“十三五”国家重点出版物出版规划项目

名校名家基础学科系列

Textbooks of Base Disciplines from Top Universities and Experts

概率论导论

（翻译版）

Introduction to Probability

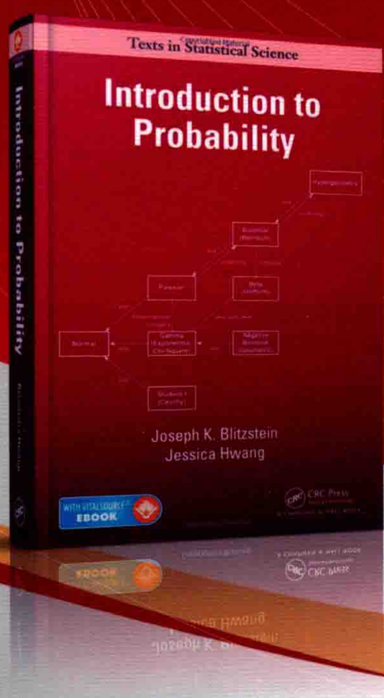
[美] 约瑟夫·K.布利茨斯坦（Joseph K.Blitzstein） 著

哈佛大学

杰西卡·黄（Jessica Hwang）

斯坦福大学

张景肖 译



CRC Press
Taylor & Francis Group

机械工业出版社
CHINA MACHINE PRESS

“十三五” 国家重点出版物出版规划项目

名校名家基础学科系列
Textbooks of Base Disciplines from Top Universities and Experts

概率论导论

Introduction to Probability

（翻译版）

[美] 约瑟夫 · K. 布利茨斯坦（Joseph K. Blitzstein） 著
哈佛大学
杰西卡 · 黄（Jessica Hwang）
斯坦福大学
张景肖 译

机械工业出版社

本书源自著名的哈佛大学统计学讲座的专用讲义，介绍了帮助读者理解统计方法、随机性和不确定性的基本语言和工具，并列举了多种多样的应用实例，内容涉及偶然性、概率悖论、谷歌的网页排名算法（PageRank）及马尔可夫链蒙特卡罗方法（MCMC）等。本书还探讨了概率论在诸如基因学、医学、计算机科学和信息科学等领域的应用。

全书共13章，分别介绍了概率与计数、条件概率、随机变量及其分布、期望、连续型随机变量、矩、联合分布、变换、条件期望、不等式与极限定理、马尔可夫链、马尔可夫链蒙特卡罗方法、泊松过程等内容。本书用容易理解的方式来呈现内容，用实例来揭示统计学中基本分布之间的联系，并通过条件化将复杂的问题归结为易于解决的若干小问题。书中还包含了很多直观的解释、图示和实践问题。每一章的结尾部分都给出了如何利用R软件来完成相关模拟和计算的方法。

本书可作为高等院校本科生概率论课程的教材，也可作为相关科研人员的参考书。

Introduction to Probability/Joseph K. Blitzstein, Jessica Hwang/ISBN: 9781466575578

图书在版编目（CIP）数据

概率论导论：翻译版/（美）约瑟夫·K. 布利茨斯坦（Joseph K. Blitzstein），杰西卡·黄（Jessica Hwang）著；张景肖译. —北京：机械工业出版社，2018.12

书名原文：Introduction to Probability

“十三五”国家重点出版物出版规划项目 名校名家基础学科系列

ISBN 978-7-111-61054-0

Ⅰ.①概… Ⅱ.①约… ②杰… ③张… Ⅲ.①概率论-高等学校-教材 Ⅳ.①O211

中国版本图书馆 CIP 数据核字（2018）第 227484 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）
策划编辑：汤 嘉 责任编辑：汤 嘉 陈崇昱 任正一
责任校对：王 延 封面设计：张 静
责任印制：张 博
三河市宏达印刷有限公司印刷
2019 年 1 月第 1 版第 1 次印刷
184mm×260mm · 28.75 印张 · 1 插页 · 708 千字

标准书号：ISBN 978-7-111-61054-0
定价：118.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务	网络服务
服务咨询热线：010-88379833	机 工 官 网：www.cmpbook.com
读者购书热线：010-88379649	机 工 官 博：weibo.com/cmp1952
	教育服务网：www.cmpedu.com
封面无防伪标均为盗版	金 书 网：www.golden-book.com

译者序

概率论是研究随机现象数量规律的数学分支，同其他数学学科一样，概率论既有逻辑推理的严谨性，又有实际应用的广泛性。

哈佛大学教授约瑟夫·K. 布利茨斯坦和斯坦福大学教授杰西卡·黄所著的《概率论导论》是一本通俗易懂，但又不失专业严谨性的概率论启蒙书。本书内容十分丰富，作者对概率论这一学科的基础知识进行了系统、全面的描述和总结。本书最初是著名的哈佛统计学讲座的专用讲义，为理解统计学、随机性和不确定性提供了重要的基础。书中，作者采用多种案例，对要阐述的定理与命题进行了生动而形象的解释，同时结合概率论的知识探讨了其在诸如基因学、医药学、计算机科学与信息科学等领域的应用。在全书的叙述中，作者利用图表和现实世界中的实际例子，以生动形象且鼓励发散性思考的方式将书的内容传递给读者，尤其是在第 3 章随机变量及其分布中，作者采用事例来阐明统计学中基本分布之间的联系，并通过一定的方法将复杂问题转化为易于理解的多个小问题。

R 是一款可以免费使用的数据分析软件，被广泛地应用于概率论和统计学领域，本书在每一章最后都专设一个小节用于讲解如何利用 R 软件来实现相关模拟和计算的方法。

本书共 13 章。第 1 章介绍概率的定义和计数方法，并简单介绍 R 统计软件的获取与使用方法。第 2 章到第 6 章介绍概率论的主要研究对象以及基本概念，包括：条件概率、随机变量及其分布、常见的连续型和离散型随机变量、用于描述分布的期望、方差以及一般化的矩、矩母函数等。为使读者更容易理解这些概念，书中给出大量实际案例，并给出详细的分析思路和步骤，以便于读者直观地理解概率论中一些看似违反直觉的结论。第 7 章介绍了用于刻画多个随机变量间关系的联合分布的概念。第 8 章介绍随机变量的变换，由此可通过将已知分布进行变换得到其他分布。第 9 章介绍条件期望的概念，它可作为简化期望求解过程的有力工具。第 10 章介绍了复杂概率和期望的近似求法，包括：蒙特卡罗模拟法、不等式约束法以及极限理论近似法。第 11 章介绍用于证明大数定律应用于非独立随机变量原理的马尔可夫链，其本质是一种具有马尔可夫性质的离散事件随机过程。第 12 章介绍用于模拟复杂分布的 MCMC 算法及其应用实例。第 13 章介绍泊松过程的定义和性质，泊松过程实质是发生在时间或空间上的随机事件的计数过程。

本书一直是美国哈佛大学、斯坦福大学概率论课程使用的教材，作者在教学方面也有其独特的想法。译者认为，本书可作为我国国内所有高等院校开设概率论这一课程的专业（如数学专业、统计学专业、工科专业等）的概率论入门基础教程，也可作为有关概率论方面的参考书。

在翻译本书的过程中得到了王伟华、张海涛、马文博、李涛、郭凯迪、严云贵、常宇、唐浩开等同学的很多帮助，在此深表感谢。另外感谢本书的编辑一直以来的大力协助。

限于译者的水平，译稿中存在的问题和不当之处，敬请读者批评指正。

译者

于中国人民大学

前　言

本书通过现代的观点来介绍概率论，为理解统计方法、随机性和不确定性奠定了基础。书中包含了丰富的应用案例，从基本的抛硬币问题和偶然性的研究到谷歌的网页排名算法（PageRank）以及马尔可夫链蒙特卡罗方法等。由于概率论是一门经常被认为是反直觉的学科，所以书中给出了很多直观的解释、图示和案例以证明这个观点的偏颇。每章的结尾部分还结合 R 软件来更详细地探讨这一章的思想（R 软件是一种用于统计计算和模拟的免费软件）。

本书取材于哈佛大学的视频公开课 Stat110（从 2006 年起，这门课程每年均由 Joseph 讲授），课程视频可在 stat110. net 网站上免费获取。其他附加的补充材料，诸如 R 代码及标记了Ⓢ的练习题的解答也均可在该网站获取。

掌握微积分是学习本书的一个前提，而对统计学的基础则没有要求。数学方面的主要挑战不在于完成微积分求解，而在于能够在抽象的概念和具体的例子之间转换。

本书的主要特征概括如下：

1. 案例。书中的定义、定理和证明都是通过案例来呈现的，这种呈现既保留了数学的精确性，又概括性地对现实世界的一些现象做出了解释。通过那些让概率分布广泛地在统计建模中使用的案例来探究概率分布。我们尽可能避免冗长乏味的推导，取而代之的是致力于给出解释和直觉判断来说明为什么那些主要结论是正确的。事实证明，通过深刻理解来替代死记硬背的方法可以提高学生对内容的长期记忆力。

2. 图。由于图本身就能表达很多内容，所以我们通过图来补充定义，使得那些主要概念与让人印象深刻的图相联系。在很多领域中，一名初学者与一名专家的差距常被描述如下：初学者总是努力去记住大量看似不相关的事实和公式，而专家则会领悟出一个统一的结构，在这个结构中仅通过少量的原理和思想就可将那些事实连贯地联系在一起。为了帮助学生领会概率论的结构，我们特别强调了思想间的联系（同时从语言上和视觉效果上加以巩固），并在大多数章节的结尾部分给出了概念与分布的循环、扩展图。

3. 概念和策略的双重教学。我们的目的在于让学生在读本书时不仅能够学习概率论的概念，同时还能够掌握广泛适用于概率论之外的一系列解决问题的策略。对于书中的例子，相同的问题经常会给出多种不同的解答方法。我们对求解的每一步都进行了解释，同时也对如何思考并选择采用的方法进行了评述。

我们对诸如对称性和模式识别这样的重要策略进行了明确的标记和命名，并且通过给出了标有☣（生物危害标识）的内容来消除常见误解。

4. 实践问题。本书包含大约 600 道不同难度的练习题，目的是为了让学生加强对内容的理解，同时强化他们解决问题的能力。这些练习题中有些是策略实践问题，根据主题进行了分组以促进对特定主题的实践，而有些则是混合型实践问题，在这些实践问题中需要综合一些前面章节中的内容。大约 250 道练习题已有详细的在线解答以供线下实践及自学使用。

5. 模拟、蒙特卡罗方法和 R 软件。很多概率问题都因计算太难而不能精确求解，并且在任何情况下，对所给答案进行核查都是很重要的。我们介绍了通过模拟来研究概率论的方法，并证明了借助简短的几行 R 代码就足以对一个看似复杂的问题进行模拟。

6. 聚焦现实世界的关联性和统计思维。书中所有的例子和练习题都有明确的现实背景，都聚焦于如何为进一步学习统计推断和统计建模打下坚实的理论基础。我们简要介绍了重要的统计思想，例如抽样、模拟、贝叶斯推断和马尔可夫链蒙特卡罗方法及其应用领域，包括基因学、医学、计算机科学和信息科学等。对例题和练习题的选择都是为了突出概率思维的力量、适用性及其美之所在。

致谢

感谢我们的同事、Stat110 的教学助理和数千位 Stat110 的学生所给出的与这门课程和这本书相关的评论及想法。特别要感谢 Alvin Siu、Angela Fan、Anji Tang、Carolyn Stein、David Jones、David Rosengarten、David Watson、Johannes Ruf、Kari Lock、Keli Liu、Kevin Bartz、Lazhi Wang、Martin Lysy、Michele Zemplenyi、Peng Ding、Rob Phillips、Sam Fisher、Sebastian Chiu、Sofia Hou、Theresa Gebert、Valeria Espinosa、Viktoriia Liublinska、Viviana Garcia、William Chen 和 Xander Marcus 对本书的反馈。尤其感谢 Bo Jiang、Raj Bhuptani、Shira Mitchell 和那些匿名的审稿人针对本书草稿所给出的详细评论，及 Andrew Gelman、Carl Morris、Persi Diaconis、Stephen Blyth、Susan Holmes 和 Xiao-Li Meng 关于概率的无数次富有深刻见解的讨论。

CRC 出版社的 John Kimmel 在本书的写作过程中提供了极好的编辑方面的专家意见，对他的支持深表感激。

最后，对我们的家人致以最深的谢意，感谢他们对我们的爱和鼓励。

Joseph K. Blitzstein 和 Jessica Hwang
分别于马萨诸塞州剑桥市和加利福尼亚州斯坦福市
2014 年 5 月

目录

第1章 概率与计数

运气、巧合、随机、随机性、不确定性、风险、怀疑、时运、机会——这些词你应该听到过无数次，但对它们的用法可能很模糊。然而，尽管概率普遍存在于科学和人们的日常生活中，但它可能是非常反直觉的。如果我们在不知道正确与否的情况下依靠直觉，就会出现预测不准确或面临预测过于自信而带来的严重风险。本书旨在搭建一个有原则的量化不确定性和随机性的逻辑框架来介绍概率学。同时，本书还致力于强化读者自身的直觉，包括当人们的初始猜想与逻辑原理一致或不一致时的情况。

1.1 为什么要学习概率论?

数学是一种确定性的逻辑；而概率论是一种不确定性的逻辑。概率论在很多领域都有广泛应用，它为理解和解释差异、噪声中的分离信号以及复杂现象的建模等提供了一种工具。以下是从概率论不断扩展的应用领域中选取的一些简单示例。

1. 统计学：概率论是统计学的基础和语言，它使得许多强大的方法通过利用数据去了解这个世界成为可能。

2. 物理学：爱因斯坦曾说过："上帝不掷骰子"，但是目前对量子物理学的理解大量涉及自然界最根本层次的概率。统计力学是物理学的另一个重要分支，它建立在概率论的基础上。

3. 生物学：遗传学和概率论密切相关，例如在基因遗传和随机突变模型中均大量涉及概率。

4. 计算机科学：随机算法在运行时会做出随机选择，并且在许多重要应用中，它们比现在已知的其他任何基于确定性的方案更简单高效。概率论也在研究算法性能以及机器学习和人工智能中起着至关重要的作用。

5. 气象学：天气预报是（或应当）以概率的形式来推断和表达的。

6. 赌博：早期很多关于概率论的研究都是旨在解决赌博和机会游戏的问题。

7. 金融学：前面的例子可能存在冗余的情况，需要指出的是，概率是计量金融的核心。例如，随时间变动的股票价格建模以及金融工具的定价等主要基于概率。

8. 政治科学：近些年来，政治科学越来越偏向统计和数量化。例如，纳特·西尔弗（Nate Silver）在预测选举结果方面取得的成功，他在2008年和2012年的美国总统大选中，使用概率模型来实现民意调查并且进行模拟（参见Silver [25]）。

9. 医学：随机临床试验（患者被随机分配接受治疗或安慰剂）的发展近几年来已经改变了传统的医疗研究。就像生物统计学家大卫·哈灵顿（David Harrington）所指出的，"一些人认为这可能是20世纪科学医学中最重要的进步……现代科学让人开心的讽刺之一是通过将偶然变异因子引入到研究设计中，随机试验可以'调整'对照实验中的可观察和不可

观察的异质性。”[17]

10. 生活：生活是不确定的，而概率就是一种不确定性的逻辑。然而，当我们在做每一次决定时都搬出一个概率计算公式是不现实的，对概率进行深刻思考有助于我们避免一些常见错误，充分理解巧合，并且做出更好的判断。

概率论为原则性解决问题提供了方法，同时也会产生陷阱和悖论。例如，我们在本章可以看到，即使是莱布尼茨和牛顿（二人都在17世纪独立发现了微积分），也无法避免一些概率论上的基本错误。在本书中，我们会用到以下几种方法来避免这些潜在错误。

1. 数值模拟：概率论的魅力在于，常常可以通过模拟试验来研究问题。例如可以直接运行一个模拟，看看到底谁是对的，而不用无止境的争论。本书每章结尾部分都会给出一些在R语言（一种免费的统计计算软件）中如何计算和模拟的例子。

2. 生物危害：研究中常见的错误对于理解什么是概率的合理推理有着很重要的作用。在本书中，常见错误就叫作生物危害（Biohazards），并用☣表示（因为犯此类错误会对健康造成伤害）。

3. 完整性检查：当通过一种方法解决问题后，我们通常会尝试使用另一种方法来解决同样的问题，或者检查所得到的答案在简单和极端的情况下是否都是合理的。

1.2 样本空间

概率论的整个框架是建立在集合的基础上的。假设做一个试验，试验结果是所有可能结果集合中的一个，在试验之前，不知道会是哪一个结果。做完试验之后，其结果会具体化为实际的结论。

定义 1.2.1（样本空间和事件） 一个试验的样本空间 S 就是这个试验可能出现的所有结果的集合。事件 A 是样本空间的一个子集，称某事件 A 发生，当且仅当 A 中的某个结果实际出现。

一个试验的样本空间可能是有限的，可数无限的或不可数的（参见附录A.1.5中关于可数集和不可数集的解释）。当样本空间为有限的时，我们可以将样本可视化为圆点，如图1.1所示，每个圆点代表一个结果，一个事件就是一些圆点的集合。

做一次试验就等同于随机选取一个圆点。如果所有的圆点都是相同的，那么所有的圆点被抽中的可能性就相同。这个例子是后面两节讨论的重点。在1.6节，我们令每个圆点质量不同，给出了概率的一般定义。

集合理论在概率论中用处广泛，它为表达和处理事件提供了丰富的语言；附录A.1回顾了集合理论。集合的运算，特别是集合的并、交、补，可以轻松地基于已定义事件来创建新事件。这些理论让我们可以用不止一种方式来表达事件。对于同一事件，通常存在一种最为简单的表达方式。

例如，S 是一个试验的样本空间，令 $A,B\subseteq S$ 表示事件，则 $A\cup B$ 就是一个事件，这个事件发生，当且仅当 A、B 中至少有一个事件发生；事件 $A\cap B$ 发生，当且仅当 A、B 两个事件同时发生；事件 A^c 发生当且仅当事件 A 不发生。另外还有德摩根（De Morgan）律：

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

它表示两个事件至少其一发生的反面是两个事件都不发生，两个事件都发生的反面是两个事

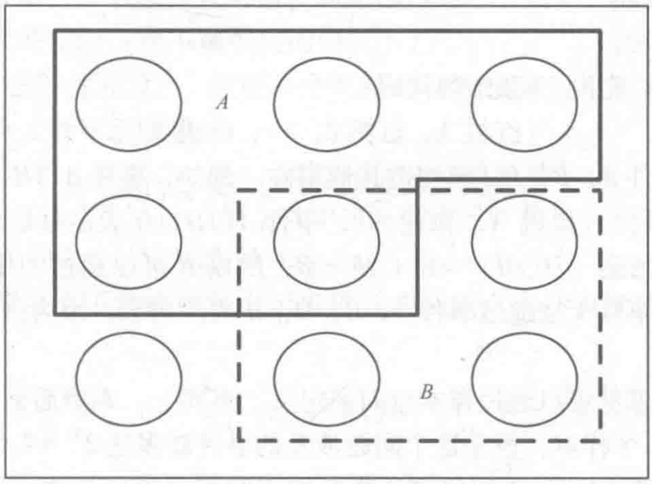


图 1.1　一个用点来表示的样本空间，其中用实线框表示事件 A，用虚线框表示事件 B。

件至少其一不发生。对于两个以上的事件的交和并，也有相似的结果。

如图 1.1 所示，A 是一个包含 5 个圆点的集合，B 是一个包含四个圆点的集合，$A\cup B$ 表示在 A 或在 B 的圆点集合，由 8 个圆点组成，$A\cap B$ 表示同时在两个集合的圆点集合，只有 1 个圆点。A^c 表示不处于 A 中的圆点集合，共有 4 个圆点。

样本空间的概念广泛且抽象，所以记住一些具体例子对理解这个概念非常重要。

例 1.2.2（掷硬币）　掷一枚硬币 10 次，正面用 H 表示反面用 T 表示，一个可能的结果为 HHHTHHTTHT，且样本空间就是所有长度为 10 的由 H 和 T 组成的字符串的集合，令 H 为 1，T 为 0，所以一个结果就可以用一个数列$(s_1,s_2,\cdots,s_{10})$表示，其中 $s_j\in\{0,1\}$，样本空间即为这些数列的集合。下面我们看几个事件。

1. 设 A_1 为一个事件，表示第一次掷硬币结果为 H，代表的集合如下：

$$A_1=\{(1,s_2,\cdots,s_{10}):s_j\in\{1,0\}\text{且}2\leqslant j\leqslant 10\}。$$

这是样本空间的一个子集，所以可以将它看作一个事件；事件 A_1 发生等价于第一次掷硬币的结果为 H。同样地，可以设 A_j 为事件第 j 次掷硬币结果为 H，其中 $j=2,3,\cdots,10$。

2. 设 B 事件表示至少有一次掷硬币结果为 H，可用如下集合表示为

$$B=\bigcup_{j=1}^{10}A_j。$$

3. 设 C 事件表示所有掷硬币结果都为 H，用集合表示为

$$B=\bigcap_{j=1}^{10}A_j。$$

4. 设 D 事件表示至少有两个连续的 H，用集合表示为

$$D=\bigcup_{j=1}^{9}(A_j\cap A_{j+1})$$ □

例 1.2.3（任意抽取一张纸牌）　从标准的 52 张纸牌中任意抽取一张。此时样本空间 S 为全部 52 张纸牌的集合（所以这里有 52 个圆点，每个圆点代表一张牌）。考虑如下四个事件。

- A：点数是 A。
- B：颜色为黑色。

- D：花色为方块。
- H：花色为红桃。

H 作为一个集合，它由 13 张纸牌构成：

$$\{红桃\ A，红桃\ 2，\cdots，红桃\ K\}。$$

我们可以通过事件 A，B，D，H 构造其他事件。例如，事件 $A\cap H$ 表示抽取的纸牌为红桃 A；$A\cap B$ 等价于事件｛黑桃 A，梅花 A｝，事件 $A\cup D\cup H$ 表示抽取的纸牌为红色或者点数为 A。同样地，注意到 $(D\cup H)^c=D^c\cap H^c=B$，所以 B 可以通过 D 和 H 来表示。另一方面，纸牌为黑桃这个事件无法通过事件 A，B，D，H 表示出来，因为这几个集合无法细致到能划分出黑桃和梅花。

还有许多其他的事件可以通过样本空间来定义。事实上，本章后面介绍的计数方法表明虽然样本空间只有 52 个样本，但是这个问题涉及的事件却多达 $2^{52}\approx 4.5\times 10^{15}$ 个。

如果抽到的纸牌为 joker，该如何表达呢？由于我们讨论的样本空间是错的，因此无法表达；这说明假设试验结果必须是样本空间 S 中的一个。 □

如前面的例子所示，事件可以用语言或集合符号来描述。通常语言描述更容易解释，而集合符号更容易操作。设 S 为样本空间，s_{actual} 为实际的试验结果。下面给出了语言描述与集合符号之间的转换表。例如，对于事件 A 和 B，陈述语句"A 含于 B"指当事件 A 发生时，事件 B 也发生；就集合而言，翻译为 A 是 B 的子集。

语言描述	集合符号
事件	
样本空间	S
s 是一个可能的结果（元素）	$s\in S$
A 为事件	$A\subseteq S$
事件 A 发生	$s_{\text{actual}}\in A$
必然发生的结果	$s_{\text{actual}}\in S$
根据原有事件得到新事件	
事件 A 或事件 B 发生	$A\cup B$
事件 A 与事件 B 同时发生	$A\cap B$
事件 A 不发生	A^c
A 或 B 发生，但不同时发生	$(A\cap B^c)\cup(A^c\cap B)$
事件 A_1，…，A_n 至少有一个发生	$A_1\cup\cdots\cup A_n$
事件 A_1，…，A_n 都发生	$A_1\cap\cdots\cap A_n$
事件间的关系	
事件 A 含于事件 B	$A\subseteq B$
A 与 B 互不相容	$A\cap B=\varnothing$
A_1，…，A_n 是样本空间 S 的一个划分	$A_1\cup\cdots\cup A_n=S$，$A_i\cap A_j=\varnothing$，$i\neq j$

1.3 概率的朴素定义

历史上，概率最早定义为使事件发生的所有可能的方法数除以试验所有可能出现的结果

数。由于它有很强的限制性，依赖强大的假设因此称其为朴素定义；但是，理解该定义是非常重要的，且有着强大的作用（在不被误用的情况下）。

定义 1.3.1（概率的朴素定义） 假设一个试验的样本空间为有限集合 S，设 A 为这个试验的一个事件，事件 A 概率表达如下：

$$P_{\text{naive}}(A)=\frac{|A|}{|S|}=\frac{\text{会导致 } A \text{ 的结果数量}}{S \text{ 中总共的结果数量}}。$$

（这里用 $|A|$ 表示集合 A 的大小；参见附录 A. 1. 5。）

在点空间中，朴素定义下的概率就表示事件 A 中的圆点占所有圆点的比例。例如，如图 1. 1 所示

$$P_{\text{naive}}(A)=\frac{5}{9},P_{\text{naive}}(B)=\frac{4}{9},P_{\text{naive}}(A\cup B)=\frac{8}{9},P_{\text{naive}}(A\cap B)=\frac{1}{9}。$$

考虑事件的补集，有

$$P_{\text{naive}}(A^c)=\frac{4}{9},P_{\text{naive}}(B^c)=\frac{5}{9},P_{\text{naive}}((A\cup B)^c)=\frac{1}{9},P_{\text{naive}}((A\cap B)^c)=\frac{8}{9}。$$

一般来说，

$$P_{\text{naive}}(A^c)=\frac{|A^c|}{|S|}=\frac{|S|-|A|}{|S|}=1-\frac{|A|}{|S|}=1-P_{\text{naive}}(A)。$$

我们在 1. 6 节中将看到关于补集的很多结论在概率论中也同样适用，且不仅限于概率的朴素定义。当我们试图求出一个事件的概率时，一种较好的策略是，首先考虑是直接确定事件本身的概率容易还是确定它的补集概率更容易。德摩根律在这里就变得非常有用，因为它使得处理交和并变得更容易。

朴素定义有很强的限制性，它要求样本空间 S 必须是有限的，且每一个样本都有相同质量。当人们在没有任何判断和论证的情况下，就假设各结果有相同的可能性，“它只有两种可能，发生或不发生，但我们不知道哪个，所以可能性各为 50%”，此时该定义就会被错用。这种推理除了有时会给出荒谬的概率外，甚至还会产生自相矛盾。例如，它会得到结论在火星上有生命的概率为 1/2（“有或者没有”），同时也会得到火星上有智能生物的概率为 1/2 的结论，但由 1. 6 节中介绍的概率性质可以证明，很显然这里存在矛盾——后者概率应该严格小于前者概率。但在以下几种重要的问题中，朴素定义都是适用的：

- 当问题存在*对称性*使得结果可能等同出现时。假设一枚硬币正反面落地的概率都为 50%，这是符合常识的，因为在物理上硬币是均匀对称的。[㊀] 一副标准的、洗好的扑克牌，假设所有排序有相同可能性是合理的。不存在一些特殊的牌更倾向于靠近前面的位置；每个特定位置对于 52 张牌中的每一张都有相同的可能性。

- 当把试验的每个不同结果*设计*为具有相同可能性时。例如，考虑在总量为 N 个人的总体中抽取 n 个人进行一项调查。一种常见的目标是获得一个*简单随机样本*，即令这 n 个人从所有大小为 n 的子集中以相同可能性随机抽取。如果成功的话，就说明朴素定义是适用

㊀ 参见 Diaconis，Holmes 和 Montgomery［8］的物理学结论，抛掷一枚硬币，抛掷前后出现相同面的可能性为 0. 51（接近但略大于 1/2），Gelman 和 Molan［12］解释了为什么一枚硬币即使制作成两面重量不相等，出现正面朝上的概率依然接近于 1/2（对于常规的掷硬币行为；不包括允许硬币旋转等）。

的，但事实上由于各种复杂的原因，这种设想可能很难实现，比如说无法获取一个完整、精确的个人信息联系表。

• 当朴素定义被作为一个有效的零模型时。在这种设定下，我们假设朴素定义只适用于观察将会产生怎样的预测，然后再通过比较观测数据和预测值来评估这种等可能假设是否成立。

1.4 如何计数

计算一个事件 A 的朴素概率涉及计算事件 A 包含的圆点个数和样本空间 S 中总共的圆点个数。通常我们需要计算的子集数目庞大。本节介绍一些计数的基本方法；更多方法可以参考组合学的书籍（组合学是研究计数问题的一个数学分支）。

1.4.1 乘法法则

一些情况下，可以直接根据一个基本但是通用的法则——乘法法则进行计数。概率论和统计中常遇到*有放回抽样*和*无放回抽样*，本节将介绍乘法法则针对这两种情形会自然产生相应的计数规则。

定理 1.4.1（乘法法则） 考虑一个复合试验，它由两个子试验 A 和 B 构成。假设试验 A 有 a 种可能结果，每一种情况又都对应于试验 B 的 b 种可能结果。那么这个复合试验共有 ab 种可能结果。

想了解为什么乘法法则是正确的，这里可以想象一个如图 1.2 所示的树状图。假设该树有 a 个主分支，分别代表着试验 A 的 a 种结果，每一个主分支又延伸出 b 个分支代表试验 B 的 b 种结果。所以总共就有 $\underbrace{b+b+\cdots+b}_{a}=ab$ 种结果。

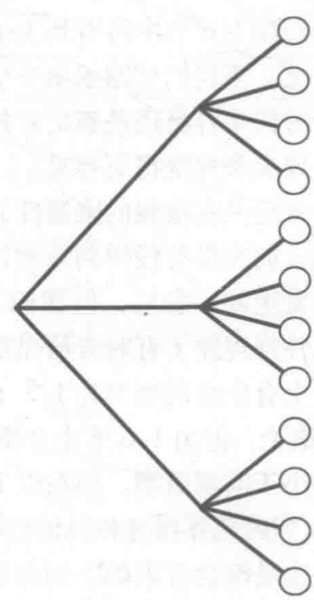


图 1.2 这个树状图可用来说明乘法法则。如果试验 A 有 3 种结果，试验 B 有 4 种结果，则试验总共有 $3\times4=12$ 种可能的结果。

注 1.4.2 通常很容易想当然地认为试验是有先后顺序的，但是乘法法则并没有要求试验 A 必须在试验 B 之前实施。

例 1.4.3（冰淇淋甜筒） 假设某人正在购买一支冰淇淋甜筒（ice cream cone），甜筒（cone）有蛋卷（cake）或华夫饼（waffle）两种选择，冰淇淋有巧克力（C）、草莓（S）和香草（V）三种口味（flavor）可选。决定的过程可以用如图 1.3 所示的树状图来表示。

由乘法法则可知，共有 $2\times3=6$ 种选择。这是个较为简单的示例，但却值得仔细思考以推广至更为复杂的情况，后面我们还将会遇到一些复杂的问题，它们在已知的空间中无法绘制出树状图，但从概念上讲还是可以将其看作冰淇淋甜筒问题。这里有一些值得注意的问题：

1. 无论是先决定甜筒（cone）类型还是冰淇淋口味（flavor）类型，试验都是总共有