牛津大学出版社镇社之宝

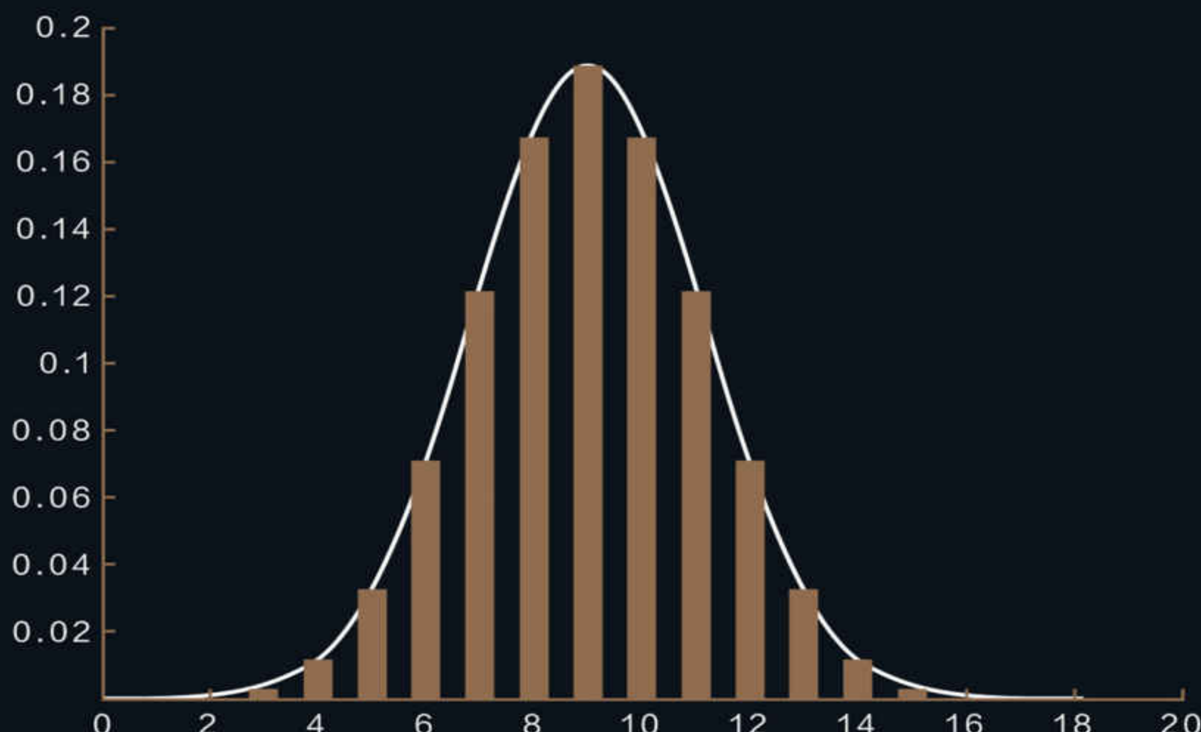
概率

牛津通识课

三小时读懂概率如何帮你做好选择

PROBABILITY: A VERY SHORT INTRODUCTION

[英] 约翰·黑格 著　　徐大成 译



东方出版中心

牛津通识课：概率

[英]约翰·黑格 著
徐大成 译

东方出版中心

图书在版编目（CIP）数据

牛津通识课. 概率 / (英) 约翰 · 黑格著；徐大成译. -- 上海：东方出版中心，2020.12
ISBN 978-7-5473-1744-0

Ⅰ. ①牛… Ⅱ. ①约… ②徐… Ⅲ. ①科学知识－普及读物②概率－普及读物 Ⅳ. ①Z228②O211.1-49

中国版本图书馆CIP数据核字（2020）第248911号

上海市版权局著作权合同登记：图字09-2020-1022号

牛津通识课：概率

著　　者　［英］约翰 · 黑格
译　　者　徐大成
责任编辑　江彦懿
封面设计　汪芝灵

出版发行　东方出版中心
地　　址　上海市仙霞路345号
邮政编码　200336
电　　话　021-62417400
印 刷 者　天津联城印刷有限公司

开　　本　787mm x 1092mm　1/32
印　　张　7
字　　数　80千字
版　　次　2021年2月第1版
印　　次　2021年2月第1次印刷
定　　价　32.00元

目录

致谢

我要感谢莱顿·沃恩·威廉姆斯(Leighton Vaughan Williams)和迈克·史密斯(Mike Smith)，因为他们提供给我关于赛马赔率的信息；我要感谢莱斯·米勒(Les Miller)提供他对轮盘赌的仿真结果；我要感谢伊恩·麦克海尔(Ian McHale)对如何估计足球世界杯决赛圈中32支队伍的胜率的详尽解释；我还要感谢德里克·罗宾逊(Derek Robinson)在图表方面对我的帮助。我要特别感谢一位匿名评论者，他鞭辟入里、富有建设性的建议帮助我消除歧义，更连贯地编排内容和提取这个学科的中心思想。

对那些认为自己是书中的信息、观点和逸事的来源，但未被我提及的人，我恳求原谅，我引用了太多资料以至于不能列出全部出处。这不是一本其中每个断言都能找到其来源的学术专著，本书意在提升读者们对概率这门学科是什么、如何发展和如何应用的兴趣。

所有书中遗留的谬误和缺陷都来自我自己。

约翰·黑格

2011年8月

01 基本原理 Fundamentals

概率的视角

概率是不确定性这一概念的形式化表述。误打误撞效应显然到处都是。从生物学上说，我们都是父母基因随机混合后的产物。像是石油泄漏、火山喷发、海啸、地震等灾害，或是中彩票这样令人愉悦的事情，都会随机且显著地影响人们的生活。

许多人具有良好的理解概率的直觉，但在你对某件事情有了某种先入为主的观点，而后来一些具有不完全明显的相关性的新事实被披露出来的时候，这种理解就会让你误入歧途。的确有一些臭名昭著的有关生日、二孩家庭、有三个选择的电视节目游戏的“诡计问题”(trick questions)，它们似乎被设计成说服你这门学科是有违常识的。其实概率并不违背常识，只要清除掉或者考虑到这些问题中所有隐藏的假设，合理的答案就会浮出水面。只不过概率的确需要清楚的思维过程。

概率的广泛应用促进了这门学科概念和方法的发展。1944年6月的诺曼底登陆[1]能够发生，就是因为当时人们认为有利天气出现的概率相对较高。荷兰的工程师们在建造保护其国家免受海洋侵袭的堤岸时，必须考虑发生严重洪水的概率。一种新型治疗方法是否比先前的方法更能帮助一名患者多生存五年？你需要交多少钱来给自己、车辆、房子或财产上保险取决于早期索赔的可能性。你所做的大多数决定：在学校学习什么、选择谁作为人生伴侣、在哪里居住、从事什么工作都是在有不确定性的情况下进行的。就像皮埃尔-西蒙·拉普拉斯(Pierre-Simon Laplace)在1814年所说的那样：

……生命中最重要的问题大多都只是概率问题。

“概率是……”这样的措辞无论何时出现，都伴随着某些假设(它们可能在不经意间被忽略了)。如果那些假设是无端的，那么这些断言就不会被人相信。我希望在这本书中假设是明确的，无论它们是含蓄还是直白。在我们将目光转向概率的种种阐述能如何被诠释之前，先描述一下产生这些阐述的不同思路。**古典概率**

概率的**古典**(classical)或者说**客观**(objective)视角经常出现在有关概率的游戏中，例如掷色子和转轮盘赌。这些过程都会产生一系列可能的结果，我们出于对称性的考虑，或者因为找不到是其中一个结果而不是另一个更会发生的原因，认为它们都是等可能的。所以我们只是对结果进行计数，并赋予它们相等的概率，这样试验中的任何事件的概率都被认为是引发它的结果占所有结果的比率。

例如，连掷两次硬币，四种可能的正反面结果是：正正、正反、反正、反反。就一枚**公正的**硬币来说，每次掷出正或反都是等可能的，所以四个结果中没有一个比另一个更可能或更不可能，每一个结果的概率都应该是1/4。其中有三个至少一次掷出正面，所以总体上讲正面出现的概率是3/4。

从一个牌堆中取两张扑克牌，有1326种结果(请相信我的话)。如果牌堆是被洗好了的，我们就认为这些扑克牌组合都是等可能的。因为其中有64种牌面由一张A和一张“十牌”(即10、J、Q或K[2])组成，所以我们得出结论，抽到这样的组合——“二十一点(Blackjack)”——的概率是64/1326，刚好不到5%。

仅从概率的角度而言，这些例子都可以转化为从装有完全相同的球的袋子中取出某个球的形式。第一个例子对应的袋子中装有4个球，3个是红球；第二个例子对应的袋子中装有1326个球，其中64个是红

球。的确，每一个对概率的客观考量的例子本质上都与从袋子或者瓮中取出一个球的问题完全相同(这就解释了学生们教材中这类例子过多的原因)。

我要强调的是，仅仅计算可能结果的数量然后计算多少个结果会引发相应的事件是不够的。一定要有令人信服的理由说明任何结果都不会比其他的更可能或更不可能发生。否则，基于彩票只有两个可能的结果：要么中奖，要么不中，你会掉入买彩票中大奖概率是50%的思维陷阱中!

试验证据——频率

我们希望在“大富翁”这类家庭游戏或者例如双色子赌博的赌场游戏中，色子的六个面中掷出每一个都是等可能的。但如果色子由不均匀的材料制成，或者它的长度、宽度和高度三者不相同，那么假定每种结果是等可能的显然不明智。在相同条件下进行的一系列投掷过程中，出现任何一个面的频率都会波动，但最终将会稳定并趋近于一个特定值。

不可能出现前1000次投掷中20%的结果是6点，而接下来的1000次投掷中这个比例跳到了60%。在这些可重复试验中，结果可能是不完全一样的，但是每一个结果都倾向于表现出某个特定的频率，**频率论者**(frequentist)认为这个频率值就是相应结果的概率。

对于一个不完美的色子，在前1000次投掷中，我们可能会得到170次6点，下1000次中，可能得到181次6点，诸如此类。我们不能从这些试验中推断出掷出6点的概率**精确值**，但是试验数据指导我们对概率进行估计，我们收集的试验数据越多，我们估计得就越准确。我们无法知道概率的精确值，但这一事实并不能否认概率的存在。

如果我从洗好的牌堆中抽取一张牌，似乎没有理由认为某种花色比其他花色更容易被抽到。每种花色都有1/4的客观概率。而且如果我放回这张牌，重新洗牌，然后再进行100次试验，我会预期每种花色的出现是同样的频繁，就是大约25次。类似地，对于投掷结果都是等可能的普通色子，投掷结果是5点的概率客观地讲是1/6。在600次投掷中，我们预期掷出5点的次数大约为100次。

在重复大量具有等可能性结果的试验时，任何特定结果相应的**频率**都预期会接近于它客观计算的概率。一个公正的硬币极少会在100次投掷中给出50次正面朝上的结果，但是直觉上我们不知道该期望投掷结果多么接近理想情况才合理。

频率观点不仅被应用于同样条件下的重复性试验，还有在即将出生的婴儿是男是女上。不考虑家庭因素，我们来看看从许多国家和文化环境中收集的覆盖了很长时间跨度的数据。一个持续的模式是：每49个女婴出生，就有51个男婴出生。鉴于无法将某个新生儿和其余的进行区分，一个频率论者会认为生男孩的概率是51%。

一些规模惊人的试验已经开展了。1894年，动物学家拉斐尔·韦尔登(Raphael Weldon)发表了将12个色子投掷2600次的结果。他的数据与六个面等可能出现的观点相抵触，因为5和6这两个数字出现得太频繁。为了辨认数字，他的色子上每个面钻了小孔，刻有5和6的面分别对着刻有1和2的面。这些色子的重心就会更接近数字较小的面，这给出了一个对观察结果中频率过大貌似正确的解释。

大约70年后，一个有大量时间的一丝不苟的人——威拉德·朗克尔(Willard Longcor)在哈佛大学顶尖的统计学家弗雷德里克·莫斯特勒(Frederick Mosteller)手下效力。在莫斯特勒的指导下，朗克尔收集了超过200个色子，并将它们中的每一个都投掷了超过20 000次，只记录结果的奇偶性——得到超过400万个数据。为了让每次投掷的环境

尽可能相同，他使用了一个铺了毯子的桌面，用一个升起来的台阶将色子弹下去。那些类似韦尔登使用的廉价色子存在微小但明显的偏差，以至于出现了太多的偶数，这并不出人意料，也是那些钻孔的原因。而对于那些使用在拉斯维加斯赌场的高精度色子，上面表示数字的点不是轻轻画上去的就是极薄的圆盘贴上去的，就没有可检测到的偏差。这些色子各种结果的频率与在古典视角下等可能结果的概率是一致的。

“二十一点”专家皮特·格里芬(Peter Griffin)挖苦地说，他在拉斯维加斯玩的1820局牌中，庄家牌堆顶上要么是十牌，要么是A的情况出现了770次。而抽到这些对庄家有利的牌的客观概率是5/13，所以格里芬怀疑自己是否被欺骗了——随机概率只会让发牌者抽到这种好牌大约700次。

2002年3月，马拉维有6202名五岁以下的儿童被认为疑似患上了肺炎，其中523名儿童死亡，死亡率为8.4%。已知没有某些特殊情况让这段时期不同于以往，一个频率论者就会推断：一名患上肺炎的马拉维儿童的死亡率是8%~9%。从客观角度来说，关于马拉维患有肺炎的儿童的死亡率的一般性陈述仍是一种推测，尽管基于这样确凿的证据：如果随机从那些**特定的**6202名儿童中选择一名，他的死亡概率是8.4%。

我们将会在后面更深入地讨论频率数据和客观概率的关系。

主观诠释

布鲁诺·德·菲内蒂(Bruno de Finetti)是概率这个领域中最有影响力的思想者之一，他曾写过：

> 概率不存在。

作为概率理论方面的教授，他并不是在将自己研究的学科比作海市蜃楼，而是在驳斥例如“正面朝上的概率是1/2”这种**绝对性**的陈述。对于他来说，每一个包含概率的陈述都是观点的表达，这种表达基于一个人自己的经验和知识，并且有可能在更多的信息被发现的时候发生变化。

考虑如下五个断言：

英国板球队队长会在下一次国际板球对抗赛猜对硬币；

奥斯卡金像奖最佳男主角奖获得者，无论是谁，都会在下一年再次获奖；

没有奥斯陆出生的人曾经获得过奥运会击剑金牌；

理查三世(Richard III)应该对“塔中王子[3]”的死负责；

如果拉尔夫·纳德(Ralph Nader)没有成为候选人，阿尔·戈尔(Al Gore)本会在2000年被选为美国总统。

对于这其中的每一个推断，我们都能够给出自己的**可信度**(degree of believe)、**个人概率**

(personal probability)或者**主观概率**(subjective probability)。这将会是一些非负数，并且不大于1，就是说它是一个介于0%和100%(含)之间的百分比。

0和1分别代表着两个极端——**不可能**和**必然**。我确信在本世纪内足球世界杯必然会再次由非洲国家举办。我认为年龄小于20岁的人不可能获得诺贝尔物理学奖[4]。

评估主观概率

上面的五个断言各具有不同的性质，关于它们我们有多种不同的佐证。对第一个断言来说，我们能用正面和反面的对称性加以反驳；对第二个断言来说，我们可以参考1929年以来的奥斯卡奖历史记录，前两个情形都能在很短时间内确定其真实与否；第三个断言，无论是真是假，都可以通过盘点奥运会获奖记录来确定；第四个或真或假，但我们永远都无法确定；我们不能让历史重来去探明第五个断言是真是假。

后面会有一些具体的例子来阐释主观概率是如何被评估的。除了这些观点之外，有至少三个一般性评估主观概率的不同方法。一个就是做出一个事件发生与否的合理赌注。但是这个方法不总是对每个人有用：有些人原则上抵制赌博，还有一些人不考虑进行可能导致个人损失的行为。而且对于那些愿意赌博的人来说，他们的**合理赌注**也可能会随着他们站在打赌双方的哪一边而变化。

第二个评估某件事可信度的方法就显得有些刻意了。你会选以下哪一个：猜某一个事件是否发生，或者猜牌堆顶上第一张牌的颜色是红色还是黑色，猜对了获得5英镑。如果你更喜欢后者，就说明你认为此事件的可信度在50%以下。

假设我们继续比较如下两种情况的预期，这个事件是否发生，还是猜第一张牌的花色，猜对获得5英镑。后者有25%的可能会发生，所以你对这两种情况的选择，会反映你认为这个事件的可信度是比25%低，还是在25%到50%之间。

更加精细地比较这些数值会让我们无法确定更偏好哪一边。你对这件事的可信度就会接近于那个选牌的客观概率。你也许会想要使用装有20或者100个完全一样的球的罐子来明确地评估这个事件的可信度，而不是计算分数很困难的有52张牌的牌堆。

这里给出一些具有合适精确度的结果。2010年，网球运动员约翰·伊斯内尔(John Isner)和尼古拉斯·马胡(Nicolas Mahut)进行了温布尔登网球锦标赛[5]史上最长的比赛。经过计算，他们在下一年再次成为对手(这的确发生了)的精确概率是2/285，或许应该四舍五入到“略低于1%”。但是《星际迷航》[6]的一集中，史波克先生(Mr. Spock)告诉柯克(Kirk)他们逃脱的胜算“大约是7824.7比1”，这就很荒谬了。

第三种方法，想一笔金额大小合适的钱，别太少以至于对你来说无关紧要(比如一便士)，也别太多以至于拥有了它就会对你的境遇产生巨大的影响(对大多数人来说是100万，对于比尔·盖茨就要数额大一些了)。我觉得10英镑就行——把它叫作单位金额。

现在假设，不知何故某个事件的真实与否会在明天揭晓：并且如果这个事件是真实的你会获得这个单位金额，如果它是假的就什么也得不到。但有一种提议是：不用等到明天，今天你会获得单位金额中确定比例p的一部分，但对你来说今天或者明天得到这笔钱没有什么差别。

如果p特别小，你可能就会拒绝这个提议，并且更愿意等待；如果p接近于单位1，你可能会接受提议中这个确定比例的金额。但是中间会存在一个p的值让你在接受这个提议与等待明天结果被揭晓中摇摆不定。这个p就是你认为这个陈述或者事件的可信度。

这里我提供我自己的对上述五种断言的主观答案。我认为没有合理的原因来解释为什么一方比另一方更可能在板球掷硬币中获胜，所以给出的第一个数字是50%；浏览奥斯卡奖的历史，不仅是演员奖，其他类别的奖项也只是零星地在相邻年份中重复颁发——可能现在参选人更多了，所以我给出3%，或者更低；挪威人不以善于击剑著称，但是重剑、花剑、佩剑这些击剑项目自1896年以来一直出现在所有的夏季奥运会中，一些奥斯陆出生的人也许曾经在某次获得过金牌，但是我很怀疑——这里我给出的数字大约是95%；出于对白玫瑰郡[7]的偏爱，而不是客观证据，对于第四个断言我给出10%；对于第五个断言，考虑到每一个州的投票情况和纳德获得的貌似合理的票数，我给出20%。

在这儿停一下，给出你们自己对这五个断言的意见。在事情不确定的时候，你越善于评估概率，你在生活中做的决定就越可能让你开心。

赔率

无论我们使用古典概型、频率诠释还是可信度，**赔率**(odds)这个词语在描述概率的时候经常出现。我们可能会说用公正的色子掷出6的赔率是“1赔5”——在一系列投掷中，每一次我们掷出6，预期都会有五次掷不出。如果一个结果预期比它的反面更有可能发生，例如排名更高的网球选手获得比赛的胜利，那这个结果就被叫作**有胜算的**(odds on)。

概率和赔率之间有确定的对应关系，我们能够简单地将其相互转换。思考一下频率将会很有帮助。如果概率是20%，或者说1/5，我们预期这个事件在五次机会中发生一次，所以赔率是“1赔4”。如果概率是75%，我们预期它会在四次中发生三次，给出“3赔1”的赔率。如

果赔率是“5赔6”，这就表明每五次事件发生，就会有六次事件不发生，所以概率是5/11。

你不必拘泥于数字。在洗好的牌堆顶上的牌是K或者Q的概率是2/13。这可以被表述为“2赔11”，或者同样精确的“1赔5.5”。喜欢哪个就用哪个。

虽然短语“赔率是1赔1”从来不被使用，但是它很有含义。它表明期望一个事件发生和不发生的机会是五五开，所以它的概率是1/2。然而，我们会板着脸说“赔率是均衡的”。

需要解决的问题

对于如何使用概率，我们没有重大的分歧，但是我们曾经讨论过的三种方法的信徒们可能会从不同的角度分别论述它们的价值。每一个观点都有其用途。为了理解这门学科的运作方式，无论从哪个思路我们都要探寻适当的观点。

客观方法被限制于有限多个结果的情况，所有的这些结果都被判断为等可能的。但是没有硬币或者色子是完美地对称的，基于什么我们可以把这些不完美当作无关紧要的元素而不去考虑它们呢？甚至于我们能否确定可能结果的数量呢？例如假设我们有一个装有两个球的罐子，这两个球要么均为白色，要么均为黑色，要么一白一黑。我们是否可以说有3种等可能的情况，或者球在按顺序被置入的时候，实际上是白白、白黑、黑白或者黑黑这4种等可能情况？持不同看法的人会对两个球均为黑球的概率给出不同的答案。或者假设你到达了一条分出三条岔路的路口，两条路通往新城，第三条路通往海港，做一个“随机选择”：你去往海港的概率是1/3(三个出口中的一个)还是1/2(两个目的地中的一个)？

一个频率论者希望处理可重复试验，它在完全相同的情况下能够不限次数地进行。试验结果的数量不需要是有限的——想想掷同一枚硬币直到正面连续出现3次，或者在一根棍子上取一个随机的点。但是，无论我们多么小心，试验环境都不会保持**绝对**一致，而且任何极限值都只能做估计。怎样描述这种估计中的误差？宣称误差在2%以下的概率是99%，就会引入循环论证——我们需要知道概率是多少，以便定义它！

一个国家入侵另一个国家的概率，或者特定的一次心脏移植成功的概率，这类问题中的情形只出现一次，而且备选结果不能被减少为有限列表中的等可能的结果。客观和频率方法对这些事件都无能为力。这就需要主观概率了。

主观主义者必须确保她相信的事情都是自洽的。例如，在英国国家彩票[8]中，一个机器从列表{1， 2， 3，…， 49}中选取6个数字，苏西也许会倾向于认为约1400万种选择均是等可能的。那么，当问到下面哪一个更有可能的时候：

(a) 抽取的数字中没有超过44的；

(b) 那些抽取到的数字中不包括连续数字。

她或许会在想了一会儿之后选择二者中的某一个。但是只要她选择了这些事件中的**任何一个**，她都会因为自己的观点不能自洽而愧疚，因为合理的计算显示这两个事件发生的可能性正好是相等的！对于这种不自洽性，主观概率方法仅仅要求它被解决，但是并未给出确切的解决方式。

因为相比具有有限多种等可能选择的情形，我们希望考虑更宽泛的情况，在考虑不能不限次数地重复试验的情况时，我们将主观概率

方法作为默认选项。而且一旦有客观或者频率方法的支持，我们将会更加坚信我们的观点。

解读

借用“袋子中的球”的视角，一些事件的概率被当作是袋子中红色球的比例。所以仅当袋子中没有红色球的时候，概率的值才是0。在这种情况下，这个事件永远不会发生。类似地，概率为单位1对应着每一个球都是红色，所以这种情况下这个事件每次都会发生。只有0和单位1这些值，才可能确凿地被试验证据证明是**错误的**：如果事件发生了，它的概率就不可能是0；当它没有发生，它的概率就不可能是单位1。而且这从频率或者主观方法来说也是对的。假设概率是某些中间值，比如说3/4。

我们首先来处理一个十分细致的问题。无论一个轮盘赌轮被设计得多么好，从本质上讲所有标着数字的格子被转到的概率**精确**相同是不可能的。赌场需要的是这些概率足够接近理想情况，以至于不大可能分辨出任何数字的概率比其他数字更多或者更少。类似的说法也适用于色子、硬币和纸牌。所以类似于“概率是3/4”的说法，意味着对于所有实际目的来说概率都足够接近3/4。否则，一个书呆子就会沾沾自喜地告诉你，他**知道**概率不是3/4，并且不害怕引起争执。

在可重复试验的背景下，我们期望从这个断言中得到什么信息呢，“得到红球的概率是3/4”？值得强调的是，我们并不会期望如果进行4次这个试验(每次取球之后放回)，我们会精确地在其中3次抽到红球。可能的情况是，4次重复试验根本没抽到红球，或者甚至每次都是红球。但是在一系列漫长的重复之后，我们的确期望红球出现的频率接近3/4。

漫长的重复试验有多长，或者结果需要多么接近3/4？没有一个确定的、非黑即白的答案。如果在前40次重复试验中，我们只有20次抽到红球，我会强烈质疑概率是3/4的断言，但是如果接下来的40次中得到28次红球的结果，那些质疑就会被削弱。相信或者不相信这个断言可能会在相当长的一段时间内是临时立场。假设试验条件的确一直保持不变，我们使用**所有**收集到的数据来做决定——试验次数过少会引起误导。

稍后我会提供一些准则，并且证明它们。我们以重复100次试验为例，假设概率是一个中间值，接近一半。计算这个数字与由数据得到的真实频率的**差值**：如果这个差值超过0.1，我会对这个断言产生一些怀疑；如果差值超过0.15，我会产生强烈质疑。在重复试验1000次而不是100次后，我期望结果有更强的一致性，所以用0.03和0.05代替原来的数字。如果假想的数字接近0或者单位1，比如说10%或者90%，我也会期望更强的一致性。在重复试验的基础上，特定的概率更容易让人信服，而**不是**某个所谓的值。

对于一个主观评估，例如明天降雨的概率是60%，情况是怎样的呢？我们不能数百次地再现今天的天气情况，然后检查降雨是有多频繁。这种“试验”只能够进行一次。但是我们也许可以通过检查这个数字产生的过程来检验这个断言。天气预报员使用天气规律的模型来得到他们的结论，即使他们的电脑屏幕上的数字是31.067%，他们也会聪明地给出大约的数字，你会听到“降雨的概率是30%”。所以现在你就能收集不同日期的数据，看看经验证据——在去年给出降水概率为30%的83天中，有多少天真的下雨了？只要那个比例合理地接近于30%，你对这个方法的信心就会增强，所以接受对“明天”的降雨概率就是个理性回应了。

概率是在不确定情况下做决定的关键。如果你真诚地相信特定的一件事或者一个论断的概率是单位1，那么你应该按照它无疑会发生一样来行事；如果你真诚地相信概率是0，那么就按照它好像绝不会发生一样来行事。

如果你认为概率是从0到1之间的某个值，那么就按照你预期它会发生的比例来行事。例如，如果你的判断是概率为60%，想象你会面临这样的情况100次，在60次中(但是你不知道是哪60次)这个事件会发生，而在40次中不会发生。努力理解，考虑到这种权衡，决定你的行动。如果你猜测概率是80%，说明你预期这个事件会更频繁地发生，你的行动可能就会不同。

就像大主教约瑟夫·巴特勒(Bishop Joseph Butler)1736年在他的《宗教的类比》(Analogy of Religion)中写的那样："对于我们来说，概率正是生活的准则。"

[1]　诺曼底登陆(The D-Day Invasion of Normandy)是第二次世界大战时西方盟军在欧洲西线战场发起的一场大规模攻势，为"霸王行动"的一部分。

[2]　这四种牌面都算作10点，原文为"ten-card"。

[3]　塔中王子(the Princes in the Tower)，指英格兰国王爱德华四世(Edward IV)的两个儿子：爱德华五世(Edward V)和约克公爵(Duke of York)。他们被理查三世关进伦敦塔之后失踪。

[4]　原文为："I think it is impossible for someone under twenty years of age to win a Nobel Prize."本书成书于2014年马拉拉·优素福·扎伊获得诺贝尔和平奖之前，经与原书作者John Haigh沟通，此处添加"诺贝尔物理学奖"的限定。

[5]　温布尔登网球锦标赛(The Championships, Wimbledon)，网球运动中历史最长和最具声望的公开赛之一。

[6]　《星际迷航》(Star Trek)，美国系列科幻娱乐影视剧。

[7]　指约克郡，白玫瑰为其与约克王朝的共同象征。理查三世为约克王朝最后一任国王。

[8]　英国国家特许经营的彩票，开始于1994年。

02 概率的运作 The Workings of Probability

除了主观的、客观的和频率的理解方法，还有其他理解概率的视角。例如，一定要坚持将一个概率对应于某一个数字吗？我们是否可以说一个概率更大，或者一件事情的可信度比另外一件事情更高？我们真的必须提出一组公理——不言而喻的事实——并据此建立一套理论？

许多杰出的作者都认为建立两个独立的理解概率的方法是有用的，一个是可信度，另一个是古典概率。两者应该具有相同的逻辑规律，不自相矛盾，但两者对于概率是如何生成的和被理解的可以不同。任何理论都应该与古典观点一致，基于可重复实验都会给出等可能的结果，所以我们将着眼于这些案例，寻找概率必须遵循的规则。

加法定理

从洗好的牌堆里面取一张牌。我们认为抽到所有牌都是等可能的，所以求出任何事件的概率——例如抽到梅花、黑桃或者A——就是计算这些事件占总事件的比例。我们如何求出两个事件之中的**每一个**发生的概率呢？

如果两个事件的所有可能结果中没有任何相同，我们称这两个事件是**相互排斥**(mutually exclusive)或者**不相容**(disjoint)的。“抽到黑桃”和“抽到梅花”这两个事件是不相容的，但是“抽到黑桃”和“抽到A”这两个事件不是，因为“抽到黑桃A”同时属于这两个事

件。当两个事件**互斥**，这两个事件中任何一个发生的结果总数就是其分别发生的结果数之和，所以我们有一个简单的结论：

当两个事件互斥时，至少一个事件发生的概率是两个事件各自发生概率的和。

这就是概率的**加法定理**(the Additional Law)。它显然适用于所有我们能够以古典视角观察的试验：用袋子中的球作类比，这个定理可以被理解为抽中红或蓝球的结果总数是红球的总数和蓝球的总数之和。而且在任何可重复试验中——例如掷色子或者旋转轮盘赌轮——两个不相交事件的频率和一定是至少一个事件发生的频率。所以加法定理从频率角度看也成立。

一个持主观视角的人也能接受这个定理。否则，存在两个互斥事件，称为A和B，加法定理对它们不成立。这种情况下，主观主义者会面临三个赌局：一个赌事件A发生，一个赌事件B发生，一个赌事件A和B至少一个发生。而对他来说每个赌局都是公平的，他会接受，但他如果参与全部三个赌局，则必定会输钱。加法定理就禁止了这样的矛盾。

加法定理可以拓展到包含大量事件的集合中，前提是这些事件中任意两个都没有相同的结果——它们是**两两不相交的**(pairwise disjoint)。即使一个事件包含了1 000 000种不同的结果，其发生的概率也仅仅是每种结果单独的概率的和。但是假设结果的个数不再有限，例如连续掷一枚硬币直到正面出现时掷硬币的次数。

这个试验可能的结果组成一个无限长的表{1， 2， 3， 4，…}，表中的每一个数值都对应着其非零的概率。正面出现时掷硬币次数为

偶数的概率是多少？在{2， 4， 6， 8，…}中的结果会让这个事件发生。我们能通过计算其概率和来计算这个事件的概率吗？

数学上这个加法计算没有很大困难，但这个操作已经超出了古典概型的范围，古典概型只能处理**有限**多个可能的结果。这种**无限**长的表中的事件概率的加法定理是不是概率在起作用，人们对此还没有达成共识。有利于将其包括到概率的范畴的是：我们也许可以得到更多种类事件的概率。不利于将其包括到概率的范畴的是：对这个事件的概率计算不是古典理论的一部分，我们需要在计算中小心会带来陷阱的步骤。这个问题没有正确或者错误的答案。

我是一个实用主义者，我满意于加法定理这样的拓展应用，而且我从没有对这样拓展带来的结果失望过。这种态度是大学中讲授这门学科的大多数教材中给出的标准诠释。但是德·菲内蒂从谨慎的角度建议避免进行这种拓展，有一部分人也这样认为。

乘法定理

掷一枚普通的硬币，你会预期猜对正面或者反面的次数是总次数的一半。洗牌之后预测牌堆顶的牌是红色或者黑色，你也会估计猜对的次数是总次数的一半。当你同时猜测掷硬币的结果和牌堆顶牌的颜色，有多大可能两个**都**正确？

假设做这种双重试验100次。预期你大约50次猜对掷硬币的结果，当你猜对之后，预期你继续在一半的次数中猜对牌的颜色。这意味着你大约有25次两个都猜对，看起来得到了25%或者1/4作为两个都猜对的概率。对这样的试验来说，两个都正确的概率就是将两个事件单独成立的概率相乘。

10个大小和材质均相同的球被标有数字0到9，它们中的任何一个被抽中都是完全随机的。所以球上写着较小数字(0~4)或者较大数字(5~9)是等可能的。其中5个数字用绿色写成，另外5个用蓝色写成，所以绿色和蓝色也是等可能的。我们猜颜色或猜数字较小还是较大都分别有50%的概率。那么球上既标有较小数字又是绿色数字的概率是多少？

前文关于硬币和扑克牌的论断意味着答案是1/4，但是想一会儿你就能发现这不对。在有10个球的情况下，不可能其中的1/4(2.5个)是较小的绿色数字!正确的答案取决于哪一些数字是绿色的，哪一些数字是蓝色的。不妨假设1至5是绿色，其余是蓝色。

这种情况下，10个数字里面有4个(1、2、3和4)是较小的绿色数字，所以前述事件的概率是0.4。但就像我们处理第一个问题时一样，我们也可以使用两步走的过程：100次重复试验中，我们预期得到较小数字50次。5个里面有4个较小数字是绿色的，所以在我们得到较小数字的情况下，我们预测绿色次数占4/5。总体上讲，我们预期得到较小的绿色数字40次，再次指向了0.4这一答案。

在硬币和扑克牌的问题中，掷硬币的结果对扑克牌的结果没有影响。我们不会因为得知了硬币掷出正面就改变头脑中抽到红色牌的概率——在**给定**第一个事件发生的条件下，另一个事件的**条件概率**(conditional probability)就是它的正常的概率。如果这成立，这两个事件被称为是**独立的**(independent)，两个事件同时发生的概率就是两个事件单独发生的概率的乘积。

对于10个球的问题，两个事件同时发生的概率也作为乘积出现，其中第一个乘数是一个事件(较小数字)的概率，而第二个乘数是得到较小数字时得到绿色数字的条件概率。所以两个计算在形式上是相同

的，唯一的不同就是第一个事件的结果会否影响第二个事件。这两个计算中，我们都用到了概率的**乘法定理**(the Multiplication Law)：

两个事件同时发生的概率就是第一个事件发生的概率与第一个事件发生时第二个事件发生的概率的乘积。

独立性

我们用“独立的”这个术语来描述一种情况：第一个事件的发生并不影响我们对第二个事件概率的评估。假设这是成立的。但假如我们知道了**第二个**事件已经发生，这可能影响我们对第一个事件概率的评估吗？

不会。一个事件的发生与否不会影响另一个事件发生的概率，第二个事件是否发生也并不会影响第一个事件的概率。当两个事件中任何一个发生与否均不会对另一个事件的概率产生影响时，这两个事件是独立的。要计算两个事件同时发生的概率，就将它们各自的概率相乘。

彼此不相互影响的事无疑是独立事件，例如突尼斯今天下雨和巴黎新生儿的性别。但有时独立性并不明显。使用一个公正的普通色子，考虑事件“得到偶数”和“得到3的倍数”，它们的概率分别是1/2和1/3。只有得到6的时候两个事件同时发生，概率是1/6。因为1/2和1/3相乘等于1/6，这两个事件是独立的。得到偶数的概率并不会在我们得知是否得到3的倍数之后改变(反之亦然)。

现在，当你有一个8面色子或者10面色子的时候，考虑相同的问题，色子的每个面都分别被标记了1~8或者1~10。再进行相应的算术过

程：你会发现在其中一种情况下两个事件**是**独立的，但是在另一种情况下**不是**。判断独立性时，直觉是有用的，但是并不总是足够的。

在两个因素并不独立的时候假设它们是独立的，是评估概率过程中最常犯的错误。假设在一所大学的研究生院中一半的学生是女生，并且1/5的学生学习工程学科。随机选择一个学生：这个学生是女生的概率会被认为是1/2，这个学生学习工程的概率会被认为是1/5。然而你会发现这个学生是个女工程师的概率远远小于这两个值的乘积——1/10。

具有重叠的事件

加法定理说明了如何计算两个事件中至少一个发生的概率，只要这些事件是互斥的。如果二者不互斥会怎么样？例如，随机抽取一张卡片，是黑桃或者A的概率是多少？黑桃A同时属于这两个分类，所以如果我们只是将各自的概率相加，我们就会将黑桃A计算两次。为了计算两个事件中**至少一个**发生的概率，并纠正可能会被重复计算的结果，就将各自的概率相加，然后减去两个同时发生的概率。

如果两个事件是互斥的，就不可能同时发生，所以这个额外项的值为0，我们就回到了原先的加法定理。

让我们来看看在先前两个例子中这种观点的实际应用。在猜硬币和红牌黑牌的问题中，我们至少猜对一个的概率来自计算过程1/2 + 1/2 - 1/4，等于3/4。在另一个例子中，随机地抽取有标号的球，是较小数字或者绿色的概率是1/2 + 1/2 - 0.4 = 0.6。

抽到黑桃或者A的概率就是13/52 + 4/52 - 1/52 = 16/52，这可以被证明，因为52张牌中刚好有16张满足条件。

最后一个计算过程会警示你不要提前进行算术简化。的确13/52与1/4相等，4/52与1/13相等，但是要将1/4和1/13相加，你最好用原始的分数。将一个像5/13这样的美观的分数写成它的丑陋的近似小数(0.384 615 38……)几乎没有好处。

两个以上的事件

由许多事件组成的集合中的一些事件发生与否，并不影响任何其他事件的概率时，这些事件就被描述为独立的。在这种情况下，乘法定理意味着，以任何方式从这个集合中选取事件，它们均发生的概率仅仅是它们各自的概率的乘积。

但是在3个或者更多的事件**不是**独立的时候，我们如何得到它们均发生的概率呢？例如惠斯特纸牌和桥牌[1]游戏中，纸牌都被随机洗好然后等量地分给四个玩家。他们全都恰好抽到一张A的可能性有多大？

考虑4个单独的事件：甲恰好抽到一张A，乙恰好抽到一张A，丙恰好抽到一张A，丁恰好抽到一张A[2]。显而易见，这4个事件并不是独立的，因为如果任意3个均发生的时候，最后一个就一定发生了。我们会通过一个**三段式**过程来得到它们均发生的概率。

首先，我们计算甲恰好抽到一张A的概率。假设所有可能的分牌方式都是等可能的，我们有了一个计算练习：计算可能的分牌方式的总数，然后计算甲恰好抽到一张A的方式有多少种。相信我，这个概率算下来略低于44%。

假设甲只有一张A(因此有12张非A)。这就给其他的玩家留下了3张A和36张非A，同时乙随机地抽取了其中的13张。对这个较小的牌堆，相似的计数过程指出乙恰好抽到一张A的概率略低于46%。乘法定理告

诉我们，两个事件，即甲和乙均恰好抽到一张A发生的概率是这两个值的乘积，略高于20%。

现在我们假设甲和乙每个人都恰好有一张A。然后丙从剩下的两张A和24张非A中随机地取13张牌：他恰好得到一张A的概率是52%。

最后一步就是再一次使用乘法定理，将最后的两个计算过程组合起来：甲、乙和丙均恰好抽到一张A的概率略高于10%。如果这发生了，丁就不可避免地得到最后一张A，所以我们找到了我们所寻求的答案。

虽然分牌是随机的，最公平的A的分配却是颇不可能的。这个答案本身没有什么实际意义，但我们使用的**方法**是具有普遍意义的。为了得到集合中每一个事件均发生的概率，就应该将整个过程分成若干阶段。找到一个事件的概率；然后，**假设**这个事件真的发生，再找到第二个事件的概率；再**假设前两者**均发生，找到第三个事件的概率；然后**假设前三者**均发生，找到第四个事件的概率——以此类推。最终，将所有的数据相乘。

在其他情形中我们是否也要依照这个过程？假设我的交通行程有三个阶段，并且我可以评估它们各自不延迟的概率：但所有阶段都会受天气影响，并且一个阶段延迟与否会改变其他阶段延迟的概率。在制造业中，一件生产设备的安全依赖于数个并不独立工作的组件——其中的一些可能使用相同的供水系统，另一些可能由同一个不可靠的员工做了不充分的测试。在手术中，可能会出现问题的事情和其他事情是否独立会对全部过程能够顺利进行的概率产生巨大的影响。

如果事件是相互独立的，那么它们全都发生的概率就只是它们各自概率的乘积。但我们很少会足够幸运地处于这种情况，实际上在一个分段的评估中，随着工作的推进概率在不断变化，这是一种常态。

三个或者更多的事件中**至少一个**发生的概率是多少呢？加法定理的确可以拓展到这种情况，然而因为这个表达式实在是难以处理，我不会在这儿写下来。它的推导过程与之前描述过的将乘法定理应用于许多事件均发生时所使用的过程一样：一步一步来。

将**独立**当作**互斥**是常见的错误，反之亦然。随机地抽取一张牌的例子会帮助你认识到如何避免它。这里，“抽到一张黑桃”和“抽到一张梅花”这两个事件是互斥的，但绝不是相互独立的，因为如果二者之中任何一个发生了，另一个就不可能发生，所以两个同时发生的概率是0。同时“抽到一张黑桃”和“抽到一张A”是相互独立的(是吗？)，但显然不是互斥的。

记住：加法定理用来计算至少一个事件发生的概率，乘法定理用来推导它们全部发生的概率。

有时人们说：计数真的只有1、2、无穷大。这个说法揭示了一条真理，如果我们可以完成从处理一件事到处理两件事的过渡，那随后到第3、4、5等的过渡相比而言就不那么重要了。这个道理当然是对加法定理和乘法定理都成立的。

一个巧妙的把戏

任何事件要么会发生，要么不会发生。总概率被分成了事件发生和不发生两部分。所以如果我们能够找到事件不发生的概率，把这个概率从100%中减掉，就能够推断出它发生的概率。

举例来说，我们要计算掷两次公正的色子时，至少得到一个6的概率。任何结果都写成表示第1次和第2次的投掷结果的一对数字，例如(5， 2)或(4， 4)，并且我们认为所有这样的结果都是等可能的。每次投掷都会有6个可能的结果，以至于总共有6×6=36种结果。我们的

事件在没有色子是6的时候不发生，一共5×5=25种情况。没有6的概率是25/36，所以至少一个6的概率是11/36，比1/3略小一点。

这就引出了1654年布莱瑟·帕斯卡(Blaise Pascal)和皮埃尔·德·费马(Pierre de Fermat)解决的一个点数问题的初级版本。我们必须掷多少次色子才能使我们至少掷出一次6是**更有可能**的，即掷出一个6的概率比1/2大？我们刚刚看到，掷两次是不够的。

每一次额外的投掷都会让可能的结果数增加6倍，同时未掷出6的结果数乘以5。所以3次投掷一共有216种结果，而且其中125种(超过一半)没有包含6，3次投掷也是不够的。然而4次投掷得到1296种结果，并且其中只有625种是未掷出6的，少于一半。这就使得包括6的结果多于不包括6的结果，所以这时包含6更有可能。4次投掷就足够了。

实际上帕斯卡和费马分析的游戏中不只包含了掷一个色子，还包含了同时掷两个色子的情况；并且设问若要使**两个6**同时出现至少一次更有可能，需要多少次重复同时投掷两个色子。解法是相同的，但是原始的计算过程是艰难的。如今我们可以借助小型计算机或者袖珍计算器来快速地得到结果，然而直到17世纪，较便捷的对数和计算尺才被使用。直到第24次投掷没有双6产生的可能性都更大，但是第25次投掷就将扭转这一局面。

大多数具有“计算这些事件中至少一个发生的概率”这种格式的问题，都可以用这种方式解决：计算它们均不发生的概率，然后从单位1中减掉这个概率。

[1] 惠斯特纸牌(whist)和桥牌(bridge)均为经典的纸牌游戏。

[2] 原文中，甲乙丙丁分别为：Anne、Brain、Colin和Debby。

03 历史概要 Historical Sketch

开端

在1600年左右的佛罗伦萨，有一种关注三个普通色子总点数的游戏。所有色子掷出1(即总点数为3)和所有色子掷出6(即总点数为18)这两种情况出现得最少，其他大多总点数都接近于这个范围的中间值。你应该能发现得到9点有6种方法(例如6 + 2 + 1、5 + 2 + 2，等等)，得到10点也是有6种方法。通常认为，这就“应该”使色子总点数为9和10出现的频率一样。但是一段时间之后，玩家们注意到总点数为10出现得比9明显多。他们就此向伽利略(Galileo)请教一个解释。

伽利略指出他们计数的方法有缺陷。将色子涂成红色、绿色和蓝色，并按涂色的顺序列举出结果。从3 + 3 + 3得到总点数为9需要三个色子具有相同的点数，只有一种方式能够使其发生，(3， 3， 3)。但是5 + 2 + 2的组合可以通过(5， 2， 2)、(2， 5， 2)或(2， 2， 5)中产生，所以这个组合倾向于出现得比前者频繁3倍；6 + 2 + 1通过(6， 2， 1)、(6， 1， 2)、(2， 6， 1)、(2， 1， 6)、(1， 6， 2)和(1， 2， 6)产生，所以这个组合有6种途径产生。一个合理的寻求不同总点数出现频繁程度的方法需要考虑这种因素，而且这种因素确实使得获得10点比9点有更多的方式。佛罗伦萨的赌徒们(Florentine gamblers)学习了关于概率的重要一课——一定要学会**正确地**计数。

1654年夏天，帕斯卡(在巴黎)和费马(在图卢兹)就**点数分配问题**(the problem of points)进行了一次通信。假设史密斯和琼斯约定进行一系列的比赛，首先赢得3局的是获胜者；但不幸的是，当史密斯领先琼斯的比分为2：1时比赛必须中止。该如何分配赌金？

那时这样的问题已经被提出了至少150年了，仍没有令人满意的解答，但帕斯卡和费马各自独立地找到了一个解决方案，对任意的目标得分和任意的比赛意外终止时的比分，都能够在两人之间**公平**地瓜分赌金。他们使用了不同的方法，但是得到了相同的结果，两人都对对方的才华表示赞赏。对上述具体的问题，应该按照3∶1的比例分配，史密斯得到3/4的赌金，琼斯得到1/4的赌金。

他们解法的关键是假设在未来的任何对局中两个玩家获胜是等可能的。他们就每一个玩家计算了能够使其获得最终胜利的可能的假想对局结果数量，并提议按照这两个数量的比值分配赌金。换句话讲，假设两人在接下来的游戏中旗鼓相当，赌金应该按照每个玩家在经历一系列对局后最终取胜的概率来分配。对概率的系统研究由此拉开了序幕。

这个问题能被概率的客观方法解决，但是帕斯卡考虑得更多。他提出了一个有关上帝是否存在的赌局。“上帝存在，或不存在，缘由无法回答。在无限远的彼岸掷一枚硬币，正面或者反面就要出现。你赌哪边？”

他提出，如果上帝存在，相信或者不相信带来的不同，就是在天堂获得无限的幸福与在地狱忍受无限的痛苦之间的区别；如果上帝不存在，相信或者不相信只会带来尘世生活中细小的差别。所以一个不可知论者应该强烈倾向于相信上帝存在。

在这个赌局中，“正面”或者“反面”出现的概率大小是具有个人色彩的选择，不能从对称性抑或计数证据中推导出。所以帕斯卡也是概率的主观方法的先驱者。

瑞士的伯努利家族

在17世纪和18世纪，来自巴塞尔的伯努利家族[1]的成员在数学(包括概率)领域取得了重要进展。家族内的竞争起到了鞭策作用：他们中的一个会提出难题，另一个就会回应，难题的提出者会说他发现了所谓的解决方案中的瑕疵，等等。

关于概率的游戏激发了许多对概率运作的早期关注。在这些游戏中，无论是掷色子、发牌，还是掷硬币，一些“试验”会在本质上相同的情况下被重复进行。之前提出过一个自然的问题：一个结果被观察到的概率和客观概率有什么关系？

雅各布·伯努利(Jacob Bernoulli)在其遗作《猜度术》(The Art of Conjecturing)中，用他的例子巧妙地进行说明，给出了一个答案。假设罐子中60%的球是白色的，其余的是黑色的，随机抽取一个球。伯努利证明，只要试验抽取至少25 550回，每一次试验中抽到白球的比例落在58% ~ 62%的范围**外部**，就会有至少1000次试验中抽到白球的比例落在这个范围**内部**。不规范地说就是：在多次抽取的条件下，我们观察到白球的频率会压倒性地倾向于接近它的客观概率。

类似的分析过程适用于任意能在相同条件下不限次数地重复的试验，一个试验的结果不会对其他试验结果产生影响。每一次试验中，某些特定的结果代表着成功，它们的客观概率是一个固定的值p。这个概念现在被称为**伯努利试验**(Bernoulli trials)。在p这个值附近取任意区间，你愿意它有多小就有多小(±2%或±0.1%，都无所谓)。然后给出你想要让成功的频率落在这个区间内部比落在其外部高多少(100倍还是100万倍，怎样都行)。伯努利的方法证明了只要试验重复足够多次，**任意**这样的要求都会被满足。观察到的频率会像你期望的那样尽可能地接近于客观概率，只要给出充足的数据。这个断言被称为**大数定律**(the Law of Large Numbers)。

在1975年，一个主要致力于促进概率和数理统计发展的国际学会被命名为“伯努利学会”，以向这个家族致敬。

亚伯拉罕·棣莫弗

亚伯拉罕·棣莫弗(Abraham de Moivre)以胡格诺派[2]难民的身份在英国定居，依靠国际象棋和他的概率知识谋生。艾萨克·牛顿(Isaac Newton)那时已经50多岁而且事务非常繁忙，为了岔开有关数学的咨询，他说：“去找棣莫弗吧，他比我对这些事情了解得更清楚。”棣莫弗的《机会的学说》(Doctrine of Chances)在1718年以英语出版，1738年的第二版包含了伯努利工作中的主要进展。了解他的成就要思考一个具体的问题：如果一个公正的色子被投掷1000次，我们能合理地预期数字6的产生频率与平均频率之间有多大偏差吗？

棣莫弗提出了一个对这类问题具有广泛应用的公式。他高超的洞察力表现在，他意识到数字6的实际数量与期待的平均数量之间的偏差，可以用投掷次数的**算术平方根**来进行最适当的描述。

如何夸大这个发现的重要程度都不为过。当你听说一个民意测验(opinion poll)中一个政党的支持率是40%，它经常会附加一个暗示，这只是一个估计，真实的支持率“非常可能”在一个范围中，比如38% ~ 42%。这样的区间宽度告诉你最初数字40%的精确度，而如果你想要更高的精确度，就需要更大的样本：这个平方项意味着要将精确度变为2倍，样本需要扩大4倍!我们有一个“报复式”的收益递减法则——要达到原来的2倍效果，我们必须投入原来的4倍精力。

棣莫弗的方法可以用考察一个公正的硬币投掷20次时有多少次正面朝上来说明。基于所有的例如正正正反正……正反正反这样的，长度为20的序列都是等可能出现的，我们可以绘制出图1。其中垂直条的高度表示大约100万种序列中有多少个恰好包含0、1、2、……19、20

个正面。这些数字各自的客观概率就正比于这些高度。棣莫弗证明了经过这些竖条顶点的最佳拟合的光滑连续曲线非常接近于一个特别的形状，现在通常称之为**正态分布**(normal distribution)。

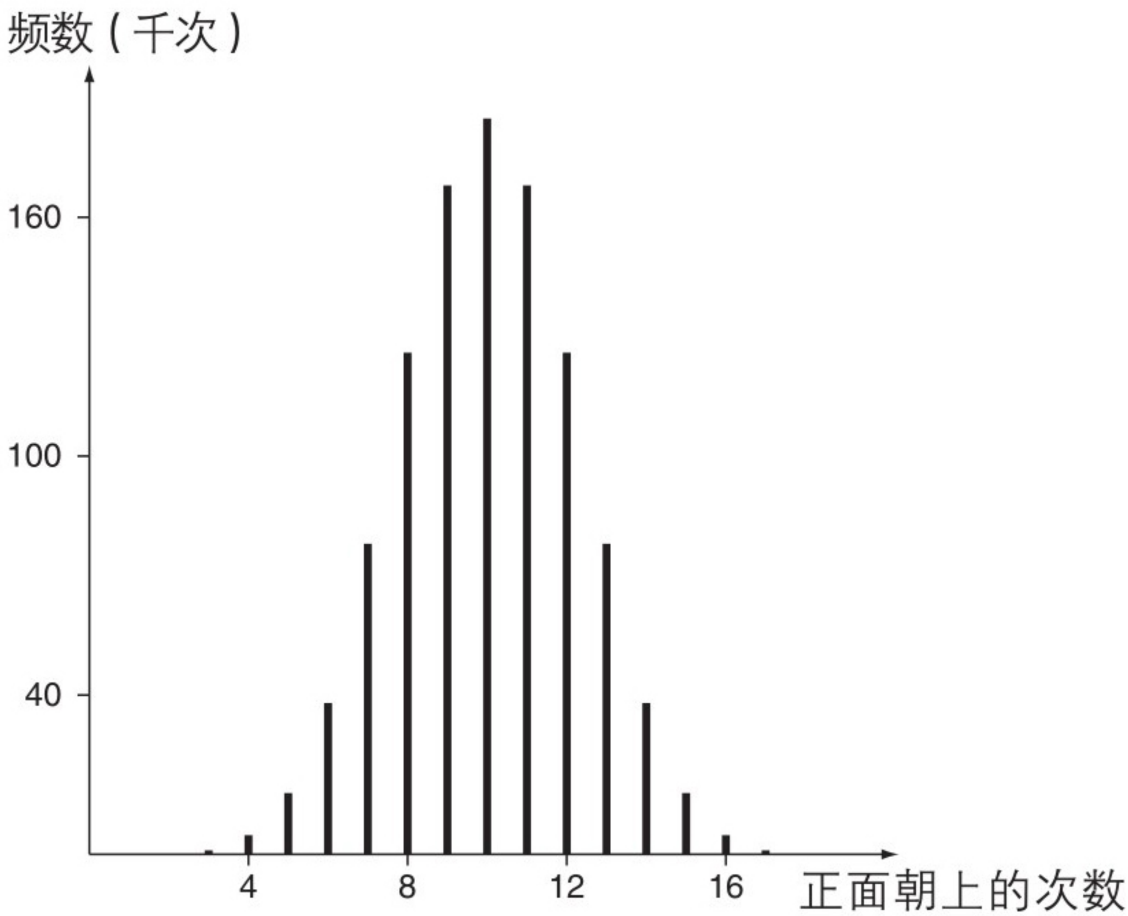


图1　20次投掷中正面朝上的相对频率

这种曲线会生成于所有的多次掷硬币过程中，并且还可以包括掷出正面的概率不等于1/2的情况。所有曲线之间有一个简单的关系，所以棣莫弗可以就一个基本的曲线制作一个简单的数表，并能在任何情况下使用。整体的成功频率在一个确定的限制范围中，现在就能够简单地获取对这样的事件的发生比例的估计——需要的仅仅是获胜的概率和将要进行的试验的次数。将一个公正的色子掷200次，你想知道数

字6出现次数在30~40间的可能性有多大吗？或者一个公正的硬币在100次投掷中掷出60次以上的正面的可能性有多大？没问题——棣莫弗有解决方案。

假设我们知道一群人死亡时的年龄，所有人都活到了至少第50个生日。棣莫弗的工作可以回答这样的问题：“如果一个50岁的人在70岁之前死亡是更有可能的，我们能够观察到这些数目的各种变化的可能性有多大？”虽然这十分有用，但是它不能回答新兴的人寿保险业提出的关键问题：“我们有多么确信一个50岁的人在他70岁之前即死亡是更有可能的？”

逆概率

托马斯·贝叶斯(Thomas Bayes)是一个在数学领域有建树的长老会牧师，他的思想现在比在其生前更受重视。他的《机遇问题的解法》(Essay Towards Solving a Problem in the Doctrine of Chances)在他死亡三年后的1764年出版，给出了初步处理主观概率的一般方法和从数据中推断概率的保险精算师问题的一个解决方法。这本书也包含了一个处理概率的重要工具，被称为**贝叶斯法则**(Bayes' Rule)。

为了举例说明这个法则，设想我们掷一个公正的色子两次。已知第一次掷色子点数是3，很容易地就能够得到总点数是8的概率，因为这个事件会在第二次投掷点数为5时发生。我们不假思索地就能给出解答为1/6。但是将问题调转一个方向，设问：给出总点数是8，第一次掷出3的概率是多少？答案远远不那么简单了，但是我们可以应用贝叶斯法则来得到结果。在掷色子的标准模型下这个概率为1/5。

对于刑事审判中处理证据的方法，**逆概率**(inverse probability)这个概念至关重要。假设在犯罪现场找到的指纹被鉴别

为属于一个已知的人——史密斯。如果史密斯是无罪的，发现这个证据的概率很可能是非常低的。但是法院判决的依据不是“已知史密斯是无罪的，发现这个证据的可能性有多大”而是“已知发现了这个证据，史密斯无罪的可能性有多大”。贝叶斯法则是获得答案唯一合理的方法。我们将会在后面的章节中看到这个法则是如何帮助我们做出正确决定的。

贝叶斯展示的洞察力在很多年中被忽略了，但是他的确指出了中心问题：如果在一系列的伯努利试验(例如掷色子)中，成功的概率是未知的，但是试验和成功的次数都分别是已知的，这个不可知的概率落在指定区间内的可能性有多大？而另一位极其优秀的数学家拉普拉斯的计算优于贝叶斯。

从1774年试探性的开始到1812年的理论综合体，拉普拉斯逐渐地完善着他的分析，并最终给出了解答贝叶斯问题的一系列明晰的公式。例如，利用巴黎男性和女性的出生人口数目，他得出结论，毫无疑问男性出生的概率高于女性——他估计这结论错误的概率是10^{-42}。

贝叶斯被安葬在伦敦的邦丘原野公墓(Cemetery of Bunhill Fields)，在皇家统计学会(the Royal Statistical Society)附近。其墓地曾经被修复过，来表达全世界统计学家对贝叶斯的敬意。

中心极限定理

将一些伯努利试验的结果写成由胜利(Success)和失败(Failure)组成的序列，例如FFFSFFFSSFSFF……现在将每个S用数字1代替，每个F用数字0代替，得到0 001 000 110 100……这表明了一个巧妙地理解这些试验中胜利的总数的方式：序列中的这些数字的**和**(同意吗？)。棣莫弗利用他所谓的正态分布曲线，给出了一个描述这个和的分布的良好近似方法。

对于一个巨大的数值序列，我们要考虑的可能只是其中随机变化的各个值的和。例如，负责垃圾处理的政府部门主要感兴趣的是整个城镇中的垃圾总量，而不是来自每个家庭的数量。当一位园丁播种红花菜豆时，他关心的不是每个豆荚的大小，而是总产量。一个赌场基于它的全部赢得的钱来评估其经济收益，不论个别赌徒的收益如何。将着眼的事物看作大量随机数据的和，这经常是卓有成效的。

拉普拉斯拓展了棣莫弗的工作以便能涉及像这样的情况。他建立了**中心极限定理**(Central Limit Theorem)，该定理说明了在很多情况下，大量随机数据的和是棣莫弗的正态分布的理想近似状态。我们不需要某个单独数据如何变化的细节，**整体数据**变化的模式会紧密地贴合这个正态法则。

为了利用这个想法，我们只需要两个数字：第一个是全体数据的平均值，第二个是一个简单地表示它的变化程度的数据。知道这两个数据，任何一个概率都能够从棣莫弗的表格中找到。

下面谈到卡尔·弗里德里希·高斯(Carl Friedrich Gauss)，他是和牛顿、阿基米德(Archimedes)并列的顶级数学天才。当时他正在研究如何处理观测恒星和行星位置时产生的误差。他提出，平均而言误差为0——观测中向左偏差和向右偏差是等可能的——**并且**误差大小遵循这个正态分布，他因为这个方法在数学上很简洁而使用它。但是拉普拉斯看到高斯的书时，将这个结果引用到了自己的书中，同时提出，因为一次观测中的**全部**误差以许多随机因素堆积总和的形式出现，这样误差**应该**遵循正态分布法则。高斯蹩脚的理由“数学上的便利”被拉普拉斯更加有说服力的“数学理论表明……”所替代。

“正态”这个术语应用在这个分布上是不恰当的。它暗示我们应该预期我们遇到的任何数据都遵循这个形式，但是远非如此。为了避免这种暗示，并且为了纪念一个伟大的人，我们将会转而使用另一个

术语**高斯分布**(Gaussian Distribution)。如果你可以说服自己，你关注的值可以貌似可信地被当作大量随机的较小数字的和，这些较小的数字的来源都是不相关的，那么这个中心极限定理告诉我们可以预期这个值遵循高斯分布。

观测中的误差真的遵循这个规律吗？亨利·庞加莱(Henri Poincaré)——对数学各分支具有全面知识的最后一位数学家——说：“人人都相信它，因为数学家误以为这是观测中的事实，而观测者认为这是个数学原理。”

西莫恩·德尼·泊松

西莫恩·德尼·泊松(Siméon Denis Poisson)出名是因为一个含有他名字的**分布**——概率在一个平均值周围变化的方式。在物理学家欧内斯特·卢瑟福(Ernest Rutherford)及其同事的工作中——计算7.5秒长的时间间隔内有多少个α粒子从放射源中发射——相关的例子出现了。这个数字从0到十几不等，平均值稍小于4。图2展示了两个经典的实验，说明(在这些情况中)有四五个粒子发射。卢瑟福认为这些发射事件都是随机的。

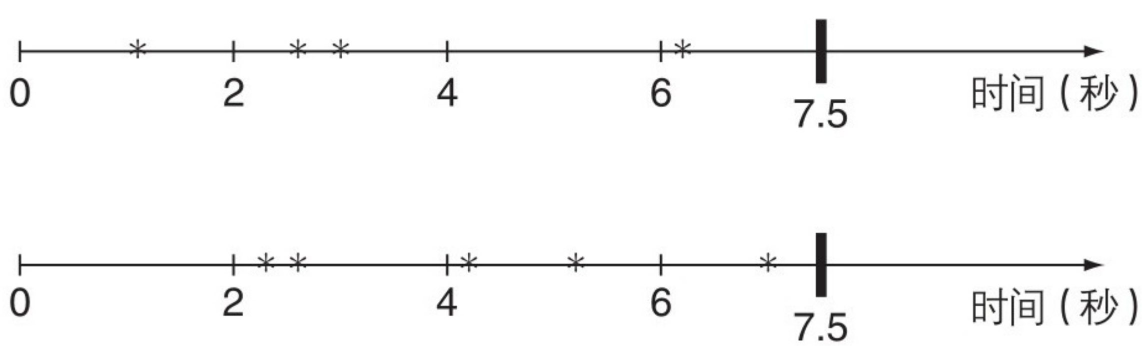


图2　α粒子的发射时间

将7.5秒切成极多个极小的时间间隔，小到我们可以忽略其间发生一次以上发射的概率。除了几个发生了一次发射，其他所有的小间隔都没有出现发射这一事件。在各个小间隔之中，将一次发射事件看作游戏胜利，所以粒子发射的总数就是胜利的次数——又是伯努利试验。

极小间隔胜利的概率实际上和其长度成比例，所以随着单个间隔长度的缩小，间隔的数目增多，每一个间隔发生胜利事件的概率减小。泊松计算出了小区间长度缩小至0的过程中，发生0、1、2……次发射的全部的确切概率。

这个**泊松分布**(Poisson Distribution)就经常出现在我们计算事件“随机”发生概率的时候，至少是以一种良好近似的方式。它恰当地描述了卢瑟福的实验数据；它适用于第二次世界大战中投掷在伦敦南部不同地区的炸弹数量；它看起来是一本书中每1000个字中错印字数的有效模型。如果你同时从两个洗好了的牌堆发牌，正面朝上，平均而言你会发现发牌恰好有一次是一样的，但是实际的匹配次数会非常接近于泊松分布。有一个可怕的例子，在一项长达20年，涉及几代普鲁士特兵团学员的追踪调查里，被自己的马踢死的军官数量也满足这个分布。

所有这些例子都符合一个相同的模式：大量的机会，每个机会中事件发生的概率很小。每当你正在研究的现象符合这种模式，泊松分布就很可能对它有用。

俄罗斯学派

一个数学定理具有这样的特征：如果某一个假设是成立的，那么一个预期的结论就会产生。我们主要的兴趣在于应用这个预期的结论，所以所需的假设最好不要太复杂。有时候预期的结论只存在于非

常具有限制性的假设的情况下，或者极其难被证明：之后的研究者也许会找到使用相同假设的更简单的方法，或者在较少的限制性条件下得到相同的结论。最好的情况是结论在非常宽松的假设成立时，能被简短而优美地证明。巴夫尼提·切比雪夫(Pafnuty Chebychev)的工作给出了这种理想情形的良好案例。

切比雪夫展示了如何在更广泛的情况中应用大数定律。最初的大数定律和伯努利试验有关，它描述了在一系列试验中事件发生次数的比例能够多么合适地用于估计事件发生的概率。如果我们想估计入伍士兵的平均身高，或者一个家庭一周的消费，我们似乎可以很明显地从相关人员中抽取合适的样本。但是这种估计有多合适呢？切比雪夫的工作给出了误差足够小以使得估计可靠的概率。

很多统计结果都是这些想法的应用。

切比雪夫最知名的学生是安德雷·马尔可夫(Andrey Markov)，马尔可夫的教学启发了又一代极有天赋的俄国人。马尔可夫将他的想法应用于诗歌和文学作品。在将亚历山大·普希金(Aleksandr Pushkin)的《叶甫盖尼·奥涅金》(Eugene Onegin)中的元音字母(vowel)和辅音字母(consonant)分别替换为字母v和c之后，马尔可夫得到了一个只含有这两个字母的序列。在原始的基里尔字母[3]中，元音字母占有43%的文字比例。在一个元音字母之后，另一个元音字母出现的频率是13%，而在辅音字母之后，元音字母出现的频率是66%。在预测一个字母之后的字母是元音还是辅音的过程中，他发现，已知当前的字母时可以忽略它前面所有的字母，因为它们基本不构成影响。

这个“可以忽略”的特性广泛存在。有一些例子：赌徒手中的一系列连续的赌金数额；特拉维夫(Tel Aviv)每天的天气(干或者湿)；在每一个顾客离开时队列的长度；连续世代的基因组成；两个相连的

容器中的气体扩散过程。如果知道序列中的前一个值，要预测随机变化序列的下一个值的时候，我们都可以忽略更前面的那些值，那么这个序列被称为具有**马尔可夫性质**(Markov property)。描述这种序列的理论已经很好地发展起来了，这些理论也是许多概率成功应用的基础。

马尔可夫在政治上非常活跃，对数学史也知之甚详。1913年，俄国政府组织了罗曼诺夫改革300周年的庆祝活动，马尔可夫相应地开展了对伯努利发现第一个大数定律200周年的纪念活动。

这里我偏离一下主题，讨论一下在20世纪早期，法国人埃米尔·博雷尔(émile Borel)的工作。回想伯努利试验中的大数定律：在大量试验中，实际事件发生的频率有极大的特别接近于事件发生概率的可能性。但是这仍然留下了这样的可能：在无限次的试验中，实际事件发生的频率极其偶尔地会落到事件发生频率附近任意给定的公差带之外。但是博雷尔的工作完全消灭了这种概念难题。任意给定公差带，一定会有一个时刻(我们不知道什么时刻，但是的确有那么一个时刻)，在这个时刻之后，实际事件发生的频率就会永久地停留在公差带**内部**。这被称为**强大数定律**(the Strong Law of Large Numbers)。

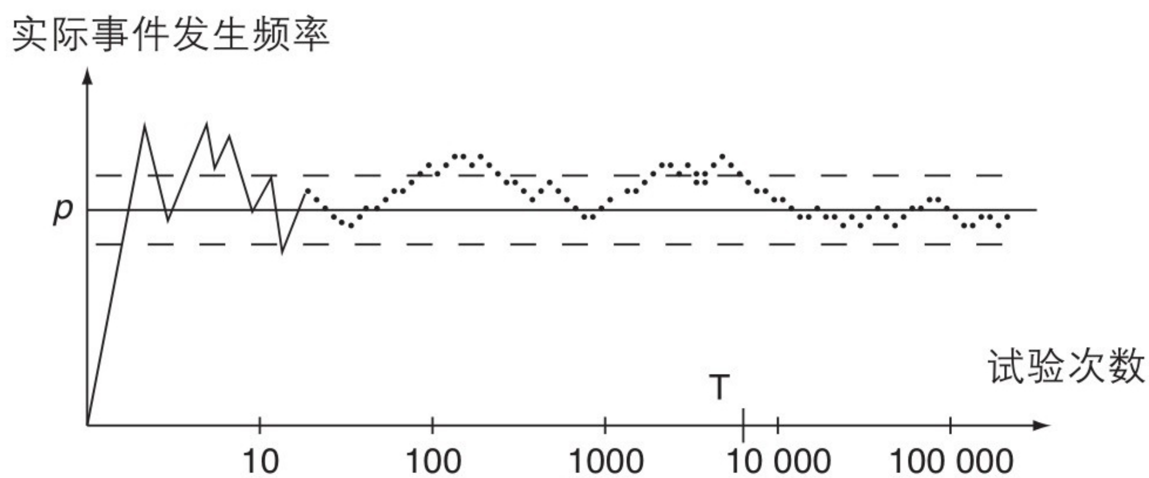


图3　强大数定律图示。p是事件发生的概率，虚线表示公差带。在T次试验之后，实际的事件发生的频率永久地停留在公差带内部

这个强大数定律也会延伸到更广阔的情境中。我们可以将大数定律的含义浓缩到一个不正式的习语中：

从长远来看，平均统领一切。

在1924年，亚历山大·辛钦(Alexander Khinchin)发表了命名极佳的《重对数律》(Law of the Iterated Logarithm)。就像伯努利和拉普拉斯的早期工作一样，这个理论应用于一个以和的形式出现的随机数量时，它可以给出更加精确的有关这个和会多么接近其平均值的信息。

在大约300年中，概率工作的前沿进展都是来自一些特别的方法。然而到了1933年，杰出的苏联科学家安德雷·柯尔莫哥洛夫(Andrey Kolmogorov)使用了近期发展出的**测度论**(measure theory)中的理念，将概率科学确立在了令人满意的逻辑框架中。所有已知的理论都可以重新容纳进柯尔莫哥洛夫的设定中，并能给出精确程度作为后续计算过程的催化剂。

柯尔莫哥洛夫和辛钦还有他们的学生鲍里斯·格涅坚科(Boris Gnedenko)一起极大地拓展了拉普拉斯关于随机数量和的工作。提高纺织业和其他制造业中机器的可靠性的方法，生产线上产品的质量控制，还有解决拥堵导致的问题都是他们研究的动机。

柯尔莫哥洛夫是一个卓越的研究者和教育者。他在1987年去世，当时的苏联总书记米哈伊尔·戈尔巴乔夫(Mikhail Gorbachev)还调整了自己的工作日程，以便能够出席葬礼。

更近的近代

战争经常会驱动科学发展。1939—1945年的世界冲突促进了运筹学的发展，其中许多成就都起源于对概率论理念的合理利用。为了使一艘补给船不被敌军潜艇击沉的概率最大，一系列数据的组合和计算给出了舰队比单艘船更好，大舰队比小舰队更好的结论。当这个结论被付诸实施，损失就显著地降低了。布莱切利园(Bletchley Park)的密码解读工作现在大概已经广为人知了。然而我们经常忽视贝叶斯公式在确定恩尼格玛密码机(Enigma machines)卷轴最有希望的布置方式中的应用。

在1950年，威廉·费勒(William Feller)出版了一本介绍概率的著作，并在1957年和1968年再版。这本书是我心目中有史以来最好的非虚构著作。凭借着直觉和严密论证的结合，这本书直接地或者间接地引发了人们对这门学科的巨大兴趣。随后，乔·杜布(Joe Doob)使用术语“**鞅**”(martingale，这个词原本指每次损失后将赌金加倍的投注策略)来描述那些在未来某时刻的平均值与现在的值(大致上)相等的随机量。他揭示了鞅的主要性质并给出了一些密切相关的概念：这些工作普遍有用，因为它指出许多有实用性的随机量都会包含在这个理论研究的范畴中。之后我们会举例说明概率这个概念是如何在一系列领域中被有效地应用的。

许多专攻概率的学术期刊已经发行，其中一些还产生了分支期刊，从来没有听说它们缺少可以发表的材料。现代计算机的算力已经转变了计算概率的模式：它们的运算速度和存储容量极大地拓展了可解决问题的范围。在早期，许多概率问题只受一个因素影响，比如说时间或者距离，人工的精确计算总是可能的；而现在，那些概率随时间、空间的三维和其他因素的影响而变化的复杂问题也已经被成功解决了。

即便如此，使沟通交流更便利才是计算机对概率论发展最巨大的影响。TeX[4]语言已经成为数学和许多科学写作的标准语言框架。研究者在互联网上发布他们的想法和观点，学术文章可以在家中或者办公室中通过万维网(the World Wide Web)轻易地取得。

[1] 伯努利家族(The Swiss Family Bernoulli)来自瑞士的一个商人和学者家族，有很多艺术家和科学家出自其中。

[2] 胡格诺派(Huguenot)， 16世纪至17世纪法国基督新教归正宗的一支教派，1685年被法王路易十四宣布为非法。

[3] 基里尔字母(Cyrillic alphabet)又称西里尔字母、斯拉夫字母，是在使用斯拉夫语族语言的大多数民族中通行的字母书写系统。

[4] 正式标志为TEX，中间的E有点下沉，但在无格式纯文本中写为TeX。

04 概率试验 Chance Experiments

对于以概率作为结果的任意试验——买彩票、投注赛马、相亲、接受医学治疗，我们用**分布**这个词来详细说明其所有可能的结果，以及与它们相关的概率。我们讨论泊松分析——大量重复试验中多少稀有事件会发生——的时候提到过这个词。

“分布”是分析概率试验中结果变化范围的中心概念。坦率地说，我们需要知道可能的结果的范围。为了给出这些结果概率的合理数值，我们必须讲清楚我们的假设，并且期望它们对于我们想要考察的试验是合适的。

离散分布

首先，我们来看看那些可能的结果能够被写成一个列表的情况，每个结果都有它们自己的概率。术语**离散分布**(discrete distribution)适用于这种情况。

最简单的情况就是我们认为结果具有相同可能性时计算结果的数量。这里使用**均匀分布**(uniform distribution)这个术语，因为总体的概率均匀地分散在各个结果上。许多试验都被认为满足均匀分布——轮盘赌、掷色子、扑克牌、选择彩票中的中奖号码等。精确的计数给出了合适的答案。

术语“伯努利试验”描述了一系列发生概率均为常数的独立试验。在伯努利试验次数固定的情况下，有一个简单的公式叫作**二项分**

布(binomial distribution)，分别给出了事件发生恰好0、1、2……次的概率。这个公式只依赖于试验的次数和事件发生的概率。当你依次浏览这些结果的时候，它们的概率先是升高到一个最大值，然后逐渐降至0。泊松分布也遵循这个模式。

我们能计算20次掷色子中数字6出现次数的二项分布；或者一个学生对30道多选题中的5个选择随机作答的时候，蒙对个数的二项分布。但是我们**不能**预测一个桥牌选手的13张手牌里梅花张数的概率：虽然每一张单独的卡片都有1/4的概率是梅花，但是连续的牌不是独立的，因为下一张牌是梅花的概率会被所有前面的结果影响。

永远要留意(通常是小号字)附属细则。使用二项分布需要3个条件：固定试验次数，每个事件与其他事件相互独立，并且事件发生的概率是常数。

在一系列伯努利试验中，事件首次发生的时候经过了5轮试验的概率是多少？这种情况发生的唯一方式就是前4次试验中事件都未发生，随后1次试验中事件发生；因为所有的试验都是独立的，问题的答案就是将这些结果分别的概率乘在一起，给出了一个令人愉快的简介表达式，这就是所谓的**几何分布**(geometric distribution)。

事件首次发生需要恰好1、2、3……次试验的概率稳定地下降。下一个概率值等于将现在的概率值乘以一次事件未发生的概率，一个小于单位1的固定值，每次都是这样。因此，无论事件发生的概率是多大，事件首次发生时经过的试验次数的最可能的值就是1。

假设在板球比赛中，连续的击球构成了伯努利试验。一位投球手，将事件发生理解为他投球成功，他可以乐观地想：他开始投球的时候，下一次投球成功最可能的就是这一次；相反，一个具有相同视角的击球手就得听天由命地接受他这一局最有可能的持续时间就是他

面对这一个球的时间。（就算是最好的击球手，记录表明他们最有可能的总得分总是0!）

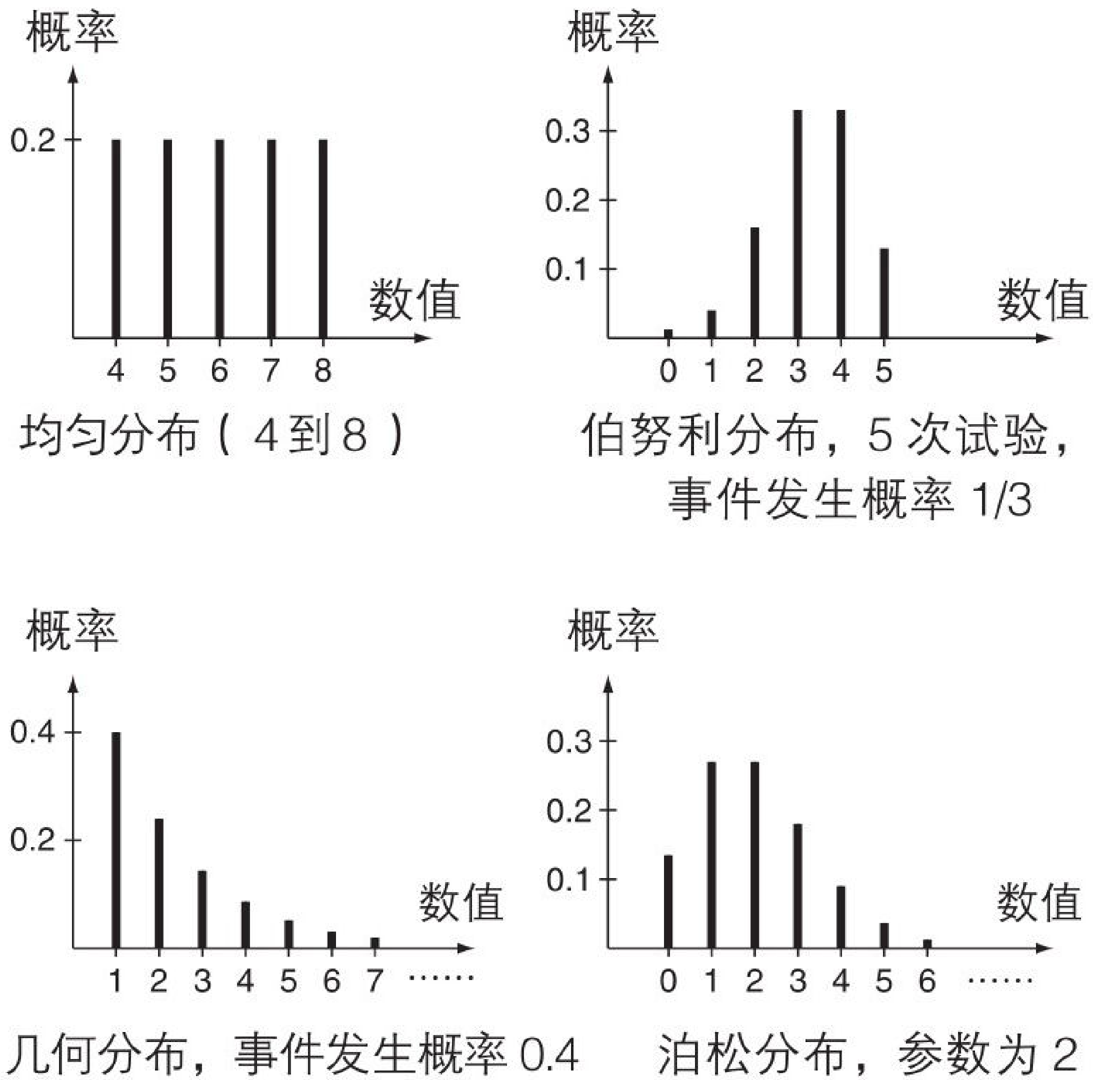


图4　一些常见的离散分布

图4举例说明了一些常见的离散分布。对于每一个可能的数值，竖线的高度给出了它的概率，并且这些高度的和总是单位1。

连续分布

我们如何拓展古典的概率观点来解决在一个长度为80cm的木棍上随机选取一个点的试验？可能的结果组成一个**连续统**(continuum)，而不只是一个列表。

“随机”意味着所有单独的点都具有相同的概率值。但如果这个相等的值超过了0，那么，在取了足够多的点之后，它们的总概率就会超过单位1，这是不可能的。每个单独的点的概率一定是0，我们也不能使用像图4一样的图片了。我们需要将概率、片段或者区间相关联，而不是将概率和单独的点相关联。

为了对80 cm的木棍的每一部分一视同仁，所有具有相同长度的片段一定有相同的概率。想象一下将木棍砍成8个相等的片段：按照定义，一个“随机的”点落在每个片段上的概率一定相同，举例来说，落在20~30 cm的片段上一定具有1/8的概率。

图5a展示了下一步操作，这可以用口头禅“面积表示概率”表述。标注了h的水平线的高度是设定好的，这条线下阴影部分的面积是单位1，这呈现了一个事实，我们可以百分百地确定随机点落在区间0~80 cm中的某处。接着图5b展示了如何确定随机点落在32~52 cm的片段上的概率，只需要计算对应的阴影面积即可。简单地说，这个概率是1/4。

要得出随机选择的点落在木棍两端10 cm内，或者中间20 cm内的概率，我们就可以使用图5c，并且依据加法定理，要求的概率是三个阴影面积的和，也就是1/2。

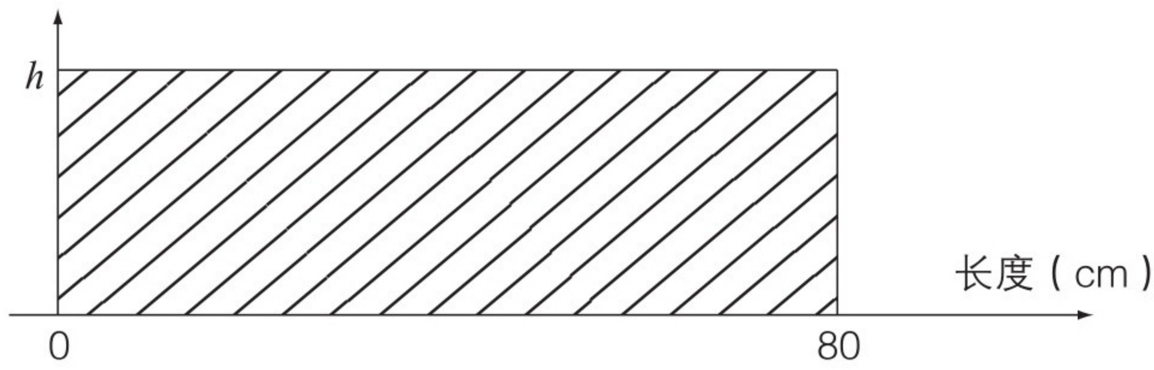


图5a　阴影面积是单位1

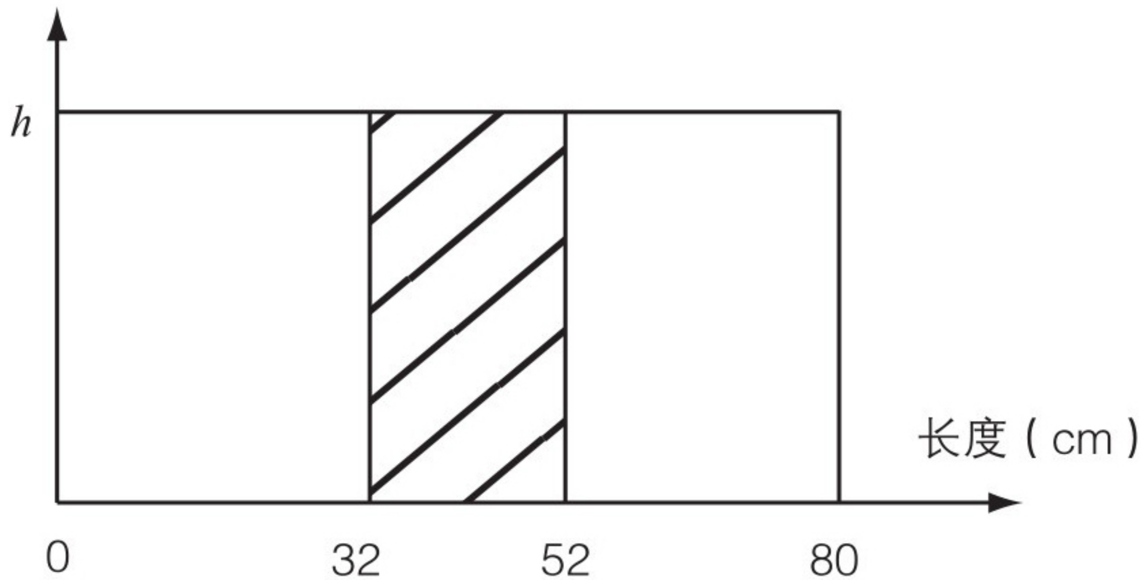


图5b　落在32~52cm之间的概率是1/4

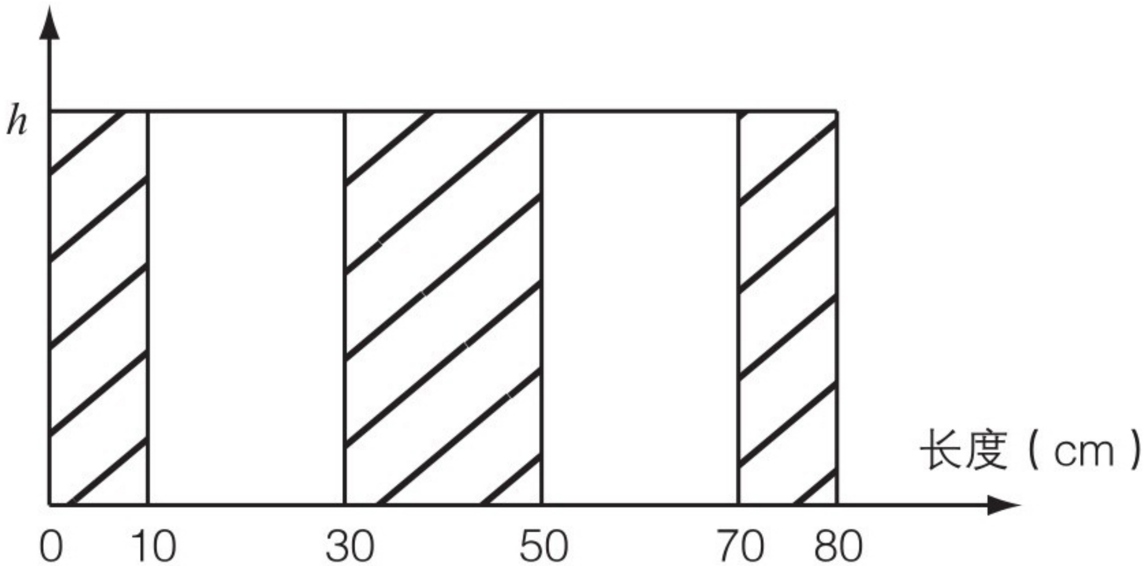


图5c　见正文

图6展示了对结果取连续值的另一些情况下相似的解决方式，例如一段特定的高速公路上下一次事故发生需要的时间。我们会在下面论证展示图片上的曲线在这种情况下是合理的，但核心观点是图线的尺度是特意选择好的，以至于标注了“时间”的直线以上，和以点E为起始端点的曲线以下的总面积是单位1，因为我们可以百分百确定我们考察的这段时间一定取非负值。

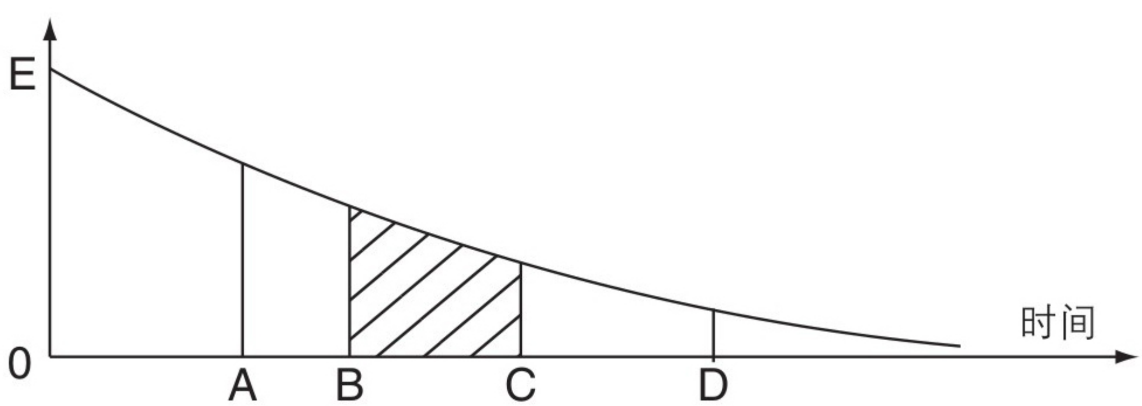


图6　连续分布

时间至少是B但不大于C的概率就是阴影的面积。我们可以用类似的方式得到考察的时间落在任意给定区间内的概率，还能像之前一样根据加法定理，得到落在更复杂区间内的概率。

一个能按照这种方式生成概率的曲线被称为**概率密度**(probability density)。已知面积的计算方式是“长乘以宽”，任何直线的宽度都是0。因此图6中在点A或者点D的竖直线的“面积”都是0，所以这两个单独的点具有0值的概率，就像之前提过的那样。但是点A的密度曲线比点D高，所以点A**附近**的值比点D附近的值更可能。简单地说，图片表明具有或高或低的概率的区域。在这里我们使用**连续分布**(continuous distribution)这个术语。

在所有这些试验中，因为单独一个点具有的概率值为0，我们可以稍微草率一点：无论一个区间包括了两个端点或一个端点，抑或都不包括，结果的概率都是一样的。

为了限定一个概率密度，一条曲线一定必须具有两个特性：不能取负值，在曲线下的全部面积必须是单位1。这些保证了对概率的所有计算能得出合理的结果。

许多概率密度函数出现得足够频繁以至于可以被赋予名称。对于从给定的一个区间内选取随机点，密度函数在这个区间内完全平直，就像图5中的一样：简单地说，所有相同长度的片段具有相同的概率。再一次，我们叫它**均匀分布**。

假设我们对一些特定事件在多长时间后发生感兴趣。例如，^{210}Pb是一种铅的不稳定同位素，“它的半衰期是22年”这个断言被印在物理教材上。它的意思是，如果我们有一块这种物质，22年后只有原来的一半保持原样，其余的都通过辐射衰变成其他物质了。

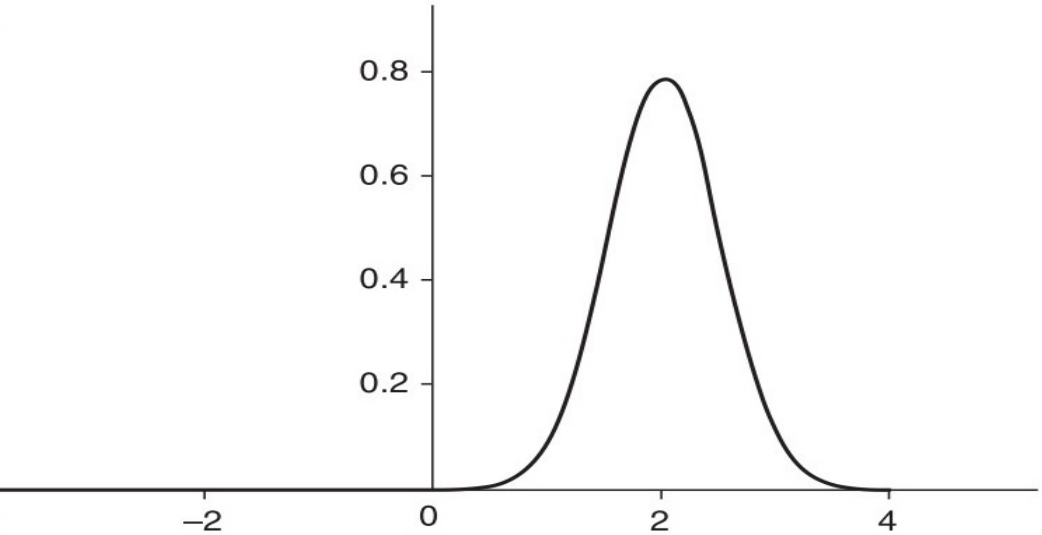
这块物质由巨量的原子组成，所有这些原子的行为都是独立的。如果关注单个原子，它通过放出一个粒子而衰变。我们不知道什么时候这个过程会发生，但是因为在22年内这块物质中的一半的原子都衰变了，所以这个**特定的原子**在这个时间段内发生衰变的概率是50%。假设它在5年后还没有发生衰变：这时，它就是剩余的^{210}Pb块中的一个原子，所以它在未来22年衰变的概率也是50%，并且如果它在接下来的3年中没有发生衰变，情况也一样，以此类推。

一个给定的原子的衰变时间只有遵循所谓的**指数分布**(exponential distribution)的时候，上述情况才能发生，它的概率密度的一般图形展示在图6中，曲线的高度按照确定的比率下降。类似的场景也应用在交通事故中：过去的一周内没有事故发生，那似乎不可能对未来的事故概率产生影响，所以我们预期交通事故下一次发生的时间也遵循指数分布。

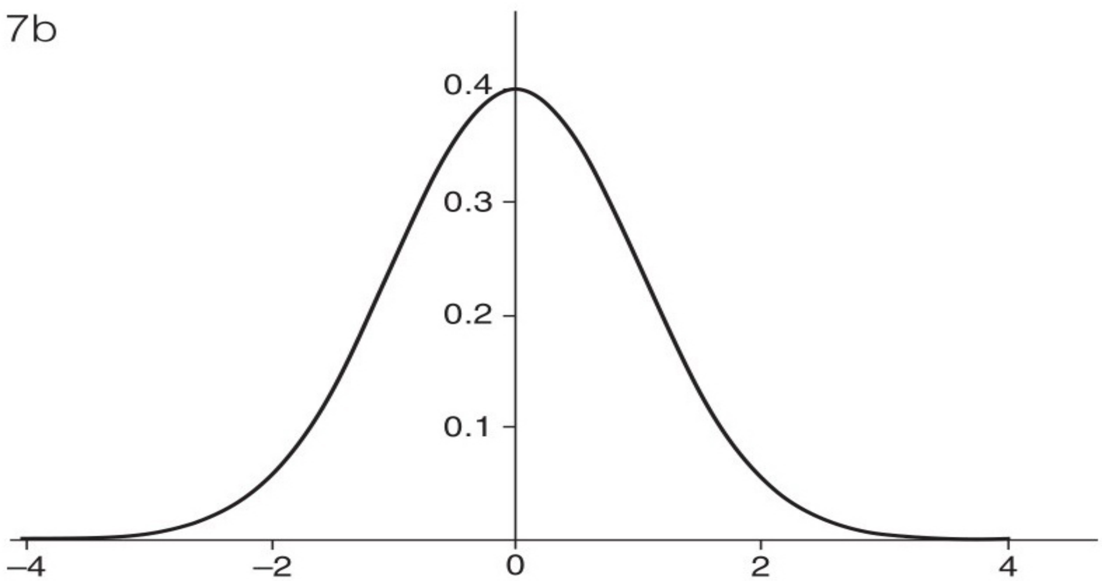
这个分布和泊松分布密切相关。只要事件本质上是随机发生的——暴风雨中的闪电、细胞复制中的自发突变、顾客来到邮局——在固定的时间段内这样的事件发生的**数目**倾向于遵循泊松分布，那么一对事件之间的**等待时间**的概率分布就具有这种指数形式。

最重要的连续分布是我们已经提到过的**高斯分布**。就像图7展示的那样，这个分布家族中的成员关于单独的一个峰对称，并且在两边快速下降，然而永远不会达到0。两个数字就可以告诉我们任意一个分布实例在这个家族中的归属：一个数字表示峰的位置，另一个数字描述散布程度——较小的散布值导致像图7a那样的高且窄的图形，较大的散布值给出像图7c那样的矮且宽的图形。这个家族成员的任何位置的概率都可以借由这两个数字来与图7b的分布相关联而得到，这个分布的峰在0，标准散布值为单位1。棣莫弗创制了对应数表之后，这些对应关系就很容易得到了。

7a



7b



7c

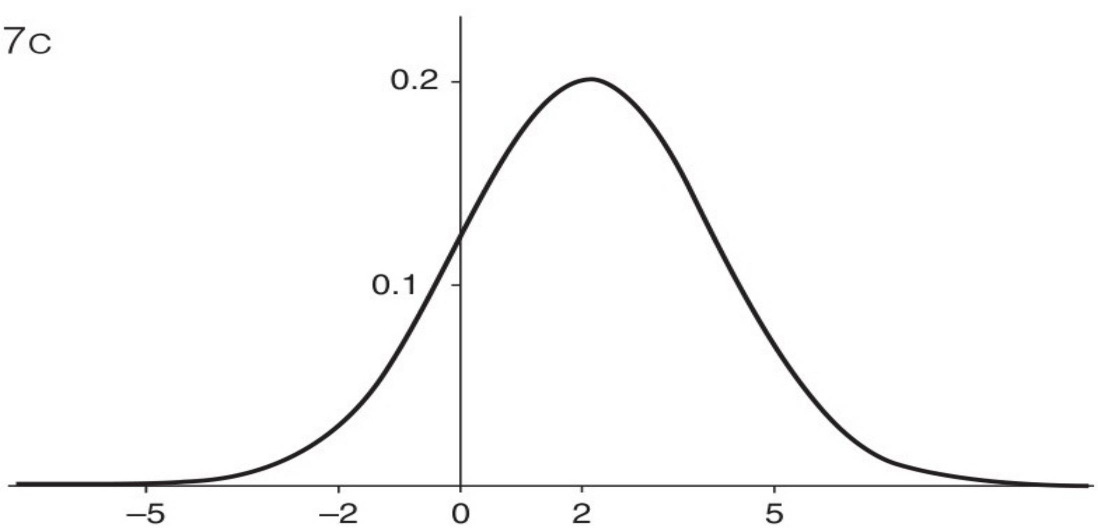


图7　高斯分布

一个问题的再解决

你也许已经注意到了一个问题。已知期望的结果组成的集合是有限的，或者是一个像{1，　2，　3，…}这样的无尽的集合，那么即使这个集合中的一些成员的概率是0，任何概率是0的事件也不会发生。然而，对于**连续分布**，即使每个单独的点的概率是0，它们其中的一个在试验进行过程中也是**会发生**的!我们不再能够认为“不会发生”与“概率为0”具有相同的意义。

为了解决这个问题，我们来考虑从装有100万块完全相同的大理石的盒子中随机选取一块。只有在提前猜对了结果的情况下，我们才会感到惊讶，因为猜对的概率只有一百万分之一。但是，无论抽中了哪块大理石，虽然的确出现了某个概率只有百万分之一的结果，但我们也不会感到惊讶。

把盒子做大一点——10亿块或100 000亿块大理石——实际产生结果对应的概率可以无限接近0——但是它的确发生了。这与在一条连续的线上选取一个点的过程并没太大区别：对于任意的点来说，它的概率是0，但是它们其中的一个的确将会发生。

我们接下来开始说明，在一个可重复试验中，如果猜对结果的概率是1/6，我们可以期望按顺序进行的6次试验中有1次猜对。将事件发生的概率除以100万，我们预期等待正确结果出现的次数就被乘上了100万。具有极小概率的结果的确会发生，但是越来越罕见。

如果概率下降到0，我们可以预期要等待比任意有限长都要长的时间——那它就是不会发生!在**提前指定**的情况下，认为任何概率为0的事件都不会发生是合理的。

平均值

已知一个概率试验中结果的分布，我们就可以计算我们想要的任何概率。但是有些时候，所有的这些细节都成了障碍——只见树木，不见森林[1]：所以我们想要提取出分布的主要特征。

举例说明，假设可能出现的结果只有2、3和7，分别对应概率60%、10%、30%。我们预期在100次重复试验中，2这个值会出现大约60次，3出现大约10次，7为剩下的30次。所有这些数值的和是120 + 30 + 210 = 360，所以所有这100个结果的平均值是360/100 = 3.6。这个值就是数值2、3和7的**加权和**(weighted sum)，权重就是它们的概率。

无论我们有什么样的分布，相似的计算都会得出大量重复试验结果的集中趋势。“集中趋势”是一个宽泛的词，对于这类计算的结果，我们更喜欢使用**平均值**(mean)这个术语。可能有一些捷径：如果值在一个范围内均匀地分布，平均值就在两个端点的正中间；在一系列伯努利试验中事件发生的次数的平均值就是试验次数和事件发生概率相乘。

掷一个公正的色子，得到4的概率是1/6。所以在600次投掷过程中，我们应该可以得到大约100个4：简单计算表明，连续出现的4之间的平均间隔是6。大小为1/6的概率导致平均间隔是6，这不是一个巧合。任何间隔的长度就是下一个事件发生的等待长度，所以在一系列伯努利试验中，我们就有了令人愉快的结果：

等待一个事件发生所需的平均时间是事件发生的概率的倒数。

在连续分布中，想法是类似的，但是加权和是由一种名为**积分**(integration)的数学方法来得出的。对于高斯分布，峰处就是平均值。按照整体特定的频率发生的随机事件发生的平均时间是一个指数分布：平均时间就是频率的倒数，这并不奇怪。

除了“平均”和“平均值”，术语“期望”和“期望值”也会被使用。掷一个公正的硬币12次，“期望”正面朝上的数目为6；掷一个公正的普通色子，“期望”得到的分数是3.5。当然正是因为一次投掷中反面朝上的期待值是0.5，我们实际上不能**期望**得到一半的反面朝上!文字都很是奇妙。

平均值非常友好：和的平均值通常是平均值的和，无论不同的和是不是独立出现的。大数定律告诉我们，从长远看，平均值占据主导：如果你买一张彩票花费1英镑，其中一半的钱都进入奖池中，那么，无论奖金的分配结构如何，你的**平均**收益都是50便士，从(非常)长远看，这就是你能得到的。

离散程度

用一种简洁的方式描述一个分布的离散程度通常是有用的。我们可以计算每一个值和平均值的差值，然后得到这些差值的(适当地加权的)平均值。但是，就像所有计算上努力展示的那样，这种方法是不成功的：负的差值不可避免地抵消了正的差值，最终结果总是0。

但是无论一个差值是正是负，我们都可以在将它平方之后得到一个正的值。所以我们可以通过将这些平方值加权来得到离散程度。得到的这个值就叫作**方差**(variance)。如果分布集中在平均值的附近，那么方差就会比较小；当有合理的原因使一些值距离平均值比较远的时候，方差就会变得很大。

当考虑以美元计收入的分布的时候，平方值的单位就是“平方美元”，不管它究竟是什么意思。将方差取算术平方根就可以得到原始计量中的单位，这就得到了**标准差**(standard deviation)。

平均值和标准差合在一起，经常能给我们理解一个概率分布的快速且有用的方式。在高斯分布中，仅用这两个数字就足够计算所有的概率!就像点金石一样，当分布为高斯分布的时候，在大约68%的试验中，结果在平均值周围1个标准差范围内；在超过95%的试验中，结果在平均值周围2个标准差范围内；而400次中大约有一次结果是在3个标准差之外。

这些数字就是在第1章中给出的，我们能够有理由地期望事件发生的概率和事件发生的实际频率有多么接近的相关参考的基础：关键就是中心极限定理，它说明了作为大量随机成分的和而出现的数量预期接近遵循高斯分布。

在图7中，展示了3个高斯分布的概率密度函数，这几张图的平均值分别是2、0和2，标准差分别是1/2、1和2。

但是注意：虽然和的平均值总是平均值的和，但是对方差和标准差来说可不是这样。如果和的组成部分恰好是独立的——比如说在拉斯维加斯一家赌场7天分别的收益——那么和的方差就的确是分别的方差的和，否则就会偏高或者偏低。直接将标准差相加几乎不会给出任何有意义的结果。

极端值分布

在概率的某些应用场景中，我们关注的重点在于随机数量的最大值或者最小值。例如，线或者电缆的强度依赖于最弱的纤维；洪水防护设施要考虑的是下一个一百年中预期发生的最大规模的洪水；**生存**

分析(survival analysis)这个学科调查一段给定时间后的剩余人口。极端事件可能很少发生，但是当它们发生了的时候，结果就变得很重要了。

最简单的看似可信的模型假定存在一些独立的随机变量，每个都分别遵循一个特定的分布。例如每一年中，对一家保险公司的索赔。对接下来的50年中它可能会收到的最大的总索赔额有多大，保险公司有一个经历了漫长的数学推导的可用结果：无论在每一年中索赔额如何变化，在很大的年代跨度中，最大索赔额一共只有三个可能的种类，它们被称为极端值分布，具体的名字是弗雷歇(Fréchet)、冈贝尔(Gumbel)、韦布尔(Weibull)。有一个合理的数学原理，如果有一个关于最大值的理论，就一定有一个相对应的关于最小值的结果。所以如果感兴趣的东西是最小值，也存在相似的结论。

能够对这三种分布进行一些限制是非常有帮助的。通过估计极端值的平均值和方差，从三种分布中选择一种最接近于实际数据的，就能计算分布中的另一些概率的合理估计，比如真实情况中极端和破坏性事件的概率。

[1] 原文为“we can t see the wood for the trees”。

05 理解概率 Maki ng Sense of Probabi l ities

这里我会介绍在我们面对不确定性问题的时候，概率的观点是如何起作用的，还会描述一些误解了概率的情况。

赔率？

回忆之前我们说过，概率可以用赔率(odds)的形式来描述，反之亦然：概率是1/5和赔率是4比1是相同的。不幸的是，“赔率”这个术语也已经被赌徒们篡改成了一个相当不同的意思——如果你选择的马胜出了，你获得的赌资。所以当你听说“海都之星(Sea The Stars)”在2009年德比大赛(Derby)[1]中以11比4的赔率胜出，这只是说明了对于押在这匹马上的每4英镑赌注，它如果赢了收益就是11英镑。“11比4”没有与胜利概率之间的必然联系。这个比例只依赖于赌徒对于马胜出概率的主观评定。对于像“11比4”这类的表达，术语“赔付金额(pay out price)”是一个更准确的语言使用方式，但遗憾的是，我们不得不接受在赌博情境中对于“赔率”一词的滥用。

如果两边都没有经济收益上的优势的话，赔付金额就被称为“**公平的**”。也就是说，赌注的平均值是0。从洗好的牌堆中正确地抽出指定花色牌的公平赔付金额是3比1，这也就是抽中的赔率。

因为商业赌博不会在没有庄家收益(house advantage)的时候运行，就是说，商业赌博都是不公平的。对一家英国赌场的轮盘赌来说，所有的37种结果都是等可能的，赌其中一个数字的赔付比例只有35比1，而不是公平的36比1。所以一次37英镑的赌局回报的平均值只

有36英镑，这就得出庄家收益率——对于任何赌局庄家都预期会赢——为1/37，大约2.7%。

这个优势对大多数你能想象得到的轮盘赌局都是一样的：无论你是赌一对数字，还是3个、4个、6个、12个，对于你下注的每37英镑，你的平均回报总是36英镑。在拉斯维加斯，标准的庄家收益还会更大一些，因为一个多加的格子，双零(double zero)会给出38种结果，但是赔付金额和在英国时相同。38美元的平均收益通常是36美元，庄家收益率是2/38，或者说5.3%。

赌马的另一种形式是通过同注分彩系统。这里，押在所有的赛马上的赌注被放入彩池中，以固定比例——常见的是80%——被押了获胜马匹的人按照他们下注的比例分享。无论哪匹马获胜，分彩收益率就是20%。

赛马中，赌注登记人(bookies)的收益的大小取决于哪匹马获胜了。虽然赌博者在单次比赛中都既可能收益也可能损失，但最近的数据给出了一个令人清醒的事实：在赔付金额为6比4的时候，赌博者预期会损失他们赌本的10%；赔付金额在5比1的时候，预期损失大约13%；在10比1的时候，平均损失超过了23%；如果你投机地定下50比1的赔率，预期你会损失2/3的钱。

这个现象被称为**热门冷门偏差**(favourite-longshot bias)。赌博者在更受欢迎的马匹上输钱会比高赔付金额的输钱**更慢**。2009年，蒙莫梅(Mon Mome)在英国国家赛马障碍大赛(the Grand National)中以100比1的赔率获胜之后，博彩公司都很高兴。**绝对风险，还是相对风险？**

假设在一群特定的人中，未来50年内大肠癌的发病概率是1/1000。我们预期在10 000人中，会有10个人罹患大肠癌。一种新的

药物会将这个概率减小至1/2000：意味着如果药物投入使用，10 000个人中只有5个人会因癌症身亡。医药公司就会将新闻稿加上“患癌风险降低50%”的大标题。这的确是精确的：对于每个人来说，风险的确会减半。

这个计算方法描述了**相对风险的降低**，并且经常会因为对数据的过于乐观的解读而饱受诟病。假设初始的风险是1/10 000 000：将它减小50%得到新的风险数字1/20 000 000，但是在前后两种情况中，这个风险是相当的小，以至于在10 000个人中，我们预期两种情况给出相同的患者数字——0。尽管风险被减半了，药品也几乎不会产生什么作用。

但是假设这群人患癌的风险是比较高的，比如说40%。一种能将概率减少到20%的药物不但有资格使用那个大标题，甚至会被称为重大突破，因为在10 000个人中，患癌人数整整减少了2000人。

考察绝对风险而不是相对风险的变化，通常是更有意义的。在上面的第一个情景中，绝对风险从0.1%变成了0.05%，降低了0.05%；在第二个情景中，降低了一个极小的数字0.000 005%；而在最后一个情景中，降低了令人印象深刻的20%。

一个合理的描述方式是陈述为了阻止一个恶性病例的发生，平均有多少个病人应该使用药物——**需治疗人数**(the Number Needed to Treat)或者说NNT。这个数字越小越好，它也就是绝对风险的变化的倒数。在上面的例子中，NNT分别是2000、20 000 000和5。

为了阻止一个恶性病例的发生而去治疗20 000 000人，很难说这是正当的。NNT，连同对治疗费用和疾病影响的严重程度的了解，让我们能做出关于分配卫生保健资源的合理决策。

组合小概率

在大量均具有微小概率的事件中，有多大可能至少有一个发生？在考虑灾难发生概率的时候，这会是一个相关的问题。如果无数组件中的任何一个失效，复杂的系统或者机器就会失效；两架飞机会相撞吗？一座核电站的堆芯可能会熔毁吗？所谓的**波莱尔-坎泰利引理**[2](Borel-Cantelli Lemmas)给出了一些提示。数学结果告诉我们，在很多情况下，重要的是这些微小的概率的**和**：如果这个和是无穷大的，那么灾难一定会发生。

结果就是我们永远不会满意于当前的安全标准。持续进行改进是必不可少的。因为无论我们的安全标准有多高，在某一个月中，一定会有不为0的失效概率：而且无论这个值多么小，如果它保持不变(或者它降低得比较慢)，大量月份的概率和就会变得无穷大，灾难一定会在某一时刻发生。

一个持续改进的充足计划也不会保证一定能避免灾难，但是满足于现状就是坐以待毙。

一些误解

(a)当一个医生告诉患者一种特定的药物有30%的概率会引起令人不快的副作用，他的意思是他预期有30%服用了这种药的患者会遇到副作用。然而，这个患者也许会相信当她服用这种药物的时候，在30%的次数中有副作用产生。医生考虑的是他看到的所有患者，这个患者考虑的是她服药的次数——他们的**参照类别**(reference class)是不一样的。

(b)公众会如何解读电视天气播报员说的“明天芝加哥下雨的概率是30%”？播报员们预期他们的听众会进行频率解读，即长远地看，如

果天气情况类似于现在看到的那样，在30%的情况下，第二天会下雨。

但是当被问及此事的时候，即使在那些对于“30%概率”这个说法感到满意的电视观众之中，解读也是各种各样的。一些人认为他们被告知城市中的降雨区域超过30%；其他人认为芝加哥会在一天中30%时间中下雨；还有一些认为30%的气象学家预期会下雨!一小部分人认为一定会下雨，30%这个数字意味着降雨的强度。播报员们提及的事件和听众们理解的事件会有很多不匹配。

(c)如果随机选择至少23个人聚集在一起，其中有两个人生日相同的概率就会超过1/2。人们第一次听说这个事实的时候通常会感到惊讶，但是在计算证明了这个断言是正确的之后，他们会被说服。然而，一小部分人仍然不会被说服，因为他们以为自己被告知，如果另外22个人被聚集起来，其中的一个和**他们**生日相同的概率超过1/2。听仔细!

(d)假设有一枚被认定是公正的硬币，连续9次投掷中均反面朝上。或许是受到了某些要求正面立即掺和到连续出现的反面之后的“平均法则”的影响，一些人就会宣称下一次投掷几乎确定地是正面朝上。这样的法则并不存在。大数定律的确会暗示正面和反面朝上的比例是相等的，但在**长远的**情况下，任何的连续9个反面的序列都会被前面或者后面的1000次投掷抵消。

或者一些人会(正确地)指出出现连续10次反面朝上的概率小于1/1000，所以当他们看到9次连续的反面之后，可能就会“推断”下一次也是反面是非常不可能的。但是这是伪逻辑。即使我们有连续的9次反面朝上，第10次反面朝上的概率也是1/2。这是事件的绝对概率和条件概率之间产生的混乱，它也在1996年的一个著名事件中体现出来：赛马骑师弗兰基·戴图理(Frankie Dettori)在雅士谷(Ascot)赢得了前6场赛马，因为从来没有人会赢得一天之中的7场赛马，戴图理赢得

最后一场几乎“必须”是不可能的。但是他的确赢了。极少有人会在7场比赛中的前6场均胜利，但是一旦有人做到了这一点，他可能也会赢得最后一场。

问你自己：我是在评估20件事情发生的绝对概率，还是已知前19件事已经发生的情况下，评估第20件的条件概率？

(e)报纸经常会在巨大的时间压力下出版发行，所以一些文章包含有荒谬的内容也容易理解。但是有3个报道是需要批评的。

当从超市买来的6个鸡蛋全都是双黄蛋，报纸就会宣称一件令人震惊的事件发生了。1000个鸡蛋中，只有1个有这样的特性，所以得到一整盒的双黄蛋的概率就是连乘6次后的一个非常小的值了。这个数字结果是如此小，以至于如果你每秒打开一个盒子，你预期要用300亿年才会遇到一个只装有双黄蛋的盒子。

但是这个计算只有在6个鸡蛋都是从双黄蛋比例是1/1000的大量鸡蛋中被独立选择而来的时候才有意义。但是事实并不是这样。在装入盒中之前，鸡蛋都是被按照尺寸分类了的。有些盒子上甚至被标上了装有双黄蛋的标签……

针对某一个英格兰足球队长私人生活的指控曝光，一名记者对于球队经理的4种可能的反应给出了“可能性”数值：

(a)将他从球队中开除——1/10；

(b)留在球队中，但是要求他辞职——3/10；

(c)留在球队中，但是废除队长职务——6/10；

(d)不采取行动——8/10。

这4个估计的概率中的任何一个单独都是貌似可信的。但是因为它们都关联着互斥的结果，它们的和就不能超过单位1。然而这些“概率”加在一起是1.8。

第3个报道中说，在英国国家彩票中赢得至少50 000英镑的人的名字，大多包括了约翰(John)、大卫(David)、迈克尔(Michael)、玛格丽特(Margaret)、苏珊(Susan)和帕特里夏(Patricia)。目前为止看起来还好，但是因此应该在购彩票集团中引入有这些名字的人就是很荒谬的了!

描述未知

从一个洗好了的牌堆中抽出一张牌，我预期红牌和黑牌是等可能的，我可以很自信地将红牌的概率赋值为“1/2”。使用一个稍微令人费解的方式，我会说我将抽到红牌的概率的**分布**百分之百地分配给1/2——百分百显示了我的信心。如果我确定概率是一个确定的数字，那么我就将这个数字附加上百分之百。

但是我经常无法选择单独一个数字：我对错过中转火车的最佳估计可能是3/4，但是2/3和4/5也貌似是可信的，甚至接近于极限值0和单位1的值也不是完全不可能的。我可以用一个0到单位1的连续分布来描述我对这个不确定概率的感觉。

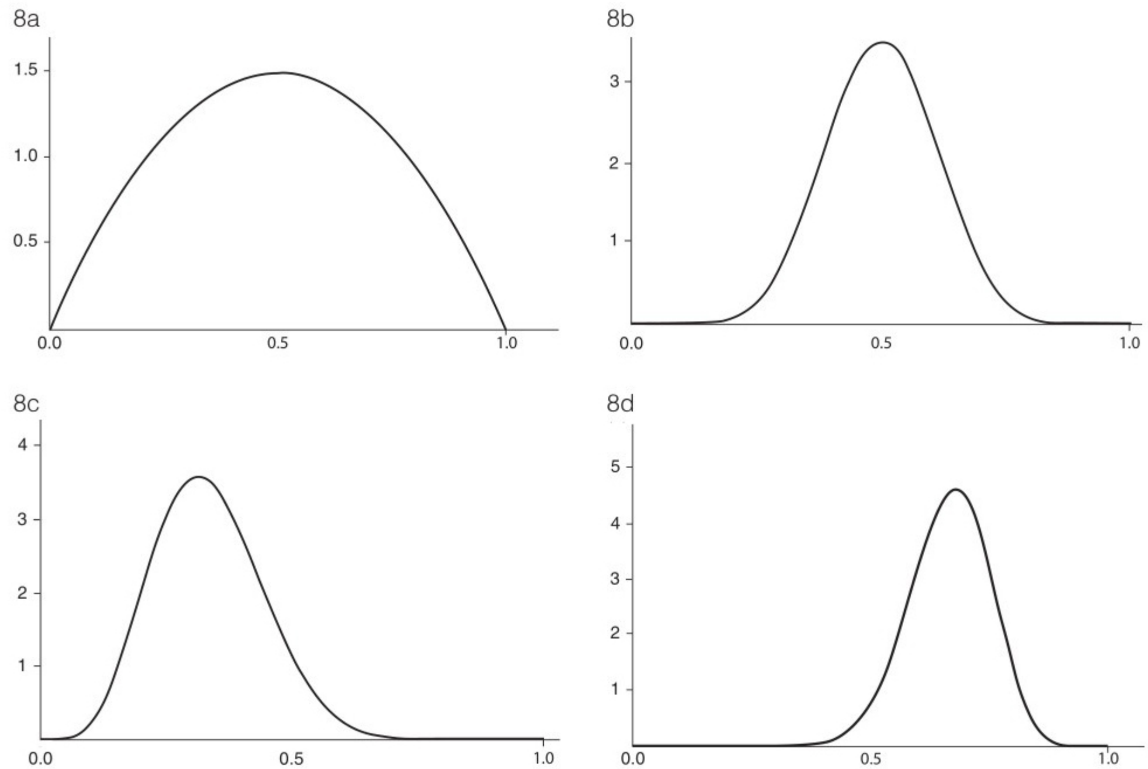


图8　贝塔分布的一组图像

完全未知的概率——非常罕见的情况——可以被图5中的连续均匀分布描述。更常见的情况是，首先有一个中间值，它是你以单个数字对这个概率给出的最佳猜测，然后你会直觉地认为更大和更小的值的分布会降到0。图8展示了具有这种性质的被称为**贝塔分布**(beta distribution)的一组图像。

图8a是我们对概率的值知之甚少的情况，我们将1/2当作最可信的值，但是像1/5一样小和4/5一样大的值也是很有可能的；在图8b中，我们对这个值接近于1/2更有信心了，同时也不排除极端值；在图8c中，我们最相信的值接近于1/3；图8d中的分布强烈地集中在了2/3周围，而且对于这个值小于1/2的期望是非常低的。

我的车加了10升油的情况下能开多远？当预热之后匀速行驶的情况下，我预期能开90英里，但如果我在几周内进行过几次短程行驶，就更可能是60英里了。在这两种情况中，这个距离都会有一些不确定性，可以用连续分布来表示。为了找出是哪种分布，想象一下将这些汽油分成10毫升一杯的小份。一共就会有1000个这种杯，总的距离就会是这1000杯中每一杯行驶的距离之和。回忆中心极限定理，大量随机量的**和**趋向于遵循高斯分布。

对于恒定速度的高速公路行驶，我会选择中心位于90的高斯分布，方差小所以峰很尖。对于在城镇中的情况不定的行驶，我也会用高斯分布描述，但是中心在60而且分布比较宽来显示较大的不确定性。

效用

有位仙子让你进行一次选择。她给你1英镑，或者她掷一个硬币，如果你猜对了正反面，她就会给你10英镑，否则你什么都得不到。你会做出什么选择？

备选的选项是确定得到1英镑或者一个公平的赌局，回报是5英镑。几乎所有人都会更喜欢后者，但是将问题中的金额都乘上1 000 000，人们的选择就必然会变化。比起什么都得不到与得到500万英镑概率对半来说，确定能够得到100万英镑，更有吸引力。这种差别背后是**效用**(utility)这个概念起作用。

对较小的总金额，翻倍**的确**意味着翻倍，但是如果100万英镑会为你和你的家人带来相当程度的快乐，将这个值翻倍就不会使快乐翻倍。无论你选用什么单位来计量货币金额的“价值”，通常效用和金额之间的关系曲线的形状会像图9所示：图线总会上升，一开始像一条直线，但后来增长得越来越缓慢了。

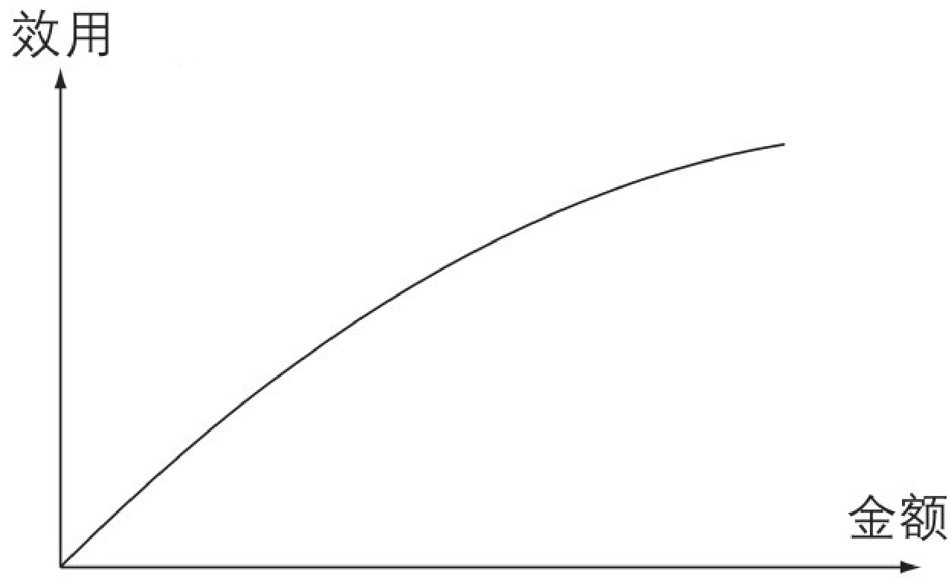


图9　一般的效用曲线的形状

“效用”解释了为什么房主和保险公司会同意总额250英镑是合理的年度保险费，这笔保费能够确保你在火灾、沉降或者洪水中免受25万英镑的房屋损失。保险公司的大额赔款意味着，对于任何单独一栋房子，这笔赔款可以看作与相关的数目较小的赔款预期的和是相等的。只要在任意一年中需要赔款的概率小于1/1000，保险公司在这笔交易中的期望值是正的，它就会盈利。另一方面，没上保险的房主就需要面对巨额的负效用，如果房子遇险，损失25万英镑，所以对于她来说，主动放弃250英镑来阻止这种可能性就是一笔合适的交易。

给诸如电视机、微波炉等的损坏上保险通常是坏主意。损失的总额很小，负效用和赔款本质上是相等的，保险公司会收取大额的保险费来保证盈利。应该将假想的保费存入银行作为修缮基金，而不是买高保费保险。很少有人会因为这么做而后悔。

如果你能成功地创建你自己的效用函数，你能够用它在不确定性存在的情况下做出选择。对于每一项行动，计算各种结果的**预期效用**

(expected utility)，即用对应的概率对效用加权，然后选择预期效用最大的一项行动。

无论具体情况如何，这就是概率论者做选择时候的通用秘诀。

[1]　全称Epsom Derby，为每年在英国举行的赛马比赛。

[2]　概率论中的一个基本结论。

06 人们玩的游戏 Games People Play

许多休闲游戏(recreational games)都是能力和机会的组合。能力是能提升的，但机会就与运气有关了。对于这里讨论的所有“游戏”，你会很容易相信它们是有限多个结果，并且它们都是等可能的。因此在这一章中，除非明确声明，我们都会使用处理概率的经典方法：计算可能结果的个数，任何事件的概率就是它们发生的结果占总结果数的比例。

我的目标是展示在不确定性存在的情况下，概率能如何帮助一个玩家做出正确的选择。观众对概率的理解也会增加观赏游戏的享受和乐趣。

彩票

在英国国家彩票中，最常见的模式是6/49。涂有不同号码的49个橡胶圆球在塑料桶中被搅拌打散，然后随机选择出6个来。赌博者花费1英镑来选择6个号码，如果这个选择中包含了至少3个中奖号码，他们就会赢得奖金。但是因为只有50%的彩票收入会成为奖金，彩票玩家平均收入的值远低于赌场赌博或者赛马。

彩票对人们的主要吸引力是巨额的中奖预期，虽然它们很难得到——曾有一张英国彩票开出2200万英镑的大奖，美国彩票的奖额已经超过了3亿美元。计算可知，用一张彩票中得头奖奖额的概率，在英国大约是1/14 000 000，在欧洲百万彩(the Euromillions Game)中小于

1/116 000 000，在美国超级百万(the USA Mega Millions)中大约是1/176 000 000。

为了体现这些概率到底有多么小，我们单看英国彩票。实际上，一个随机选出的40岁男子在一年内死亡的概率是1/1000。所以他在一天中死亡的概率是大约1/365 000，在一小时内死亡的概率是1/9 000 000，所以概率低到1/14 000 000表示的就是他在接下来的35分钟内死亡的概率了!对于美国超级百万彩票来说，在相同的假设下，这名男子中得头奖的概率和他在接下来3分钟内死亡的概率几乎相等。

尽管有极低的回报和几乎没有的赔率，“效用”能够给出买彩票行为的合理解释。用1英镑作为交换，无论如何你都会得到平均50便士，另外50便士你就用来买了梦想未来奢侈生活的权利、做了慈善事业，还有像我一样确信你在浪费钱的人的潜在的羡慕。这些权利的确算是一些效用。

我们应该假设未来的彩票开奖结果和先前的结果是相互独立的——一个无生命的橡胶球不会“记住”它是否“应该”被抽中。除了作弊，没有任何方法能够改变中奖的概率。但是你**的确可以**影响你可能获得的奖额的大小，因为在这些彩票规则中，奖池确定比例的部分会在每个奖励等级中被相应的中奖者们平分。我们有了一个训练某种能力的机会。

某些很小或特别的确定的数字(例如生日)会比其他数字更经常被选择，并且许多彩票玩家将自己的选择平均分配到自己买的彩票中，可能是误认为这么做就进行了“随机”选择。结果就是，几个较大数字，或者相距不远的一串数字，或者边缘的数字，比较不经常被选择。如果你能够确定其他的玩家做出了什么类型的选择，并且做一些**不同的**选择，你中奖的概率不会受影响，但是一旦你中奖了，你就会赢得比平均值更多的奖金。

别基于“没有其他人会想到这个”要小聪明，例如选择{1， 2， 3， 4， 5， 6}，或者选上次开奖时候的中奖号码。他们的确会想到。在英国国家彩票首次开奖的时候，大约10 000人选择了前6个数字。在2009年9月，相邻两次保加利亚彩票的中奖号码完全相同：第一次没人选那个号码，但第二次就有18个人选了。

已知其他玩家会像他们以前做的那样继续频繁地选择他们特定的号码，对于英国的6/49类型的彩票来说，下面的过程会帮到你。取普通的52张一副的纸牌然后去除3张。将剩下的纸牌与数字1到49一一对应，洗好牌，然后抽取6张纸牌。这是一种完全随机的抽取6张纸牌的方法。人类不可能不借助外物进行这种随机选择，我们需要这类辅助。

然后对于这6张牌，作如下规定：

(a) 它们加在一起至少达到177(为了偏向较大数字)；

(b) 在写在彩票上的时候，它们分成了2、3、4或者5个集群；

(c) 它们中有3、4或者5张的值落在彩票数值的边缘上；

(d) 它们不会在彩票上形成任何明显的规律。

如果其中任意一个条件不被满足，就将这6张纸牌放回牌堆，彻底地洗一遍牌，然后重复这个过程。

即使你遵循这个秘诀，你仍然预期会输钱——因为全部的回报只占彩票公司收款的50%这个事实是没法克服的。但是这样你就不太可能同整个世界分享你的头奖了。

电视游戏

“黄金球”(Golden Ball)在2007年首播。最后两位玩家会面对11个金球，其中一些对应一些奖额，另一些叫杀手球，不对应任何奖额。玩家从中选取5个球，生成潜在奖金，选中的任何杀手球会将之前选取的球代表的金额减小到原来的1/10。因此在选取了一个50 000英镑的球之后再取两个杀手球就会让奖额变成500英镑。

所有的球表面上看起来都是一样的，所以玩家完全是随机选择。从11个物体中选取5个一共有462种方式，所以从中选取到5个最有价值的球的概率就是1/462。在最初的288期节目中，这只发生过1次。

假设一个球名义上价值1000英镑：即使不考虑杀手球，选中它的概率也只有5/11，所以它的真实的价值就是455英镑。选取任意杀手球都会降低总奖额——在有3个杀手球的情况下，平均值的计算结果是255英镑。

当选取完5个球，实际的奖金已经知道的情况下，两个玩家都会做出一个彼此保密的决定，或是将自己的奖金**分享**给另一个玩家，或是**夺取**所有的奖金。他们俩会同时公布他们的选择：如果都选了分享，那么两个人分享奖金；如果其中只有一个选择了夺取，这名玩家就会获得所有奖金；如果两个人都选了夺取，两个人都不会得到任何奖金。

这是在博弈论中非常著名的一个情景，它被称为“**囚徒困境**”(the Prisoner’s Dilemma)。假设你的对手选择了分享，那么你选择

夺取就会获得较多的钱。如果你的对手选择了夺取，你怎么选都无所得。所以无论另一个人选择什么，你可以说选择夺取就永远不会输。经常发生的情况是，两个人都选了夺取，那么唯一的赢家就是不出任何钱的电视节目制作公司。

不同版本的“成交与否”(Deal or No Deal)已经在70多个国家播放。在英国，22个密封的盒子装有从1便士到250 000英镑不等的钱。这些盒子被随机分配给22个玩家，其中一位叫艾米的玩家将会在这一天参与游戏。她自己的盒子直到游戏结束都一直是关着的。她首先选择5个其他的盒子，之后盒子中的金额会被展示出来。庄家这时会出价来交换艾米的盒子中的金额。要是接受的话，她说“成交(Deal)”，则游戏结束；说“不成交(No Deal)”，则拒绝这个出价。若游戏持续进行，更多的盒子会被打开，庄家会开出新的出价，等等。

出价时，仍在游戏中的确切金额是已知的，所以他们的平均值是很容易计算的。在最初的几轮中，正常情况下庄家的出价会远远小于这个均值，但是艾米必须将她自己的效用函数牢记心上：如果她十分想要5000英镑，而出价是4500英镑，她就应该理智地接受，即使剩余盒子的平均金额超过20 000英镑——如果坚持不成交的话，她可能最后只得到1便士。

22次中有1次，艾米会被分到最高金额的盒子，但是她很少会赢得那么大的金额。效用可以给出一个令人信服的解释。在最后一次决定中，剩下两个盒子，其中一个是250 000英镑，另一个也许是2英镑。如果庄家出价80 000英镑，虽然这比均值125 001英镑小很多，也只有最有勇气的或者最富有的艾米才会拒绝。千鸟在林，不如一鸟在手……

已知庄家总是会给比剩余的盒子的平均金额少的出价，长远地看大数定律保证了“成交”的选手获得的金额比他们盒子中的少。所以

庄家的确会支付出价，但是也会在选手成交之后获得盒子中的金额，就会长远地获利。

“金钱本色(The Color of Money)”被认为是最令人紧张的电视节目，它在2009年只播出了4期。但是它给出了计算概率过程中应用加法和乘法定理的极好的例子。

1000英镑、2000英镑，一直到20 000英镑被随机分配到20个不同颜色的盒子中，玩家葆拉要达到一个预先设定好的目标金额，例如64 000英镑。为了达到这一点，她可以选择至多10个盒子，一次选一个。如果她(不知情地)选择了14 000英镑的盒子，1000英镑、2000英镑，一直到14 000英镑的数字就会以稳定的速度在屏幕上变换，每显示一个数字她都可以叫停。如果她及时叫停了，就会积累下最后显示出来的那个数字，但如果她等得太久，就什么都攒不下。如果在选取了10个盒子之后，她还是没有达到既定目标，她就什么都赢不到。她应该采用什么策略呢？

除了颜色，所有的盒子都是相同的，所以葆拉从每一轮中剩下的盒子中做选择是完全随机的。在她的最后一轮中，剩下了11个盒子，她的策略显而易见：例如，如果她想要9000英镑来达到目标，而且恰好6个盒子价值9000英镑或者以上，她就会计划在9000英镑出现时叫停，她成功的概率是6/11。但是在早期的几轮中她应该怎么办呢？

或在还有两轮的时候，剩下的12个盒子装有(以1000英镑为单位)1、4、5、6、9、10、12、13、15、17、19、20，而她需要额外的15 000英镑。当她看到7000英镑时叫停是不合理的；如果7000英镑曾经出现过，她就知道这时候盒子中至少有9000英镑，所以她就应该在9000英镑时叫停，显然这是更好的策略。她也可以将自己的选择限制在这剩下的12个数字中。相同的论证也同样适用于早前的几轮中——她叫停的最佳时机总是在剩下的某一个盒子里的金额出现的时候。

如果葆拉的确想要在9000英镑出现的时候叫停，她可以这么说：“12个盒子中有8个至少有这么多，所以我成功的概率就是8/12。而且如果我这轮的确成功了，在最后一轮中我只需要6000英镑，这时候11个盒子中有8个是我想要的。乘法定理告诉我们这两件事情同时发生的概率是(8/12) × (8/11) = 64/132。第一轮中有4个盒子的金额少于9000英镑，所以第一轮我什么都得不到的概率是4/12；这时我在最后一轮中就需要15 000英镑了，这个概率是4/11。再一次根据乘法定理，这种情况的概率是(4/12) × (4/11) = 16/132。这两个胜利是互斥的，所以加法定理告诉我们胜利的总概率是80/132。”

她也可以对她其他的选择策略进行相似的分析，例如在第一轮中期待6000英镑或是12 000英镑。我请你来做这些计算——附录中描述了她的最佳选择。

在这个电视节目的筹划阶段，有人出过一个主意，请一位专业数学家来提供及时的策略建议。葆拉可能会讲她会在8000英镑的时候叫停，这名专家也许会说：“这个选择不错。如果你这么做了，你赢钱的概率就是75%。但是如果你计划在11 000英镑的时候叫停，你的成功概率就会是80%。”

你可以想到会发生什么!专家说得都是对的，但如果葆拉更改了策略，却最终失败了，同时的确她最初的直觉会奏效。就一定会有一些小报叫嚷：“数学研究员夺走了军事英雄遗孀的64 000英镑。”

在我们调查过电视游戏节目中的数学之后，得知他们从未引入过数学角度建议员，我们都松了一口气!

纸牌游戏

大数定律意味着我们会在长时间的游戏中公平地得到好牌或者坏牌，所以最终游戏结果展现的是能力水平。我们来看3种流行的纸牌游戏。

在“二十一点”中，庄家必须遵循关于何时发牌的固定规则，玩家想做什么就做什么。直到爱德华·索普(Edward Thorp)开始赢得数额巨大的赌金之前，赌场运营者们都相信没有任何其他系统能够击败他们内在的优势。他们的逻辑中存在一个致命的瑕疵：虽然他们可以期望在牌堆中有6或者8副牌的时候赢得奖金的1%~2%，但是几局过后局面可能会转向对玩家有利。赌场忽略了使用剩下的牌的**条件概率**，而只考虑到了由完整的牌堆计算得到的概率。

索普开发出了一个追踪牌堆中剩余牌的方法。当数值较大的牌占比较高的时候，可能规则会变成强迫庄家抽牌，最终导致其点数超过21点而爆牌输掉。在相同的情况下，玩家却可以选择不抽牌。只要牌堆中数值较大的牌的比例较低或者适中，索普就会下尽可能小的赌注，而当牌堆组成对他更有利时，他会下更大的注。简单，但是高效。

当牌堆的组成的确对玩家有利的时候，他应该下多少注呢？约翰·凯利(John Kelly)在索普分析研究的几年前给出了精确的答案：下注的资金应该与他的优势大小一样。这个策略会让他资金增长的速率最大。

例如，假设他有1000英镑而且游戏对他有一点点有利——他获胜的概率是51%，失败的概率是49%。他的优势就是2%，所以他下注的比例应该是总资金的2%，即下注20英镑。下一次，他就会有980英镑或者1020英镑，如果他的优势保持在2%，按照出现的不同结果，他应该下注19.60英镑或者20.40英镑。如果他过于贪婪——在凯利指出比例仅2%时下注总资金的10%，尽管他有一定的优势，他最终还是会破产。他

的资金是有限的，下的赌注过高以至于不能承受不可避免的失败的趋势。

赌场会采取措施来识别和禁止熟练的数牌者，从来没有人因理解概率的出色能力而获利。

我们注意到，在考察法庭上的证据会如何改变我们相信有罪或无罪的程度时，正确的方法是采用贝叶斯公式。在例如惠斯特桥牌的纸牌游戏中，使用这个规则会提升你做出正确决定的概率。简便起见，我使用法律相关的词汇，用“有罪”来表示一个特定的对手同时持有一组特定的牌，比如说红桃Q和K，同时用“无罪”表示她至多持有这些牌中的一张。

使用计数的方法，我们可以找到她持有全部两张牌的情况占全部情况的比例，这样就会给出对于“有罪”概率的最初评估。按照标准的习惯，最好将这个概率转化成它等效的赔率。最开始的计算完成后，我们可以说我们得到了(有罪的)**先验赔率**(Prior odds)。

当纸牌游戏进行下去，相关的**证据**就会浮现出来——但她或许是在耍花招。为了考察这些证据是如何影响“有罪”的赔率的，我们来计算一个叫作**似然比**(Likelihood Ratio)的量。首先，假定“有罪”(她同时持有K和Q)，评估这些证据出现的概率；然后，假定“无罪”(她至多只有其中一张)，评估这些证据出现的概率。似然比就是第一个值和第二个值的比值。

我们现在就可以推断出**后验赔率**(Posterior odds)，也就是说在考虑到证据的情况下“有罪”的概率，使用贝叶斯公式，也就是：

后验赔率=先验赔率×似然比

这个定律的形式是清晰直白的：似然比越大(也就是说，你的对手的确是有罪的时候，证据越容易出现)，有罪的赔率就越大——但是这个公式精确地告诉你相关证据如何影响有罪的概率。

要看这个理论是如何运行的，来思考一个现实的情况：我们的对手要么是同时持有K和Q(有罪)，或者她只持有K(无罪)；先验概率告诉我们这些选择差不多是同等可能的。如果她是有罪的你最好就出A牌；如果她是无罪的，你就应该出一些其他的牌。证据这时出现了，她出了K。

如果没有证据，你就只能凭空猜测，而且你也就有一半的概率会赢。如果她是无罪的(她只持有K)，那么证据(她出K)出现的概率就是100%；但是如果她是有罪的(她同时持有K和Q)，她可能也就会出Q而不是你看到的K，所以证据出现的概率就是50%。这两个数字比值是1/2，所以贝叶斯公式告诉你后验概率就是1/2。也就是说，她无罪(只有K)的可能性是有罪的2倍。所以现在出A就有2/3的概率获胜。

在正确使用概率的情况下，你应该希望能做得更好，如果你能够以2/3的概率获胜，就别以1/2的概率。你不能保证每局都赢，但是这样做你可以提升获胜的概率。

桥牌玩家把这种情况称为“限制选择原则(the Principle of Restricted Choice)”——如果对手只有一张K她就必须打出，如果同时有K和Q，那她可以做选择。她的确打出了K的这个事实，让概率向她不得不这么偏移了。

如今，最流行的扑克的形式是德州扑克[1](Texas Hold Em)。每个玩家发2张牌并且设法使自己手里面的牌被最有效地打出，5张公共牌随后会被陆续发放。公共牌发放之后，为了最有可能击败其余两组牌，哪一组牌是最好的？

手牌A：两张梅花2，黑桃2；

手牌B：黑桃A，方片K；

手牌C：红桃J和10。

这当然是一个陷阱问题：在仔细计算之后，手牌A会以52%的概率击败手牌B，手牌B以59%的概率击败手牌C，手牌C以约53%的概率击败手牌A。所以你就会选择A而不是B，选择B而不是C，但是你同时也会优先选择C而不是A！如果你的对手从这3组牌中选取任何一组，随后你在剩下的牌组中随机选择1组，你都有超过50%的概率获胜。

扑克游戏需要的远远不止概率技巧。你必须对对手有可能持有的牌，以及你什么时候应该虚张声势做出评估。但是有些时候概率十分有用。假设赌注总额是50个筹码，还有1张公共牌等待发放。这时你发现，如果最后一张牌是黑桃，你就能打出同花顺，这样你一定会赢；如果它不是黑桃，你的一个对手就会赢。你应该在游戏中押下更多的筹码吗？

忽略掉你已经在赌注总额中放了多少筹码。重要的是未来的情况。你能看到6张牌的牌面——2张在你手里，4张在桌上的公共牌中。在46张未知的牌中，9张黑桃会让你赢得胜利，其余的会导致失败。在赌注总额已经是50个筹码的情况下，再多付10个筹码来看最终分发的纸牌是值得的吗？再付20个呢？

通过计算得到你一定要多押的x个筹码带来的平均收益(或者损失)，你可以得到一个阈值x。它将会从长远角度带来收益。附录中给出了这个问题的答案。

[1] 简称德扑，最流行的扑克衍生游戏之一。

07 在科学、医学和运筹学中的应用 Appl ications in Science, Medici ne, and Operations Research

我们会根据情景使用不同方式来对概率进行评定或者诠释。但是，就像大卫·汉德(David Hand)在他的《牛津通识课：统计》(Statistics: A Very Short Introduction)中写的那样，“……微积分是一样的”，换言之，概率的操纵方式是不变的。

你头脑中要牢记这个学科的中心思想：加法和乘法定理、独立性、将客观概率和频率联系起来的大数定律、在将随机数求和时候使用的高斯分布、其他的一些经常出现的分布函数、反映总体情况时有用的平均值和方差。我们可能不指望我们对相关概率知道得像前几章那样精确，但是一个对于问题大致正确的回答对于作出合适的决定有良好的指导意义。就像统计学家乔治·博克斯(George Box)所说的那样，“所有的模型都不是完全正确的，但是有一些是很有用的”。

下两章中举例说明了概率的应用，这些应用以章节标题粗略地分了组。

布朗运动和随机游走

1827年，植物学家罗伯特·布朗(Robert Brown)观察到，在液体中悬浮的花粉粒子似乎随机地在动来动去。将近80年之后，阿尔伯特

• 爱因斯坦(Albert Einstein)对其给出了一个解释：花粉粒子被液体中的分子持续地撞击。这种运动当然是发生在三维空间中的，但是为了创建一个令人满意的模型，我们首先思考在一条直线上的运动。

假设每一步运动都是具有固定长度的，有时向左有时向右，每一次运动都是独立的。这个概念就叫作**随机游走**(Random Walk)。在许多次跳跃之后的位置只取决于向两个方向的跳跃次数的差值；从起始点计算的距离平均值和方差与进行跳跃的次数成正比。

下面进行一个微妙的计算：在一个固定的时间段中，**增加**跳跃的频率，并且**降低**每一次跳跃的距离。在这两种因素平衡的情况下，极限情况就是连续运动，运动经过的随机距离遵循高斯分布(依据中心极限定理)，这个分布的平均值和方差都与时间段长度成正比。如果向左和向右的运动是等可能的，平均值就会是0。

爱因斯坦对于布朗的观察作出的解释，是粒子在三维空间中运动，基于上文给出的原因，在每一个方向上的运动都遵循高斯分布。他对原子和分子的行为作出了预测，这些预测推动了一些实验，这些实验消除了有关原子和分子存在的所有悬而未决的问题。

术语“**布朗运动**(Brownian motion)”本应该专指在液体中的粒子的实际运动，但是它也被用于指代描述这种运动的数学模型。

随机数

“随机数”这个词指代的是下列两种想法之一。第一，就像理想情况下的色子游戏或者轮盘赌游戏，一个数从一个有限的列表中被取出来，所有的这些数都是等可能的。第二，就像从一个随机的一点折断一根木棍一样，一个点被从一个连续的区间中取出来，没有任何一小段是比其他的小段更容易被取点的。取出这样的随机数中的一段长

序列的方法具有广泛的应用，序列中的每一个值与其他的值都是相互独立的，下一部分会举例说明这一点。

在1955年，一本名为《一百万个随机数》(One Million Random Digits)出版了。它的确就是它的书名描述的那样：一页又一页的0到9的数字，被分块以便阅读，但是连续的数字是完全不可预测的——无论前面的数字是什么，你都会有1/10的概率猜对下一个。如今，现代计算机已经有内置的软件来获取与之相同的结果。输入一个初始值——随机数**种子**(seed)，一个确定的数学公式就会给出下一个值，它被当作新的随机数种子，以此类推。这个过程中毫无随机意义可言，并且如果每一次都是用相同的随机数种子，就保证会生成完全相同的数列。但是，基于数学公式的巧妙选择，生成的序列成了统计检验中的基础，而且对于任何目的和用途来说，它**看起来**就好像是随机的一样。我们用**伪随机数列**(pseudo-random sequence)这个术语来称呼它。

无论在这个过程中有多么小心，一定会有一些方法中的谬误将会引起对随机数应用场景的担心。但是依赖大量受人尊敬的科学家的经验，我还是会认为我电脑中根据需求产生的随机数是可以接受的。(显而易见的内部人员欺诈的风险导致了这些方法不能应用于彩票摇奖，或者英国溢价债券。)

蒙特卡罗方法

连续37次转动标准的欧洲轮盘赌轮会得到多少不同的数字？理论上讲，这个个数会是1~37之间的任何数，但是那些极端的值会十分罕见，最常出现的不同数字的个数是多少？

这个问题第一次呈现在我面前的时候，我没有立即看出有什么简单的方法能解决它。转动赌轮37次会得到37^{37}(一个有59位的十进制

数)种可能的结果，而且当你试图罗列，例如有28个不同数字的结果的时候，你将会很快失去热情。一个更加吸引人的方法是进行一个所谓的**蒙特卡罗模拟**(Monte Carlo simulation)。

在这里，计算机产生的随机数流将会用来模拟37次转动赌轮的结果，之后计算机将会计算有多少不同的数字出现了。这个过程将会重复100万次，结果是24个不同数字出现了203 739次，而23个数字只出现了199 262次。最接近的竞争对手是22或者25个数字，都出现了不到160 000次。大数定律告诉我们不同的结果出现的频率将会稳定在它们的各自的概率，而且这些数字的确本质上证实了这件事：最有可能的结果就是有24个不同的结果会出现，概率刚刚超过20%。

几天后，我羞愧于当时没能找到一种标准的方法来解决这个问题!我能够计算出对于任意的X，转37次赌轮得到X个不同数字的精确的概率，以证明上面的结论是正确的。但是这不会让应用于这类问题的模拟失去效力——快速而粗糙的结果也会是很有用的。的确，模拟给出的结果与精确计算得到的结果保持一致这个事实，增加了我对计算机随机数生成器按预想运行的信任。

一个更加严肃的蒙特卡罗方法的应用是在聚合物化学(polymer chemistry)中。一个分子是由大量的原子被随机扭曲的长链连接构成的。原子们只能出现在均匀分割的晶格中，并且关键的一点是，没有两个原子能够出现在同一个位置。从分子的一端到另一端的距离可能有多远?

我们可以认为原子被放置在一个醉汉走过的位置，这个醉汉摇摇晃晃地随机经过三维空间中的晶格，但是因为某些原因不能在同一个位置经过两次。没有不能重复经过同一个位置的这个要求，数学家们可以给出很好的解答，但是这个限制条件似乎将问题复杂化到了理论无法解决的地步。然而，哪怕一个不专业的计算机程序员也可以写出

一个对这个复杂、曲折的链的合理的模拟，而且在100万、1000万，甚至10亿次重复之后，得到一个所需的精确答案。(回忆棣莫弗的工作，精确度与模拟次数的**算数平方根**成正比。)

假设你想要估计一片不规则形状叶片的面积。画一个包围了这片叶子的矩形框，然后模拟大量的随机分散到矩形内部的点的位置。你的估计就是用矩形区域的总面积乘以落在叶子边界内部的点的占比。

作为最后一个举例介绍的应用，假设保罗(Paul)想要建一个新的加油站。如果他安装4个加油泵，这是最小的可行的泵数，就会有至多8辆车的等待区；每个额外的泵减少2个等待区，所以如果他安装了最多的8个泵，就没有等待区了。为了计算多少个泵会让他的收益最大，他可以进行对安装了4、5、6、7或者8个泵的情况的模拟。

他应该知道潜在的顾客前来加油的比例和一辆车停在泵旁边加油的时间的分布，还有安装费用、运行费用与边际收益。他也应该考虑到如果没有加油泵空闲，或者队伍过长时，一个潜在的顾客直接开过去不来加油的概率。找到或者估计所有这些数字都是相对简单的，而且用计算机进行模拟会比在几个月中以不同的泵数进行实地试验**便宜很多**。因为他可以每一次都使用相同的随机数种子，他就可以在精确相同的情况下运行所有的模拟，对比不同的估计从而增加收益。

为什么叫“蒙特卡罗”呢？除了随机数和赌场游戏之间的明显的联系，这个名字其实原本用来指代军事领域对这些方法的应用，这其中就包括了早期的核弹的研制[1]。

电码中的错误

摩尔斯电码(Morse code)告诉我们如何只用两种符号，比如说0和1来传输消息。但是其中一些符号也许会被损坏，以至于原本发出的0

接收到的时候就变成了1，反之亦然。甚至在较低的错误率下，接收到的消息也会与发送的消息有天壤之别。我们如何应对这种情况？

假设每个传送的符号都有相互独立的较小概率被损坏。我们可以重复发送这些符号，但是稍微一想就会发现，发送00和11而不是0和1根本不能解决问题：如果01或者10到达，确认到底传输的是00还是11就全靠猜测。我们会猜对一半，但是重复发送符号意味着我们可以预见其中会产生两倍的错误，所以两种因素大部分相互抵消了。但是我们考虑一下传输000和111而不是0或者1。

采用“少数服从多数”的原则来进行解码，所有的{000， 100，010， 001}都被理解为0，其他4种可能的情况被理解为1。如果只有1%的发送的符号被损坏了，那么当000被发送出去，利用二项分布告诉我们有99.97%的概率上述的4种序列被接收到。这意味着错误率从1%降到了0.03%，降低了超过30倍。如果每个数字重复5次，我们可以得到更好的结果，但是消息变长会增大开支。最佳选择将会依赖内在的错误率和传输的速度。

羊膜穿刺术

准父母(同时也是统计学家)樊娟娟(Juanjuan Fan)和理查德·莱文(Richard Levine)在考虑樊娟娟要不要接受羊膜穿刺术(amniocentesis)——可以检测她的胎儿是否患有唐氏综合征[2](Down’s Syndrome)。他们的经历可以作为其他类似情况的模板。

基于樊娟娟的年龄和简单的血液检查，我们可以给出胎儿患有唐氏综合征的概率为1/80；超声图像检查的结果是令人满意的，借助贝叶斯公式计算后，得到的患病概率减小到了1/120。羊膜穿刺术是一个侵入性的手术——一根中空的针刺入腹腔抽取羊水样本；如果作为唐氏综合征特征的21号染色体被检测出多余复制，就一定能够确诊，但

是羊膜穿刺术这种检测有一定的风险导致流产，在这个情境中风险概率估计是1/200。如果确诊胎儿为唐氏综合征后，父母们一定会选择流产。他们应该接受这个检测吗？

樊娟娟和莱文通过使期望效用最大化的逻辑分析过程得出他们的决定。可能出现的最坏结果是没有患唐氏综合征的胎儿流产，其效用赋值为0；最好的结果，是出生的胎儿不患有唐氏综合征，其效用赋值为单位1。不进行羊膜穿刺术，生下来患有唐氏综合征的孩子，赋值效用x；进行了检测，的确检测出了唐氏综合征的效用y应该大一些。(最后一个情况中检测导致的流产就无关紧要了，因为无论如何胎儿都会被流产。)

他们计算出了进行检测和不进行检测的期望效用。如果第一个值超过第二个，就应该进行检测，在这种情况下，就是要求y大于(119/200) + x，粗略地要求y大于0.6 + x。

如果樊娟娟和莱文认为确诊唐氏综合征之后进行流产的效用小于0.6，那么进行检测就会毫无意义。他们给生下尽管患有唐氏综合征的小孩的效用x赋值越高，那么y的阈值就会越高。如果这个效用x超过0.4，那么他们就一定不会接受检测。

选择x和y的合适的值需要一定的思考，而且如果基本的参数——接受检测的流产概率为1/200，不接受检测生下唐氏综合征婴儿的概率1/120——发生变换，最后计算得到的判断准则就会变化(参见附录)。简单地讲，如果胎儿患有唐氏综合征的概率小于意外流产的概率，接受检测就不合理了。真的是这样吗？

樊娟娟和莱文讨论了他们面临的难题，他们赞同的效用值的选择让他们决定接受检测。结果是美好的：没有多余的染色体，也没有发生流产。

血友病

血友病(haemophilia)是一系列伤口产生后无法凝血的疾病的统称。控制凝血的基因位于X染色体上，这个基因异常的概率小于1/5000。因为女性有两个X染色体，所以她们只会在两个染色体上的基因均异常的时候才会发病，发病概率小于1/25 000 000，但是男性只有1个X染色体(和一个Y染色体)，所以几乎所有的病例均为男性。

如果男性患有血友病，这在他做父亲之前就已经确定了，但是一个女性有可能有一个正常的X染色体和一个携带异常凝血基因的X染色体。这些女性被称为携带者，而且她把异常基因传递给孩子的概率是50%。女儿继承了异常基因会成为携带者，儿子继承了异常基因就会成为血友病患者。维多利亚女王(Queen Victoria)肯定是一个携带者，因为她的儿子利奥波德(Leopold)是一个血友病患者，而她的5个女儿中至少2个是携带者，她的另外3个儿子没有患这种病。

假设贝蒂(Betty)有一个患有血友病的兄弟，贝蒂有7个孩子，其中包括安妮(Anne)。那么安妮是携带者的概率是多少？

为了解答这个问题，得到贝蒂是携带者的概率就足够了，安妮的患病概率一定是这个值的一半。因为贝蒂的兄弟是血友病患者，所以贝蒂的妈妈是携带者。但就这个信息而言，贝蒂是携带者的概率是50%。而且如果安妮的**任意**一个兄弟患有这种疾病，贝蒂就一定是携带者，所以我们来看安妮的所有兄弟都正常的情况。

贝叶斯公式非常适合于这种情况。设“有罪”意味着贝蒂是携带者，因为有罪的先验概率是50%，先验赔率就是单位1。如果贝蒂是无罪的(不是携带者)，证据(所有的兄弟都是健康的)的概率就直接是100%。但是如果贝蒂是有罪的，每个安妮的兄弟都独立地有50%的概率避免错误的基因，所以每个健康的兄弟都会让似然比减半。将后验赔

率转换为后验概率，贝蒂是携带者的概率就相继是1/3、1/5、1/9、1/17……，对应于她有1、2、3、4……个全部健康的兄弟。

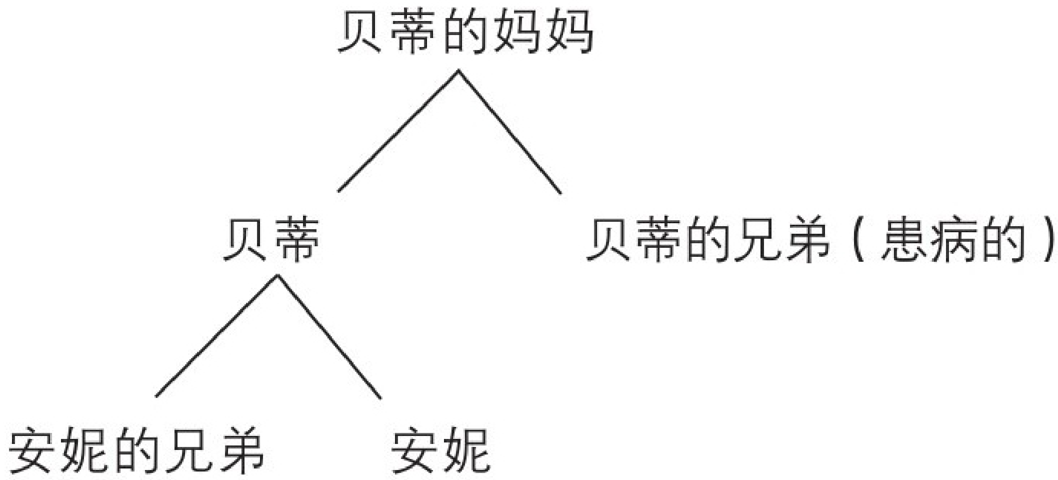


图10　家庭成员关系

作为消遣你可以思考一下，假设安妮同时有姐妹，她们其中一些有儿子，而且安妮的这些侄子们都不患病。这将如何影响安妮是携带者的概率？将你的答案与附录中的相对照。

流行病学

术语**群体免疫**(herd immunity)表示一个事实：如果足够多的人免疫了，那么即使传染病进入社群，流行病也不会发生。因此即使那些没有接种疫苗的人也非常不可能感染疾病。为什么是这样的，以及我们如何知道“足够多”意味着多少？

通常来说，那些感染者将会把疾病传染给其他人，但是那些痊愈了的就会获得对这种传染病的免疫。所以我们把人们分类成易感人群、受感染人群和去除人群，最后一个指的是那些接受疫苗注射、感染后康复、物理隔离或者已经死亡的人。将这四种情况归为同一类看

似有一些生硬，但残酷的事实是，仅从流行病传播的角度，这四种情况是完全等价的。

为了考察流行病是如何发展的，设S为易感个体数，I为受感染个体数。这两个数字的乘积给出了受感染个体与易感个体可能接触的总数。城市中拥挤的全体居民要比同样多的农村中分散的全体居民更频繁密切地接触，而一个易感个体在一次接触中被感染的概率依赖于疾病的传染性。总的来说，在任意很小的时间段内，一些易感个体被感染的概率形如$\beta \times S \times I$，其中β同时依赖于疾病的传染性和人们混合的情况。

在这个很短的时间段内，任何一个受感染个体可能会被移入去除人群中。因此受感染人群减少一个成员的概率将会与受感染的人数成正比，所以我们采用$\gamma \times I$的形式，其中γ依赖于感染被治愈、隔离或者致死的速度。

我们已经分别得到了受感染人数上升和下降的概率。它们($\beta \times S \times I$和$\gamma \times I$)之间的平衡将决定流行病是否会发生。这十分类似赌博。如果游戏局势对你不利，每一个赌局都会减少而不是增加你的资金。你的资金遵循着随机游走的规律，不可避免地漂移到0。但是如果游戏局势对你有利，例如坏运气在游戏初期没让你破产，此时的随机游走就会使你从所有输钱的趋势中抽离出来，使你的资金远远超过0。要赢得大额赌金的话，游戏对你有利这个条件是必要的，但不是充分的。

在这个流行病的情境中，这就意味着流行病(等同于大笔收益)只会出现在受感染者数目变化是增加远多于减少时。用符号描述，就是$\beta \times S \times I$远大于$\gamma \times I$，这也等同于要求易感个体数S超过比例γ / β，这个比例被称为易感个体数的阈值(threshold)。这就是我们一直寻找

的：即使疾病传入人群中，流行病也只会发生在易感个体数超过这个阈值的情况下。

威廉·科尔马克(William Kermack)和安德森·迈肯德里克(Anderson McKendrick)在1927年发表了这一结果。将易感人群总数控制在这个阈值以下可以阻止流行病的发生，两种明显不同的方法能达到这一效果。第一，接种疫苗来减少易感个体数；第二，找到增大阈值的方法。这个阈值是一个比值，分子增加的时候，阈值就会增加——加快治愈的速度，或者更迅速地将人们隔离——或者让分母减小——也许可以降低疾病的传染性，或者确保人们更少地相互接触，这可以通过暂时关闭学校或者推迟有大批人群聚集的运动赛事进行。我们还可以评估这些对策的收益，以便决定采取什么策略。

相同的原理对处理动物中的流行病也是奏效的。阻止牛群中口蹄疫暴发的第一步就是限制牛群的活动，这是通过降低分母来提升阈值。与此同时也会伴随着大量屠宰(在人类中不可行!)，这样不仅降低了易感个体，也会提升阈值的分子。

上述分析也揭示了为什么我们更希望两次儿童流行病暴发的间隔时间是相对有规律可循的。在一次流行病将易感个体数量减少到阈值以下之后，没有足够免疫的新生儿会逐渐将易感个体数提升到阈值以上，这就为下一次流行病的暴发创造了条件。两次流行病暴发的间隔越长，易感人数就会越多，暴发就会越严重。

了解概率知识可能不会治愈疾病，但可以缓和疾病的影响。

批量测试

一支军队想要确定1000名潜在的新兵中有哪些是对某种疾病易感从而不适宜服役的，可以进行血液检测，但每次检测花费50英镑。用

少于50 000英镑可以完成这项工作吗？

如果易感人群只占相当小的比例，那么答案就是肯定的。选取数字K，收集K名新兵的血液混成一个样本；然后测试这个样本。如果结果是阴性的，那么所有的这些抽血了的新兵就都是健康的，不需要进一步的检测了；否则(检出阳性)这组人里面至少有一个是阳性，所以我们再进行K次检测，每个人进行一次，来确定到底谁是阳性。如果我们足够幸运的话，一次检测就够用了，但我们有可能需要进行K + 1次检测。我们希望检测次数比每个新兵单独检测的次数低。

对一次收集多少个人的样本是最佳选择依赖于检测到阳性的概率。假设这个概率是1%。那么如果我们一次混合10个样本，那么它是一个阳性样本的概率大约是10%，同时这个样本为阴性的概率是90%。所以90%的情况中我们只需要1次检测，但是10%的情况我们需要11次检测。这致使每次需要平均2次检测。混合10个样本将会把检测10名新兵的花费从500英镑减少到100英镑。如果我们将这1000个士兵每组10人地分成100组，然后混合他们的血液样本，我们预期将会节省最初估计花费50 000英镑的80%。

更加精细的计算显示，如果检测出阳性的概率的确是1%，将每组按11人批量检测会比10人稍微好一点，但是其中的差别很小。但是批量检测分组大小的最佳选择对于检测出阳性的概率是非常敏感的。如果我们预期2%的新兵会被检测出阳性，那么每组8人的时候花费最少；在预期的阳性率为5%的时候，每组5人花费最少；而在预期的阳性率为10%的时候，每组4人最合适。(又一次地，使用二项分布可以得到这些结果。)

在第二次世界大战中，这个朴素的想法为美国节省了最初预期检测花费的80%。

航班超额预订

即使必须对那些支付过机票但被赶下飞机的乘客进行补偿，航空公司还是会常规地售出比飞机载客量更多的机票。其中的原因涉及简单的经济学：航空公司完成一次航班的开销几乎不受乘客数目影响，但是每个空座位都会使收益减少。如果预期不是所有订了某次特定航班机票的乘客都会乘机，那么航空公司如何计算出最合适的超额预订数量？

假设飞机上有100个座位，每个座位的费用是100英镑，但是如果因飞机上位置已满而拒绝一名乘客登机需要赔偿其200英镑。航空公司需要对一名预订了机票的乘客实际乘机的概率进行较好的估计。对于前往旅游目的地的包机来说，这个概率将会接近于100%，但是对于那些旅行计划中有更多可变性的乘客来说这个概率会比较低。有关频率的数据将会帮助航空公司估计这些概率。

假定每个预订机票的乘客都独立地有80%的概率会最终登机。如果售出120张票，总的收益就是12 000英镑，而且，虽然平均只有96名乘客会登机，但是有一定的概率，大概15%，超过100名乘客露面，至少有一名乘客留在地面。(这些数字再一次由二项分布给出。)这种情况下航空公司因为超额预订而需支付的赔偿平均为80英镑。多售出5张机票将提升500英镑收益，而平均全部赔偿只提升了275英镑，所以这个策略预计会收益更多。平均收益最多的是售出128张票——比售出125张更多，额外的收益300英镑也已经比平均赔偿295英镑更多了，同时售出129张票会比这稍差一点。

假设某些乘客比另一些乘客更有可能实际登机是更现实的，而且一组一起订票的乘客要么会全部出现，要么全部不出现。这些细节会被容纳到模型中，航空公司会继续售出座位，直到多售出一个座位预计的赔偿超过了额外收益的增加。

排队

概率发展得最好的应用之一是在各种排队中。最初的需求源自试图理解电话线路中的阻塞现象——为表扬丹麦电话工程师艾格纳·埃朗(Agner Erlang)的工作，电话通信流量大小以他的名字命名。排队论为1948年和1949年的柏林空运[3](The Berlin Airlift)的胜利作出了贡献，对排队的系统性的学习在接下来的20年中蓬勃发展。

大卫·肯德尔(David Kendall)引入了一种记号，形如A/B/n，现在普遍作为在顾客逐个地加入队列的时候，对于队列的描述方法。第一个组成部分A指的是两个顾客到达的间隔时间的分布，B指的是为1个顾客提供服务所需时间的分布，n是服务员的数量。

例如，在表达式D/D/3中，“D”是确定性(deterministic)的简写，它意味着完全没有随机性。顾客到达队列末尾的间隔时间是恒定的，所有的服务时间都恰好相等，并且有3名服务员。这种队列在概率的领域中可能并不让人感兴趣，因为其中没有变化性。但若假设有大量的顾客，在很短的时间间隔内他们中的每一个都有很小的概率加入队列，所以顾客以一个总体来说平均的时间间隔加入队列，但是顺序完全随机。这种情况下使用符号M，为了纪念安德雷·马尔可夫，所以M/D/2意味着顾客们随机地到达并且选取两个服务员中的一个，这两个服务员完成工作花费的时间是固定的。

我们想知道队列是如何运作的。最主要的问题是顾客们将会等多久，服务员多么频繁地无所事事，还有为了改善情况我们可以做什么？“服务员”可能是重症监护病床，同时“顾客”就是等待看护的患者。

如果平均每5分钟有一个顾客到来，一共有3名服务员，那么除非平均服务时间少于15分钟，否则就将会产生一个无限长的队列，总体

运营就无法维持了。所以我们必须假设考虑到了服务员数量之后，平均服务时间小于顾客加入队列的时间间隔。这两个平均时间的比值被称为**流量强度**(traffic intensity)，是一个介于0和单位1之间的数值。

在理想情况中，顾客排队时间比较短而且服务员将会一直工作。但是这两种要求正好是矛盾的。以简单的情况为例，有一个服务员而且顾客随机到达。如果流量强度是0.9，计算显示我们预计也许平均有5名顾客在等待，在10%的时间中队伍中没有人。如果流量强度增加到0.98，服务员将会只在2%的时间内空闲，但是平均队伍长度将会升高到25。大多数的顾客将会认为这是个比较差的安排。除非服务员们有更多的“空闲时间”，否则顾客们将会生气或离开放弃这里的服务，或者又生气又离开。

队列的运作不只是依赖于流量强度。在其他条件相同的情况下，服务时间的可变性越强，预期队列将会越长。在只有几个服务员的情况下，顾客加入唯一的一条队伍之后分配给6个服务员(例如火车站)，与顾客可以选择加入某一条队伍之后等待(例如超市)是不一样的。在某些情况下(例如叫救护车)某些顾客可能会有更高的优先级。许多的队伍遵循着“先到者先服务”的规则，但是在不易腐烂的货物存放在架子上等待使用时，可能会遵循“后到者先服务”的规则。某些队伍会汇入其他队伍，服务员们工作速度不同，一群顾客可能会同时到来。明察秋毫的排队论理论家们可以解答几乎所有你能想到的实用模型中的核心问题。

[1] 20世纪40年代，科学家冯·诺伊曼(John von Neumann)与斯塔尼斯拉夫·乌拉姆(Stanislaw Marcin Ulam)等人于美国洛斯阿拉莫斯国家实验室(Los Alamos National Laboratory)为核武器计划工作时，发明了蒙特卡罗方法，此方法的名称来源于乌拉姆的叔叔经常光顾的位于摩纳哥的蒙特卡罗赌场。

[2]　又称21-三体综合征、先天愚型，是一种因为21号染色体的三体现象造成的遗传疾病。

[3]　英美等国为解除冷战初期苏联对西柏林的物资封锁而展开的空运行动。

08 其他应用 Other Applications

这一章中我们不关注概率在掷色子、赌场赌博和其他在自然科学中的种种应用。这一章中为了展示概率的普遍存在，我找出了它在法律、社会科学、体育运动和经济学中的应用。

法律事务

尽管20世纪最著名的英国法官之一——丹宁勋爵(Lord Denning)有一个数学学位，但是很少有律师对概率抱有好感。有关概率这门学科有关的措辞在法庭中被自由使用，这应该是令人震惊的。在民事案件(例如诽谤案件)中，陈述“概率是均衡的”就相当于明确地将分割线设在了50%。但是在刑事案件中，只有在陪审团“确认”有罪后他们才能定罪。在数字上没有共识。其中一些人在她们80%确定有罪时定罪，另一些人会在95%或者更高时确定。这些明显是主观概率。而且尽管无论在什么样的罪行判定中都使用相同的措辞，有些人会对相对较小的罪行采取较低的证明阈值。这也就是为什么比起逃票，恶性谋杀案更难以定罪。

假设一名专家证人做证说被告的DNA与在案发现场找到的DNA相符，而后者与随机选取的一个无辜者相符的概率是几百万分之一。陪审员们对这个证词的理解可能存在两个不同的问题。第一个是他们也许会认为证词等价于说在犯罪现场发现的DNA不属于被告的概率是几百万分之一。第二个是他们会将所有这些很小的数字等量齐观，即使一千万分之一与十亿分之一的一百倍是不同的。

第一个问题被称为“检察官谬误”(prosecutor’s fallacy)。很明显这个错误就相当于认为已知DNA相符的情况下无罪的概率，与已知无罪的情况下DNA相符的概率，是相等的。这是个逻辑上的无稽之谈。已知轮盘赌轮是公正的时候，转到0的概率与转出了0的情况下赌轮公正的概率并不相等。向陪审团提供与犯罪现场DNA相符的市民的数量会避开这个陷阱。在有大约6000万人口的情况下，如果相符的概率是两百万分之一，就会有30人左右相符；如果概率是两千万分之一，大约有3个人相符，不太可能超过半打。但是不要忽略了“随机选取”这句话：与罪犯亲缘越近的人，越可能相符，这个证据对被告不利的程度就越弱。

通过考虑到贝叶斯公式计算证据有效性的方法，可以避免第二个错误。在这个证据呈递之前，你对被告是否有罪有一定的看法。如果证据出现的可能性在有罪时是无罪时的10倍，那么有罪的赔率就会在证据出现之后乘以10；如果证据出现的可能性在无罪的时候是有罪的时候的3倍，那么有罪的赔率就会减少到原来的1/3，以此类推。在存在DNA证据的情况下，假设有罪时证据出现的概率经常是100%，这就让证据的影响变得清晰起来：有罪的赔率就应该乘以那个是“几百万”的具体数字。

随机化回答

一名班主任想要查明高年级学生吸食大麻的比例。直接询问无法得到可信的答案，但是一种称为**随机化回答**(randomized response)的技术是可用的。大意就是负责记录回答的老师并不知道实际上询问的问题是什么，所以大麻使用者就不会担心被认出而可以诚实地回答问题。

80张卡片每张都写着“我吸食大麻”，另外20张都写着“我不吸食大麻”。每张卡片都被装入完全相同的信封中，这100个信封在一个

大袋子中被充分混合。学生们应该看到这些操作，所以他们明白袋子中装着两种不同版本的问题和它们的比例。

安吉拉随机选择了一个信封，打开了它并自己看了问题，然后仅仅回答“同意”和“不同意”中的一个。然后她将卡片放入信封，将信封返还到袋子中，并摇晃袋子为下一个学生做准备。

假设1/3的回答都是“同意”。因为学生们都是随机地抽取信封，“同意”来自80%的使用者，与20%的非使用者的诚实回答。几行代数计算显示这与2/9的学生是使用者的情况是一致的。班主任得到了他的答案，没有单独哪个学生被认出来。

或者，把“我不吸食大麻”替换成一个无关的题目，它被“同意”的比例是已知的。如果一个早期的调查估计有一半学生拥有一只宠物，而且没有理由将吸食大麻和拥有宠物联系在一起，在那20张卡片上的问题就可以是“我拥有一只宠物”。然后，如果1/3的回答是“同意”，我们估计7/24的学生是大麻使用者。

得出这些答案的计算过程会在附录中给出。

每次提问时问的是哪个问题的不确定性致使最终估计中存在一些不精确度。装有敏感问题的信封的占比应该尽可能高，但是要足够低以使真正的大麻使用者相信给出诚实的回答不会招致麻烦。在高达95%的卡片上设置敏感问题是不会有效果的。

世界反兴奋剂机构

世界反兴奋剂机构(The World Anti-Doping Agency，WADA)致力于促进体育运动成为一项健康的活动，他们识别使用能提升表现的药物的运动员，并将之排除在比赛之外。但是无论使用什么方法，任何

检测程序都可能会给出两种相反的错误：认为一名无罪运动员使用了禁药，和让一名禁药使用者通过检查。

不幸的是，减少两种错误之一的概率的方法经常会造成另一种错误的概率增加。例如，检测测量睾酮(testosterone)与表睾酮(epitestosterone)的比例。正常情况下肌体产生相等数量的这两种物质，但是那些通过注射睾酮以作弊的运动员就会有较高的T/E比。T/E比高于某个特殊的值，比如6比1的运动员，将会被禁止参赛。然而，自然情况下T/E比就会变化：它会随着月经周期(menstrual cycle)变化，如果你得流感了，它也会增加。将T/E比临界值设定得太高则没有药物作弊者会被检出，设置太低则许多无罪的运动员将会被错误地指控。

假设某一次检测出错误的概率是1%。这意味着如果这个运动员是无罪的，那么他们不通过检测的概率是1%；如果他们是禁药使用者，那么他们通过检测的概率也是1%。萨姆没有通过测试：她是无罪的概率是多大?

问题这样描述，那么你几乎会脱口而出认为是“1%”——这个测试每100次出现一次错误，所以如果她没有通过测试，出错的可能性就是1/100。你要克制这种冲动。唯一的正确答案是：“我不知道，这个概率可能是任何值。我们需要知道在全体受试者中使用禁药作弊的比例。”

比如，假设这个比例是1%左右。那么在10 000个运动员中预计有100个禁药作弊者，和9900个无罪者。在检测中，我们预计只有1个禁药使用者通过了测试，其余99个都不能通过测试。但是9900名无罪者中的1%，也就是说99名运动员，也会无法通过测试。在那些无法通过测试的运动员中，一半是无罪者：那么萨姆是无罪者的概率就是50%。

如果禁药作弊者比例不同于1%，结论就会变化。如果它升高，萨姆无罪的概率就会降低，但如果它降低，她无罪的概率就会升高。禁药使用者的比例越低，这项测试就会越不令人满意，尽管它显然表现出色。

在我们考虑如何检测出机场中的恐怖分子的时候，相同的逻辑也会奏效。无论什么筛选设备都不会是完美的，但是假设一名真正的恐怖分子规避掉这些检测的概率是很低的，比如说1/10 000，同时一名无辜的人被带进小屋进行严密审讯的概率是1/100 000。那么一个被揪住的人有罪的可能性有多大？

我们在不知道准乘客中恐怖分子的比例的时候没法回答这个问题。试作1/1 000 000——考虑到希思罗机场(Heathrow Airport)每年客流量为5000万，这高得恐怖。但是这个数字使我们确信，即使有50名潜在的恐怖分子，他们全被检测出也是极有可能的。

但不幸的是，500名无辜的乘客也将会被扣押!在那些被检测系统拦截的人中，只有少于10%是恐怖分子。而且如果恐怖分子少于50名，被拦截的人确实有罪的概率甚至更低。检测方法想要有效的话，准确性需要更高。

足球比赛结果(1)

在英国，赌足球比赛的结果会引起人们的极大兴趣。无论赌局多么奇异都会有人捧场——赌第一个球的进球时间、所有进球球员所穿球衣的号码之和、比赛中判多少次黄牌和红牌——但最引人注目的是比赛结果是主队获胜、平局、客队获胜这三种中的哪一种。一个理性的下注者会评估这三种结果各自的概率，而他是否下注、下注大小，还有赌本的赔付金额(payout prices)都会基于这些评估。

但是下注者如何推断出他对不同结果的可信度呢？2009年5月，统计学家大卫·斯皮格霍特(David Spiegelhalter)接受BBC广播节目《更多或更少》(More or Less)的挑战来分析两天后进行的10场英超联赛。对每一场比赛，他都基于对每一支队伍的进攻能力及其对手的防守能力的考察，给出了每支队伍平均的进球数。例如，一个强势的主队，如阿森纳(Arsenal)在对阵斯托克城(Stoke City)的比赛中估计平均会进球2.1个。

没有任何一支球队会进球2.1个，这个数字只是假想数量的比赛中进球数的平均值。最重要的步骤是评估在单场比赛中进球数为0、1、2、3……的概率，斯皮格霍特使用了泊松分布。多年以来的数据显示，这是一个描述实际进球数在其平均值附近变化情况的好方法。在阿森纳平均进球数为2.1时，没有进球的概率是12%，进1球是26%，进2球是27%，进3球是19%，等等。

斯托克城的数据显示他们的平均进球数是0.67。这相当于没有进球的概率是51%，只进1球是34%，进2球是11%，等等。我们姑且相信每支队伍的进球数都是无关的。所以比赛结果为2比1的概率来自主队进2球的概率乘以客队进1球的概率——在上述情况中为27%×34%，大约为9%。

在这种方法中，任何可能的比分结果都被估计了。然后主队获胜、平局、客队获胜各自的概率就可以用加法定理计算得到，即将能得到这三种比赛结果的比分的概率分别加在一起。这就得到了阿森纳获胜的概率是72%，斯托克城获胜的概率是10%，平局的概率是18%。比分结果2比0的概率最高，是14%。

别得意!10场比赛中最高概率的比分才出现2次，而8场比赛的比分都是已被认定为不是最可能的。赌徒已经把钱押在了每个“最可能的

结果”和每个“预测的”确切分数上，即使比赛比分结果未揭晓，他也会笑得很开心。

我们不能让阿森纳进行同一场比赛100次来计算他们获胜的频率，那么我们如何调和阿森纳获胜72%的可信度与概率的频率观点呢？回忆我们是如何评价天气预报的准确性的，播报员说明天下雨的概率是30%。只有一个明天，明天要么下雨要么不下雨。然而，我们可以查看她给出30%降雨概率所对应的所有场合，并计算第二天确实下雨了的频率。我们应该根据她播报的整体记录，选择相信或不相信她对明天天气的播报。在足球比赛中，我们也可以对于整个赛季中进行的所有比赛做类似的计算。在这些比赛中(也许共有40场)，对其中一些比赛我们给出了接近72%的概率——我们可以检查“预测的”结果真的发生了的频率是否在72%左右，这可以作为一种验证我们方法的方式。

一个赌徒通过使用这些方法就能够赢钱了吗？赔付金额很大程度上依赖于每种比赛结果被下了多少注，最大的金额通常都会被下注到一支队伍的获胜或者另一支队伍的获胜。忠诚的粉丝们不会倾向于下注到平局上。如果评估得到的平局概率是25%，而且赔付高于3比1，赚钱的机会就在这里。

所以不要以为最好的下注方式就是押在有最高预测概率的比赛结果上!

足球比赛结果(2)

在2010年国际足联世界杯决赛前夕，统计学家伊恩·麦克海尔(Ian McHale)发表了他的计算结果，为32支球队分别指定了赢得世界杯的非零概率。他认为西班牙队最有可能获得世界杯，尽管只有11.6%的可能性；紧随其后的是巴西队，他对其给出的概率是10.3%。

为了计算得到这些结果，麦克海尔使用了与上述计算类似的方法来计算每场比赛的结果。然而，他并未直接计算每场比赛不同结果的概率，他依靠蒙特卡罗模拟。

英格兰队每场比赛的平均进球数为1.5，泊松分布模型给出没有进球的概率是22%，进1球的概率是33%，等等。计算机的随机数生成器选取0、1、2、3……中的某一个值和与其对应的概率，并对英格兰队的对手做相同操作，得到一些模拟比分结果例如2比2平局。对每一场预定的比赛都进行类似的模拟，得到模拟分组表格，然后进行淘汰赛阶段的比赛直到决赛。这个过程重复了100 000次，每一支队伍成为总冠军的模拟数都被记录下来。西班牙“获胜了”11 633次，因此有前述的11.6%这一数字。像往常那样，大数定律是这个方法合理的理由。

那一年西班牙队的确获胜了!麦克海尔的概率是“正确的”吗？我们不知道，如果能进行无限次重复的联赛，也许西班牙队会在其中的65%获胜。但是能证明其方法合理的最好证据就是，庄家们设定初始赔付金额时沿用同样的方法来吸引下注者。

布莱克-斯科尔斯模型

股票市场上的股价会波动，有些时候没有明显的原因。如果今天价格是5英镑，你不知道下个月会是什么价格。然而，你可以购买**期权**(option)——在一个指定的未来时间按照**行使价**(strike price)5.2英镑购入(或者出售)的权利。如果在未来的那个时候，市场价少于5.2英镑，你应该不会行使期权来购买；但是如果市场价高于这个价格，通过行权买入并立即出售你就可以立即获利。相应的论述也适用于认沽期权。那么这些期权的合理价格是什么？

费舍尔·布莱克(Fischer Black)和迈伦·斯科尔斯(Myron Scholes)于1973年解决了这个问题。他们工作的核心是假设股价会随

机变化，但是变化方式是一种与高斯分布有关联的特定方式。认购期权与认沽期权的合理价格都依赖于当前价、行权价格、干预时间、现行利率和相关价格的波动性(用一段时间内的标准差来衡量)，但是**不**依赖于平均的价格预期值!

最后一点似乎是令人惊讶的，但这就是现实的运作规律。这也是非常有用的，这意味着我们避免了任何因估计价格趋势而引入的不确定性。如果你想要知道一些特定期权的合理价格，有免费的软件可用——只需要在你最喜欢的搜索引擎中输入“布莱克-斯科尔斯”。给定当前价格和行权价格，如果时间间隔变长，利率变高，或者价格波动性升高，一份认购期权的合理价格会随之增加。

刚才那句话与你的直觉相符吗？第一个情况似乎的确是合理的，因为你等待的时间越长，相关价格上涨的概率就越大，但是另外两个情况就更加微妙了。波动性的增高会使得价格上涨的概率更高，但同时也会让价格**下跌**的概率更高——但是前者的效应更强一些。

通过考察这一年中约250个交易日内价格的变化，可以衡量得到波动性。为了使估计是可靠的，应该向模型提供足够的数据，但是也不能向过去延伸得太远，因为那些数据与现在的情况不相关。对于波动性的不良估计会导致对期权合理价格估计的不合理。

只有在假设合理的情况下一个模型才是有效的。如图7所示，认为价格波动符合高斯分布就意味着严重事件的概率是非常小的，例如价格下降了3或4个标准差。当这些事件的真实概率被严重低估的时候，模型就是无效的，而模型导出的结论就根本没有合理的基础。在第4章中提到过的极端值分布可以用来解决这个问题。

投资分配

A公司与B公司都预计会盈利。在低利率的情况下，A预计会收益20%， B预计会收益40%；在高利率的情况下，情况反转——A会盈利40%， B是20%。假设尼克是一名规避风险的投资者，而玛丽喜欢冒险。

如果高利率和低利率可以被认为是同等可能的，两个公司看起来都很吸引人，收益平均都是30%。依据两人对风险的态度，尼克会将资金分成两等份投给两家公司，而且一定会获得30%的收益，无论实际利率是高是低；玛丽会把所有钱投到其中一家公司，并希望会获得40%的收益，但也得接受只获得20%收益的结果。

假设情境中C公司取代了B公司，其会在低利率的时候有10%的收益，或者在高利率的时候有50%的收益——同样的平均为30%，这一点与B公司一样。但是将A与C组合对两名投资者都没有意义：尼克只钟意于A，玛丽将所有的钱投入C。

两种情况的本质区别是，A与B的收益是**负相关**(negatively correlated)的——在一个比较高的情况下，另一个比较低；但是A与C的收益是**正相关**(positively correlated)的——它们一同变好或者变坏。“相关性”(Correlation)以一个在-1(完全负相关)和1(完全正相关)之间的值来衡量。如果每家公司的资产波动变化都独立于另一家，它们的相关性就是0。

规避风险的投资人倾向于分散持股，所以任何损失都会被其他地方的收益所弥补。它们希望持有负相关的资产。但是其中有一个无法避免的逻辑陷阱：如果X与Y负相关，Y与Z负相关，那么X与Z就似乎是正相关了!

然而，还有一线希望。所罗门·博纳(Salomon Bochner)的数学工作结果证明了在很大的投资组合中，每一对投资都是负相关的确是**有**

可能的。但是投资可选的数目越大，两两负相关的情形就越难以达到。

09 有趣而棘手的问题 Curiosities and Di lemmas

在这本书的开头，我曾经说过一些概率问题第一眼看上去有违常识。随着故事的逐渐呈现，相关的例子已经讲述过了。这里呈现一些直觉会产生误导的情况，但是足够小心的话，这些表面上的矛盾都是可以解释的。概率这门学科已经完全不含有真实的悖论了。

但是即使概率知识能够帮助我们做出合理决定，我们也许仍然会发现，就算是考虑某个特定事情的概率也可能会遇到棘手的问题。

帕隆多悖论

格雷厄姆·格林(Graham Greene)的小说《败者为王》(Loser Takes All)是一本好书，但是它基于一个错误的前提：有一些巧妙的基于数学的下注组合方法能让玩家占有优势而不是庄家。但数学家们已经证明，在每一场赌局单独来看都是对庄家有利的时候，无论怎样组合都不会扭转局面对玩家有利。对不起啦，朋友们。

胡安·帕隆多(Juan Parrondo)告诉你，你务必十分明确地阐述如下一般论断：在所有赌局都对一方有利的时候，无论何种情况我们都不可能找到一种组合让另一方有优势。我在这里描述一种他的思想的变体，其来自迪恩·阿斯图米安(Dean Astumian)，他描述了一种依托于一张画有5个格子的纸板的简单游戏，如图11。(这不是一个**认真的**游戏。它的存在只是为了阐述上述观点)

输	左	起始	右	赢

图11　阿斯图米安的游戏纸板

你需要一种生成可能性为1%的随机事件的方式：也许是一个装有99个白球和1个黑球的袋子，或者一个会等可能地停在100个小格中的转盘。游戏开始时，在标有“起始”的格子中放一个标记物。每一次移动都将标记向左或者向右移动一格，如果在到达赢之前没到过输就算胜利。

一共有两组基本规则，我们叫它们A和B。在规则A中，你总是会从起始移动到左；你一定会从右移动到赢；你会在左使用转盘，有1%的概率移动到输，99%的概率移动到起始。在规则B中，你在起始使用转盘，有99%的概率移动到右，1%的概率移动到左；在右，你总是从右移动到起始；你在左时，情况和规则A中一样——转盘给出1%的概率移动到输，99%的概率移动到起始。

对这个游戏的分析是很简单的，在规则A中，没有规则允许移动到右；你在起始和左之间往复移动，直到随机概率使你从左移动到输。在规则B中，你经常在起始和右之间移动，而偶尔移动到左。最终，在其中一次经过左的时候，随机概率使你移动到了输。在这两种游戏中，移动到赢的概率都是0。

在一个新的游戏规则C中，你需要一枚公正的硬币。在每次移动前，掷一次硬币：如果正面朝上，使用规则A；如果背面朝上，使用规则B。

结果是你在规则C中取胜的概率超过了98%!很容易就可以阐明为什么游戏对你十分有利：如果你到了左，你非常有可能安全回到起始。

从起始，有一半可能你使用规则B，有99%的概率到达右；而在右，你有一半概率使用规则A，必然会取胜。

遵循着规则A和B，你**一定**会输：在两种规则之间交替，你几乎每次都会赢!**悖论中**构筑一个排除了上述例子的数学命题需要非常严谨的语言描述，这也就说明了格林的结论其实摇摇欲坠。

2 + 2 = 4，还是2 + 2 = 6?

假设我们用一个公正的硬币来实现伯努利试验，也就是说，每次投掷硬币结果都是独立的，且正面(Heads)反面(Tails)是等可能的。典型的结果是HHTHTTTHT……要掷出H的平均等待投掷次数是2；但是要掷出HT，或者HH的平均等待投掷次数是多少呢？

直觉上讲，答案是4，因为我们预计会等2次投掷来得到第一个标志H，等2次投掷得到另一个标志T。我们等待HT的平均投掷次数的确是4，但对于HH**不是**这样。为了看到这种样式，平均投掷次数是6!

有这种不同是因为，要得到HT，认为我们预计得等2次才得到H是正确的，而要再等2次才得到T从而完成这个样式。2加2等于4。但是对于HH，在我们得到第一个H的时候，下一次投掷有一半的可能性会掷出T，我们就得重新开始了——之前得到H的所有次数就都浪费了。得到正确答案的计算过程在附录中。

H与T两者中一个先出现是等可能的；那在HH与HT之间呢？再一次地一个比另一个出现得早是等可能的，因为我们必须等到第一个H出现，之后下一次投掷决定了最终结果。然而，在HH与TH之中，后者首先出现的可能性是前者的3倍!原因很简单：序列由HH开始的可能性是1/4，但是除非这样，那么不可避免地TH首先出现(思考一下这是为什么)。

彭尼赌局游戏(Penney-ante)就是基于上述的观点。你请你的对手选择8个可能的长度为3的一组结果中的任何一个，比如HHT，或者THT等，它们都可能会是连续3次投掷公正硬币的结果。之后你选择一个不同的结果，一个中立的人重复投掷硬币，选择了首先出现的结果的那个人获胜。

尽管表面上看你大方地允许你的对手先选择，但是这个游戏是对你有利的——如果你知道你应该做什么的话。无论她选择了什么，你都可以选择一个有至少2/3可能性比她的样式先出现的样式!获胜的秘诀在附录中。

给我点暗示……

1. 三张形状大小完全相同的双面卡片装入一个袋子中。其中一张两面都是蓝色，另外一张两面都是粉色，最后一张一面是粉色，另一面是蓝色。随机选择一张卡片，可以看到它的一面是粉色。另一面更有可能是蓝色还是粉色呢？或者说可能性相等？问题交给你吧——下面有回答。

2. 细致的计算表明，从洗好的牌堆中分发出来的一副13张的桥牌，有26%的可能性包含2张或者更多张A。你给露西发牌。对于问题：“你的手牌中至少有一张A吗？”她的回答是“是的”。在另一个情形中，你给蒂娜发牌，并问：“你的手牌中有黑桃A吗？”她的回答也是“是的”。哪一副手牌更有可能包含2张或者更多张A？或者说可能性相等？答案在下面。

3. 假设1000名男性和1000名女性都有令人满意的资质，但有480名男性和仅240名女性获得大学录取资格。这是不是明确的性别歧视——男性被录取的概率是女性的2倍？

答案是什么呢？在粉/蓝卡片问题中，看到粉色就明确地排除了两面均蓝色卡片的情况。所有3张卡片都是完全相同的，只有两张剩下了，粉/粉和粉/蓝。这些卡片中的一张，背面是蓝色的，而另一张，背面是粉色的。似乎粉色和蓝色是同样可能的。

这个推理过程是草率的：粉色的可能性是蓝色的两倍，你可以通过重复进行十几次这个试验来验证这件事。更好的理解是，注意到这些卡片中有3个粉色面，所有这些面都等可能地被看到。但是只有一个粉色面的背面是蓝色——而有两个粉色面的背面是粉色(你可以使用贝叶斯公式，但是那就是杀鸡用牛刀了)。

一名贫困的研究生，瓦伦·韦弗(Warren Weaver)，同时也是**信息论**(Information Theory)的创建者之一，就曾经不断地和其他学生玩这个游戏并赢钱，教育了要他们了解概率的效用。

在牌组问题中，我们知道两个情况中手牌里都有至少一张A，许多人都会认为蒂娜和露西有2张或者更多的A是等可能的——所有的A都是等可能被抽到的，所以为什么蒂娜承认她有特定的一张黑桃A就会带来不同呢？请你丢掉这些想法，来做正确的计算。

对于露西来说，在有至少一张A的手牌中，我们可以算出有2张或者更多A的比例——大约37%。对于蒂娜，除了黑桃A，她还有另外12张手牌，从剩余的51张中随机选取。手牌中包括另一张A的可能性是56%：蒂娜远比露西更有可能持有2张或者更多A。

你的怀疑心告诉你第三个问题正确的答案是“否”。假设在英语系，950名女性申请者中的20%，以及50名男性申请者中的10%被录取了；在商学院，所有的50名女性得到了录取资格，但是950名男性中只有一半被录取。求和得到：240名女性和480名男性被成功录取，但

是，**在每一个院系**，女性的录取率都是男性的2倍。有歧视的话也是针对男性的，而不是女性!

确实，在真实世界中，伯克利[1]研究生院的几千名申请者中，44%的男性被录取，而只有35%的女性被录取。然而，当申请数据被分配给不同的院系的时候，男性和女性的录取比例就几乎没有差距了。但是不同的院系的录取率的确不同，而那些对两种性别都只录取很小比例的院系，女性申请者最多。

这个反直觉的结果是**辛普森悖论**(Simpson’s Paradox)的一个例子。它展示了相较于对绝对数字的操作，对比例的操作是很危险的，这种情况到处都会发生。

这一切都不只是好玩。除非你知道数字的真正含义，不然你没有正当理由来说你会用数字。

你真的想知道吗?

我曾经说概率是在不确定性中做决定的关键，我也不会收回我说的话。但更加精确地理解概率或者在新的情境中理解概率都会带来一些令人不舒服的难题。

现在，个人可以对自己整个遗传密码进行测序，但是诺贝尔奖获得者詹姆斯·沃森(James Watson)和哈佛大学心理学家史蒂芬·平克(Steven Pinker)都选择不去知道他们携带的一种被叫作APOE的基因的版本。有一个epsilon4版本的这种基因会让患上阿尔茨海默病[2]的概率上升4倍，而有两个这种基因会让概率升高20倍。(矛盾的是，有这种epsilon4基因也与一个人某些年轻时的益处相关。)另一名诺贝尔奖获得者，克雷格·文特尔(Craig Venter)，知道他的确有一个epsilon4基因。一家研究实验室有**从不**向志愿者透露其APOE基因情况

的政策，理由是基于现在人类掌握的知识，没有治疗可以减轻其带来的消极影响。

但是一些商业公司也许会对你的APOE基因的情况(实际上是你的全部基因组)**非常**感兴趣。如果你的基因组成暗示着你早逝的概率非常高，它们也许会愿意大幅提高养老金——但是也可能会要求更高的医疗保险费。拥有某个人全部基因信息的公司可能会“提供”定制服务，完全按照客户的预期寿命量身定制。

约翰和汤姆都是65岁，每个人都会花费15 000英镑来购买养老金；比如说正常的剩余预期寿命是15年，但是约翰的基因暗示着长10年的寿命，而汤姆是缩短10年。不考虑基因情况的A公司对两人每年都提供同样1000美元的养老金。但是B公司考虑到了基因信息，向汤姆每年提供3000美元的养老金，但只向约翰每年提供600美元。

回想那条格言：从长远来看，平均统领一切。两个人都会接受更高的出价，所以A公司将会给像约翰这样的人支付25 000美元，每次都会损失10 000美元；同时B公司将会向汤姆和他这个类型的人支付15 000美元，所以收支平衡。A公司将会倒闭。而B公司会生存下来。

如果只有像B那样的公司才能生存下来，那么我们可以预见到那时会有许多不幸的人，他们要么是根本无法支付医疗或者旅游保险，要么因为被提前告知储蓄不足，导致退休计划被严重扰乱。

律师在诘问中应该只去问那些他们已经知道答案的问题。在你想要对自己的基因组进行测序的时候，要确保自己对你可能得到的消息有充分的准备。考虑一下人生中的所有阶段：在孩子出生时的基因组情况的打印件也许会带来晴天霹雳；想结婚时，你和与你订婚的人是否应该去了解你们孩子患有严重疾病的可能性？你的老板是否应该有权利因为你患某种疾病的风险较高而拒绝你的晋升？高级公职人员，

比如总统或者首相的候选人，是否应该公开他们的基因组信息，以便投票者进一步了解任何基因层面的不稳定因素？

随机选择一名英国女性，其罹患乳腺癌的可能性是12%。但是如果她继承了特定的被称为BRCA1或者BRCA2基因的突变，这个概率就会升到60%。一名有3个孩子的母亲，在一名姐妹有这种突变的情况下，接受医学检测，而如果她接受了检测而且收到坏消息，她的女儿(如果有女儿的话)应该在什么年龄被告知她们每个人都有50%的概率继承了这种突变？

无论你在这些令人不舒服的情形中感受到了什么，要记得这只是“概率”而不是“事实”。如果艾玛有这种突变的概率是10%，而菲奥娜的概率是60%，结果也有可能是艾玛患有乳腺癌而菲奥娜没有。如果她们知道自己有这种突变的概率，她们也只能按照自己的方式来处理这件事情。在此重复一下决策论的核心信条：合理的决定是能最大化结果的平均效用。你永远不能确定自己采取的行动会带来最好的结果，但是你已经充分利用了你所拥有的信息。你不能要求更多了。

全书完

[1] 此处应指加利福尼亚大学伯克利分校(University of California, Berkley)。

[2] 阿尔茨海默病(Alzheimer’s disease)，俗称早老性痴呆症，是一种发病进程缓慢、随时间不断恶化的神经退行性疾病。

附录 书中提出的问题的答案

第2章：对于有8或者10个面的色子，掷出偶数的概率是1/2，掷出6的倍数的概率分别是1/8和1/10。掷出3的倍数的概率分别是1/4和3/10，对8面色子来说，掷出偶数和掷出3的倍数这两个事件是独立的，而对于10面色子不是这样。

第6章：在“金钱本色”难题中，你首先应该选择积攒下12 000英镑。如果成功了，你下一次就会希望存下3 000英镑；如果失败了，你下一次就必须尝试存下15 000英镑了。你总体的获胜概率是(6/12) × (10/11) + (6/12) × (4/11) = 84/132 = 7/11，比其他的策略都高。

在有46张相似扑克牌的游戏中，46次重复抽牌中有9次会是黑桃，你会净入账50块；在其余的37次中，你会输掉额外的赌金。如果你多押12块或者更少，长远地看你将会有盈利，但是如果你多押13块或者更多，预期你会输的比赢的多。

第7章：流产的概率是1/m，生出患有唐氏综合征的婴儿的概率是1/n， m和n都很大而且m＞n。只需要看是否有y＞x+n/m。

假设安妮的姐姐或妹妹西莉亚有n个儿子，全都健康。如果贝蒂不是一名携带者，西莉亚的所有孩子都一定是健康的；如果贝蒂是一名携带者，那么西莉亚不是携带者(因此她的儿子们都健康)的概率是1/2，而西莉亚是一名携带者，但是尽管如此她的儿子们都健康的概率是$(1/2)^{n+1}$。所以在有安妮的侄子们都健康这一信息的情况下，贝蒂是一名携带者的后验赔率，就是先验赔率(即利用安妮的兄弟都健康这一信息计算得到的值)乘以后面这些来自安妮姐妹的概率的和[1/2 +

$(1/2)^{n+1}$]。这个值就转化为贝蒂是一名携带者的概率，将其减半就是安妮是一名携带者的概率。

第8章：在随机化回答的例子中，假设吸食者占比为x，而且所有的答案都是诚实的。那么，在第一种情况中，全部回答“同意”的比例是0.8 x + 0.2 (1 - x)；令其等于1/3，解方程得x = 2/9。在第二种情况中，回答“同意”的比例是0.8 x + 0.2/2；令其等于1/3，得到x = 7/24。

第9章：设x为掷出HH的平均次数。为了得到第一个H，平均需要等2次。在这之后，我们至少需要额外的1次投掷；而即使这样，有一半的可能我们需要从头再来。所以x = 2 + (x/2)，即x = 6。

在彭尼赌局游戏中，如果你的对手选择了HHH，你应该选择THH，而你获胜的概率是7/8；如果她选择了HHT，你再一次地选择THH，获胜可能性为3/4；如果她选择HTH，你选HHT，如果她选择THH，你选TTH——在这两种情况中，你有2/3的可能性获胜。利用对称性就可以得到应对TTT、TTH、THT和HTT的最佳选择。

术语对照

A

absolute risk　绝对风险

Additional Law　加法定理

airline overbooking　航班超额预订

Alzheimer’s disease　阿尔茨海默病

amniocentesis　羊膜穿刺术

Astumian’s game　阿斯图米安的游戏

axiom　公理

B

Balls in a bag/urn　袋子/罐子中的球

batch testing　批量测试

Bayes’s Rule　贝叶斯公式

Bayes, Thomas　托马斯·贝叶斯

belief, degree of　可信度

Bernoulli family　伯努利家族

Bernoulli Society　伯努利学会

Bernoulli trials　伯努利试验

beta distribution　贝塔分布

bets　赌注

binomial distribution　二项分布

birthdays　生日

births　出生

blackjack　二十一点

Black-Scholes　布莱克-斯科尔斯模型

Bletchley Park　布莱切利园

Bochner, Salomon　所罗门·博纳

Borel-Cantelli Lemmas　波莱尔-坎泰利引理

Borel, émile　埃米尔·博雷尔

Box, George　乔治·博克斯

breast cancer　乳腺癌

bridge　桥牌

Brownian motion　布朗运动

de Finetti, Bruno　布鲁诺·德·菲内蒂

de Moivre, Albraham　亚伯拉罕·棣莫弗

density　密度　参见probability density

dice　色子

diffusion　扩散

dilemmas　难题

discrete distributions　离散分布

disjoint　不相容的

distribution　分布

Doob, Joseph　乔·杜布

E

eggs鸡蛋

Einstein, Albert　阿尔伯特·爱因斯坦

engineering students　工科生

epidemics　流行病

Erlang, Agner　艾格纳·埃朗

errors　误差

expected utility　预期效用

expected value　预期值　参见mean value

exponential distribution　指数分布

extreme-value distributions　极端值分布

F

fair price　公平价格

Feller, William　威廉·费勒

fencing, Olympic　奥运会上的击剑项目

Fermat, Pierre de　威廉·费勒

Florentine gamblers　佛罗伦萨的赌徒们

fiying bombs　飞行炸弹

football results　足球比赛结果

Fréchet distribution　弗雷歇分布

frequencies　频率

frequency approach　频率方法

G

Galileo, Galilei　伽利略·伽利莱

H

I

impossibility　不可能

independence　独立性

insurance　保险

inverse probability　逆概率

Iterated Logarithm, Law of　《重对数率》

K

Kelly, John　约翰·凯利

Kendall, David　大卫·肯德尔

Kermack, William　威廉·科尔马克

Khinchin, Alexander　亚历山大·欣钦

Kolmogorov, Andrey　安德鲁·柯尔莫哥洛夫

L

Laplace, Pierre-Simon　皮埃尔-西蒙·拉普拉斯

Large Numbers, Law of　大数定律

Likelihood Ratio　似然比

Longcor, Willard　威拉德·朗克尔

lottery　彩票

M

N

Nobel prize　诺贝尔奖

normal distribution　正态分布　参见Gaussian distribution

Number Needed to Treat　需治疗人数

O

objective approach　客观方法

odds　赔率

opinion poll　民意测验

option　期权

Oscars　奥斯卡

P

pairwise disjoint　两两不相交

Parrondo’s paradox　帕隆多悖论

Pascal, Blaise　布莱兹·帕斯卡

payout price　赔付金额

Penney-ante　彭尼赌局游戏

personal probability　个人概率　参见subjective probability

Poincaré, Henri　亨利·庞加莱

points, problem of　点数分配问题

Poisson distribution　泊松分布

Poisson, Siméon Denis　西莫恩·德尼·泊松

poker　扑克

polymer chemistry　聚合物化学

posterior odds　后验赔率

prior odds　先验赔率

Prisoners’ Dilemma　囚徒困境

probability density　概率密度

prosecutor’s fallacy　检察官谬误

pseudo-random sequence　伪随机数序列

Q

queues　队列

R

radioactivity　放射性

random choice　随机选择

randomized response　随机化回答

random numbers　随机数

random walk　随机游走

relative risk　相对风险

Restricted Choice, Principle of　限制选择原则

Richard III　理查三世

risk　风险

roulette　轮盘赌

Rutherford, Ernest　欧内斯特·卢瑟福

S

safety　安全

sex discrimination　性别歧视

share portfolios　投资分配

Simpson’s paradox　辛普森悖论

soccer, World Cup　足球世界杯

Spiegelhalter, David　大卫·斯皮格霍特

square root　平方根

standard deviation　标准差

Star Trek 《星际迷航》

Strong Law 强大数定律

subjective probability 主观概率

sums 和

survival analysis 生存分析

symmetry 对称性

T

tennis 网球

terrorist detection 检测恐怖分子

Thorp, Edward 爱德华·索普

tiny probabilities 微小概率

traffic intensity 流量强度

trick questions 轨迹问题

TV game shows 电视游戏

U

uniform distribution 均匀分布

utility 效用

V

variability　变化

variance　方差

volatility　波动性

W

WADA　世界反兴奋剂机构

weather　天气

Weaver, Warren　瓦伦·韦弗

Weibull distribution　韦布尔分布

Weldon, Raphael　拉斐尔·韦尔登

Z

zero probability　零概率

[1]　原文误作“Gorbachev, Mikhael”

读客

激发个人成长

多年以来，千千万万有经验的读者，都会定期查看熊猫君家的最新书目，挑选满足自己成长需求的新书。

读客图书以“激发个人成长”为使命，在以下三个方面为您精选优质图书：

1．精神成长

熊猫君家精彩绝伦的小说文库和人文类图书，帮助你成为永远充满梦想、勇气和爱的人！

2．知识结构成长

熊猫君家的历史类、社科类图书，帮助你了解从宇宙诞生、文明演变直至今日世界之形成的方方面面。

3．工作技能成长

熊猫君家的经管类、家教类图书，指引你更好地工作、更有效率地生活，减少人生中的烦恼。

每一本读客图书都轻松好读，精彩绝伦，充满无穷阅读乐趣！