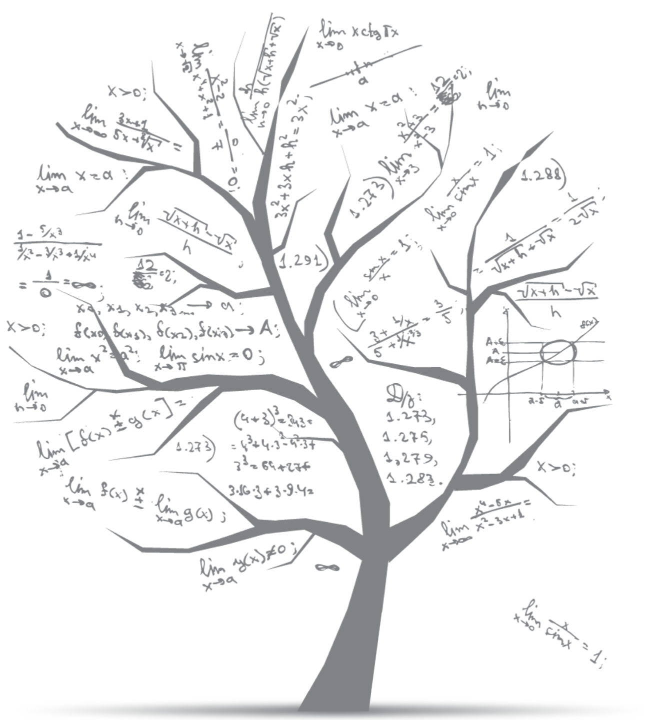
TURING

图灵程序
设计丛书

程序员的数学 2
概率统计

[日] 平冈和幸　堀玄 / 著　陈筱烟 / 译



人民邮电出版社
北　京

图书在版编目（CIP）数据

程序员的数学. 2，概率统计 / （日）平冈和幸，
（日）堀玄著；陈筱烟译. -- 北京：人民邮电出版社，
2015.8
（图灵程序设计丛书）
ISBN 978-7-115-40051-2

Ⅰ. ①程… Ⅱ. ①平… ②堀… ③陈… Ⅲ. ①电子计
算机—数学基础②概率统计 Ⅳ. ①TP301.6②O211

中国版本图书馆CIP数据核字(2015)第176373号

内 容 提 要

本书沿袭《程序员的数学》平易近人的风格，用通俗的语言和具体的图表深入讲解程序员必须掌握的各类概率统计知识，例证丰富，讲解明晰，且提供了大量扩展内容，引导读者进一步深入学习。

本书涉及随机变量、贝叶斯公式、离散值和连续值的概率分布、协方差矩阵、多元正态分布、估计与检验理论、伪随机数以及概率论的各类应用，适合程序设计人员与数学爱好者阅读，也可作为高中或大学非数学专业学生的概率论入门读物。

◆ 著　[日] 平冈和幸　堀　玄
译　陈筱烟
责任编辑　徐　骞
执行编辑　高宇涵
责任印制　杨林杰

◆ 人民邮电出版社出版发行　北京市丰台区成寿寺路11号
邮编　100164　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
北京　印刷

◆ 开本：800×1000　1/16
印张：26.5
字数：492千字　2015年8月第1版
印数：1-4 000册　2015年8月北京第1次印刷

著作权合同登记号　图字：01-2013-8603号

定价：79.00元

读者服务热线：(010)51095186转600　印装质量热线：(010)81055316
反盗版热线：(010)81055315
广告经营许可证：京崇工商广字第0021号

译者序

说到与程序设计关系紧密的数学学科，肯定有不少人首先就会想到概率论与统计学吧。从信息论到机器学习，从模式识别到数据挖掘，概率与统计的概念和原理活跃于计算机科学的各个领域，发挥着重要的作用。同时，概率与统计也都是应用性极强的学科。它们源于现实需求，在不断发展成熟后又反过来推进了各类问题的解决。纯粹的理论推演无法展现概率论与统计学的全部魅力，计算机科学的出现，为概率统计搭建了绝佳的应用舞台。

话虽如此，充满了大量概念与公式的概率与统计理论并非总是轻松有趣。概率论中既有简单明了的结论，也有与直觉相违的原理。只有准确地理解并掌握所有这些，才能真正在解决实际问题时灵活运用它们。幸运的是，这本《程序员的数学2：概率统计》另辟蹊径，借助不同于传统数学教材的独特编排思路，将概率泛化为面积与体积，强调随机变量的函数属性，深入浅出，巧妙地阐述了概率论与统计学的基本理论。全书穿插了大量形象有趣的实例，并辅以难度恰当的练习题，使没有经过大学数学系课程专业训练的读者也能理解与掌握概率与统计的精髓。

既然题为“程序员的数学”，本书自然也专门介绍了一些与计算机科学相关的主题。本书涵盖了随机数理论、贝叶斯公式、马尔可夫链、信息熵等计算机科学专业课程中涉及的知识点，不但能够作为相关课程教材的补充，还能启发读者理解甚至设计相关的计算机算法，解决与概率统计相关的计算问题。本书的习题量不大，但都很典型，是检测相关知识点理解程度的理想材料。为了深入掌握本书内容，建议读者准备纸笔，随时通过笔头推算与演练，必能获得更好的学习效果。

数学是一门严谨的学科。翻译以数学为主题的书籍，不仅需要准确地理解原文，将其转换为恰达的中文，更重要的是严格精确地表述各条数学表达式的含义。中日数学表述体系间存在的细微差异无形中为此增加了难度，遑论国内数学教材对部分名词术语的翻译尚未达成统一，一字一句，都必须推敲琢磨。原则上，本书选择了当前学界较为流行的中文术语与公式表述方式，关键的公式与理论也经过了审校与确认。然限于水平，谬误之处在所难免，敬请广大读者给予批评与斧正。

本书能顺利完稿，离不开许多人的帮助与支持。家人的理解与关心，使自己能够始终安心地翻译。当我在生活与工作中遇到困难时，好友的支持帮助我攻克了一个个难关。最

后还要感谢图灵公司的编辑提出了大量有价值的建议与意见，帮助本书顺利完成并最终问世。最后，希望对概率统计有兴趣的读者们都能从本书中获益。

陈筱烟

2015年6月于上海

前　言

本书的编写目标

本书的目标读者是那些希望在自己感兴趣的领域中运用概率统计知识，但并不打算成为数学专业人士的人。本书名称沿用笔者的另一本书《程序员的数学3：线性代数》，但内容更加普适而通用。

说起概率论与统计理论，很多人都会对此抱有一些负面印象：计算公式复杂，检验类型繁多，辛苦学习很久后却发现真正对工作有帮助的其实是那些电子表格软件。不少人在运用统计理论时态度也很消极，他们只是为了让别人相信自己的结论客观合理而被迫使用那些繁琐的处理。

然而概率统计理论并非没有价值。事实上，近年来为了更好地处理数据，人们开始在各个学科中积极应用概率统计理论。数据挖掘、垃圾邮件自动筛选、文档的自动分类与类似文件搜索、非法使用的鉴别（例如根据信用卡消费历史得出持卡人的消费模式，以检测非正常消费行为，或监视网络流量识别可疑的访问模式等）、语音识别与图像识别、通信工程（例如信号清晰的高音质通话技术）、基因分析、金融工程中的资产组合分析、受生物学启发得到的高效信息处理方式（神经网络、遗传算法等）以及基于蒙特卡罗模拟的循证式日程规划等各领域中都活跃着概率统计理论的身影。

为了了解这些技术，我们必须掌握一些概率统计基础知识。然而，目前面向非理科背景读者的教材少之又少，他们往往不知道应该怎样学习概率统计知识。

概率论的专业教材需要一定的数学功底才能阅读，普通读者可能难以理解；而轻松易懂的入门书又常常不成体系，或是无法提供足够的信息。如果要应用概率统计理论，我们需要掌握更多的基础知识。例如，在实际应用中我们常会遇到多个事件同时发生的情况，如何才能胸有成竹地处理这类问题，而不是仅凭直觉或模糊的概念妄加猜测呢？为此，除了了解基本的计算步骤，我们还需要深刻理解以下两条概念。

- 概率是面积与体积的泛化
- 随机变量是一种以变量为名的函数

不了解这方面的读者可能难以理解这两句话，不过它们已经成为现代概率论的基础。

另一方面，与统计相关的参考书通常会罗列各类分布与检验方法。可惜这种编排方式难以帮助我们将概率统计理论应用至上述当前热门的领域。然而，直接阅读专门介绍垃圾邮

件分类的技术书也并非良策。换个角度讲，我们不可能仅靠从应用中学到的知识来改进已有的应用方式。如果基础知识不足，在实际应用中就不能深入理解问题，从而无法触及问题的本质。何况这些技术至今仍在不断发展，只学习一些阶段性的未成熟的技术并没有太大的价值。相比研究这些具体的检验方法与应用技术，我们更应该首先强化相关的基础知识，以应对(包括当前方式在内的)更广泛的应用情景。

如果读者已经具备了一定的数学基础(不但能够理解数学表达式，还能领会那些抽象而巧妙的描述方式)，教材的选择就会轻松很多。注重学习效率的教材、基础知识详细的教材、讲解生动有趣的教材，不一而足，大家可随意挑选。然而，由于存在这一门槛，普通读者缺乏合适的教材，于是能够掌握稍微系统一点的概率统计知识的人就一下子少了很多。

为此，本书在编写时考虑了以下三个原则。

- 精选了大量非数学专业读者也应当掌握的知识点
- 内容的深度与大学非数学系学生的学习能力相符
- 知识的讲解力求详细具体

因此，与传统教材相比，本书的着力点与讲解思路都较为独特。尽管本书面向初学者，却保持了一定的深度。虽然前半部分花了较长篇幅详解条件概率的概念，但省略了各种类型的分布，也没有具体介绍各类计算技巧及特征函数等内容。这是权衡了各个知识点的重要程度之后做出的决定。此外，有些部分的讲解看似冗长，实则语言明晰，非常易于理解。

在讲解各类估计与检验方式等统计学核心知识点时，本书也下了一番功夫。本书希望读者不仅能了解它们的用法，还能理解其背后的原理，因此详细介绍了相关的基础概念与思维方式。上文提到过，本书并不打算罗列所有的统计方法(尤其是数据的采集方式与区间估计等内容)。如果读者想要学习各种具体的估计与检验方法，请参考专门的统计学教材。本书将着重说明这些方法背后的理论依据，解释这些方法背后共通的原理。

0.1 在阅读本书前，需要预先掌握哪些数学知识[①]

- 需要具备高中理科级别的数学知识(包括向量、微积分等的概念与基本计算能力)
- 偶尔会用到大学级别的数学知识(主要是多元微积分)，不过本书会为不具备这方面知识的读者提供必要的解释
- 不会使用大学数学系级别的知识(主要是测度论)。如果讲解时无法避免，本书将明确提示，并给出易于理解的说明，帮助读者建立直观印象

此外，第5章与8.1节需要用到一些大学程度的线性代数知识(例如，需要通过正交矩阵将对称矩阵对角化)。如果读者之前不了解线性代数理论的应用价值，也许会惊叹它的巨大威力。

最后，附录A简单总结了求和符号、指数与对数的一些基础知识，读者如有需要可适当参考。

① 该部分在正文中称为“问答专栏”。——译者注

0.2　为什么要在写线性代数教材的同时再写一本概率统计教材

撰写这两本教材都是因为该学科的初学者与专家之间存在人才断层现象。在这两个领域中，一些有经验的人能够轻松解决的问题，初学者却常常感到过于复杂。产生这一问题的根本原因在于，无论是矩阵还是概率，不同水平的人对它们的理解程度往往大相径庭。

例如，只要具备一定的线性代数基础，就一定会知道矩阵的本质是一种映射。然而，考虑到初学者不一定能够理解这种抽象的解释，入门教材一般会使用更浅显的说明方式。因此，笔者希望打破这一惯例，教授初学者矩阵更加本质的含义，即，

矩阵是一种映射！

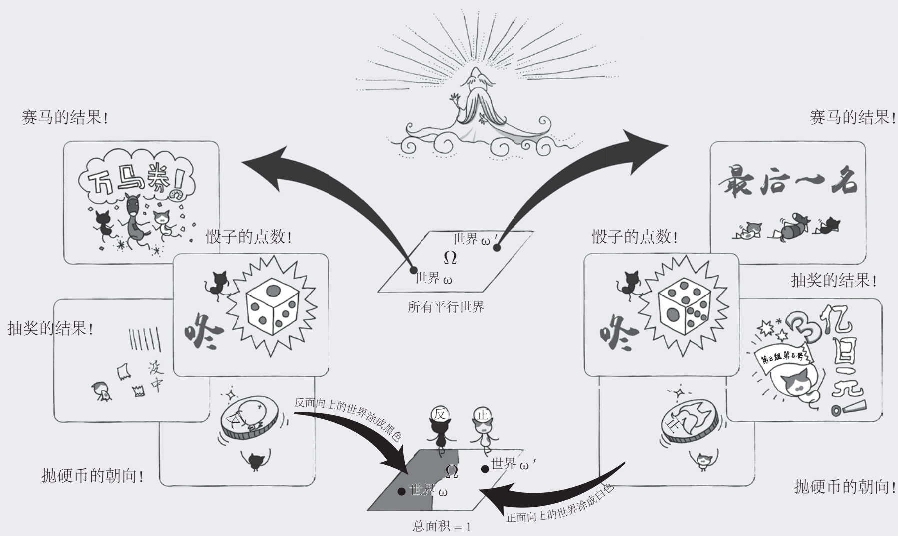
笔者相信，这本书证明了如下几点。

- 只要讲解详细，说明充分，初学者也能理解抽象概念的本质
- 即使不打算成为数学方面的专家，掌握这些本质也大有裨益
- 事实上，首先阐明概念的本质，往往更利于读者理解

本书也将以此为目标。对于具备一定的概率统计基础的人来说，以下这句话是众所周知的常识。

概率是一种面积！

本书希望让更多的人理解这一概念[①]。1.3 节将借助“上帝视角”具体讲解这句话的含义（它也许与读者的想象有些差别）。在理解了这句话之后，概率论中的很多问题都会迎刃而解，敬请期待。



① 准确地讲，概率是一种测度。测度可以简单理解为面积与体积等可测量的量的泛化概念。

本书的构成

本书包含以下两大部分。

- 第 1 部分：聊聊概率这件事
- 第 2 部分：探讨概率的应用

其中，第2部分中的各个主题相对独立，读者可依个人兴趣与喜好有选择地阅读。

不过，这并不是这样分类的真正原因，这背后还有一个更为深层的理由，即“不希望读者混淆观测值与该观测值背后的生成机制”。

我们来看一个极为简单的例子。

- 背后的生成机制：算命箱中有七成是吉，三成是凶
- 观测值：5人抽签后，4人抽到了吉，1人抽到了凶

我们可以从7/10与4/5这两种角度来讨论吉凶的比例，但两者的含义与意义完全不同。我们可以直接观测到观测值，但无法观测该值生成的机制。要真正理解并应用统计方法，必须明确区分两者。

第1部分虽然会告诉读者生成机制，但仅讨论由该机制生成的观测值具有怎样的性质。在充分理解这部分内容之后，第2部分的重点将同时关注生成机制本身，讲解观测值怎样通过其背后的生成机制生成。我们姑且称前者为正问题，后者为逆问题。由于无法得到完整的信息，逆问题往往更难解决，且在讨论逆问题前，我们必须对正问题有深刻的理解。

现实问题往往属于逆问题，如果为了吸引读者而在正问题中混入逆问题，就很可能使读者感到混乱。为此，第1部分将集中讨论正问题，在读者对概率的基本概念了然于胸之后，再开始研究难度更大的逆问题。

0.3 为什么第 1 部分的例子多与赌局或游戏相关

上文所说的观测值与生成机制正是产生这种情况的原因。赌局与游戏背后的生成机制即是它们的规则，而这些规则我们可以事先知道。因此，它们非常适合作为第1部分的实例。

在线资源

读者可以从图灵社区的网站上下载在各章末专栏的实验中执行的Ruby脚本[①]。

谢辞

原启介、高桥信、小山达树与堀英彰审读了本书的原稿，并对内容、编排与表述方式提供了很多有价值的建议。[②] 小柴健史给伪随机数部分提出了宝贵意见。使用过本书Beta版的学习小组也提供了很多有益的反馈。kogai、studio-rain、mia、pige以及其他一些匿名网友为本书在互联网上公开的草稿提出了许多勘误与问题。欧姆社开发部的诸位从各方面为本书的写作提供了支持[③]，使本书达到了仅凭笔者一人绝对无法达到的水准。mUDA为本书的关键概念绘制了形象有趣又令人印象深刻的插图。笔者在此向诸位表示衷心的感谢。本书得以面世，离不开大家的努力(因笔者水平所限，本书难免还有不足之处，望读者见谅并批评)。

笔者

2009年10月

① http://www.ituring.com.cn/book/1254，点击“随书下载”。——编者注

② 阅读尚不成熟的原稿十分辛苦。多亏了几位的早期审读，笔者才能对包括本书整体框架在内的方方面面做出巨大改进。

③ 大家除了准备邮件列表、版本管理系统、错误追踪系统等基本的软件环境，还提供了可以直接通过make指令确认运行结果的实验模拟程序，技术水平非常高。

目 录

第2部分 探讨概率的应用

第1部分

聊聊概率这件事

第1章 概率的定义

A：根据我的调查，明年经济景气的概率是**71.42857...%**。

B：你怎么能得到那么精确的数字？

A：模拟了**7**种设定后，其中**5**种情况下经济将会好转，于是从**5/7**得到**0.7142857...**

B：呃……槽点太多都不知道该怎么回答你了。总之我们先了解一下概率究竟是什么意思吧。

概率的难点在于它不够直观。如果我们的头脑中没有具体的概念，仅凭直觉想象，就很难深刻理解概率的本质。

本章将形象地描述概率这一抽象概念，将其转换为易于理解的具体问题并分析。本章所讲内容的要点在于“概率即是面积”。如果读者能掌握这种思考问题的方式(代入所谓的上帝视角)，很多问题就将迎刃而解。

作为构筑本书基础的一章，本章将首先介绍概率论的“舞台”——三元组$(\Omega, \mathcal{F}, \mathrm{P})$，然后再介绍“主演”——随机变量与概率分布。

1.1 概率的数学定义

概率是什么？这个问题很难回答。

- 掷骰子时，点数为1的概率是1/6
- 打扑克牌时，抽中黑桃的概率是1/4

有了这些概率，我们很容易就能得到如下结果。

- 掷骰子600次后，大约有100次结果为1
- 抽扑克牌400次后，大约有100次抽中黑桃

那么，下面这样的概率是否存在呢？

- 明天某地下雨的概率
- 1192年6月6日某地曾下过雨的概率

“明天下雨的概率是30%”，这样的结论究竟有没有意义呢？如果某事件仅发生一次，尽管我们依然可以讨论该事件的概率，但前文这样的解释已不再适用。更何况“曾下过雨”是在讨论过去的事件，就算说话者本人不清楚当时的情况，是否下过雨早已是既成事实。这类既定事件与概率一词有何关联吗？

“是否有意义？”这类问题，如果较起真儿来，那就没完没了了，讨论将无法展开。因此，在数学中，我们暂且不管问题是否有意义，先抽象地定义了概率的概念。通常，概率的具体定义为“满足下列条件的三元组 $(\Omega, \mathcal{F}, \mathrm{P})$ 称为概率空间。这些条件包括……”[①] 之类，如是云云。总之，根据这一定义，我们就能对概率执行各种运算并得到符合要求的结果。因此，我们还是别对这种抽象的定义抱有怨言了。不过，并非人人都是数学专家，要让业余爱好者们也能读懂这些定义显然不切实际。为此，我们将设法统合概率的直观印象与抽象定义，这也是本章的目标。

? 1.1　为什么面向业余爱好者的书非要使用如此抽象的定义

本书执意介绍三元组 $(\Omega, \mathcal{F}, \mathrm{P})$ 出于以下两个理由。

第一个理由很简单，就是因为有趣。以新的视角看待问题不是一件令人激动的事吗？如果要换用一种与过去全然不同的思考方式考虑问题，自然会感到费力。不过一旦跨越了这一障碍，就会豁然开朗，不由得便会觉得干劲十足。

另外，学习三元组能够帮助我们迅速理解概率论中的很多理论，借助 $(\Omega, \mathcal{F}, \mathrm{P})$，很多原本抽象难懂的概念与性质都将变得一目了然。例如，本书的第2章、3.3节、3.5节等都能通过三元组解释。

1.2　三扇门（蒙提霍尔问题）——飞艇视角

如果直接讲解 $(\Omega, \mathcal{F}, \mathrm{P})$，读者可能不容易理解，因此，我们先讨论下三扇门游戏。这也称为蒙提霍尔问题（Monty Hall problem），是一个争论不断的著名问题。以该问题开始本节的讨论或许有些绕。不过本节的主题正是如何处理这类复杂问题，因此还请读者耐心阅读。

---

① Ω 是希腊字母 ω（Omega）的大写形式。$\mathcal{F}$ 是 F 的异体。读者暂时不必在意这些细节问题。

有些读者可能已经对蒙提霍尔问题有所了解，但本节将侧重于说明它与$(\Omega, \mathcal{F}, \mathrm{P})$的关系，仍有一读的价值。

1.2.1　蒙提霍尔问题

图1.1中有三扇门。其中只有一扇是正确的门，打开后将能获得一辆高档豪车。另两扇门是错误选项，门内只有山羊。从门外无法获知哪一扇才是正确选项。挑战者需要从三扇门中选择一扇打开。

在决定选择某扇门后，还剩两个选项，其中至少有一个是错误选择。此时，（知道正确答案的）主持人打开了没被选中的门中错误的那个，让挑战者确认了门后的山羊，并询问："是否要重新选择？"

挑战者是否应当重选，还是应该坚持最初的选择？又或是两种做法没有区别？

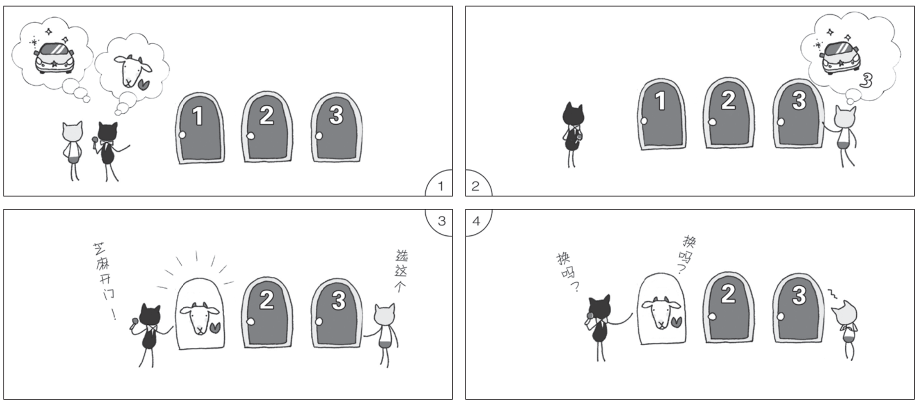


图1.1　蒙提霍尔问题

这就是蒙提霍尔问题。请读者先考虑下自己将如何作答。为使概率的讨论更清晰明了，我们制定了以下前提条件。

- 主持人通过骰子决定将高档豪车藏在哪扇门后（点数1、2表示在门1中，3、4表示在门2中，5、6表示在门3中）
- 挑战者通过骰子决定选择哪扇门（点数1、2表示门1，3、4表示门2，5、6表示门3）
- 如果挑战者选中了正确答案，主持人仍将通过骰子决定打开哪一扇门（即使它们都是错误答案，其中点数1、2表示门1，3、4表示门2，5、6表示门3）

1.2.2　正确答案与常见错误

聪明的读者可能很快就得出了正确答案。试考虑以下情况。

> 在挑战者做出第一次选择之后，有1/3的概率正确，2/3的概率不正确。这很容易理解，无可争辩。那是否应该重新选择呢？我们来仔细看一下规则。
>
> - 如果第一次选择正确，重选必定错误
> - 如果第一次选择错误，重选必定正确
>
> 也就是说，“第一次选择错误”的概率就是“重选后正确”的概率。即重选的正确率是2/3。重选更加有利。

不过，即使能够做出正确的选择，我们有时也很难向那些判断错误的人解释清楚这样选择的原因。例如，有人可能会有下面这样的草率见解。

> 在游戏开始时，存在三种可能。
>
> $$\begin{cases}\text{门1是正确答案（概率1/3）}\\\text{门2是正确答案（概率1/3）}\\\text{门3是正确答案（概率1/3）}\end{cases}$$
>
> 假设挑战者选择门3，而主持人打开了门1。于是，第一种情况将不再成立，只剩下两种可能。
>
> $$\begin{cases}\text{门2是正确答案（概率1/2）}\\\text{门3是正确答案（概率1/2）}\end{cases}$$
>
> 此时，重新选择门2与继续选择门3的概率似乎都是1/2。

得到了这一错误结论的人，无论怎样向他解释，都难以认同我们之前的说法。他们会觉得“虽然你说的也有道理，但我的想法也没有错”，讨论将不了了之。那我们应该怎么办呢？

1.2.3　以飞艇视角表述

概率是一种很抽象的概念，如果我们仅凭直觉判断，很难清晰理解它的本质。为此，我们希望换一种视角来表述概率，尽量把问题转换成一种可以实际衡量的形式。具体来讲，就是放弃使用骰子。这种思维模式的转换，就是现代数学中概率定义的基石，也是本书要传达的核心思想。在刚接触这种思考方式时，读者可能会有些不适应，不过一旦习惯之后，

许多概率问题都能迎刃而解，因此，让我们放下固有观念，试着体会这种新的视角吧。

如图1.2所示，假设我们准备了大量游戏会场。在一个巨大的广场中分设了360个游戏会场，每个会场将同时进行游戏。你现在乘坐在飞艇中，从上空俯视这些会场。与之前的规则不同，我们这次为每个会场准备了剧本，主持人与挑战者都将按照剧本表演。所有人的行动都已事先确定。不过，每一个会场的剧本内容各不相同。

在视角转换前，主持人通过掷骰子来决定哪一扇门是正确答案。由此，我们将剧本设定为360个会场中有120个会场的门1是正确答案，120个会场的门2是正确答案，120个会场的门3是正确答案。接下来由挑战者进行选择，但我们同样不再使用骰子。取而代之的是，在门1是正确答案的120个会场中，有40个会场的剧本要求挑战者选择门1，40个会场选择门2，40个会场选择门3。门2是正确答案的120个会场和门3是正确答案的120个会场也同样如此。以上内容整理如下。

	挑战者选择门1	挑战者选择门2	挑战者选择门3
门1是正确答案	○40个会场	×40个会场	×40个会场
门2是正确答案	×40个会场	○40个会场	×40个会场
门3是正确答案	×40个会场	×40个会场	○40个会场

图1.2　飞艇视角

之后，主持人将打开一扇错误的门。请读者特别注意此处。如果挑战者的选择错误(×)，剩下的两扇门分别是正确答案与另一个错误答案，此时主持人只能打开错误的那扇门。另一方面，如果挑战者的选择正确(〇)，剩下两扇门被主持人打开的概率各为一半。综上，选择结果整理如下。

	挑战者选择门1		挑战者选择门2		挑战者选择门3	
主持人	打开门2	打开门3	打开门1	打开门3	打开门1	打开门2
门1是正确答案	〇20个会场	〇20个会场	—	×40个会场	—	×40个会场
门2是正确答案	—	×40个会场	〇20个会场	〇20个会场	×40个会场	—
门3是正确答案	×40个会场	—	×40个会场	—	〇20个会场	〇20个会场

接下来，我们从飞艇上向下眺望，数一下各种类型会场的数量。如果挑战者坚持最初的选择，共有120个标记为〇的会场选择正确，240个标记为×的会场选择错误。另一方面，如果挑战者改选了其他门，标记为×的240个会场反而将得到正确答案，标记为〇的120个会场最终选择错误。根据这一结果，我们很容易得出改选更加妥当的结论。只要统计一下各种结果的数量即可得到这一答案，不存在任何歧义。

通过飞艇视角，我们甚至可以数出之前的轻率结论究竟错了几处。挑战者选择了门3而主持人打开了门1的会场共有60个。其中，门3是正确答案的会场仅有20个。概率并非各占一半。

1.2　为什么要准备那么多会场，18个会场不就足够了吗

确实如此。不过，更多的会场能够帮助读者理解之后的内容，因此本章设定了较多的会场数量。

我们再次强调一下飞艇视角的强大之处。

- 准备了大量会场，每个会场将会同时进行游戏
- 在各个会场中，所有人仅仅是按照事先确定的剧本行动
- 只要剧本设定合理，整个广场中所有会场的结果能够完美模拟原本的事件发生概率
- 只要从飞艇上俯瞰并统计会场的数量，就能明确判断各种结论的正误

这种方式的优势十分明显。如果我们通过直觉来说明概率这种抽象的概念，往往很难把问题解释清楚。如果换用飞艇视角，抽象的情况就被转换为了具体的统计问题。

1.3 虽说是转换表述方式，其实是由概率问题转变为了确定的事件，看起来完全不同。这样的转换真的合适吗

这种想法非常合理，但我们无需担心。转换后我们仍然能够根据需要做出概率形式的解释，与原来的表述方式没有区别。例如，我们可以通过转盘来选择需要观察的会场。

我们需要准备一个写有1至360这360个号码的巨大转盘。在转动转盘后，如果停留在124号，我们就观察第124个会场。虽然我们准备了360个会场，但只会观察其中的一个。如果观察者不知道这种机制，他将怎样看待自己正在观察的会场呢?

在360个会场中，有120个会场的门1是正确答案。从观察者的角度来看，门1是正确答案的概率是$120/360 = 1/3$。门2与门3同样是$1/3$的概率。于是，实际的情况与通过掷骰子决定没有区别。

这一机制很好地模拟了原本的情况。不过，这样的解释其实仍不充分，在2.2节，我们将进一步深入讨论。

1.4 蒙提霍尔问题还有更巧妙的解释吗

有。还存在诸如“考虑有100扇门而主持人将打开其中98扇的情况”等解法。这些解释没有使用数学算式就巧妙地解答了问题。这类解法与本节这样中规中矩的说明孰优孰劣，需要具体情况具体分析。本节希望介绍一种更加通用的思考方式，不仅能解决蒙提霍尔问题，还能分析其他的情况，因此解法较为稳妥。今后，我们将根据需要选择合适的方式。

1.3 三元组(Ω, $\mathcal{F}$, P)——上帝视角

在之前的1.2节中，我们已经在飞艇上从空中俯瞰了各种会场中不同角色的行动。也许这与上帝眼中的世界无异。会场即“世界”。也就是说，如图1.3所示，上帝能够置身各种平行世界之外俯视着它们。

在图1.3中，平面上的每一点都是一个不同的世界。每个世界中都有银河系，有地球，上面生活着许多人。然而，每个世界中发生的事件并不相同。骰子在一个世界中的结果为1，在另一个世界中则可能是5。这是由于这些世界具有不同的剧本——没错，我们假设每个世界都会预先准备专用的剧本。剧本中记录了某个世界从古至今乃至未来发生的所有事件。无论是什么，一切都只是根据剧本发展而已。对于某个特定的世界来说，所有的结果都已确定，不存在任何随机事件。

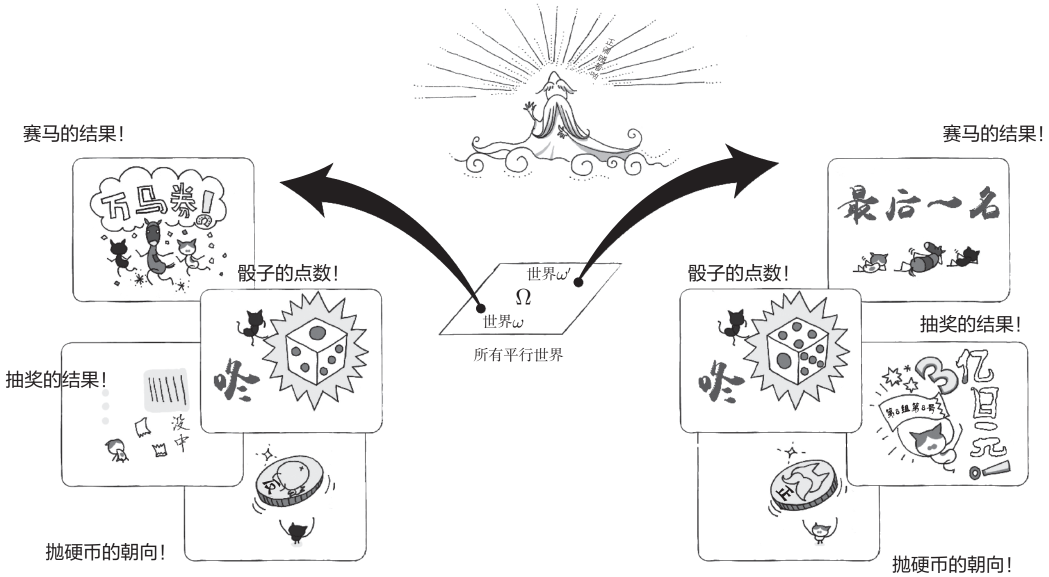


图1.3　上帝视角

那么，我们该如何讨论概率的问题呢？只需将它理解为计算会场的数量即可。不过，如果平行世界如上图的方式呈现，我们将无法计算它们的数量。因此，我们将测量该平面的面积，而非计算世界的数量。我们以计算抛硬币正面向上的概率为例来说明这个问题。为此，我们首先假定正面向上的世界（这个世界的剧本写有“硬币正面向上”这样的情节）为白色，反面向上的世界为黑色，并涂上颜色区分。图1.4是涂色的结果，该图中白色区域的面积为0.5，用于表示“正面向上的概率”。我们假定整个平面的面积为1。

现在我们来计算一下第二次抛硬币时正反两面的出现概率。依然是白色表示正面，黑色表示反面，结果白色的面积为0.5，它表示再次抛硬币时正面向上的概率。需要注意的是，黑白两色的涂色形状将发生变化。上一次结果是正面向上并不意味着这次也会有同样的结果，每次的结果不会始终相同。重叠两次的涂色情况，就能总结出两次抛硬币的结果。对了，下面这几句话至关重要，请读者用心体会。第一次与第二次的概率（区域的面积）相同。然而，这并不表示第一次与第二次抛硬币的结果是等价的，这是因为，两块区域的形状并不相同。本书将着重分析这种多个因素或现象之间的关系。如果要灵活运用现代的概率统计知识，分析现象之间的关系是一个关键。

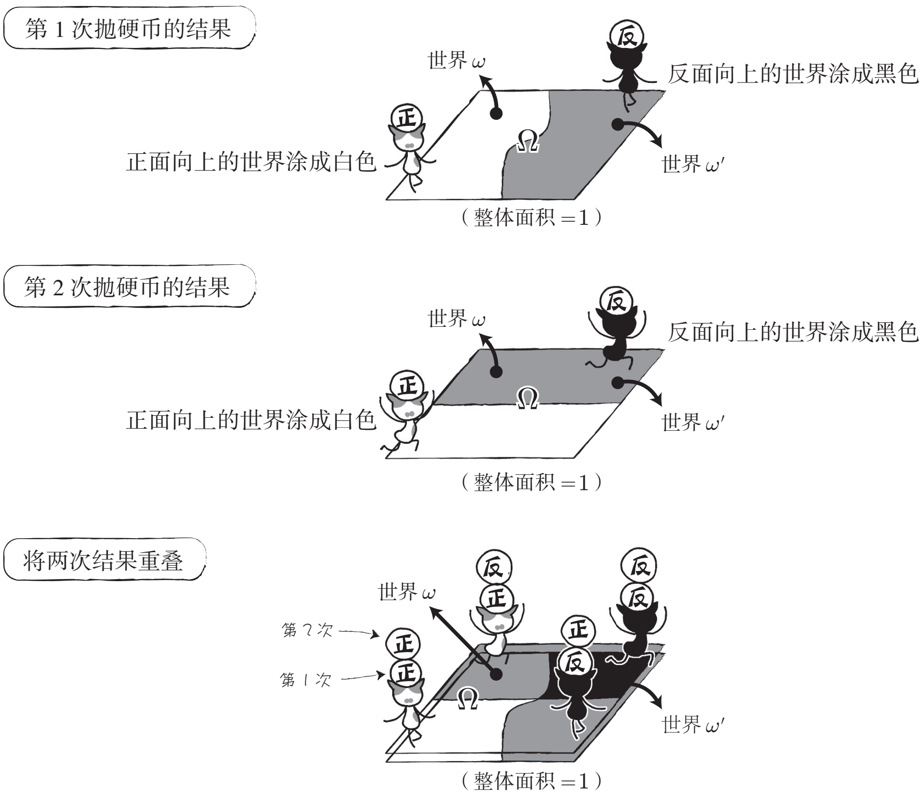


图 1.4　通过涂色区分平行世界（概率 = 面积）

这张图中有以下几条重点设定。

- 每个世界的抛硬币结果完全确定
- 然而人们无法知晓自己住在哪个世界

例如，如果第一次抛硬币的结果为正面，就能知道自己居住在图 1.5 中实线范围内的某个世界中，但无法确定是其中哪一个世界。同时，该范围内的世界在第二次抛硬币时仍会得到不同的结果。因此，人们无法确定下一次抛硬币的结果。请读者与问答专栏 1.3 作比较。

读者在初次听到这种设定时或许会感到奇怪。不过像这样明确区分确定与不确定的含义之后，接下来的讨论将更加明了。因此请读者务必努力理解这些内容①。

① 假设未来已经确定，以及上帝之类的说法，都是为了便于读者理解，与宗教之类的自然没有关系。这也许会与读者的信仰有些冲突，但请允许本书继续使用这样的说法，以更方便地理解并解决概率问题。

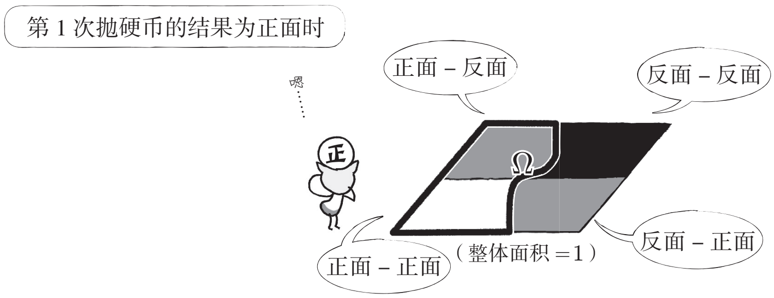


图1.5　人们无法知晓自己住在哪一个世界

至此，所有准备都已完成，终于能进入正题了。首先，我们来看一下三元组$(\Omega, \mathcal{F}, \mathrm{P})$的含义。话虽如此，其实我们已经解决了大半问题，接下来只需对已经遇到的概念标上记号即可。习惯上，我们会以希腊字母ω(Omega)来表示具体每一个世界。与之对应的所有平行世界组成的集合由大写的Ω表示。Ω中的每一个元素都是一个单独的世界。这是$(\Omega, \mathcal{F}, \mathrm{P})$的第一个元素。接下来，$\Omega$的子集$A$(如果不习惯这个词，可以理解为$\Omega$内的区域$A$)的面积将由$\mathrm{P}(A)$表示。在前面的例子中，P(白色区域)$=0.5$。用于表示面积的函数P是$(\Omega, \mathcal{F}, \mathrm{P})$的第三个元素[①]。对于所有平行世界$\mathrm{P}(A)=1$，这是前提条件。第二个元素$\mathcal{F}$不太容易理解，这里暂且跳过。

最终我们可以得到以下结论：只要知道由所有平行世界组成的集合Ω与用于测量Ω中区域面积的函数P，就能对概率问题进行讨论。借助这两个元素，概率问题被转换为了“区域与面积”的问题。这种方式的特点在于，不确定的概率问题成为了确定的数学问题。概率的概念不够直观，在将它转为面积这样含义清晰的值之后，我们就可以进一步展开明确的讨论。

（上帝视角）	完全确定的面积问题	$\longrightarrow$	面积的答案
	$\Updownarrow$		$\Updownarrow$
（人类视角）	结果不确定的概率问题	$\dashrightarrow$	概率的答案

① 这里称P为函数，其实它与大家熟悉的“输入数字后返回数字”的函数有些不同，是一种“输入集合后返回数字”的逻辑。

1.5 为什么非要使用平行世界这样夸张的说法？会场与飞艇的比喻更加容易理解嘛

之所以这么做，是为了准备无限多的备选情况。例如，飞行距离与捕鱼量（重量）都是连续量，我们无法逐个数出这些值所有的可能情况。此时，无论我们准备多少会场，都无法对应所有的备选情况，因而无法模拟原本的概率。此外，如果概率是 $1/\sqrt{2}$ 这样的无理数值（无法以整数 / 整数的形式表示的值），即使我们准备再多的会场，也不能准确表现这一概率。因此，我们必须改用更加通用的形式。

1.6 我不太理解由所有平行世界组成的集合 Ω 是什么概念。请再解释一下吧

读者只需暂且按照插图示例，将其理解为面积为 1 的正方形即可。1.7 节将作更详细的解答。

1.7 感觉我们是不是把问题过分简化了？去除了一些关键的要素后，还有讨论的价值吗

这不正是数学常用的解决方式吗？

例如，行星探测器与星体的距离时刻都在变化，如果我们以时间为横轴，以距离为纵轴作图，就能通过一张静态图像表示探测器从过去到未来的运动情况。于是，时间这种特殊概念，也能转换为数学形式，借助普通的数字与函数表述。

本节的内容也是如此。概率存在多种可能，如果我们以 $(\Omega, \mathcal{F}, \mathrm{P})$ 表述概率，就能以一种静态的方式呈现所有的可能情况。于是，概率这种特殊概念就像这样转换为了数学形式，能通过普通的集合、数字与函数表述。

顺便说明一下，每个世界 ω 称为样本，由所有平行世界组成的集合 Ω 称为样本空间，Ω 的子集称为事件①。不过，这些术语过于正式，本书不会使用。

1.4 随机变量

1.3 节已经构建了舞台。接下来，我们讨论一下如何在这一舞台上表现随机量，并整理之前出现过却还没有详细讲解的内容。

这种随机量称为随机变量。通俗来讲，它是一种会随机改变的不确定量。不过，如图 1.6

① 有时，ω 也称为样本点或基本事件，Ω 也称为基本空间。事实上，事件的定义与 $\mathcal{F}$ 有关，有一定的限制条件，不过在此我们不做进一步讨论。

所示，我们现在处于上帝视角。该视角下，我们可以俯瞰由“所有平行世界”组成的集合 Ω。其中每个世界 ω 中的事件都已确定，只是根据剧本表演而已。

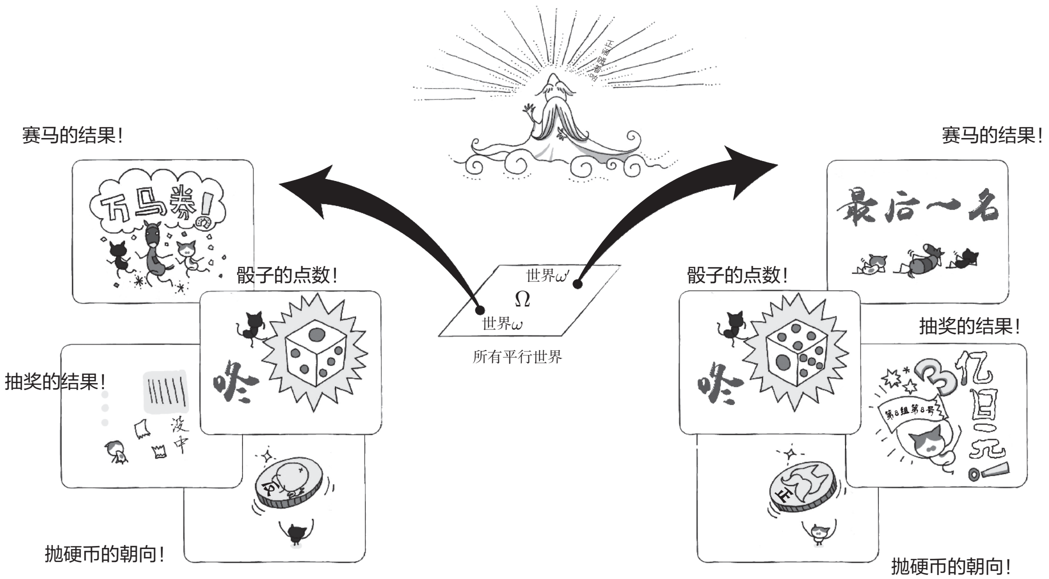


图1.6　上帝视角（与图1.3相同）

从 $(\Omega, \mathcal{F}, \mathrm{P})$ 的角度来看，随机变量只是 Ω 中的函数而已。请读者观察图1.7。如图所示，对于 Ω 中的各元素 ω，函数 $f(\omega)$ 将返回相应的整数，这些整数值就是随机变量①。f 只是一个函数，返回情况完全确定。除了定义于 Ω 中，它与诸如 $g(x) = x + 3$ 之类的函数没有区别。不过，如果人们无法知道自己处于哪个世界 ω，就不能确定 $f(\omega)$ 的值。这是因为 $f(\omega)$ 的值根据 ω 而变化，而此时 ω 无法确定。读者能够理解这里的讲解吗？

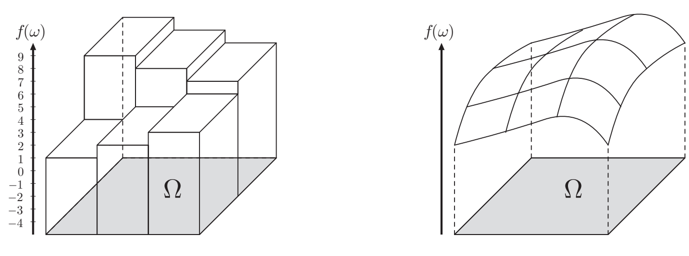


图1.7　随机变量是 Ω 中的函数。左侧是整数值（离散值）的随机变量。右侧是实数值（连续值）的随机变量

① “返回值”是程序设计中的术语。也就是说，函数 f 在以参数 ω 调用后将执行一些处理，并将结果返回给调用方。

这一点非常重要，因此我们需要再详细说明一下。请读者在阅读时思考此时是否随机。从上帝视角来看，随机变量只不过是一种完全确定的函数 $f(\omega)$。但从人类的视角来看，由于不确定自己身处哪一个世界 ω，因此也无法得知 $f(\omega)$ 的值。如果读者能够在两种视角之间自由切换，接下来的概率问题讨论应该不会有太大问题。

（上帝视角）	完全确定的面积问题	$\longrightarrow$	面积的答案
	$\Updownarrow$		$\Updownarrow$
（人类视角）	结果不确定的概率问题	$\dashrightarrow$	概率的答案

此外，习惯上，我们会通过 $X(\omega)$ 这样的大写字母来表示随机变量。因此在后文中，我们将尽可能使用大写字母。

接下来，我们来看一些例子。

如图1.8所示，Ω 是一个正方形。从集合的角度来讲，它是一个由0至1间的实数组成的二元组集合。也就是说，Ω 中的元素呈 $\omega=(u, v)$ 的形式（$0\leqslant u\leqslant 1$ 且 $0\leqslant v\leqslant 1$）。P是传统意义上的面积。整个 Ω 的面积为1，完全符合要求。

我们将按照图1.9来定义随机变量 X①。

$$X(u,v)\equiv\begin{cases}\text{中选} & (0\leqslant v<1/4)\\ \text{落空} & (1/4\leqslant v\leqslant 1)\end{cases} \tag{1.1}$$

随机变量 X 的值是中选与落空这两种选项之一。这类既非整数又非实数的值也能作为随机变量定义。例如，对于 $\omega=(0.3,0.5)$，则 $X(0.3,0.5)=$ 落空。而对于 $\omega=(0.2,0.1)$，则 $X(0.2,0.1)=$ 中选。那么，X 的值为中选的概率是多少呢？由于 $X=$ 中选对应的面积是1/4，因此 X 有1/4的概率值为中选。

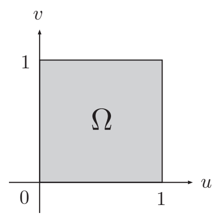


图1.8　由所有平行世界组成的集合

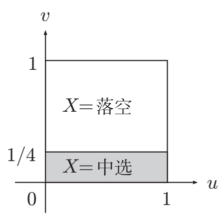


图1.9　随机变量 X

① 该式中的 $\equiv$ 表示定义。由于这是一种定义，因此只是命名了随机变量而已。有些书也会使用 $:=$ 或 $\triangleq$ 这样的记号。此外，在其他领域中，$\equiv$ 可能具有不同的含义。

为帮助理解，我们再来看一下另一个随机变量的例子。

$$Y(u,v) \equiv \begin{cases} \text{中选} & (2u+v \leqslant 1) \\ \text{落空} & (\text{其他}) \end{cases} \tag{1.2}$$

由图1.10可知，Y有1/4的概率中选，3/4的概率落空。

最后，我们再看一个实数值的随机变量。

$$Z(u,v) \equiv 20(u-v) \tag{1.3}$$

请问：

- Z的取值范围是什么？
- Z大于等于0且小于等于10的概率是多少？

对于第一个问题，根据u与v的范围可知，Z可以是-20至20之间的实数。对于第二个问题，如图1.11所示，当$Z(u,v) \leqslant 10$时，(u,v)表示的面积为$3/8(=1/2-1/8)$，因此概率为3.8。

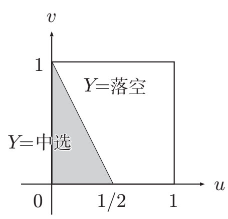


图1.10　随机变量Y

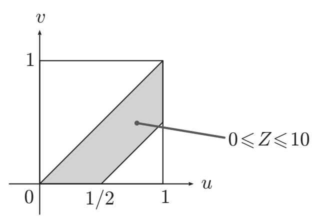


图1.11　随机变量Z

上面的例子暂时只写了随机变量的表达式，作为练习，请读者用平行世界及剧本等词复述这些例子。请闭眼思考“在上帝视角下情况将会如何？”以及“从普通人的视角来看，情况将会如何？”这两个问题。

需要读者多加注意的是，虽然X与Y都是1/4的概率中选，3/4的概率落空，但X与Y本身并不相同。如何通过平行世界与剧本等概念来解释两者，也是需要读者思考的问题。

对于以整数或实数为对象的随机变量X与Y，我们只要通过直观的方式来理解$X+1$、$3X$、$X+Y$，或XY等表达式的含义即可。从普通人的视角看来，“随机值X”加上1后得到的就是$X+1$。从上帝视角来看，$X+1$表示的是各个世界中值为$X(\omega)+1$的函数。

1.8 前文图中的那些与表达式对应的区域是怎样画出来的呢

要理解区域边界的方程式并不难。例如，对于方程式“$Y(u,v)=$ 中选”，我们实际上是要求 $2u+v\leqslant 1$ 表示的区域。将表示该区域边界的方程式 $2u+v=1$ 变形后，我们将得到 $v=-2u+1$。该方程式表示一条直线，其斜率为 -2，y 轴截距为 1。我们首先画出这条直线，之后判断该区域位于直线的哪一侧。我们只需随意取一个点检查即可轻松得出结论。例如，代入 $(u,v)=(0,0)$ 能得到 $2u+v=0$，因此点 $(0,0)$ 位于 $2u+v\leqslant 1$ 表示的区域中。而代入 $(u,v)=(1,0)$ 将得到 $2u+v=2$，可知点 $(1,0)$ 不属于该区域。于是，该区域与点 $(u,v)=(0,0)$ 位于直线的同侧。（可能会有读者指出这里悄然利用了连续性的原理，不过能了解这些概念的话，应该也就不会提出这个问题了吧。）

顺便一提，随机变量在英语中称为random variable。板书中有时会缩写为r.v.，以防读者不理解改缩写的含义，这里特别提醒一下。

1.5 概率分布

随机变量涉及具体的平行世界。与之相对地，概率分布的概念更为宽泛，它只考虑面积，不涉及具体的平行世界。在不会产生歧义时，我们可以将概率分布简称为分布。简单来讲，下表表示的就是一种概率分布。

表 1.1 老千骰子的概率分布

骰子的点数	掷出该点的概率
1	0.4
2	0.1
3	0.1
4	0.1
5	0.1
6	0.2

我们也可以用图 1.12 的形式来表示概率分布。这种方式也许更易于理解。

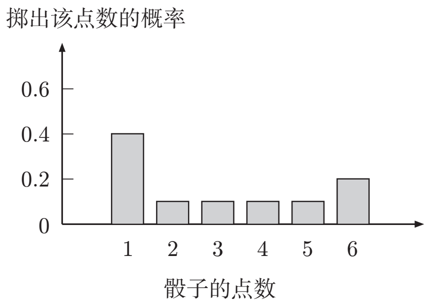


图1.12　老千骰子的概率分布图

两种表述方式没有本质区别，我们需要注意的是概率分布与随机变量之间的差异。对于随机变量，哪一个世界中将得到哪一个值都已确定，而概率分布不涉及事件具体发生在哪一个世界。

只要知道随机变量，我们就能计算相应的概率分布。例如，表1.2与表1.3分别是之前式1.1与式1.2中 X 与 Y 的概率分布。

表1.2　X 的概率分布

值	该值出现的概率
中选	1/4
落空	3/4

表1.3　Y 的概率分布

值	该值出现的概率
中选	1/4
落空	3/4

虽然 X 与 Y 是不同的随机变量，但是它们的概率分布相同。我们可以用下面的表达式来表述概率分布表。

$$\begin{cases} \mathrm{P}(X = \text{中选}) = 1/4 \\ \mathrm{P}(X = \text{落空}) = 3/4 \end{cases}$$

下面的写法含义相同。

$$\mathrm{P}(X = k) = \begin{cases} 1/4 & (k = \text{中选}) \\ 3/4 & (k = \text{落空}) \end{cases}$$

两种写法都表示 $X =$ 中选的概率为1/4，落空的概率为3/4。

对于更为一般的随机变量 X，其概率分布的求法应该也非常直接。下面的表达式表示“X 为 k 的概率”，因此这类方程式的集合就是相应的概率分布。

$$\mathrm{P}(X=k)=\text{“}X(\omega)=k\text{时区域}\ \omega\ \text{的面积”}$$

只要得到随机变量 X，我们就能求出相应的概率分布，但反过来却不成立，仅凭概率分布，我们无法求出随机变量的值。之后这句提醒也将反复出现，因此读者不必担心会遗忘。

我们需注意概率分布的两条性质。

- 每一项概率都大于等于0且小于等于1
- 所有概率的和必定为1

从上面采用的面积的角度来看，这是理所应当的两条性质(在集合 Ω 面积为1的前提下)。

不过，即使没有列出所有的公式，我们只要知道了所有值的出现概率，就能表述概率分布。例如，下面的数学表达式也能表述概率分布。

$$\mathrm{P}(X=k)=\frac{12}{25k}\quad(\text{其中}\ k=1,2,3,4) \tag{1.4}$$

此时，与其说“式1.4是随机变量 X 的概率分布”，不如说“随机变量 X 遵从式1.4中的概率分布”。

顺便一提，我们对式1.3中那样的值连续的随机变量 Z 也定义了概率分布。不过这种概率分布的表述略有些复杂，第4章我们再详细讨论。

1.6　适于实际使用的简记方式

通过方程式表述概率对格式的要求较高，如果读者不习惯这种写法，在书写时或许会有些困难。因此，我们经常会使用一些便于书写的简记。本节将介绍一些常用的简记方式。熟悉之后，阅读简记也并非难事，不过对初学者来说，这些简记可能会引起混乱。因此，如果读者在阅读时感到有些困惑，即使老师也使用简记表述概率，也请先改用原本的表示方法以避免误解。

1.6.1　随机变量的表示方法

原则上，随机变量应以大写字母表示。在使用 $X(\omega)=a$ 这种写法时，我们可以清楚地区分出 X 是随机变量，而 a 只是普通的数字。不过，我们并不一定非要遵守该原则。如果情况比较复杂，字母将不够使用，此时使用其他名称也无可厚非。

话虽如此，但若使用 $X(\omega)$ 或 X 之类过分简单的记法，阅读者可能会产生误解。如果使

用 $X=a$ 的写法，其中又含有“不确定的随机数 X 可以取值 a”的含义，是一种基于普通人视角的表述方式。请读者在必要时改用上帝视角，以 $X(\omega)=a$ 的形式表述，即“在点 ω 处，函数 X 的值为 a”①。

有时我们也可以图省事，同时用字母 X 表示随机变量与普通的数字。随机变量是“Ω 上的函数”，而后者只是前者“可取的值”，两者并不相同，不过我们可以不作特别区分，都以 X 表示。读者必须根据上下文判断该字母究竟表示哪种含义。在学习概率知识时，我们应尽可能避免采用这种表述方式。

1.9　26个英文字母仍然不够用吗

不够用，26个英文字母很容易就会用完。我们不能任意使用这些字母。

许多字母已经存在习惯用法，这些字母被赋予的特定含义有助于数学表达式的理解，人们看到这些字母时，能够下意识读懂它们的含义。但另一方面，由于存在这种习惯，我们无法随意使用所有的字母。例如，对于“函数 $f(x)=3x+1$”与“函数 $x(f)=3f+1$”，虽然它们除了名称不同外并没有实际的区别，但读者会感觉第二种写法有违常规。

尽管没有明确规定，我们仍建议读者遵循习惯用法，以便阅读。应该遵循习惯还是坚持使用大写字母表示随机变量，需要具体情况具体分析。(虽然与随机无关)集合与行列式也常会使用大写字母，请读者注意区分使用，不要发生混淆。

1.6.2　概率的表示方法

本书将三元组 $(\Omega, \mathcal{F}, \mathrm{P})$ 中 P 所对应的面积定义为概率。根据该定义，我们只能使用“P(Ω 的子集)”这种表示方法来表示概率。不过，我们也能根据需要使用形如“P(条件)”这样的表示方法。例如，$\mathrm{P}(2\leqslant X\leqslant 7)$ 表示概率 $\mathrm{P}(A)$，其中 A 是“由所有满足 $2\leqslant X(\omega)\leqslant 7$ 的 ω 组成的集合”。同理，对于 $\mathrm{P}(X=3)$ 表示的概率 $\mathrm{P}(A)$，集合 A 是“由所有满足 $X(\omega)=3$ 的 ω 组成的集合”。

不过，这种表示方法也有不足，如遇到 $\mathrm{P}(X=3)=0.2$ 的情况，连续出现的 $=$ 不太美观。为避免这种问题，我们有时也会将 $\mathrm{P}(X=3)$ 改写为 $P_X(3)$ 的形式②。如果不会产生歧义，我们还可以进一步省略，只写出 $P(3)$。

问题在于，同时存在两个随机变量 X 与 Y 时，我们应该怎样区分表示它们各自的概率分布。$\mathrm{P}(X=3)$ 或 $P_X(3)$ 的写法虽然稳妥，但如果数量过多也容易看花眼，书写也比较麻烦。

---

① $X=a$ 的表述方式存在歧义，因此如果没有上下文，读者可能会产生误解。
- 理工科的学生一般会将它理解为随机变量。本书今后也会采用这种表示方法
- 数学系学生可能会认为“对于任意 ω，随机变量的值总是为 a”。从普通人的视角来看，这一随机变量不再随机，只有 a 这一种可能值(这类值恒定的变量同样属于一种随机变量)

② 有些书中，$\mathrm{P}_A(B)$ 这样的表示方法可能用于表示不同的含义(在事件 A 发生时事件 B 的条件概率，详见2.3节)。

换用其他的字母，以诸如“将 X 的概率分布记为 P，Y 的概率分布记为 Q” 的形式来表述概率分布是一种较为妥当的做法。此时，$P(3)$ 表示 $\mathrm{P}(X=3)$，$Q(3)$ 表示 $\mathrm{P}(Y=3)$。这种方式的优点是不必使用下标，书写更加方便，但也有需要引入新字母的不足。

在实际使用中，我们有时还会使用更加简略的表示方法。这些表示方法容易产生歧义，因此不建议初学者使用。不过如果不知道这些写法，读者可能无法理解某些书采用的简记方式的含义，为此，本书将对它们作简单介绍。那就是可以将 $\mathrm{P}(X=x)$ 简记为 $\mathrm{P}(x)$，$\mathrm{P}(Y=y)$ 则可以简记为 $\mathrm{P}(y)$。这种表示方法存在歧义，对于没有上下文的 $\mathrm{P}(3)$，我们将无法辨别它的具体含义。

本书有时还会使用 $\mathbf{Pr}(\cdots)$ 或 $\mathbf{Prob}(\cdots)$ 等写法代替 $\mathbf{P}$。它们的含义相同。

1.7 Ω是幕后角色

为了描述由所有平行世界组成的集合，我们引入了字母 Ω 来表示这种夸张的说法。不过，它基本上是一个幕后角色。本节将解释 Ω 为何不太会出现在前台大出风头的原因。

1.7.1 不必在意Ω究竟是什么

在前几节中，我们将 Ω 定义为面积是1的正方形。不过，其实即使我们将 Ω 的定义改为其他形状，概率论的原理也不会发生变化。例如，如图1.13所示，我们可以将 Ω 定义为面积为1的圆盘。之所以这种改变不会产生影响，是由于我们不关心 Ω 及其中区域 A 的形状。我们只对面积 $\mathrm{P}(A)$ 感兴趣。只要 $\mathrm{P}(A)$ 的值符合预期，无论 Ω 是圆是方，抑或是更古怪的形状，都不会妨碍我们讨论概率问题。

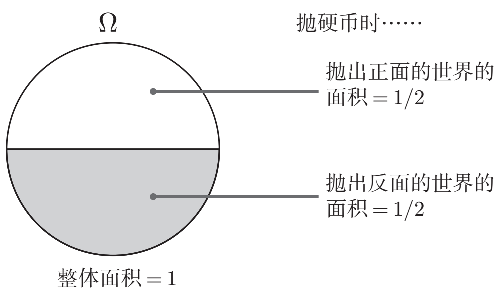


图1.13 由所有平行世界组成的集合 Ω 即使是圆形，相关的概率也不会发生变化

我们甚至无需在意概率是否以面积表示。例如，如图1.14所示，即使我们将Ω定义为体积为1的立方体，只要将P视作相应的体积问题就没有不同。该图右侧将Ω定义为了球体，情况依然如此。因此，只要将P定义为与面积和体积具有相同性质的概念即可。不过，我们仍须遵守P(Ω)=1的前提条件。

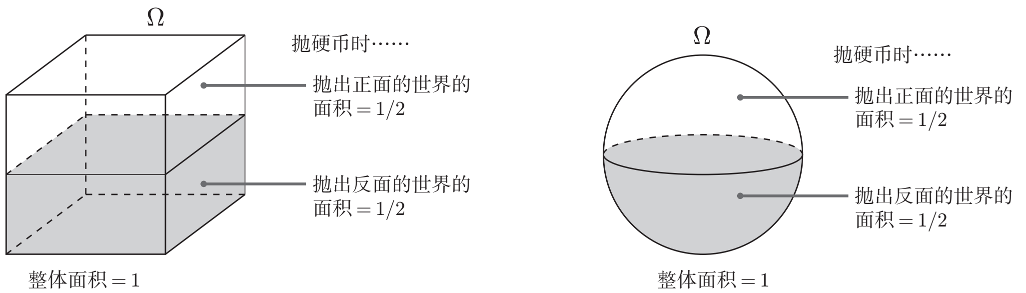


图1.14　由所有平行世界组成的集合Ω即使是立方体或球体，相关的概率也不会发生变化

请读者注意，虽说面积之外的方式也能够表现概率，但本书今后仍将在插图中使用正方形，并以面积描述概率问题。

1.7.2　Ω的习惯处理方式

以上就是概率的数学定义。我们已经知道，概率仅与Ω中区域的面积(也可用测度这一更加普遍的概念)有关。我们只需关注面积的值，而不必在意Ω本身。

不过，在通过这种方式解决实际问题时，存在两种不同的处理方式。

第一种方式与我们迄今为止的做法相同，将使用宽泛的Ω概念。也就是说，我们将准备一个通用的巨大舞台Ω，并将每个问题设定为其中不同的随机变量。如果采用上帝视角，通过平行世界的方式来解释，这种处理方式非常合理。原则上，本书将采用这种方式。

另一种方式将针对每个具体问题设定更小的集合Ω。一些内容较浅的概率论书籍常会采用这种方式。例如，试考虑投掷一次硬币的情况，我们只需将Ω设定为仅有两个元素的集合{正面, 反面}即可[①]。1.2节中的蒙提霍尔问题总共只有$3\times3\times3=27$种不同的结果组合(正确的门、挑战者选择的门以及主持人打开的门)，因此我们只要定义一个含有27个元素的集合就能解决问题(其中有些组合不可能实际出现，只需将它们的P定义0为即可)。

---

① 如果按下面的方式定义，采用面积表述或体积表述的结果相同。其中$\emptyset$表示空集{}。
$\mathrm{P}(\emptyset)=0$, P({正面})$=1/2$, P({反面})$=1/2$, P({正面,反面})$=\mathrm{P}(\Omega)=1$

$$\Omega = \{(1,1,1), (1,1,2), (1,1,3), (1,2,1), \ldots ,(3,3,3)\}$$

1.10 我读了一本专门论述概率论的教材，里面没有明确定义 Ω 的含义。还真是不专业啊

不，采用第一种方式的教材通常都不会专门定义 Ω 。

Ω 终究仅是为了存在而存在，只要能够模拟希望讨论的问题，我们不关心 Ω 的具体形式。因此，我们没有必要对 Ω 作具体的定义。

基于这一理由，在讨论概率论问题时，我们无需每次都介绍 Ω 的具体含义，最多提一下我们将用到 Ω 这样一个集合即可，之后无需作太多说明。

1.7.3 不含Ω（不含上帝视角）的概率论

本书在前面的讲解中着重强调了上帝视角。

- 概率是在全体平行世界的集合 Ω 中划分出的区域面积
- 随机变量仅仅是 Ω 上的函数

由于概率分布只是一列概率一览表，因此无需引入上帝视角也能解释。不过这种方式也有不足，随机变量的概念很可能会变得无足轻重。尽管我们也能通过语言描述这一概念，但却无法跳出常规视角，难以像上帝视角那样以数学的方式明晰地讨论概率问题。

如果问题仅涉及一个随机变量，这并不是什么大问题。如果同时存在多个随机变量，问题就会变得有些棘手。请读者回忆我们之前反复提醒的一点，即“即使概率分布相同，相对应的随机变量也可能不同”。当问题较为复杂时，常规视角下的语言和直觉将难以解决问题，上帝视角更为有效。并且，对这类复杂随机变量组合的分析也是现代概率统计学中的一个关键。

1.8 一些注意事项

1.8.1 想做什么

仅凭语言难以讨论概率问题，很容易陷入双方各执一词的无休止的争论中。如果只用语言表述，讨论可能会莫名其妙地转化为哲学问题，不知不觉中做出错误的判断。

因此，本书试图一同介绍纯粹的数学定义，并在后文中结合数学来讨论问题。为此，我

们必须讲概率这种抽象概念转化为集合、数字与函数等已有的数学概念，使读者可以把握问题的本质。

以上就是本章的目的。

1.8.2　因为是面积……

在强调了“概率就是面积”这种基于上帝视角的观点后，很多问题将会不言自明。下面就是一个例子。不过从本书的角度来讲，这个例子可能有些画蛇添足。

不过，随着概率学习的不断深入，在接触到各种各样的问题后，读者可能会忘记“概率即面积”这一要点。以防万一，本书专门辟出一节强调概率的面积定义。下文中的 ○○ 与 ×× 表示任意的“条件”。

首先我们来看一下概率的取值范围。如图1.15所示，由于概率是一种面积，因此它不可能为负。此外，由于整体的面积为1，因此概率的最大值是1。

$$0 \leqslant \mathrm{P}(\text{○○}) \leqslant 1$$

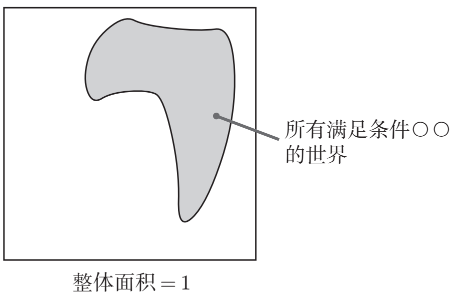


图1.15　概率的取值范围

并且，“满足 ○○ 条件时的概率”与“不符合条件 ○○ 的概率”的和应为1。

$$\mathrm{P}(\text{○○}) + \mathrm{P}(\text{非 ○○}) = 1$$

例如，

$$\mathrm{P}(X = 3) + \mathrm{P}(X \neq 3) = 1$$
$$\mathrm{P}(X < 3) + \mathrm{P}(X \geqslant 3) = 1$$

两者的含义相同，请读者习惯这种表述方式。

$$\mathrm{P}(\text{非 ○○}) = 1 - \mathrm{P}(\text{○○})$$
$$\mathrm{P}(X \neq 3) = 1 - \mathrm{P}(X = 3)$$
$$\mathrm{P}(X \geqslant 3) = 1 - \mathrm{P}(X < 3)$$

无论采用哪种方式，只要读者可以想象出图1.16中的画面，问题就会一目了然。

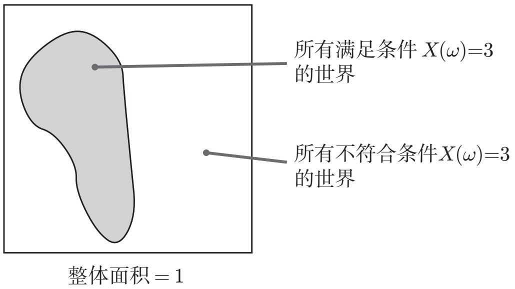


图1.16　不符合 ○○ 条件时的概率

请读者注意以下问题。如果 ○○ 与 ×× 不可能同时发生，“ ○○ 或 ××”的概率就是两者概率之和。

$$\mathrm{P}(\text{○○或××}) = \mathrm{P}(\text{○○}) + \mathrm{P}(\text{××})$$

该式之所以正确是因为，如果这两个条件不可能同时成立，两者表示的区域就不会存在重叠。图1.17是一个例子，我们可以很容易从中看出这一性质。

$$\mathrm{P}(X = 3 \text{ 或 } X = 7) = \mathrm{P}(X = 3) + \mathrm{P}(X = 7)$$
$$\mathrm{P}(X < 3 \text{ 或 } X > 7) = \mathrm{P}(X < 3) + \mathrm{P}(X > 7)$$

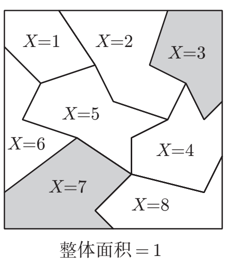


图1.17　“$X = 3$或 $X = 7$”的概率

希望读者注意的是，该性质必须满足所有条件不会同时发生的前提。如果该前提不成立，我们就不能通过简单的加法来计算概率。例如，我们假设骰子的点数为 X。此时 $\mathrm{P}(X$ 为偶数，或 X 能被3整除$)$ 并不等于 $\mathrm{P}(X$ 为偶数$)+\mathrm{P}(X$ 能被3整除$)$。如图1.18所示，加法将重复计算两者面积的重叠部分。

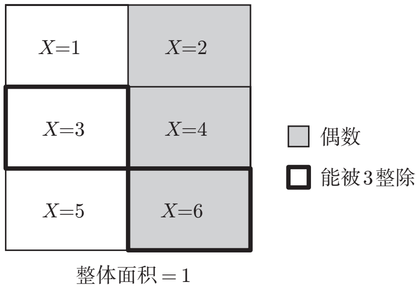


图1.18 如果面积重叠

1.8.3 解释

最后笔者需要作一些解释。本书的目标是教授一些有一定深度的概率统计知识。笔者假设读者并不希望成为专业的数学家，因此反复思考应当优先讲解哪些内容，最终决定采用以下方针。

- 不涉及测度理论及概率论公理
- 不涉及面积的定义
- 假定面积的性质与日常生活中的面积相同

本书并没有使用严谨的数学定义。如果读者感觉本书非常专业，或许是因为还没有学习过真正的数学（除了数学系的专业，很多人直至大学毕业可能都没有接触过真正的数学）。

事实上，不对概念作明确定义就讨论其性质本身并没有意义。如果没有严格定义面积就讨论面积问题，我们很难获得严密明晰的结论。本书最终决定采用这种讨巧的做法。不过另一方面，笔者也承认本书介绍的数学并非真正的数学。

如果读者希望学习真正的能登大雅之堂的数学，请再另外学习一下测度论与概率论的公理。在查找教材时，可以尝试搜索勒贝格积分等术语。

专栏　蒙提霍尔问题的模拟

图1.19试图借助计算机，以伪随机数组模拟蒙提霍尔问题①。在执行一万次坚持最初选择的方案后，中选率约为33%，落空率大约67%。而如果执行另一种方案，改选另一扇门，则中选率约为67%，落空率约为33%。根据该结果可知，改选其他的门的胜率的确更高。

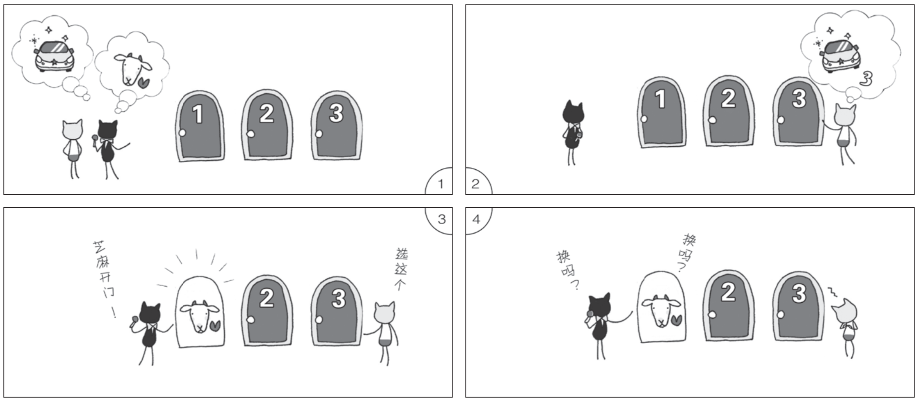


图1.19　蒙提霍尔问题（与图1.1相同）详情请参见1.2节

```
$ cd monty ↲
$ make long ↲
(no change)
./monty.rb 10000 | ../count.rb
O: 3303 (33.03%)
X: 6697 (66.97%)
(change)
./monty.rb -c 10000 | ../count.rb
O: 6674 (66.74%)
X: 3326 (33.26%)
```

① 关于程序的获取方式，请读者参见前言中的在线资源部分。在下面的执行示例中，$之后的粗体字部分需由用户输入。$本身不必输入。第7章讲解释伪随机数组的定义以及通过伪随机数组模拟问题的意义。关键字是蒙特卡罗方法（7.1.3节）。