HZ BOOKS
华章教育

CENGAGE Learning

统计学精品译丛

（原书第3版）

数理统计与数据分析

Mathematical Statistics and Data Analysis

(Third Edition)

（美）John A. Rice 著
加州大学伯克利分校

田金方 译　谢邦昌 审校

机械工业出版社
China Machine Press

统计学精品译丛

（原书第3版）

数理统计与数据分析

Mathematical Statistics and Data Analysis

(Third Edition)

（美） John A. Rice 著
加州大学伯克利分校

田金方 译 谢邦昌 审校

机械工业出版社
China Machine Press

译 者 序

《Mathematical Statistics and Data Analysis》是美国加州大学名誉教授 John A. Rice 所著的一本优秀的概率论与数理统计教材，1988 年由 Thomson Brooks/Cole 出版，并于 1994 年再版，2003 年机械工业出版社购买了该书在中国的影印版权，发行了影印本，2007 年本书的第 3 版问世. 书中直观而深刻的统计思想，简明而翔实的数据分析实例，新颖而丰富的图形工具和计算机技术使其别具风格，开创了概率论与数理统计教程著述方式的先河，引领了数理统计发展的方向，深受广大读者喜爱和专家学者的好评，至今，已被美国、英国、加拿大和中国的许多大学选为概率论与数理统计的教材或参考书.

John A. Rice 教授（1944—）在加州大学伯克利分校获得博士学位，并一直任教于该校统计系，现为统计学名誉教授，美国数理统计学会成员，发表过多篇理论和应用统计学论文，其研究兴趣集中于海量和需要高强度计算的随机数据的分析方法，例如时间序列. 他的近期研究工作主要集中在两个天文项目上：探测太阳系外围地区（柯伊伯带）的物体和探测伽马射线脉冲星.

译者于 2003 年看到本书后，深为其内容和特色所吸引. 自 2004 年春季至今，译者在为面向研究生和本科生所开设的概率论与数理统计和统计学等多门课程中连续使用这本书. 同时，在面向财经类专业研究生开设的统计学课程的讲授中，也系统介绍了本书的基本理论和方法，利用 Excel、R 和 SAS 等统计软件包实现了教程中的数据分析实例和习题，众多学生受益匪浅. 积多年使用该教材的经验以及各类不同层次本科生及研究生对该教材的反映，我们深感这不仅是一本不可多得的概率论与数理统计教材，也是一本与经济学、管理学、医学、天文学、生物学、工学、社会学等其他学科紧密结合，展示统计学应用的优秀教科书或参考书. 随着第 3 版的问世，其内容更加丰富和完善，涵盖了目前前沿的统计分析方法，时间不仅没有使其过时，相反随着岁月的流逝，得到越来越多同行的关注. 如果我们希望找到能够借以站立的巨人肩膀，那么这本著作将是一个很好的选择.

根据本人粗浅的理解，简要概述本书的特色和贡献如下：

- 内容丰富，几乎涵盖了所有经典和前沿的概率论与数理统计理论和方法.
- 讲述材料的方式以数据分析为主，注重统计的实务和应用.
- 借助于经管、生物医学、金融、社会等领域的实际问题，增强读者对理论的理解和方法的使用.
- 强调图形工具和计算机技术，反映了计算机在统计学中扮演的越来越重要的角色；将自助法与传统的推论性过程结合起来，增加了蒙特卡罗方法.
- 叙述过程化繁为简. 本书既避免从理论到理论，又防止理论与实际脱节，而是理论构建在寓意深刻的背景内容下，对其逐步补充和加强，并与通俗的分析方法结合在一起. 这种方法不是抽象或美学的思考，同时，也没有回避学生应该知道的数学内容，适合于统计学的实践要求.

- 为使概念更加清晰，书中提供了大量的示例，而且还有丰富的习题，以增强读者的计算能力.

本书适合作为统计学、数学、其他理工科专业以及社会科学和经济学专业高年级本科生和低年级研究生的教材，同时也可供相关领域技术人员参考. 译者向广大读者推荐这本书，旨在希望它不仅成为读者学习概率论与数理统计学科的“捷径”，而且也能成为迈向其他相关学科前沿领域的“阶梯”.

在翻译过程中，我努力做到“信、达、雅”，但由于水平有限，译稿难免存在不当之处，请博雅之士不吝赐教，在此预先表示感谢，并于今后重印校正.

本书是在机械工业出版社王春华编辑的热心促动下翻译完成的，对其认真负责、精益求精的工作表示感谢. 此外还要感谢翻译过程中提供宝贵意见的同事和同学们，他们帮助我不断提升本书的译文水平. 感谢我的家人和朋友，感谢他们的理解和支持.

田金方

2011 年 3 月 7 日于山东经济学院

E-mail: tianjinfang2009@hotmail.com

前　　言

读者对象

本书适合于统计学、数学、自然科学和工程专业的低年级和高年级本科生，或一年级研究生，以及具有一定统计学基础的社会科学和经济学专业的学生阅读. 读者必须修读了包含泰勒级数和多元微积分在内的一年微积分课程，以及初级的线性代数课程.

本书的目标

这本书反映了我对第一门统计学课程的认识，而这对很多学生来说可能是最后的统计课程. 这样的课程应该包括数理统计的一些经典内容（如似然法），以及描述统计学和数据分析的一些内容，特别是图形显示、试验设计和复杂的实际应用. 它还应该体现出计算机在统计学中所起的不可或缺的作用. 这些主题适当地交织在一起，可以将现代统计学的本质展示给学生. 分别讲授两个主题的课程 —— 一个是理论，一个是数据分析，对我来讲似乎有点造作. 此外，很多学生仅学习一门统计学课程，而没有时间学习两门或两门以上这方面的课程.

数据分析与统计实践

为了将上述主题融合在一起，我一直在努力地撰写一本能够紧密结合统计实践的教科书. 只有分析实际数据，才能使我们明白形式理论和通俗数据分析方法所扮演的角色. 我围绕着各种问题组织了这本书，这些问题都需要使用统计方法来解决，此外书中包含很多实际例子，借此引入和介绍理论内容. 这样安排的优点是理论构建在寓意深刻的背景内容下，对其逐步补充和加强，与通俗的分析方法结合在一起. 我认为，这种方法是适合于统计学的，其历史发展主要是由实践需要来促进的，而不是抽象或美学的思考. 同时，我也没有回避学生应该知道的数学内容.

第 3 版

本书第 1 版于 1988 年问世，第 2 版于 1994 年出版. 尽管本书基本的目的和结构没有改变，但是新的版本反映了统计学科的发展，尤其是计算方面的革新.

这一版最显著的变动是对贝叶斯推断的处理. 我将最后一章的材料做了迁移，分散于之前的各章中，这是由于很多老师很难讲授到这一章. 现在贝叶斯推断首先出现在第 3 章的条件分布中. 然后，在第 8 章与频率学派方法同步讲解，那里的贝叶斯方法可以非常自然地解决最大似然估计量. 第 9 章假设检验的引言部分现在以贝叶斯公式作为开端，然后再转向奈曼–皮尔逊范式. 这样做的一个好处是似然比的至关重要性更突出. 在应用中，我强调无信息先验，说明频率学派和贝叶斯学派得出的定性结论具有相似性.

概率论章节新增了基因组学和金融统计的例子. 这些材料除了与相应的主题相关外，还可以

很自然地强化基本概念. 例如，连接函数 (copulas) 强调了边际分布和联合分布之间的关系. 其他变动包括第 10 章探索性数据分析中散点图和相关系数的介绍，以及第 14 章中利用局部线性最小二乘进行非参数平滑的简介. 本版新增了将近 100 道习题，主要集中在第 7 ~14 章，同时还包括几个新的数据集，有些数据集完全可以用于计算机实验室上机操作. 此外, 还修改了前面版本中解释含糊不清的一些段落.

概要

当然，我们可以从目录中找到完整的大纲，这里，我仅仅强调几点，并指出教师讲授课程时需要取舍的章节内容.

前 6 章包含概率论的内容，特别是与统计学密切相关的内容. 第 1 章以非测度论的观点介绍概率论的基本内容，以及初等组合方法. 在这一章和其他概率章节中，我尽可能地利用现实世界的例子，而不是使用球与盒子的抽样模型.

第 2 章介绍了随机变量的概念. 我选择将离散型和连续型随机变量放在一起讨论，而不是把连续情形推迟到以后再进行介绍. 本章介绍了几个常见分布. 这样安排的好处是它能为后面的章节提供一些讨论和介绍的内容.

第 3 章继续讨论随机变量，但是转向联合分布. 教师可以跳过雅可比行列式，这不会有损课程的连续性，因为它们很少在本书的其余部分出现. 如果教师乐意之后做些回溯工作，可以在讲解时跳过 3.7 节极值和顺序统计量的内容.

期望、方差、协方差、条件期望和矩生成函数共同构成第 4 章. 教师可以跳过条件期望和预测，尤其是没有计划讲解稍后的充分统计量时. 这一章之后的部分介绍了 δ 方法 (误差传播方法), 这个方法多次出现在统计学的章节中.

第 5 章在非常严格的假设条件下证明了大数定律和中心极限定理.

第 6 章汇编了与正态分布有关的常用分布，以及利用通常的正态随机样本计算所得统计量的抽样分布. 我没有在此浪费过多的时间，但确实介绍了统计学章节所必需的知识点, 学生很有必要学习这些分布.

第 7 章是有关抽样调查的内容，以非常规但比较自然的方式导入统计学的研究议题. 很多学生在学习抽样调查内容时感到比较模糊，而恰恰在抽样调查中很自然地提出了一系列比较特殊的具体统计问题. 从历史上看，抽样调查涉及了很多重要的统计概念，并可以将其用作传播介质引入在后面的章节中深入介绍的概念和技术，例如：

- 作为随机变量的估计量的思想，具有与之相关联的抽样分布.
- 偏倚、标准误差和均方误差的概念.
- 置信区间和中心极限定理的应用.
- 通过研究分层估计量揭示试验设计的概念以及相对效率的概念.
- 期望、方差和协方差的计算.

抽样调查不受欢迎的原因之一是其计算十分令人讨厌. 然而，这种讨厌也有其长处，学生可以在这样的计算中得到锻炼. 教师可以灵活地掌握介绍本章概念的深度. 比率估计和分层部分是可选的，初次讲授时完全可以跳过，或稍后再讲这些概念，这并不影响课程的连续性.

第 8 章介绍参数估计，它是由拟合数据的概率律问题引起的，其中介绍了矩方法、最大似然方法和贝叶斯推断方法，同时还介绍了效率的概念，证明了克拉默–拉奥不等式. 8.8 节介绍了充分性的概念及其一些衍生问题. 可以跳过克拉默–拉奥下界和充分性的内容. 在我看来，充分性的重要性通常被过度强调了. 负二项分布的内容也可以跳过.

第 9 章介绍了假设检验及其拟合优度检验的应用，这配合第 8 章的内容.（这个内容还会在第 11 章深入讨论.）这里还简要展示了图方法. 如果课时有限，教师可以跳过本章最后的 9.6 节（泊松散布度检验）、9.7 节（悬挂根图）和 9.9 节（正态性检验）.

第 10 章介绍了几种描述性方法，其中的很多技术都会在后面的章节中出现. 本章强调了图方法的重要性，并介绍了稳健性的概念. 将描述性方法放在本书的后面似乎有点怪异，这样做是因为描述性方法通常有其随机性的一面，三章之后再介绍之可以使学生有足够的基础知识去研究各种汇总统计量的统计行为（例如，中位数的置信区间）. 我在讲授课程时，会较早地介绍这部分内容. 例如，在抽样调查实验中，我让学生制作抽取样本的箱形图和直方图. 教师可以跳过生存函数和危险函数.

第 11 章介绍了两样本问题的经典分析方法和非参数方法. 假设检验的概念第一次出现在第 9 章，在此做了更深一步的介绍. 本章的末尾讨论了试验设计并解释了观测研究的一些内容.

前面 11 章是初级课程的核心，涵盖了估计和假设检验的构造理论、图和描述性方法以及试验设计的内容.

教师可以自由地选择第 12 章到第 14 章的内容. 特别地，没有必要按照书中给定的顺序讲解这些章节.

第 12 章利用方差分析和非参数技术讨论了单因子和二因子试验设计问题. 多重比较问题第一次出现在第 11 章末，在此进行了深入讨论.

第 13 章简单讨论了分类数据分析，介绍了齐性和独立性的似然比检验，并叙述了麦克尼马尔检验. 最后，通过前瞻性和回顾性研究的讨论引入了优势比的估计问题.

第 14 章讨论了线性最小二乘. 首先介绍了简单线性回归，接着利用线性代数讨论了更一般的情形. 我选择运用矩阵代数，但尽可能地将其维持在简单和具体层面上，没有超过初级一学期(每学年分为四学期制度中的一学期) 课程所讲授的内容. 特别地，我没有介绍一般线性模型的几何分析内容，也没有试图将回归和方差分析统一起来. 在这一整章中，理论结果伴随着更多基于残差分析的定性数据分析步骤. 在本章末，我通过局部线性最小二乘介绍了非参数回归.

计算机使用和习题解答

计算是现代统计不可或缺的一部分. 它是数据分析的本质，可以帮助我们理清基本概念. 我的学生使用开源软件包 R，将其安装在自己的计算机上就可以使用. 也可以使用其他的软件包，但在这本书中，我没有讨论其他的软件程序. 原书配套的 CD 内容可从华章网站 (www.hzbook.com) 下载，其中包括书中涉及的数据.

这本书包含大量的习题，从例行的基本概念强化题到具有一定难度的分析题. 我认为习题解答，特别是非常规的习题，是非常重要的.

致谢

我要感谢很多人，他们直接和间接地促成了第 1 版面世. Richard Olshen、Yosi Rinnot、Donald Ylvisaker、Len Haff 和 David Lane 在教学中使用了早期版本，他们提出很多有益的意见. 他们和我自己课堂中的学生提供了很多建设性的意见. 助教，尤其是 Joan Staniswalis、Roger Johnson、Terri Bittner 和 Peter Kim，解答了很多习题，发现其中的很多错误. 很多审稿人给出了有益的建议：Rollin Brant，多伦多大学；George Casella，康奈尔大学；Howard B. Christensen，杨百翰大学；David Fairley，俄亥俄州立大学；Peter Guttorp，华盛顿大学；Hari Iyer，科罗拉多州立大学；Douglas G. Kelly，北卡罗来纳大学；Thomas Leonard，威斯康星大学；Albert S. Paulson，伦斯勒理工学院；Charles Peters，休斯敦大学；Andrew Rukhin，马萨诸塞大学安默斯特校区；Robert Schaefer，迈阿密大学；Ruth Williams，加州大学圣地亚哥分校. Richard Royall 和 W. G. Cumberland 热心地提供了第 7 章抽样调查所使用的数据集. 我在休假时有幸在国家标准局度过了愉快的一年，那里的统计学家让我留意到书中其他几个数据集. 我深深地感激编辑 John Kimmel，他的耐心、毅力和信念促成这本书的出版.

使用过本书第 1 版的很多学生和教员给出了坦诚的评论，这极大地影响了第 2 版的修订. 我要特别感谢 Ian Abramson、Edward Bedrick、Jon Frank、Richard Gill、Roger Johnson、Torgny Lindvall、Michael Martin、Deb Nolan、Roger Pinkham、Yosi Rinott、Philip Stark 和 Bin Yu. 我要向无意间遗漏的同仁表示道歉. 最后，我要感谢 Alex Kugushev 在进行修订时所提供的鼓励和支持，感谢 Terri Bittner 在校正和解答新的习题时所做的细致工作.

很多人促成了第 3 版的问世. 我想感谢如下这些审稿专家：Marten Wegkamp，耶鲁大学；Aparna Huzurbazar，新墨西哥大学；Laura Bernhofen，克拉克大学；Joe Glaz，康涅狄格大学；Michael Minnotte，犹他州立大学. 我深深地感激很多读者，他们慷慨地花费大量时间指出书中的错误，并提出了很多改善结构安排之类的良好建议. 特别地，Roger Pinkham 发送了很多有益的电子邮件信息，Nick Cox 指出了大量的语法错误. Alice Hsiaw 详细评述了第 7~14 章. 我还想感谢 Ani Adhikari、Paulo Berata、Patrick Brewer、Sang-Hoon Cho Gier Eide、John Einmahl、David Freedman、Roger Johnson、Paul van der Laan、Patrick Lee、Yi Lin、Jim Linnemann、Rasaan Moshesh、Eugene Schuster、Dylan Small、Luis Tenorio、Richard De Veaux 和 Ping Zhang. Bob Stine 贡献了金融数据; Diane Cook 提供了意大利橄榄油的数据; Jim Albert 提供了篮球数据集，很漂亮地解释了回归向均值的问题; Rainer Sachs 提供了可爱的染色质分离数据. 我要感谢编辑 Carolyn Crockett 坚强的毅力和耐心，使这一版修订的愿望得以实现，还要感谢这个充满活力且高效的工作团队. 我要向无意间遗漏其姓名的其他人表示道歉.

John A. Rice

目 录

第1章　概　　率

1.1　引言

尽管概率 (几率或随机性) 的起源思想十分久远，但其数学形式的严格公理化定义却是在近代才出现. 概率论的很多基本思想起源于博弈问题. 21 世纪以来，概率的数学理论应用在很多现象上，一些代表性的例子如下：

- 概率论已经被用在遗传学上，用来构建基因突变模型，计算自然变率，同时在生物信息学上也扮演着重要的角色.
- 气体分子运动理论有很重要的概率成分.
- 在设计和分析计算机操作系统时，系统中的各式队列长度可以用随机现象来刻画.
- 已经有很成熟的理论将电子设备和通信系统中的噪声用随机过程来处理.
- 大气湍流中的许多模型使用概率论的概念.
- 在运筹学中，经常建立货物存货需求的随机化模型.
- 保险公司使用的精算科学高度依赖于概率论工具.
- 概率论可以用来研究复杂系统，提高它们的可靠度，例如现代商用或军用飞机的设计.
- 概率论是金融学的奠基石.

类似于这样的例子可以列举很多.

本书介绍概率论和数理统计的基本内容. 第一部分将概率论作为随机现象的数学模型来阐述，第二部分讲述统计学的有关内容，从本质上讲，它主要关注于数据的分析步骤，尤其是那些具有随机特征的数据. 为了理解统计理论，读者必须有一个很好的概率论背景.

1.2　样本空间

概率论研究那些发生结果具有随机性的现象. 一般地，这样的现象称为试验，所有可能的试验结果全体称为相应于该试验的**样本空间**(sample space)，记为 Ω, 其元素记为 ω. 举例如下：

例 1.2.1　一个人开车去上班，他要穿过三个交通信号灯，在每个信号灯处，他要么停下 (记为 s)，要么通过 (记为 c). 所有可能结果全体形成的样本空间是：

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}$$

例如，其中 csc 表示这个人在第一个信号灯处通过，然后在第二个信号灯处停下，最后在第三个信号灯处通过. ■

例 1.2.2　主计算机打印队列中的工作数目可以看成是随机的，这里我们取样本空间为：

$$\Omega = \{0, 1, 2, 3, \cdots\}$$

也就是所有非负整数组成的集合. 在实际中，打印队列长度可能有一个上限 N，因此，这样的样本空间定义为：

$$\Omega = \{0, 1, 2, 3, \cdots, N\}$$ ■

例 1.2.3 地震是不规则运动的反应，有时也可以随机化. 例如，某个特定地区，就震级超过给定阈值的地震而言，连续两次之间的时间长度可以看做一个试验. 这里的 Ω 是所有非负实数的集合：

$$\Omega = \{t | t \geqslant 0\}$$ ■

我们经常感兴趣于 Ω 的特定子集，用概率的语言称之为**事件**(event). 在例 1.2.1 中，上班者在第一个信号灯处停下是一个事件，即 Ω 的子集：

$$A = \{sss, ssc, scc, scs\}$$

(经常用斜体大写字母表示事件或者子集.) 在例 1.2.2 中，打印队列中的工作数目小于 5 个这一事件表示为：

$$A = \{0, 1, 2, 3, 4\}$$

代数的集合论可以直接用在概率论中. 两个事件 A 和 B 的**并**(union) 表示事件 A 或者事件 B 至少有一个发生的事件 C：$C = A \cup B$. 例如，如果事件 A 表示上班者在第一个信号灯处停下 (见之前表述)，事件 B 表示上班者在第三个信号灯处停下，

$$B = \{sss, scs, ccs, css\}$$

那么事件 C 表示这个人在第一个信号灯处停下或者在第三个信号灯处停下，由 A 或者 B 或者都在两者中的元素组成：

$$C = \{sss, ssc, scc, scs, ccs, css\}$$

两事件的**交**(intersection)，$C = A \cap B$，是事件 A 和 B 都发生的事件. 如果事件 A 和 B 如之前所述，那么事件 C 表示上班者在第一个信号灯处停下，同时又在第三个信号灯处停下，因此它由既在 A 中又在 B 中的元素构成：

$$C = \{sss, scs\}$$

事件的**补**(complement)，A^c，是事件 A 不发生的事件，因此由样本空间中不属于 A 的元素构成. 上班者在第一个信号灯处停下的补事件表示他在第一个信号灯处通过：

$$A^c = \{ccc, ccs, css, csc\}$$

读者可能从上述集合论的表述中回想起一个比较神秘的集合：空集，通常记为 $\varnothing$. **空集**(empty set) 是不包含任何元素的集合，是不含任何试验结果的事件. 例如，事件 A 表示上班者在第一个信号灯处停下，事件 C 表示他连续通过三个信号灯，$C = \{ccc\}$，那么事件 A 和 C 没有共同的试验结果，我们可以将此表述为：

$$A \cap C = \varnothing$$

在这种情形下，我们称事件 A 和 C 是**不相交的**(disjoint).

文氏图常用来可视化集合运算，如图 1.1 所示.

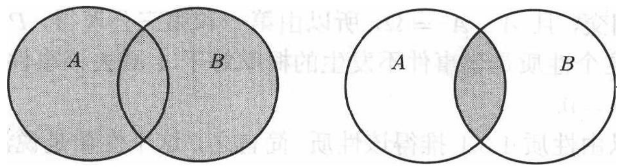

图 1.1　$A\cup B$ 和 $A\cap B$ 的文氏图

集合论的一些运算律如下：

交换律：

$$A\cup B=B\cup A$$

$$A\cap B=B\cap A$$

结合律：

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

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

分配律：

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

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

三者之中，分配律是最不直观的，读者可以利用文氏图理解它的直观解释.

1.3　概率测度

样本空间 Ω 上的**概率测度**(probability measure) 是定义在 Ω 子集上的实函数，且满足如下公理：

1. $P(\Omega)=1$.

2. 如果 $A\subset\Omega$，那么 $P(A)\geqslant 0$.

3. 如果 A_1 与 A_2 是不相交的，那么

$$P(A_1\cup A_2)=P(A_1)+P(A_2)$$

更一般地，如果 $A_1,A_2,\cdots,A_n,\cdots$ 是相互不交的，那么

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

前两个公理显然是合理的. 因为 Ω 包含所有可能的试验结果，所以 $P(\Omega)=1$. 第二个公理仅仅说明概率是非负的. 第三个公理陈述了以下事实：如果 A 与 B 是不相交的，也就是说，它们没有共同的试验结果发生，那么 $P(A\cup B)=P(A)+P(B)$，并且这个性质可以推广至无穷. 例如，打印队列中有一个或者三个工作的概率等于队列中有一个工作的概率加上有三个工作的概率.

下面概率测度的性质都是公理的结论.

性质 1.3.1　$P(A^c)=1-P(A)$.

因为 A 和 A^c 不相交，且 $A \cup A^c = \Omega$，所以由第一和第三公理得，$P(A) + P(A^c) = 1$，由此导出该性质. 简言之，这个性质是说事件不发生的概率等于 1 减去该事件发生的概率.

性质 1.3.2 $P(\varnothing) = 0$.

因为 $\varnothing = \Omega^c$，所以由性质 1.3.1 推得该性质. 简言之，这个性质是说没有任何试验结果发生的概率等于 0.

性质 1.3.3 *如果* $A \subset B$，*那么* $P(A) \leqslant P(B)$.

这个性质是说如果事件 A 发生时，事件 B 必然发生，那么 $P(A) \leqslant P(B)$. 例如，如果天下雨时一定多云，那么下雨的概率就小于或者等于多云的概率. 正式地，性质 1.3.3 证明如下：B 可以表示成两个不相交事件的并，即

$$B = A \cup (B \cap A^c)$$

那么，利用第三公理

$$P(B) = P(A) + P(B \cap A^c)$$

从而

$$P(A) = P(B) - P(B \cap A^c) \leqslant P(B)$$

性质 1.3.4 *加法定律*：$P(A \cup B) = P(A) + P(B) - P(A \cap B)$. ■

这个性质很容易从图 1.2 中的文氏图看出. 如果将 $P(A)$ 和 $P(B)$ 加在一起，则 $P(A \cap B)$ 被计算了两次. 为了证明之，我们将 $A \cup B$ 分解成三个相互不交子集的并，如图 1.2 所示：

$$C = A \cap B^c$$

$$D = A \cap B$$

$$E = A^c \cap B$$

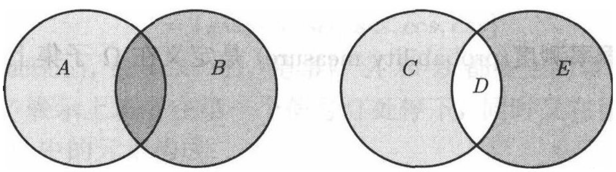

图 1.2 解释加法定律的文氏图

利用第三公理，我们有，

$$P(A \cup B) = P(C) + P(D) + P(E)$$

同样，$A = C \cup D$，C 和 D 也是互不相交，所以 $P(A) = P(C) + P(D)$. 类似地 $P(B) = P(D) + P(E)$. 将这些结果放在一起，我们可以看到

$$\begin{aligned} P(A) + P(B) &= P(C) + P(E) + 2P(D) \\ &= P(A \cup B) + P(D) \end{aligned}$$

或者

$$P(A \cup B) = P(A) + P(B) - P(D)$$

例 1.3.1 设想抛掷一个质地均匀的硬币 2 次. 令事件 A 表示第一次掷得正面，事件 B 表示第二次掷得正面. 样本空间为：

$$\Omega = \{hh, ht, th, tt\}$$

我们假设 Ω 中的每一个基本元素是等可能发生的，概率为 $\frac{1}{4}$. 事件 $C = A \cup B$ 表示第一次或者第二次掷得正面. 很显然，$P(C) \neq P(A) + P(B) = 1$. 确切地，因为 $A \cap B$ 是第一次和第二次都掷得正面，所以

$$P(C) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.5 - 0.25 = 0.75 \qquad \blacksquare$$

例 1.3.2　《洛杉矶时报》(1987.8.24) 的一篇文章讨论了艾滋病毒感染的统计风险问题：

> 对病毒携带者性伙伴的一些研究表明，没有保护措施的单一性行为感染未携带病毒者的风险惊人的低 —— 可能从百分之一到千分之一不等. 平均来看，该风险为五百分之一. 如果与病毒携带者进行 100 次性行为，感染的几率增加到五分之一.
>
> 从统计上来看，与同一个感染者进行 500 次性行为或者与 5 个感染者进行 100 次性行为将导致百分之百的感染率 (这是从统计的角度来解释，现实不一定如此).

按照这个推理，与同一个感染者进行 1000 次性行为将导致感染率等于 2(这是从统计的角度来解释，现实不一定如此). 为了找出这个推理的谬误之处，我们考虑两次性交的情况. 令事件 A_1 表示第一次性交时感染了艾滋病毒，事件 A_2 表示第二次性交时感染了艾滋病毒. 那么，两次性交感染上艾滋病毒的事件为 $B = A_1 \cup A_2$，且

$$P(B) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leqslant P(A_1) + P(A_2) = \frac{2}{500} \qquad \blacksquare$$

1.4　概率计算：计数方法

有限样本空间上的概率计算特别简单. 设 $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$，并且 $P(\omega_i) = p_i$. 为了计算事件 A 发生的概率，我们只需将 A 包含的基本事件 ω_i 发生的概率相加即可.

例 1.4.1　设想抛掷一质地均匀硬币两次，记录下出现正面和反面的次序. 样本空间为

$$\Omega = \{hh, ht, th, tt\}$$

如上一节例 1.3.1，我们假设 Ω 中每一试验结果发生的概率是 0.25. 令事件 A 表示至少有一个正面出现. 那么 $A = \{hh, ht, th\}$，$P(A) = 0.75$. $\blacksquare$

这是一般情形的简单例子. Ω 中的元素具有等概率性，因此，如果 Ω 有 N 个元素，那么每一个元素发生的概率都是 $1/N$. 如果事件 A 通过 n 个互斥途径中的任一种方式发生，那么 $P(A) = n/N$，或者

$$P(A) = \frac{\text{导致}A\text{发生的方式个数}}{\text{所有试验结果个数}}$$

注意，这个公式仅当所有试验结果是等可能发生时才成立. 在例 1.4.1 中，如果仅仅记录两次试验出现正面的次数，那么 Ω 就变为 $\{0, 1, 2\}$. 这些试验结果的发生不是等可能的，$P(A)$ 就不再是 $\frac{2}{3}$. $\blacksquare$

例 1.4.2 (辛普森悖论)　一个黑色的盒子里面装有 5 个红球和 6 个绿球，一个白色的盒子里面装有 3 个红球和 4 个绿球. 允许你先选择一个盒子，然后从选中的盒子里面随机选一个球. 如果选中红球，你就会得到一个奖品. 那么应该从哪个盒子里面选取呢？如果从黑色的盒子里面选

取，选中红球的概率是 $\frac{5}{11}=0.455$(你能抽到红球的方式个数除以所有的试验结果个数). 如果从白色的盒子里面选取，选中红球的概率是 $\frac{3}{7}=0.429$，因此，应该从黑色的盒子里面抽取.

现在考虑另外一个游戏，又有一个黑色盒子里面装有 6 个红球和 3 个绿球，一个白色盒子里面装有 9 个红球和 5 个绿球. 如果从这个黑色盒子中选取，得红球的概率是 $\frac{6}{9}=0.667$，而如果从这个白色盒子中选取，概率是 $\frac{9}{14}=0.643$. 因此，你还是应该从黑色盒子中抽取.

在最终的游戏中，将第二个黑色盒子中的球加入第一个黑色盒子中，第二个白色盒子中的球加入第一个白色盒子中. 再一次选择抽球的盒子. 你应该选择哪一个呢？直觉告诉我们应该选择黑色盒子，但让我们计算一下各自事件发生的概率. 黑色盒子现在有 11 个红球和 9 个绿球，因此从中抽中红球的概率是 $\frac{11}{20}=0.55$. 白色盒子现在有 12 个红球和 9 个绿球，因此从中抽中红球的概率是 $\frac{12}{21}=0.571$. 所以，你应该选择白色盒子. 这种反直觉的结论就是*辛普森悖论*的例子. 实际生活中发生的例子见 11.4.7 节. 更多有趣的例子参见 Garder(1976). ■

前面的例子很容易计算试验结果发生的数目，并计算概率. 为了计算更加复杂形式的概率，我们必须介绍计数试验结果的系统方法，这就是下面两节的任务.

1.4.1 乘法原理

下面陈述了非常有用的乘法原理.

乘法原理 *如果一个试验有 m 个结果，另一个试验有 n 个结果，那么这两个试验共有 mn 个可能的结果.*

证明 记第一次试验结果为 $a_1,\cdots,a_m$，第二次试验结果为 $b_1,\cdots,b_n$. 两次试验结果就是有序数对 (a_i,b_j). 这些数对可以用 $m\times n$ 矩形数组表示，数对 (a_i,b_j) 构成该矩阵的第 i 行和第 j 列元素. 在这个数组中，共有 mn 个元素. ■

例 1.4.1.1 扑克牌有 13 种面值和 4 种花色. 这样的面值和花色组合共有 $4\times13=52$ 种. ■

例 1.4.1.2 一个班级有 12 个男生和 18 个女生. 老师选取 1 个男生和 1 个女生作为学生会的代表. 他有 $12\times18=216$ 种可能的选择方式. ■

扩展的乘法原理 *如果有 p 个试验，第一次有 n_1 种可能的试验结果，第二次有 n_2 种，$\cdots$，第 p 次有 n_p 种可能的试验结果，那么 p 次试验共有 $n_1\times n_2\times\cdots\times n_p$ 种可能的试验结果.*

证明 该原理可以通过归纳法由乘法原理证得. 我们看到 $p=2$ 时该原理成立. 假设 $p=q$ 时结论亦成立 —— 也就是，前 q 次试验有 $n_1\times n_2\times\cdots\times n_q$ 种可能的结果. 为了完成归纳法的证明，我们必须说明原理在 $p=q+1$ 时也成立. 视前 q 次试验为具有 $n_1\times n_2\times\cdots\times n_q$ 种试验结果的单一试验，利用乘法原理，我们可以推得 $q+1$ 次试验有 $(n_1\times n_2\times\cdots\times n_q)\times n_{q+1}$ 种试验结果. ■

例 1.4.1.3 一个 8 位数二进制单词是一个 8 位数字序列，每一位数字或者为 0 或者为 1. 有多少个不同的 8 位数单词？

第一位有两种选择，第二位有两种，等等. 那么共有