理性派

身边的数学译丛

生活中的概率趣事

[瑞典] 彼得·欧佛森 (Peter Olofsson)◎著　　赵莹◎译

WILEY

机械工业出版社
CHINA MACHINE PRESS

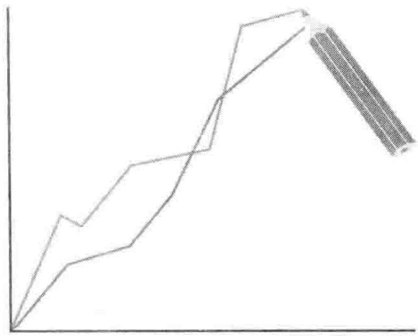
身边的数学译丛

Probabilities:The Little Numbers That Rule Our Lives

生活中的概率趣事

[瑞典] 彼得·欧佛森 (Peter Olofsson) ◎著

赵莹 ◎译



机 械 工 业 出 版 社

这是一本内容丰富且可读性很强的科普书，作者言简意赅地为读者描绘了一个神秘的概率世界，书中避免了冗长的数学推导和复杂的公式，取而代之以妙趣横生的例子，为读者展示了概率在日常生活中所起的作用，这些例子在具备娱乐性的同时又富有代表性。比方说，其中有一些是我们生活中不易察觉但与概率密切相关的例子，如：生日问题，购物的最优策略，等车时间问题等，此外，还有一些违反直觉的例子，如：蒙提霍尔悖论、辛普森悖论、决斗的策略等。同时书中也介绍了许多概率统计的应用及其原理产生的背景，如：贝叶斯法则在医疗诊断中或法庭断案中能提供的帮助等。

本书既适合学生增加学习兴趣，又适合教师作为教学参考。同时，数学爱好者以及概率统计应用的科技人员也能从中获益。

北京市版权局著作权合同登记 图字01-2013-1400号

图书在版编目（CIP）数据

生活中的概率趣事/（瑞典）欧佛森（Olofsson，P.）著；赵莹译. —北京：机械工业出版社，2014.8
（身边的数学译丛）
ISBN 978-7-111-46263-7

Ⅰ.①生… Ⅱ.①欧… ②赵… Ⅲ.①概率 Ⅳ.①O211.1

中国版本图书馆CIP数据核字（2014）第061343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 任正一 韩效杰
责任校对：薛 娜 封面设计：路恩中
责任印制：李 洋
北京市四季青双青印刷厂印刷
2014年8月第1版第1次印刷
170mm×230mm · 14.5印张 · 233千字

标准书号：ISBN 978-7-111-46263-7
定价：38.00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
社服务中心：(010)88361066	教材网：http://www.cmpedu.com
销售一部：(010)68326294	机工官网：http://www.cmpbook.com
销售二部：(010)88379649	机工官博：http://weibo.com/cmp1952
读者购书热线：(010)88379203	**封面无防伪标均为盗版**

你是否**留意**过生活中**众多事件**背后的**秘密？**

在**概率**的世界里**挑战**你的**直觉**。

译者的话

与大多数科普读物一样，本书的作者用最朴素的语言为读者描述了概率这样一个深奥复杂的世界。同时用化繁为简，深入浅出的写作手法、平易的叙述方式将每一个概率理论娓娓道来，译者在翻译过程中深深地感受到了作者的用心。

俗话说："没有规矩不成方圆。"世界按照自然规律在运行，只有当你懂得了这些规律，掌握并学会运用这些规律时，你的生活才能更加舒适便利。本书最大的特色在于作者在每一章节之中都会巧妙地把概率的理论融入生活的点点滴滴中，让读者切身体会到概率这一小小的数字贯穿于我们生活的方方面面，无处不发挥它的作用。希望读者通过阅读本书，学以致用，将概率知识运用到日常生活中去。

特别感谢中国人民大学统计学院王星老师对本书稿进行了审校。由于译者水平有限，译文错误及不妥之处请读者批评指正。

前　言

我们的生活与小数字们如影相随，本书是有关它们的故事。试着回想你最近一次听到“概率”、“机会”、“胜算”、“随机性”、“风险”或者“不确定性”这些词是什么时候？想必不会是很久以前的事吧。在这本书中，我将要向读者讲述关于这些概念的原理以及如何运用它们去更好地了解我们所在的社会。此书并不是一本教材，所以它既没有定义、定理、也没有习题。我写这本书的主要目的是寓教于乐，你当然也可以从中汲取些许知识。书中偶尔会有一些小练习，但这些练习都已经巧妙地贯穿于正文当中，也许你自己都没有意识到已经完成了这些练习。

首先我要对我的夫人致以衷心的感谢。感谢 Αλχμήνη 利用闲暇时间提供各种生活素材协助我完成作品，尤其是书中关于希腊词汇的运用技巧以及我早年那些难忘的旅行经历，读者可以在书中读到这些小故事。同时也非常感谢瑞典哥德堡查尔姆斯理工大学 Olle Häggström 教授的帮助。他通读了整篇手稿，并且做出了许多深刻、准确又客观的评论。如果读者在本书中看到一些觉得非常愚蠢的话语，很有可能 Olle Häggström 教授审稿时已经指出过，但我还是固执地保持了原样。感谢苏赛克斯大学的 John Haigh 和威立出版社的 Steve Quigley，Kris Parrish 和 Susanne Steitz，还有许多其他匿名的评论意见。还要感谢 Sheree Van Vreede 出版服务处专业的编辑工作，以及特克斯技术公司的 Amy Hendrickson 对我遇到的技术问题进行耐心迅速地解答。

本书的主要内容是在 2005 年那个纷乱的秋天完成的。我们在当年八月上旬从休斯敦搬到了新奥尔良，这个时间非常不凑巧，因为三周之后卡特里娜飓风就袭击了这里。我们不得不又搬回了休斯敦，但随后飓风丽塔来袭，于是我们只能在德克萨斯州西部与新墨西哥州交界的沙漠中避难。相比于飓风，沙尘暴真是客气许多！2006 年一月我们终于搬回了新奥尔良，这个城市非常

的美，铁架烤牡蛎更是代表性的珍馐美味。我非常感谢那些在这个秋天给我们提供住宿，并在各方面给予我们帮助的好心人。正是由于你们热情的帮助，本书才能付梓。感谢休斯敦莱斯大学统计系的凯西恩索公司和德克萨斯城的大陆学院汤姆英语公司为我提供了办公地点。最后，感谢我的博士论文导师查尔姆斯理工大学的 Peter Jagers 教授，从一开始写这本书时就一直给予我指导。

彼得·欧佛森

目 录

第 1 章

计算可能性：算对了还是算错了

概率影响着我们生活中的方方面面，本章介绍了概率论的基本原理和法则，并用其解释空难是否会再次发生、抽奖是否公平等实际问题。同时概率也是表达不确定性的一门艺术，下面请跟随作者来一探概率世界的奥妙。

1.1 关于概率学家

不管你是否愿意承认，概率的确主宰着我们的生活。如果你曾经过着赌徒式的生活，那你必然已经痛苦地意识到了这一点。而对于我们这些过着平凡生活的人来说，概率也会时不时地影响着我们。这些影响随处可见，如确定保险费用、新药物的引进、民意调查、天气预报和法庭上出示的 DNA 证据。不仅如此，概率还关系到我们每一个人。你的父亲遗传给你的是 X 染色体还是 Y 染色体？你遗传了祖母的大鼻子吗？从更专业的角度来看，量子物理学家告诉我们世间万物都是由概率来刻画的。他们整天都在研究薛定谔波动方程、海森堡不确定性原理，这些术语对我们来说艰深晦涩，但是至少从中可以得出一个结论：物理学的基本定律是在概率的基础上讨论的。事实上，牛顿物理定律也要归功于概率论。在日常生活中我们常常会说“对这件事，我 99% 地确定”，“只有百万分之一的可能”。当发生了不寻常的事情时，我们就会反问道“这件事情有多大的概率发生呀?”

我们中的一些人以概率为生，涉及发展新的概率理论，探索它的新应用前景，并将这些知识传授给人们，有时还会写一两本相关的书。我们自称概率学家，但数学家和统计学家的称呼要比概率学家响亮得多，通常你可以在一所大学的数理统计专业发现我们的身影，但却没有办法找到概率专业。实际上我们和数学家、统计学家都能沾上点边，但我们通常不愿承认这一点。如果我在一场鸡尾酒会上说自己是一位数学家或者统计学家时，人们大概都会兴味索然地离开。但如果我说自己是一位概率学家……，好吧，我承认大多数人还是会离开。因为这听起来就像提线木偶戏里的瑞典厨师正用一些生涩的词来吸引你的注意。但至少还有一部分人会留下来听我介绍这个我即将带你们走进的世界。

现在，让我们把自己想象成一位语言学家。我们首先从概率（probability）这个词开始研究它的含义。概率一词的拉丁词根是 probare 和 habilis，前者的意思是去试验，去证明，或者是去批准；后者表示才能，技术和能力。它们组合起来的单词——probable，最初是用来表示“值得肯定的”。在后来的使用中probable一词才渐渐有了“可能的”、“合理的”这一层意思，这才跟随机性产生了联系。在我的母语瑞典语中，偶然性一词对应的单词是sannolik，它在字面上的含

义就是“跟真理一样的”。德语中 wahrscheinlich 一词也表示这个意思。这和英语中概率一词还是有细微差别的，在韦氏词典中列举出了这些细微的差别。对我们来说，一个具体概率常用来描述一件事情发生的可能性，而概率论指的是概率这个学科。

概率伴随着随机性而产生。许多人就什么是随机性这个问题已经进行了深入的探讨，在此就不再赘述了，否则整本书就会陷入无尽的哲学讨论中。法国数学家皮埃尔-西蒙·拉普拉斯（1749—1827）曾将其经典地总结为：“概率是由我们的未知和已知组成的。”在拉普拉斯的启发之下，我们达成了共识：**当你遇到不确定性时，你就要使用概率了。**比如：

- 抛硬币、掷骰子或者转轮盘；
- 观察股市、天气或者美国橄榄球超级杯大赛；
- 想探知你家的后花园究竟有没有油井，火星上是否有生命，猫王是否还活着。

这些例子各不相同，在第一个例子的 3 种游戏中，结果都是等可能地出现，每种结果出现的概率可以简单用结果种类数的倒数来计算，也就是有 1/2 的概率抛到硬币的正面，有 1/6 的概率掷出数字 6，1/38 的概率得到数字 29（美式轮盘共有 38 个数字，包括 1 至 36 号、0 号以及 00 号）。这些都是显而易见的。我们能够计算出各种不同数字出现的概率。比如，在掷骰子时得到偶数的概率有多大？由于在 6 个数字中有 3 个偶数，因此结果就是 3/6 = 1/2。这些都是古典概率学的例子，它们也是数学家们研究的第一类的概率问题。其中法国数学家拉普拉斯和帕斯卡被认为是最为杰出的代表，他们在 17 世纪的来往信件被认为是最早开始对概率问题进行的系统研究。

下面我将列举三个例子来说明我们是如何将数据运用于概率的。根据观察得出当前的天气状况有 20% 的概率会下雨，那么我们就说今日降水概率为 20%。这个概率会随着天气数据的不断搜集而变化，我们称之为统计概率。在 2006 年的美国橄榄球超级杯大赛中，我在休斯敦德州人队上下了赌注，赔率是 800：1。这意味着庄家认为休斯敦德州人队夺冠的概率低于 1/800。庄家能够得到这个结论除了由于自己曾经在休斯敦度假并几乎中暑休克之外，还通过很多数据测算从而得出结论。

第三个例子与前两个例子有所不同，原因在于结果已经确定，只不过是你不

知道它而已。比如究竟有没有油井这个问题。当你在开始挖之前，你始终想要知道发现石油的概率。某一位地质学家可能会告诉你这个概率是75%。但是这个比例并不是说一年中有九个月油井都在你的后花园，而剩下的三个月在你的邻居那里。它只是表示这位地质学家认为你挖到石油的概率很高。另外一位地质学家认为你挖到石油的概率是85%，这也只是在数值上的变化，内涵依然不变，也就是你挖到石油的概率真的很大。我们把这些称为主观概率。但问不同的人“猫王活着的概率”这个问题，得到的答案要么是0%要么是100%。有谁会说猫王有25%的概率活着呢?

了解一些关于比例的知识对我们计算主观概率大有裨益。假设你在匹兹堡的姑姑简给你打电话告诉你她的新邻居人很好，并且有一份跟恒星、占星家或天文学家有关的工作。如果没有足够的信息，这个邻居是一位天文学家的概率有多大?在几乎没有任何信息的前提下，你会说是50%吗?有些人会这样认为。但是你必须要考虑这个事实：在美国占星家的数量是天文学家的四倍，所以20%的概率更加现实。不要因为一件事情是“或者……或者……”就认为它的概率是一半对一半。在这种情况之下，安迪·鲁尼的50-50-90规则显得无比睿智，他说：“当你有50%的机会猜对一件事时，那么也许有90%的可能你猜的是错的。”也就是说，如果两件事机会均等，那么猜对事件发生的可能性微乎其微。

1.2 概率学家的玩具和语言

概率学家都喜欢玩硬币和骰子。从柏拉图式理念的角度上看，我们喜欢一切出现跟抛硬币和掷骰子一样的等可能事件的实验。假设现在有一个家庭有四个孩子，随机选择孩子。四个孩子全是女孩的概率有多大?如果用抛硬币的方法，就是连抛四次硬币每一次都出现正面的概率。很多概率问题都可以用抛硬币的方法解决。但重复使用这种方法会让大家觉得枯燥无味。所以我们就用掷骰子、转轮盘、从盒子里拿球或抽扑克牌这些方法来替代，在本书后面有关博彩那章将会详细介绍这些游戏的玩法。因为概率最早就是从博彩中起源的。

概率是表达不确定性的一门艺术。“抛硬币抛到正面的概率是1/2”是一个精确的表述。它表示抛到正面和反面的概率是一样的。这也是在长期反复试验的

情况下考虑概率的一种方式。如果你连续不断地抛硬币，从长远来看差不多有50%的概率出现正面，50%的概率出现反面。这是你能够确定的事，但你不能确定的是下一次抛的硬币究竟是正面还是反面。

概率学家们也有自己的术语。例如，我们通常会把不确定的情形称为“试验”。这种情形可能是一次真实的抛硬币或掷骰子试验，也有可能是完全不同的股市起伏或者是温布尔登网球决赛。试验结果可能是“正面朝上”、“6”、“沃尔沃的股价上涨”、“比约·博格赢了”（这些都是现实试验的一种真实结果）。这些结果就叫做“事件”。简而言之，事件是指发生在试验中的结果。它可以是一个简单的结果（比如骰子的数字6）也可以是一组结果（比如骰子中的偶数）。关于事件这个概念的数学描述就是：事件是所有可能的结果的集合中的一个子集。数学家们把这个结果称为集合的一个元素。统计学家用“结果”和“事件”来强调与现实生活中发生的事情的联系。在公式中，通常用大写字母来指代具体的事件，用字母“P”来表示概率。因此 $P(A)$ 表示的意思就是“事件 A 发生的概率”。

所有可能出现的结果的集合被称为样本空间[○]。在有些情况下，样本空间的选择多种多样。假设你现在抛两枚硬币，需要知道两枚硬币同时正面朝上的概率。这时可能出现的正面朝上的硬币个数为 0、1、2。你可能会将这三个数字选为样本空间，然后平均下来每种结果出现的可能性为 1/3。但是，如果你不断重复这个试验，过一段时间你会发现得出两枚硬币同时朝上的概率小于 1/3。问题就在于你选择的样本空间包含的三个结果并不是等可能地出现。现在让我们区分一下这两枚硬币，将其中的一枚涂成红色，另外一枚涂成蓝色。它们将会等可能出现四种结果：同时正面朝上；红色的正面朝上，蓝色的反面朝上；红色的反面朝上，蓝色的正面朝上；同时反面朝上。让我们以一种更为简单明了的方式把这四种情况表现出来：HH、HT、TH、TT[㊀]。四种情况中有一种是两枚硬币同时正面朝上，所以正确的概率应当是 1/4。图 1-1 所示为抛两枚硬币出现的四种结果。

○ 样本空间是数学家们的标志性术语。1931 年，奥匈帝国战斗机飞行员理查德·冯·米泽斯在他的德文著作《概率微积分》中创造了 Merkmahlraum 一词，在德语中表示样本空间的意思。

㊀ H 为英文单词 head，表示正面朝上。T 为英文单词 tail 表示反面朝上。

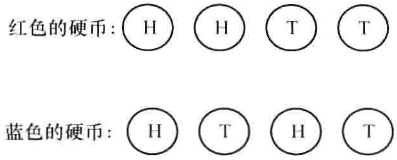


图 1-1 抛两枚硬币等可能出现的四种结果

接下来这个问题也是相同的。假设要掷两次骰子，得到两次的点数之和等于8的概率有多大？首先我们需要明确，掷两次骰子点数之和可以是2、3、…、12，但是它们出现的概率是不一样的。为了找出每种可能出现的结果，我们同样需要用前面提到的区分硬币的方法来区分一下这两个骰子，也就是将它们分别涂上红色和蓝色。点数之和等于8，存在着以下三种情况：2 +6，3 +5，4 +4。我们首先认为36 种结果之中有三种情况能够得到总和等于8。但是我们还需要进一步作出区分。比如将2 +6 这种情况区分为蓝色骰子点数为2，红色为6 的情况和蓝色骰子点数为6，红色为2 的情况。这样区分之后，我们就会意识到一共有五种情况可以得出点数之和为8 的结果。因此，两次的点数之和等于8 的事件的概率就是5/36 了。详细的样本空间分布如图1-2 所示。

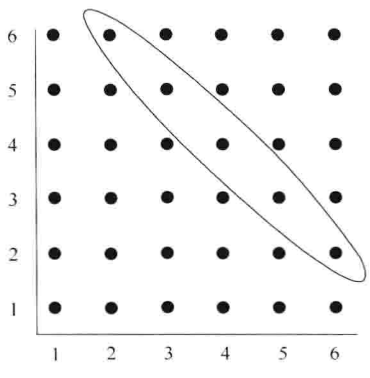


图1-2 这个样本空间是掷两次骰子可能出现的36 种结果。点数之和为8 的事件用椭圆标注出来了，总共有5 种可能。因为得到2和6 与得到3 和5 同时都存在两种情况，而得到4 和4 只存在一种情况。

图 1-2 掷两个骰子的样本空间

接下来我们再举一个类似的例子。假设一个家庭有三个孩子，那么只有一个女儿的概率是多大呢？实际上，这个家庭可能有0、1、2、3 个女儿，而这四种情况出现的概率是不一样的。我们把这三个孩子按照出生的顺序进行排列区分，比如说BGB 代表第一个孩子是男孩，第二个孩子是女孩，第三个是男孩。那么，

可能会出现 8 种不同的结果：

BBB，BBG，BGB，GBB，BGG，GBG，GGB，GGG

现在我们能够很轻松地算出只有一个女儿的概率是 3/8。现在来算一算这个随机选出来的女孩的同胞没有姐妹的概率是多少呢？这种情况看起来也很相似。她没有姐妹，意味着这个有着三个孩子的家庭有只有一个是女孩的概率是 3/8。相信这个结论吗？你不应该相信。这种情况是完全不同的，因为我们并不是随意选了一个有三个孩子的家庭，而是选择三个孩子中至少有一个是女孩，因此，出现 BBB 的情况是不可能的。那么概率变成了 3/7 吗？思考一会再往下看吧。

我希望你思考之后的答案是否定的。要解决这个问题我们需要一个完全不同的样本空间。我们用星号 * 来标明这个确定的女孩，那么将会出现如下 12 种情形：

BBG*，BG*B，G*BB，BG*G，BGG*，G*BG

GBG*，G*GB，GG*B，G*GG，GG*G，GGG*

因此，这个女孩没有姐妹的概率为 3/12 = 1/4。提醒大家注意之前的那些结果在这个样本空间中是如何被拆分的。三个孩子都是女孩的情形，即出现 GGG 的情况有三次，因为任何一个女孩都可能是我们指定的那个。我们计算出来的结果表明有三个孩子的家庭中只有一个女孩的占 37.5%，而 25% 的来自三个孩子家庭的女孩并没有姐妹。

那么三个孩子性别相同的概率又是多少呢？首先考虑一下这个不太完整的论证方法：其中两个孩子必定是同性别的，而对于第三个孩子来说是男孩或是女孩的概率是一样的。因此三个孩子同性别的概率是 1/2。这个例子就是投掷硬币问题的变形。早在 1894 年，英国贵族出生的业余科学家弗朗西斯·高尔顿爵士就用投掷硬币问题来说明草率思维的危害（在后文将会进行详细介绍）。让我们用第一个样本空间来发现错误，证明正确的概率应该是 1/4。

有一个古老的博弈问题跟这一事件异曲同工。甲、乙二人参与投掷三颗骰子的游戏，如果三个数相加之和为 **9**，则甲赢；而如果三个数之和为 **10**，则乙赢。如果既不是 **9** 也不是 **10**，那么就继续投掷。这种游戏规则公平吗？

有以下 6 种情况，三个数相加之和可以为 9：

1 + 2 + 6，1 + 3 + 5，1 + 4 + 4，2 + 2 + 5，2 + 3 + 4，3 + 3 + 3

同样有 6 种情况可以使其和为 10，分别是：

$$1+3+6,\ 1+4+5,\ 2+2+6,\ 2+3+5,\ 2+4+4,\ 3+3+4$$

看起来这个游戏非常公平，但是从长远来看乙肯定会逐渐赢甲。这是为什么呢？

当你下定决心要玩这个游戏时，首先需要确定两种方式投掷结果的概率是相同的。就像之前考虑两个骰子的问题一样，假设三个骰子有三种不同的颜色，分别是红色、绿色和蓝色。如果我们按顺序来投掷这三个骰子，可能得出的结果是 (1，1，1)，(1，1，2)，(1，2，1)，(2，1，1)，(2，2，1)，一直到 (6，6，6)；我们可以立刻知道总共有 $6\times6\times6=216$ 种可能性。让我们来看有多少种情况得出组合 $1+4+4$，结果为 9。这种组合对应这三种可能出现的情况：(1，4，4)，(4，1，4)，(4，4，1)。接下来讨论 $1+2+6$ 这种组合，它对应有 6 种可能：(1，2，6)，(1，6，2)，(2，1，6) (2，6，1)，(6，1，2)，(6，2，1)。总之，如果三个骰子掷出来的是不同的数字，那么一共有 6 种可能性；如果两个骰子掷出来的是相同的数字，那么有三种可能出现的情况；如果三个数字都是一样的，那只有一种可能。

现在我们知道总和为 10 有 27 种可能，但总和为 9 只有 25 种可能。决胜关键点就在于出现 $3+3+3$ 只有一种情况，但出现 $3+3+4$ 有三种情况。详情可以参见图 1-3。因此，在出现的 52 种有关输赢的情况中，27 种情况是乙赢，概率大约是 52%，而甲只能在剩下的 25 种情况出现时赢，换而言之甲赢的概率是 48%。虽然差异不大，但足够让一些庄家以此谋生了（风险投资也是凭这种方式运行的）。

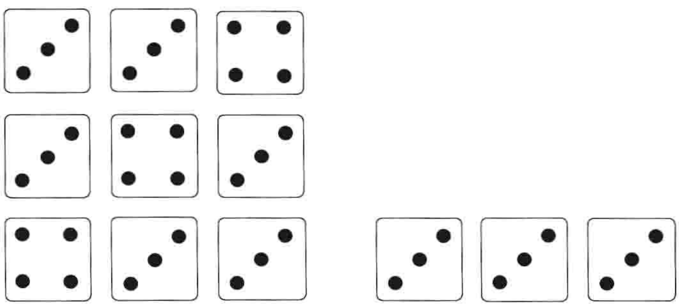


图 1-3　出现两个 3 一个 4 有三种情形（左图），但是通过三个 3 得到 9 只有一个情形。（右图）。