

仅供个人阅读研究所用，不得用于商业或其他非法目的。切勿在他处转发！

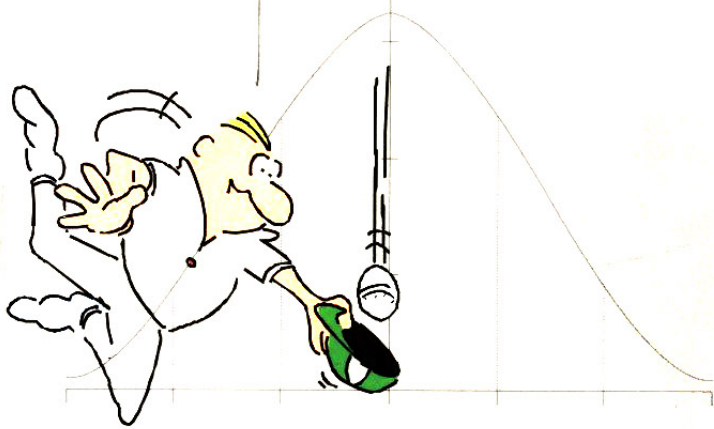
本电子书制作者

WHAT ARE
THE CHANCES?



机会的概率 有多大

[美] 巴特·K. 霍兰 著
陈 林 译



暨南大学出版社

WHAT ARE
THE CHANCES?



机会的概率 有多大

[美] 巴特·K. 霍兰 著
陈 林 译



暨南大学出版社
Jinan University Press

中国·广州

© 2002 The Johns Hopkins University Press

All rights reserved. Published by arrangement with The Johns Hopkins University Press, Baltimore, Maryland. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, or by any information storage and retrieval system, without permission in writing from Jinan University Press and The Johns Hopkins University Press.

图书在版编目 (CIP) 数据

机会有多大/[美]霍兰(Holland, B. K.)著;陈林译. —广州:暨南大学出版社, 2005. 11

书名原文: What are the chances? —Voodoo deaths, office gossip & other adventures in probability

ISBN 7-81079-626-7

I. 机… II. ①霍…②陈… III. 概率论—普及读物 IV. 0211-49

中国版本图书馆 CIP 数据核字 (2005) 第 111556 号

出版发行: 暨南大学出版社

地 址: 中国广州暨南大学

电 话: 总编室 (8620) 85221601 85226581

营销部 (8620) 85227972 85220602 (邮购)

传 真: (8620) 85221583 (办公室) 85223774 (营销部)

邮 编: 510630

网 址: <http://www.jnupress.com> <http://press.jnu.edu.cn>

排 版: 暨南大学出版社照排中心

印 刷: 湛江日报社印刷厂

开 本: 787mm × 1092mm 1/16

印 张: 11.5

字 数: 12.5 千

版 次: 2005 年 11 月第 1 版

印 次: 2005 年 11 月第 1 次

印 数: 1—5000 册

定 价: 18.00 元

(暨大版图书如有印装质量问题, 请与总编室联系调换)

目 录

内容简介 / 1

序 言 / 1

致 谢 / 1

1 轮盘赌与大瘟疫

概率在预测中的作用 / 3

连锁反应 / 7

多变性和预测 / 10

从可预测的概率中获益 / 12

What are the chances?

— Voodoo deaths, office gossip & other adventures in probability



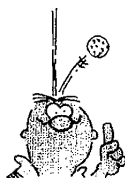
- 赌徒的谬误 / 15
- 随机变量及其分布函数 / 17
- 二项式方程 / 21
- 预测罪行累犯 / 28
- 我们的失误与星球有关吗 / 32

2 你肯定有问题

- 检验结果 / 46
- 天体分布与街头小面包房 / 51
- “正常”智力 / 57
- 爆米花与抽样平均分布 / 61

3 你可押注的生命表

- 毕生的交易 / 69
- 保险与亵渎神灵 / 71
- 格兰特生命表 / 73
- 从死亡率到预期寿命 / 75
- 寻找普遍规律 / 79
- 与自己的存活率打赌 / 80
- 就是你这种类型 / 83
- 可供选择的受孕方式 / 86



4 稀有事件

- 是不是令人震惊 / 93
- 稀有事件的统计 / 97
- 泊松分布 / 103
- 使不可能的事变为必然之事 / 111
- 奇迹 / 116

5 等候的游戏

- 使你发狂 / 123
- 一些简单排队的分析 / 127
- 更复杂的模式 / 131
- 排队中的心理学 / 134

6 股票经纪人与气候变化

- 人迹罕至的路 / 141
- 赚钱从这里开始 / 148
- 一场激烈的争论 / 161
- 对“维纳过程”的简短、直率评价 / 167
- 死海和热带植物 / 168
- 完美的模型 / 172



内容简介

我们的生活受概率支配，但是概率究竟是什么东西呢？本书中杰出的统计学家兼作家巴特·K. 霍兰带领我们周游概率世界。将现实生活中的故事编织在一起——从中世纪欧洲黑死病的传播、普鲁士骑兵军团中被战马踢死的人数、智力测验的结果、伏都教死亡魔咒，到我们为什么在迪斯尼乐园排队骑木马——霍兰用日常事件中惊人的概率实例，抓住了读者的想象力，这些概率对我们的生活产生深刻影响，但却受控于一个数字。

正如霍兰解释的那样，即使是偶然事件也受制于概率规律，并遵循所谓的统计规律。他告诉我们保险业、司法系统、医学研究、航天工程和气象学等不同领域如何成功地应用这样的规律，以获取更大的利益。不管你是对高中代数仅

仅有些模糊记忆，还是每天使用微积分公式，这本书提供的概率影响力实例都会让你感到有趣又吃惊。

本书译者为暨南大学外国语学院陈林老师，武汉大学大学英语教学部何博老师参与了第一章部分内容的翻译。

序 言

我一直没能想好自己最喜欢研究什么课题，以致现在还无法作出决定。也许这是出自某个专业领域的现职教授（在其 50 多岁的时候）的一种奇怪的坦白。即使在学术领域都潜在着各种各样既有用又有趣的工作可供选择，注重某一特定课题而排斥其他课题似乎是非常主观的。另一方面，专业化也是职业市场的需要，杂而不精的三脚猫是不会有成就的，要达到研究高产就要求专业化。

对于有科学头脑的人来说，解决这一难题的方法是在概率和统计方法方面接受良好的训练。这个见解既不是我随意得出的，也不是凭直觉得出的，而是逐渐、偶然得出的。在读大学期间，我学了许多门人类学方面极富启迪性的课程，接着学了人口统计学。这样的学习使我惊讶地发现在研究社

会现象的同时，规律是如何从个人行为中产生的。随后，我在生物学方面也发现了同样的现象：概率规律可以归纳、预测像种群遗传学和对药物反应这些差异如此之大的现象。在经济学、物理学、地质学、心理学等学科领域，你都会发现，概率是非常有用的，它对人类的理解思维起着十分重要的作用。

目前，我主要将概率统计用于医学研究，为医生提供实验中的有用信息，这些信息即使不是很混乱，似乎也是相当难解释。但是，我从来没有失去过对其他研究领域的兴趣，而总是将概率应用的有趣（更准确地说是奇特）的例子记录下来。正是 Trevor C. Lipscombe 和约翰斯·霍普金斯大学出版社，才使我终于有机会将我所喜欢的一些实例写出来。我希望读者会发现这些实例能引起兴趣，并且发现这是学习概率的基本概念和统计方法的有效方式。当你看到概率对人类生活这么多方面产生如此巨大的影响时，也许会与我一样感到惊讶。

致 谢

我的编辑 Trevor C. Lipscombe 博士是第一个建议我用历史、医学和日常不同的例子写一本这样的书，来阐述概率统计的原理（因此，如果你不喜欢这本书的话，你应该责备他，而不是我）。幸运的是，我们有许多共同的兴趣，因此有足够的机会从所收集的资料中交换一些趣闻轶事，共同讨论如何将其安排到书中。我们对语言使用的观点也是一致的，这种一致性使我很容易接受他那娴熟的编辑建议。

我最忠实的读者是 Jean M. Donahue 博士，我要感谢她对书稿作极富深刻见解的审阅。作为科学家、教师、诗人和妻子，她分享我的工作和我对书的嗜好。在这一点上，我真的很幸运。她的父母 Nora 和 Charles L. Donahue 在我感到强烈的写作冲动时，帮我照料孩子，使我安心写作。因此他们对

本书能够完稿显然作出了直接的贡献。

我还要感谢 John Bogden 博士和 William Halperin 博士 (William Halperin 博士在新泽西医学院我们系连续当了几任系主任) 对我的鼓励。此外, 我的同事 Joan Skurnick 博士提出了不少有用的建议, 并提供了那个她在暴风雪的日子里所遇的有趣故事, 这个故事被作为有关随机游动那一章的例子。

最后, 我要感谢新泽西铁路运输公司——由于火车晚点才使我有时间来写作; 感谢我的朋友 Elizabeth R. Nesbitt 为 Esmeralda 命名; 感谢乔治·F. 史密斯图书馆工作人员让我获得了许多难得的资料。

！ 轮盘赌与大瘟疫



概率在预测中的作用

那情景真是令人毛骨悚然。一具具肿胀不堪、发紫发乌、散发着腐臭味的尸体由弩炮大力掷向空中，沿着抛物线的轨迹飞落到被围困的城池内。该城即克里米亚半岛上的卡法城（今乌克兰境内的费罗多西亚）。在 14 世纪，它是热那亚商旅的大本营。那时它正遭受蒙古军队的围攻，一如此前的数次被袭。在 1344 年的围攻中，这座城市还近乎坚不可摧，然而仅仅两年之后情势就发生逆转，这次伴随着中亚铁骑而来的还有黑死病。大批的鞑靼侵略军死于此病，堆积如山的尸体同时又造成严重的卫生问题。于是有军事天才提出了解围妙计。这些蒙古人随军带有一种被称为“投石机”的强大弩机，通常用其投掷沉重的石块来摧毁城墙和塔楼等石

砌的防御工事。而今，“人体导弹”取代石块雨点般落到坚守在城墙内的人的身上。一个名叫加布里埃尔·穆斯的目击者在一份拉丁文手稿中描述道：很快地，如山的死人堆里增加了大批誓死守城的基督徒们的尸体，只有那些侥幸得以逃生的人躲过了恶臭与疾病。

卡法城的故事并不只是人类细菌战的一个早期案例。一些历史学家和流行病学家相信这次战役标志着鼠疫开始由中亚传入欧洲。那些逃回欧洲的热那亚人很可能经由所乘船只上的老鼠及老鼠身上的跳蚤（这些跳蚤在咬人的同时将耶尔森氏鼠疫杆菌传入人体血管）将病菌带回了家乡。无论起源如何，1348年的欧洲大鼠疫的确是从地中海沿岸港口城市爆发并传播开来的。从当时僧侣的著述及教区死亡记录可以得知：被鼠疫夺走生命的人的总数占欧洲人口的25%~50%。但是，我们永远也无从确定黑死病传入欧洲的实际路径。

虽然黑死病爆发于数个世纪以前，但它所提出的问题至今仍为我们关注。为什么会“爆发”流行病？为什么科学家无法预测出某种“旧”型流行病（比如流感或麻疹）下一次爆发的时间、地点和规模？为什么他们更难预测出诸如艾滋病之类的“新”型流行病的发生？天气预报必须依赖于大气环流及海洋的恰当模型，而流行病预报的困难就在于很难建立起精确而科学的感染模型。只有当特定的一系列事件发生后流行病才爆发；由于每一事件都有一定的发生概率，因此疾病的爆发也就具有一个平均或预期的发生频率。为了预

测流行病，我们需要建立准确的流行病发展的数学模型。要建立数学模型，就需要了解整个链条中的每一环节及其各自的发生概率，然后将所有环节的发生概率与特定区域所有相关人员的总数相乘，即可测算出整个事态发展的预期结果。举个简单的例子，让我们设想某一疾病在人际间传播的途径，比如说通过打喷嚏感染流感病毒。如果人群中每一感染者都平均接触并传染一个健康人（流行病学家称之为“易感者”），一场流行病就会爆发。要是每个感染者平均接触并传染一个以上的易感者，每一个新的感染者随后又接触并传染数个易感者，如此循环往复，则流行病将会大规模蔓延开来。倘若每个被感染者平均下来不能“成功”传染一个新个体，则流行病会渐渐衰亡。一个简单的链状概率模型便能反映上述情况。

另一个密切相关的例子源于生活中的幽默。假设你编了个笑话讲给几个好友听，要是这笑话一点儿也不好笑，它就不会传开。但如果你的朋友听后捧腹大笑，而且其中每个人在 24 小时内转述给其他两个人听，则 24 小时后在此单链状模型中听到该笑话的人数就增加到 2；依此类推，48 小时后听到的人数又会增加到 4，3 天后增加到 8，7 天后就会有 128 个新人听过你的笑话了。听起来很惊人不是吗？别忙，让我们再看看——到第二周结束时，听过这笑话的人数将超过 16 000 人，到月底时这个数字将达到 2.5 亿人（大致相当于美国总人口数）。真的会有这么多人听到吗？有多少次你

刚刚兴冲冲开个头，就会有人说“我早听过了”或“这笑话一点不好笑”？事实上，人口数量是有限的，同时，一些人会对某一疾病具有免疫力或是对某一笑话无动于衷，这些因素极大地影响了我们所建立的试图用来解释某个笑话或疾病如何得以传播的数学模型的计算结果。情况若非如此，对人类而言则既有好消息又有坏消息：好消息是我们大家都可以靠连锁信发家致富，坏消息如同 1348 年的黑死病也许会将整个欧洲人口灭绝殆尽。

这种连锁反应机制的另一有趣例子是闲话在工作场合的传播。不过，流言蜚语从本质上与传染病有更多共同之处。你听到了一条绘声绘色的小道消息，于是告诉了几个密友，他们随后又会讲给别人听。一传十、十传百，传到最后这个消息会与最初时大相径庭。譬如，你听到有关克瑞格和莫琳的某桩趣事并讲给其他人听，一个月之后你听到的却是一段添油加醋的有关格里格和诺琳的故事。你能听出这是个被传得走了样的版本吗，或是会把它当作新闻大热门立即通过电子邮件转告你的朋友？用生物遗传学的语言来描述，闲话已经发生了变异，然后如同发生变异的病毒，它可以再感染那些感染过初始病毒的人。流感发生的情形就是这样，因此疫苗每年都必须更新以有效预防新出现的病毒变种。

连锁反应

连锁反应对于理解科学领域的许多重要变化过程至关重要。同样的模式——通过链条上各环节相继发生的几率来计算所有可能发生结果的概率——也主宰着核物理中的链式裂变反应。当一种放射性物质的重原子核受到一个中子撞击时会分裂成两个质量较轻的原子核，同时释放出数个中子；中子再去撞击另一个原子核，就会引起新的裂变，释放出新的中子，以此类推，裂变不断持续下去形成链式裂变反应。在自然界中，放射性衰变时时都会发生，中子被释放出来但链式裂变反应却并没出现：由于放射性同位素无法高度聚集在一起（非放射性同位素不能像放射性同位素那样分裂并释放中子和能量），反应会逐渐衰减。因此，所释放出的中子大体上都会“温和”地穿过物质而不造成原子核的裂变，从而也不会有更多中子和能量的释放。这些中子即使不被用来激发连锁反应，也能发挥其他重要作用。放射性同位素被广泛运用于医药领域：因为某些放射性物质在注射或口服后会被人体的特定骨骼、器官或组织所吸附。此时将感光乳胶底片或其他辐射敏感装置放置到接近人体的部位，被释放出的大量粒子会在底片上造影成像，使得癌症和其他疾病的诊断成为可能。此外，放射性同位素还可用于治疗某些癌症，因为

辐射能够杀死癌细胞，同时治疗所针对的人体组织对其的选择性吸收也是一种很好的特质。

二战期间由莱斯利·格罗夫斯将军领导的美国军方曼哈顿计划的核心任务之一就是找出生成持续的链式裂变反应的途径，从而制造出原子弹。致力于此计划的科学家们分离出放射性同位素后将其浓缩，从而增加中子释放的概率及随之而来的整个链条上的核裂变反应。同时，要预测并提高每一个所释放的中子轰击并分裂下一个原子核的几率，从而引发更多中子和能量的释放，科学家还必须测量出原子核的直径，并压缩原子使之尽可能彼此靠近。当时，包括罗伯特·奥本海默在内的一批世界顶尖级物理学家为此倾注了他们的全部心血。1945年7月16日清晨5:30分不到，第一颗原子弹在新墨西哥州的洛斯阿拉莫斯爆炸成功。一个月之内，“大块头”和“小男孩”（原子弹名）就在日本上空被投下。

通过类比可以得知：核裂变连锁反应的发生机制与疫病的流行机制如出一辙，持续裂变/传染的前提是某一个体的原子/人必须撞击/传染靠近其的另一个体。在人类历史的早期阶段，大规模疫病即使爆发过也是极为罕见的。在人类从事采集和狩猎的史前时代，人口分布非常稀疏。一个小部落通常由为数不多的几个家庭组成，倘若某个部落成员感染了麻疹病毒，整个部落就会被传染。但疫病爆发期间他们遭遇并感染其他部落的几率就很低。平均说来，那时感染上传染性疾病的人几乎是没有机会遇上并传染另一易感人群的。因

此疫病很快就会停止传播。有证据表明，最初此类疾病鲜有发生，而后随着人口的持续膨胀，传染的概率不断增加，疫病的发生愈来愈频繁。公元前 4000 年的底格里斯河和幼发拉底河流域由于农耕文明的到来而出现了星星点点的小城邦，古病理学家从该时期人类遗骸上找到了麻疹遗存的证据。病理学家们对古埃及木乃伊的解剖发现：公元前 1000 年左右，当牛这种动物刚刚被人类驯化并放养时，古埃及人就知道结核病的存在了，这种传染病很可能是从结核病菌肆虐的牛圈传入人类社会的。只有在达到一定规模的人类聚居地，传染病的人际间流行才具备了最理想的条件：在较大的人口群落内，某个人成为传染源的可能性也较大，同时他与那些通过生育和移民而不断得以更新和流动的易感人群接触的机会也更大。至少需要一个村子的规模才能爆发一场流行病。

在任一给定规模的人口内部，某种疾病的易感人群所占的比例对该病流行的最终规模起着关键作用。对于北美洲印第安原住民而言，天花并非本土“特产”，所以没有人能依靠感染哪怕是最轻微的天花病毒而获得对其的免疫力。因此，当欧洲殖民者踏上美洲时，几乎所有印第安人都是天花的易感者。许多欧洲人因为有过天花接触史而具备了对其的免疫力，另一些人则本身携带有活性病毒。1763 年，英军北美总指挥官杰弗里·艾莫斯特爵士与负责宾夕法尼亚拓垦地军事行动的亨利·布科上校在多次通信后拟订了一项计划，

阴谋向印第安原住民部落提供带有天花病毒的物品。历史再次重演：面临强敌和病魔双重威胁的军队想出了以毒攻毒的计策。在信中布科如是汇报：当地人“肆意践踏我们的新拓居地，毁坏庄稼，屠杀成年男女和儿童”，同时军队正经受一场天花的劫难。布科和艾莫斯特议定此项计划旨在用“和平馈赠”的方式安抚（和消灭）敌人，所谓“和平馈赠”就包括从匹特堡天花医院搜罗来的毛毯、手绢等物品。在1763年7月16日的信中艾莫斯特写道：“你完全可以利用毛毯来姑息安抚这些印第安人，同时动用其他一切手段将其斩草除根。”于是，此前在俄亥俄和杀瓦努部族历史上闻所未闻的一种疾病——天花——爆发了。究竟是带菌的馈赠品制造了这场疫情，还是当地人在与白人殖民者的其他接触中偶然染病，答案尚不明了。但从1763年下半年至次年年初，大批大批的印第安人被天花夺走了性命。

多变性和预测

如果导致传染病爆发的因素广为人知，为什么我们依据现有概率模型预测疫情的工作不尽如人意呢？从统计学角度分析，这些模型中存在太多参数，每一参数的变化繁多，因此一项预测就存在太多的不确定性。通过长期观察研究，科学家可以对某一微生物体非常了解。他们知道它的表层分子

构造，洞悉其生物毒素的化学组成，掌握其新陈代谢的规律，甚至可以精确描绘其 DNA 序列——然而，谁知道一个病毒感染者接触到一个易感者的概率是多少呢，无论是在两河流域还是在纽约地铁系统内？谁又知道这个易感者被感染（这是他再传染其他人的先决条件）的概率是多少呢？易感者染病的概率部分取决于病菌的传染性，而病菌的传染性又是由其 DNA 构成所支配的不同化学和结构特征决定的，因此具有多变性。打个喷嚏会喷出多少病菌？其中有多少会被人体吸入？导致感染的最小“剂量”又是多少？至于这后一个参数，专门从事细菌战研究的人已经估计出多种细菌的“最低致病量”。不过，在自然环境下，人与人体质的差异（甚至同一个人在不同状况下体质的差异）也会影响感染的概率。如此众多的参数所固有的多变性自然决定了概率极大的不确定性。

对于依赖于传染病媒介（即携带有病菌并传染给人类的动物）的疾病，其传染途径上的每一环节都已经被科学地鉴别出来了，比如鼠疫。相关有机体的生命周期也为人知晓。但是我们对于各个参数的估量却有太多不确定性，而且能够估量准确的可能性如此之小，以致疫情预报几乎是不可能的。在极为特殊的情况下，比如不计其数的鼠疫受害者和老鼠在欧洲大陆四处流窜时，或是沾染了天花病毒的毛毯被“慷慨馈赠”给从无病史的人群时，你不需要太多信息就能够意识到威胁正在逼近。可惜现实情况鲜有如此一目了然

者。即使有，通常也是到情势危急时才认识到（而且这种认识是非常粗略的）。事实上，流行病学所涉及的因素不计其数，对于这些因素的定量认识甚至远比核物理学领域的少。正因为如此，我们可以在计算链式裂变反应的基础上成功地研制出原子弹，但却无法准确预测出流行病的爆发。

从可预测的概率中获益

颇为奇怪的是，某些事件序列很少由我们可以实际估算出的生物学或物理学参数支配，相反，几乎无一例外地取决于我们所谓的“随机选择”；然而在某种意义上，它们又属于最可能准确预测的现象范畴。以赌博为例。我们不需要科学家的队伍来研究硬币、纸牌或是轮盘赌盘的物理特性，但赌场或彩票商却可能预先准确估算出中奖人数及长远收益。这里无须多少详尽的科学知识——仅仅凭借统计规则性即可，因为没有什么个别现象的变异率会对结果产生重大影响。比起生活中众多其他现象，轮盘赌要简单明了得多。

轮盘赌中的事件序列与流行病中的事件序列还有以下显著不同：每一次转轮盘，每一次出结果，都与前一次不相关联。轮盘赌机并没有记忆力，只要转盘没有被人作弊“固定”在某个数字，只要游戏规则是公正的，那么这一盘你是赢是输对你下一盘不会有任何影响。而在传染病流行模式中

情况截然不同：朝你打喷嚏的那个人是否从前一个朝他打喷嚏的人那里感染了病毒，这一点显然至关重要！

18 世纪致力于永动机发明的法国哲学家、科学家布莱兹·帕斯卡是轮盘赌的发明人。他设计的这个装置让成千上万的人发财致富，同时又让成千上万的人倾家荡产。发财的人接二连三地走好运，倒霉的人则一而再再而三地走背运。如果轮盘赌机没有记忆力，如果每次转盘彼此互不关联，怎么会有接二连三走好运的事发生呢？如果运用概率法则，我们能否预测这样的事件序列呢？要回答这些问题，让我们仔细分析一下一连串好运背后的秘密吧。为了说明仅凭统计概率就能导致这类情况发生，我们要选择一种简单纯粹、完全靠运气的活动来建模。足球或是垒球这一类的集体运动项目不适合我们的分析，因为在这类项目中有太多错综复杂的因素同时起作用，比如与之对阵的球队顺序如何、比赛所需时间的长短或是球迷热情程度的高低等，都会对球员的场上表现产生影响。而且全体球员必须密切配合，这更为建模增添了难度。这一切都会紧随我们所谓的“纯属运气”（即影响球队成绩的内在随机变数）发生。所以还是让我们选择一种靠运气的游戏使模型尽可能简单化，看看我们能否预测出好运频频出现的几率吧。

蒙特卡罗赌场的轮盘赌机上共有 37 个小槽，编号从 0 到 36 转盘每转一次停下后，盘上的小金属球就会落进其中某个小槽。赌注可以押在单数或双数上。当然，如果我们只

考虑 1 至 36 这些数字，其中单数和双数各 18 个，那么我们自然认为赌场的经营会赚赔相当：因为平均说来，一半的赌注会押在单数上，另一半会押在双数上，而赌场会把从这一半上赚到的钱赔到那一半上去。然而，0 这个数字才是确保赌场经营轮盘赌只赚不赔的秘诀。0 既不是单数也不是双数，如果金属球落进 0 号小槽，赌场就会将押在单数和双数上的所有赌注尽收囊中。因此 37 个小槽中有一个能保证赌场坐收渔利。金属球落进 0 号槽的概率为 $1/37$ ，轮盘每天大约转 500 次，那么平均下来赌场每天会有 13 至 14 次的机会通吃整场赌注。美国赌场的轮盘赌更有利于赌场赚钱，轮盘赌机上既有 0 号槽又有 00 号槽，因此所中数字既非单数又非双数，从而赌场通吃所有赌注的概率为 $2/38$ 。无论哪种情况，赌场的收益率都是很可观的（ $1/37$ 意味着 2.7% 的收益率， $2/38$ 则为 5.26% 的收益率）。从长远来看，这些百分数可以被视为相当准确的预测。不过，偶尔也会出现令赌客和赌场都瞠目结舌的事件序列。

1873 年，蒙特卡罗一家名为“纯艺术”的赌场就发生了这样的事件。一个名叫约瑟夫·贾格斯的英国工程师赢了一笔巨款。他的助手提前一天到赌场，记录下当天出现的所有数字。贾格斯仔细研究这些数字，试图摸索出其中隐藏的非偶然性规律。六台轮盘赌机中有五个运作都非常正常，但第六台上却有九个数字被选中的几率远远高出一般概率。第二天贾格斯来到赌场，在那台赌机上专押这九个数字。到第

四天结束时，他已经赢了 30 万美元。贾格斯去世的那一年，英国杂耍演员查尔斯·科本恩凭借一曲“那家伙专抢蒙特卡罗的银行”而声名大噪。

贾格斯之所以交好运，并不是赢在数学上，而是赢在物理学上。那台轮盘赌机上有一条小裂缝，正是这条裂缝让那九个数字出现的频率高于统计学的估算。从那以后，蒙特卡罗赌场里的轮盘赌机每天都要由专业技师检查调试，以确保所有数字被选中的几率相同。

赌徒的谬误

蒙特卡罗赌城曾见证过一次极不寻常的轮盘赌，而且这次并非由于轮盘上有裂缝。1913 年 8 月 18 日，双数连续出现了 26 次。考虑到轮盘每一次旋转可能出现的 37 个数字中有 18 个双数，某一双数在一轮中出现的概率就是 $18/37$ 或 0.486（单数同此），连续出现 26 次双数的概率为 $18/37 \times 18/37 \times 18/37 \dots$ ，连乘 26 次，结果是 0.000 000 007，约等于 $1/142\ 857\ 000$ 。假如真的有谁胆大到连续 26 次坚持把赌注押在双数上，同时极有先见之明地适时停止下注，那会多么幸运啊。然而，怀着单数迟早会出现的期待，蜂拥而至的赌徒们在不同时刻纷纷放弃了双数，直至最后无人能从这场赌局中获益——除了赌场之外。

双数连续出现 26 次能够称之为“运气”吗？是不是那个受到命运眷顾的人偏偏愚蠢地将天赐良机拒之门外了呢？请记住：一台轮盘赌机每天大约转动 500 次，而且每天都有 4 台、5 台甚至更多台同时开转。一年四季赌场几乎天天营业，迄今已运营 125 年之久。终有一天会出现不同寻常的结果，而这种结果的产生完全不依赖任何特别的机制：除了运用概率法则将逐次概率相乘，其他任何解释都是多余。要预测下一步会发生什么情况简直是白费气力，因为这种结果太过稀有——简直是百万分之一的偶然几率，在基数够大的情形下，甚至是五亿分之一的几率！从这个意义上说，结果是可预测的——我们估算各个数字出现的概率，假如所分析的样本足够大，就可以根据数学推算找出极端偶然事件的出现几率。比起预测某种罕见疾病的爆发或者整个棒球赛季的输赢结果，预测轮盘赌中某种异乎寻常的结果的出现要容易得多。很难想象还有比 $18/37 \times 18/37 \dots$ 如此连续相乘更简单的概率链模式了。

我们之所以能够进行这样的运算，完全是因为在轮盘赌机正常公正地运转时，每次转盘的结果是互不相干的；由于每一次转盘的结果都与前一次无关，每一数字出现的概率是固定不变的，所以我们无须了解其他情况。但是这种完全的无关性对于赌徒而言却意味着绝望：既然这一把的输赢结果对下一把毫无影响，他们就无从知晓下一把可能发生什么情况，诸如“单数早该出现了”之类的想法被视为赌徒们惯有

的谬误。在轮盘赌中，小球在轮盘上的每一次转动都是一次全新的开始，无论是在第一次还是在第一百万次。这确实有点矛盾：当所有可能结果均考虑在内时，依照概率原理所作出的预测，我们可以准确得知小球转动可能出现的某些奇特结果的概率；对于单次结果的具体细节，准确的预测却是不可能的。这些统计学数据仅仅有助于预测出现某种结果的概率。

随机变量及其分布函数

如果能借助随机变量的分布函数来研究随机事件，对随机现象产生的独立结果的分布模式的预测就会简单许多。分布函数是一种非常便捷的工具，我们可以将其视为一序列的随机结果及每一结果的概率的组合。在日常生活中随意抛掷两枚普通骰子得到的所有结果组合就是一例随机分布函数的例子。由于每一枚骰子有 6 个面（数），所以可能出现的数字组合为 6×6 即 36 种，而且只要没有在骰子上做手脚，这 36 种组合中每一组出现的几率都是相同的。但是，不同数字组合相加所得的数值均在 2 ~ 12 之间，而这些和才是我们掷骰子的结果。由于其中某些和是可以通过不同组合的数字相加所得的，所以显然这 11 个和出现的几率不会相同。表 1 - 1 描述了掷骰子的 11 种结果以及各种结果是如何组合

的及其分布的概率。

表 1 - 1 掷骰子的可能结果分布

总数	获得该总数的方式 (第一枚骰子,第二枚骰子)	获得该总数的概率
2	(1,1)	1/36 或 0.028
3	(1,2);(2,1)	2/36 或 0.056
4	(1,3);(3,1);(2,2)	3/36 或 0.083
5	(1,4);(4,1);(2,3);(3,2)	4/36 或 0.111
6	(1,5);(5,1);(2,4);(4,2);(3,3)	5/36 或 0.139
7	(1,6);(6,1);(2,5);(5,2);(3,4);(4,3)	6/36 或 0.167
8	(2,6);(6,2);(3,5);(5,3);(4,4)	5/36 或 0.139
9	(3,6);(6,3);(4,5);(5,4)	4/36 或 0.111
10	(4,6);(6,4);(5,5)	3/36 或 0.083
11	(5,6);(6,5)	2/36 或 0.056
12	(6,6)	1/36 或 0.028
合计		1

许多随机变量的分布函数由于过于复杂而无法以表格形式表现。其中有一种被称为“二项式分布”。之所以称之为“二项式”，是因为它所描述的随机结果的概率是两种范畴 / 类别合为一体的。但是，在最简单的情形下，例如在抛硬币时，正面朝上和背面朝上的几率是相同的，分布函数的值仍可以用轮盘赌的发明人帕斯卡所研制的一种表格形式来计算。表 1 - 2 就是这种被称为“帕斯卡三角形”的表格形式。

表 1-2 帕斯卡三角形

试验的次数	概率分子	概率分母
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256

以下是解读该表的内容：假设抛硬币的次数（或称试验数）为 1。在表上同一行三角形顶端的两个 1 是分子部分，分别代表正面朝上和背面朝上两种随机结果。随机结果的总数显然为 2，即分母部分永远为 2，那么正面朝上和背面朝上的概率都为 $1/2$ 。如果试验两次（即抛两次），则如表中下一行所示，可能出现的结果有 3 种。其中两种结果用 1 代表，其出现的概率分别为 $1/4$ ，它们属于“两个正面朝上”和“两个背面朝上”的特殊情况。第三种可能结果是“一正一反”，有两种不同方式出现（先正后反或先反后正），因此以 2 表示，其概率为 $2/4$ 即 0.5。最后，假设我们要抛四次硬币，则试验数为 4。第一次抛可能出现的结果有两种，正面朝上或者背面朝上，第二、三、四次亦然。因此，我们可能见到随机结果的组合总数为 $2 \times 2 \times 2 \times 2 = 16$ ，亦即分母为 16。帕斯卡三角形列出了分子部分，那么我们可以据此计

算出 16 种组合的概率：四次正面均朝上的概率为 $1/16$ ，一次正面朝上三次背面朝上的概率为 $4/16$ ，两次正面朝上两次背面朝上的概率为 $6/16$ ，三次正面朝上一次背面朝上的概率为 $4/16$ ，四次全是背面朝上的概率为 $1/16$ 。注意，由于该三角形呈中轴对称，所以选择哪一面为正面哪一面为背面无关系。在计算器出现以前，能够建立这样的一种表格极其有用，而且它非常便于记忆。只要在二角形的外边写下数字 1，则其他任何数字都可以通过其左上方和右上方的两个数字相加求和而得。这一计算过程适用于任意大的试验次数。

不过，倘若情况不是抛硬币这样正反结果各一半又如何呢？尽管帕斯卡三角形是一项伟大的发现，但它只能适用于“结果五五对开”的事件，何况随机事件的结果千千万万（试验次数也是一样），一张能覆盖一切可能情形的表格恐怕得印成厚厚几大摞书，这显然是不现实的。下面所举的这个例子将有助于我们认识方程式作为一种数学表达方法带来的经济实用性。只需通过一个小小的方程式，我们就能生成所有可能的二项式分布，不论试验次数是多少，也不论随机事件的概率是多少。假设某一家制药企业宣称有一种新药如依医嘱服用则治愈率将高达 80%。某医生现有三位互不相识、彼此也不会发生任何联系的病人，他们都依医嘱服用了此药，但结果仅有一人病情好转。很令人吃惊吗？这个临床观察的治愈率（33%）是否与制药公司的宣传相悖？让我们用二项式方程来回答这些问题吧。

二项式方程

假定 p 代表一次成功治疗的概率，据制药企业称 $p=0.8$ 。如果 q 代表一次治疗失败的概率，则 $q=1-p=0.2$ 。试验次数用 n 表示，此处 $n=3$ ；治愈成功的结果总数为 r ， $r=1$ 。问题是，在基本治愈率为 0.8 的情况下，3 例病例中 1 例治愈的概率是多大？二项式方程为：

$$P(r) = \frac{n!}{r! (n-r)!} p^r q^{(n-r)}$$

在此例中方程如下：

$$P(1) = \frac{3!}{1! (2!)} 0.8^1 0.2^2$$

此方程中的感叹号当然不是要求我们必须感情充沛地大声念出 n 。实际上此处的“！”应该称为“阶乘”，表示从某一给定数字到 1 之间的所有整数连续相乘；因此， $3! = 3 \times 2 \times 1 = 6$ 。依规定， $0!$ 和 $1!$ 都等于 1。简单运算一下就可得出 $P(1) = 0.096$ 。由此可以得知，即使该药物的基本治愈率为 80%，在任何三个病患中仅有一例治愈的概率也接近 10%。但建立一个相关的二项式分布函数使我们得以准确计算每一种可能结果出现的概率。掌握这些概率是在某一随机事件出现前作出定量预测的基础，同时也是判断这些随机事

件出现的可能性有多大或多小的前提。

为什么这个方程式会有用，而又为什么我们在计算蒙特卡罗轮盘赌双数连续出现 26 次的概率时不需要借助于它呢？事实上我们已经在运用二项式分布函数来计算，但并不需要借助方程式来表述这一过程。由于没有单数出现且 $(n - r) = 0$ ，所以分数部分为 $26! / (26! 0!)$ 。 $0! = 1$ ，则 $26! / 26! = 1$ 。在结果一致为双数或单数时，分数部分始终等于 1，在这种情况下运算时无须考虑分数部分了。

该二项式计算了产生某一随机结果可能出现的所有组合数。在诸如结果全为双数的一致情况下，只有一种可能组合：即赌盘每一次转动小球都必须落入双数槽。相反，倘若 26 次转盘所得结果中仅有 1 次为单数，就会有 25 种可能组合：单数可能出现在第 1 轮、第 2 轮、第 3 轮……以此类推，直至第 26 轮。在计算 1 次为单 25 次为双的轮盘赌概率时，二项式可以简化为 $26! / (25!)(1!)$ ，即 26。

同样地，当我们计算 $18/37$ 的 26 次方时， q 值的指数部分 $(n - r)$ 为 0，原因在于我们没有一例结果为单数的情况需要纳入考虑范围之内。所以计算某一连续序列的双数结果的概率是很简单的，你只需取单个偶数的结果的概率“乘以自身”，有多少个偶数就乘多少次；不过，许多人没有意识到，上述情况只是众多更为复杂的二项式分布中极其特殊、极其简单的一种。

二项式为我们提供了一个轮盘赌赌盘转动的模式，它精

确地展示出连续多次转盘中能够出现的随机结果（单数或者双数）。不过，二项式在描绘连续多次抛硬币的随机结果时就不是那么绝对精确了，因为世界上多数硬币在铸造时都很难做到绝对均衡——硬币两面不同图案所耗用的材料在质量上的细微差异会导致某一面的重量稍许大于另一面。二战时期德国占领丹麦期间一位名叫 J. E. 克里奇的英国统计学家曾被作为战犯投入监狱。狱中时光漫长难挨，他便做起抛硬币的试验，把一枚硬币抛掷 10 000 次，结果有 5 067 次是正面朝上——换言之，正面朝上的概率不是 50% 而是 50.67%。实际上，这一随机变量的期望值与“五五对开”的平分概率还是略有出入的，而且每一次抛掷的结果都是互不关联的。所以，那些抛硬币赌博的人在连续多次押不中后，若是认为手气很快会转好（不过是该来的迟来了罢了），那就大错特错了。然而，要预测新生儿中男女性别的比率，一分为二的二项式模式就远远不够了。

首先，据人口统计学家研究，新生儿男女性别比率为 104 ~ 107:100，并且这一比率无论在何种社会形态或哪个历史时期均相当普遍。据报道，这一出生性别比率有时候会高达 110:100，抑或更高，例如在中国农村的某些地区。一些观察人士相信这一上升比率反映出了当地杀害女婴的现象，这一现象的根源在于民众中严重的重男轻女思想在作祟，同时也与政府对超生家庭采取处罚措施的强力人口控制政策有关。另一些人则认为政府的生育政策虽提供了诱因，但未能

及时上报并登记新生女婴人数也导致了官方统计数据（而非现实人口）向男婴倾斜。暂且抛开中国不谈，一个生物学上的不争事实是普遍的男婴出生比为 $105/(105 + 100)$ ，即 51.2%；女婴的为 $100/(105 + 100)$ ，即 48.8%，两者之间的差别达到 2.4%。因此，倘若你总是打赌生的是男孩而对方打赌生的是女孩，那么从长远上看你的胜率大约与蒙特卡罗轮盘赌中庄家的赚头相当。

不过，男性从出生到百岁乃至超过百岁的每一年龄段上的死亡率都远高于女性，因此随着年龄一岁岁的增长，男性出生时的剩余人口比率会逐年减少，直至最后性别比率发生逆转，寡妇的数量远远超过鳏夫。男女出生率和死亡率的这种动态平衡引起统计学家约翰·彼得·苏密赫的极大关注，他于 1741 年出版了一本专著，其英文译名为《由人口统计学推导出的神的旨谕》。他在书中写到，出生人口中过剩的男婴数恰恰证明了造物主那洞悉万物的智慧，“唯有如此才能弥补由于男性天生的鲁莽冲动、超强的劳动负荷、危险的工作使命、战争、航海以及移民海外等诸多因素所造成的较高男性人口损失……同时，（造物主）也借此维持两性人口的比率均衡，使得每一个人在适婚的年龄段都能顺利地找到一个配偶。”当然即便是在苏密赫本人的时代也并非人人得以成婚，况且在婚后这种差别死亡率会消失，因而孀居的年数会低于平均值。苏密赫的书中并没有提及上帝安排中的这些方面。

以上所引苏密赫的评述来源于我的朋友兼同事阿诺什·莎娜瑞安 1988 年发表于《社会生物学》上的翻译，此后不久她不幸过早逝世，那时距她取得博士学位才不过几年时间。她的文章详尽考察了影响新生儿出生性别比率的诸多因素，其中包括母亲的荷尔蒙促性腺激素的水平，这种激素水平来自遗传，不同家族、不同种族之间各有差别。比如，她提出和欧洲人种相比，有着非洲血统的人群“由于体内促性腺激素的水平较高，其生育女孩的几率相应较高，因此新生儿男女性别比也较低”。多次百万名新生儿普查结果已经证实了这一预测。普查显示非洲裔美国人新生人口数中男女性别比在 102 到 104 之间。此外，妇女在月经周期早期或末期受孕则生男孩的比率较高，同一家庭内头几胎为男孩的可能性也较大，而且父亲年纪较轻也容易生男孩一些。所以“抛硬币”的二项式模式并不适用于男女出生比，不仅仅因为男婴女婴出生的几率并非一分为二，还由于同一母亲多次生育史之间并非毫无关联。同一母亲的荷尔蒙水平或人种特征在这次怀孕与下次怀孕时的区别显然没有随机抽取的两位女性在孕期时的大。

无论如何，在某些时候，像二项式这样的简单概率分布是无法提供对随机结果的可靠预测的。因为在某些情形中，影响事件结果的因素多种多样，而且每一种都有其未知的概率分布。那么要建立一个好的统计学模型，所需掌握的就远不止单单一个成功的概率和试验的总次数。在这些情形中，

模式化的工作只能限于一个粗糙的不甚严密的模型，所作预测也是很不完美的。但是，一个不完美的数学预测总比什么都没有强，而且它通常确有实用价值。

当统计学上的这种预测针对的是人的行为模式而不是人体生理机能或临床医学时，这种情形更为普遍。例如，从事市场营销的人深知，居住在同一邮政编码区域、有着某些相同人口统计学数据的潜在消费群体与其他群体相比，会特别倾向于购买或排斥某一种特定商品。尽管并非每一位潜在消费者都会“咬钩”，但对于他们消费倾向的正确预测足以提高一场市场营销战的收益，降低盲目投向那些不可能购买你产品的人群的花费不菲的商品目录、促销电话或各类广告的费用。同样地，你可以依据购买百分比（即消费统计概率）来预测群体的消费倾向，但无法预测某一个体的消费决策。没有哪个称职的市场营销人员会试图预测个体消费者的行为，他们所能预测的是那些大的消费群体的行为，而正是这种预测决定了他们的效益高低。因此，那些被你视为“垃圾邮件”的主动投递上门的信函在成功的市场营销战略中，都是经过精心选择后有针对性地寄给你的。

据我所知，有两种职业的确以预测特定个人的行为特征为己任：一个是法医精神病学，另一个是占星术。那么他们声称所作预测准确有效的理由有多充分呢？实际上，包括法医精神病学家在内，所有医生一直以来都被要求作出病情预测。患者不断地询问医生病情会如何发展，下一步治疗怎么

做等等；同时，医生对不同情况下病情发展趋势的预测无疑将决定采取何种治疗方案。倘若我们对某种疾病知之甚少，所采取的治疗方案又是以前未曾尝试的，那么其中不确定因素就更多，预测的准确性就更值得质疑。举例说来，对于某些目前以常规手段无法治愈的癌症病例，我们可以用一种新药进行临床试验，以检测该药是否能延长病人的生命或者改善他的生存质量。当然，在此之前我们一定有某种理由期待它会奏效，否则就不会贸然进行试验；另一方面，该药的治疗效果必定存在很大的不确定性，否则也没有必要先进行试验了。在其他情况下，预测还是很准确的。链球菌性扁桃体炎患者在用抗生素治疗后几乎百分之百会康复。不过，所有预测多多少少都会有些不确定性。即使是链球菌性扁桃体炎，偶尔也有病人就医太迟而导致严重感染，对抗生素产生抵抗性；或是病人本人对链球菌具有极高易感性从而产生免疫系统紊乱。因此，我们经常听到医生们说：“我想你会好起来的。”可这只是一种大概的预测或最好的估计。他们绝对不会说你百分之百能够康复，除非逼不得已。在生理健康和个人行为模式必须同时考虑在内的情形下，比如在精神病学领域中，预测是最难的。

预测罪行累犯

法医精神病学是同时涉及精神病学与法学的医学分支学科。我们常常请求法医精神病学家来判断释放某个在押于狱中或精神病院的人是否具有危险性。一旦犯人被释放回社会大环境，那些情节严重到行为人必须被处以刑事监禁的社会不良行为是否会重犯？法医精神病学家就是要对此作出相关预测。很自然，有关不同精神病症状的人其暴力行为发生频率的统计数据对预测至关重要。例如，与抑郁症患者相比，患有妄想型精神分裂症的病人通常对他人更具威胁性。但在法医鉴定领域，对某一病例有针对性的预测远比这样广而言之概括说法有价值，它必须考虑这个特定对象的人格的多重方面、他的病情发展史、他曾经实施的暴力行为及其发生背景。

在这些因素方面很可能个人之间存有很大的差异，而这些因素对作决定具有重要的影响。实际上决定总是充满了模棱两可的东西，以至于那些令人尊敬的法医精神病专家在法庭上就对立面进行激烈的辩论，这导致荷兰法律规定在法庭上的陈述必须简洁。对每一个精神病专家来说，总是还有一个水平相当而意见相仿的专家。相对来说决定一旦作出，就比较直接，细菌对各种抗生素的敏感度甚至可以预先确定。

决定的模糊性相对小一些，因为个人间的差异不可能对治疗的结果有重大的影响。

那么法医精神病专家对未来的行为究竟能做到何等的预测呢？两项研究可以提供一些线索。第一项研究发表在 1972 年《美国精神病学杂志》上，该研究涉及到关在 Dannemora 纽约州精神病医院的病人。其中的一个囚犯 Johnnie K. Baxstrom 提起诉讼要求释放，他说自己被监禁的公开理由是为了治疗，而他为这种未经授权的治疗而遭受监禁是违反宪法的。此外，他的权利受到侵犯，因为“将 Baxstrom 留在 Dannemora 的行政决定是在未对 Baxstrom 作任何审判前作出的，尽管作证的精神病专家意见一致作出了相反的结论，但是这一行政决定还是作出了……而且按美国最高法院 1966 年的决定，根本没有理由说 Baxstrom 不能够转到一般医院”。最终 Baxstrom 和其他 969 名病人被释放。尽管原先因预计这些人随时都有危险而将他们还押在 Dannemora 候审，但是后来所有 970 人都在普通精神病医院治疗。

转到普通医院四年之后，有一半人出院过正常生活，另一半人继续留院或后来重新住院。970 人中超过 3/4 的人在四年中从来没有袭击过医务人员或公众，其余的人也只是发生过一些轻微的违规（尽管从法医学的角度来看大多数这些违规还是较少发生的）。问题是：这种由于法律原因而出人意料的释放所提供的证据是否表明对送到 Dannemora 的 970 人的预测是完美的呢？有一半人在普通社区内成功生活，另

一半人在普通医院而不是在以犯人为主的精神病医院治疗。尽管原先的精神病专家极力关心公众的安全而认为将这些病人予以监禁完全是合理的，但是超过 3/4 的人并没有表现出危害的迹象。于是有些人根据这些结果得出结论，要对某些抓获的有暴力倾向的精神病人有足够的控制，也许只需给予在普通医院环境或社区环境下的恰当治疗（而且已经发现从美国精神病医院释放的类似人员）。一方面，似乎超过 3/4 的 Dannemora 病人被剥夺了自由，毫无必要地被监禁，因为他们被视为危险人群；另一方面，作为一个群体，由于他们比普通人更富有攻击性，所以是危险的。这种情形对公共政策所呈现的困境在某种程度上来说是由于只预测由某一群体，而不是由某一个人的统计特性造成的。

1972 年 Kozol 和同事发表了另一项研究，该研究预测医生对危害性的判断力。这一次研究的结果是关于那些留院治疗，后来被释放的性犯罪者的重新犯罪率。一组人员是由法庭根据医生的判断而作出的，重新犯罪率为 6.1%；另一组是法庭不听医生的判断释放的，累犯率为 34.7%。因此，似乎医生比司法机构对危害性的判断要好得多，但是他们的判断也不是完美的——他们也有 6% 的犯错率。虽然这个百分比不大，但在预测上则是个严重的错误，因为 6% 的这些人仍然在攻击他人。使用医生的预测来作出释放的决定对社会来说是否恰当？撇开惩罚的争端问题，所有的性侵犯者是否都应该永远被监禁？拘禁 94% 的人是为了预防 6% 的人重

犯。这样公正吗？或者是否可以接受这样一种方法，既可以
让 94% 的人获得自由，也可以让那些继续重犯自己可耻罪行
的人获得自由呢？

众所周知，某些裁定的暴力罪犯，特别是性罪犯也许会
由司法系统释放，回到社会，而这些人员重新犯罪的可能性
是很高的。这种意识引发了一场政治运动，这场运动使那些
要求一旦被裁定有罪但没有被监禁的性罪犯必须公开列入档
案的法律条文得以通过。这场运动之所以获得强力支持，完
全是由新泽西州的 7 岁女孩 Megan Kanka 被残忍地奸杀这一
暴行所引起的。她的父母还是第一次听说因这次致命的袭击
而被捕的邻居是一个两次被判犯有性犯罪的家伙。随后各个
州被要求在 20 世纪 90 年代中通过性罪犯登记法案，以符合
美国联邦指导原则。在大多数地方，在作为立法催生剂的这
个事件之后，这些新的规定被非正式地称作 Megan 法令。

被释放的犯人重新犯罪的风险程度因人而异，而 Megan
法令规定风险的程度决定公布信息的量和公告的形式。当一
个罪犯被释放时，作为一个必要的程序之一部分，就是要向
法庭提交风险程度的建议。这个建议是根据“性罪犯登记法
案风险评估表”而作出的。这份评估表或问卷表将风险依据
分数制度归类为三级。如果一个罪犯定为三级，包括罪犯的
姓名、年龄、地址、犯罪史和相片的信息也许会被张贴，甚
至被发布到互联网上。例如透过衣服的不必要接触为 5 分，
性交为 25 分；前次犯罪不到 3 年，加 10 分；受害者为两

人，加 20 分；受害者为三人或更多，则加 30 分。有四种情况自动归类为三级：先前犯有严重的性犯罪、受害者受重伤或死亡、罪犯采取进一步的犯罪威胁或者临床判断这个人由于特定的精神不正常而缺乏自控力。这是一个维护公众利益的政策例子，这一政策是由于公众对精神病专家对罪犯不尽如人意的预测判断和犯人过早释放的担心而推动的。但是自相矛盾的是，那些鼓吹 Megan 法令的人也不得不承认统计学上的预测还是有一定价值的，因为折中的解决办法在某种程度上来说还是需要根据风险的统计分类而作出公告。

个人之间总是存在无法预测的差异性，即使是几乎完全相同的犯人和精神病病例。起作用的因素总是不同的，由于一系列的原因，总是以犯罪而结束。精神病学专家在对个人的危险性的评估方面也许总是存在差异，不可能有一个完美无缺的系统评估表来让我们释放所有“危险性得分”低于临界值的犯人。因此这类问题不可能有一个永久、客观正确而且意见一致的解决办法，因为有关群体预测的统计结果必然会对法医学的判断产生影响。

我们的失误与星球有关吗

占星术是另一个试图利用人类行为预测的领域，就像法医精神病学一样，应该受制于相同的统计证据标准。从古时

候起，人们就注意到占星术在预测特定而明确的人生大事方面的数值很低。典型的例子是生活在公元前 106 年至公元前 43 年的古罗马元老院议员西塞罗的批评。在他的书《定命》中（我用的是 Falconer 的翻译，著者注），西塞罗特别提到柏拉图的学生 Eudoxus，顶级学者无疑都会将他认为是占星术的鼻祖。在 Eudoxus 的著作中留下了以下观点：“当卡尔迪亚（古巴比伦）占星术家声称能从一个人出生那天的星座位置预测其未来时，千万不要相信。”西塞罗对 Eudoxus（生活在公元前 400 年的希腊人）的高度评价随着时间的过去已经被证实。Eudoxus 不仅证明了几何学的主要定理，而且绘制了第一张希腊星球天体图，是最早运用同心球数学原理的几何模式来计算行星的运行的人。这是一种重要的计算工具，也是一种观念上的进步。

西塞罗阐明了怀疑占星术的许多理由。这种原理不可能是真实的，因为正如他所说的，在同一时间对地球上的所有观察者来说，这些星球不可能在同一个位置上，而是根据他们所处位置而发生变化。他也注意到孩子们的相貌、姿势、习惯和生活方式像其父母。这些显著的特征和那些“区分印度人和波斯人，埃塞俄比亚人和叙利亚人”的身体上和心灵上的差异“更多的是受到当地的环境影响，而不是月球位置的影响”。对他来说，比这种普通推理更重要的是占星术缺乏预测力：对他来说，“在同一时刻出生的人在性格、事业和命运方面各不相同这一事实很清楚地表明出生时间与确定

人生轨迹毫不相干”。他强烈地质问：“在 Cannae 战场上所有牺牲的罗马人诞生时的星位都一样吗？但是结局却只有一个。” Cannae 之战对罗马军队来说真的是一场毁灭性的失败。在公元前 216 年，率领迦太基部队大约 50 000 士兵的汉尼拔，击败了规模更大的罗马军队，8 万罗马士兵有 6 万多被杀死。西塞罗认为，如果诞生时的星位不能够用来预测 Cannae 战场上的死亡，就不能说它具有预测价值。

不言而喻，这是统计学的观点。罗马士兵在 Cannae 战场上的死亡概率不取决于他何时出生（概率对所有出生日期和时间都是一样的）；相反，在死于那场战斗中的人中，没有看到一个特别的星座有什么优势（概率并没有提高）。近期反对占星术的证据显然来自统计学，相关的信息均有所记载和保存。1985 年发表在英国《自然》杂志上的一项研究阐明了这一观点，并说明统计学证据反对占星术的优势。

与西塞罗不一样，进行这项当代研究的研究小组为了避免检验“科学家对占星术的看法，而不是检验在有名望的占星术界普遍使用的占星术”，咨询了该领域地位高的职业星相学家。这些人都是从国家地宇研究机构提供的名单中挑选出来的，该机构是得到全球星相学家尊重的一个组织。28 名星相学家同意参加，并提供表明从普通公众中选拔的 100 多名受过高等教育的自愿被试的出生时天体位置的生辰图表，以及随之产生的性格描述。尽管有西塞罗的批评，但是星相学家们确实知道在相同时间点上的星球在不同位置是各不相

同的；因此，按要求向他们提供了准确的地点以及出生日期和时间。这一信息只要是以出生证、医院记录或者是与出生同时的其他资料，都包括在研究范围之内。

该研究有两部分，首先，每一份出生记录都会有一个生辰图表和一份由星相学家描述的被试应具备的个人性格特征的书面解释。其次，所有名字均从这些解释性文档中去掉，然后给每个被试提供三份性格描述：自己的一份和两份随机挑选的其他人的资料。每个被试的任务是选出对自己的描述。就概率来说，我们预料他们有 $1/3$ 的概率可以选对。这是科学家的展望，而星相学家觉得他们的工作会得到 50% 以上的确认。

这些配对的预测率代表了估算，当然没有人为了证明自己的观点会说配对率必须是刚好 $1/3$ 或者 $1/2$ 。抽样的波动要考虑，这是随机概率所产生的一种现象。例如虽然投掷硬币最终正反面的几率为 50:50，投掷 10 次，也许刚好获得 5 次正面、5 次反面，但是发现一面为 6 次，而另一面为 4 次也毫不奇怪。事实上 10 次一组，每组总是刚好 5 次正面、5 次反面倒是令人吃惊的。问题是在单个的试验中，人们的期待围绕某一特定的数字，实际结果越接近那个数字，研究结果越支持挑选这个数字的观点。

在检验占星术的成果时，85 个个案中的被试和星相学家的资料都是可以利用的。有多少被试辨认出根据自己的生辰信息而得出的性格特征描述呢？83 个被试中有 28 人配对选

择正确，或者说是大约 1/3（实际上是 33.7%）。因此，除了概率吻合之外没有什么特别的东西。也许占星术没有从生辰信息中提供可供辨认的描述，但是，对占星术成就的这种衡量是弱勢的。星相学家们，或者甚至科学家们也许会争辩说这个结果只能说明人们不擅长挑选对自己的描述。

该研究的第二部分涉及的是对不同结果的检验。每个被试要填“加利福尼亚心理学调查表”（CPI），这是一种从大约 1958 年建立起来的 400 多个有关各种不同的喜好、情形和性格特征的题库。所有的问题都以陈述的形式出现，要求判定每个陈述是否真实。H. G. 高夫在 1994 年的《心理学报告》一书中公布了一些典型问题的例子。这些问题包括：

- “如果薪水合适的话，我愿意跟随马戏团周游或参加狂欢节。”
- “我从来不与陌生人打扑克。”
- “在做某事之前，我会考虑朋友对此事的反应。”

“加利福尼亚心理学调查表”在一些二级指标中有许多平分等级，这些二级指标为支配 / 被动、自控力、容忍度和灵活性等。这一测试不直接问个人的性格表现，结果通过与心理学家有关性格特征的判断比较，由外部因素确定。因此这个检验受到心理学家的尊重；而且这是从所获得的性格特征中特别挑选出来的，因为《自然》杂志研究项目组所招募

的星相学家都一致认为所评估的性格特征最接近“占星术所能辨认的属性”。

在该研究第一部分的镜像里，给星相学家提供了含有一个出生信息表和三个加利福尼亚心理学调查表数据的内容，其中一个数据来自提供了出生信息表的人，另外两个是随机挑选的。星相学家的任务是拿着出生信息来决定这三个加利福尼亚心理学调查表数据的内容与同一个人相匹配。科学家的预言是星相学家匹配成功率与概率所预期的 $1/3$ 一致。而星相学家觉得如果被试匹配的成功率达到 50%，就证明占星术是有用的。总共有 116 张生辰表要与加利福尼亚心理学调查表匹配。因此，科学家估计有 39 例会匹配成功，而星相学家预期会配对 58 例或更多。最终选对加利福尼亚心理学调查表的数字是 40。此外，星相学家也被要求在一个 1~10 的尺度范围来评定自己对每一个配对有多大信心，所有答案都倾向在 8 左右，但是配对与配错的比例并无区别。

偶尔在社交场合碰到的人会问我占星术的“星座”。几年前，我常常告诉他们，这是一种伴有一闪而过“领悟”的启示，同时还加上一句这样的评论，“当然，这就是为什么有这样或那样的性格特征的原因”。我总是想，如果真的如此显而易见，他们非要这么问就有点奇怪了。为什么不直接走到我跟前说，“你肯定是狮子座”或“你是双鱼座，对吗？”这样我就不必再回答星座问题，而是更具有挑战性，因为我非常好奇想看到结果。现在再有人问我的星座时，我

通常回答道：“你告诉我。”我一直关注着各种答案。迄今为止，所有的星座都提到过，而且数目相当。这就是结果的平均分布，因为在条形统计图表上，代表星座的条形大致是一样高，没有一个星座会占优势。我经常想更进一步地给一个错误的星座，看看我的这种欺骗行为是否会察觉。但到目前为止，我一直抵御着这种诱惑。

我自己的星座预测经验以及星相学家们不能够从出生记录预测加利福尼亚心理学调查表的结果，意味着某些有关个人的性格特征的东西分布于占星术各个范畴。不管这些范畴是广的，比如星座范畴的选择，还是窄的，比如各种可能存在的出生信息结构，占星术各个范畴内的个人性格特征或多或少是平均分布的。如果你用条形图来表示带有某一特定性格特征人数的比例，用一个单独的条形来表示每一种占星术范畴，每个范畴的条形高度相当。同样，如果西塞罗制作一个表示在 Danae 战斗中死亡率的条形图，条形的高度也会是相等的，就像比较星座或者占星结构一样。当在不同范畴间有数据呈平均分布态势时，某特定范畴内个人的成员资格对有关那个人特定结果的可能性不会提供有用的信息。

另一方面，有些数据的分布对特定结果的可能性提供了许多的信息。我们已经熟悉二项式分布及其对概率估算的作用，但是许多情形涉及连续数据。例如，血糖水平有连续分布，但是在糖尿病患者身上这种分布转向高端。眼球内的压力升高，表明患有像青光眼这类变性疾病，而且眼内的压力

越高，病情越严重。临床医学有许多这样的测量，社会科学里也有许多，它们都遵循一种称为正态分布的“铃型”曲线。当测量真的遵循正态分布，组群内的各种测量结果的概率计算就容易多了，而且预测结果也会改善。正态分布的应用是下一章节的主题。

2

你肯定有问题



A great performance from the Yankee's Derek Jeter helped crush the Baltimore Orioles at the stadium yesterday. ①

当我把上面这句话敲出来时，我的文字处理器自动查出语法和拼写错误并在 Jeter 这个词下面画了横线。这位球员的姓本来拼写是正确的却被标记为拼写错误，而另一个输入错误却没有标出来：单词 form 实际上应该是 from。任何可以把事物分成两类的系统都能够出现两种错误，一种是“假对”，另一种是“假错”。被标上拼写错误的 Jeter 一词就是“假对”——他的名字是正确的，本不应该被标成是有问题的。而本应该被标记为错误的 form 一词却没能标出就是“假错”的一个例子。由于这种因文字处理程序无法识别而造成的错误还很普遍，所以现在编辑校对还是必不可少的，

① 此句话的意思是：昨天在运动场上来自北佬队的 Derek Jeter 的一记妙招把 Baltimore Orioles 打得落花流水。但由于输入错误，把 from 打成了 form。（译者注）

因为人工可以识别 Jeter 输入是正确的而 form 是错误的，因而至少目前来说人工比计算机要可靠得多。

由于出现“假对”和“假错”现象是不可避免的，其后果不可忽视。比如，它们会严重影响疾病的正确诊断和治疗。验血就是一个很好的例子。以前，找医生看病就意味着在患者身上放些水蛭或是放血以除去病人体内的杂质。现在看病则是先抽一试管血然后拿到化验室去化验。几小时或几天之后，结果出来了，也许有些数值表明异常，就是说这些值超出了正常范围。你的医生就会说：“这个地方有问题，你血液中的胆固醇和钙偏高。”血清中这两样成分的增高分别表明你患心脏病和代谢紊乱的危险加大。然而化验得出的异常结果就一定说明你的身体有问题吗？难道化验结果就不会有“假对”或“假错”的现象吗？如果有，这些错误的普遍性又是怎样的呢？

首先我们需要知道所谓“正常”在这种情况下意味着什么，因此我们得知道医学中“正常”的范围是什么，而一个人在什么情况下属于“异常”。现在让我们看这样一个典型的过程：给一大群健康的人抽血，并且像胆固醇或钙这样的变量的数值分布也已确立（在第一章里我们已经将“分布”定义为一系列的随机结果及每一结果的概率的组合）。有了化验值的分布，下一步就是决定极限值。通常分布在中间的 95% 的值都不重要，而只有 5% 处于极端的值用于确定正常范围的界限（即观察结果中有 2.5% 为最大值，2.5% 为最小

值)。

尽管可能会有一些实际上正常的人的化验指标远离中间值甚至达到极限值而需要复查，因为他们在其他典型分布上具有更高的概率，这种人群往往意味着患有疾病，也就是说他们不像典型的正常人。

这样一种普通的程序并非基于对疾病过程本身的周详了解。科学家并不能确定某个正常人的化验值的分布应该是多少或是基于病理学来判断一个人是否患有疾病。他们的判断是基于所观察的典型的人口指标。显然，采用这种方法就会有 5% 的绝对正常的人被判定为达到了“极限”指标。更糟的是，复查次数越多，就越可能得出错误的判断，使得一个根本没有病的人的化验结果却是“异常”的。在体检中的“假对”往往都是因为这种现象造成的，尤其是在对某个病人进行一系列体检时更容易出现“假对”现象。

我们可以使用双称名分布法来量化“假对”出现的几率。假设对一个健康的人进行了 18 项指标的体检，那么他的身体指标在任何体检中都会有 95% 的可能判定为正常。如果我们假设这些指标测试中的上下浮动与一个人的身体状况无关，那么这 18 项指标中判定为正常的几率应为 0.95^{18} ，各个指标的总概率应当为 1.0，异常指标的总概率则为 $1 - (0.95^{18})$ 。这样，至少一个实验室得出“异常”的化验结果的几率就将会高达 60% 以上。所以，如果你去看病并做了许多项体检，就很有可能有些项目的结论是值得怀疑的，

需要复查。

这无疑需要重新体检，事实上倘若每一项指标都是单独检查的，只要没有其他意外，出现“假对”或“假错”的机会是很少的。另外，复查往往能更好地判定出有病还是没病。当然这种体检费用往往会很高，所以一般人不会连续做很多项指标的检查。比如，在检查引起艾滋病病毒 HIV 指标时，一般先做酶联免疫吸附剂化验（ELISA），复查时才做蛋白质印迹和病毒培养检查，因为前者比后者要便宜一些，不过效果要差些。

检验结果

那么我们凭什么说一项检查比另一项检查的效果差呢？一项诊断检查的质量要从这样几个方面来评价。一是灵敏性，即该检查对疾病判断的正确率。换句话说就是，判断患者真正有病其检查结果呈阳性的概率。与灵敏性相伴的另一测量是针对性，即判断患者真正无病其检查结果呈阴性的概率。灵敏性和针对性的计算需要使用一组已知的“真阳性”和“真阴性”，而判定“真阳性”和“真阴性”则是通过另外一个假定为准确的测量或测试得到。这个另外的测试被称为“黄金标准”。例如，检查 ELISA 指标以判定是否携带艾滋病病毒 HIV 时，就可以看艾滋病的临床检查结果是否与制

造商的检验标准一致。

不过，作为病人，他们不会对实验测试的敏感性和针对性感兴趣。他们只是想从另外的角度了解检查结果的可靠性。当医生说检查结果出来为“阳性”或“阴性”时，病人想知道的只是该结果是否是真的。在这里有两项相关的测量：第一项是“呈阳性”（即实际上患病的概率）的预测值，第二项是“呈阴性”（即实际上没病的概率）的预测值。

预测值如何确定呢？敏感性和针对性基本上是由测试项目的物理与化学性质决定的，而这些性质的估算由制造商在出售产品前提供，因而检验标准应当尽量接近“黄金标准”。然而，由于预测值并不仅仅取决于敏感性和针对性，所以无法预先得知预测值。预测值还取决于被检查的疾病的普遍性——也就是说，还决定于实施该项检查的人群特点的重要程度。

也许这听起来很奇怪，不过先看看下面这个例子。我们检查一个疾病，其敏感性为 0.95，针对性为 0.95。根据已知“阳性”或“阴性”的范围，该检查可以说对此疾病正确性判断的几率为 95%。假定该疾病在对 1 000 人的检查中的真普遍性为 40%。如此之高的普遍性也许很令人惊讶，但是要想知道进行某个疾病检查的人们是疾病临床疑似患者所表现出的病征或症状的子集。而且即使在无症状的人群中疾病的普遍率也可能确实有那么高，例如，据估计在新泽西州的纽瓦克用静脉吸毒的人群中，HIV 感染率高达 50% 以上。

表 2-1 在高疾病普遍性环境下的疾病状况和检查结果

		真实的病情		
		阳性	阴性	总数
检验结果	阳性	380	30	410
	阴性	20	570	590
	总数	400	600	1 000

因此被检查的人群可分为表 2-1 中的几类。在 1 000 人中有 400 人患病，600 人没病。而这 400 有病人群中，有 95%（或者说有 380 人）的人被正确地诊断为呈“阳性”，600 没病人群中有 95%（或者说有 570 人）被正确地诊断为呈“阴性”。但是请注意，400 个真正有病的人中会有 20 人会被误诊为“阴性”，而 600 个没病的人中则有 30 人会被误诊为“阳性”。这些人误诊为“阳性”或“阴性”的人他们自己包括他们的医生都不会意识到是被误诊，因为他们目前除了依靠实验室的检查结果外，也没有其他的检查方法了。从病人的角度来说，在这 410 个被医生诊断为“阳性”的人中，380 人的诊断结果是对的，即呈“阳性”的预测值为 $380/410$ 或 0.927。也就是说，当病人的检查结果出来是呈“阳性”的，有 92.7% 的几率是对的，说明他患有疾病。对呈“阴性”的检查结果来说，590 人中有 570 人的结果是对的，预测值为 0.966，或者说有 96.6% 的正确率。

不过再看看表 2-2，就会发现这还取决于普遍性这一因素。假设敏感性和针对性是一样的，都是 0.95 我们把普遍

性从 40% 改为 2% ,即 1 000 人中有 20 人有病 , 980 人没病。这 20 人中间会有 19 人被正确地诊断为呈“ 阳性 ” , 而 980 人中有 931 人会被正确地诊断为“ 阴性 ”。如表 2 - 2 中所示 , 误诊为“ 阳性 ” (49) 和误诊为“ 阴性 ” (1) 的人数可用减法计算出来。

表 2 - 2 在低疾病普遍性环境下的疾病状况和检查结果

		真实的病情		
		阳性	阴性	总数
检验结果	阳性	19	49	68
	阴性	1	931	932
总数		20	980	1 000

那么“ 阳性 ”的预测值为 $19/68$ 或 0.279 , “ 阴性 ”的预测值为 $931/932$ 或 0.999 。这样 , 呈“ 阴性 ”的检查结果几乎都是对的 , 而呈“ 阳性 ”的检查结果的正确率还不到 28%。那又是怎么回事呢 ? 在前面我们看到的检查结果中 , 呈“ 阳性 ”的正确率是相当高的。敏感性仍然不变为 95% , 但是疾病的普遍性大大降低意味着即便是在所有的疾病病历都被测定为是“ 阳性 ”的情况下 , 诊断出呈“ 阳性 ”的正确率也会大大降低。而且 , 对更多无病人群而言 , 在真正呈“ 阳性 ”的人数很少的情况下 , 出现“ 假阳性 ”的比率会更大 , 即检查结果会错误地诊断出更多呈“ 阳性 ”的人。

在这方面具有判断启示 : 医生 (和病人) 在解释检验科的化验结果时必须使用判断力。具有讽刺意味的是 , 对疾病

预测概率的评估尽管可能是主观的，但是在确定究竟给予看似客观的检验报告多少信赖度方面，必须发挥作用。这就意味着医生应该非常清楚地了解病人取自何种人群（有关风险因素等）。在类似该病人的人群中其流行性越低，“实验室呈阳性”的正确性就越低，更有可能是一种假阳性。这还具有决策性启示：用于鉴别某一疾病的人群应该是处于患此疾病的高危群体，以免花费时间和资金仅仅为了获得少量的真正阳性（而大量的假阳性只有到情况弄清楚之后才不必进一步处理，不再需要其他花费，不必再担忧）。但是，当美国被因意识到起初艾滋病的流行而引起的焦虑所困时，有人建议在美国的每个人都应该做艾滋病检查。如果 1% 的美国人口感染艾滋病，剩余 99% 的人口甚至在很小的误差率的情况下也会有大量的假阳性，那么在全国范围内的检验计划中检测出的大多数阳性很可能都是假阳性（依据准确的敏感度和特征）。这是不能作为检验计划范例的。

如果流行上涨而不是下降，则趋势倒转：阳性的预测值提高，则阴性的预测值就成问题了。真阴性的数量最终会很小，即使通过检验而正确鉴定为阴性，在绝大多数的阳性结果中任何小误差率将会导致假阴性在所有的阴性结果中占多数。用一个极端的例子来说，在艾滋病收容所里，每一个阴性检测有可能都是假阴性检测，阴性的预测值为零。

当然医疗界的许多测量方法并非仅仅是阴性或阳性，即使有关病人的病情是以这种二元方式记载的。医生说病人是

否患有高血压、青光眼等，但是实际上大自然提供的往往是各种参数值的连续统一体，而诊断所使用的是武断的截止点。在某一特定点上，治疗高血压是重要的；在另一特定点上，眼内的压力有必要治疗以防止青光眼所造成的损害。了解潜在的数值分布是作出医疗决定和确定什么时候某种药物在转换某一参数的标准值方面起了作用的基础。其中最普通、最熟悉的潜在分布称为常态曲线。

天体分布与街头小面包房

常态曲线按其来源有时也指高斯曲线或错偏差曲线。卡尔·弗里德里希·高斯是最先广泛使用分布的人之一，它可以用来计算各种结果的概率。1777年出生在德国的神童高斯是一名杰出的数学家和天文学家。他计算天体的距离、大小和位置，并使用一种新式望远镜企图得出更精确的月球直径。在计算过程中，他注意到了一些奇怪的现象。由于偏差，他反复测量的结果总是不同，但是偏差以一致的方式出现。大多数的测量结果与平均数稍有不同，偏差越大，出现的可能性就越小。高斯确定偏差概率是根据常态曲线分布的，常态曲线形状见图 2-1。水平轴(x)代表测量结果，垂轴代表的是这些测量结果的相对频率。

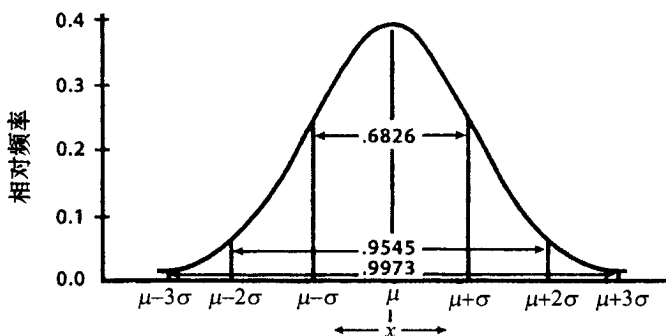


图 2 - 1 正态分布曲线下的区域

常态曲线是单峰的（一个驼峰）、连续的（任何 x 值都有可能，不仅仅是某些断断续续的数值），可以向两边延伸而不接触 x 轴。后一种特性意味着曲线下的区域（AUC），不论何种测量，不管有多么极端，都不可能为 0。当然 AUC 在 x 的极端值是很小的。超过 99% 的 AUC（实际上均大于 99.7%）包括在 x 值的范围之内， x 值的范围从低于平均值的 3σ 到高于平均值的 3σ 。

希腊字母 σ 代表的是统计值，称为标准差，也叫 SD。它是通过从每个 x 值中减去平均值（ μ , mu ）所得。这种方法给出一组偏离平均值的数据，这样我们就可以检验这些偏差，看看到底这些偏差究竟是大还是小。毕竟有些测量是易变的，而另一些测量则是可再现的。为了确定变化量，有一个能够反映这些偏差的数据是非常有用的，但是有些偏差有正号，有些则有负号，这些偏差将会相互抵消。（平均数刚

好在所有观测数据的中间，曲线是对称的，因此很容易看出实际上这个总数之和总是 0。) 为了避免这个问题，偏差值在求和之前先将这些偏差变成平方（求和的过程由希腊字母 Σ 代替）然后除以 N （所观察的总人数）再得平均数。由于在得平均数之前这些偏差值已经变成平方，所以现在采用平方根以恢复先前的数值。这个结果就是标准差：

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

高斯计算出在正态分布中，68% 所观察的结果落在平均数 1 个标准差上和 1 个标准差下之间。大约 95% 的结果落在 $\mu + 2\sigma$ 范围之内。如前面所提及的，平均数任何一边的 3 个标准差几乎都在曲线下区域（AUC）。高斯发现更为普遍的是，如果有平均数和正态分布的平均差，那么就有可能计算出在 x 轴的任何一点看到某个观察结果的概率。

不仅天体测算遵循正态分布规律，涉及检验生产过程结果的质量控制专家在分析飞机部件的宽度和装 1 升容量的瓶子中酒的重量的平均数和可变数时，也通常使用正态分布。部件与部件之间、瓶子与瓶子之间总有些差异，所以通过检查分布可以看出这些变化是否在可接受的公差极限范围之内。

早在高斯量化他所发现的系统离差之前，人们就已了解到在测量和生产过程中的差异是可以预测到的。通过常量与标准量度的比较来尽量控制差异。历史上，君主们都有各种

不同币值的标准金币，他们希望皇家造币厂所生产的金币在重量上与之相匹配。他们知道完全一样的匹配是罕见的，所以他们公布了允许极限——超过变化极限会导致欺诈的指控。另一方面，有时候在建立测量单位本身时，需要考虑这些变化。许多单位起初是用来界定距离的，这些距离近似于某些人体部位之间的长度。例如，“码”指的是从一个成年人的鼻尖到他伸直的手中最长的手指尖的距离。在埃及和圣经时代，基于前臂的腕尺（注：古时的长度单位）用于建筑业。英尺（foot）作为单位是另一个明显的例子。这些单位非常有用，因为它们总是随时可用，不需要特别的器材，但是很显然这些单位差异很大，因此无助于商业目的，通常会导意见不一。

在启蒙运动时期，在巴黎，技术和政府方面都涉及到关键问题，如 1 米的起始点应刻在金属条的什么位置，并用来作为全法国的标准？但是即使在那之前，已有人看出可以使用个人间的差异平均数来作为标准单位。在 1584 年，J. Koelbel 提出了一种计算路德（长度和面积单位，现已过时不用）及线性步幅的方式，他的方法发表在其名为《几何学》的书上。这本书是为测量员写的，而不是为数学家写的，书中将几何学定义为“测量地球的科学”，而不是像欧几里得几何学那样为有关空间数据的定理和验证。Koelbel 提出选择 16 个刚从教堂走出来的人，让他们排成一列，左脚一一紧挨着，十六只脚的排列长度就是“正确的法定路

德”这个长度的 $1/16$ 就是“正确的法定的脚长（即 1 英尺）”这里我们可以看出，为了得到更为真实的值，将随机抽取的个人脚的长度差异平均是很有必要的。真值就是所观察到的平均数，就像正态分布一样。

高斯曾经论证过人们可以用曲线来表示所观察的结果的预期变化，并说明这种偏差的可能性有多大，发现常态曲线有许多用处，有些对人类作用不大，而有些则对人类大有益处。法国伟大的数学家 Jules Henri Poincaré（1854—1912）用正态分布证明了他被当地面包房欺骗的情况。他习惯每天买一条重量为 1 千克的面包，Poincaré 知道如果面包平均重量为 1 千克，面包一次的分量不够不能证明是有意欺骗，因为下一个面包有可能会比平均重量重一点点。因此，他将一年当中所买的面包一一过秤，发现了一个平均重量为 0.95 千克的正态分布。平均起来他被骗了 5%，于是他向有关部门投诉，面包商受到警告。第二年的数据使 Poincaré 又投诉，他说这个商人继续骗他。数据又是单峰的，峰频为 0.95 千克，但是分布不是对称的，当分布区域右边与以前相同时，峰左边的曲线有所削弱，在高度和长度上均比上次短。怎么解释呢？面包师傅只在一个方面作了变化：意识到 Poincaré 投诉面包分量不够，面包师傅于是总卖给他最大的面包。当警察造访面包店时，面包师傅对 Poincaré 居然还意识到投诉有效感到很惊讶。

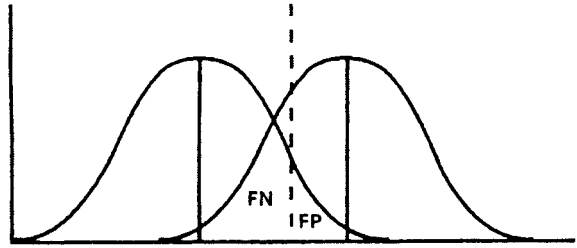


图 2-2 两条正态曲线重叠时，假阴性（FN）和假阳性（FP）的频率

正态曲线可以帮助我们理解为什么使用实验室检验单的医生仍然很难确定某个特定病人是否患有某种疾病，即使手头上有精确的连续数据。图 2-2 是该问题的示意图。左边是一条正态曲线，代表那些没有患此病的人的门诊测量分布。例如，这可能是不患青光眼病人的眼内压力分布。另一条曲线是那些确实患有此病的人的测量结果分布。这两条曲线重叠——实际上这两个组的人分离显然是不完整的。

图 2-2 中相交于这两条正态曲线的垂直虚线可以看做一条标准线：如果你去看眼科医生，而你的眼内压力高于这一点，你将被认为患有青光眼。你患有此病的可能性是很大的，所以从这个意义上来说这是一种好的检验方式。但是有一些人尽管检测结果与标准线右边相符，但不会得此病，他们所占的区域标为 FP，表示假阳性；也有一些人，患有此病，单检测结果落在垂线的左边，标为 FN，表示假阴性。在图 2-2 中，敏感性也有所表现，即在标准线右边得此病的人群中该区域的比例，特异性是标准线左边不得此病人群

的检测结果的比例。

也许有人建议将标准线往左移动，以包含那些检测结果为阳性的人中更多得此病的人。这样的调整确实可以通过对更多的患病区域内的检测结果的正确分类、增加敏感性，但是这样做会牺牲特异性——少量的病症阴性可以被分类出来，而假阳性则会增加。将这条标准线往右移，则会出现相反的结果：假阴性会降低，特异性会提高。但是在阳性区域右边重叠区得此病的人更少，因此敏感性会下降。在这种情况下，不可能选择一个标准值准确地区分哪些人会得此病，哪些人不会得此病。

通常没有依据医学检验结果来区分病人的单个、客观的最佳标准。在实践中，对结果分类的标准取决于某个不正确决定的结果。有时候进一步检验以及可能出现的错误治疗的代价和副作用超过某种小毛病的危害。而其他时候，及时治疗某种疾病又是非常重要的，实验室检验结果中的假阳性可以接受，以后它们可以被排除，不被认为是真正的病例。

“正常智力

许多基本的社会重要事件，而不是特殊的医学重要事件，也是正态分布的，这里我们也发现不同年龄组的人的重叠。例如完成各种各样的智力任务的能力呈正态分布：偶尔

某个人做某事也会特别聪明，或者特别无能，但是能力水平的极端例子要远远少于更为普通的能力。那些参加智力测验的人最后得到一个将他们置于一个能力统一体中的数字分数，这些结果一般也是正态分布，集中在平均分 100 左右，标准差为 15。因此 68% 的人的智商在 85 ~ 115 之间，2.5% 的“最聪明的人”智商高于 130 ($\mu + 2\sigma$)，而 2.5% 的“最愚笨的人”的智商低于 70 ($\mu - 2\sigma$)。

但是 IQ 即“智商”代表了什么？智力测验是测验一个人从事某些特定任务的能力，比如完成语言类比、在一个序列中挑选下一个适当的数字或者找出合成几何图形的平均数以生成另一个图形。很有趣的是，在构思一个智力测验题时，选择哪些类型的抽象推理题并不重要，个人在这些不同的任务方面的得分是非常接近的。这些问题的总得分为 g ，即普通智力。

在某种程度上，各个问题分数间的密切关系使我们相信智商作为测量某种以不同形式呈现的先天能力是可靠的。但是从另一方面来讲，这种一致性也说明智商测验只能测量某种特定的推理能力。例如，估计朋友的情感状态的能力、选择恰当的言辞或行动来安慰某人的能力，或者影响一群人、煽动政治集会的能力等，所有这些能力均因人而异。它们确实是各种智力形式的反映，并不包括在“普通”智力指标之内。

然而许多学业问题来源于智商测验分数上的这类能力，

这些分数可以很好地预测学生在校学习的成功与否，不管其学习结果是以学业成绩还是以最终接受教育的年限来衡量。智商对这种结果的预测要强于社会地位。兄弟姐妹的学习成绩有这样一种趋势：他们一起长大，但是由于智商不同，所以学习成绩也不同。相反，双胞胎分开来抚养（例如，由于过继领养），尽管环境不同，其智商和学习成绩则比概率所预测的更相似。然而现在不清楚在比这些狭义的教育测量更普遍的成功方面，智商的预测值是什么；因为很显然迄今为止还没有一致公认的、能够反映人生成就的全面测量办法。

智商的最大争议之处在于非洲裔美国人和祖籍欧洲的美国白人的智商数值正态曲线的差异。根据美国心理学会的《心理学百科全书》（2000版），前一组的平均值大约是85，后一组大约是100。如果两组的标准差是15，白人的平均值要比非洲裔美国人标准值高1个标准差。因此，只有16%的非洲裔美国人的智商超过100，而白人则有50%的人的智商超过100。

但是我们还不能从这种差异的存在就推断出科学的结论，认为这是由两组特定的基因特征造成的。考虑一下这种情况：在相同的环境下长大的人，其智商上的差异显然源于基因。再考虑一下相反的情况：正如心理学家彼德·格雷所说：“如果你是在普通的中产阶级家庭长大，而我则是在储藏室里长大，那么我们之间的智商差异多半是由于我们的环境差异造成的。”

当然，非洲裔美国人和欧洲裔美国人属于基因结构不同的群体，而且他们都是社会群体，受制于不同的社会环境。在美国历史上，对非洲裔美国人的偏见和歧视使人们普遍认为他们智商低、学习成绩差。研究表明，被强势种族人群视为低人一等的弱势少数民族在所有存在等级的国家，其智商都要比强势民族低 10 ~ 15 分。例如，日本人身上某种特定的不可区别的基因子集在社会上被认为是低人一等的。在日本，那些部落民到 1871 年才获得公民地位，从那个时候开始，日本一位开明的天皇颁布了一个法令，禁止公开嘲笑部落民。但是即使在今天，这个种姓的群体始终处于社会底层的不利境地，绝大多数处于仆人地位；与部落民做朋友会招致大多数日本人的鄙视。和那些在美国具有非洲血统的人一样，日本的部落民在智力测验方面的得分显然要比大多数人低得多。而在美国的日本移民中，部落民与其他日本人在智力测验与学校考试中的得分是一样的。一旦贱民的地位和无助感得以消除，这种差异也就消失了。在美国，几乎没有人知道某个日本人是部落民，也没有人在意此事。一旦得到平等的社会机会，部落民的智力测验的常态曲线呈上升的变化，并与普通日本人的常态曲线相吻合。

爆米花与抽样平均分布

当你在微波炉中爆玉米花时，正态分布就可听得见。开始一会儿，什么也没发生，但随后你偶尔听到零星的爆裂声，接着一次可以听到几声，爆裂声间隔时间越来越短，直到大量的玉米同时爆裂，中间没有间隔。震耳欲聋的爆裂声数量不断增长，然后逐渐减少直至你一次只听到一两声爆裂声。最后爆裂声逐渐消失，隔几秒钟才听到一声爆裂声。当爆裂声很少时，你就可以从微波炉中拿出纸袋，享用你亲手做的零食了。

这个过程的曲线看上去很像我们在图 2 - 1 上看到的那条曲线一样， x 轴表示“爆裂时间”， y 轴表示需要此爆裂时间的玉米粒数量。左边的尾部表示极少数的玉米粒即刻爆裂，然后逐渐增加，接近平均数；右边的尾部也是对称的，整个过程通常不用探究：当你将纸袋拿出来时，肯定会有没有爆裂的玉米。

一家爆米花公司也许将 10 000 磅的玉米粒全部储存在一个筒仓里，玉米粒有一个平均的爆开时间。但是当把这些玉米粒装入 20 000 个半磅的袋子里时，每个袋子的玉米粒爆开的时间是不同的。每一个袋子都是一个样本，它的平均数受到抽样的多样性影响。当你爆一包爆米花时，你就可获得一

个抽样平均数，即对真实总平均数的估算。所有抽样平均数的总评必须是真实平均数，特别是假如在同一时间爆所有的玉米袋时，你可能会想象听到的是一种什么样的景象。所有玉米总平均爆裂时间就是爆裂声的最高点。这一点是有正式验证的，即数学家们所熟知的“中心极限定理”。

我们曾提到从正态分布中所提取的任何观察概率，只要知道 μ 和 σ 值就可以计算出来。同样地，抽样平均数也是呈正态分布的，它们的概率也是可以计算的。虽然抽样平均数几乎都接近总平均数，但是偶尔也会一包裹有多一点的玉米粒爆裂得慢或快。这袋玉米的平均数离总平均数越远，这种情况就越不可能出现。质控人员尤其关心如何使一个公司的产品质量从第一次购买到下一次购买保持一致，因此他们严密监控抽样平均数的分布。

当只有一个样品，而特定的总平均数又不知道时，抽样平均数则是总平均数的最好估算。尽管这仅仅是一种估算，但却是统计学家们所称的“无偏估算”；因为没有特别的原因会使其高于或低于真实总平均数，所以总的来说它还是正确的。但是抽样变化的成分（在总平均数里是不存在的）肯定是存在的。

样本的规模越大，抽样的变化越小，抽样平均数就越可能更精确地反映总平均数。毕竟因为样本规模接近群体大小，所以它包含越来越多的群体；即使样本包含一些异常的观察数据，这些结果也不会对平均数有太多的改变，只要样

本基本上包括了所有的观察数据。而当样本规模缩小时，异常样本的可能性就增加。设想一个爆米袋里面只有四粒玉米。如果碰巧袋内有一粒玉米爆裂的时间很短或很长，则对平均数就有极大的影响。

我们通过创建一个被称为标准误差（简称 SE）的统计量来考虑样本规模所带来的影响，这是从样本数据计算而来的标准差，除以样本总数的平方根。一旦作出这种调整，所有一般的高斯概率统计均可进行。当要计算一个抽样平均数而不是某个特定的观察数据时，可以简单地用标准误差（SE）来代替标准差（SD）。例如，从正态分布中提取的 68% 的样本将会有从总平均数上一个标准误差至总平均数下一个标准误差这个范围内的平均数。

非常奇怪的是，样本不必非要来自呈正态分布的群体数据，这是为了使抽样平均数遵循一种正态分布，以便于高斯概率计算。中心极限定理也证明了抽样平均数的这一特性，但是思索片刻就会发现它没有起初那么违反直觉。假设我们在逐个地查看 6 个选自完全平均分布的观察数据的样本，当用图表示时就形成一个大的矩形。如果是个随机的样本，最有可能的情形是这 6 个结果会或多或少均衡地散布于 x 值范围内，即抽样平均数大致等于总平均数。最罕见的情形是随机样本将包含 6 个观察数据，它们紧紧地串聚在一个极端或另一个极端，因此这些样本会产生与其抽取的总平均数相比较显得极高或极低的平均数，这是十分稀少的。在 x 值范围

内分布既不均匀，也不完全纠集在两极的抽样平均数与真实的总平均数还是有一定的差别，这些样本出现的频率为中频。因此，即使当潜在的群体根本不是正态分布时，抽样平均数最终还是呈平均分布。

在工业和医疗应用方面，真实的总平均数是不知道的，只能获得单个样本的数据。例如，要检验用一种新型灯丝生产的灯泡的寿命有多长，可能要让一批灯泡通电照明，直至它们全部“熄灭”。这种“破坏性试验”是非常昂贵的，通过毁坏所有灯泡以显示如何比较真实平均数和抽样平均数，来检验一个生产厂家的全部产品自然是不值得的。在其他情形中，比如检验一种新的药物，一种样本只提供在这种药品卖给民众前的治愈率信息。这个样本也许很大，也富有代表性，但它仍然是样本。人们设计了一种称作“可靠区间”（简称 IC）的工具，以便在只有一个抽样平均数的时候，来推测潜在群体的平均数。

以下是一段有关可靠区间的陈述：“我们有 95% 的把握说我们这个牌子的灯泡的连续照明时间为 3 500 ~ 3 600 小时。”这种说法是怎么得出的？这种说法又意味着什么？再看一看图 2 - 1，假设这是抽样平均数的分布。由于我们这里处理的是抽样样本，而不是个别的观察数据，所以我们可以用标准误差（SE）来代表图中的标准差（ σ ），而图中的相应概率保持不变。让我们考虑一个特定的样本，其平均数刚好比总平均数少 $2SE_s$ 。对于这个样本，实际上对任何平均数

落于 95% 这个中间区域的样本来讲，真实的总平均数将落在以下的区域间隔：样本平均数 $\pm 2SE_s$ ；而且由于 95% 的样本集中在抽样平均数分布的中央，并具备那样的特性，所以根据抽样平均数创立的 95% 区间将包括总平均数。因此，如果你随机抽取一个 $\pm 2SE_s$ 的抽样平均数，有 95% 的可能这个区间包括样本抽取的那个群体的真实平均数，因而被称为 95% 可靠区间。对于它包含真实平均数这点，你有 95% 的信心。

有时候平均数不是从样本中估算出来的唯一尺度，由标准差 (σ) 代表的变化也无从所知。在这种情况下，标准误差的分子是 s ，即抽样的平均差 (SD)，用来估算总体标准差。在这些情形中，额外的抽样变化被引入方程式，即使不确定性的估量现在仍不确切。另一种分布类似于正态分布，可以用来解决这个问题。尽管形状像铃型曲线，但是 t 分布比正态分布曲线更“胖”。这顾及了更多的不确定性：当 s 而不是 σ 用来估算标准误差时，可靠区间需要更宽一些。因为不确定性是随着样本的大小变化的，所以 t 分布“肥胖”；在较大规模的样本中， t 分布变得消瘦些。例如当样本大小为 5 时，基于 t 分布的 95% 可靠区间由抽样平均数 $\pm 2.447SE_s$ 。当样本大小为 20 时，95% 可靠区间是抽样平均数 $\pm 2.093SE_s$ ；样本数为 30 时，可靠区间是抽样平均数 $\pm 2.045SE_s$ 。从数学角度来看，我们说随着样本数量的增加， t 分布接近正态分布。在许多涉及庞大群体的临床实验

中，即使数据可以提供标准误差 s 的估算，人们还是使用 z 分布， t 分布和 z 分布基本上能给出相同的可靠区间。

可靠区间不一定局限于 95% 的可能性，90% 和 99% 的可靠区间也是常用的。可靠区间的百分比越高，区间肯定越宽。一个 68% 的可靠区间显得非常狭窄，因为只需平均数的 $\pm 1SE$ 范围即可。可是谁会想要一个只有 68% 确定性的估算呢？虽然一个 100% 的可靠区间是最理想的，但这是毫无实际价值的：其范围从负无穷大到正无穷大。要完全确定某一群体的特定参数值，一定要包括各种可能的值，范围要无限。因此在建立可靠区间时，要在估算更大的确定性与更大的精确性方面取得一种平衡。

当然我们不可能把不确定性从生活中去掉，一想到在测定飞机部件这样的过程和医学试验中检测药品效率这样的事情上还有相当的不确定性，可能会使人深感沮丧。但是抽样时，即使在像“抽签”这样的随机事件上也是有范式可寻的。在概率波动上发现的这种规律可以通过各种分布预先进行定量计算，这显然是数学上最不寻常的发现之一，也是最具有洞察力的发现，当代医学、工程学和社会科学都非常依赖这一发现。

3

你可押注的生命表



毕生的交易

1997年8月一位叫让娜·卡儿门德的法国妇女去世了，享年122岁。她是个土生土长的阿尔勒镇人，1888年尚·凡·高在该镇住了一年。凡·高在该镇留下的大量作品给美术史家们留下了深刻的影响。就是在阿尔勒镇凡·高创作了《装有十四朵太阳花的花瓶》、《星光之夜》以及一百多幅其他作品。凡·高给让娜·卡儿门德的影响也很深，那时她在其叔叔开的商店遇见凡·高时才13岁。即使是在童年时代，她对这位艺术家的印象是“邈里邈遑、脾气古怪、不易相处”。到1927年查尔斯·林德伯格横渡大西洋时，让娜·卡儿门德已经52岁了。

在20世纪60年代中期，让娜·卡儿门德与一位律师达

成一笔看似互惠的买卖。这位律师购买她的一套房子，每月低价支付房款。按协议，付款于她去世时终止，然后他才得以搬入此房。这样，她在晚年便有了一笔源源不断的生活来源，而他则可以较为廉价地获得一套住房，但不能确定何时他可以拥有产权。

在支付了 30 多年的房款之后，这位律师先于她去世，享年 77 岁。他的家人继承了这份协议，他们将会得到这套房子，但要获得房子，他们必须承担原先的协议，继续每月支付房款，直至她去世。她去世的年龄超过这位律师去世的年龄 45 岁。

显然，结果证明对于这位律师来说，以这种“廉价的方式”去获得一套房子并非明智之举，事实上他从来都没有在这套房子里住过。但是，基于普通人的一般寿命，这是一笔好交易的这种愿望完全是合情合理的。他无法知道和他做这笔买卖的妇女竟然会活得如此之久，实际上这也是当时有记载的最长的寿命记录。让娜·卡儿门德也无法预料自己的寿命，尽管她觉得自己饮食中大量的橄榄油以及适当饮一些波尔多葡萄酒有重要的功效（当今大多数流行病专家也赞同这一观点）。

虽然个人的寿命无法预测，但当从不同的人群中收集数据并进行分析时，就会出现非常有规律的模型。某年的平均寿命（也称“预期寿命”）与下一年度的平均寿命的变化是微乎其微的，某一特定人群的死亡年龄分布都有其显著而又

能重现的性质。因此，不同年龄段的死亡概率的估算是非常可靠的，这些估算也是保险业的基础。（在英国称作“人寿保险”。保险业可以使人对生活中某种事情放心，并能告诉你这种事有可能何时发生。）卡西诺赌业也是基于概率知识，但是对于保险公司来说这套房子的“获益”要远远高于任何的轮盘赌。与这位不幸的律师不同，保险公司不是赌一个人的寿命，而是千百万人的寿命，因此，显著偏差的频率是可以预测的，因而不会是灾难性的。

保险与亵渎神灵

缺乏可靠的死亡数据意味着保险业在一段时期内不涉足寿险，只专注于其他的保险。例如，大约公元 200 年罗马法学家 Ulpian 就已计算了各个不同年龄段的人生存的比例。尽管 Ulpian 这方面的著作已失传，但他的数据在 Justinian 的《法规汇编》中所引用的一段有关养老保险金计算方法的文字中得以保存（《法规汇编》是一部源自公元 533 年的罗马法律汇编。）虽然这些统计数字的利用率有限，但是古希腊和古罗马时期的保险是针对海运货物的。巴比伦和腓尼基时代也一样。西塞罗（公元前 160 年—公元前 43 年，古罗马政治家、演说家、哲学家）在公元前 49 年曾把他的日用品的海上运输交由私人保险公司担保。保险费用依据承保人的

险损经验和货物的价值而定。某些举世无双的艺术品总被认为是不可予以保险的，人们认为同时产生的多重损失对私人承保商来说巨大。因此，保险人不愿意因多起船损而破产。例如，公元 58 年克劳蒂斯皇帝给商人安排了抵御暴风雨的损失，很有点像今天的联邦机构负责赔偿保险公司赔付能力外的灾难损失。

尤其是在欧洲中世纪时期对寿险的反对达到了巅峰，其他保险仍然盛行，包括为去圣地朝圣者保险：投保人付了保险费，如在途中被俘，则由保险公司支付赎金。但是基于投保人死亡的保险则是不可接受的，因为那时的欧洲已是基督教世界了。雅克·迪帕克尔在他 1996 年版的《死亡率表》中写道：

推断生命长短与传统的基督教死亡观念不符。在封闭的中世纪思想体系中，死亡有着神圣的特点，不仅不能成为推测的目标，而且企图借助法律来控制生命也是不合时宜的，是对神灵的亵渎。每个人的命运隶属于神圣上帝的意志，神可以在任何时候终止其生命，要么出于对善的回报将其招入天堂，要么为了惩恶将其打入炼狱，使其遭受永久的毁灭。这排除了所有的预测，甚至所有的计算。

从中世纪到文艺复兴时期，这种态度变得不是那样强硬

了，而且社会组织、银行、财会和其他金融机构变得越来越稳定，越来越复杂，提供人寿保险也逐渐变得越来越可行。与此同时，反对人寿保险的意见也失去了根基。因此，尽管保守的统治集团基于道德的原因反对人寿保险，但必须通过法律来禁止这种保险，因为光是道德说教是不够的。例如，1570年在西班牙人寿保险成为非法，在1598年的荷兰人寿保险也成为非法。但是，思想的变化很快。不到100年，对死亡率的计算以及对寿命的预测已经建立起来，而且是以数据、概率和统计学为坚实基础的，似乎也没有人认为应予以反对。神学家或宗教上层人物均未对此提出异议或出面阻止。“更妙的是，”迪帕克尔写道，“在人类各种现象中的统计规律的发现很快被解释为神谕存在的一种新的证据。”

格兰特生命表

1687年牛顿的《自然哲学的数学原理》出版，让世人认识到宇宙的各种数理规律。另一个英国人约翰·格兰特使人们对生死概率的数理规律的看法发生了彻底的变化。格兰特1620年出生在伦敦，他一开始的工作是做买卖，尤其是做零星服饰用品交易。他担任过多个公职，并获得过少校军衔。1662年格兰特出版了《关于死亡表的自然的和政治的考察》，此书一出版即大受欢迎。1663年他当选为英国皇家

学会会员 至 1665 年该书第四版印刷发行，这一年伦敦发生大火，而一年前大瘟疫最后一次肆虐英国。

死亡表是根据伦敦堂区工作人员呈交的有关死亡人数的报告而编制的报表，这些报告含死亡年龄和死亡原因的分布。这些报表的编制在伦敦至少始于 1532 年，从 1625 年开始便以印刷形式发放，这有助于政府追踪瘟疫及其他流行病的传播情况。格兰特仔细研究了这些数据的质量及其可能的缺陷，借助于均数与分布情况来分析这些数据。而他所做的最有价值的贡献则是构建了当今统计学家们所称的寿命表。他认为这种寿命表可以说明并能控制死亡的隐含规律的存在。

表 3-1 格兰特关于存活率的数据

每 100 个新生儿中	
在满 6 周岁时	64 人存活下来
在满 16 周岁时	40 人存活下来
在满 26 周岁时	25 人存活下来
在满 36 周岁时	16 人存活下来
在满 46 周岁时	10 人存活下来
在满 56 周岁时	6 人存活下来
在满 66 周岁时	3 人存活下来
在满 76 周岁时	1 人存活下来
在满 86 周岁时	0 人存活下来

资料来源：格兰特. 关于死亡表的自然的和政治的考察. 伦敦，1662

格兰特不仅向人们展示了一连串实际观察到的人群中的特殊数字，而且向人们说明死亡是怎样减少一组有共同特点的人群数量的。挑选一个假设的“整数”人群来制作与生命表进行对比的死亡表，将观察所得的死亡率用于这个整数，他发表了表 3 - 1 中的数据，一半以上的人于 16 岁之前死亡。

现在我们起用该表是从 0 岁开始，而不是从妊娠开始，不同年龄范畴都会用到。但是在追踪适当的共同群体经历时，运用不同年龄段死亡率的概念是相同的。

戈特弗里德·威廉·莱布尼茨曾在微积分发明者的归属上与牛顿有过激烈的竞争，他对生命表的分析有重要贡献。他设想出如何在获得像格兰特这样的数据时计算出不同年龄段的预期寿命。这一发展使当今更为精确的生命表成为可能，这些生命表是有关人类生存各方面信息的来源，内容非常丰富。

从死亡率到预期寿命

什么是预期寿命？如果你仔细观察一张当代生命表，那么这一统计量的含义及其与死亡率的关系就会更加清楚。表 3 - 2 是一张最近的生命表。这些数据是关于 1996 年美国女性的死亡率，来自国家疾病控制中心的一个分支机构——国

家健康统计中心网站 (www.cdc.gov/nchs/data)。大多数国家生命表都是几年之后出版的。这些数据的依据是死亡证，因此要确保几百万死亡人数的精确性和完整性，然后将其以计算机格式保存，以便分析，这确实是项艰巨的任务。

表 3-2 节略性生命表提供了有关 1996 年美国女性死亡率数据

$X \text{ to } X + N$	${}_N Q_x$	l_x	${}_N D_x$	${}_N L_x$	T_x	E_x
00 ~ 01	0.006 59	100 000	659	99 435	7 907 507	79.1
01 ~ 05	0.001 35	99 341	134	397 043	7 808 072	78.6
05 ~ 10	0.000 83	99 207	82	495 812	7 411 029	74.7
10 ~ 15	0.000 93	99 125	92	495 426	6 915 217	69.8
15 ~ 20	0.002 20	99 033	218	494 654	6 419 791	64.8
20 ~ 25	0.002 42	98 815	239	493 488	5 925 137	60.0
25 ~ 30	0.003 11	98 576	307	492 128	5 431 649	55.1
30 ~ 35	0.004 30	98 269	423	490 336	4 939 521	50.3
35 ~ 40	0.006 08	97 846	595	487 848	4 449 185	45.5
40 ~ 45	0.008 58	97 251	834	484 325	3 961 337	40.7
45 ~ 50	0.012 69	96 417	1224	479 247	3 477 012	36.1
50 ~ 55	0.020 36	95 193	1 938	471 421	2 997 765	31.5
55 ~ 60	0.031 50	93 255	2 938	459 363	2 526 344	27.1
60 ~ 65	0.050 68	90 317	4 577	440 808	2 066 981	22.9
65 ~ 70	0.074 84	85 740	6 417	413 497	1 626 173	19.0
70 ~ 75	0.116 07	79 323	9 207	374 780	1 212 676	15.3
75 ~ 80	0.174 95	70 116	2 267	321 360	837 896	12.0
80 ~ 85	0.277 21	57 849	6 036	250 275	516 536	8.9
85 以上	1.000 00	41 813	41 813	266 261	266 261	6.4

生命表总是以一个年龄段为一行来排列的，有些生命表非常详细，以年龄数提供信息。当整张表以一年为年龄范畴时，此表称作非节略性生命表。表 3-2 即为节略性生命表，

第一列显示的是每行的年龄，从 X 岁（此年龄段开始的岁数）到 $X + N$ 岁（ N 为年龄跨度）

那么 10 岁到 15 岁的死亡概率是多少呢？严格地说，这是特定年龄的死亡概率。在表 3 - 2 中由 Q 表示，与其他的概率符号一样： P 是这一年龄段的存活率， $Q = 1 - P$ 为余率。这些比率是从死亡证和人口的出生日期计算出来的，无须其他东西来计算生命表。剩下的事就是用格兰特的推理方法和莱布尼茨的改进方法来进行演绎，但是要得到 Q_x 这一列数据也并非那么简单。某一年 X 岁死亡的人在那一年初并非都是 X 岁整，有可能死于 X 岁的这一特定人群随着这一年的过去也在变化，因为有些人达到那个年纪，而另一些人则安全地步入下一岁数。因此，有必要作出调整以确保死亡人数与在某一特定岁数有死亡风险的准确人数相关。这个净值就是 ${}_5Q_{10}$ ，即 10 至 15 岁的死亡概率是 0.000 93。

假设我们以 100 000 人为例，在 X 岁时还有多少人还活着？这在表 3 - 2 中第 3 列有显示，与格兰特发表的数据相符，以 l_x 标记。这 100 000 人称作生命表的基根。最初的一组将减少 0.006 59 点，或者说在生命第一年有 659 个婴儿死亡。死亡人数 659 在标记为 ${}_N D_x$ 的这一列显示。从同一行的 l_x 列中减去某一特定年龄组的 ${}_N D_x$ 就可以得到下一行新的 l_x 。这里我们用 100 000 减去 659，得到 99 341，这就是 l_1 的值。

在这一点上，我们知道每个年龄组的人数，因为我们知道在这一年龄组的死亡人数，因而我们也可以获得存活的人

数。现在我们可以计算一个重要的统计数字，这就是该年龄组的累计生存人年数（标记为 ${}_N L_X$ ）。它之所以重要是因为预期寿命表示已活到该年龄的人平均预期还可以活多少年，因此计算需要每个年龄组的人年数。编制 ${}_N L_X$ 这一列，可以用该年龄组起点年龄 X 岁的人数乘以年龄组跨度。由于步入该年龄组的人期间死亡，所以年龄组的人年数必须减少一点。假设该年龄组期间的死亡是平均分布的，那么合理的近似值是那些期间死亡的人的年龄跨度减半。这个基本原则必须遵循，但是死亡风险不会绝对平均分布的。当获得准确的死亡年龄分布数据时，便可用来编制 ${}_N L_X$ 。

假如你从该表的底端开始往上累计计算 ${}_N L_X$ 的值，便可得到 ${}_N T_X$ 的值。例如，在最后一个年龄范畴内 ${}_N L_X$ 的值和 ${}_N T_X$ 的值是相同的，但是 T_{80} 是85岁时 ${}_N L_X$ 的值（266 261）与80岁时 ${}_N L_X$ 值（250 275）的和，即 $T_{80} = 516 536$ 。有了这一累计数字，便可以计算出平均预期寿命。第一行 T_0 的累计数是同一批人在0岁之后（即出生之后）所生活过的人年数总和，因此我们可以发现同一批人中的某一成员可以存活平均年数。而这个特殊的平均数叫做出生时平均预期寿命（ E_0 ），由 T_0 除以 l_0 而得。更笼统一点地说就是在任何年龄（ X ）

$$E_0 = T_X / l_X$$

寻找普遍规律

表 3-2 中在死亡率和平均预期寿命上有一种与年龄相关的形态，这种形态在各地大规模人口的生命表中均有所反映。生命第一年的死亡率，即婴儿死亡率较高。实际上这个死亡率与随后岁数的死亡率间的差别太大，因而即使在节略性生命表中第一年的数据几乎总是单列的。有些婴儿受早产或先天性缺陷的折磨，这些疾病以及对生命有害的各种各样的畸形在第一年显得尤为明显，影响最大。随后死亡率急剧下降，在 5~15 岁间为最低。在这个年龄段严重疾病稀少，父母的监护又使致命事故较少发生。接下来青年期的各种放荡行为、偶发事故、驾驶导致 ${}_Nq_x$ 值直线上升，随着年龄的增长，死亡率有规律地快速增长。有趣的是，在表 3-2 中 40 岁时的死亡率超过婴儿的死亡率。

各个国家不同时期的生命表中的特定指标和其他详细数据是不同的，这些差异是由社会条件决定的，如是否获取适当的营养、干净的饮用水和医疗照顾等。在大多数国家由于生活水平的逐渐改善，某年至下一年度特定年龄的死亡率不会有显著的变化，但俄罗斯最近出现的相反情况则完全是个例外。在阅览特定生命表时，前段所述的 ${}_Nq_x$ 值的形态仍然是显而易见的。人类社会这种形态的普遍性一定是源于生存

老化的生理特征。格兰特的预期已经得到随后经验的证实。

显然，平均预期寿命从一个年龄组到下一个年龄组呈下降趋势，但是在表 3-2 的 E_x 列中有个明显的奇怪现象。例如看 60 岁的女性，平均还可活 22.9 年。然而，如果她活到 65 岁，则她的预期寿命并不是减去她所活过来的这 5 年，即不是 17.9 年， E_{65} 是 19 年。这种现象有时被人口统计学家们诙谐地总结为“你活得越长久，你的寿命将越长”。人们认为这是身体虚弱多样性的结果，常与选择有关。由于患心脏病、癌症，甚至遭遇交通事故的倾向具有多样性，那些有可能遭遇这些的人要比其他人更早经历这些事情。因此，那些不易受这些问题影响的人构成了随后年龄范畴中不成比例的小部分人数。这种现象在高龄段更为显著（也许是因为健康因素尤为重要），但不幸的是它不会无限地延长。

与自己的存活率打赌

当购买固定年限的人身保险，如 1 年或 5 年，实际上你是在和保险公司打赌。你赌的是你在此期间将会死亡，而保险公司赌的是你在此期间有很大的存活概率。保险公司只有在你死亡的情况下，才输掉这场赌博，并给予赔偿。保险公司也有各种各样的生命表，并相应地对保单进行定价。例如，在表 3-2 中一个 35 岁的妇女在未来 5 年内的死亡率是

0.006 08。因此，如果一家保险公司对 35 岁妇女每份 1 000 美元、保期为一年保险收取 6.08 美元的保费，那么这家保险公司刚好有足够的钱来支付这一组别人的赔偿费。当然，保险公司出售的保单价格比这要高，含有管理成本，如办公室和员工的费用，并要保留一定的利润盈余，还要储备一笔资金，以防出现被保险人死亡率波动上升的可能性。不过，对保险公司来说，估算出实际的保价相对还是容易的，因为有生命表，所以保费的其他要素就可以计算出来。此外，通过要求被保人在签发保单前进行体检，保险公司就可以确保子集人口更加健康，而不是完全依据生命表所反映的整个人口情况。因此保险公司很少有破产的。

当然，概率的波动对保险公司来说是至关重要的，但几乎不会威胁到人寿保险的利润。为了避免大量人数的同时赔付，由于战争因素或不可抗拒因素而导致的死亡通常被特别排除。“不可抗拒因素”是法律语言，包括对城市造成毁灭性的地震，这里不包括自然死亡，甚至由传染病引起的死亡。很显然，我们的司法体系所使用的语言完全有别于中世纪的概念。

排除了这些因素，自然的死亡率波动就不重要了。全国的死亡率是基于千百万人口统计而得出的，偶尔出现百岁以上老人或连续死亡的情况，对平均预期寿命的影响是微乎其微的。甚至爆发疾病，包括艾滋病，在西方国家对全国的生命表也不会产生影响（据统计，北美 0.58% 的居民感染艾

滋病病毒)。保险公司的秘诀是有一群从整体上看能够代表全体国民的投保人，并尽可能达到全国规模。保险范围越大、越全面，受取样波动和连续死亡的影响就越小，因而所观测到的死亡率更能接近逐渐发生的变化。

如果概率被保险公司彻底理解，并按有助于他们的方式重新洗牌，那么购买人寿保险岂不是不明智的赌博吗？其实不然。即使赢取这场赌博的可能性很小，人寿保险在用于保护家人抵御经济风险时还是有用的。当大笔投资能产生年均10%的收益时，大多数金融顾问建议家庭挣钱者应购买价值相当于被保险人年均收入10~15倍的保险。如被保险人死亡，那么这个家庭的经济状况基本上保持不变。由保险资金提供的投资收入大致取代了死者的收入。因此，购买人寿保险是财政计划的重要组成部分，由保险公司来有效保护人们抵御破产在社会上是最理想的。

虽然我所描述的这种人寿保险是一种直接的赌博，但却有许多的变化。一方面投保人希望他们进入后续年龄段时有权更新自己的保险。当任何特定人群的年龄增长时，患病的和保险公司要排斥的人将会越来越多，因而可变更的保险费用就要额外高一些。

另外一种变化称作保费递减保险。普通保险的保费是随死亡率变化的，在较高年龄段保险费剧增，此时已无住房和孩子教育费用的支出。为了使保费恒定不变以使保险费支付逐年递减，通过变更保单，这两个变化可同时兼顾。

另一种产品适用于那些坚持固定保费的人，这就是终身死亡保险。这种保险的保险费是固定不变的，但是保费要高于初级承保范围。扣除的免赔额构成被保险人在缴足若干年保险费后不愿继续投保而收回的现金价值；对那些继续投保（继续生存）至更高年龄范畴的保险而言，现金价值可用来应付额外的风险，特别是保险公司投资超期时。此外，终身保险还为保险公司带来额外的现金，尤其是万一发生资金短缺或需要建新的办公楼时，这就显得更为重要。

就是你这种类型

有些风险因素的分布比艾滋病病毒还要广泛。在这种情况下，将增加的风险和高度的流行性结合起来有助于分开计算那些高风险人和低风险人的生命表。尤其是与汽车交通事故相关的保险的定价依据购买保险的人的类型而不同，年龄、性别、居住状况及过去发生事故的记录在客户分类、预测风险方面都起着重要的作用。

另一个例子涉及吸烟。许多保险公司希望保险费能反映因吸烟者的高死亡率而增加的费用。以较低的价格向非吸烟者提供保险也可获利，非吸烟者自然从被划为低风险人群而获益。因此，虽然我说过个人的寿命是不可预测的，但是这并非绝对的。生命表给我们提供了平均数（某一类人的预

测)，这些数据对于赌博来说已足够了。一个人的寿命长度也许与平均寿命有显著的差异，但是对一个人的风险因素了解得越多，就越能减少这个变量，对特定种类的人所预测的死亡率也就越准。

风险组别的差异是非常重要的，图 3 - 1 是根据美国癌症协会的癌症防治研究项目中大约 50 万的随访而计算出来的每十万个男性中的年均死亡率。最底端的线显示的是非吸烟者的患肺癌死亡率。尽管一生中患肺癌的风险略有上升，但对非吸烟者来说，风险确实很小。

另一个极端则由图 3 - 1 最顶端的线显示，在随访研究中追踪连续吸烟者的经历，他们的风险远远高于非吸烟者。事实上在 80 岁时，每十万人中患肺癌的死亡人数超过 1 500 人，每年死亡率多于 1.5%，这一比率远远大于最低曲线。

与连续吸烟相比，戒烟是有好处的。在图 3 - 1 中，从上往下看，中间的曲线代表那些更早戒烟的人群的死亡率变化。但是请注意，这种风险不会回到从不吸烟的人的水平。例如，一位于 50 岁初戒烟的人在 70 岁时死于肺癌的风险比从不吸烟的人在 70 岁时死于肺癌的风险要高好几倍，比在 30 几岁戒烟的人也要高一倍。尽管正在戒烟者死亡率没有下降，停止吸烟还是非常重要的，因为相对于那些持续吸烟者的风险来说，戒烟相对会带来好处。

有些终生吸烟者不得肺癌，而有些非吸烟者则早年就死于肺癌。大多数人也许认识某位像我婶婶玛丽这样的人，一

生大部分的时候吸烟很厉害，却在不到 101 岁时才去世，但是他们也许不了解许多这样的人。虽然玛丽超过了平均数，但是吸烟过量的朋友和亲戚均患有严重的疾病，或年轻时就去世了。尽管我们不能预测个人会发生什么事情，我们却肯定可以通过概率分析来推断将有可能发生什么事情。图 3-1 中的概率对于各个组别的人来说都是可以预测的，因为 1 000 名终生吸烟者死于肺癌的比例远远高于非吸烟者，他们被归入低风险组，而不是高风险组。

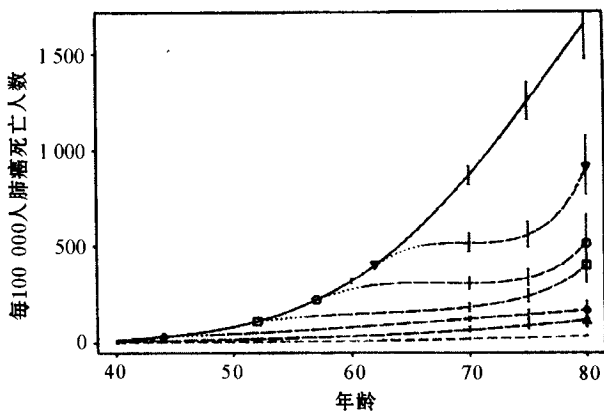


图 3-1 根据吸烟者开始吸烟年龄为 17.5 岁，每天吸烟 26 支，按年龄段计算男性肺癌死亡率，现在仍以吸烟者为实线、先前吸烟者为虚线、从不吸烟者为点线。这 5 组戒烟者按戒烟时的年龄及在 80 岁时的区分如下：△代表 30~39 岁，◇代表 40~49 岁，□代表 50~59 岁，○代表 55~59 岁，▽代表 60~64 岁。资料来源：M. Halpern, B. Gillespie, K. Warner. 曾经吸烟者的肺癌死亡率绝对风险形态分析。国家癌症研究院学报，1993 (6)：457

可供选择的受孕方式

生命表方法的使用决不局限于对死亡率的分析。任何事件只要有跨时间的发生率分布，就可用生命表进行分析。如在工程学方面，机械或电子元件的寿命通常是研究的对象，生产商提供替换保证，因为他们提前知道其产品的使用寿命。有了这种知识，他们就可以设定保修期的长度，这样保修期就可在出现多处故障、维修、零部件更换成本上升前终止。这就是为什么你的那些小型电器产品的保修期刚好在坏之前就过期的原因。此外，厂家对许多像电冰箱、高级音响等电子产品也出售维修合同。维修合同的价格依据产品的寿命表而定，设定的价格超过这批产品与故障相关的付出成本，方法类似于人寿保险费的设定。但是，许多人发现这些高价的维修合同有点令人讨厌，因为几年的保修费用可能只占更换零部件成本的极少一部分。

生命表不仅可以用来研究死亡率，而且可以用来研究事件的形态，所研究的事件不必尽善尽美。生命表也可用来研究所探究事件的等待时机，最明显的例子就是怀孕和生产。

在不孕研究上，生命表的分析是常见的。在健康营养良好的普通人群中，大约 85% 想怀孕的已婚夫妇婚后一年内会怀孕，另外约 10% 第二年会怀孕。此外，通常还有某种医学

原因导致不孕（尽管有时仅仅是运气不好）。受孕率最高的夫妇最有可能立即怀孕。在死亡率方面，以前有人说过“活得年岁越大的人将活得越久”，但是就怀孕来说，则可能是“等得越久的人将要等得时间越长”。随着时间的推移，还未受孕的人群包括逐渐增多的完全不孕的人（以及那些长久等待怀孕的人）。

表3-3 用黄体酮栓剂治疗的病人的生命表分析数据

时间(月份)	到月底时的累计受孕率
0	5.7
1	20.0
2	28.6
3	31.5
4	37.2
5	48.6
6	60.3
7	69.9
8	83.3
9	91.7
10	100.0
11	100.0

资料来源：D. L. Murray, L. Reich, & E. Y. Adashi. 生殖与不育，1989（1）：51；数据经美国生殖医学协会同意引用。

不仅怀孕概率可用生命表来研究，而且为了增加这种概率的治疗效果也可用此方法分析。如不孕症的一个可能原因是所产生的黄体酮不足。即使在某些情况下可以怀孕，但黄体酮不足很难保住胎儿。这种不足可以通过使用黄体酮药栓

来克服。医生是怎么知道这一点的呢？支持这种治疗方案的证据来自表 3-3 中的生命表。表中显示黄体酮的补充可以导致接近或超过正常受孕水平的比例。对不孕症的治疗通常是用生命表来判断的。

表 3-3 中显示的是经特别挑选的小样本。本研究的抽样人群仅包含已知由于黄体酮水平低而患不孕症的人，研究的人数较少，不到 100 人。在不到一年的时间里累计怀孕率达 100%对门诊病人来说是不现实的，因为这个试样量排除了治疗对那些不起作用的小比例精确估算（并非所有的门诊不孕病人都有这种特殊问题）。

人口学家也用生命表来分析社会因素对怀孕概率的影响。如社会科学工作者注意到在较贫困的农村，通常由于各种传统的、宗教的价值观和农业劳动力的需求，避孕方法的使用率较低，大家都希望有大家庭。与此同时，婴儿死亡率很高。这就导致在这种情况下，妇女在小孩较早夭折后尽快再次怀孕。这种“孩子替代的假设”是否有数据支持呢？

表 3-4 中显示的数据来自 1991 年尼泊尔的生育率、计划生育与健康调查。第一列所列举的是自生产头胎后间隔逝去的时间，另外两列显示的是生产第二胎的相关概率。概率的统计对那些头胎小孩死亡的人和头胎小孩存活的人分开进行。在每一个逝去的时间段，如果头胎小孩死亡的话，生产的概率就更高，所以孩子替代假设有这种类型研究的数据支持

表 3-4 二胎生育的累计概率

自头胎小孩出生后的月份数	头胎小孩	
	存活	死亡
0	0.000 0	0.000 0
9	0.003 5	0.022 8
12	0.017 4	0.113 5
15	0.038 0	0.200 6
18	0.066 7	0.267 9
21	0.115 9	0.346 6
24	0.188 8	0.478 5
27	0.257 1	0.557 1
30	0.325 3	0.614 5
33	0.391 4	0.669 3
36	0.463 6	0.723 7
39	0.523 9	0.754 8
42	0.574 7	0.779 5
45	0.619 9	0.799 8
48	0.654 9	0.829 9
51	0.689 1	0.851 8
54	0.716 4	0.861 7
58	0.753 8	0.875 5

资料来源：Sameer Rajbhandary 尼泊尔的生育率与儿童死亡率的
计量经济学分析，博士论文，2000

当我们看有关出生率与死亡率的统计数字时，我们不仅对其有规律的可预测性形态感到惊讶，而且也对其所描述的潜在事件出现的相对频率感到惊讶。育龄期年轻妇女在特定年份生育的可能性大于死亡的可能性。但是死亡也不是稀有事件，表 3-2 中最低的死亡率是 5~10 岁的孩子，为

100 000人中死亡 83 人，或者说不到 1%。一个购买彩票、希望中头奖的人，在购买彩票这一年死亡的概率大大高于兑取头奖的概率，这对任何一个参赌的人来说都是一个令人不安的观察。但是，当我们观察稀有事件时，比如像彩票的头奖及其他发生概率极小的事件时，我们发现抽样的变量显著增加，因而增加了预测的困难，尤其是与有规律性的生命表进行比较。而且对于越是稀有的一连串事件的解释就特别成问题了，因此就发明了对稀有事件的统计方法，我将在下一章中予以讨论。

4

稀有事件



是不是令人震惊

曾经有一个美妙的夜晚，突然一个可怕的念头掠过我的心头：我的祖母刚刚去世。很难解释其中的缘由，但我对这一直觉感到惊讶，很快这一直觉就得到证实。果然不一会儿，电话铃响了。我带着沉重的心情拿起电话，镇定地接受这一噩耗，然而电话却是修理工打来的。这个电话根本与我祖母无关，她后来又活了好多年。

谢天谢地，这样的事情很少重复发生，因而不太引人注意，如果这些事情引起惊人的结果，比如某人想象自己或梦见自己收到死亡或获奖的通知，而且不久之后，确实这样的事情发生了，这就引人注意了。这些结果不仅值得说明，而且有时候还会当作超自然力的证据或者超人给受惠人启示的

证据。

这种解释是没有必要的。毕竟在任何特定的日子总有千百万人会产生类似的想法，而肯定会发生某件事情。实际所发生的事件在所有的想象中只占很小的比例。由于概率的原因，有些想象的事件一定会证明是很精确的。担心亲人死亡很普遍，因此在任何特定日子，死亡肯定会夺走少量亲人的生命，而有很多人在为此担心。

做梦的原理也一样，在大城市的特定夜晚，人们做着无数的梦。许多梦是无意义的，但也有人确实梦见了熟人或可能发生的事。有些梦确实与现实所发生的事吻合，这一点也不奇怪，只要可供选择的梦与可匹配的、潜在的事件的数量都很大。你的任何熟人也许会买一辆新车、订婚、换工作或者赢得赛马。但是，估算与实际所发生的事件的相符比例是困难的，因为传到你耳朵里的相关说法仅仅是那些成真的梦只占一小部分。更多的是不相符的，常听到有人说“我在梦中见到了老朋友，但实际上并没有见到”。没有巧合的乏味的梦境是没有人传说的，而一个人做了一个梦，这个梦又得到应验，他就会讲述这个经历。这样的梦往往会使人毛骨悚然，因为它们通常被视为超自然力的展示。“在我一辈子所做过的成千上万的梦中，有些具有可预见性的梦应验了。”作出这样的评价不是很常见的，人们不想把自己惊人的经历看做是寻常的巧合。

有时不能将这些结果不假思索地认为是巧合，因为因果

联系是存在的，不管这种可能性看上去有多小。典型的例子就是伏都教巫术死亡现象。反对伏都教信仰的西方医学工作者观察到施咒语，随后便出现被施了魔法的人得病，然后出现死亡这样的现象，这些现象在医学文献中记载了下来。对任何特定个人来说，死亡率通常是很小的，但被施了魔法的人的死亡率就更高。肯定有某种东西在起作用，绝不是随机概率或偶然巧合。例如，肯尼斯·戈尔登于 1977 年在《美国精神病学杂志》上发表的一篇文章展示了下面这个病例：

一名来自阿肯色州靠近小石城的 33 岁黑人男子被送到阿肯色医科大学接受神经治疗。病人最近经常发作，变得越来越易怒，逃避家庭。在神经病房无法安全留住时，病人被转移到精神病房，在那里他变得更加暴躁、神志不清，甚至说胡话。当人们接近他时，他变得非常害怕，开始产生幻觉。在注射了 100 毫克的氯普马嗪（氯丙嗪）之后，他最终镇静下来，但仍有必要被限制在床上。然而神经学研究，包括大脑神经扫描都证明他一切正常。

住院两周后，病人心跳停止，所有抢救努力都失败，尸体解剖没有发现死亡原因。

而病人的妻子却提供了死因。他得罪了村里的一个妇人，人们认为这个女人被赋予了施诅咒的魔力，因此，依照

他们的信念，他的死亡是不可避免的。

1992年，C. K. Meador博士在《南方医学杂志》上发表了另一篇文章，记述了两名医生所观察到的事情。一个自认为在与一名伏都教牧师争吵时被施与了死符咒的人已停止了进食，变得非常虚弱，要求住院。他依靠导管喂食，一直处于昏迷状态，接近死亡。医生发现其器官并无病变。

在惊恐万分的病人妻子和亲人面前，这主管医生向病人“透露”说，自己最近和那位伏都教牧师大吵一架，在医生的刑讯逼供下，这位伏都教巫师泄露了病人疾病的天机。由于伏都咒符，病人体内的一只蜥蜴正在蚕食他的食物及其脾脏。解释了之后，医生给病人注射了一针催吐针，通过奇妙的办法，最后吐出了一条蜥蜴。病人后来睡着了，第二天醒过来时胃口极大，一星期之后便出院了。

这些疾病以及其他医学文献上记载的相似病情都是有生理原因的，他们都是由于巨大精神压力而导致的生理影响结果。对于在被施与了伏都教咒符后而导致心脏病发作而暴死的现象，神经病学专家已提出了各种各样的说法。在任何情况下，过度的害怕可能会引起致命的心脏病，正如我们经常看到相关的报道，在年迈的抢劫案受害者和人质中常会发生这样的猝死。显然这不是巧合，伏都教咒符导致的死亡是有某种夸张的负面安慰剂作用造成的，而不是伏都教巫师所声称的直接机理造成的。在“被施与魔咒”的人中，暗示的法力是生理疾病根源的证据是很有说服力的：因伏都教

魔法而死的病例只有在受害者及其朋友和亲人都相信伏都教牧师的魔力会致人死亡时才会发生。他们必定都相信特定的咒符是致命的，因而会像对待任何即将死亡的人一样对待这个受害者，以使这种幻觉圆满而有效。因此，这就有了一种自然解释。这种推理不局限于“发生于其后者必然是其结果”这一谬误。

稀有事件的统计

有许多种情形，其中导致某一特定现象的机理根本弄不清楚，或者存在严重的分歧。此时完全基于这种罕见度要把概率和偶然性作为这些事件的根源来加以区分是困难的。对完全由于概率的作用而出现的罕见事件的经典例子与猴子敲击键盘有关。我的电脑键盘有 55 个键，随意敲击键盘的猴子大约有 $1/55$ 的概率击中用来打印某个特定字符的键。（我这样说大约是因为所有的键并非一样大小，有些键的位置比较容易敲击，等等）。我的文字处理程序告诉我本书前一章有 41 400 个字符。所以随意连续击键 41 400 次就可打印出这一章节的概率是 $(1/55)^{41\ 400}$ 。这是一个非常小的概率，但也不是 0 如果有足够的猴子和时间，就可指望最终形成这一章节的字符序列是可以任意打出的（对任何一个作者来说都是一种愚见）。

该例子的启示是，你碰到一篇文本，于是便判断它是人的意志力的结果，而实际上它是事件的任意结果。但是，这个启示显然是不真实的。当你读一篇出版的作品时，你知道这个作品绝不是任意敲击字母就可生成，或者随意选择一些单词就可碰巧形成可读性的文本。（有人也许会争辩说，20世纪20年代左右在巴黎兴盛的达达派文艺运动确实产生了随机从布袋里抽取单词而组成诗歌的例子。此外，有些读者发现詹姆斯·乔伊斯的后期作品非常接近随意之作，尽管存在着这些作品究竟是否有可读性这样的问题。）

现实一点说，任何人都知道可读的书绝不是随意之作，之所以知道这一点是因为我们对猴子和人了解得很多，不仅了解概率。当制造出上面的例子时，当有人观察到完全不可能发生的事时，就可能真的出问题，而当面对大量的不确定因素时，你的概率知识就显得极为有限了。

例如，假设有一种癌症不遗传。显然这种癌症的出现是偶然的（医生称为“非流行性的”），人口中得病率约0.5%，这是一种罕见的病。但是，住在靠近垃圾堆附近的一个人家却有三个孩子得了癌症。这似乎不是碰巧的偶然，我们可以说住在垃圾堆附近是危险的。毕竟在随机选择的一组三兄弟中，三个人都患癌症的概率是 0.05^3 ，或者说是 $1/800$ ，这与大多数人对一件不可能发生的事的看法是一致的。然而在一个像美国这样大的国家里，超过一百多万的三个家庭有三个孩子。如我们设想这些家庭 $1/8\ 000$ 个家庭有三

个孩子都患癌症的话，则 100 万个家庭里就有 125 个这样的家庭。

这些以每三个病例为一小组的预算是以概率为基础的。如果这 125 个家庭像任何其他 125 个随机挑选的美国家庭一样分布各地，许多家庭会在好的地段安家，但还是有一些家庭将住在靠近垃圾堆、高压电线、化工厂及相类似的地方。那些人只会将自己的经历认为是真实的。家居住在垃圾堆附近，多个孩子患了罕见的癌症，还有什么比接近垃圾堆和这种疾病的关系更明显的呢？

流行病专家都知道当小概率加到大分母上时，在没有因果联系的情况下也会出现令人惊讶的巧合事件，判断这种联系是否是偶然的不仅需要对所选的样本有所了解，而且需要对这种疾病在人口中的整个类型有所了解。因此，如果“三兄弟同患病”的概率一般是 $1/8\ 000$ 个家庭，但是对住在靠近垃圾堆的家庭来说则是 $1/800$ 的话，我们就可以说那些受影响的家庭相对来说其危险性要高 10 倍。这还不足以作为偶然联系的证据，尤其是在没有已知的机理时，可是与挑选的单个例子相比，这肯定算是好的证据。之所以需要这种相对的证据是因为在完全没有任何关联的情况下，受影响的人群和不受影响的人群，他们的概率都是 $1/8\ 000$ ，但是有三人患癌症的家庭仍然会对这种疾病与他们附近任何看上去可疑的东西有“联系”记忆犹新。他们所说的“联系”可以称为轶事性的证据，这种说法在那些提供这些观察的人看来

是侮辱性的。所以，这种证据的最大价值也许可以提出用科学比较来回答的问题，这些比较涉及不同组别中出现的病例数量和概率。而轶事性的报告本身对接触影响所带来的危险没有提供足够的信息，最糟糕的是它会误导人，充其量也只是一种不能令人信服的证据。

在没有受过科学训练的人中，许多人反对这种结论（或非结论）。首先，有些人坚信科学家能够根据任何现有的证据作出判断，而不管信息是多么稀少。“不能够确定”通常被认为是不可接受的答案。第二，人们不喜欢自己的经历受到低估。还有一个关键原因就是为什么有些人憎恶对一连串“稀有事件”的怀疑态度。他们不熟悉一种源自概率原理的重要现象：偶然性是突现的。

如果偶然性是突现的似乎是非直觉的话，想象一下抬头望着夜空，看到星星井然有序地排列在一个栅栏里。在这个栅栏里星星分布在方框的各个角落，因此天空就呈现出成千上万的方框。看到这种现象你也许会感到惊讶，很难想出一个偶然性较小的安排。在某些方面与这个栅栏相对的肯定是随机分布：有些地方是空的，有些地方则布满了星簇。只有排列非常有规则的栅栏才不会凝集成簇。

很明显随机分布疾病的地点会出现有的地方没有疾病发生：在随机分布的图上会出现分布不均的散点。互补的情况是地图上没有发病的地方必须由多发病的地方来进行平衡。

从这些发病集中的地方采集数据再把这些信息用到不发病地

区，然后认为这些孤立的集中病例是令人震惊的、罕见的，这是不科学的。疾病集中出现的程度是否令人震惊必须要看引起整个疾病类型的环境而定。

为什么必须考虑整个环境，而不是考虑导致特定家庭出现多个病例的稀有个案呢？可以考虑以下这个类比：掷一排硬币，要使10个硬币全部为头像的概率不到 $1/1\ 000$ ，所以一旦掷出的结果是头像的话，那倒是令人惊讶的，但这仅仅是因为单个序列的10次投掷而令人惊讶。假如你拿起一个硬币，连续掷10次，如果全部是头像的话，那才是真正令人惊讶的事。但是，如果做几百万个投掷10次的试验，你事先就知道有几千次投掷的结果将是头像。在这种情况下，挑出一次投中10个头像，就坚持说投中10个头像肯定是有原因的，决不仅仅是偶然的，这肯定是不对的。

要确定某一疾病是否有特定的潜在生理原因，而不是偶然发病还是有问题。假设在进行比较的两组人中患病率实际上是不同的，但差异不大。一个组的患病率也许是 $1/8\ 000$ ，另一个组为 $1/8\ 001$ 。两组是否不同？ $1/8\ 000$ 对 $1/8\ 100$ 又会怎样？我们究竟应在哪画一条线来确定我们观察不同的患病率？

当差异单纯由于偶然性而超出了我们所预期的波动范围时，就要画这条线了。概率上也有偶然的波动，即使掷一枚硬币真的会有平均50%的概率，但是某特定的一组十次投掷也许结果与此不同，尽管出入不大。当概率很小的时候，其

影响也会加剧。疾病额外增加一个病例也许会使概率从 $1/10\ 000$ 加到 $2/10\ 000$ ，绝对值只增加一点点，但相对风险却翻了一倍。并不是说我们应该忽视相对风险，仅仅将绝对风险认为是有意义的，因为现实生活中的流行病学研究显示，少量接触也会极大提高患病概率。甚至抽烟也会将肺癌的绝对死亡率大约提高到 $1/100$ ，与非吸烟者相比，吸烟者患肺癌的风险要高 10 倍。

协和式超音速飞机的坠毁为我们提供了一个稀有事件固有的极端变化的例子。在 2000 年 6 月 25 日前协和式飞机从未出现坠机事件，所以所观察到的此类事件的概率为零，这么说来这种飞机是天空中飞行最安全的。波音 737 系列每百万次的飞行中，致命事故为 0.33 次，其飞行记录已达 3 100 万次。空中客车 320 型飞机每百万次飞行中，致命事故为 0.55 次，飞行记录为 7 300 万次。（注意：飞机安全检测工程师是以每一架次为基础，而不是以乘客所乘坐的里程为基础来估算事故率的，这是因为飞机的整个行程中最危险的部分是起飞和降落阶段，每一架次的航班不管飞行距离多远，都有起飞和降落。此外，从工程师的角度来看，一架飞机的可靠性或机械故障的概率与所载乘客的数量无关，虽然从可靠性角度来说，乘客人数也是非常重要的。）

这次事故后，协和式飞机成了安全记录最差的飞机了。超音速飞机的机种并不多（大约十几种），相对来说飞行的架次较少。截至 2000 年，协和式飞机大约飞行 80 000 架次，

一次坠机事件就将事故概率从 0 提升到 $1/80\,000$ 或者每百万次飞行发生 12.5 次事故。在研究极其稀有事件时，特别是当分母相对来说较小时，比率可能会起伏，因为他们取决于大量的抽样变化。总之，小比率的估算是不太精确的。

泊松分布

有一种分布在概率很低的时候来处理概率波动的问题时非常有用。甚至在基数不确切的时候，我们也可将这种分布用于庞大的人群，我们所要确切知道的是所发生的事件的数目（如病例数或流星击中地球的次数）。然后，可用泊松分布的数学方法来确定这种观察到的数据与其他典型的预期数据是否有显著的差异，是否高于或超出概率波动的范围。这种分布是以法国数学家西蒙·丹尼斯·泊松（1781—1840）来命名的，他是皮埃尔—西蒙·拉普拉斯的学生（后面将会谈及拉普拉斯）。

泊松分布与第一章中讨论过的二项式分布关系密切，使用二项式分布需要特定试验中所发生的事件或击中的概率及试验的次数，从这个数据中可以算出这组试验中各种不同结果的概率。现在我们感兴趣的极为稀有事件还是由用于试验次数（ n ）的概率（ P ）所产生的。泊松分布中的预期值或者平均数 m 与二项式分布相同： $m = n \times p$ 。实际上，如果保持 m 常量，将 p 降低到极小数，大幅增加 n 值，二项式分布

就成了泊松分布。在微积分上，泊松分布是二项式分布的一种特定限制情况： $n \rightarrow \infty$ ， $p \rightarrow 0$ 。

表 4-1 当泊松期望值是 1 时，观察没有“击中”的概率

n	p	$P(1, 0)$
5	0.200 0	0.328
10	0.100 0	0.349
50	0.020 0	0.364
100	0.010 0	0.366
500	0.002 0	0.367 5
1 000	0.001 0	0.367 7
5 000	0.000 2	0.367 8
10 000	0.000 1	0.367 9

在处理稀有事件时一个特殊的问题是，即使在相当大的样本中，这个事件也可能根本观察不到。这时从一组观察中计算出潜在的概率是特别困难的。泊松分布在这方面特别有用，因为计算时既可以基于 m ，也可以不知道具体的 p 和 n 值。表 4-1 显示其中的原因，也表明泊松分布是对二项式分布的限定。

假设 $m = 1$ ， p 和 n 的各种各样组合会导致这个期望值。表 4-1 中显示了一些这样的组合，以及没有看见击中和基数扩大而击中的概率。稀有事件每隔多久在样本中消失？

最后一列中的类目读为“当 $m = 1$ 时，观察到 0 次事件的概率分布”，这个数字就是 $(1 - p)^n$ ，因为我们计算的是在连续 n 次试验中不发生的概率。注意： $P(1, 0)$ 中的变

化起初较快，随后在达到一定限度时，逐渐变慢。这个限度叫做 $P_p(1, 0)$ ：泊松概率。 $P_p(1, 0)$ 等于 e^{-1} 。

为什么 e^{-1} 是 $P_p(1, 0)$ 的极限呢（ e 是自然对数的底数）？在表 4-1 中， $np = 1$ ，所以 $p = 1/n$ 。因此，我们可以将 0 次成功的概率由 $(1-p)^n$ 改写成 $(1-1/n)^n$ 。但是从定义上说，当 $x \rightarrow \infty$ 时， e 是 $(1+1/x)^x$ 的极限，用 $-n$ 替代 x 并相应地改变符号，则等于说 e^{-1} 是 $(1-1/n)^n$ 的极限，也是 $(1-p)^n$ 的极限。

可以使这种观察变得更为普遍，当 m 不等于 1 而是其他期望值时， $P(m, 0)$ 的极限是 e^{-m} 。 m 的值可以通过 p 的任意序列和改变 n 的值而得到改变。例如，假设 m 为 2，表示所有的 n 值都翻了一番，现在用 $-2n$ （不是 $-n$ ）在公式里代替 x ，来表示 e 。如以前一样，通过符号的变化，我们可得到有关 e^{-2} 的极限，更普遍一点就是有关 e^{-m} 的极限。

在第一章中我们见到二项式分布的公式：

$$P(r) = \frac{n!}{r!(n-r)!} p^r q^{(n-r)}$$

公式里的 r 是概率计算中的击中或成功的次数，我们可以调整这个公式，从而很容易得到泊松概率，包括除 0 次击中外的其他结果。我们可以利用这种事实，将小概率用于极大的样本上，也就是说在相当大的试验次数（ n ）中，成功的次数（ r ）是极少的，好比 1 000 000 - 1 仍然是大约 100 万， $n-r$ 仍然大致是 n 。在这种情况下，二项式分布 $p^r q^{(n-r)}$

基本上是 $p^r q^n$ ，特别是当失败的几率 q 接近 1 时。但是我们刚才也演示了 $q^n = (1-p)^n = e^{-m}$ 。因此， $p^r q^{(n-r)} = p^r e^{-m}$ 。

由于 n 和 r 相对的大小相约以及 $m = np$ ，这个公式的分数式部分也变得简化了。泊松分布的二项式就变成了 $P_p(m, r) = m^r e^{-m} / r!$ 。我们将看到，这个公式特别容易使用，尤其更适用于完整的二项式公式中。一个原因是二项式公式需要表示人群规模 n 的阶乘积，而 n 可以是数以百万计的。要获得这些阶乘积可能是相当困难的。

但是，获得大规模的阶乘积的值并非不可能，在斯特林的公式里，当 n 很大时，有一种有效的方法来计算 $n!$ 的近似值。这种方法实际上起初是由亚伯拉罕·棣莫弗尔提出的，他于 1730 年出版了一本关于概率的书《分析杂谈》。詹姆斯·斯特林也在他自己的 1730 年出版的书《微分法或无穷级数的简述》中，用同样的方法但更为方便的公式描述了这种方法。后者的公式在棣莫弗尔的 1738 年版的书中有所提及，并将其归功于同时代的英国同行。有了斯特林的公式（和任何类型的计算器）就可得到 $n!$ 为 $[(2\pi)/(n+1)]^{1/2} e^{-(n+1)} (n+1)^{(n+1)}$ 。近似值非常好，并随着 n 值的增加而改善，当接近无穷大时，错误的比例则接近 $1/(12n)$ 。

在数学上这个无处不在的奇特的 e 值得一提，许多人觉得它有奇妙的作用。我们已见过其在二项式分布与泊松分布及斯特林计算阶乘积的公式之间的重要作用。或许 e 最熟悉的应用涉及钱以复利率（ r ）经过一段时间（ t ）之后的增

长：这段时间之后的总数等于 e^n 。放射衰变以与钱复利增长的相反方式出现，即 e^{-n} 。《伊莱·马奥尔的 e ：一个数字的故事》解释了这个数字在数学这么多领域的应用和历史。

在任何事件上，泊松限制函数是计算稀有事件概率的最简便方法，以下是该应用的一个重要例子，用来检验注射疫苗的影响。假设 100 个人注射预防某种疾病的疫苗，该疾病每年在人群中通常有 1 人会感染该病（比如住集体寝室的人）。因此， $m = 1$ 。在注射了疫苗后的一年里，这 100 人中无人得此疾病（因此， $r = 0$ ）。不注射疫苗，完全凭运气，无人得此疾病的泊松概率是：

$$P_p(1,0) = m^r e^{-m}/r! = 1^0 e^{-1}/0! = e^{-1} = 0.37$$

换句话说，如果你想在 100 人中平均找到一个病例的话，37% 的 100 人样本是没有这种病例的，主要是因为在处理这种小比例的过程中抽样起伏不定造成的。所以，在某一特定年份里没有发现任何病例是毫不奇怪的。现在，假设在一个更大的样本中做试验，比如 1 000 人，预期患病的比例大致相同， m 的值为 10。假设在这一组中也没有观察到任何病例，那么我们就可得到：

$$P_p(10,0) = m^r e^{-m}/r! = 10^0 e^{-10}/0! = 0.000\ 045$$

因此，如果这种疫苗没有效果，那么在 1 000 人中发现

无病例就会是令人惊讶的。很可能由于个人不得此疾病的原因，“1/100”会下降至 0；但是当人群为 1 000 时，其概率为 45/1 000 000，预期 10 个病例没有得这种疾病疫苗的临床试验的样本规模是用这种推理提前设置的，所以这种疫苗的效果将超出抽样起伏范围，因而是可以察觉的。

这个公式中最重要的是要确切知道 n 的值， p 的值并不需要，只要人群数是相对稳定的，只需假设 m 值。其实际好处是在不会随时间发生快速变化的大规模人群中，年复一年的预期起伏变化可以预测，并可区分频率中的真正变化（那些有特定原因的变化）。泊松分布的魔力不仅限于医学上的应用，例如，部队的规模通常每年基本上是稳定的，除非爆发战争。不幸的是，有些士兵即使在和平时期也会死亡。在 19 世纪，普鲁士骑兵偶尔会被马踢死。在 18 世纪末，骑兵分为 16 个不同的兵团，各兵团所有年份的 m 值是 0.7，平均死亡率还不到 1 起。非常罕见，但确实发生过多起死亡。在某一特定年份里，11 个兵团曾发生过 3 起死亡事故，甚至有两次发生过 4 起死亡事故。这些超出的死亡数字不被看成是由于部队缺乏训练而是在战马附近粗心大意所造成的。死亡的数字基本上与泊松分布所预测的频率相同，所以他们与小比例所预期的抽样起伏变化是相吻合的（见图 4-1）。

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	-	2	2	1	-	-	1	1	-	3	-	2	1	-	-	1	-	1	-	1
I	-	-	-	2	-	3	-	2	-	-	-	1	1	1	-	2	-	3	1	-
II	-	-	-	2	-	2	-	-	1	1	-	-	2	1	1	-	-	2	-	-
III	-	-	-	1	1	1	2	-	2	-	-	-	1	-	1	2	1	-	-	-
IV	-	1	-	1	1	1	1	-	-	-	-	1	-	-	-	-	1	1	-	-
V	-	-	-	-	2	1	-	-	1	-	-	1	-	1	1	1	1	1	1	-
VI	-	-	1	-	2	-	-	1	2	-	1	1	3	1	1	1	-	3	-	-
VII	1	-	1	-	-	-	1	-	1	1	-	-	2	-	-	2	1	-	2	-
VIII	1	-	-	-	1	-	-	1	-	-	-	-	1	-	-	-	1	1	-	1
IX	-	-	-	-	-	2	1	1	1	-	2	1	1	-	1	2	-	1	-	-
X	-	-	1	1	-	1	-	2	-	2	-	-	-	-	2	1	3	-	1	1
XI	-	-	-	-	2	4	-	1	3	-	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	-	4	-	1	-	3	2	1	-	2	1	1	-	-
XV	-	1	-	-	-	-	-	1	-	1	-	-	-	-	2	2	-	-	-	-

训练年数	骑兵兵团每年伤亡人数	
	训练中	预测数
0	144	143.1
1	91	92.1
2	32	33.3
3	11	8.9
4	2	2.0
5年以上	-	0.6

图 4-1 按年份和兵团划分的死亡原始数据 (1875—1894 年)。上表显示的是被战马意外踢死的士兵数, 下表显示的是有特定死亡数的军团数目和根据泊松分布推算出来的死亡数 (右边这列)。资料来源: L. von Bortkewitsch. 小数字的规律. 莱比锡 Teubner 出版社, 1898. 24

在第二次世界大战期间, 伦敦的猛烈空袭提供了“泊松分布应用”的一个例子。1946 年 R. D. 克拉克以此为题在

《英国精算学会学报》上发表的一篇文章描述了被德国 V-2 型炸弹轰炸破坏的类型。他说：“在炸弹袭击伦敦期间，经常有人断言炸弹轰炸点是集中的”，而不是任意分布于伦敦大都市上空，似乎有些特定区域是轰炸的特别目标，或者说炸弹似乎由于机械的原因集中在一起掉下来。前一个解释似乎暗示控制飞行线路的出乎预料的精确，而后一种解释好像有点奇怪，假设每一个炸弹都是在其火箭推进器上飞越欧洲的。为了确定是否真的会发生连续轰炸，“伦敦南部的 144 平方千米被划分为 576 个 1/4 平方千米的区域，然后分别计算遭炸弹 1、2、3…次袭击的平方千米数。在考察的时段内，该地区受袭击的炸弹总数为 537。于是，与通过计数而获得的实际数目相一致的预期区域数便可用泊松分布公式计算出来”。表 4-2 显示的是实际结果。

表 4-2 与泊松预期值相比较的伦敦地图标格区域 V-2 型炸弹数

每个区域的炸弹数	预期区域数(泊松)	实际区域数
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
5(以上)	1.57	1
总数	576.00	576

资料来源：R. D 克拉克泊松分布的应用，英国精算学会学报，72卷，1946 (355)：481

最通常的结果是没有炸弹，其次是一枚炸弹，但是也有区域真的不幸遭到大规模的袭击。尽管给人有连串性印象，可是密集轰炸并不比人们预料的更普遍。观察到的遭受 4 枚或 5 枚以上炸弹袭击的区域几乎与泊松分布所预期的数目相等，其他方面稍有差异。

使不可能的事变为必然之事

当人们极力避开一些稀有事件时，如患某些奇怪的疾病、被马踢倒、遭飞机炸弹袭击以及其他一些梦寐以求的稀奇事件，这些事件的泊松分布概率不一定非要覆盖到。很少有像选择在如此巨大的人群中赢取抽奖号码这样稀有事件能够激起如此热切的希望。赢取大奖的机会是很小的，但是有千百万人参与，也许是好玩有趣，或者是出于经济上的绝望或受到欺骗，很少有人相信自己真的会抽中大奖。

但是，1992 年在澳大利亚墨尔本一个叫国际投注基金的投资集团确实希望赢取一个数百万美元的大奖，该集团的组织者注意到弗吉尼亚州的彩票有一种玩法是从 1 ~ 44 数字中抽取 6 个数字，他们注意到抽中 6 个正确数字的机会还算不错。6 个数字的可能组合总数为 $44! / (6! 38!)$ 这个数字等于 7 059 052。只有一个组合是获奖的数字组合，所以任意挑选的 6 个数字组合获奖的概率为 $1/7\ 059\ 052$ 。

这家投资集团决定在弗吉尼亚购买 7 059 052 张彩票，每张彩票均有一个可能的 6 个数字组合。购买一张随机挑选的中奖组合彩票有可能获奖 2 700 万美元，彩票每张面值 1 美元。因此，对于投资 7 059 052 美元（不包括购买 7 000 000 张彩票的管理费——我想这笔费用也是相当可观的）来说，回报有可能超过三倍，这笔钱很快可以得到，但也没有风险。国际投注基金之所以有钱是因为它的资金来自澳大利亚、新西兰、美国和欧洲的 2 500 个小投资者，每个投资者平均投资 3 000 美元，每 3 000 美元可以产出 10 800 美元。

弗吉尼亚的彩票获奖实际上并不是即刻支付的，而是分期付款，20 年付清，在此期间每年还是有 135 万美元的回报，投资 3 000 美元的人在 20 年间可以每年获得 540 美元。

《纽约时报》上报道 2 月 25 日中奖的一篇文章从一名会计师的角度描述了这种结算。注意：在这 20 年间的支付中是没有任何利息的；由于通货膨胀，每次随后的支付款价值要比前一次低；此外，如果你获得头奖的话，这笔资金可以用来赚取 20 年的收入。因此，《纽约时报》认为这是一种很好的投资，即“700 万美元的投资可以获取大约 16% 的回报率收入”，而且通常只有风险最大、最投机的投资才有接近 16% 的收益率。

中奖收益实际上只比此稍高一点，在弗吉尼亚彩票中还有各种不同组合的二等奖、三等奖和四等奖，但是这些对于

头奖的收益是微乎其微的。二等奖是 899 美元（这样的奖有几百个，因为需要选择中奖的号码较少），三等奖为 51 美元（这样的数字组合有成千上万个），四等奖为 1 美元，刚好让你购买下一次中奖的机会。总之，握有各种彩票就可以中得总数为 27 918 561 美元的奖金。

购买每一种彩票组合，收取获奖彩金看起来好像是铁板钉钉的事，但还是有三个风险因素会威胁到这种投资风险的概率。一种是被认为是不现实的担心：在某些情况下彩票主办方可能会不予发奖。第二种潜在的问题由于可能性更大会导致巨大焦虑：其他人碰巧抽到同样的获奖号码，奖金就会在多于一个获奖者中平分。1992 年 2 月 15 日在抽奖进行中时，已经开奖 170 次，其中大多数时候，无人中头奖。事实上这种结果已经出现过 120 次了，头奖将在剩下的 50 次开奖中开出，但是中奖则会在这些开奖次数中由 10 人分享。与另一个人分享获奖，收到 18% 的投资回报，在 1992 年与其他投资相比较来说还是被认为是可以接受的回报率（尽管不能与当时的纽约股票交易相比）。如果售出 3 张中奖号码的彩票，回报率大约是 5%，这种回报率（或者其他由于中奖人多而导致更低的回报率）从投资者角度来说基本上是一种损失，因为 700 万美元如果投资在其他地方，可以获得更高的回报率。

对投资利率造成威胁的第三种因素是也许不能以快速及时的方式完成彩票的购买。这会导致花费大量的钱购买大批

彩票但不能获得足够的奖金。投资集团因而认真做好准备工作，尽最大努力使彩票的购买能够顺利进行。买彩票前，他们按要求填好 140 万张纸条，每一张纸条可以购买 5 注彩票（也就是数字组合）他们有 72 小时购买彩票的时间，各組人马在 8 家杂货连锁店的 125 个店购买彩票。杂货店的雇员不得不轮班以创记录的速度打印彩票。一家杂货店在 48 小时之内，刚好在开奖前售出 75 000 张彩票，一家连锁店为了将其各分店的压力降到最低点，以这样一种方式来销售彩票。连锁店总部接受购买 2 400 万彩票的银行支票，将打印这些彩票的工作分布到各个分店，彩票由专人送取。

结果投资集团的工作人员时间还是不够，在开奖前，他们只能完成购买 500 万张彩票，而不是 700 万张彩票。但是，他们挺幸运，在彩票销售时间行将结束前，他们得到了一张中奖票，避免了平分奖金，从而绕过了两个主要的潜在问题。他们花了几天时间才找到这张彩票，但是最终他们还是将彩票呈交给彩票发行方。

彩票发行方感到左右为难，因为对公众来说靠买取所有的数字组合，而不是靠概率来赢取中奖似乎不公平。此外，公众抱怨说由于投资集团的活动，他们无法买到彩票。彩票发行方举行了听证会，会上一名比萨外送员说：“谁也不愿意排队排在一个在那里待了三四天的人后面。”有报道说机器被贴上“发生故障”的标签，但实际上在没有购买者的情况下，它被用来打印整批购买的大量彩票。最后，官方发表

声明说为了增加中奖机会而购买几百万张彩票并没有任何错误。但是，彩票委员会马上更新规则，给予那些站队的人优先权，涉及每个商店售出的最高彩票数的有关规则也在考虑之中。

彩票发行方确实给投资方设置了真正的威胁，他们停止支付还有另一个原因。他们找到一条法规以取代潜在规则来使中奖彩票失效，这就是一条州法律要求整个交易的发生必须以此为前提，即机器打印彩票。这条法规的目的与销售几百万张彩票本身无关，相反它是用来阻止中间人购买大量的彩票，然后在其他地方出售。有些人也许会从中间人手中以高于 1 美元的价格购买彩票，以换取不必在商店排队等候的便利。这种倒买彩票的“小额利润”是非法的，因为彩票含有调节价，同时也存在潜在伪造彩票的可能性。但是，由于有些彩票是在公司总部付款，而在公司的杂货店内打印的票，州官员起初说他们可能拒绝支付奖金。杂货连锁店说从来不知道有这条规定，投资者律师也证实这一说法。另外，国际投资基金说很难确定其手中的中奖彩票是否在总部付款购买而在其他地方打印的。最后，由于担心一味延长法律纠纷，同时很难确定中奖彩票究竟是在何处购买，弗吉尼亚彩票委员会官员在 3 月 15 日宣布 2 月 15 日的头奖还是发给这家投资集团。

有没有其他办法能够确定或者几乎使你能够确定会中奖呢？研究表明，除了有些人相信数字的“轮回”或者在特定

的彩票中某些被抽中的数字的趋势，这些实际上仅仅是人们所希望的巧合。1996年11月8日《纽约每日新闻》上的一篇文章声称“揭开中彩之诀窍”。文章中充斥着统计数据，没有一句话涉及可以解释这些数据的内在波动变化。例如，文章中提到46这个数字在纽约的彩票中奖中出现过22次，高于第二个常出现的数字（4曾出现过18次）。但是，这个只考虑了一年的有价值的数字，没有提交人们在这一系列时间内以可观察的比例或频率形式来表现所期望的变化计算。该文还说：“有些名字也很有运气，叫玛丽、玛利亚或相似名字的妇女经常有可能中奖。男人中，叫约瑟夫的人多半会中头奖。”这是运气吗？像玛丽、玛利亚这样的名字难道在纽约不是最普通的名字吗？这不正好说明他们中奖的概率正好是她们在普通人口中所占比例的反映吗？难道“幸运”就预示着一个叫玛丽的人中奖的概率就比我们所依据比例的希望更大吗？还有，应该用比较的方法。就我个人来说，如果中奖最多的名字是Esmeralda而不是Mary（玛丽）的话，我认为会更令人惊讶，也更具有特殊幸运的特征。

奇迹

1950年的一个晚上，内布拉斯加的比阿特丽司的一个教堂的唱诗班经历了一次既令人烦恼又不可思议的巧合事件。

令人烦恼的巧合之事是那天晚上每一个人都碰巧彩排迟到。这是非常令人惊奇的，因为唱诗班由 15 个人组成，所有的人并非都是被同一场暴风雪耽误的。有一个人耽误是因为睡过了头，另一个人耽误是因为汽车发动不起来，还有一个人耽误是因为要完成几何家庭作业。所有的这些耽误并非都完全是独立的，因为有的家庭不止有一个唱诗班成员，迟到具有家庭特征。但是还有 10 个家庭的迟到明显可以看成是独立的。

10 个家庭都被耽搁，某一特定约会全都迟到的概率有多大？这是不太好界定的，因为我们对一次迟到的几率没有好的估算办法。在抽彩方面，我们可以知道所抽中的某一特定数字或数字组合的确切概率。在许多医疗应用方面，我们可以依赖大量公布的历史来提供对某特定疗法有效反映的概率估算。这里我们只能猜测，但是假设唱诗班成员有 10% 时间迟到的话，十个独立的人一起迟到发生的概率为 0.1^{10} ，等于十亿分之一。即使迟到比较普遍，但是 10 个人在一个晚上同时迟到的事还是极为罕见的。如果一个家庭迟到的概率是 0.2 的话，10 个家庭同时迟到的概率大约是一千万分之一。

令人烦恼的巧合是每个人都迟到了。但是，这也是令人惊讶的巧合，因为就在大家都迟到的那个晚上，发生了一个怪异的故事。唱诗班原定开始练习的时间过后几分钟，一声爆炸摧毁了教堂。在 1950 年 3 月 27 日《生活》杂志上的一

篇文章里，唱诗班的所有成员都疑惑，这一系列奇怪的迟到是否是天意之作。

很少有教堂被摧毁，特别是在一个特定的夜晚，一连串的非常稀有的情形耽误了所有的唱诗班的成员。对于教堂毁坏，我们可以给多大的概率呢？我们需要用教堂毁坏的概率乘以十个迟到的概率，粗略估算为一千万分之一，或者是十亿分之一，以便得到这些迟到在那一刻发生的联合概率。当然，我们所需要的这个数字是不可能得到的（尽管我认为保险公司也许有估算数据），但是肯定是一个非常小的概率。据说值得注意的是，英格兰 Durham 天主教堂（建于中世纪时期）也是在一场怪异的暴风雨中被毁坏的。毁灭就在 Durham 主教公开质询玛丽是否真的是处女之后不久发生的。这样的事件不是奇迹吗？当然概率肯定是非常小的。

经常有人（甚至统计学家）祈求上帝来解释一系列极度稀罕的事件。约翰·阿巴思诺特是爱尔兰人，内科医生，医学院毕业后发现自己对统计学和数学的兴趣远高于对医学的兴趣。在 17 世纪后半叶，他将荷兰天文学、物理学家克里斯蒂安·惠更斯写的一本关于概率的书翻译成英文，并加上了自己的游戏概率的例子。这本书合作出版的时候，就成了概率领域中第一本用英语写的书。从那以后他不断地出书，教授数学，并于 1704 年成为皇家学会会员。1710 年他在皇家学会的《自然科学会报》上发表了“为上帝辩护”的文章。

阿巴思诺特注意到第一章提及的男婴出生率高于女婴，尽管比例小，但却是始终如一的。他看了一系列载有伦敦 82 年来每年洗礼命名记录的数据，这些数据表明 82 年里每年男婴比女婴出生多的概率不是很大。而且他声称这种情况很可能“从古至今一直发生着，不止在伦敦，而是在世界各地一直发生着”。因此，他得出结论说这种现象的发生概率“将接近一个无穷小数，至少比分数要小，从其源头来看，控制性别比例的不是概率，而是艺术”。

究竟是谁的“艺术”？为什么要以这种方式来控制新生儿的性别比例呢？阿巴思诺特是这么说的：“我们必须看到男性容易遭受的外界事故（他们必须冒险去寻找食物）确实给他们构成巨大的威胁，而这种死亡，外加各种疾病导致的死亡要远高于女性，为了弥补这一缺失，全能的上帝未雨绸缪地让造物主带来了比女婴更多的男婴，而且这个比例几乎是恒定的。”男性在任何年龄段的死亡率高于女性，过去是这样，现在仍然是这样。对性别比率的这种解释是 30 年前牧师兼制图家约翰·彼得·苏密赫发表的，他的文章基于勃兰登堡和普鲁士的统计研究，曾在第一章中有过讨论。

当代科学趋于揭示直接的自然原因（越直接越好），而不是中间原因或终极原因。例如，我们对出生时的性别比率恒定性的解释集中在荷尔蒙是如何调节的，他们又是如何确定婴儿性别的。由于这种还原科学的结果，我们对医学和其他许多问题的掌控有了提高，因为直接改变自然威力的能力

使科学如虎添翼，尤其是当这些变化具有可观察的理想效果时。因此，认为某些不可能的事也许是神秘天意的看法传播得已没有先前广泛了，一连串事件的极度稀缺本身并不能排除仅仅是偶然巧合还是有自然原因。而且，很难想象某件事件的巧合概率为零，纯粹是神的旨意。另一方面，有些非常罕见的偶发事件的发生，我们根本没有机会碰巧观察到。我所指到的观察到的事件是比如像空气中所有分子都向房间一边移动这样不可能的事件。如果我们只有依赖统计推理，没有无可辩驳的证据来说明究竟是自然原因还是上帝旨意，那么这些最稀奇事件的源头必定是无法确定的。

5

等候的游戏



使你发狂

如果你乘公共汽车的话，肯定有时会疑惑为什么在等得不耐烦之后公共汽车才接二连三地到达。它们为什么不能间隔均匀一点呢？火车的间隔差别也很大。若你自己开车的话，也同样会经历突然聚集起来的塞车（甚至在没有交通事故的情况下也这样），而这塞车也会像形成时一样很快消散。数学可以预测这种日常经历，其中的一个分支——排队论就描述了一群乘交通车上下班的乘客这样的现象，并解释了为什么他们的旅途从来不是平稳、均衡的。

尽管乘客有怀疑，但公共汽车成群结队地到站并非是司机的初衷。假设公共汽车以均衡的间隔发车，比如说每隔十分钟，但问题是乘客不是以均衡的间隔开始等车的。在间隔

过程中，公共汽车站聚集的乘客数量变化是任意的。即使他们都想在同一时间准时到达车站，一个人也许睡过头，另一个人也许起得特别早，如此等等。在某个特定时刻，其间隔期内聚集的乘车人数远远超过平均数，因而经过的这辆公共汽车就要载许多人。等所有的乘客都上车、付车费等事情会使公共汽车耽误一点时间，这一耽误缩短了拥挤的公共汽车与随后发出的车辆之间隔的距离。由于已经耽误，司机发现拥挤状况越来越严重，如果司机按时到站的话，下一站等车的人更多，于是拥挤的公共汽车变得越来越慢。与此同时，由于后面的公共汽车要载的乘客少于整个两车间隔期所聚集的乘客，所以开得越来越快，越来越近。随着旅途的继续，乘客量的变化更加明显，公共汽车间的距离和时间变得越来越短。当然，这种趋势也可以逆转。紧跟在后面的乘客量少的公共汽车也许突然间得搭载大量乘客，他们都是在那辆拥挤不堪的公共汽车开走后不久开始等车的，比如说许多人同一时间放学或下班。罗伯·伊斯特威和杰芮米·温德汉姆合著的书《公共汽车为什么三辆一起到站？》描述了上述一系列事件及许多其他排队和成群结队的例子。

无论在城市还是在高速公路上行驶，汽车的交通流量也会“拥挤在一起”。在城市街道上，信号灯绿灯亮时会放行一定数量的汽车。当交通流量适中的时候，绿灯的时间长度通常足以让红灯亮时所积压的汽车通过。其实，交通信号灯的时间通常就是要达到此目的。但是，由于汽车是任意进入

这一系统的，有的时候车辆太多，一次绿灯无法全部通过，那么就会积压相当多的车辆，尤其是当接近红灯时发生排长龙现象。而且，在交通高峰时段，这样的长龙是典型的，即使红绿灯被调整以跟上交通流量，也无济于事。另一个可能的问题是在一个主要的十字路口，如果绿灯长时间阻挡交通流量，也不是随意可以调节红绿灯的。

在没有红绿灯的高速公路上纯粹由任意的交通密度变化造成塞车也许更出乎意外。但是，由于短时间的交通缓慢，大多数司机在全速行驶之后，不得不减速。当你确定不可能开得像你原先想象的那么快时，你就决定顺其自然。可是这时交通变得顺畅，前面的司机恢复到原来的速度，均衡的车距也得以恢复。当你经过先前造成交通堵塞的现场时，根本看不到任何交通事故的迹象或其他导致速度减缓的原因。

交通密度也是引起交通堵塞的主要原因，因为这种现象在交通舒畅时是观察不到的。但是，当路面汽车数量饱和时，拥挤的车距也许会达到司机容忍的极限，这时如果一个司机感到前面的汽车太近，他会轻轻地踩一下刹车，以使汽车保持较为舒适的距离。当前方又有车进入高速公路，或者有人减速躲避小鸟、瓦砾或凹坑时，就会发生这种情况。一旦一个司机踩了刹车，在他后面的司机也会感到距离太近，于是也会减速。这样，连续几辆汽车都会减速，只有当某一司机碰巧觉得不减速仍然有令人满意的“安全距离”时才结束。但是，这种减速也会造成“凸起效应”，于是就会出现

相反情况：第一辆汽车一通过减速起始点（比如出口的路面小凸点）后就加速行驶，后面汽车连续调整速度和车距。对远处的人来看，似乎产生了事故影响，然后这些影响又消失，经过刚才发生塞车现场时，没有见到任何能够提供实质性解释的塞车原因。

有些人希望通过在家工作或家庭计算机终端来避免排队。这可以减少对这些人的折磨，也可以减少对出行车辆的损耗，但你仍然无法摆脱排队理论：像电话和因特网这类设施的线路也需要排队理论管理。

从历史上来看，与电话线路有关的问题首先刺激数学来探讨排队理论。在 1909 年，一名叫 A. K. 爱尔朗的丹麦工程师将这一理论引入统计学文献，他是哥本哈根电话公司的雇员，他要在通话高峰期为避免严重堵塞确定公司所需的电路数字选择（而在那个时代这是由接线员来完成的）。很显然，这里有个平衡的问题：线路的占用是有成本的，因此，一定要确保所有的电话都能够完成的重要性，要权衡由此引发的昂贵的过度负荷的问题。以概率基础来考虑这个问题，在特定的时刻，话务量概率为 0 肯定是不存在的。但是话务量的极限是不可能的。（世界上每一个人真的会在同一时刻向某个特定邻居打电话吗？）不值得花精力和资金去建用来处理最不可能发生的情形的设施。爱尔朗意识到最理想的是用数学模式来计算电话拥塞的不同水平的概率分布，只要知道电话流量的性质和变化即可。打电话的等候时间和排队的长度

也是系统的重要特征，对它们进行量化是重要的，因为它们也是衡量该系统是否充足的指标，也是顾客满意度的重要组成部分。

一些简单排队的分析

调控排队现象的这个公式有着广泛的应用范围。例如，附近公园有一个美丽的湖泊，晴天有游船供出租。这些船只能在一个地方租，即在存放船只的小棚屋内，顾客进入、登记、交押金、领取救生衣和阅读安全规则说明。然后，工作人员帮助顾客将船拖入水中。在没有其他顾客的时候，从顾客进入停船棚屋到他在水上划动所费时间为 5 分钟。用排队论的语言来说，这 5 分钟时段称作服务时间，以 s 表示。服务速度为 $1/s$ ，所以，在这个例子中 $s = 1/5$ ，等于每分钟接待 0.2 个顾客，或者说每小时接待 12 名顾客。

这个基本的服务速度被认为是恒定的，尽管抽样观察不可避免会有些不同的变化。有的接待时间可能会比下一个长一点或短一点，但是它们基本上被一个基本的服务速度支配着，没有连续效应：所有的起伏都是任意的，前一个接待中所发生的事对后一个接待并没有产生影响。因此，我们可以说接待时间都可设定为独立的，同等分布的，其中服务速度 $\mu = 12$ 是指数分布中的一个平均数。

当然，还有另外一种方式估算队列积聚情况和排队所需时间：到达率。这个比率告诉我们，每个单位时间内进来多少个顾客，通常用希腊字母 λ 表示，而 λ 通常被认为跟着泊松分布的。事实上，真实世界有很多情形涉及到达过程，对于这些过程，泊松分布是一种恰当的模式：时间段通常被分成若干间隔，在间隔期内，到达一个顾客是正常的，多到顾客是可能的也是较少见的，而许多顾客到达则是十分罕见的。

假设我们的服务接待时间为 5 分钟，每隔 7 分半钟到一个顾客，也就是说每小时到 8 名顾客。于是，我们就有了 $\mu = 12$ ， $\lambda = 8$ 。那么，停船棚屋的工作人员到底有多忙呢？他可以每小时接待 12 个顾客，但只来了 8 个顾客。因此，他们是 $8/12$ 或 $2/3$ ，或者说 67% 的时间里处于忙碌状态。这就是所谓的使用率 ρ 。通常 $\rho = \lambda/\mu$ 。在这个系统里，平均总的等待时间，包括接待时间是： $W = 1/(\lambda - \mu) = 1/(12 - 8)$ ，因此，平均等待时间是 0.25 小时，即 15 分钟。平均排队等候时间——排除接待的纯等候时间，由下面公式所得： $W_q = \rho W = \rho/(\mu - \lambda) = 0.67/(12 - 8) = 0.1675$ 小时，即大约 10 分钟。这些数字似乎是反直觉的，接待时间比到达的间隔时间要短，那为什么工作人员不能在下一个顾客到达之前完成前一个顾客的接待工作呢？其原因是无法保证新到达的顾客刚好在工作人员完成前一个顾客的接待，正准备接待下一个顾客时到达。这就是为什么平均等待时间肯定比到

达间的平均时段要长的原因：到达的顾客有 0.67 的概率是在工作人员正在帮助他人时到达的。

正如等待时间所表明的那样，尽管到达间隔时间要比服务接待时间长，还是有一个“预备队伍”。在某一特定时刻，系统中顾客平均数为： $L = \lambda / (\mu - \lambda) = 8 / (12 - 8) = 2$ ，这包括这个人的服务要求得到满足。在这种情况下，他后面的人的纯等候排队时间为： $L_q = \rho L = \rho \lambda / (\mu - \lambda) = 1.34$ 。

假设这是一个异常美丽的春天，而且是周末。接待服务速度也许保持不变，但每小时到达的顾客人数现在是 10 人，而不是 8 人。排队和等候时间会怎么样？ μ 仍然等于 12， λ 等于 10， $\rho = 0.833$ 。在某一特定时刻， $L = 10 / (12 - 10)$ ，所以在这一系统中现在有 5 个人，而不是 2 个人。高比例也反映在队列中： $L_q = 0.833L$ 或者说 4.165（即有 4 个人在排队等候，第五个人在接受帮助，他的服务时间还剩 16.5%）。要通过这个系统的等候时间为： $W = 1 / (\mu - \lambda) = 1 / (12 - 10)$ ；平均等候时间翻倍，达到半小时。平均站队等候时间——纯等待时间 W_q ——现在是 0.833（0.5 小时），或者说 0.417 小时，即从前一例子的 10 分钟增加到现在的 25 分钟。

每小时来 12 个顾客会怎样？ λ 和 μ 都等于 12 他们的比率 ρ 等于 1——这些工作人员将在 100% 时间里忙碌着。等待的时间和队列的长度就会是 $12 - 12$ 或者为 0；等待的时间和队伍的长度都是无限长，如果排入这样的队列，你将

永远得不到招待。有时候这是你自己亲自遇到的真实情形！解决问题的办法就是增加人手。事实上，顾客服务至上的行业可用这种模式来提前检验其系统会停止运作的条件，确定适量的人手以避免在不同到达率和服务接待时间的情况下排长队，或整个系统瘫痪。“人手”不必局限于工人，各个地点所需的自动取款机的数量事先可用这个公式所提出的同样的原理和数据来确定。如果 p 显示一个人会在 100% 时间内都很忙的话，则需要 2 个人才能使这条队列动起来；如果 p 是 2 则人手必须增加到 3 人，以此类推。

有时候人手增加到等待时间非常合理，但是顾客仍然不满意，这种情况在杂货店里经常发生。超市里排队也是很烦人的，因为通常队列都是一样长，你还要选择哪个收款员收款速度最快。就我来说，我是经常选错人的。无论站队时我那队情况如何，很少有比其他队先结束的。我看到其他人在别的队里离我很近，但都先于我结束。毫无疑问，你们都有过同样的经历。为什么会发生这种情况？

分析这种情形不需要十分复杂的排队模型，只要概率推理即可。假设有 10 名业务同等熟练的收款员，10 排长度相当的队伍供选择。实际上这 10 排队伍处理的长度是相当的，确切地说预期的收款服务时间是相同的。这反映了这样一个事实：买了一大堆商品的人劝阻其他人不要排此队，其他地方就会聚集几个商品少的顾客，其他收银队伍处理了足够的少额顾客时，这个大户后面又会聚集起人来。这样，在这种

稳定的流动状态中，我们可以认为轮到收款员接待你的时间都是相等的。

那我们为什么经常会作出站错队的决定呢？可以这样来想一想这个问题：即使有 10 个相同的队伍，10 个相同的收款员，但只有一个是最快的。由于一些无规律的不可测因素（如收款员需要更多的零钱、袋子、核对价格或偶尔与某个顾客调一下情），各个队伍之间存在一些变化。还有一些随意的因素，像难以扫描的物品、疲劳及速度放慢。考虑到这些随机因素，我们的目的似乎是要从这 10 个队伍中任意挑选一队成为赢家（最快的一队），从概率上来说十有八九你是不会在这个队伍中的。

更复杂的模式

当然，我们一直在讨论的模式是简化了的，当额外的因素相关的话，就要考虑进来。在租船的例子中，我们想当然地认为顾客的人数是无限的，他们在附近等待的区域也是无限的，可供租的船的数量也是无限的。但是，显然任何一个停船棚屋或者湖泊能够容纳的船只数量是有限的，而且会有一个最佳的船只配置数。在特定时间，湖面如果看上去很拥挤，租船看起来似乎不太理想，顾客到达率就会下降。排队论也经常用来模拟医生诊所的等待时间，也会出现相同的问题。

题。候诊室的规模相对来说总是不够大，非急诊病人也许会看到拥挤的诊所后，决定到其他地方就诊（或者转为某天电话约诊）。复杂的排队模型通常需要综合这些反馈。

同样，租船模型依赖反映分布平均值的参数，但是如果平均数有很多变异的话，这些参数会造成虚假的模型。假如服务时间差异很大——也许一些“常客”在停船棚屋实际上不需要多少时间，而有些新来的顾客会问一些有关可接受的付款方式、一人以上的票价或者一些令人费解的免责规则的问题。当 p 值较高时（或在一天结束时），有一些接待服务时间变化是可以预料的。顾客人数也许比预想的多，或者也许根本不按泊松分布出现。

许多非常复杂的排队模式是不能用数学模式处理的，但是排队论的结果可以为商业提供实际的应用。用“蒙特卡罗法”（以这个著名的赌场而命名），通过计算机模拟就可获得结果。

在蒙特卡罗模拟中，计算机可以生成代表在各种不同情景中的虚拟人物。例如，可能会从泊松分布或其他分布中随机产生一个与观察到的数据相配的数，这可以得到一个虚拟的人到达的时间。从指数分布的数字中随机选择的结果可以提供他的服务时间。用一组新的数字就可生成下一个人的数据。对于一组数以千计或百万计的庞大虚拟人群，这也可以连续处理。在每个新人的数据产生时，计算机将新人的数据输入这个系统，然后对他们进行“服务”，能计算到那个

时间点的时候会有多少人排在队中。计算机在连续加入虚拟人物时，可以不断处理所有的数据，而且可以在数秒钟内处理几百万的虚拟个人和服务时间。

虽然各种分布的方法被用来描述客流和服务类型，但是由于这些类型是从概率分布中随机选择的，模拟的顾客可能会拥塞或稀疏，有时介于两者之间。这种实际发生的模型比不考虑随机变化的公式所提供的模型要好，而且这种模拟能够提供平均拥塞和极度拥挤状况出现的概率计算。此外，在该模型中服务时间也可以改变，以说明条件的变化——在一定数量的顾客接待完之后，服务时间加长（说明服务者可能疲劳），或缩短以表明增加人手之后服务速度加快。当队列超过一定人数时，后来者被阻止排队，以说明等待区域有限；当人数减少时，新来的顾客又可加入。在长队中顾客因讨厌排队而退出的概率可以具体规定，并用于那些排队者中。总之，计算机可以让研究者玩他们开发的所谓玩具模型，以发现真实模拟，进而找到解决现实问题的可行方法。

排队论也有助于公司赚钱。用排队论提出的模型，无论是基于公式还是基于模拟，为了获取最大的利润，最大限度地使顾客满意，公司想方设法尽可能完善顾客服务系统。有几个因素必须考虑，例如，服务者的利用率（ p ）应该在 0 与 1 之间。当 p 值接近 0 时，队伍的长度和等待的时间就短，因为当一个顾客来的时候服务员在忙的概率小；相反，当 p 值接近 1 时，队伍的长度和等待时间显著增加。从商业

的观点来看 接近 0 的 p 值不是很理想的：顾客也许认为服务优良（不用等待），但是提供这样的服务通常成本很高，而产生的回报收益却很低。似乎最理想的是很高的利用率，因为高成本的雇员和自动柜员机不会坐在那里无所事事，但是顾客将接受的是劣质服务，也许会跑到竞争对手那里去，无论如何，他们是不会站在长队里等待的，因而收入也就失去了。

顾客人均服务成本的变化与一列长队的成本是成反比的。假如两队交叉，则组合成本是最低的。这些公司通常都有专职人员研究与他们经营有关的各种分布与参数，以便模拟队列，估算成本。当今，这种分析的一个主要用途是用于估算一个电话客户所需的电话线路数，就像当年爱朗首次提出用数学分析排队一样。

排队中的心理学

正如我在杂货店排队感到恼火所表明的那样，在使用排队论时，仅仅使用纯数学是不够的。对公司来说，排队中的心理学是确定可接受的排队结构的一个重要组成部分。有些站队很少有人放弃，例如，脱掉衣服、穿着睡衣排队等候磁共振扫描的人排队时就几乎不离开队伍。另一方面，等候磁共振扫描时，每多等一分钟都要比等冰激凌甜筒或等候划船

难受，更令人难以接受。

没有任何说明的等待似乎比说明原因的等待显得要长，没有确切等候时间的等待看上去也显得更长、更恼人。在大多数市郊往返的铁路系统中，乘客们认为那些不给任何说明的短暂等候是最烦人的事，即使这些等候对于到达的时间不会产生多少影响。顾客对于这些令人烦心的小事的不满会导致收入的丧失，因为顾客为了避免这种紧张的气氛，就会选择其他的交通工具。

商业管理部门很清楚，无所事事的等候更烦人，因而等候的时间比有点事情可做的等候显得要长。在纽约吃中饭的时候，各银行的排队都特长。为了通过改善等待时间的知觉以获得竞争的优势，曼哈顿储蓄银行曾经在午饭时间尝试过提供轻快的音乐娱乐，结果顾客的满意度提高了。我在纽约经常去的那家拥挤的银行已经安装了电视屏幕，播放最新新闻、天气预报和财经报道，因此它至少在你等待的时候有些有趣的东西可以观看；人们认为等待的时间浪费较少，因为至少你看到了最新消息。同样，许多市郊往返的铁路系统在车站站台上加装了新闻屏幕。在并没有改善等待时间的情况下，顾客的满意度提高了（而且又是收入的额外来源）。

排队心理学的一个最奇怪的例子是休斯敦国际机场成功处理顾客在行李输送带旁长时间等待的抱怨问题。机场管理部门经常接二连三地接到对等候行李的批评。这些抱怨涉及工作日上午高峰期到达的航班，这些航班满载赴商业约会的

乘客（因此，他们更难容忍质量差的服务）。部门经理们决定特别留意这个问题，因为批评之声非常普遍而且尖锐。他们花钱增加行李搬运车的数量，并找专家征求怎样改善服务的建议。采取了这些措施之后，他们想知道顾客的经历是否可接受。他们有两种形式的资料来确定这一点。一种是从离开飞机到拿到行李所花费的平均时间，结果为 8 分钟，这一时间在航空运输业中被认为是可以接受的。另一种是仍不减弱的投诉登记表。

后来专家们采用了一种新的方法，推断出所涉及的实际等待时间并非是问题的根源，而是这一情景中的心理方面的因素导致这一结果。当他们仔细考虑拿着托运行李的顾客的经历时，注意到这 8 分钟的时间段是这样构成的，从飞机走到行李输送带花 1 分钟，接着在行李输送带旁要花 7 分钟。同时，行李输送带位于机舱门与出租车站之间，这样可以使等候托运行李的人看到其他航班的乘客迅速离开。于是，机场方面采取了一个解决办法，该办法并没有改变从机舱门到出租车总共所需的时间，只是改变了等待时间的心理。上午高峰期降落的飞机被引致机场行李提取区对面的门口，现在从飞机舱门走到行李输送带所需时间为 6 分钟，而实际站着等候托运行李的时间减少为 2 分钟，顾客的不满就这样消失了。

但是，有的时候当你在排队等候时看着别人会增加你的满意度，而不是降低满意度。例如，游乐场所设计者特意将等候骑木马的队伍围绕吸引人的木马排列，这样观看那些游

玩的人本身提供了一个迷人的广告。以这样一种方法，很少有人会因没有耐心而放弃排队，因而收入也就增加了。此外，迪斯尼乐园有长长的双列队，迂回进出，使排队的人时刻处于运动之中。因此，即使等候的时间长，人们也高兴，因为他们不必静静地站在那里。

乐园的队列通常不仅仅是围绕游乐项目而列，他们也会在一个地方前后而站，这样对站队的人数心中有数。通常这也是特意安排的，运用了顾客心理的另一个方面。想骑木马的愿望在某种程度上来说与顾客看到有那么多人想玩该项目有关。这种心理作用是非常强烈的，甚至在骑完木马之后依然存在。与在需要排队等候的日子骑木马所带来的乐趣相比，顾客普遍认为在不用等候、不拥挤的日子骑木马所带来的乐趣显然要低。

在上述各种模式中，有一条明确的路径提供了一个计算模块。排队论能够提供有助于我们理解诸如驾车和顾客服务之类现象不断变化的密度所产生的影响的估计，但是计算必须针对特定的队列，比如特定的公交线路和加油站。如果路径没有提前确定又会怎样呢？如果你所选择的路径受不确定因素的影响，这些因素不确定其速度，而只确定其方向的话，走完某一特定的路程需要多少时间呢？已经出现了大量的数学理论来研究这一类“随机漫步”，而且在许多不同的领域，如经济学和气象学中得到应用。这类随机漫步是下一章将要讨论的主题。

6

股票经纪人与气候变化

人迹罕至的路

真是白茫茫的一片！你离开野营地，徒步去峡谷探险，突然发现自己遇上了一场巨大的暴风雪，没有任何路标引路。青灰色的天空映着脚底的雪地，色彩搭配得如此之完美，使你无法看见地平线。各个方向的景色看上去都一样。眼前的雪下得很密，甚至环顾四周都变得越来越困难，狂风使人连站直走路都很困难。但你必须站直，因为现在跌倒就意味着被雪埋葬，死于严寒。因此，你只能继续不停地走，希望有机会遇到围绕大峡谷的墙，因为你知道沿着这些大峡谷的崖壁，每隔一段距离都有可供使用的紧急避难所，这样就能够到达安全的地方。峡谷的崖壁都在 30 英尺以内，但是穿越这 30 英尺的距离并非易事。你不能只往一个方向走，

因为你根本无法知道走的路是否正确，而且狂风的干扰有时使你蹒跚而行，往一个方向走一两步，往另一个方向走半步，然后连续走几步，再往相反的方向走一小段距离。你往正确方向迈出一步的几率与后退一步偏离方向的几率是相同的。你正在走的就是科学家所谓的“随机漫步”。

半小时之后我们将在哪里找到你呢？（我希望你是在庇护所内品茶。）我们不可能绝对肯定，但可以估算出你到达某一特定距离的概率。对于刚出发的人来说，某一个方向和另一个方向同样的好，所以如果我们标上他们在半小时内所能到达方位点的相同概率，则这些点就可形成一个圆圈。如果这个圆圈表明概率大的距离，则半径就小；反之，半径就大。到达偏远区域的可能性很小。

很容易看出两个极端代表最不可能走到的距离。有一种最糟糕的情况：两步从来都不能排成列，你只能在原地蹒跚，反复在原路上折转，因此所走的距离为零。另一个极端是碰巧每一步都是连续不断的，完全成直线往前走。所以所走的距离是最远，相当于所走的步数乘以你的平均步幅的距离。就像投掷硬币一样，第一种情况就是在半小时内投掷的结果是硬币的正面头像与反面交替出现，第二种情况则投掷结果总是正面头像，这两种情形出现的可能性极小，但并非不可能。

最有可能的距离涉及所走的步数和平均的步幅。还有更多的要计算，因为所走的步子不都在同一个方向。为了便于

计算，我们可以把折转看作是勾画一系列三角形、反复穿越先前走过的路。假设你设法在半小时内走了 25 个直线线段，平均每个线段为 6 英尺，但你实际并没有走到 $25 \times 6 = 150$ 英尺的距离，而是 $\sqrt{25} \times 6$ 英尺（即 $5 \times 6 = 30$ 英尺），因为直线距离是切入所有三角形的大斜边。因此勾股定理可以用来确定你走了多远，而这能为我们提供最佳的距离估算：走 150 英尺的非定向性行进距离为 30 英尺。

以下是数学原理：假设你要跨出第 i 步，位移是 x ，你向左（距离 $-d$ ）和向右（距离 d ）迈步的概率是相等的。

通常情况是， $(x_i) = \frac{1}{2}(d) + \frac{1}{2}(-d) = 0$ ，因此你没有往前走。

但是， $(x_i^2) = \frac{1}{2}d^2 + \frac{1}{2}(-d)^2 = d^2$ ，所以在走了 N 步之后，你距你出发地的平均距离为 Nd^2 。

但是，随机行走可以用来分析比暴风雪中行走更多的情形。在自然界和人类行为中有很多现象遵循随机行走的规律。最初的一个应用源自一位 19 世纪的苏格兰牧师、植物学家罗伯特·布朗的观察，他对为了准备显微镜检查而悬浮在溶剂上的孢子的激烈而近似随机的运动作出了评价。这些孢子的运动似乎是被一种在普通显微镜下观察不到的力踢来踢去，一会儿这样，一会儿那样。据报道，水中的微生物也有类似的运动，除了微生物表面上选定的方向外，它们的路径还有随机的成分。这种运动也可以是灰尘中的微粒、烟雾

以及悬浮在空气中的其他微小物质，这其中似乎也有看不见的力在确定这些微粒的路径。

起初对布朗所报告的以及其他相似观察存在过激烈的科学争论，有些人说微粒的运动是无生命物质中的生命力的证据，通常这会通过自发生成产生生物。另有许多人说这种运动仅仅是对流的作用，加热液体的流动（或气体）在使微小物体运动。还有一些人用邻近微粒，无论是孢子、微生物还是灰尘的运动是无关联的这一事实来反驳对流论观点。因此，微小的对流不是其运动的源泉，或者说至少必须在附近有以相同方式运动的几个微粒。热确实与运动有关，因为冷却会降低这些运动的速度，使这些运动变得不那么剧烈，减少运动数量和频率，而加热则刚好会产生相反的效应。最终人们认识到这些运动是原子相互撞击的结果，原子总是处于无规则的运动状态。很显然，温度越高，运动则越剧烈，因为热能使原子更频繁地向四面八方运动，且具有更多的能量。这种运动后来被称为布朗运动，其数学原理和物理原理受到广泛的研究。虽然爱因斯坦最著名的是他的相对论，并且因光电效应的研究工作而获得诺贝尔奖，但是他于1905年写的关于布朗运动的文章则是他的被引用最多的科学文本。事实上，悬浮微粒的布朗运动是支持原子存在的关键性的可观察的证据。

由于这种运动完全是无规则的，任何一个微粒的路径都是不可预测的，徒步行走者在暴风雪中的任何超过单个路径

都可以预测。从无规则行走的数学性质可以预测一个系统的通常行为。我们已看到一个例子，其中无规则行走产生了离中心 30 英尺的距离，但是这种“最佳的估算”是一个平均数。这就意味着如果你非常不幸反复多次重复这种“试验”，每次半个小时旅途结束时，都会在地上留下一个 X ，多个 X_s 便会形成一个最大直径为 30 英尺的圆圈。

试验的次数越多，平均结果的确定性就越大。四五次顶风行走也许会得出与 30 英尺完全不同的距离，四五百次的行走则会显著地降低偏离 30 英尺的概率，四五百万次的试验则会极大地降低这种概率。所以当研究对象是气体或液体中的分子，而不是人的时候，那么这种系统的一般特征是完全可以预测的。在标准大气压和温度状态下，一升气体中有 6.02×10^{23} 个分子，这就是统计行为规则。从原子要素的微观行为来重新获得物理学的宏观规律尝试，产生了物理学的一个分支，即统计力学。尽管统计力学产生于 19 世纪，但在当今仍然生机勃勃。

统计特征规律支配着像混合溶液这样的现象。如果把一种色彩鲜艳的溶液轻轻地倒入一大杯的清水中，不搅拌，无规则的布朗运动会使色彩扩散至整个杯子。可以预料的最最终结果是一种完全均衡的颜色，而我们的经验当然也与这种统计预测相一致。混合之所以发生是因为每秒钟会发生万亿次的原子撞击——数量之多以至于这些撞击方向一致，我们的估算受到阻碍的概率是微不足道的。有色液体的前部以受每

次布朗运动的“冲击”（一个很小的数字）所产生的平均距离控制的速度向前移动，撞击次数（一个非常大的数字）的平方根用徒步行走者所走的距离方式来进行预测。一个非常小的数字乘以一个非常大的数字产生一个中等数字，因此布朗运动结果的溶液混合是一个相对缓慢的过程。（这就是为什么马丁尼酒必须要摇、要搅拌，而不是等其自行扩散。）

偏离统计特征规则的可能性极小，在日常生活中多被忽略，你不可能指望在暴风雪期间看到几百万人徒步行走，而且碰巧以精确的直线蹒跚前行。同样，当一种彩色液体完全扩散于一种清澈液体中时，我们就认为它达到了均衡状态。看到混合的液体自行复原，即杯中的液体分成清澈的和彩色的两部分时，你会感到惊讶。彩色分子的一般分布位置偏离整杯水中的平均分布的可能性很小，因为彩色分子每次受到向下的挤压时一般也会受到附近另一个不同分子向上推的作用而达到平衡。但是如果所有的作用力碰巧在特定时刻都是向上的话，那么彩色分子就会同时由杯底向杯口移动，而杯底就变成清澈的了。举一个可能更加惊人但完全可能的例子，如果你现在坐着的房间里的所有晃动的空气分子排列成一条直线，一直延伸到房间的另一边的话，会发生什么情况呢？简单地考虑一下这种情况，假设房间分成两半——半是你所在的一边，还有另一边。我们忽视分子穿越房间的距离及所需的时间，仅仅以这种方式来考虑：每一个分子可能在你这边，也可能在房间的另一边。我们说你房间这边的

所有分子碰巧晃动进入房间另一边的联合概率是 0.5^x ,其中 x 是分子的数量——几百万亿个分子。

我不了解你的计算器，但在我的计算器上尽管使用科学符号作为压缩的方式显示小数点后的 99 个 0 的数字，但是仍然从 $0.5^{1\,000\,000\,000}$ 到 $0.5^{1\,000\,000\,000}$ 这个数字比 0.5 的几百万亿次方要大得多，而且我们还没有为了需要而排除从另一半房间扩散到你房间的分子。如果将此排除在外，则概率更小。

在高峰期纽约市挤满行人的街道就是一个很好的例子来说明碰撞的原理：与自己在没有多少行人的其他时间内行走相比，大规模的碰撞可以使通行速度减缓。但是多年上下班期间从中央大车站走出来的反复经历与光能粒子从太阳放射出来所经历的遭遇相比就显得微不足道了。太阳的半径大约是 420 000 英里，以光的速度，几秒钟足以使粒子横越太阳，并与其脱离，不管这颗粒子是如何产生的。但是与原子（和亚原子）发生无数的撞击，粒子在撞击中每次直线行进的距离大约只有 1 厘米，然后粒子在其他方向被弹回来。接着发生一种无效的传输方式——“随机漫步”光粒子脱离太阳平均需要几年时间（取决于其起始地点）。然后以直线路径，只需 8 分钟时间便可到达地球。

行人密度高峰期的纽约市街道为我们提供了一个有关这一原理极好的例子，与你在其他时间人少的情况下行走相比，大量的碰撞会极大地减缓行走的时间。

赚钱从这里开始

让我们暂时回到地球来考虑一下大多数人关心的实际问题：赚钱。你不是一颗企图摆脱太阳的光粒子，而是一个试图积累足够的钱而脱离劳动的人，这样你在退休之后就可以过上舒适的生活。也许你正在拼命工作来积累财富，这样可以早一点不用工作，而且又富有（或者你早已做到了这一点）。许多人追求这样的目标，通过研究股市，试图使自己的投资回报超过股市的行情，这样他们的股票增长超过股票“篮”里的平均数，如道·琼斯工业指数或金融时报股票交易指数。每天在主要的报纸上和互联网上你都可以看到表示股市指标近期表现的图表。当然，这些图表每天都不一样，总的表现如图 6-1 所示。媒体上的许多报道主要是这些数据的分析，包括说明近期的趋势以及这些趋势与近期特殊新闻事件或者近期的态度变化，有时也称作“大众心理”的相关性。这些图表也用来研究周期的起伏变化，因为商业周期可能是未来股票价格的非常重要的决定因素。总之，你认为有助于你正确紧盯未来股票价格的东西都有助你赚钱：你可以知道何时买入、何时达到高位、何时抛售。

如果过去的历史有指导意义的话，那么股市确实是一种好的投资。股市的经历也就 100 多年，如果挑选相距 20 年

的两天，你就会发现股市投资（比如用琼斯工业指数来衡量）超过“固定”投资（我指的是那些比较保险，但是利率固定的债券、票据以及其他的金融工具）。股票也超过黄金和商品的价格。

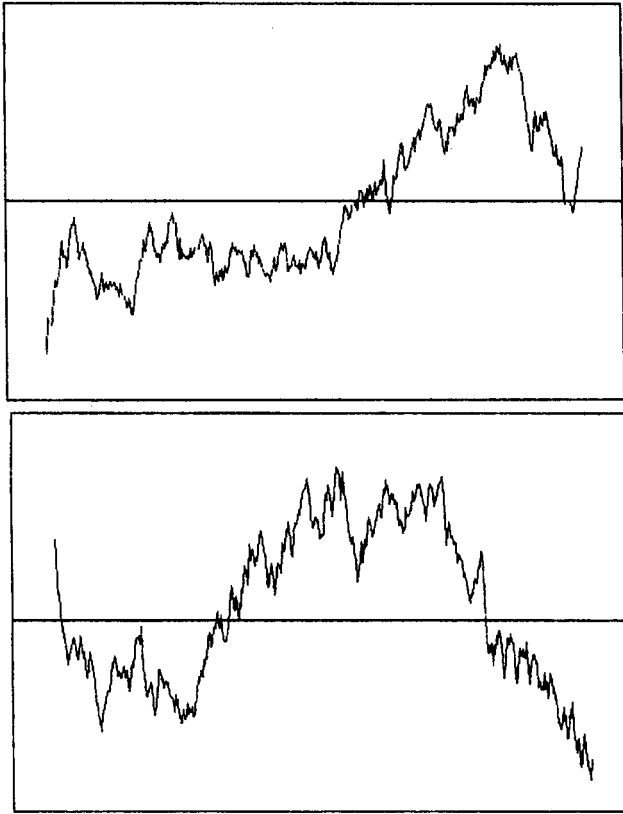


图 6 - 1 历时计算的两种变化类型：是股市价格波动还是随机漫步

当然，不能保证过去的历史有指导意义，毕竟有长期记录的股票交易只是存在于在那个时间段内成为主要的工业化国家之中。随着前所未有的扩张和繁荣，公司的股份也出现了前所未有的增长。然而在所有短期内的股价反弹中，股市交易中的证券价格有普遍上扬的趋势，你也许会说短时间的波动在向上的趋势线附近发生，这些波动集中在这条趋势线上，这一趋势似乎可能会继续下去。

但是大多数人对股市的突出表现感兴趣，即对高于线上的表现感兴趣。由于个人的财政收入在于利润的最大化，所以很少有投资者满足于购买道·琼斯或金融时报股票“篮”内的股票，不闻不问 35 年，然后由于其时代的经济发展，带着其所得的获利退休。相反，人们不停地买进卖出，随时获取、积累微利，而股票经纪人也在那建议你是否交易、何时交易最有利。甚至那些股票投资受限于其顾主提供的养老金计划的人，也有机会来决定什么时候为了利用预期上涨的股票而牺牲有固定收益的债券，以增加他们投资组合中的股市比例。一些大的机构具有法律的和信用的职责来投资几千人，甚至几百万人的养老基金，这样可以使投资的回报最大化。他们也研究股市的趋势和起伏变化，因为在投资上要做到像“强于平均数”这样突出，还是有许多的竞争。

假设道·琼斯指数的轨迹真的是对总的增长趋势线的无规则起伏，甚至在没有得到股票经纪人的有用信息的情况下，随机选择一天投资道·琼斯股票，而在随机选择的另一

天卖出股票的人有 50% 的机会操作得比股市要好，但是也有 50% 的机会正好相反，这也是真的。成千上万这样的人从投资中获利，无论是使用显灵板，还是用工作中的朋友的“内部消息”来作决定，他们都认为自己受益于好的建议。那些接受了相同来源的随意而毫无内容的消息的人不知道什么地方出了差错。如果股票经纪人提出建议（甚至是随意的建议），他们有半数的客户最终会庆幸自己有一位如此有洞察力的经纪人来投资他们的储备金，使他们变得如此富有。

股本共有基金是分析师们通常挑选的股篮，人们认为他们有机会超越股市行情。在年末，他们中的有些人要比其他人干得更出色，而那些干得最出色的人获得广泛的建议，吹嘘自己的获利极具竞争力。问题是甚至在随机挑选的共有基金中，也是有些表现更好，而有些比一般的还要差。你有可能连续几年在某种投资方面比大多数做得出色，而你不能肯定这是否算是一种侥幸，就像投掷硬币时连续投中五个头像那样。但是，大约 20 年前，用斯坦普耳公司评出的 500 家股价指数来衡量的话，股本共有基金所产生的利润普遍低于股市利润。有些失利可能是由于服务费用，但有些失利源自预测股市的基本方法。目前，大约 70% 的共有基金的利润要低于斯坦普耳 500 股票。当交易成本普遍较低的时候，就会出现专门的共有基金的激增，有些共有基金集中于科技行业和互联网。越来越多种类的共有基金的存在，而不是大量的同类基金，使得有些（或者许多）基金更有可能集中于股

市的特定行业，如技术股或制药股，这些股票当时正经历赢利扩张，但是这种趋势并非必然会持续下去。

现在该提一下初看起来令人惊讶的随机漫步的一个特点。如果我们观察一下真正随机漫步的一小部分，我们必然会看到一部分根本不像随机漫步，在这一部分似乎方向性很强，或者也许是循环的，仿佛我们看到一个人在沿着一条特定的路径向前行走。我们也许错误地得出结论：我们不是在观察随机漫步，特别是如果那一部分是供我们观察的唯一数据时，更是如此。

假设有人告诉你图 6-1 代表最近几个月两种特别股票的价格，它们会在特定时间点上看上去成为你的理想投资吗？图 6-1 实际上表示的是硬币投掷中随机漫步的特征，而不是股市价格。趋势线上的每个上升点位是一个“硬币正面”结果，每一个下跌点位是一个“硬币反面”结果。垂轴表明连续的正面超过反面。图中最高点和最低点之间的中位线上，两个结果是相当的。在这一点下方，反面超过正面，而在这一点的上方，则正好相反。记住当上升趋势超越这条线时，并不意味着有“连续”大量的正面头像，而是指往这个方向是累积盈余。即使有许多正反面的混合，彼此间的盈余也可以累积，并持续下去。碰巧在这些序列中有一个反弹，但明朗的趋势是显著的：硬币正面部分的比例在许多地方是明显上升的。但是图 6-1 实际上只是成千上万次硬币投掷的曲线图的小部分摘录，仅仅是局部的趋势。整个曲

线图也许会延长至几英里长，曲线图上的硬币正反面数量几乎是完全一样的，上升和下降的次数也一样。随机漫步确实能产生各种趋势，甚至各种循环，但是趋势和循环有可能被曲解。特别是在短时的趋向资料中，简直不可能区分这些数据是否取自随机漫步，或者它们是否是正在发生的某种有控制的活动或现象的曲线图。

伯顿·马尔基尔的畅销书《漫步华尔街》考察了这样一个问题：是依据其过去的表现，还是分析其商业潜力来挑选个人股票是否有可能让投资者的操盘胜过随意选择股票呢？马尔基尔指出如果深思熟虑地，而不是随意地挑选股票是有益的话，那么在同一种股票的连续价格变化之间必然有一种相关性。注意这种连续的相关性没有必要是完美无缺的，但是它必须要比人们偶然在随机漫步中，如图 6-1 中所描绘的那些结果，所发现的相关性要好。在图 6-1 中仔细挑选恰当的起点可以使你相信投币中的正面头像的百分比在增加，所以，如果一种趋势为真时，那么这种趋势就必须与随机漫步中可能看到的区分开来。马尔基尔观察到连续的价格变化并非是完全独立的，它们确实趋向某一特定方向，而不是巧合，但是这种连续相关性是很微弱的。投资者得出的结论是“买入持有”是最佳的策略：让股市的一般上升趋势增加你所持有的股票价值，忽略特定股票的特殊趋势（不管这些价格变化的原因是什么），因为要把这些变化与偶然性区分开来几乎是不可能的，所以要预测它们的未来价格也基本上

不可能。一般趋势是对你的股票价值的更为可靠的预测因素。

马尔基尔认为“不应该在有效股市中坚持未被利用的交易机会”，或者说至少存在这些机会的证据是微弱的。在这种情况下，有效市场会使股市一有新信息就会在股价上有所反映。股票的价格运动是任意的，因为对股价影响最大的是那些基本上无法预测的外部消息，如战争、暗杀、政治问题和时尚与娱乐的变化等，甚至公司内部的各种事件也会随时成为股价的重要组成部分：在某个时期非常有效的管理方法在另一个时期可能并不恰当；经理也许会调整或及时更换，或者不更换。但是经周密考虑的投资方式能够给你带来什么好处呢？不多。一旦得到相关的信息时，每个人对该信息的影响的判断会通过这一信息对供求的影响而体现在股票价格上。随着判断的再调整，股价逐渐反映人们对这一偶然新闻事件的影响的一般估算。因此，利用信息超越普通投资者是根本不可能获利的，因为每个人都有同样的信息。甚至那些反趋势、不理睬公司灾难传闻、并从自己的乐观个性中获利的人，下一次也不见得会更幸运。因为他的决定是偶然的（并不预示他下次决定的成功），或者他是根据信息而作出决定（而他的下一个决定涉及新的偶然失误、不确定因素和判断错误，每个人都会面临这些同样的情形）。现在也没有办法来分辨好的决定是否偶然因素还是来自好的判断。

看来成功的方法是要获得准确无误的内部消息，但这是非法的，因为这不公平，会扰乱股票市场。只有每个人都拥

有同样的信息，股价的变化才会反映出一只股票的真正价值。所以监管者希望看到有效股市内的非欺诈性的合适价格。例如，在 2000 年 10 月，新的监管条例要求所有的美国公司要同时向各方发布其股票价格潜在收益的信息。收益的公告、收益计划、专利的授予等公告，只有在提交给华尔街分析师之后才能向公众公布。过去你只有到经纪人那里才能得到有关某个公司的全面资料，但这已是过去的事了。在 2000 年 10 月 20 日的《纽约时报》上，美国证券与交易委员会主席亚瑟·李威特（Arthur Levitt）发表了自己的看法：“这种靠眨眼、点头和耳语的体制所造成的挥发度、不确定性，甚至内部交易对我们股市的危害远远大于尊重投资者才智的体制，投资者不需要中间人来向他们诠释金融信息。”

曾经要求获得相同的资讯一直是很困难的。当威灵顿（Wellington）在滑铁卢打败拿破仑时，这一消息由一只信鸽送到内森·罗斯恰尔德（Nathan Rothschild）手中，他是伦敦一名富有的投资者，他设法得到了这条消息。罗斯恰尔德甚至比英国政府还要早知道这一胜利的消息。很显然，一旦大家都知道这一胜利消息，英国的股票价格就会疯涨。罗斯恰尔德是一位投资大师，他有着绝佳的策略，不是立即全部买入股票，然后从即将上升的股价上获利。他跑到股票交易市场，把他所有的股票，不论价格多少，全部抛售。其他投资者看到他以这种方式对一个信息作出反应，都以为威灵顿战败了，于是也纷纷抛售自己的股票。股票价格大跌，只有这

那个时候，罗斯恰尔德才买入他所要的全部股票，结果他从自己故意发动的抛售和他预先预料的胜利消息必然给股市价格带来的令人眼花缭乱的变化中，可以双重获利。

即使没有个人操控股市，还是有许多由“群众心理”所导致的财富激增或骤降的例子。毕竟像装饰用的钻石或某个画家的油画这样的商品之所以值高价，是因为人们对其价值达成了一致的意见（也让其他人相信其价值）：需要某一商品的人越多，这一商品就越值钱。

16世纪中叶，郁金香从土耳其引入荷兰，从而引发了郁金香花球茎的投机市场。人们开始出价购买越来越稀有、越来越时髦的品种，这些品种中只有少数才会结球茎。17世纪初，个别稀有、受人追捧品种的球茎非常昂贵，可以用来交换房子、购买工厂（如酿造厂）、作为新娘的陪嫁。通常郁金香球茎是根本不会转手的，某种奇特的郁金香球茎还长在田地就已被购买。随着对郁金香球茎的需求越来越大，价格只会上涨的这种想法成了一种必然会实现的预言。投资郁金香的回报持续上涨，而且涨幅巨大，在17世纪30年代中期达到顶峰。1637年突然出现投机泡沫，价格上涨马上就会结束的这种感觉突然像传染病一样传开了，成为第二个必然会实现的预言：市场几乎一夜之间将郁金香球茎评价为一种毫无价值的投资。没有发生任何客观的事情，仅仅是有关郁金香球茎的“价值”的简单变化。许多家庭由于这种逆转而破产，或者失去因购买郁金香球茎而抵押的房子。郁金

香的价值消失得如此之快，完全依赖大众所共有的一种幻想，以至于当今的投资专家观察到一旦某一投资的自我加强的有利条件终止，就会爆发所预期的投机泡沫时，他们便发出“又一次郁金香事件”的警告。

短时期内社会力量，比如相信郁金香的市场价格仍会持续下去，可以产生非随机性的价格趋势，即你当然可以利用你对郁金香需求的了解，而不是忽略这个资讯，暂时更加有效地投资。这样做肯定比你随意投资享有更好的回报。问题是当放在一个较长时间的尺度上来看时，投资者在诸如郁金香这样的事情上所犯的糊涂以及随后的清醒，却是另一个任意的、不可预测的光点。你既不能预测到投资者究竟寻求什么新潮的东西，也不能预测这种狂热何时开始、何时停止。

当态度上的变化涉及的不是特殊的投资，而是有关市场可能或应该如何运作的信念时，自身永存的趋势和信念有时也会由于表面上的原因而出现。比如，监管股市公平交易的管理条例上的改变是经广泛协商同意的，而在某种程度上来说是它们任意的规章；但是它们有可能导致非任意的趋势，也许能永久地改变市场。市场如何运作的变化之所以会发生，仅仅是因为人们广泛自愿地接受了新颖的经济学概念和理论。一个例子就是布莱克—休斯（Black-Scholes）的期权定价理论。这一理论开始为一种数学模型，有些人认为它准确地反映了股票期权的价格。起初追随这一理论的人从中赚了钱，所以其他人就跟着做。现在这一理论已被广泛接受，

以至于股票期权定价的所有决定都是依据布莱克-休斯公式而作出的。人们认为这是正确的，因此这是一种作为评估期权价值方法的“强制成真”的模型：因为投资者认为这一理论是正确的，也只有布莱克-休斯价格，才会被认为是公平的期权价格。

让我们澄清一下什么是股票期权。“购买选择权”使你有权购买预先确定价格的某只特定股票。购买选择权是一种预先支付，在特定日期截止的契约权。你不一定非要行使期权，但是假设股票 A 售价为 200 美元，你觉得股票 A 的价格会上涨，因此就买了两个月后价位为 220 美元的股票期权。如果这只股票那时的售价低于 220 美元，则花钱买了个无用的期权；如果这只股票售价为 250 美元，则“锁定”220 美元价格的期权值为 30 美元；如果这只股票值 300 美元，则期权值为 80 美元，依此类推。

布莱克-休斯的期权定价是由下面的公式决定的，这一公式是由迈伦·休斯、罗伯特·默顿和菲西尔·布莱克发明的。这一等式被认为是非常重要的，以至于休斯和默顿因此而获得 1973 年经济学诺贝尔奖（1973 年布莱克已经去世，诺贝尔奖是不颁发给逝世者的）。

$$C = SN(d_1) - Le^{-rt}N(d_1 - \sigma\sqrt{T})$$

什么是购买股票期权的公平价格？很显然这个价格 (C) 必须根据该股票的初始价格 (S)、你以后购买该股票

的认购价格 (L) 以及该期权失效的时间 (T) 来确定。还有另外两个因素起作用, 一个是在同一时期你在某种安全而固定的投资上所赚取的主要利率, 这个需要被包括进来, 以使你在其他地方有潜在收入, 因而放弃购买期权。有时经济学家将这称为“货币的时间价值”, 这种时间价值由复利期限 e^{-rt} 来反映。最后一个因素是股价潜在的升值 (或贬值), 这是不可知的, 但是可以通过衡量股价的挥发度来考虑这一因素。 SD 用来计算这个变化, σ 明确代表来量化的股票回报的连续复利 SD 。

一旦所有这些因素都经过评估, 前面的那个公式就可用来设定期权的价格: 这只是股票销售价减去少量的预售价后的一小部分。 D_1 期限有助于设定这小部分价格。这涉及前面计算过的参数, 可以得到:

$$d_1 = \frac{\ln(S/E) + (r + \sigma^2/2)T}{\sigma \sqrt{T}}$$

最后, $N(d_1)$ 归入布莱克-休斯的公式是因为在提取某个特定的计算出来的 d_1 值时存在随机可变性, d_1 是从正态分布的各种可能观察到的 d_1 总体中的一次抽样观察。 N 表示从正态分布中提取 d_1 的概率 (正态分布已在第二章中有所论述)。

期权定价公式已得到广泛的认可, 因此当每个人给期权定出同样的价格, 这些价格不随意发生变化时, 对股市就会

产生巨大的非随机性的影响。所有投资者以相同的方式来获取利率的最大化的这种统一行为显然背离了带有投资者随机可变性的定价结构。（“放空”的原理也一样，可以让你有权以预先设定的价格卖出股票；也有这类相关的公式。）期权价格现在是可以预测的了，谁也不能因获知此信息而获得好处，它们已经被并入了投资市场运作结构中。谁也不能从垄断股票期权市场或者能够像罗斯恰尔德那样以特有的洞察力进行投资这类情形中获利了。

1991年1月的一个夜晚，在威灵顿的胜利使罗斯恰尔德暴富176年之后，美国和它的盟国部队开始轰炸伊拉克。第二天早晨，正如马尔基尔所叙述的，“有非常清晰的迹象表明我们将很快获得胜利，所以道琼斯工业指数以比前一天收盘价高80点开盘，股市价格的调整是非常快的。”当然，对我来说似乎有了当代电子媒体和24小时在线即时新闻、网上股票交易的出现，股票价格正在变得越来越接近于由所有人都能随意获取的全球实时新闻事件来决定。这种趋势使所有投资者的判断趋于一致，限制了利用信息的机会，甚至使打破平局回报率变得更加困难，但是有一种抵消力。仅仅几年前，交易成本还是大宗、频繁买卖的主要费用，可是这些成本现在降了下来。这就导致交易量高涨，或许还会导致一种新的挥发度，因为额外交易已没有任何障碍（伴随着实时消息的可利用性，有越来越多的因素促使更多的交易）。不规则的事件链数量可增加，可随机产生新的赢家和输家。有

些人认为勤奋地关注即时新闻和股票价格将会使那些更为聪明的个人决定制定者突破股市，这是因为他们时常保持好的判断。但是这类个人行为的结果的有代表性的抽样还不能用来无限地表明随机漫步最终会被投资者避免。

另一个理论在股票价格方面也未经证明。有一套称作混沌论的数学理论确认，对于一组只涉及少数几个输入因子的简单公式而言，有可能产生出完全随机的输出结果。这种混沌过程的结果可满足所有的随机统计检验，但是结果完全是确定性的。因此，知道了公式，一组特定的输入总是与一组特定的输出结果相对应的，从而使预测成为可能。有些人认为，同样不规则的股价起伏遵循某些特定的公式原理，所以股价的起伏变化可以从当前的价格和其他的因素中预测出来，但这种观点还没有得到证实。还没有人能具体说明所需的公式和数据，然后成功地预测出股票价格的未来走向。

一场激烈的争论

股票价格变化的例子表明，区分随机漫步和受特定因素影响的趋势可能是非常困难的。当然这个问题还有许多其他的例子，包括各种各样的具有实际应用意义的例子。对地球气候的变化将会给人类各方面带来麻烦的关注，激发了对全球变暖的研究。因此，许多我们称之为“激烈”争论的东西

聚焦于这样一个问题上，即所观察到的趋势是源自随机漫步，还是源自某些受人类活动影响的潜在过程。

地球近期已变得越来越暖了，这种趋势在过去 150 年，尤其是 1900 年以来的数据上表现得很明显。图 6-2 显示的是单个年份的变化和以五年的移动平均数所显示的相同数据，对于某一特定年份来说，五年移动的平均数就是那年的读数与两边邻近两年的数据平均所得。移动平均数的使用被称为修匀过程，由于某个特定数据点在连续几个平均数（随其他数值）中使用，随着平均焦点向前移动，修匀过程则会减少孤立的起伏变化，让潜在的趋势更为明显地显现出来。

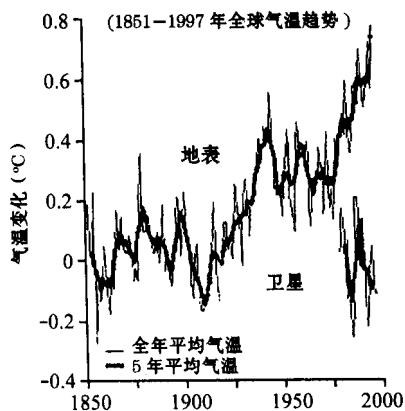


图 6-2 全球气温趋势，度 量与 1851—1870 年（地表）及 1979—1980（卫星）平均气温有关。资料来源：地表资料——菲尔·琼斯，（英国）东英格兰大学和大卫·帕克，英国气象局；卫星资料——约翰·克里斯蒂，亚拉巴马大学和罗伊·斯宾塞，（美国）国家航空和宇宙航行局。

地表温度用以从整个数据串中心的平均值中所得的摄氏温度来表示差异。初看起来，这种差异是微不足道的，因为它们完全包含在大致平均低 0.3°C 至高 0.8°C 范围内。但是，如果整个地球在大约一个世纪左右的时间内变暖 0.5 至 1°C ，地球就会保留大量的热量。这种能量并入全球气候系统会产生大量的风暴，并在其他变化中改变雨水的分布形态，特别是如果这种变暖不均匀，从气候系统中产生新的对流气流。此外，由于图 6-2 列举的是气温平均变化，它模糊了全球各地变暖的极大的不平衡性。根据美国环保总局的网站上的消息，单是 1951—1993 年间，在较为凉爽纬度的北美，地表平均气温上升了 6°C ，而在其他地区也观察到气温下降同样多的温度。

顺便说一下，你可能会看到气象学家要获得这些数字必须使用大量的统计数据。在全世界成千上万个地方，气象站工作人员为了连续性而优先收集了在所有 150 年时间内或大多数时候在运转的气象站的数据。为了得到海洋表面的温度，人们使用在相关时间段内从船上获得的几百万次的观察数据。地球的表面被分成 1 度或更小度数的经纬度栅格。对于特定年份来说，在某一特定方格内的各次观察结果（不管是陆地温度还是海上温度）被平均。然后，在考虑了方格也许会覆盖不同面积的区域这一事实后，将所有方格进行平均，就可得到地球的总气温平均数。

地表气温的数据组合起来，有广泛的基础，因而是可靠

的，对这些数据的信心受到邻近气象站间的一致性支持和陆地与海上测量结果的一致性的支持，也受到当地气温数据与当地在冰川面积、树木生长及其他现象的变化的一致性的支持。

卫星数据可汇编用来研究大气低层的气温。尽管这些数据的趋势与地表温度是一致的，但是迄今为止还仅仅是一个短暂的趋向，起伏变化和趋势还很有限。因此，有关地球大气层变暖的数据还存在较大的争议，而全球地表变暖的数据则是非常明确的。

假设我们接受全球气温正在上升这一观点，为什么全球变暖存在争议呢？一方面，确实存在着上升的趋势，但有可能这仅仅是由“随机漫步”造成的。毕竟我们在图 6-1 看到由投币而产生的曲线图，它们也有在长时间内完全因概率波动引起的连续硬币正面或硬币反面图案，尽管每次投掷的 50% 的潜在概率没有发生变化。同样，我们最初的期望也许是观察的时间序列不会显示趋势，由于硬币正反面的概率是相等的，所以会在“0 线”上下波动。因此，许多科学家一直在不断探索以确定是否可以将全球地表的气温序列与“随机漫步”区分开来。

在 1991 年，A. H. 戈登（Gordon）在《气候学报》上发表了一张显示气候方向累积变化的序列（见图 6-3）。上面三条曲线显示的是基于年数据的全球地表气温趋势和分开的两个半球的地表气温趋势。这些线条上特定年份的向上移

动表明在这些序列里，与前一年相比，气温上升，否则情况相反。从 -10 到 10 的数值范围说明此序列中的上涨年份或下降年份。气温连续几年持续偏高的温和的（但绝不是均衡的）趋势导致气温在（顶上的）总曲线图上持久地处于 0 度线以上的区域，北半球的情形也一样。在南半球，这条曲线在正值区停留一会儿，然后一直是处于负值区。最底下的曲线表示的是一种称作 Keno 随机赌博游戏的结果，这个序列仅仅以单双数来“衡量”。随着气温的变化，尺度上表明当前是单数还是双数占优。

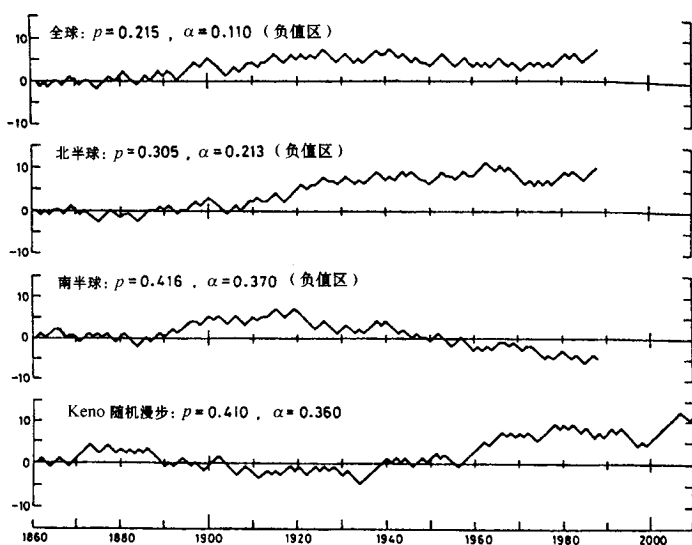


图 6-3 全球和半球温度累积变化序列与“随机漫步”的比较（经美国气象局同意，予以复制）

我们也许会预料在像上述这类曲线上，位于 0 轴线上或线下的时间比例将会是一半。四个序列中，没有一个的长度分配得如此均匀。事实上，“随机漫步”趋于偏离平均分布线，这里 Keno 赌博的“随机漫步”也不例外： α 值为 0.36，这说明 36% 的 Keno 的“随机漫步”是花在曲线一边的（即在线下），这种不平衡的概率（ p ）是相当大的，达到 0.41，因此这个结果并不令人惊讶。南半球的气温数据有类似的 α 值（0.37）和 p 值（0.416），也无法与赌博的概率结果区分开来。北半球所显示的结果也并没有好到哪里去， α 值等于 0.213， p 值表示 30% ~ 31%（将近 1/3）的“随机漫步”的结果将有 21.3% 或者说 1/5 多的时间停留在负值区。

在综合的全球气温数据中可以看出与 Keno 赌博的“随机漫步”的最大差异，其中 11% 的观察点位于负值区。这说明在一年的 89% 的观察中，温度保持在平均度数以上（不管是往平均度数的方向上涨或下降）。但是在一个序列中获得 11/89 的分离的概率是不太自然的：其概率为 0.215，因此 1/5 多的“随机漫步”会有这种始终向上的趋势。

这看上去很惊讶，但我们是用一种特殊的方式在看任意独立的硬币投掷：我们不只是在数硬币的正面和反面，或者向上或向下，而是在特定的某一点上我们得到的硬币正面多还是反面多。如果首先全部是正面，然后全部是反面，则每一面所预期的可能性均为 50%，即以长时间增长，接着以长时间减少的形式出现。当然，这是一个极端的例子。但是一

一旦出现一段稀有的正值趋势，则有可能会保持很长时间，直到相应发生同样稀有的负值趋势。因此，盈余（一旦开始）就会在“随机漫步”中保持下去。

对‘维纳过程’的简短、直率评价

你也许注意到，温度的起伏变化是连续性的，而戈登则将事情简化为独立的步骤，要么是上升，要么是下降。其数学原理类似于二项式数列，提供了一种符合我们目的的解释：或者有一种上升的趋势，或者没有，不考虑大小。这个分析恰当地包括了“随机漫步”，因为“随机漫步”是有步骤的。有一个称为“维纳过程”的类似数学公式，在概念上相当于“随机漫步”，但是是用来检查连续的起伏变化的，是根据研究像主要依据随机事件的布朗运动此类过程的美国数学家诺伯特·维纳的名字命名的。他在第二次世界大战中非常活跃，将自己的才能用于优化防空炮位的布置、从雷达信号中过滤噪音、发明机器编码。有的时候是用“维纳过程”，而不是“随机漫步”来模拟全球变暖，但相对来说这是极为罕见的；就气候变化来说，每年的变化幅度是很小的，最令人感兴趣的一个问题实际上是个非常简单的问题：地球是否正在变暖？

死海和热带植物

不管怎样，戈登下结论说全球的气温序列不可能与“随机漫步”区分开来，4年后韦恩·A. 伍德沃在《气象学报》上发表了一篇重要的补充文章，来检验这些数据，他们检验了全球气温的几种可能模式，有一种模式具有通过点呈现的线性趋势：

$$t \text{ 时间的气温} = \alpha + \beta t + z_t$$

其中 α 和 β 是恒量， α 是从线上截取的，而 β 则是斜率，表明单位时间内趋势的倾斜度，选项 z_t 允许曲线周围的随机趋势增加或减少。另一种模式类似，只是增加了一个二次选项，以使线条呈曲线而不是直线：

$$t \text{ 时间的气温} = \alpha + \beta_1 t + \beta_2 t^2 + z_t$$

最后，他们检验了一种自回归模型的结果，其中对所使用的线条，无论是直线还是曲线，都没有预先设定的公式，相反，所使用的是一种移动的平均数。每个连续数据点均可从移动的平均数加上少许的趋势值就可预测出来，趋势值是随机挑选的。如果最后一个变化是上升的，则上升；而如果情况相反，则下降（就好比我们的测量，而不是仅仅投币）。

由第三种模型预测的全球气温时间序列与实际数值最相符。

这一模式的含义是全球气温序列有随机趋势，而不是由一个确定性等式描述的某种万能的固定程序的结果。气温序列更具有开始是随机性的，随后趋于固定的数据特征。每年它们在轴线上上下反弹比预先确定的（增加或减少一点干扰）变化要多，但是反弹的方向则主要取决于紧靠前面的趋势。因此伍德沃和格雷评论说，当前所观察到的趋势也许在未来会减少，他们还补充说：“这里所显示的结果也说明，只要有 30 多年或 40 多年的可靠气温数据（或类似的资料），就可以得出一个较为确定的结论，即能否预报这种趋势会继续下去。”那些认为由人类活动所产生的“温室气体（二氧化碳）”正在导致变热，确信这种趋势将会继续的人，感到所观察到的全球变暖现象不久就会显现出来。

进一步探究这个问题是很重要的，因为关于全球变暖还有许多问题没有弄清楚。我们真的认为图 6-3 中所描绘的变化是一种显著的变化吗？在白垩纪时期，地球是非常暖和的，我们称为热带植物的东西到处长得非常茂盛。冰川期来了，其中的原因还没有完全弄明白。大陆漂移、火山灰或其他灰尘的盛行、各种各样的云层和冰层都是其中的因素。在地球轨迹的形状和对气候有影响的地球倾斜度上也有变化。例如，在地球轨迹的椭圆形和地球倾斜度上的周期性变化也会影响所接受到的太阳能。事实上，在五六百万年前，地中海还是一片沙漠，这一点可以通过化石以及那时形成的而现

在还在海底的矿石（石膏和石盐晶体）得到证实，通过直布罗陀海峡来的大西洋洋流受到地质变化过程的阻碍，水分蒸发率超过来自降雨和河流的水分补充。引起如此高水分蒸发率的一个原因是那时地球围绕太阳运行轨迹的倾斜度，这产生了异常高的热能。1999年 Wout Krijgsman 在《科学》杂志上发表的一篇文章解释说，“较干旱的地中海气候”原因是在 50 万年的时间里，星罗棋布的水潭比死海更苦涩，这些水潭退出了孤立的、地表富含盐分的地中海盆地”。

在过去 50 年间，这种趋势也可看见，有一半是“人造的”——这是否因为自然的波动，比我们所愿看到的要多呢？也许我们最好将近期的经历和几个世纪前的经历比较一下，以便查清人类活动对这一相对稳定的背景的影响。当然，这个时期的数据无法从各地气象站获得，但是可以从花粉化石所证实的植物种类和明显的树木年轮的生长速度中推理出来。还有许多其他渠道可以获得信息。深海沉淀物中有主要由碳酸钙（ CaCO_3 ）组成的贝壳。这种化合物中的氧气可能由几种同位素组成，海水有一个易受气温依赖而蒸发的影响过程。当气候寒冷时，氧气中的重同位素更普遍，海水中的重元素被蒸发得就少。因此海水过去的温度可以从取自海洋底部的贝壳核心部位的同位素比例来确定。冰川深层中发现的冰块，其核心部位的气泡中的氧气里类似的同位素比例也可用来估算过去历史或史前的气温。与人类活动相关的甲烷、二氧化碳和其他气体——有时称作“温室气体”——

的不断增加，在冰川核心内也有所证实。

所有这些数据来源可以提供收敛性证据，强化了这样的结论，即在过去几千年间，在北半球，迄今为止 20 世纪是最热的。确实过去 1 000 年中，四个最热的年份分别是 1990、1995、1997 和 1998 年。但是，有人注意到在中世纪曾经有过“小冰期”，他们认为这种变化仅仅是正常的复原循环的一部分。上升趋势的速度及最终的大小很快就会对这个问题给出更具权威性的答案，即全球变暖究竟是由温室气体这种人类对地球的污染造成的，还是这仅仅是无论如何都会发生的自然循环的一部分。

其他的争议主要围绕如果这种趋势持续的话，全球变暖会带来什么影响的问题。格陵兰岛、南极洲和其他地方的雪会融化吗？这是否会导致海平面上升？新的海洋对流形态会改变海洋生命，影响食物供应吗？大气层的新形态会改变降雨量吗？这会对不同地区的农业生产产生有利的还是不利的影晌呢？地球的气象系统是非常复杂的，还不清楚需要什么样的参数来作预报，更不确定的是，所有这些参数的实际值是多少。毫无疑问，会有使趋势更小化或使其更明显化的反馈环。了解地球气候的原理、影响气候的因素、气候的趋势走向这类问题使我们想起第一章曾讨论过的传染病预报过程中所遇到的那些问题。

完美的模型

曾经人们坚信带有计算精确的参数的公式可以用来预测所有的自然系统。那是在启蒙运动时期，欧洲社会受过教育的人对许多自然现象真的可以通过数学函数来预测这一发现兴奋不已。17 世纪末 18 世纪初见证了理解和数学预测的成功，如牛顿的《自然哲学的数学原理》出版、诸如行星引力的倒数规律这类关系的解释、行星轨迹的预测及彗星轨道的预测。这种思维方式所取得的非凡成功持续了一个多世纪，直接导致法国数学家皮埃尔-西蒙·拉普拉斯在 19 世纪初得出这样的观察结果：“在特定时刻，能够理解赋予自然生机的各种力量和人类各自环境的智力，如果能提供足够的数据进行分析，就可在同一个公式内包含宇宙最大天体的运动和最轻的原子运动。对于这样的智力来说，任何东西都是确定的，未来和过去一样，都会展现在其眼前。”

这是一个真诚的说法，认为公式控制着所有的自然事件，而且这是一个有趣的推测，认为可以使用公式来从过去的事件中预测所有将来事件的进程，并可反推出当前事件的起源。但是，这种全知肯定总是远离我们的经验：为了作出预测，你需要每一个相关的公式，并要对每个所需的参数完全了解。现在的观点正视拉普拉斯的想象。你不可能了解为

了完美预测像地球气候这样的自然系统的行为所需的一切因素。例如，为了得到完美预测所需的有关气温的必要数据，我们必须将温度计覆盖地球表面和整个大气层。当然这是不可能的，而且也会改变地球的反射率，进而改变地球的气候，所以我们根本不可能完全了解在图 6 - 3 中任何一年中的地球温度。我们甚至对最完美的模型也提出统计上的不确定性，而这种不确定性在我们模拟气候时所需的许多参数中都存在。此外，我们如何提交如此巨大的信息以供分析？有人估算过，一台足够精确模拟地球温度的计算机必须要非常大，需要宇宙中所有物质才能适应它。毕竟现在电脑需要 24 小时才能精确地预测出 36 小时的天气，这些预测也不可能是完全没有错误的。虽然需要一台宇宙型电脑这话有点过头（尤其是如果量子计算处理成为可能）但是关键是拉普拉斯的梦想原则上似乎是不可能的。

罗伯特·梅先生是一位杰出的科学家、牛津大学教授，曾担任过英国首相的主要科学顾问。他起初是在悉尼受教育，学的是物理学，现在专门研究数学模型在生态学和流行病学上的应用。但是梅（他是我以前的导师，他的名字本身指的就是不确定性）曾经说，统计模型必须“模仿现实”：它是所发生事情的再现，是完全可以辨认的，但是已经过于简化，只局限于几个关键的（如果被扭曲的话）特征。因此，模型是人们作出判断的一个重要的辅助手段，从某种程度上来说，我们的未来取决于统计模型在各方面的明智应

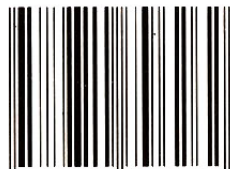
用，从药物的研制到气候学。对概率学和概率在生活中所起的作用的了解是决定行动方案的重要因素。如果我们在其中加入智慧和判断力的应用，那么也许会有一个更好的未来，或者说，至少使其变得更加成为可能。

WHAT ARE
THE CHANCES?



机会的概率 有多大

ISBN 7-81079-626-7



9 787810 796262 >

责任编辑：申作宏
黄圣英
责任校对：周玉宏
封面设计：山内

ISBN 7-81079-626-7
0·36 定价：18.00元