The world is random

一部真正有趣的
概率统计入门读物

我宁要模糊的正确，也不要精确的错误。
——沃伦·巴菲特

世界是随机的

大数据时代的概率统计学

李　帅◎著

薛定谔的猫 与 庞加莱的称
科比的强悍 与 马刺的稳定
庄家的秘密 与 赌神的绝学
东野的推理 与 谷歌的预测

清华大学出版社

世界是随机的

大数据时代的概率统计学

李　帅◎著

清华大学出版社
北京

内 容 简 介

这是一本写给初学者的书，目的是帮助读者理解大数据下概率统计等概念的意义，写作中以案例作先导，引起读者的兴趣和思考，在解答问题的过程中讲述知识。

本书共有 9 章，第 1 章和第 2 章介绍概率和随机变量的基础知识；第 3 章和第 4 章介绍统计和分布的基础知识；第 5 章是专门介绍赌博中的概率统计的一章，前四章的知识在这里得到了应用；第 6、7、8 章分别介绍了概率统计的三个重要方法——假设检验、贝叶斯定理和线性回归；第 9 章是漫谈概率统计。本书努力避开说教式的言辞，把知识融入故事中，在讲解知识的同时，带给读者阅读的乐趣。是一本难得的适合所有对概率统计感兴趣或者学习有需求的读者阅读。希望本书可以帮助读者更快速、更深刻地理解和应用大数据。

图书在版编目(CIP)数据

世界是随机的：大数据时代的概率统计学/李帅著. —北京：清华大学出版社，2017
ISBN 978-7-302-46109-8

Ⅰ. ①世…　Ⅱ. ①李…　Ⅲ. ①概率统计　Ⅳ. ①O211

中国版本图书馆 CIP 数据核字(2016)第 313733 号

责任编辑：刘志彬
封面设计：汉风唐韵
责任校对：宋玉莲
责任印制：王静怡

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　**邮　　编**：100084
社 总 机：010-62770175　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62770175 转 4506
印 装 者：三河市金元印装有限公司
经　　销：全国新华书店
开　　本：170mm×240mm　**印　张**：13.25　**字　数**：214 千字
版　　次：2017 年 3 月第 1 版　**印　次**：2017 年 3 月第1次印刷
印　　数：1～4000
定　　价：39.00 元

产品编号：072671-01

前　言

凯文·凯利在《失控》中曾提道，当高度互联的低级群体的数量大到一定程度时，群体特征便会涌现出来，这特征是群体中的任何个体都不具备的。比如，大量水滴汇集成河水、海水，便会产生让水滴“感到陌生”的新特征——漩涡和波浪。

2013年8月，谷歌公司提出了一个票房预测模型，该模型仅以单词搜索量为依据，便可以提前一个月预测电影的首周票房，准确度高达94%。更令人惊讶的是，这是一个简单的线性回归模型。谷歌是如何做到的呢？

人类对数据的处理已经进入大数据时代。可是，绝大多数的人，对数据统计等基本常识还在算术常识时代。这是一个科技的时代，相对于十年前和二十年前，全球市值最大最受人尊敬的公司Top 10，全部变成了苹果、微软、Google……这些高科技公司，任何普通人都用智能手机，任何人都在享受高科技技术带来的便利。为了更好地工作和生活，我们要了解一下这些高科技技术的常识。笔者在这方面有一些经验，所以特地编写了本书，希望以比较科普和有趣的笔调，让你了解一门新的科学，甚至进入一个新的领域。

大学本科时，我曾上过“概率论”和“数理统计”两门课，

虽然完整地学习了概率统计，却只是一知半解。攻读硕士时，我在科研工作中需要用到概率统计，方才无奈地发现，当年所学已完完全全地还给了老师。我只能匆忙地自学了概率统计，勉强能应付科研工作，但心中对概率统计的很多概念仍旧一头雾水。后来，我有幸与我的妻子走到了一起，她大学本科和硕士期间都主修“应用数学”专业，在她的帮助下，我这个概率统计的门外汉终于入门了。

硕士毕业前，我和妻子共同翻译了一部英文科普读物《让你爱上数学的50个游戏》，这本书帮助我进一步巩固了概率统计知识，也让我萌生了写书的念头。毕业后我仍从事科研工作，参与了几个与数据分析有关的项目，发现自己对概率统计的理解仍然不够深刻。于是我一口气阅读了几本概率统计的科普书，比如《深入浅出数据分析》《深入浅出统计学》和《生活中的概率趣事》，终于搞懂了“贝叶斯定理”“假设检验”等概念。看书之余，我在“简书”上写了几篇读书心得。出版社的编辑看到我写的文章，问我是否愿意写一本概率统计的科普书，说实话，能写作一本属于自己的书是我的小小理想，既然机会来了，我怎么会拒绝呢?!

开始写作前，我为自己设定了三个原则。

一是理解而非定义。概率统计的教科书里充满了数学公式，虽然数学公式能对抽象的概念做出精确的定义，但这样的定义太晦涩，难以理解。这是一本写给初学者的书，我想帮助读者理解概念的含义，而非怎么求解某个具体问题。所以，我会用解释性的语言来描述概念，而不是给出标准的定义。这么做风险很大，但我愿意尝试，希望本书可以帮助读者更快速、更深刻地理解概念。

二是引导而非灌输。从小到大，我们都承受了太多的灌输式教育，我很庆幸，自己在灌输式教育下活了下来，但我不希望“灌输”给读者任何东西。所以，我总是以案例作先导，先引起读者的兴趣和思考，然后在解答问题的过程中讲述知识。希望这么做可以为读者减负，让读者更流畅的阅读，在轻松愉快中学到知识。

三是有趣而非无趣。很多人说，“有趣”是对一个人最高的评价。我觉得，对一本书同样如此。图书销售排行榜上，小说永远是主角，因为它们“有趣”。读者喜欢故事，不喜欢说教，这是事实，更是真理。我要努力避开说教式的言辞，把知识融入故事中，在讲解知识的同时，带给读者阅读的乐趣。

写作时，我尽量坚持这三个原则，虽然期间有过挣扎和迷茫，但最终还是完成了这本书。

本书共有9章，第1章和第2章介绍概率和随机变量的基础知识；第3章和第4章介绍统计和分布的基础知识；第5章是专门介绍赌博中的概率统计的一章，前4章的知识在这里得到了应用；第6、7、8章分别介绍了概率统计的三个重要方法——假设检验、贝叶斯定理和线性回归；第9章是漫谈概率统计。

我的阅读建议是：第1、2章合并阅读，第3、4章合并阅读，在前4章阅读完成后，再阅读第5、6、7、8、9章，后5章各自独立，不需要按顺序阅读。

本书由李帅主笔编写，同时参与编写的还有黄维、金宝花、李阳、程斌、胡亚丽、焦帅伟、马新原、能永霞、王雅琼、于健、周洋、谢国瑞、朱珊珊、李亚杰、王小龙、张彦梅、李楠、黄丹华、夏军芳、武浩然、武晓兰、张宇微、毛春艳、张敏敏、吕梦琪等作者。在此一并感谢。

这是我的第一本书，其中难免出现错误，希望读者理解包涵，也欢迎读者批评指正。

如果你读过本书，想与我沟通，欢迎通过E-mail联系我：lishuaibeijing@163.com。

最后，我要感谢我的家人和朋友。感谢我的父母，陪伴我成长，帮助我养成了读书和写作的习惯。感谢我的妻子，一直理解我、陪伴我，并给我讲解了一些晦涩的数学概念。感谢刘子冲、王充山、秦培根、刘翼、孙淼、赵玮琪等老朋友，你们的支持和鼓励是我坚持写作的动力！

编　者

目　录

第 1 章

概　率

导语：我们生活的世界，是确定的还是不确定的？自古至今，人们一直试图回答这个哲学命题。一方面我们确信，苹果熟透后会从树上掉下来；另一方面我们又无法确信，抛起的硬币落到地上时，哪一面会朝上。

1.1 生还是死：这是一个概率问题

2012 年 7 月 21 日，北京大雨倾盆，事后这一天被称为“北京 7 · 21 特大暴雨”。下午两点，我接到父亲的电话，要我赶快回东北老家。家中病危的爷爷快挺不住了。

我抓起外套出了门，冒着大雨疯狂地跑进地铁，奔向北京站。

第二天傍晚五点半，我下了火车，直奔医院。病床前，我看到瘦骨嶙峋的爷爷蜷缩在那里，已经没了意识，奄奄一息。八点整，爷爷血压骤降，医生对父亲点了点头，时辰到了。我终究没能和爷爷说上最后一句话。

后来，我常会梦到爷爷。在梦中，爷爷坐在青绿色的老式沙发上，戴着折叠

式老花镜，饶有兴致地看《城市晚报》。我似乎记得爷爷已经去世了，但又分明看到爷爷就坐在那里。那一刻，梦中的那一刻，我真的分不清爷爷是生还是死。

生死与有无、对错一样，都是鲜明对立的东西，它们看似是两条平行的直线，永不相交。然而，梦中的我却分不清爷爷是生还是死。生与死真的永无相交的可能吗？

鹰溪桥上的法克尔

下面是美国小说家安布鲁斯·布尔斯的小说《鹰溪桥上》的片段节选，故事发生在美国南北战争期间，讲述的是农场主法克尔被处以绞刑的故事。

亚拉巴马州北部的铁路桥上，一个男人站在那里，俯视着桥下二十米处那湍急的流水。这人的双手被人用绳子绑在身后，一根绳索紧紧地套在他的颈部，绳索的另一端被系在他头顶上方交叉着的架子上，一段绳子松松垮垮地垂在他的膝盖处。铁轨枕木上铺着几块木板，他和要对他行刑的一名中士和两名列兵就站在上面。

那个即将被施以绞刑的男人看起来大约 35 岁，一副平民的装扮。如果从他的举止行为来看，他像是一位庄园的农场主。他五官端正——鼻子高挺，嘴唇坚毅，额头饱满，长长的黑发顺直地披在脑后，他的眼睛大而乌黑，面目和善，人们很难想象到这人即将被施以绞刑而死。

他索性睁开了眼睛，看到了他身下的流水。“如果我能把双手挣脱，”他心里这样想着，“我就能摆脱颈上的绳索，跳到河里去，然后潜到水下躲避那些子弹，拼命地游到河岸边，钻进那里的森林，就能跑回家了。谢天谢地，我家不在他们的封锁线里，我的妻子和孩子们离他们的先头部队还有些距离。”正当这些想法在犯人脑中闪过时，上尉对中士点头示意。中士从那块木板上跨到了一边。

当法克尔从桥上径直地向下坠落时，他已经没有了意识，就像是死了一样。仿佛过了很久，颈部剧烈地挤压所带来的疼痛使他从这种状态中清醒了过来，接着就感到了窒息。他知道那条绳索已经断了，他坠入了河中，那种窒

息的感觉没有加剧。他在黑暗中睁开了眼睛，看到了他上方的一道亮光。他的两只手快速的向下拍水，使身体上浮，他感觉自己的脑袋已经浮出了水面，炫目的阳光使得他睁不开眼睛。他看到了那座桥，以及给他施以绞刑的执行者，他们正大喊着用手指向这边，子弹射到水里，离他的头只有几英寸的距离，溅起的水花打在他的脸上。

法克尔猛地向水下潜去，尽量钻到水的深处。法克尔在湍急的流水中奋力地划水，他思维清晰，四肢越发有力，心里想着："上帝保佑我，保佑我能躲过所有的子弹！"

突然，他感觉自己开始一圈圈地旋转起来，像陀螺一样。水面、河岸、树林，已经离得很远的桥，还有那军事堡垒和那些士兵，都搅到了一起，变得模糊不清。水中的一处漩涡将他卷了起来，没过一会儿，他就被水流抛到了左岸边的一堆砾石上。他喜极而泣，两手抓起泥沙，一把把的往上扬，落到自己身上，喃喃地说着一些祝福的词句。他跃身而起，迅速地往坡上的岸边跑去，钻进了那片树林。

那一天，他都依照着太阳往前走，那片树林太过茂密，像是永无尽头，他到处都找不到一个可以休息的地方，甚至都找不到一条樵夫走过的小道。夜幕降临时，他已经走得精疲力竭，可是一想到他的妻子和孩子们，他又竭力地继续向前走。最后，他终于找到了一条通往他家的路。那条路像城市里的街道那样笔直而宽阔，可却像是无人从此处通行过，路的两边没有田野，也没有房屋。他的眼睛有些肿胀，没法闭眼，口中干渴，舌头也发胀起来，他把舌头伸出口外去接触空气，感受丝丝的凉意。这条没人走过的路上全是草，这些草多么柔软，软得让他没法儿感觉到脚下的路！

他站在自己家门口，所有的一切都和他离开时一模一样。当他推开门，他看到了女人的衣裙在飘动；他的妻子还是那么的清新甜美，正从门廊中走出来迎接他。她走下了台阶，脸上带着不可言喻的笑容，那种气质简直无与伦比！啊，她是多么的美丽！他伸开双臂冲过去……

——节选自《鹰溪桥上》

读到这里，我们的心中难免会有一个疑问：法克尔究竟是死了还是逃跑了？

读到法克尔掉入水中，拼命挣扎着爬上岸时，我们相信法克尔真的逃脱了。可是，怪异的树林、无人走过的路、无法感觉脚下的路，又让人心生怀疑：难道这些是法克尔的幻觉？我们希望法克尔成功逃脱，回到家中与妻子团圆，又担心一切都是法克尔的幻觉。法克尔在我们心中仿佛是一个既可能“生”又可能“死”的人！

薛定谔的猫

要测试你是否真的了解“量子物理”，只需要问你两个问题。

第一个问题：你知道“薛定谔的猫”吗？

（我猜你会点头。）

第二个问题：你知道哥本哈根学派吗？

（别皱眉了，赶快承认不知道吧。）

大多数人都知道这只著名的猫，却不知道这只猫到底是怎么来的，没错，这只猫与哥本哈根学派有莫大的关系。

哥本哈根学派于20世纪20年代初期建立，对量子物理的创立和发展做出了很多重要贡献。学派的创始人是著名量子物理学家玻尔，主要成员包括玻恩、海森堡等知名物理学家。薛定谔也是量子物理学界的鼻祖，他提出的“薛定谔方程”为量子力学奠定了坚实的基础，至今折磨着一代又一代的大学工科男。不过，薛定谔并不是哥本哈根学派的成员，这是因为他对哥本哈根学派的理论存在质疑。为了有的放矢地提出自己的质疑，他脑洞大开地想到了一个实验——“薛定谔的猫”。

“薛定谔的猫”是一个思想实验，实验的过程是，把一只可怜的雌性小猫关在一个密室里，密室里有食物也有毒药，毒药装在瓶子里，瓶子上有一个锤子，锤子由一个电子开关控制，如果电子开关被触动，锤子就会落下，砸碎瓶子，瓶子里的有毒氰化物会毒死小猫。问题是：小猫到底是活着还是死了？

实验的关键在于，电子开关是否被触动是一个随机发生的事件，发生的概率是50%。这里的50%不是“抛硬币50%出现正面”这么简单，要产生真正的随机事件，需要使用放射性元素。在微观世界里，放射性元素的衰变是宇宙都无法预知的随机事件，一个真正的有50%概率发生的随机事件。控制电子开

关的正是放射性元素，如果放射性元素发生衰变，则开关被触动，锤子砸碎毒瓶，小猫必死。

这个问题要分两种情况讨论。

情况一：我们打开密室观察，可以确切地知道小猫是生还是死。如果放射性元素发生了衰变，那么可怜的小猫一定已经中毒身亡；如果没发生衰变，那么可爱的小猫依然活着。

情况二：我们不打开密室，由于放射性元素的衰变完全无法预测，所以小猫既可能生，也可能死，我们只能认为小猫处于“生与死”的叠加状态！

用量子物理的语言来说，当我们没有观察小猫时，小猫是被“概率云”包裹的，生与死两种状态互相叠加，形成了一个“叠加态”，当我们进入密室观察小猫时，“概率云”瞬间塌缩了，于是我们只能观察到某一种状态的小猫。

一只“既生又死”的猫？这明显违背常识。薛定谔把微观世界的叠加状态平行的移植到宏观世界中，以此质疑量子物理的“完备性”，也就是说，量子物理中的“叠加态”在宏观世界中不成立。

量子物理学家玻尔曾说：“谁要是第一次听到量子理论时没有感到困惑，那他一定没听懂。”亲爱的读者朋友，你是听懂了还是没听懂呢？

我们活在当下，感知当下，环顾四周，仿佛一切都是确定无疑的。可是，此时此刻，还有很多人、很多事是你感知不到的，对你而言，它们是“不确定的”。鹰溪桥上的法克尔和薛定谔的猫到底是生还是死？这不再是一个非此即彼的问题，在谜底揭开之前，它们既可能生，也可能死，这是一个概率问题，专门研究概率问题的学科就是——概率论。

最后，我要公布《鹰溪桥上》的结局了。

他伸开双臂冲过去，正要和那美丽的女人拥抱时，他感觉到自己的颈后遭到了重重的一击，随着一声大炮的轰鸣，他的四周亮起了炫目的白光——接着，一切都陷入了黑暗和静寂。

法克尔死了，他那折断了颈部的尸体正悬在鹰溪桥后面的横木下轻轻地摆动。

——节选自《鹰溪桥上》

1.2 随机事件：翻飞的硬币

我的家乡邻近长白山，那一年，我终于登上了长白山，见到了传说中的天池。站在山顶向下望，天池宛若一面蓝色的魔镜，静如止水，莫过如此。上山之前，很多人说，想看到天池要靠运气，没多一会儿，我就明白了此言不虚。刚刚还晴空万里、阳光普照，转瞬间就是大雾弥漫，我和父亲母亲只能手拉着手站在原地，生怕在白茫茫的雾气中走失。再过一会儿，雾气缓缓消散，正当大家拿出相机要继续拍照时，乌云袭来，风雨大作，我们纷纷披上雨衣，站在寒风中瑟瑟发抖。那是我第一次感到大自然的风云变幻。

自古至今，人们都在试图回答一个哲学命题：我们生活在一个确定的世界还是不确定的世界？我们很确信，苹果熟透了，会从树上掉下来，但我们又不能确定，抛起的硬币落到地上时，哪一面会朝上。对此，哲学领域有两种不同的论断。

决定论：它是指自然界和人类社会普遍存在着客观规律和必然的因果联系，也就是说，如果我们能够发现和理解所有的客观规律和因果联系，自然界和人类社会的任何变化都是可以预知的，我们之所以还做不到，是因为我们对客观规律的认识还不够。

非决定论：与决定论相对，非决定论否认自然界和人类社会普遍存在着客观规律和必然的因果联系，认为事物的发展变化是没有客观规律的，是由事物内在的“自由意志”决定的，也就是说，人们可以自由支配自己的行为，却无法预言客观事物的发展变化和其他人的行为。

我们似乎更容易认同非决定论，毕竟世界如此纷繁复杂，我们只能控制自己，很难预知未来。但我们不能轻易否定决定论，抛开两个论断的对错之争，决定论为我们认识世界提供了新的思路。下面，我们就来做一个“抛硬币”的思想实验。

思想实验：抛硬币

抛硬币是大家十分熟悉的小把戏，足球比赛前，裁判会用抛硬币的方式让