



改变世界

的134个概率统计故事

[日] 岩泽宏和 著 戴华晶 译

CBS | K 湖南科学技术出版社

图书在版编目 (CIP) 数据

改变世界的 134 个概率统计故事 / (日) 岩泽宏和著 ; 戴华晶译.

— 长沙: 湖南科学技术出版社, 2016. 4

ISBN 978-7-5357-8840-5

I. ①改… II. ①岩… ②戴… III. ①概率统计—通俗读物 IV. ①O21-49

中国版本图书馆 CIP 数据核字 (2016) 第 012110 号

SEKAI WO KAETA KAKURITSU TO TOKEI NO KARAKURI 134-WA
BY HIROKAZU IWASAWA

Copyright © 2014 HIROKAZU IWASAWA

Original Japanese edition published by SB Creative Corp.

All rights reserved

Chinese (in simplified character only) translation copyright © 2016 by Hunan Science and Technology Press Co., LTD

Chinese (in simplified character only) translation rights arranged with SB Creative Corp, Tokyo through Bar don-Chinese Media Agency, Taipei.

湖南科学技术出版社通过博达著作版权代理有限公司与 SB Creative 株式会社获得本书的中文简体版中国大陆地区独家发行权。

版权登记号: 18-2015-066

版权所有, 侵权必究

改变世界的 134 个概率统计故事

著 者: [日] 岩泽宏和

译 者: 戴华晶

责任编辑: 杨 林 席小泉

出版发行: 湖南科学技术出版社

社 址: 长沙市湘雅路 276 号

<http://www.hnstp.com>

湖南科学技术出版社天猫旗舰店网址:

<http://hnkjcs.tmall.com>

邮购联系: 本社直销科 0731-84375808

印 刷: 湖南长沙科伦彩印文化用品有限公司

(印装质量问题请直接与本厂联系)

厂 址: 长沙市雨花区石马路 60 号

邮 编: 410007

出版日期: 2016 年 4 月第 1 版第 1 次

开 本: 850mm×1168mm 1/32

印 张: 9

字 数: 220000

书 号: ISBN 978-7-5357-8840-5

定 价: 35.00 元

(版权所有 · 翻印必究)

前言

概率论起源于 17 世纪中叶，帕斯卡发现（或者说发明）了一种前无古人的概率性的思考方式，并以此为开端改变了世界。现如今，我们的生活已经与概率密不可分，几乎可以说是难以想象离开了概率的生活。另一方面，在数学中，这一起源于分析赌博的理论并非主流派系，相较于其他领域，概率论显得颇具个性。

1900 年以后突飞猛进的统计学也让这个世界为之一变。哲学家耶安·哈金指出，卡尔·皮尔逊发现卡方验证之后的发展中，统计学是 1900 年后人类的二十大发明之一。到了 21 世纪的现在，正如小说家赫伯特·乔治·威尔斯在 1903 年所预言的那样：“统计式的思考将会和读写能力一样，成为优秀社会人士的必备技能。”在高中，统计学以“数学”的概念教育着每一代人，但正如印度伟大的统计学家马哈拉诺比斯曾经强调的一样：“统计学不是数学中的一个领域。”这样的说法才更准确，因为统计学的确是一门颇有特色的学问。

本书以具备以上特征的概率论和统计学为主体，精选了 134 则有关概率与统计的“趣闻逸事”，分为五章汇集成册。

大家在阅读教科书和参考书一类的书籍时，有时会发现题为“咖啡时间”或“茶歇时间”，有时甚至直白地写作“趣闻逸事”的文章，内容则是一些稍微偏离了主题，但却非常有趣的小短文。在苦读着教科书和参考书的正文，感到有些疲倦的时候，相信这

些趣闻逸事栏目能让您轻松地进行阅读。而在实际阅读之后，有时还能提高学习书本内容的积极性。一些正文中没能理解的严谨议题，也能在阅读后深入浅出地得到理解，时不时地还能学到一些杂学知识，实在是非常有益。

这本书不是教科书，所以读者们不必把从头到尾的每一个字都理解透彻。即使中间跳过几段，也不会出现无法理解后文的情况。阅读完毕后也没有小测试。大家只需阅读自己觉得易于理解的部分即可。比方说，您可以选择只阅读趣闻和虚构的部分用以娱乐。即便这样，阅读之后您也或许能够理解到一些概率和统计的关键点。

笔者在筛选这些内容时也有两个标准：其一是多选些想让准备学习概率论和统计学的读者感到惊讶的内容；其二则是多选些让本来就有着充足知识储备的读者也能感兴趣的内容。知识的储备量和兴趣的范围因人而异，因此没必要一字一句全部读完。

话虽如此，但也可能会有部分认真严谨的读者想要尽可能全部读完这本书，为此，这本书中会夹杂着一些诸如“本文专为对 $\times \times$ 有兴趣的读者准备”的小标题，或者会将一些难懂的部分作为正文后的补充内容，给大家做好可以跳过的信号。即使没有这些信号，对于难懂的部分（比如不擅长公式的人碰上了难懂的公式时）也可以轻快地略过，只要阅读简单易懂的部分即可。与其在自己不擅长的部分花上大量的时间，不如跳过这些难懂的部分，这样更加易于掌控阅读的速度，有时反而会促进您的理解。

前面已说了很多，总之，本书中汇集了许多其他读物中难以学到的知识，包括尚未被一般科普读物登载的故事，也有一些已经广为人知但经过笔者分析提炼的内容，还有一些深奥的内容则来源于一些大家平时不太会去看的晦涩的专业书籍。希望读者能够以轻松的心态进行阅读，从中能多获得一些新的知识。

〔日〕岩泽宏和

目录

第1章 赌博也要具备几何学的精神

—— 概率论的起源	001
001 意大利面的圈	001
002 天气预报与概率论	002
003 概率论诞生的年份	003
004 “概率”这个词汇	003
005 赌场必胜法	004
006 先驱者卡尔达诺	009
007 卡尔达诺的未解之谜——分配问题	011
008 伽利略的骰子问题	012
009 德·梅尔——创造契机的男人	013
010 分配问题的解决	016
011 帕斯卡的天才之处	018
012 费马的魔法——Dead Rubber 论法	021
013 300多年来的未解之谜	022
014 可怕的赌徒德·梅尔	024
015 概率论的专业术语	026
016 事件是什么	027
017 轮盘的倾斜	029
018 事件的分割	033
019 希腊文字	036
020 吐德哈特《概率论史》	039
021 惠更斯的活跃	041
022 赌徒破产问题	043
023 惠更斯的期待值	047
024 骰子赌博(chuck-a-luck)	048

025	期待值的计算方法	049
026	期待值的加法性	050
027	意大利面的圈的答案	051
028	统计学的开始	053
029	英国的政治数学	054
030	始于荷兰的保险数学	056
031	荷兰全盛期	058

第2章 始祖诞生之前

—— 古典概率论的完成		061
032	概率论的不幸	061
033	“神奇的一年”	062
034	牛顿与概率的交集	064
035	二项式定理	067
036	莱布尼茨的失败	070
037	古典概率论中兴的鼻祖们	072
038	雅各布·伯努利的《猜度术》	073
039	伯努利试验 二项分布	075
040	概率分布是什么	076
041	弱大数定律	080
042	天才棣莫弗的苦难	083
043	棣莫弗的诡计	086
044	诡计的后续	089
045	棣莫弗的《偶然论》	091
046	独立	092
047	52张对52张	093
048	正态分布的发现	095
049	正态分布的公式	098
050	平均、方差、标准偏差	099

051	对数	104
052	纳皮尔本身的对数表	110
053	斯特灵公式	111
054	“概率”这个术语	113
055	学号与身高的顺序	115
056	贵族蒙特莫特	117
057	treize	119
058	欧拉与概率论	122
059	法国革命时期的数学家们	126
060	古典概率论的完成者拉普拉斯	127
061	拉普拉斯《概率的解析理论》	130
062	母函数的理论	131
063	母函数在我们身边的实践案例——西克曼·戴斯	135
064	典型的使用母函数的例子	138
065	特征函数的使用方法	140

第3章 看穿面包店的小伎俩

—— 正态分布的时代	141
066 正态分布的不均性	141
067 名为“高斯分布”	142
068 斯蒂格勒定律	142
069 三大数学家	145
070 数学界的王子	146
071 出生年的记法	147
072 24岁的高斯	148
073 “少而精”	148
074 作为误差分布的正态分布	150
075 中心极限定理	151
076 高斯积分与 π	153

077	最早成就了高斯积分的是谁	156
078	高斯与概率论	159
079	高斯-库兹明分布	161
080	庞加莱的趣闻	162
081	阿道夫·凯特勒的真实故事	163
082	统计学的鼻祖——凯特勒	164
083	凯特勒指数——BMI	166
084	麦克斯韦分布	167
085	围着正态分布转的高尔顿	168
086	母群体这个词	170
087	相关和回归	171
088	秩相关系数	175

第4章 历史的下午茶

——创建了数理统计学的人们	179
---------------------	-----

089	倾斜的分布与卡尔·皮尔逊	179
090	卡尔·皮尔逊年谱	183
091	数理统计学的先驱——提勒	184
092	说到提勒	187
093	埃奇沃思	188
094	累积量	190
095	累积量和中心极限定理	193
096	推断统计学	194
097	战后日本的复苏和推断统计学	197
098	硝烟不断的20世纪统计学史	198
099	笔名	198
100	学生的t分布	200
101	样本分布论	203
102	推断统计之父——费雪	204

103	最著名的实验	207
104	随机数的书	211
105	制作随机数	212
106	奈曼-皮尔逊派的检验理论	214
107	置信区间	218
108	点估计的理论	219
109	最大似然法	220
110	最大似然法诞生的那一年	221
111	点估计量的性质	223
112	数据的终结	227
113	克拉梅尔-拉奥不等式	229
114	哈拉尔德·克拉梅尔	230

第5章 哪个模型都不对

——电脑时代的统计学	233
------------------	-----

115	约翰·图基	233
116	图基时间	235
117	快速傅里叶变换	236
118	探索性的数据分析	238
119	稳健统计	240
120	非参数统计	241
121	Jackknife法(刀切法)	242
122	Bootstrap法(自助法)	244
123	艾弗龙的骰子	248
124	贝叶斯统计学前篇	249
125	精算师与贝叶斯统计学	252
126	贝叶斯统计学与电脑	255
127	模型的正确	256

128	赤池信息量准则(AIC)	257
129	交叉检验法	258
130	广义线性模型	259
131	广义线性模型和统计工具	262
132	每个班级的事故率和广义线性模型	263
133	活着的传奇——拉奥	266
134	一切的判断都是统计学	268

活跃于本书中的主要数学家、统计学家(按出生年份排序)	269
----------------------------	-----

参考文献	271
------	-----

第1章

赌博也要具备几何学的精神 ——概率论的起源

001 意大利面的圈

在大众餐厅的一张桌子上坐着 A 和 B 两个人，两人正在等待自己的菜上桌。现在，B 的面前已经来了一盘意大利面。

A：你的意大利面，看上去很好吃啊。不知道有几根哦。

B：为什么你会关心根数？

A：假设这里面有 50 根吧！

B：喂喂，凭什么这么假设啊？

A：假设有 50 根，那面的两端就有 100 个，随机从中抽选两个端点连接起来。

B：怎么连接？系起来吗？

A：细节不用在意。把所有的端点都连起来之后就算结束。请问，你觉得能够连成圈的意大利面一共有几根？

B：这是什么问题啊？但是，嗯，要是运气非常好的话，50 根面中会有 10 根能连成圈吧？

A：嘿嘿嘿，要使用概率的平均值来说的话，只有 3 根多一点的可能性哦。

B：……等等。刚才我粗略地数了数这个盘子里的意大利面大概有

多少根，估计应该有 100 根。那也就是说，我随便猜的，能够连成圈的有 10 根左右虽然不标准，但相差也不是太远吧？

A：嘿嘿嘿，就算是 100 根，平均值也只有 3 根多一点哦。

A 所说的是实话。是不是有不少读者觉得这个数字太小，因而感到有些惊讶？概率论中有许多这样让人感到意外的事实。本章中，接下来将会向大家介绍许多有关概率论的起源的故事。上面的对话中出现的意大利面圈个数的平均值（概率论中一般称之为“期待值”）可以用一个巧妙的公式计算得出，具体的方法请看 027（p.51）中的介绍。另外，B 的直觉的确相差甚远，这也将 050（p.99）的篇尾提到。

002 天气预报与概率论

讨论到概率时，天气预报就是一个很好的象征。

包括诸多讨厌数学的人在内，人们每天都对“降雨概率”表现出了极大的关注。概率，作为一个数学上的概念——其实还是个非常高端的概念——极为罕见地、非常贴近我们的生活。当然，这时的“概率”是否有被准确地理解还是一个极大的未知数。然而，人们的确在根据“概率”提供的信息决定今天是否带伞，概率也确实实地发挥了有效的作用，影响了人们的行动，每个人至少都粗略地理解了概率大致是什么。

还有一点。天气预报中会多次提到一位概率论伟人的名字——帕斯卡因为气压的单位是“百帕（hectopascal）”。“Hecto”是 100 倍的意思，而“pascal”则是取自布莱士·帕斯卡（Blaise Pascal）（1623 ~ 1662）的名字。正是因为帕斯卡在研究压力的领域中取得了不朽的成果，他的名字才被作为气压的单位使用至今。帕斯卡和皮埃尔·

德·费马（1607 ~ 1665）一样，都是数学上概率论的创始人。

003 概率论诞生的年份

1654年，帕斯卡和费马有过一连串的信件往来。在这些信件的往来中，一种此前史上从未被人解开的问题得到了正确的解答方式——那是一个非常具有历史性的成果。那个被解开的问题用今天的话来讲，就叫作“概率的问题”。因此，他们之间的这些往来的信件也可谓是开创了近代概率论、数理概率论以及古典概率论。

当时，帕斯卡和费马是欧洲大陆最优秀、最著名的两大数学家。理应与他们比肩的笛卡尔在那不久前已经辞世，而牛顿和莱布尼茨则要在很久以后才会登场。

004 “概率” 这个词汇

数学中，概率这个概念在1654年（参照上一条）前是不存在的。1654年之后，概率的概念也不是立刻就使用了“概率”（英语中的“probability”）这个词来表示。当时，用来表示概率的是类似“运气”和“机会”这样的词语。尤其是在“机会的游戏”（英语中的“game of chance”）中，使用的是“机会”这个词。用更为通俗的语言来表达“机会的游戏”的话，正是现在的“赌博”。帕斯卡和费马这两个当事人，并没有留下任何可以证明两人曾在数学中使用过“概率”这个词的证据。

数学含义中的“概率”这个词又是在什么时候初次出现在文献中

的呢？与帕斯卡交往密切的安托万·阿尔诺和皮埃尔·尼古拉在1662年出版的《伦理学》也就是《波尔·尼亚尔逻辑》的最后一章中，为“概率（probability）”这一词赋予了数学上的含义，这通常也被认为是在文献中的首次使用。

但是，这个词也并不是作为一个专业术语使用的，它的定义还不够明确。那之后，概率这个词也没有成为专业术语的倾向。似乎一直到了18世纪，“概率”这个词才明确地成为（古典）概率论的一种专业术语。这些事情我们在后面（054，p.113）会再次提及。

至少我们可以知道，1654年那会儿还没有我们现在所说的“概率”这门概念。也就是说，帕斯卡和费马不是单纯地开发出了“概率”的正确计算方法，而是创造了“概率”这个概念自身（至少是创造它的一大原动力）。

比如说，我们可以将此事与牛顿和莱布尼茨在微积分学上的创始进行比较。微积分学所研究的是求切线、求面积和求体积这样的问题，但这种问题本身是从很久以前就存在的，微积分学的创始可以说是对计算方法进行了历史性的完成。

而相对的，概率论中，以前并不存在求概率或求期待值这样的问题。可是现代社会中，概率和期待值这个概念已经成为生活的一部分，我们甚至很难想象没有这个概念的时代。

005 赌场必胜法

——无论哪个时代，人们开始关注概率论初步研究的原因，一定都是基于赌博。

吐德哈特《概率论史》

这是发生在美国某个大学校园里的事。一位学生走进了数学老师的办公室。

学生：老师，我遇上了一点麻烦。

老师：怎么啦？

学生：今天之内，我必须筹齐 1000 美元，但我手上只有 990 美元。这 1000 美元缺 1 分也不行，必须要正好 1000 美元。明天我老家就会寄钱过来，但我必须要在今天之内筹齐。

老师：10 美元的话，随便找个人借一下不就行了。

学生：不是这样的，我刚来这所大学没多久，还没有愿意借钱给我的朋友。所以我才来求助于您……

老师：但我可是概率论的教授哦，怎么能借钱给本校的学生呢？

老师的逻辑让学生有些摸不着头脑。

老师：啊，要不这样吧。和我赌一局牌吧，21 点也行，也算是学习了概率。

学生：呃，在学校里赌博难道不是更加糟糕的事吗？

老师：不不，当然是到我的公寓里去玩了。

学生：我觉得问题的关键不是在哪里玩……

老师：对了，那就去赌场吧！合法的赌场！我开车带你去。这附近有家赌场是专门为你们这样的穷学生开设的，轮盘赌的最小赌注可以是 1 美分。用你的 990 美元作为赌注，在那里赚 10 美元就行了。

学生：但是，赌场输钱的概率不是更高吗？

老师：你所说的与其说是概率，不如说是期待值吧。那是肯定的，赌场也是一门生意，规则就是为了让赌场赚钱而设定的。

学生：那不是希望渺茫吗？