生的，因此，请你大胆地乘坐火车吧！

1.2.2 概率的性质

利用概率的公理化定义，可导出概率的一系列性质. 在概率的正则性中说明了必然事件 Ω 的概率为 1，可想而知，不可能事件 $\varnothing$ 的概率应该为 0. 切记，在数学理论体系中，只有公理、定义、假设不需要证明，其他都要证明.

性质 1.2.1 $P(\varnothing)=0$.

证明 因为 $\varnothing=\varnothing\cup\varnothing\cup\cdots$，$\varnothing\cap\varnothing=\varnothing$，由可列可加性得

$$P(\varnothing)=P(\varnothing\cup\varnothing\cup\cdots)=P(\varnothing)+P(\varnothing)+\cdots$$

再由非负性公理必有 $P(\varnothing)=0$.

在概率论中，将概率很小（小于 0.05）的事件称为**小概率事件**，也称为**实际不可能事件**.

注意：很小是一个模糊概念，没有严格的区分，要因人而定，但在许多情况下，要随试验结果的重要性，具体问题具体分析地加以确定.

小概率事件原理（又称为**实际推断原理**）：在原假设成立的条件下，小概率事件在一次试验中可以看成不可能事件，如果在一次试验中小概率事件发生了，则矛盾，即原假设不正确.

设某试验中出现事件 A 的概率为 p，不管 p 如何小，如果把试验不断独立地重复下去，那么 A 迟早必然会出现一次，从而也必然会出现任意多次，而不可能事件是指试验中总不会发生的事件. 但人们在长期的经验中坚持这样一个观点：概率很小的事件在一次试验中与不可能事件几乎是等价的，即不会发生. 如果在一次试验中小概率事件居然发生了，人们会认为该事件的前提条件发生了变化，或者认为该事件不是随机发生的，而是人为安排的，等等，这是小概率原理的一个应用. 如果我们把注意仅停留在小概率事件的极端个别现象上，那我们就是“杞人忧天”，就不敢开车，不敢吃饭，一切都不敢做了，事实上，天一定会塌下来的，但在你活着的这段时间内塌下的概率很小，杞人其实是不明白“小概率事件在一次试验中是不可能发生的”.

例 1.2.1 某接待站在一周内曾接待过 12 次来访，已知所有这 12 次来访都是在周二和周四进行的，问是否可以推断接待时间是有规定的？

解 假设接待站的接待时间没有规定，而各个来访者在一周的任一天去接待站是等可能的，那么，12 次来访都是在周二和周四进行的概率为

$$\frac{2^{12}}{7^{12}}=0.000\,000\,3\,.$$

由于小概率事件在一次试验中是不可能发生的，因此有理由怀疑假设的正确性，从而推断接待时间是有规定的.

小概率事件原理是概率论的精髓，是统计学发展、存在的基础，它使得人们在面对大量数据而需要做出分析与判断时，能够依据具体情况的推理来做出决策，从而使统计推断具备严格的数学理论依据.

性质 1.2.2（**有限可加性**） 设 $A_i \in \mathrm{F}, i=1,2,\cdots n$，且 $A_i \cap A_j=\varnothing, i\neq j$，则

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)\,.$$

证明 令 $A_{n+1}=A_{n+2}=\cdots=\varnothing$，由 $P(\varnothing)=0$ 可得

$$P\left(\bigcup_{i=1}^{n} A_i\right)=P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)=\sum_{i=1}^{n} P(A_i)\,.$$

推论 （1）对任意事件 A，有 $P(A)=1-P(\overline{A})$；

（2）对任意两个事件 A,B，有 $P(A-B)=P(A)-P(AB)$，也称为**减法公式**.

（3）若 $A\supset B$，则 $P(A-B)=P(A)-P(B)$ 且 $P(A)\geqslant P(B)$，也称为概率的**单调性**；

很容易举例说明，若 $P(A)\geqslant P(B)$，无法推出 $A\supset B$.

此推论不仅在计算事件的概率时非常有用，而且在今后一些定理的证明或公式的推导过

程中也非常有用.

性质 1.2.3（**加法公式**） 对任意两事件 A,B，有

$$P(A\cup B)=P(A)+P(B)-P(AB).$$

证明 因为 $A\cup B=A\cup(B-A)$，且 A 与 $B-A$ 互不相容，又由有限可加性得

$$P(A\cup B)=P(A\cup(B-A))=P(A)+P(B-A)=P(A)+P(B)-P(AB).$$

加法公式还能推广到多个事件的情况. 设 A_1,A_2,A_3 为任意三个事件，则有

$$P(A_1\cup A_2\cup A_3)=\sum_{i=1}^{3}P(A_i)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3).$$

利用样本点在等式两端计算次数相等可直观证明这个公式.

一般地，对于任意 n 个事件 $A_1,\cdots,A_n$，用数学归纳法可得

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n),$$

称为**容斥原理***，也称为**多去少补原理**.

例 1.2.2 设某地有甲乙两种报纸，该地成年人中 20%的人读甲种报纸，16%的人读乙种报纸，8%的人两种报纸都读，问成年人中有百分之几的人至少读一种报纸？

解 设 $A=$ “读甲种报纸”，$B=$ “读乙种报纸”，则

$$P(A\cup B)=P(A)+P(B)-P(AB)=0.2+0.16-0.08=0.28,$$

即成年人中有 28%的人至少读一种报纸.

1.2.3 概率的直接计算

事件的概率通常是未知的，但公理化定义并没有告诉人们如何去计算概率. 历史上在公理化定义出现之前，概率的统计定义、古典定义、几何定义和主观定义都在一定的场合下有了计算概率的方法，所以有了公理化定义后，它们都可以作为概率的计算方法.

1）计算概率的统计方法

在相同的条件下，重复进行了 n 次试验，若事件 A 发生了 n_A 次，则称比值 $\dfrac{n_A}{n}$ 为事件 A 在 n 次试验中出现的**频率**，记为 $f_n(A)$，n_A 为事件 A 的**频数**.

定义 1.2.3 在相同的条件下，独立重复地做 n 次试验，当试验次数 n 很大时，如果某事件 A 发生的频率 $f_n(A)$ 稳定地在 $[0,1]$ 上的某一数值 p 附近摆动，而且一般来说随着试验次数的增多，这种摆动的幅度会越来越小，则称数值 p 为事件 A 发生的概率，记为 $P(A)=p$.

概率的统计定义既肯定了任一事件的概率是存在的，又给出了一种概率的近似计算方法，但不足之处是要进行大量的重复试验. 而事实上很多随机现象不能进行大量重复试验，特别是一些经济现象无法重复，有些现象即使能重复，也难以保证试验条件是一样的. 值得注意的是，概率的统计定义以试验为基础，但这并不等于说概率取决于试验. 事实上，事件发生的概率乃是事件本身的一种属性，先于试验而存在. 例如，抛硬币，我们首先相信硬币质量均匀，那么在抛之前就已知道出现正面或反面的机会均等，所以从概率的计算途径看概率的描述性定义是

先验的，而概率的统计定义是后验的，显然两种定义并非等价. 用“频率”估计“概率”，与用“尺子”度量“长度”、用“天平”度量物质的“质量”，是完全类似的. 可以形象地说，频率是测定事件概率的“尺子”，而测定的“精度”是可通过增大试验次数来保障的.

概率客观存在的一个很重要的证据是事件出现的频率呈现稳定性，即在大量的重复试验中，频率常常稳定于某个常数，称为频率的**稳定性**，即随着 n 的增加，频率越来越可能接近概率. 我们还容易看到，若随机事件 A 出现的可能性越大，一般来讲其频率 $f_n(A)$ 也越大. 由于事件 A 发生的可能性大小与其频率大小有如此密切的关系，加之频率又有稳定性，故可通过频率来定义概率，这就是概率的统计定义.

例 1.2.3 一个职业赌徒想要一个灌过铅的骰子，使得掷一点的概率恰好是八分之一而不是六分之一. 他雇佣一位技工制造骰子，几天过后，技工拿着一个骰子告知他出现一点的概率为八分之一，于是他付了酬金，然而骰子并没有掷过，他会相信技工的话吗？

解 实际中，当概率不易求出时，人们常通过做大量试验，用事件出现的频率去估计概率，只要试验次数 n 足够大，估计精度完全可以满足人们的需求.

职业赌徒可掷一个骰子 800 次，如果一点出现的次数在 100 次左右，则骰子满足要求；如果一点出现的次数在 133 次左右，则此骰子基本上是正常的骰子，不满足要求. 至于“左右”到底多大，不同的人可取不同的值，保守、悲观的人可取小点，乐观的人可取大点，这涉及假设检验与决策问题，我们会在后面详细讲解.

人生思考：由于特制骰子出现 1 点的概率可能不是六分之一，所以在你赌博之前，一定要保证道具是正常的.

例 1.2.4 圆周率 $\pi = 3.141\,592\,6\cdots\cdots$ 是一个无限不循环小数. 中国数学家祖冲之第一次计算到小数点后 7 位，这个纪录保持了一千多年！以后，不断有人把它计算得更精确. 1873 年，英国学者沈克士公布了小数点后有 707 位的 π，但几十年后，费林先生对它产生了怀疑，他统计了 π 的 608 位小数，结果如表 1.2.1.

表 1.2.1

数字	0	1	2	3	4	5	6	7	8	9
出现次数	60	62	67	68	64	56	62	44	58	67

你能说出它怀疑的理由么？因为 π 是一个无限不循环小数，所以，理论上每个数字出现的频率都应接近 0.1，但 7 出现的频数过小，这就是费林怀疑的原因.

例 1.2.5 从某鱼池中取出 100 条鱼，做上记号后再放入鱼池. 现从该鱼池中任意取出 40 条鱼，发现其中两条有记号，问池中大约有多少条鱼？

解 设池中有 n 条鱼，则从池中捉到一条有记号的鱼的概率为 $\dfrac{100}{n}$，它近似等于捉到有记号的鱼的频率 $\dfrac{2}{40}$，即

$$\frac{100}{n} \approx \frac{2}{40},$$

解得 $n \approx 2\,000$. 所以池中大约有 2 000 条鱼.

2）计算概率的古典方法

引入计算概率的数学模型是在概率论发展过程中最早出现的研究对象，它简单、直观，不需要做大量重复试验，而是在经验、事实的基础上，对被考察事件的可能性进行逻辑分析后得出该事件的概率.

如果一个随机试验满足：

（1）**有限性**：样本空间中只有有限个样本点；

（2）**等可能性**：样本点是等可能发生的，

则称之为**古典型试验**，简称**古典概型**.

“等可能性”是一种假设，在实际应用中常需要根据实际情况去判断是否可以认为各基本事件或样本点是等可能的. 在许多场合，由对称性和均衡性就可以认为基本事件是等可能的，并在此基础上计算事件的概率，如掷一枚均匀的硬币，经过分析后就可以认为出现正面和反面的概率各为 0.5.

定义 1.2.4　设古典型随机试验的样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$，若事件 A 中含有 $k\ (k\leqslant n)$ 个样本点，则称 $\dfrac{k}{n}$ 为事件 A 发生的**概率**，记为

$$P(A)=\frac{k}{n}=\frac{A\text{中含有的样本点数}}{\text{总样本点数}}.$$

显然，古典概率也满足非负性、规范性、可加性.

扩展阅读*：（排列与组合）在确定概率的古典方法中经常要用到排列组合公式，但要注意区别有序与无序、重复与不重复. 排列和组合都是计算“从 n 个元素中任取 r 个元素”的取法总数公式，其主要区别在于：如果不讲究取出元素间的次序，则用组合公式，否则用排列公式，而所谓讲究元素间的次序，可以从实际问题中得以辨别. 例如两个人握手是不讲次序的，而两个人排队是讲次序的.

众所周知，当遇到复杂事情时，常常先将事情分类，再对每类分步，进而将复杂事情分解为多个简单事情，然后一一攻破，这就是加法原理和乘法原理.

（1）**加法原理**：设完成一件事有 m 种方式，第一种方式有 n_1 种方法，第二种方式有 n_2 种方法，……，第 m 种方式有 n_m 种方法，则完成这件事共有 $\sum\limits_{i=1}^{m}n_i$ 种方法.

譬如，某人要从甲地到乙地去，可以乘火车，也可以乘轮船，火车有两班，轮船有三班，那么甲地到乙地共有 $2+3=5$ 个班次供旅客选择.

（2）**乘法原理**：设完成一件事有 m 个步骤才能完成，第一步有 n_1 种方法，第二步有 n_2 种方法，……，第 m 步有 n_m 种方法，则完成这件事总共有 $\prod\limits_{i=1}^{m}n_i$ 种方法.

譬如，若一个男人有三顶帽子和两件背心，则他可以有 $3\times2=6$ 种打扮方式.

加法原理和乘法原理是两个很重要的计数原理，它们不但可以直接解决不少具体问题，同时也是推导下面常用排列组合公式的基础.

（1）**排列**：从 n 个不同元素中任取 $r\ (r\leqslant n)$ 个元素排成一列，考虑元素先后出现的次序，称此为一个排列，此种排列的总数记为 P_n^r 或 A_n^r. 由乘法原理可得